



Programa Interdisciplinar de Pós-Graduação em  
**Computação Aplicada**  
Mestrado Acadêmico

Thyago de Melo Tosin

Um Modelo para Integração de Informações de Bases de  
Dados Abertos, com uso de Ontologia

São Leopoldo, 2016



Thyago de Melo Tosin

UM MODELO PARA INTEGRAÇÃO DE INFORMAÇÕES DE BASES DE DADOS ABERTOS, COM USO DE ONTOLOGIAS

Dissertação apresentada como requisito parcial para a obtenção do título de Mestre pelo Programa Interdisciplinar de Pós-Graduação em Computação Aplicada da Universidade do Vale do Rio dos Sinos – UNISINOS.

Orientador: Prof. Dr. Sandro José Rigo

Co-orientador: Prof. Dr. Jorge Luis Victoria Barbosa

São Leopoldo

2016

## FICHA CATALOGRÁFICA

T714m Tosin, Thyago de Melo  
Um modelo para integração de informações de bases de dados abertos, com uso de ontologias / Thyago de Melo Tosin. – 2016.  
107 f. : il.

Dissertação (mestrado) – Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Computação Aplicada, São Leopoldo, RS, 2016.  
“Orientador: Prof. Dr. Sandro José Rigo”

1. Computação aplicada. 2. Web semântica. 3. Bases de dados abertos. 4. Governo eletrônico I. Título.

CDU 004.65

Catlogação na Fonte:

Mariana Dornelles Vargas – CRB 10/2145

Thyago de Melo Tosin

Um Modelo para Integração de Informações de Bases de Dados Abertos, com uso de Ontologias

Dissertação apresentada à Universidade do Vale do Rio dos Sinos – Unisinos, como requisito parcial para obtenção do título de Mestre em Computação Aplicada.

Aprovado em 26/02/2016

BANCA EXAMINADORA

---

Prof. Dr. Jorge Luis Victória Barbosa – Unisinos

---

Prof.<sup>a</sup> Dr.<sup>a</sup> Marta Becker Villamil – Unisinos

---

Prof. Dr. Sílvio César Cazella – UFCSPA

Prof. Dr. Sandro José Rigo (Orientador)

Visto e permitida a impressão  
São Leopoldo,

Prof. Dr. Sandro José Rigo  
Coordenador PPG em Computação Aplicada



## **AGRADECIMENTOS**

Primeiramente, devo agradecer aos meus pais, que me deram todo o suporte e educação para que eu pudesse chegar onde estou agora.

Ao meu orientador, Prof. Dr. Sandro José Rigo e ao co-orientador Prof. Dr. Jorge Luis Victoria Barbosa, pela disponibilidade, ensinamentos, orientação, incentivo e apoio neste trabalho.

Obrigado!





## RESUMO

Com a lei de Acesso à Informação (Lei 12527/2011), espera-se que nas esferas federais, estaduais e municipais estejam garantidas e facilitadas as atividades de acesso aos dados de interesse para o cidadão. As bases interligadas de dados abertos facilitam a aquisição desses dados, possibilitando que diversas aplicações sejam criadas e que consultas sejam realizadas.

Entretanto observa-se uma carência de recursos para realizar o relacionamento das informações originadas em bases de dados abertas distintas. A integração de diferentes conjuntos de dados possibilita a criação de aplicações mais ricas e relevantes. A representação formal das relações entre os dados consultados permite o uso de mecanismos de inferência e mecanismos de consulta aos dados abertos e conectados. Este trabalho apresenta o desenvolvimento de recursos para inferir e relacionar tais informações no contexto de aplicações Web voltadas para a integração de bases de dados abertos e conectados.

Um modelo foi desenvolvido como contribuição e um protótipo foi implementado como caso de uso. Neste protótipo foram utilizados os dados das compras governamentais que estavam armazenados em bases de dados relacionais, com o uso de uma ontologia, desenvolvida especificamente para este caso, os dados foram mapeados e importados para um *triplestore* Apache Fuseki em formato RDF, uma aplicação em Java EE com o uso do *framework* Apache Jena foi desenvolvida para visualização dos dados através de métodos de consulta utilizando SPARQL. Foram aplicadas três avaliações nesse protótipo: baseada em cenário, de usabilidade e da ontologia. Como resultados, verificou-se que o modelo implementado proporcionou a integração desejada e auxiliou os usuários a obterem uma melhor experiência na visualização dos dados interligados das compras governamentais com o orçamento federal.

Palavras-chaves: Web Semântica, Dados Abertos e Conectados, Governo Eletrônico.



## **ABSTRACT**

With the Law of Access to Information and sanctioned in force, access to information, at least at the federal, state and local levels, make it easy for citizens. The bases linked open data, facilitate the acquisition of these data, however there is a lack of resources for inference of relationship information. This work aims at the development of these resources to infer and relate this information in the context of Web applications integrated with open and connected databases.

The prototype implemented used the data from government purchases that was stored in relational databases, using an ontology, developed specifically for this case, the data have been mapped and imported into a triplestore Apache Fuseki in RDF format, a Java EE application using Apache Jena framework, was developed to display data through methods that consulted using SPARQL. Three evaluations were applied in this prototype: scenario-based, usability and ontology. As result, it was found that the model implemented helped people to have a better experience when viewing linked data of government purchases with the federal budget.

**Keywords:** Semantic Web, Open Linked Data, Electronic Government.



## LISTA DE TABELAS

Tabela 1 - Princípios Dados Governamentais Abertos.....	28
Tabela 2 - Principais artigos da LAI e seu tema.....	31
Tabela 3 - Aspectos dos Artigos Relacionados.....	58
Tabela 4 - Quantidades de <i>datasets</i> por Portal.....	59
Tabela 5 - Relação de Classes e seus Significados.....	70
Tabela 6 - Propriedades de Objetos.....	72
Tabela 7 - Resultado da Pesquisa Item de Despesa .....	79
Tabela 8 - Resultado Avaliação Facilidade de Uso .....	91
Tabela 9 - Avaliação Percepção de Utilidade.....	92
Tabela 10 - Principais Problemas.....	95



## LISTA DE FIGURAS

Figura 1 – Camadas da Web Semântica.....	33
Figura 2 - Grafo RDF.....	35
Figura 3 - Grafo do exemplo RDF .....	36
Figura 4 - Código XML / RDF .....	36
Figura 5 – Tripla do Modelo de Dados .....	37
Figura 6 - Exemplo Consulta SPARQL na DBPedia .....	39
Figura 7 - Resposta do exemplo em SPARQL.....	39
Figura 8 - Arquitetura do <i>Framework</i> Jena.....	42
Figura 9 - Ontologia "geopoliticabr" .....	44
Figura 10 - Ontologia de Domínio de Avaliação .....	46
Figura 11 - Página principal da aplicação .....	46
Figura 12 - Página com detalhes do curso.....	47
Figura 13 - Visão Geral da Abordagem <i>RDB2LOD</i> .....	48
Figura 14 - Tela da aplicação desenvolvida .....	49
Figura 15 - Comparação entre visualizações de dados .....	49
Figura 16 - Ontologia Organizacional.....	50
Figura 17 - Etapas do Mapeamento .....	51
Figura 18 - Página inicial do Portal do Reino Unido .....	52
Figura 19 - Página inicial do Portal de dados abertos Americanos .....	53
Figura 20 - Dados disponíveis no Portal da Transparência.....	54
Figura 21 - Página Inicial DataPOA.....	55
Figura 22 - Página Inicial do portal de Dados Abertos Federal .....	56
Figura 23 - Modelo Genérico.....	62
Figura 24 - Ontologia de Compras desenvolvida .....	70
Figura 25 - Modelo Sistêmico Proposto .....	73
Figura 26 - Modelo de Dados Compras Governamentais .....	75
Figura 27 - Dados do Órgão Instituto Federal de Roraima.....	76
Figura 28 - Possibilidades de Uso do RDF.....	78
Figura 29 - Listagem de Itens de Despesa.....	79
Figura 30 - Ferramenta de Visualização do Orçamento.....	80
Figura 31 – Trecho do Mapeamento de Dados.....	81
Figura 32 - Trecho do arquivo RDF .....	82

Figura 33 - Listagem de Órgãos.....	83
Figura 34 - Listagem de Licitações de um Órgão.....	83
Figura 35 - Relação de compras de uma licitação .....	84
Figura 36 - Dados da compra interligada com Orçamento Federal.....	85
Figura 37 - Consulta de Compras de uma Licitação .....	88
Figura 38 - Somatório do valor das compras de uma licitação.....	88
Figura 39 - Detalhes da Compra com Orçamento.....	89
Figura 40 - Resumo da Avaliação de usabilidade .....	92
Figura 41 - Arquitetura OOPS! .....	94
Figura 42 - Avaliação da Ontologia .....	96



## LISTA DE SIGLAS

XML	eXtensible Markup Language
HTTP	Hypertext Transfer Protocol
RDF	Resource Description Framework
W3C	World Wide Web Consortium
JSON	JavaScript Object Notation
TAM	Technical Architecture Modeling
SPARQL	SPARQL Protocol and RDF Query Language
URL	Uniform Resource Locator
URI	Uniform Resource Identifier
J2EE	Java 2 Platform, Enterprise Edition
API	Application Programming Interface
CIGA	Comitê Interministerial do Governo Aberto



## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>21</b>
1.1	Contexto e motivação	21
1.2	Questão de pesquisa	24
1.3	Objetivos	24
1.4	Metodologia	24
1.5	Organização do texto	25
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	<b>27</b>
2.1	Dados Abertos e Conectados	27
2.1.1	As três leis de dados governamentais abertos	28
2.1.2	Oito Princípios dos Dados Governamentais Abertos	28
2.2	Governo Aberto	29
2.3	Governo Aberto Brasileiro	30
2.4	Web Semântica	32
2.4.1	RDF	34
2.4.2	URI	37
2.4.3	SPARQL	38
2.4.4	ONTOLOGIAS	40
2.4.5	Ferramentas de apoio	41
<b>3</b>	<b>TRABALHOS RELACIONADOS</b>	<b>43</b>
3.1	Aplicação de Inferência Semântica para Exploração Conceitual de Bases de Dados Relacionais	43
3.2	Aplicando Linked Data na Publicação de Dados do ENEM	44
3.3	Avaliação do ensino superior público no Brasil: protótipo de aplicação <i>linked data</i>	45
3.4	Uma Abordagem para a Publicação de Dados Ligados Obtidos a Partir de Bases de Dados Relacionais	47
3.5	Mapeando Dados Governamentais com uma Ontologia de Organizações	50
3.6	Portais Governamentais	51

3.6.1	Reino Unido .....	52
3.6.2	Estados Unidos da América .....	52
3.6.3	Portal da Transparência Federal .....	53
3.6.4	DataPOA .....	54
3.6.5	Dados Abertos Federal .....	55
3.7	<b>Análise dos Trabalhos Selecionados .....</b>	<b>56</b>
4	<b>MODELO PROPOSTO .....</b>	<b>61</b>
4.1	<b>Visão geral do modelo .....</b>	<b>61</b>
4.2	<b>Origem de dados .....</b>	<b>63</b>
4.3	<b>Acesso aos Dados .....</b>	<b>63</b>
4.4	<b>Mapeamento de Dados .....</b>	<b>64</b>
4.5	<b>Ontologia .....</b>	<b>65</b>
4.6	<b>Notação Semântica .....</b>	<b>65</b>
4.7	<b>API de Consulta .....</b>	<b>66</b>
4.8	<b>Análise do Modelo .....</b>	<b>66</b>
5	<b>PROTÓTIPO IMPLEMENTADO .....</b>	<b>69</b>
5.1	<b>Ontologia de domínio desenvolvida .....</b>	<b>69</b>
5.2	<b>Detalhamento do modelo neste protótipo .....</b>	<b>73</b>
5.3	<b>API Compras Governamentais .....</b>	<b>74</b>
5.4	<b>API Dados Orçamentários .....</b>	<b>77</b>
5.5	<b>Notação RDF .....</b>	<b>80</b>
5.6	<b>Aplicação de consulta desenvolvida .....</b>	<b>82</b>
6	<b>AVALIAÇÃO .....</b>	<b>87</b>
6.1	<b>Avaliação por Cenários .....</b>	<b>87</b>
6.2	<b>Avaliação da Usabilidade .....</b>	<b>90</b>
6.3	<b>Avaliação da Ontologia .....</b>	<b>93</b>
7	<b>CONCLUSÕES .....</b>	<b>99</b>
8	<b>BIBLIOGRAFIA .....</b>	<b>103</b>
	<b>APÊNDICE A – QUESTIONÁRIO PARA A AVALIAÇÃO DA USABILIDADE .....</b>	<b>107</b>

## 1 INTRODUÇÃO

Este trabalho investiga possibilidades de atuação na integração de dados oriundos do conjunto crescente de dados abertos e conectados. Estes dados são disponibilizados a partir da atuação de governos e entidades, sendo divulgados com a finalidade de proporcionarem a geração de novas aplicações. Apesar de avanços promissores neste sentido, percebe-se a necessidade de investigação para fomentar a implementação de aplicações mais abrangentes (ISOTANI, BITTENCOURT, 2015).

A Lei de Acesso à Informação (Lei nº 12.527/2011) que regula o acesso às informações mantidas pelos governos nas esferas federal, estadual e municipal foi sancionada em novembro de 2011. Assim que entrou em vigor, uma enorme quantidade de dados públicos passou a ser disponibilizada na internet. Porém não existe ainda um padrão amplamente adotado para uso nos formatos de dados, bem como observa-se que os dados originados destas fontes não estão integrados com outras bases relacionadas, o que pode causar dificuldade em sua utilização mais ampla.

### 1.1 Contexto e motivação

No Brasil o governo federal tem duas grandes iniciativas para atender a necessidade de tornar público e acessível um volume cada vez maior de informação sobre a gestão e instituições públicas, que são o Portal de Dados Abertos<sup>1</sup> e o Portal da Transparência<sup>2</sup>. Porém os dados disponibilizados nestas iniciativas possuem formatos diferenciados e se originam de bases não interligadas, o que torna difícil e limitada a sua utilização. Atualmente um grande volume dos dados publicados é disponibilizado através de arquivos em formatos como PDF ou em planilhas de cálculo, dificultando sua leitura, integração e análise. Em especial as atividades de análises e integração automatizadas, realizadas por aplicações de software, são prejudicadas ou inviabilizadas neste contexto. Segundo (ANTHES, 2015), o grande desafio no desenvolvimento de serviços eletrônicos do governo é mantê-los de forma descentralizada como medida de segurança, e ao mesmo tempo preservar a compatibilidade entre aplicações e compartilhamento de dados.

---

<sup>1</sup> <http://dados.gov.br>

<sup>2</sup> <http://portaltransparencia.gov.br>

Para modificar essa situação várias empresas, governos e institutos de pesquisa têm realizado esforços para a disponibilização de dados e a produção de tecnologias web que permitam criar um ecossistema de produção e consumo de dados com o objetivo de acelerar a descoberta de novos conhecimentos e incorporar valor a qualquer informação disponibilizada livremente na Internet. (ISOTANI, BITTENCOURT, 2015).

Atualmente a internet é o principal meio de propagar informações em formato digital, principalmente com o avanço para web 2.0, onde é possível qualquer indivíduo publicar dados em *blogs* e redes sociais. Entretanto, por ser um tanto recente, estes dados possuem pouca ou nenhuma interligação, muitos deles encontram-se em bases isoladas. Como um exemplo adicional, deve ser considerada a grande quantidade de dados governamentais disponíveis na internet, gerados a partir das iniciativas do e-gov<sup>3</sup> e *open-government*<sup>4</sup>, que defendem a divulgação de dados aos cidadãos. Entretanto o consumo e reutilização destes dados ainda é limitada, tendo em vista que suas interfaces de divulgação são voltadas principalmente para extração de dados.

Os grupos de trabalho do W3C voltados para o tema de *Linked Open Data*<sup>5</sup> (LOD) propõem o uso de padrões abertos, suportados pelo W3C, para divulgação de dados na internet por meio de princípios simples, incluindo uma padronização da semântica por trás dos dados. Dados Abertos Conectados usam tecnologias da Web Semântica para tornar públicos os dados estruturados na web e criar ligações entre dados de origens adversas (HEATH; BIZER, 2011).

O anseio por facilidades de acesso e utilização efetiva das bases de dados abertas e ligadas disponíveis na web apresenta-se como a origem de uma demanda por ferramentas que favoreçam estes aspectos. Considera-se que a utilização efetiva destes dados depende da capacidade de acesso e relacionamento entre os mesmos. Atualmente existe um número reduzido de iniciativas com este objetivo. (ISOTANI, BITTENCOURT, 2015).

Com base neste tipo de dificuldade na integração automatizada dos dados, observado em diversos países, o W3C (*World Wide Web Consortium*) tem reco-

---

<sup>3</sup> <http://governoeletronico.gov.br/>

<sup>4</sup> <http://www.opengovpartnership.org>

<sup>5</sup> <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

mendado um conjunto de boas práticas para publicação de dados de uma forma estruturada e interligada. (FONSECA, LUCAS B.R., AZEVEDO, CARLOS L. B., ALMEIDA, 2014).

Atualmente a Web disponibiliza uma quantidade enorme de informações, porém existe uma grande dificuldade na sua utilização pelos usuários, devido à sobrecarga cognitiva que decorre do processo de leitura e relacionamento de informações. Por outro lado, as bases de dados abertos e conectados vêm se constituindo em uma alternativa concreta para o acesso automatizado às informações de diversas áreas, dado que adotam padrões de anotação baseados nos conceitos da web semântica. As aplicações de Web Semântica constituem uma alternativa para o uso automatizado destas bases de dados ligadas, porém esta área ainda carece de desenvolvimento em diversos aspectos, sendo um deles a disponibilidade de mecanismos para integrar informações originadas em bases diversas.

Um conjunto de conceitos representado por um modelo de dados dentro de um domínio e seus relacionamentos é tido como uma ontologia. As ontologias têm sido peça fundamental para auxiliar no estabelecimento da Web Semântica. (GERBER; BARNARD; VAN DER MERWE, 2007).

Para auxiliar no desenvolvimento de aplicações para web semântica, pode-se contar com poucos frameworks já conhecidos e amplamente utilizados, sendo que dentre eles destaca-se o viejs<sup>6</sup> e o Apache Jena<sup>7</sup>, sendo este último mais robusto e composto por diferentes APIs que interagem em conjunto para processar os dados RDF.

Dada a carência de ferramentas e frameworks para apoiar o desenvolvimento de aplicações voltadas para a web semântica, sendo que estas podem apoiar aplicações em áreas relevantes, como a Educação ou Saúde, este trabalho busca estudar e definir recursos para a integração de informações originadas de distintas bases de dados conectados.

---

<sup>6</sup> <http://viejs.org>

<sup>7</sup> <https://jena.apache.org>

## 1.2 Questão de pesquisa

A partir do contexto destacado, foi definida como questão de pesquisa para este trabalho: como utilizar bases de dados abertas e ontologia de domínio para permitir a integração de informações?

## 1.3 Objetivos

O presente trabalho tem por objetivo especificar, implementar e avaliar uma solução para interligar bases de dados isoladas e disponibilizar estes dados de forma aberta e interligada, por meio de um ambiente de exploração de dados abertos e conectados, ou seja, convertidos em formato RDF, com base em ontologias de domínio para uso comum nos temas tratados pelas bases de dados abertas selecionadas.

O objetivo geral deste trabalho consiste em conceber um modelo para integração de dados governamentais em formato aberto, com base em ontologias de domínio.

Os seguintes objetivos específicos foram estabelecidos para o cumprimento do objetivo geral proposto.

- a. Propor uma metodologia para o uso de ontologias de domínio que interliguem dados relacionados, para exploração de dados e possível cruzamento de informações.
- b. Implementar um modelo para consumo de dados oriundos de bases relacionais.
- c. Inserir notação semântica RDF nos dados.
- d. Desenvolver uma API para visualização dos dados abertos e conectados armazenados no *triplestore*.

## 1.4 Metodologia

A metodologia adotada para o desenvolvimento deste trabalho consiste na realização de pesquisa para obter informações acerca do tema definido anteriormente, em seguida especificar a solução a ser desenvolvida, nesta etapa foram definidas as tecnologias que seriam utilizadas para o desenvolvimento da solução,



após sua definição, iniciar o desenvolvimento e por último sua validação por usuários e desenvolvedores.

Segundo (KÖCHE, 2011), uma pesquisa de caráter exploratório busca descrever em detalhes o cenário e o problema tratados, demonstrando a oportunidade de seu tratamento a partir de uma proposta com base no estudo de um caso específico.

Para atingir os objetivos propostos e responder à questão de pesquisa, o método de trabalho definido por (WAZLAWICK, 2009) composto das seguintes etapas, foi adotado: a) Realização de revisão bibliográfica com vistas à ampliação da experiência nas áreas de conhecimento envolvidas; b) Estudo de trabalhos e aplicações relacionadas; c) Definição do detalhamento do modelo; d) Implementação de um protótipo para avaliação; e) Realização de estudos de caso; f) Avaliação de resultados.

Pretende-se demonstrar com este estudo o quanto a abordagem proposta favorece o desenvolvimento e avaliação de alternativas para propor um modelo para a integração flexível das informações de bases de dados abertos e conectados, com uso de ontologias de domínio.

### **1.5 Organização do texto**

O texto deste trabalho está organizado como segue. No capítulo 2 é apresentado o referencial teórico. A seguir, no capítulo 3, são citados alguns trabalhos relevantes na área de web semântica, dados abertos e conectados, portais governamentais de divulgação de dados. O capítulo 4 apresenta o modelo desenvolvido. Um protótipo implementado é descrito detalhadamente no capítulo 5 e aspectos de sua avaliação são descritos no capítulo 6. Por fim, no capítulo 7 são expostas as conclusões e indicações de trabalhos futuros.



## 2 REFERENCIAL TEÓRICO

Neste capítulo são apresentados aspectos do uso de dados abertos e conectados. Também são apresentadas informações sobre o uso de ontologias com base em padrões do W3C e outras linguagens compatíveis com o proposto neste trabalho, como por exemplo, RDF (*Resource Description Framework*), JSON-LD (*JavaScript Object Notation for Linking Data*) e SPARQL (*SPARQL Protocol and RDF Query Language*).

### 2.1 Dados Abertos e Conectados

Segundo (HEATH; BIZER, 2011), há um desejo, em constante crescimento, por acesso aos dados não disponíveis atualmente na web, ou não acessíveis através de *links* em páginas HTML desde que a web se tornou mais presente no cotidiano dos usuários. Definiu-se então como meta desejável a web de dados, que seria a extensão da web com um espaço global de dados baseados em padrões abertos.

Por definição, os dados abertos conectados se baseiam em um conjunto de regras para publicação de dados na web, com a finalidade de unificar e facilitar o acesso. As regras propostas (HEATH; BIZER, 2011) para esta finalidade são: (1) Usar URI como nome para as coisas; (2) Usar URI HTTP para que os cidadãos possam procurar por esses nomes; (3) Disponibilizar informações úteis na recuperação das URIs, usando padrões como RDF e SPARQL; (4) Incluir links para outras URIs, para tornar possível descobrir mais informações. Para assegurar diferenciação entre os dados publicados, concedendo integração desses dados com dados de outras fontes, adotam-se ontologias. As Ontologias mais populares na web possuem um conjunto de regras e uma taxonomia (BIZER; HEATH; BERNERS-LEE, 2009).

De acordo com a definição proposta por organizações como a Open Knowledge Foundation (okf.org), em manifestos como o “Open Definition” (“Open Definition”, 2014), em resumo, dados são abertos quando podem ser livremente utilizados, reutilizados e redistribuídos por qualquer pessoa.

### 2.1.1 As três leis de dados governamentais abertos

Foram idealizadas pelo ativista político David Eaves (EAVES, 2009a) e compartilhadas inicialmente na Conferência “Transparência na Era Digital” (EAVES, 2009b).

Segundo David, as três leis de dados governamentais abertos seriam as seguintes:

- a. Se o dado não pode ser encontrado ou indexado, ele não existe;
- b. Se ele não está em formato aberto da web e legível por máquina, ele não é reaproveitável;
- c. Se um dispositivo legal não permitir que ele seja reaproveitado, ele não é útil.

Apesar destas três leis terem sido definidas para o contexto do movimento de governo aberto, pode-se expandir seu uso para todo e qualquer dado na web que deseja ser identificado como dado aberto e conectado.

### 2.1.2 Oito Princípios dos Dados Governamentais Abertos

Na Tabela 1 estão listados os princípios criados por um grupo de 30 defensores públicos reunidos em Sebastopol, Califórnia, para os dados governamentais abertos (“Open Government Data Principles”, 2007).

Tabela 1 - Princípios Dados Governamentais Abertos

Princípio	Detalhes
Completos	Todos os dados públicos estão disponíveis. Não estão sujeitos à validação de privacidade, segurança ou privilégios limitados.
Primários	Os dados devem ser coletados na fonte, não podendo ser agregados ou modificados.
Atuais	Os dados devem ser publicados o mais rápido possível para poder ter seu valor preservado.
Acessíveis	Os dados devem ser disponibilizados para o maior público possível e para os fins mais variados possíveis.
Processáveis por Máquina	Os dados devem ser estruturados e disponibilizados em formato razoável para permitir o processamento automatizado.
Acesso não Discriminatório	Os dados devem estar disponível para qualquer pessoa sem exigência de inscrição ou autenticação.
Formato não Proprietário	Os dados devem ser disponibilizados em formatos abertos e que nenhuma entidade possua controle exclusivo.

Licença Livre	Os dados não devem estar sujeitos a regulamentação de direitos autorais, marcas, patentes ou segredo industrial. Restrições razoáveis de privacidade, segurança e controle de acesso podem ser permitidas na forma regulada por estatutos.
---------------	--

Fonte: ("Open Government Data Principles", 2007)

## 2.2 Governo Aberto

No momento em que se fala de governo aberto, nos referimos ao movimento iniciado pelo "Memorando sobre Transparência e Governo Aberto"<sup>8</sup>, diretivas de transparência, assinado por Barack Obama, presidentes dos Estados Unidos da América, logo após tomar posse do cargo em janeiro de 2009. A ideia básica é de estabelecer uma cooperação moderna entre os políticos, administração pública, indústria e cidadãos particulares, permitindo mais transparência, democracia, participação e colaboração. De forma geral, o Governo Aberto é visto como um parceiro natural para e-governo.

O e-governo não transforma a sociedade, simplesmente facilita a vida dos cidadãos. Pelo contrário, quando se fala de governo aberto, trata-se principalmente de valores, de repensar as administrações, seus procedimentos e seus dogmas (CALDERÓN; LORENZO, 2010).

A iniciativa OGP (do inglês *Open Government Partnership*) ou Parceria para Governo Aberto, lançada em 2011, quando oito países fundadores assinaram a "Declaração de Governo Aberto"<sup>9</sup> e apresentaram seus planos de ação, visa difundir e incentivar globalmente práticas governamentais relacionadas à transparência dos governos, ao acesso à informação pública e à participação social. Atualmente, 65 países integram a Parceria.

Segundo ("OGP - Iniciativa", 2014), para um governo ser classificado como aberto, ele deve buscar atingir quatro objetivos:

- a. Transparência: Amplificar a disponibilidade de informações sobre atividades governamentais;
- b. Participação Cidadã: Apoiar a participação civil;

<sup>8</sup> [https://www.whitehouse.gov/sites/default/files/omb/assets/memoranda\\_fy2009/m09-12.pdf](https://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_fy2009/m09-12.pdf)

<sup>9</sup> <https://www.governoaberto.cgu.gov.br/central-de-conteudo/documentos/arquivos/declaracao-governo-aberto.pdf>

- c. Tecnologia e Inovação: Fazer com que os padrões mais altos de integridade profissional na administração sejam implementados; e
- d. Prestação de Contas e Responsabilidade: Dilatar o acesso à novas tecnologias para fins de abertura e prestação de contas.

### 2.3 Governo Aberto Brasileiro

Instituído por Decreto em 2011, o Comitê Interministerial Governo Aberto (CIGA), foi responsável por orientar a implementação e elaboração dos Planos de Ação do Brasil, tendo sido composto de forma a envolver 18 ministérios e tendo sido coordenado pela Casa Civil da Presidência da República.

Possuía inicialmente as seguintes atribuições:

- a. Orientar a elaboração, implementação, monitoramento e avaliação do Plano de Ação Nacional sobre Governo Aberto;
- b. Propor ações prioritárias a serem implementadas no curto prazo no âmbito do Plano de Ação;
- c. Promover a articulação intragovernamental necessária à execução das ações conjuntas, troca de experiências, transferência de tecnologia e capacitação no âmbito do Plano de Ação Nacional sobre Governo Aberto; e
- d. Identificar ações de pesquisa e desenvolvimento necessárias no âmbito do Plano de Ação Nacional.

Dentro deste contexto, a Lei de Acesso à Informação (LAI) (Lei 12.527 de 18 de Novembro de 2011) será comentada em maiores detalhes a seguir. Esta lei abrange toda a nação brasileira, ou seja, União, Estados, Distrito Federal e Municípios, a respeito das entidades que estão subordinadas à lei, são todos os órgãos e entidades das esferas federais, estaduais, distritais e municipais, os 3 poderes (Executivo, Legislativo e Judiciário), Tribunal de Contas e Ministério Público, toda a Administração Pública Direta, Indireta e demais entidades controladas direta ou indiretamente. Por último, mas não menos importante, as entidades sem fins lucrativos que receberam recursos públicos para realização de ações de interesse público. Em tal caso, a publicidade a que estão submetidas refere-se à parcela dos recursos recebidos e à sua destinação (“LAI - Principais Aspectos”, 2011).

Destacam-se os seguintes princípios, que fazem parte de um conjunto de padrões estabelecidos com base nos melhores critérios e práticas internacionais:

- a. Acesso é regra, divulgação máxima;
- b. Solicitante não precisa justificar por que quer a informação;
- c. Limitação de exceções;
- d. Fornecimento gratuito de informação, salvo custos de reprodução;
- e. Propagação ativa; e
- f. Procedimentos e prazos para transparência passiva.

Um quadro com os objetivos para cada artigo da lei é demonstrado na Tabela 2.

Tabela 2 - Principais artigos da LAI e seu tema

Tema	Artigo
Abrangência da Lei	Arts. 1º e 2º
Garantias do direito de acesso / Diretrizes	Arts. 3º, 5º e 6º
Definição de termos utilizados na Lei	Art. 4º
Informações garantidas pela Lei	Arts. 7º e 21
Divulgação proativa de informações / Transparência ativa	Arts. 8º e 30
Procedimentos de acesso à informação	Art. 9º a 14
Prazos – Recebimento de respostas e interposição de recursos	Arts. 11, 15 e 16
Procedimentos em caso de negativa de acesso ou descumprimento de obrigações / Recursos	Arts. 11 §4º; Arts. 14 a 18 Art. 20
Informações sigilosas / Classificação de Informações	Arts. 7º § 1º e 2º Arts. 22 a 30 Arts. 36 e 39
Competências da CGU	Arts. 16 e 41
Competências da CMRI	Arts. 16 § 3º; 17; 35
Informações pessoais	Art. 31
Responsabilização de agentes públicos	Arts. 32 a 34
Abrangência da Lei	Arts. 1º e 2º

Fonte: (“LAI - Mapa”, 2011)

Considera-se Importante ressaltar o artigo 3º que destina-se a assegurar o direito fundamental de acesso à informação, o qual apresenta como itens que a divulgação da informação de interesse público deve ser executada sem alguma solicitação de usuários, ou seja, a informação tem que estar disponível para qualquer usuário. E o item de maior destaque para este trabalho está contido no artigo 8º, parágrafo 3º, onde está expresso: “possibilitar o acesso automatizado por sistemas

externos em formatos abertos, estruturados e legíveis por máquina; “. Desta forma, a lei preconiza que as informações devem estar acessíveis por sistemas informatizados.

## 2.4 Web Semântica

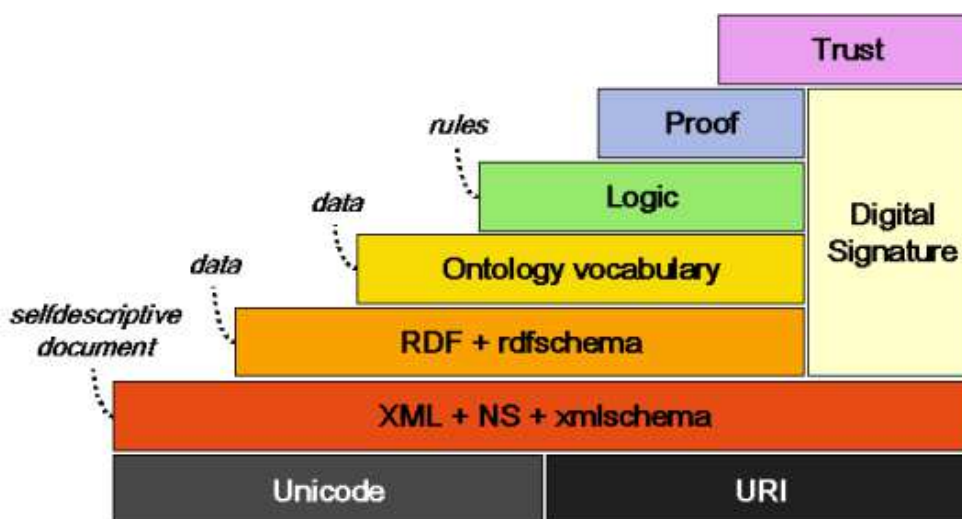
Nesta seção serão apresentadas noções gerais sobre a web semântica, padrões importantes e ferramentas disponíveis. Devido ao grande crescimento da internet nos últimos anos e com o aumento das possibilidades dos usuários em publicar e compartilhar suas próprias informações, observa-se a disponibilidade de uma enorme quantidade de dados, o que torna o seu uso uma atividade bastante complexa (BEX et al., 2014). A partir da iniciativa de instituições como o consórcio W3C e seus grupos de trabalho, a Web Semântica trata de iniciativas para facilitar a interoperabilidade dos dados e assim apoiar a criação de repositórios, vocabulários e regras para que se possam auxiliar na interoperabilidade de tais dados.

Taxonomias definem classes de objetos e as ligações entre elas. As regras de inferência permitem melhorar a análise e utilização dos dados. Para demonstrar uma ontologia na web, utiliza-se a Linguagem de Ontologia para Web (*OWL – Web Ontology Language*), que é uma extensão de *RDF Schema*. A linguagem OWL é dotada de capacidade superior de propriedades e descrição de classes em relação ao RDF, incluindo a expressão de relações entre classes, restrição de cardinalidade, características de propriedades e classes enumeradas (MCGUINNESS; HARMELEN, 2003).

A Figura 1 descreve a estrutura para web semântica descrita por Tim Berners-Lee, criador da WWW (*World Wide Web*) (BERNERS-LEE, 2015).



Figura 1 – Camadas da Web Semântica



Fonte: (“W3C Semantic Web Activity”, 2015)

Os vocábulos da Figura 1 são descritos na Tabela 2. Nela são encontradas seu significado, função e estado de sazonalidade. O vocabulário RDF é uma recomendação do W3C, sendo que ele será utilizado no modelo proposto neste trabalho e por isso foi reservado um tópico para descrevê-lo.

Tabela 2 - Tabela com os itens da proposta para Web Semântica

URI: Identificador Único de Recursos que possibilita a definição e adoção, de maneira precisa, de nomes aos recursos e seus respectivos endereços na Internet.
UNICODE: Esquema padronizado de codificação dos caracteres, que diminui consideravelmente a possibilidade de redundâncias dos dados, pois funciona independentemente da plataforma utilizada.
Signature: Conjunto de tecnologias desenvolvidas com o intuito de substituir em ambiente computacional a função exercida pela assinatura formal de uma pessoa em um suporte físico.
Encryption: Consiste de um processo em que as informações são cifradas de modo que não possam ser interpretadas por qualquer pessoa ou sistema computacional, garantindo assim a confidencialidade das informações.
XML: É uma linguagem computacional que possibilita a estruturação dos dados por meio da definição de elementos e atributos.
Namespace: Coleção de nomes, identificados por um URI, que são utilizados em documentos XML para validar elementos e atributos.
RDF Core: Núcleo que compreende as especificações do modelo e a sintaxe da <i>Resource Description Framework</i> (Estrutura de Descrição de Recursos), possibilitando a descrição dos recursos por meio de suas propriedades e valores.
RDF Schema: Utilizada para a descrição do vocabulário RDF, possibilitando a definição de taxonomias de

recursos em termos de uma hierarquia de classes.
SparQL: Linguagem computacional utilizada para realizar consultas a partir de estruturas RDF, favorecendo a recuperação de informações de maneira mais eficaz. Tal linguagem ainda não se encontra completamente padronizada, motivo pelo qual não é recomendada oficialmente pelo W3C, sendo denominada como uma tecnologia candidata à recomendação.
DLP: A DLP também é uma tecnologia candidata à recomendação e constitui a intersecção entre os dois principais paradigmas utilizados atualmente para desenvolver sistemas computacionais baseados em representação do conhecimento, Lógica Descritiva (OWL DL) e Programação Lógica (F-Logic), fornecendo uma estrutura extremamente flexível.
OWL: Linguagem computacional recomendada pelo W3C para o desenvolvimento de ontologias.
Rules: Permite a definição de regras lógicas relacionadas aos recursos informacionais
Logic Framework: Camada para a definição de regras mais abrangentes, utilizadas no tratamento das informações descritas nos níveis inferiores, possibilitando que agentes computacionais possam realizar inferências automáticas a partir das relações existentes entre os recursos informacionais, podendo inclusive inferir novas informações
Proof: Espera-se que esta camada possibilite a verificação/comprovação da coerência lógica dos recursos, de modo que os aspectos semânticos das informações estejam descritos de maneira consideravelmente adequada, atendendo a todos os requisitos das camadas inferiores.
Trust: Camada de Confiança, a partir da qual se espera garantir que as informações estejam representadas de modo correto, possibilitando um certo grau de confiabilidade.

Fonte: (RAMALHO; VIDOTTI; FUJITA, 2008)

#### 2.4.1 RDF

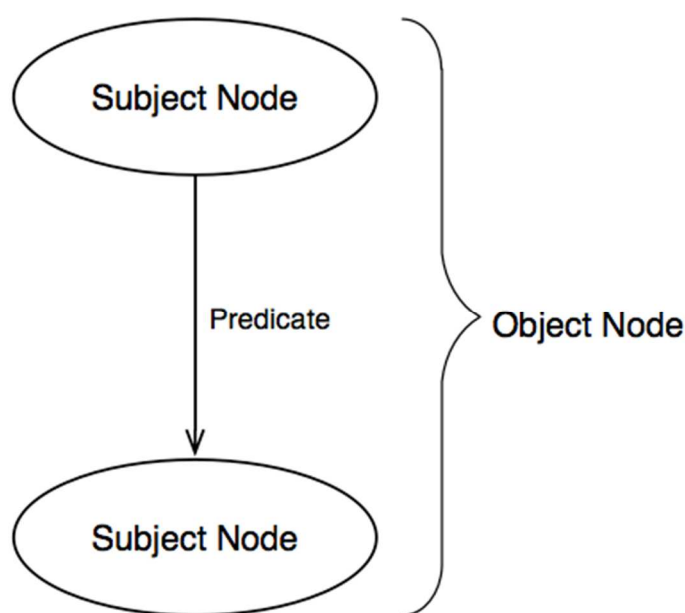
O RDF (*Resource Description Framework*) é uma recomendação do W3C desde 2004, sendo que se encontra atualmente na versão 1.1. Sua ideia é de usar, diretamente, grafos como modelo de dados. Criado originalmente para descrever metadados para páginas da web, em suma, é um padrão usado para representar informação sobre fontes na *World Wide Web*, porém tem crescido o seu leque de utilidades. O RDF é definido também como um modelo de dados universal com objetivo de estabelecer um formato de troca legível por máquina.

O RDF tenta atrair interoperabilidade diante da variedade de formatos incompatíveis existentes. Com uma semântica formal, utiliza o vocabulário *URI-based* e a sintaxe *XML-based*. Sua base é composta por uma tripla de sujeito, predicado, objeto (*subject node, predicate, object node*) que representam afirmações (*statements*).

Onde:

- *node* são endereços, variando entre local, remoto ou vazio;
- *predicates* são *IRIs*, podendo estes serem representados como uma relação entre dois nós ou como a descrição de um valor de atributo (*object node*) para qualquer assunto.

Figura 2 - Grafo RDF



Fonte: próprio autor

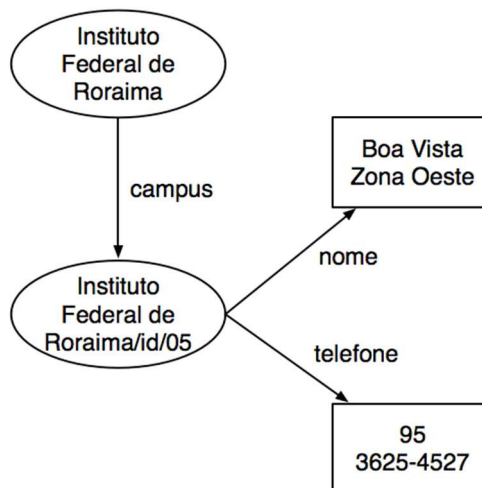
Na Figura 2 é possível observar a estrutura do RDF, definida da seguinte forma:

- *Subject node*: Qualquer coisa que possa conter um URI (*Uniform Resource Identifier*), por exemplo, páginas da web, elementos de um arquivo XML.
- *Object node*: consiste na conjunção de recurso, propriedade e valor.
- *Predicate*: um recurso que tenha nome definido e possa ser utilizado como propriedade.

Como forma de ilustrar um exemplo de utilização do RDF em uma situação prática, pode ser considerada a seguinte sentença: “O campus referenciado pelo ID 05 é chamado de Boa Vista Zona Oeste, telefone 95 3625 – 4527. O Instituto Fede-

ral de Roraima tem este campus. “. Esta sentença contém diversas informações que podem ser associadas neste contexto de nodos e relações específicas. A representação gráfica desta sentença pode ser observada na figura 3.

Figura 3 - Grafo do exemplo RDF



Fonte: próprio autor

Na Figura 4 é demonstrado um exemplo de serialização do RDF apresentado em formato de grafo, visualizado na Figura 3. A figura 4 apresenta os elementos e relações representados no modelo em forma de grafo.

Figura 4 - Código XML / RDF

```

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:s="http://description.org/schema/"
  xmlns:v="http://www.w3.org/2006/vcard/ns#" >
  <rdf:Description rdf:about="http://www.ifrr.edu.br/ifrr">
  <s:Campus>
    <rdf:Description rdf:about=" Instituto Federal de Roraima ">
      <rdf:type resource="http://description.org/schema/Organization"/>
      <v:Name>Boa Vista Zona Oeste</v:Name>
      <v:tel>95 3625 - 4527</v:tel>
    </rdf:Description>
  </s:Campus>
  </rdf:Description>
</rdf:RDF>
  
```

Fonte: próprio autor

Para facilitar a troca de dados neste formato entre aplicações, existem serviços de validação disponíveis. A Figura 5 resulta da validação do modelo de dados, através do serviço *W3 RDF Validator*<sup>10</sup>.

Figura 5 – Tripla do Modelo de Dados

Number	Subject	Predicate	Object
1	<a href="http://www.ifrr.edu.br/ifrr">http://www.ifrr.edu.br/ifrr</a>	<a href="http://description.org/RDF/Validator/run/schema/#Campus">http://description.org/RDF/Validator/run/schema/#Campus</a>	<a href="http://www.w3.org/RDF/Validator/run/Instituto%20Federal%20de%20Roraima">http://www.w3.org/RDF/Validator/run/Instituto Federal de Roraima</a>
2	<a href="http://www.w3.org/RDF/Validator/run/Instituto%20Federal%20de%20Roraima">http://www.w3.org/RDF/Validator/run/Instituto Federal de Roraima</a>	<a href="http://www.w3.org/2006/vcard/ns#Name">http://www.w3.org/2006/vcard/ns#Name</a>	"Boa Vista Zona Oeste"
3	<a href="http://www.w3.org/RDF/Validator/run/Instituto%20Federal%20de%20Roraima">http://www.w3.org/RDF/Validator/run/Instituto Federal de Roraima</a>	<a href="http://www.w3.org/2006/vcard/ns#tel">http://www.w3.org/2006/vcard/ns#tel</a>	"95 3625 - 4527"

Fonte: próprio autor

Neste estudo será utilizado o recurso RDF para criar elementos oriundos de bases de dados ligados.

#### 2.4.2 URI

*Uniform Resource Identifier (URI)*<sup>11</sup>, utilizada como identificador de recursos na internet, pode ser classificada como *Uniform Resource Name (URN)*<sup>12</sup>, *Uniform Resource Locator (URL)*<sup>13</sup> ou uma interseção das duas últimas. De forma simples, URN define o nome de uma pessoa e URL o endereço dela.

Pode-se usar o sistema do Número Padrão Internacional de Livro (*International Standard Book Number - ISBN*)<sup>14</sup> como um exemplo típico de URN. Padrão esse que se transformou na norma ISO 2108 para identificação de livros, em 1972, é utilizado o seguinte formato: (urn:isbn:0- 486-27557-4 que cita, sem incerteza, uma edição específica de *Romeo e Julieta* de Shakespeare). Para acessar esse livro, deve-se prover o endereço URL do recurso, que pode ser, por exemplo, "file:///home/user/RomeoAndJuliet.pdf".

<sup>10</sup> <http://www.w3.org/RDF/Validator/>

<sup>11</sup> <http://www.w3.org/Addressing/URL/uri-spec.html>

<sup>12</sup> <http://www.ietf.org/rfc/rfc2141.txt>

<sup>13</sup> <http://www.w3.org/Addressing/>

<sup>14</sup> <http://www.isbn.bn.br/website/>

De modo geral, uma URI é composta por um identificador de protocolo (“http”, “ftp”, “file”, etc), seguido do caractere “:” e pelo caminho para localização do recurso.

Finalmente, uma URI pode ser classificada como absoluta quando um recurso é referenciado não considerando o local no qual o mesmo está sendo referenciado (<http://www.ifrr.edu.br>) ou relativa, ao contrário, leva em consideração o contexto no qual aquele recurso é referenciado (“../script.js”, por exemplo).

### 2.4.3 SPARQL

SPARQL é uma linguagem orientada a dados, isto é, não existe inferência na própria linguagem. Ela é considerada como uma linguagem de consulta capaz de recuperar e manipular dados armazenados no formato RDF que é a base da linguagem OWL. A linguagem é normalmente utilizada em conjunto com *endpoints* sparql. Segundo (BIZER et al., 2007), os *endpoints* SPARQL podem ser classificados como genéricos, funcionando sobre qualquer base de dados RDF, que podem ser mantidas localmente ou acessada pela web, ou como específico, quando está ligado à apenas uma base de dados, não podendo estar interligada com outros *endpoints*.

A utilização desta linguagem é similar ao uso de outras linguagens de consulta amplamente conhecidas como o SQL. Deste modo, uma consulta deve ser composta pela informação, dentro da cláusula WHERE, de quais triplas devem ser analisadas e selecionadas na cláusula SELECT dos recursos a serem recuperados. A linguagem possibilita ainda a inclusão de filtros para a pesquisa. Para representar os recursos nas triplas que estão sendo consultadas, são utilizadas variáveis de consulta, as quais são identificadas por serem termos aleatórios, antecidos por um símbolo de interrogação (?), sendoque todo esse procedimento se faz necessário para elaborar uma consulta SPARQL<sup>15</sup>.

É possível configurar o *endpoint* SPARQL para retornar resultados em diferentes formatos, como por exemplo, uma tabela em formato HTML, construída através de uma transformação XSL de um arquivo XHTML resultante da consulta realizada. Quando acessado por qualquer aplicação, os resultados são serializados em formatos RDF processáveis por máquinas, como por exemplo, RDF/XML, *Turtle*

---

<sup>15</sup> <http://www.w3.org/TR/rdf-sparql-query/>

(*Terse RDF Triple Language*) e *N-Triples* (W3C RDF Core WG Internal Working Draft).

Pensando na liberdade para consultas, são utilizados endpoints SPARQL, que são interfaces utilizáveis por usuários ou aplicações em contratos de serviços para realizar consultas em uma base de dados RDF utilizando a linguagem SPARQL. Sua função nada mais é do que servir como ponto de acesso aos dados, recebendo uma requisição de consulta e ativando um motor de consulta que acessa o *triple store*, retornando o resultado. O *endpoint* pode ter como *front-end* uma aplicação Web ou *stand-alone*. Pode ser acessado via URL, na qual é disponibilizada uma interface para que a consulta seja informada e em seguida os resultados exibidos. Se caso o usuário for uma aplicação, o acesso é realizado por alguma API especializada, como, por exemplo, no caso desta dissertação, a biblioteca Apache Jena para desenvolvimento de aplicações para a Web Semântica em Java.

Na Figura 6 está ilustrado o código necessário para efetuar em SPARQL um exemplo de uma pergunta, que seria “O Rio Amazonas é maior que o Nilo?”.

Figura 6 - Exemplo Consulta SPARQL na DBPedia

```
PREFIX prop: <http://dbpedia.org/property/>
ASK
{
  <http://dbpedia.org/resource/Amazon_River> prop:length ?amazon .
  <http://dbpedia.org/resource/Nile> prop:length ?nile .
  FILTER(?amazon > ?nile) .
}
```

Fonte: (“SPARQL by Example”, 2009)

A Figura 7 demonstra, em formato XML, a resposta do *DBPedia Query Editor*<sup>16</sup> para o consulta acima.

Figura 7 - Resposta do exemplo em SPARQL

```
1 <?xml version="1.1" ?>
2 <sparql xmlns="http://www.w3.org/2005/sparql-results#"
3     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
4     xsi:schemaLocation="http://www.w3.org/2001/sw/DataAccess/rf1/result2.xsd">
5   <head></head>
6   <boolean>true</boolean>
7 </sparql>
```

Fonte: próprio autor

<sup>16</sup> <http://dbpedia.org/sparql>

#### 2.4.4 ONTOLOGIAS

Para solucionar o problema de interligação de dados originados em bases de dados diferenciadas, buscando um relacionamento mais adequado entre os conteúdos, podem ser utilizadas as ontologias. O termo foi adotado pela ciência da computação e a definição mais aceita nesse contexto é a de (GRUBER, 1993) que define ontologia como uma especificação explícita da conceituação de um domínio, visando uma classificação padronizada e formal por meio de um vocabulário controlado que representa semanticamente e exatamente os termos de um domínio específico e seus relacionamentos. O termo “formal” nessa definição conota a capacidade de ser processada computacionalmente.

As ontologias se referem a um tipo especial de objeto ou artefato que tem por objetivo representar um modelo de um determinado domínio do conhecimento na área de ciência da computação. Ontologias são meios de moldar formalmente a estrutura de um sistema, isto é, as suas entidades importantes e as ligações entre elas e que são úteis para uma finalidade definida. Ontologias podem ser codificadas em arquivos por meio da linguagem OWL - *Ontology Web Language* (GUARINO, 1998).

Na atualidade as ontologias estão sendo empregadas em inteligência artificial, web semântica, engenharia de software e arquitetura da informação, saúde, entre outras áreas, como uma forma de representação de conhecimento sobre o mundo ou alguma parte deste. Para que o compartilhamento e reutilização do conhecimento possa ser utilizado pela Ciência da Computação, é importante que os princípios presentes nas ontologias sejam bem definidos, a ponto de poder ter uma especificação formal.

A linguagem ontológica OWL é recomendada pelo consórcio W3C desde 2012, conforme já mencionado na Tabela 2, para o uso na Web Semântica. A linguagem OWL segundo (DEAN; SCHREIBER, 2004): é uma linguagem de ontologia que foi feita inicialmente para representar e definir classes. As classes são, portanto, os blocos de construção básicos de uma ontologia OWL, a qual possui um valioso conjunto de construtores para modelagem, que permite usabilidade por vários usuários.



#### 2.4.5 Ferramentas de apoio

Neste item serão comentadas algumas das ferramentas disponíveis atualmente para serem utilizadas no desenvolvimento de aplicações voltadas para a web semântica. Estas ferramentas também foram utilizadas para o desenvolvimento do modelo proposto.

A ferramenta Apache Jena Fuseki consiste de uma *triplestore*, ou seja, uma ferramenta capaz de armazenamento de dados em formato de triplas, com suporte para linguagem de manipulação.

Esta ferramenta oferece capacidade de armazenamento de dados em formato RDF gerados. Tipicamente esta ferramenta pode ser encontrada em situações onde dados obtidos em bases de dados relacionais são mapeados para o formato RDF ou em que dados no formato RDF são obtidos diretamente. O Fuseki também dispõe de uma ferramenta para efetuar consultas em SPARQL diretamente nos dados armazenados. Este servidor SPARQL é escolhido frequentemente em trabalhos por ser disponibilizado sob licença de software livre e por suportar o armazenamento dos dados em formato N-Triples, o que não ocorre em outros servidores SPARQL que disponibiliza uma versão limitada para a comunidade.

A ferramenta Protégé é um instrumento de edição e manipulação de Ontologias. É distribuída de forma gratuita, tem código aberto, não apresentando custos financeiros para sua utilização, possui arquitetura modular, o que permite a inserção de novos recursos através de *plug-ins*, desenvolvida utilizando a linguagem Java.

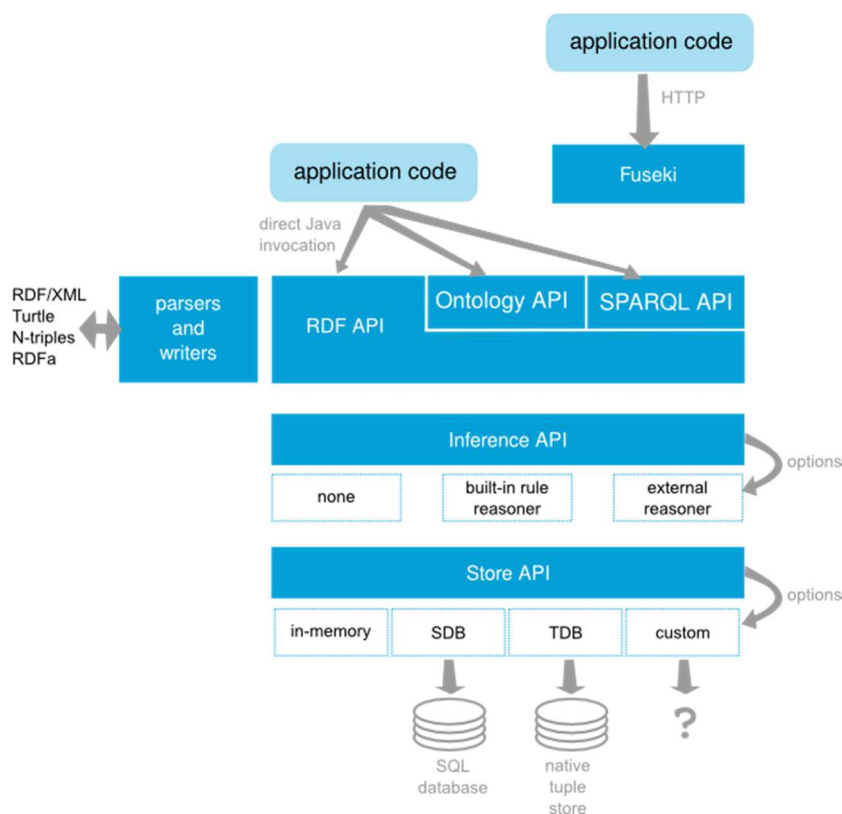
Possui fácil utilização por conta de sua interface gráfica amigável e amplamente utilizada e difundida. A ferramenta Protégé, em sua versão 5.0.0, foi desenvolvida na escola de medicina da renomada universidade de Stanford (Califórnia, EUA), pelo Stanford Medical Informatics. Essa ferramenta tem uma interface gráfica para edição de ontologias e uma arquitetura para a criação de ferramentas com base em conhecimento, podendo ser usada tanto por desenvolvedores de sistema como também por especialistas em domínio para criar bases de conhecimento, permitindo representar facilmente o conhecimento de uma área (“Protégé”, 2014).

O Apache Jena consiste de um *Framework* Java para desenvolvimento de aplicações de web semântica, originalmente desenvolvido pelos pesquisadores no

*HP Labs*<sup>17</sup> em 2000. Sendo *open-source*, isto é, de código fonte aberto, vem sendo usado extensivamente pela comunidade no desenvolvimento de aplicações de web semântica. Em 2009 seu desenvolvedor original resolveu mudar o foco das atividades de desenvolvimento para outra área, um ano depois, em 2010, foi adotado pela *Apache Software Foundation*<sup>18</sup> e em abril de 2011 foi graduado como projeto de alto nível.

Na Figura 8 pode-se verificar a arquitetura do projeto Jena e toda sua abrangência.

Figura 8 - Arquitetura do *Framework* Jena



Fonte: ("Jena", 2011)

<sup>17</sup> <http://www.hpl.hp.com>

<sup>18</sup> <http://www.apache.org>

### 3 TRABALHOS RELACIONADOS

Neste capítulo são apresentados e analisados trabalhos e iniciativas relacionadas com a temática adotada. São descritos em detalhes trabalhos que identificam, em linhas gerais, abordagens diferentes entre si, porém exerceram papel importante na elaboração do trabalho aqui apresentado. Também são expostos alguns portais de governos que possuem iniciativas com o propósito de compartilhar os dados gerados para análise e manipulação, tanto para conhecimento ou até mesmo para criação de aplicativos pela sociedade. Tais portais disponibilizam os dados abertos e conectados em vários formatos para *download* ou visualização.

Deste modo, a análise de trabalhos relacionados possui uma componente que é dedicada ao estudo de trabalhos que possam trazer contribuições e apoiar o estabelecimento do estado da arte com base na análise de sistemas e arquiteturas voltadas para atividades similares às atividades propostas neste trabalho. A outra componente é dedicada ao estudo da situação atual dos portais governamentais, no Brasil e em outros países, para identificar o contexto em que as informações e aplicações fomentando seu uso vêm sendo disponibilizados. Ao término do capítulo são expostos comentários sobre particularidades destes trabalhos e a sua influência para a definição da pesquisa proposta.

#### 3.1 Aplicação de Inferência Semântica para Exploração Conceitual de Bases de Dados Relacionais

Silva (2014) propôs uma extensão ao modelo ontológico do Orçamento Federal Brasileiro, de modo a apresentar valores de execução orçamentária relacionados com as restrições originais do orçamento planejado. Com a hipótese de que tal extensão permite aumentar as capacidades dos sistemas existentes de análise e execução orçamentária e com possibilidade de suportar a análise estratégica comparativa entre o orçamento planejado e os resultados da sua execução. Verificou a hipótese indo além do modelo ontológico anterior do Orçamento Federal que oferece apenas informações anuais do planejamento orçamentário, propondo acrescentar uma nova granularidade para séries temporais e critérios adicionais de classificação.

O autor avaliou o trabalho com uma implementação do modelo de modo a produzir datasets do orçamento em formato RDF (*resource description framework*),

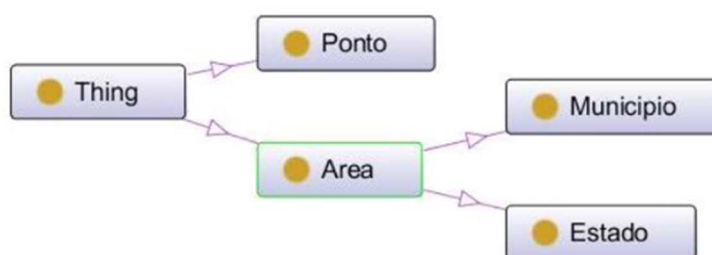
conforme recomendado pelo *World Wide Web Consortium (W3C)* como linguagem de escolha para representação do conhecimento. Desse modo, os resultados finais são publicados e acessados em um terminal de consultas denominado *endpoint*, visando dar suporte a um monitoramento público e *just-in-time* para o planejamento e para a execução do orçamento. Analisando as estatísticas de acesso, constatou que a utilização desse *endpoint* comprovava que a extensão proposta permitiu acesso ao conhecimento contido na execução orçamentária, mesmo por pessoas ou sistemas externos ao órgão de planejamento do orçamento, o que confirmou a hipótese do trabalho.

### 3.2 Aplicando Linked Data na Publicação de Dados do ENEM

Neste trabalho, (CABRAL et al., 2012) descreve o experimento que realizou com os dados do Exame Nacional do Ensino Médio (ENEM), do ano de 2008, que foram publicados seguindo os princípios dos dados abertos e conectados. Inicialmente, os dados foram tratados em um banco de dados relacional, posteriormente consolidados; os dados manipulados foram representados por uma ontologia criada pelos autores a partir de uma outra já existente. Em seguida, com o uso de ferramentas diversas, os dados foram convertidos para o formato RDF, ligados aos dados da DBPedia e publicados em um servidor de triplas. Enfim, foi desenvolvida uma aplicação Web para que estes dados fossem visualizados com o auxílio de consultas SPARQL. O experimento permitiu estabelecer um fluxo para publicação de dados geográficos, possibilitando a descoberta de novos conhecimentos.

Na Figura 9 é demonstrada a ontologia, que possui as propriedades referentes aos municípios brasileiros, na qual os autores se basearam para criar uma ontologia específica para o trabalho, denominada “enem”, que interliga os dados dos exames aos outros.

Figura 9 - Ontologia "geopoliticabr"



Fonte: (CABRAL et al., 2012)

Essa ontologia possui algumas propriedades relevantes, dentre as quais pode-se destacar o *sameAs*, para a DBPedia e os pontos cardeais, este último importante para auxiliar na visualização dos dados na aplicação desenvolvida neste trabalho.

### 3.3 Avaliação do ensino superior público no Brasil: protótipo de aplicação *linked data*

Neste artigo (SPERONI et al., 2014) os autores, a fim de ampliar e estimular a participação e controle social sobre os processos de regulação, avaliação e supervisão da educação superior no sistema federal de educação, desenvolveram uma aplicação web para visualização dos dados com auxílio de consultas SPARQL, que permite o cidadão cruzar os dados dos portais do MEC, do INEP e das Instituições Federais de Educação Superior – IFES, para verificar dados como o corpo docente, projeto pedagógico, matriz curricular, resultados do Exame nacional de Desempenho de Estudantes - ENADE, bem como outras informações gerais, de forma integrada e automatizada.

A obtenção dos dados se deu de diferentes formas. Referente ao ENADE 2012 o arquivo foi utilizado em formato CSV, contendo dados da instituição, do curso e o seu conceito. Para complementar estes dados das instituições, buscou dados na página do e-MEC<sup>19</sup>. Este sistema disponibiliza uma interface de consulta sobre instituições e cursos de ensino superior, conforme cadastro realizado pelas próprias instituições. A forma de obtenção foi com a técnica de *web scrapping*, que permite a extração, de forma automática, de pontos específicos de dados por meio da análise dos documentos codificados em linguagem de marcação, ao invés de copiá-los manualmente, o que foi o caso da obtenção dos dados das páginas dos cursos, corpo docente, matriz curricular e projeto político pedagógico, pois não se existe um padrão nos *websites* das instituições. Para enriquecer a visualização dos dados, obteve-se, em formato RDF, os dados com as coordenadas de localização dos municípios para incluir o georeferenciamento das instituições.

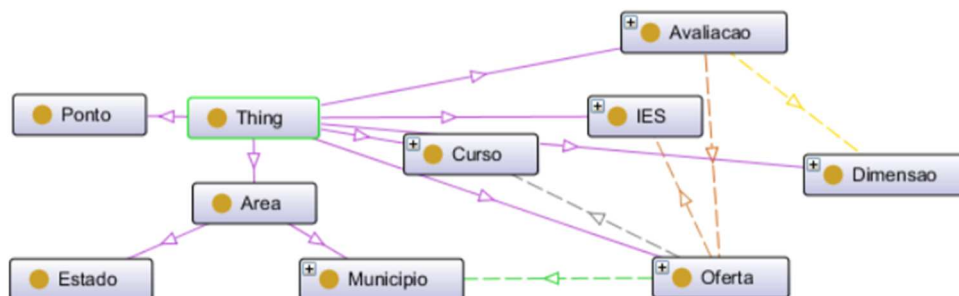
Na Figura 10 é demonstrada, com suas classes e relacionamentos, a ontologia criada para atender o escopo deste trabalho. Foi utilizada, por meio de importa-

---

<sup>19</sup> <http://emec.mec.gov.br>

ção, uma ontologia pré-existente que representa as classes e propriedades referentes aos municípios brasileiros.

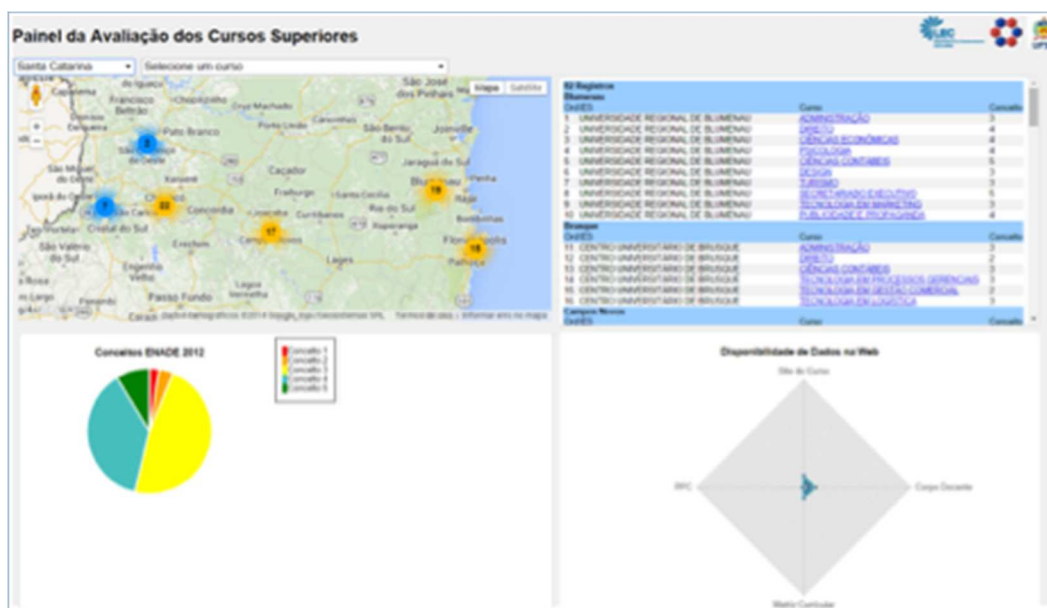
Figura 10 - Ontologia de Domínio de Avaliação



Fonte: (SPERONI et al., 2014)

Pode-se visualizar na Figura 11 a página principal da aplicação desenvolvida para visualização dos dados das instituições, com filtros por Estados e/ou cursos. Os dados nesta página foram obtidos através de consulta realizadas em um *end-point* SPARQL pelos scripts desenvolvidos em PHP e serializados em JSON.

Figura 11 - Página principal da aplicação



Fonte: (SPERONI et al., 2014)

Após selecionar um curso, a aplicação exibe uma página com as informações referentes ao mesmo, tais como dados sobre o ENADE, avaliação, instituição e etc. A Figura 12 exibe os dados do curso de Administração na Universidade Federal de Santa Catarina.

Figura 12 - Página com detalhes do curso

UNIVERSIDADE FEDERAL DE SANTA CATARINA - ADMINISTRAÇÃO - Florianópolis			
Instituição	Curso	ENADE	Avaliação
ENADE		2012	
Estudantes Inscritos:		477	
Estudantes Participantes:		397	
Conceito ENADE Contínuo:		4,15	
Conceito ENADE Faixa:		5	

(SPERONI et al., 2014)

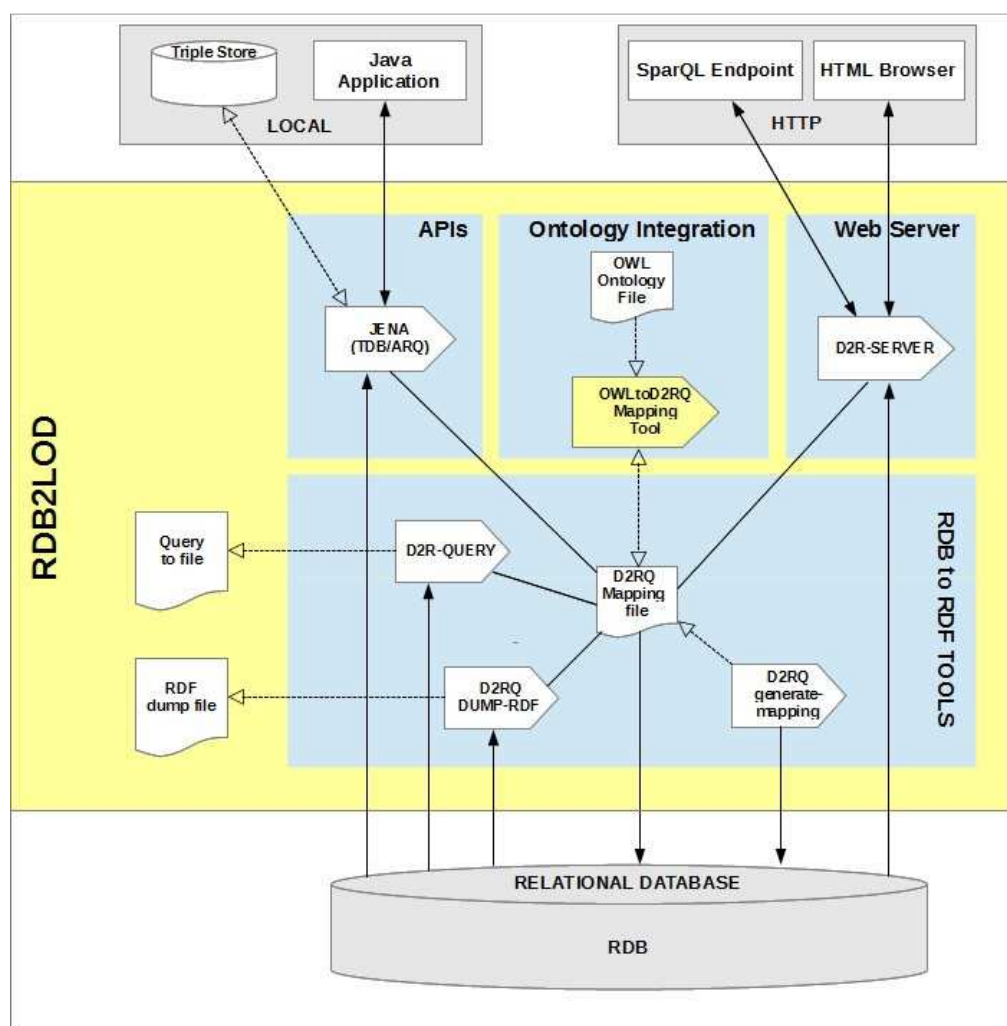
### 3.4 Uma Abordagem para a Publicação de Dados Ligados Obtidos a Partir de Bases de Dados Relacionais

Neste trabalho, Pereira (2014) apresentou uma proposta de abordagem com o intuito de facilitar e automatizar a publicação de dados abertos e ligados obtidos a partir de bases de dados relacionais, por meio da integração entre as diversas ferramentas de software aplicadas neste processo. Ofereceu ainda uma ferramenta, denominada *RDB2LOD*, que possibilita customizar, de forma semiautomática, o arquivo de mapeamento entre a base relacional e modelo de dados RDF, com a finalidade de incorporar uma ontologia de domínio, fornecida pelo usuário, neste arquivo.

Os autores encontraram vários trabalhos relacionados e propuseram como diferencial, a automatização, implementada pela interface gráfica das ferramentas para gerar e customizar o mapeamento entre base de dados relacional e RDF e para sua visualização e consulta dos dados ligados obtidos a partir desse mapeamento. Também se diferencia por ser compatível com o Sistema de Gerenciamento de Banco de Dados – SGDB *Microsoft SQL Server* e a ferramenta de implementação do método para incorporar uma ontologia ao arquivo de mapeamento, estar compatível com as recomendações de padrão do *W3C*.

Na Figura 13 é possível visualizar uma visão geral da abordagem proposta neste trabalho.

Figura 13 - Visão Geral da Abordagem RDB2LOD



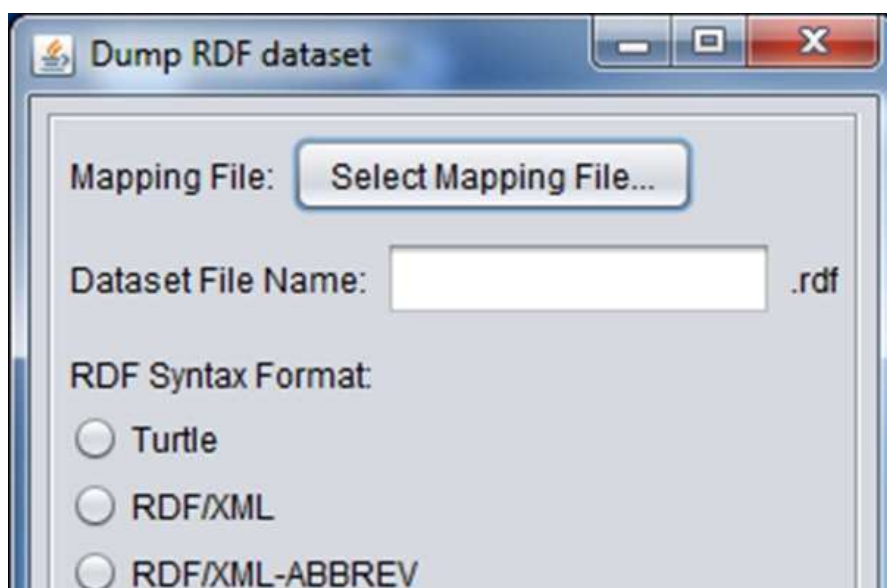
Fonte: (PEREIRA, 2014)

A camada inferior é responsável pela comunicação com a base de dados relacional, na camada acima encontra-se ferramenta proposta neste trabalho e por último, na camada superior pode-se encontrar o *triplestore* onde os dados são armazenados em formato RDF, a aplicação em Java que funciona localmente, o *endpoint* SPARQL e o navegador HTML que fazem a comunicação via *web* com a ferramenta.

A Figura 14 demonstra a tela da aplicação desenvolvida para integração das ferramentas aplicadas na abordagem. Nesta tela, podemos ver as opções de exportação dos dados já em formato RDF.



Figura 14 - Tela da aplicação desenvolvida



Fonte: (PEREIRA, 2014)

Pelo fato da aplicação desenvolvida possibilitar a inclusão de uma ontologia, os dados são gerados de forma padronizada e de fácil visualização pelo usuário. Um exemplo dessa visualização é comparada na Figura 15, onde podemos identificar de forma mais simples os dados onde as propriedades da classe foram customizados.

Figura 15 - Comparação entre visualizações de dados

Custom Mapping	
Property	Value
rdfs:label	AtendimentoAmbulatorial #122415808
temIDAtendimento	122415808 (xsd.decimal)
temIDESpecialidade	<http://localhost:3050/resource/EspecialidadeMedica/19991132>
temIDFaixaEtariaAtendimento	<http://localhost:3050/resource/FaixaEtaria/19991164>
temIDPrestador	<http://localhost:3050/resource/Prestador/19991140>
temMunicipioAtendimento	<http://localhost:3050/resource/Municipio/351230>
temPeriodoAtendimento	<http://localhost:3050/resource/Periodo/200612>
temRegiaoSaudeAtendimento	<http://localhost:3050/resource/RegiaoSaude/3511>
temValorAprovado	571.5 (xsd.decimal)
rdf:type	AtendimentoAmbulatorial

Standard Mapping	
Property	Value
db:vocab/dbo_SIA_F_PRODUCAO_PERD_PER_ID	<http://localhost:2020/resource/dbo/PER_D_PERIODO/200612>
db:vocab/dbo_SIA_F_PRODUCAO_SIAD_ESP_ID	<http://localhost:2020/resource/dbo/SIA_D_ESP_PROFISSIONAL/19991132>
db:vocab/dbo_SIA_F_PRODUCAO_SIAD_FXE_ID	<http://localhost:2020/resource/dbo/SIA_D_FAIXA_ETARIA/19991164>
db:vocab/dbo_SIA_F_PRODUCAO_SIAD_TPR_ID	<http://localhost:2020/resource/dbo/SIA_D_TIPO_PRESTADOR/19991140>
db:vocab/dbo_SIA_F_PRODUCAO_SIAF_PRB_SK	122415808 (xsd.decimal)
db:vocab/dbo_SIA_F_PRODUCAO_SIAF_PRB_VALORAPRO	571.5 (xsd.decimal)
db:vocab/dbo_SIA_F_PRODUCAO_UNT_MUNATE_ID	<http://localhost:2020/resource/dbo/UNT_MUNICIPIOS/351230>
db:vocab/dbo_SIA_F_PRODUCAO_UNT_RGSATE_ID	<http://localhost:2020/resource/dbo/UNT_REGIOES_SAUDE/3511>
rdfs:label	SIA_F_PRODUCAO #122415808
rdf:type	vocab:dbo_SIA_F_PRODUCAO

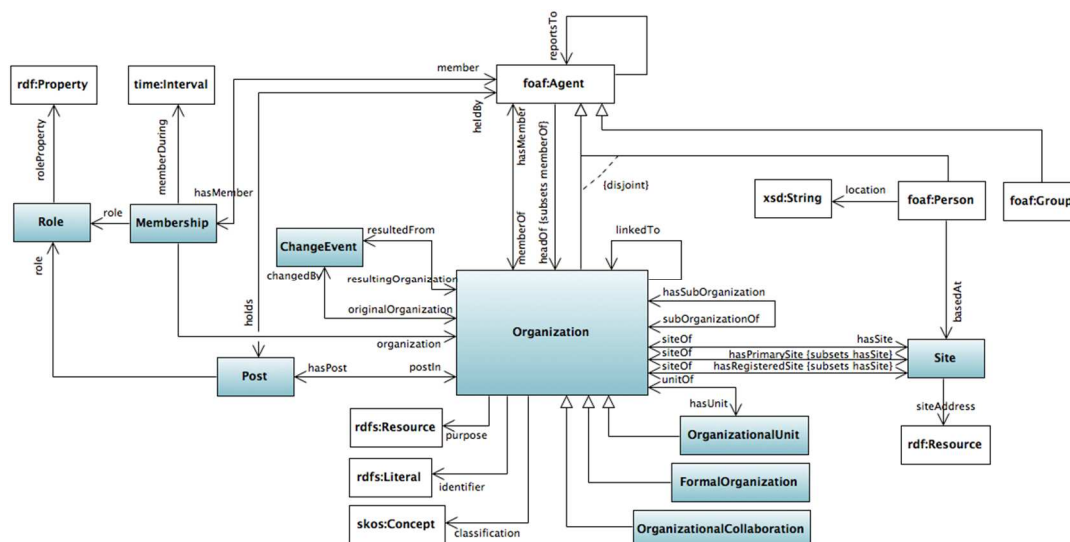
Fonte: (PEREIRA, 2014)

### 3.5 Mapeando Dados Governamentais com uma Ontologia de Organizações

Este trabalho relata um estudo de caso (FONSECA et al., 2014) na integração de dados governamentais no domínio de publicação de informações sobre organizações e estruturas organizacionais utilizando a *W3C ORG Ontology*. No estudo realizado neste artigo, foram recuperados dados publicados em seu formato corrente, um mapeamento destes dados para a ontologia de referência foi realizado e estes dados foram publicados em formato RDF conforme a ontologia.

A *ORG Ontology* em formato de diagrama não normativo, utilizada neste trabalho, é apresentada na Figura 16. Fornece um modelo genérico e reutilizável.

Figura 16 - Ontologia Organizacional



Fonte: REYNOLDS, 2014 apud (FONSECA et al., 2014)

Com o intuito de verificar tanto a integração dos dados disponibilizados pelo governo quanto a abrangência da ontologia no escopo analisado, todas as bases de dados utilizadas neste trabalho, pertencem ao governo federal. Sua escolha se deu baseada na relevância dos dados e em sua representatividade em relação a *ORG Ontology*, pois verificou-se que estes dados têm interesse público, podem ser mapeados e quando em conjunto, representam uma parte de suma significância de conceitos da ontologia.

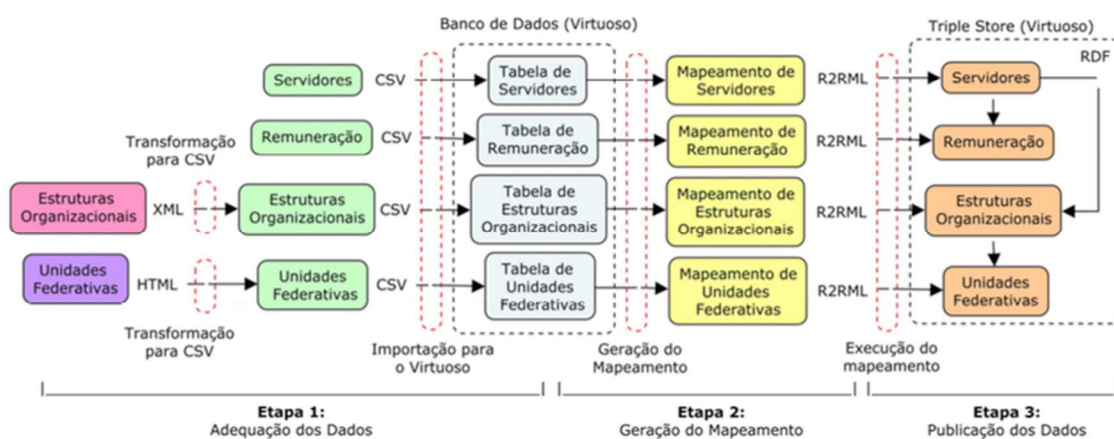
A seguir a relação das bases de dados utilizadas neste trabalho com seus respectivos formatos de dados:

1. Servidores Civis (Servidores e Remuneração), formato CSV;

- b. Estruturas Organizacionais, formato XML;
- c. Catálogo de Unidades Federativas, formato HTML.

Para o mapeamento dos dados com a ontologia, os autores utilizaram a linguagem R2RML. Na Figura 17 é exposta as etapas do mapeamento dos dados.

Figura 17 - Etapas do Mapeamento



Fonte: (FONSECA et al., 2014)

Seguindo o fluxo de trabalho, após a adequação dos dados, tem-se a geração do mapeamento e por último sua publicação. Problemas da falta de identificação única de certos dados foram encontrados, como nomes de cargos com abreviações diversas, mas que serviram para se propor soluções simples, porém eficazes. Também serviu para identificar a falta de integração entre diferentes bases de dados, um exemplo exposto no trabalho foi a de que entidades, quando possuem identificação em uma base específica, apresentam uma identificação que não é compartilhada com as demais bases do governo, fato que tornou difícil a correta identificação, separação e relacionamento dos dados.

### 3.6 Portais Governamentais

Nesta seção serão descritos e comentados aspectos de portais que visam a divulgação de dados para a comunidade em geral. Esta seção reúne exemplos de portais de governos brasileiros em diferentes esferas (nacional, estadual e municipal), bem como trata de alguns casos de portais de governos estrangeiros.

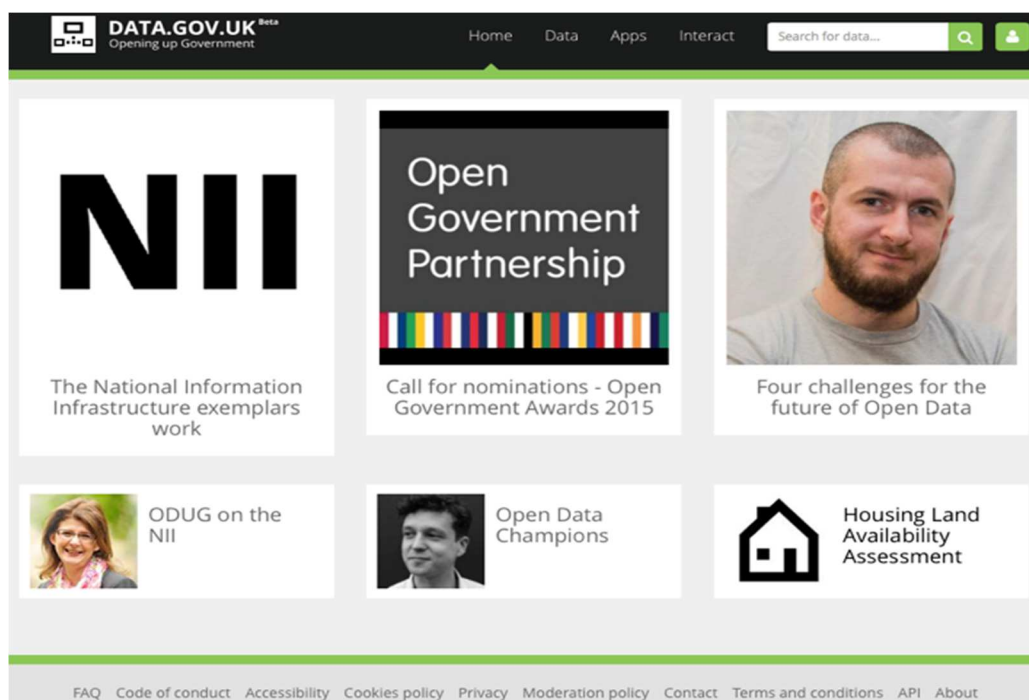
O objetivo da seção é complementar a informação dos trabalhos relacionados já apresentados, com a contrapartida existente do ponto de vista das iniciativas governamentais.

### 3.6.1 Reino Unido

Disponibilizado no início do ano de 2010, é considerado uma referência do uso de dados abertos no mundo. Conta hoje com mais de 24.500 *datasets* (conjunto de dados). Também nota-se que todos os dados podem ser adquiridos através de *download* em vários formatos. No portal [data.gov.uk](http://data.gov.uk) estão disponíveis, além dos dados, um conjunto de aplicativos, desenvolvidos pela sociedade, que utilizam os dados disponibilizados pelo governo.

A Figura 18 demonstra a página inicial do portal. Simples, funcional e moderna.

Figura 18 - Página inicial do Portal do Reino Unido



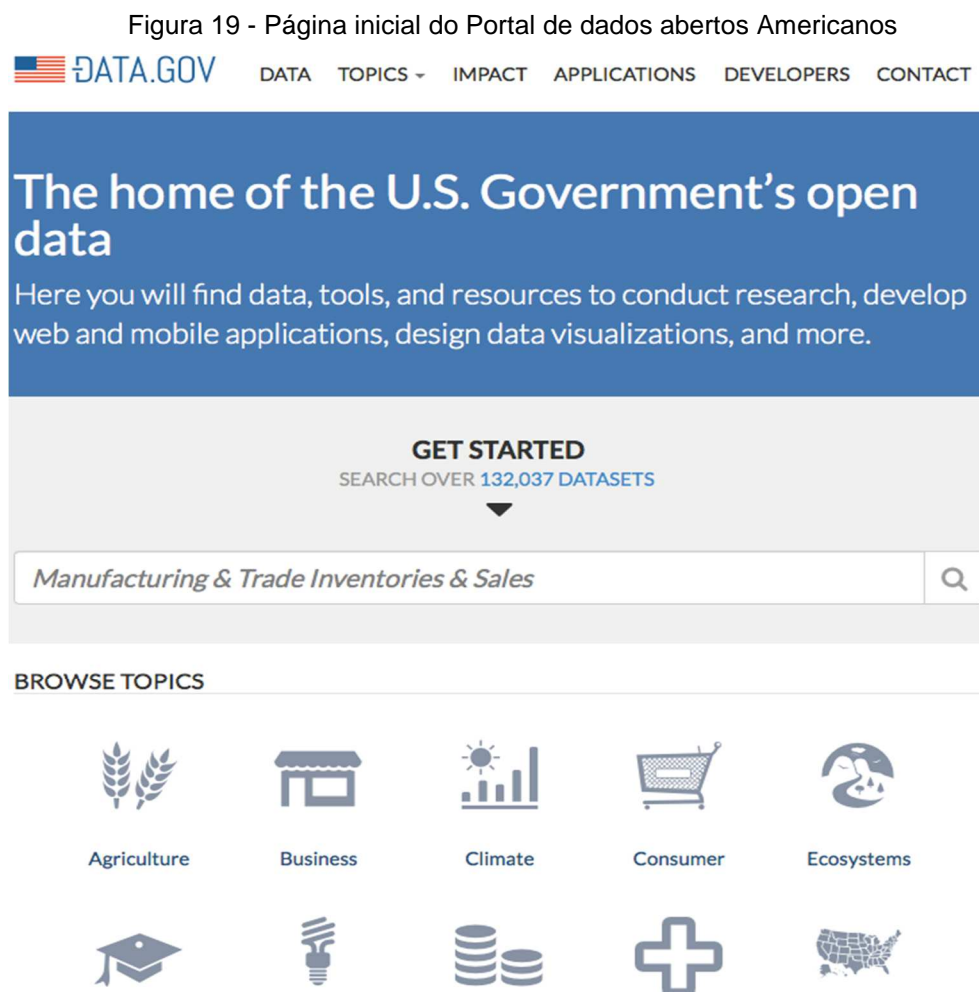
Fonte: [data.gov.uk](http://data.gov.uk)

### 3.6.2 Estados Unidos da América

O site [www.data.gov](http://www.data.gov), lançado pelo governo americano em maio de 2009, continha apenas 47 *datasets*, sendo que no momento de escrita deste texto, conta com mais de 132.000 diferentes conjuntos de dados. Também divulga os aplicati-

vos desenvolvidos utilizando estes dados, assim como observado no portal do Reino Unido.

Na Figura 19 é possível visualizar a disponibilidade de um campo para o usuário pesquisar informações nestes *datasets*, além de tópicos para navegar de forma mais rápida.



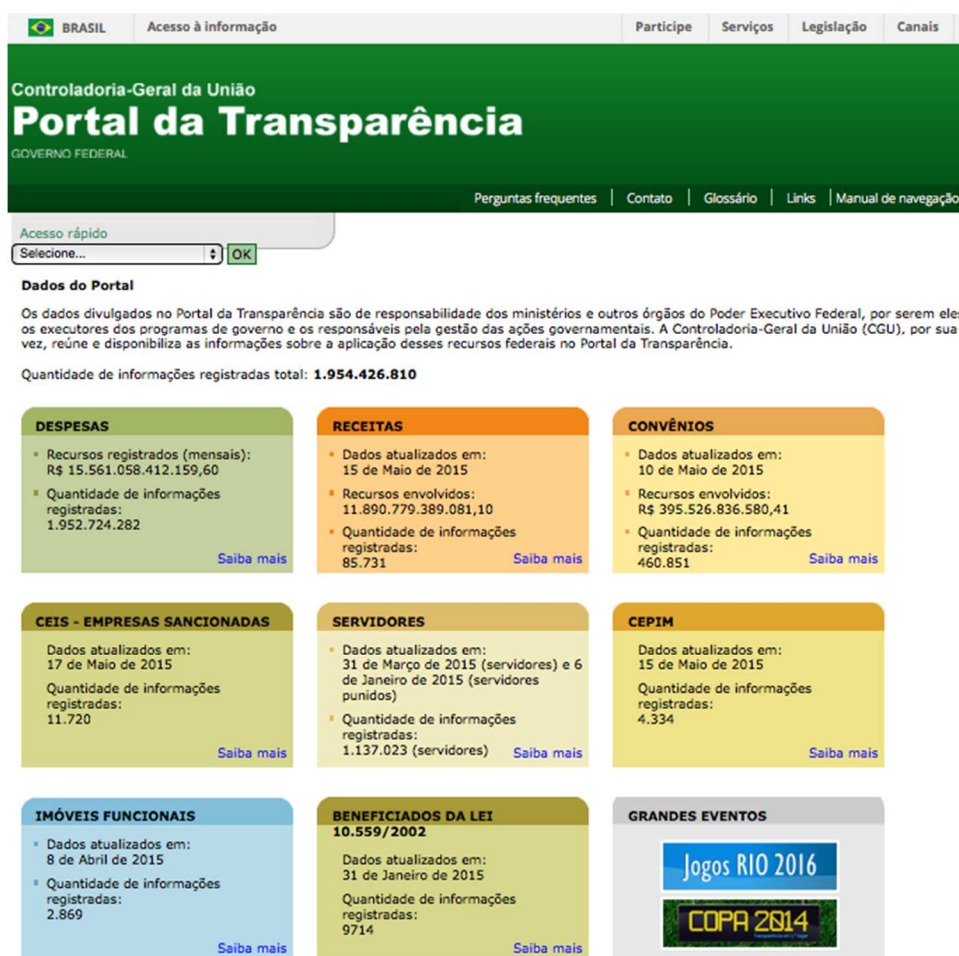
Fonte: data.gov

### 3.6.3 Portal da Transparência Federal

O Portal da Transparência Federal foi criado em 2004 pelo governo federal através de uma iniciativa da Controladoria Geral da União (CGU). Este site está localizado nesta URL: o [www.portaltransparencia.gov.br](http://www.portaltransparencia.gov.br). Este site visa apoiar a aplicação adequada dos recursos públicos. Possui o objetivo de aumentar a transparência da gestão pública, permitindo que o cidadão acompanhe a utilização do dinheiro público e ajude a fiscalizar. Em linhas gerais, o objetivo do site é atuar com medidas de transparência para mitigar iniciativas de corrupção.

Pode-se verificar na Figura 20 os dados disponíveis no portal e a enorme quantidade de informações registradas.

Figura 20 - Dados disponíveis no Portal da Transparência



Fonte: portaldatransparencia.gov.br

### 3.6.4 DataPOA

O portal DATAPOA é um projeto de dados abertos da prefeitura de Porto Alegre, capital do Estado Rio Grande do Sul, que dispõe atualmente de 59 conjuntos de dados dentre os grupos de administração, finanças, limpeza, educação, dentre outros. Seguindo o exemplo dos demais portais, divulga os aplicativos para celulares, que utilizam os dados divulgados.

Pode-se notar a objetividade do portal na Figura 21, com filtros para grupos específicos, acesso via *link* para os aplicativos já desenvolvidos e que utilizam os dados do sítio.

Figura 21 - Página Inicial DataPOA



Fonte: datapoa.com.br

### 3.6.5 Dados Abertos Federal

Esta é uma ferramenta disponibilizada pelo governo federal para que todo e qualquer cidadão possa encontrar e utilizar os dados e as informações públicas. Tem por objetivo disponibilizar todo e qualquer tipo de dado, por exemplo, dado da saúde suplementar, do sistema de transporte, da segurança pública, indicadores de educação, gastos governamentais, entre outros.

Figura 22 - Página Inicial do portal de Dados Abertos Federal

BRASIL Acesso à informação Participe Serviços Legislação Canais

Ir para o conteúdo 1 Ir para o menu 2 Ir para a busca 3 Ir para o rodapé 4

**dados.gov.br**  
PORTAL BRASILEIRO DE DADOS ABERTOS

Dados | Aplicativos | Perguntas frequentes | Contato | Sobre o portal

Pesquisa ... PESQUISAR

em 862 conjuntos de dados com 5805 recursos (o que é isto?)

**Dados em destaque**

**Obras do PAC - Programa de Aceleração do Crescimento**  
Criado em 2007, no segundo mandato do presidente Lula (2007-2010), o Programa de Aceleração do Crescimento – PAC promoveu a retomada do planejamento ...

**Reclamações do Consumidor.gov.br**  
O Consumidor.gov.br é um novo serviço público para solução alternativa de conflitos de consumo por meio da internet, que permite a interlocução ...

**SICONV - Convênios e Contratos de Repasse da Administração ...**  
A prestação de dados como um serviço governamental traz vantagens pra toda a sociedade, incluindo o próprio governo. O Ministério do Planejamento, ...

**Dados por etiqueta**

INDE (770) Geoespacial (636) Geoserviço (636)  
Cartografia (636) PGI (169) Indicador (169)  
Previdência Social (101) AEPS (100) Benefícios (44)  
Benefícios Previdenciários (43) Contribuintes (21)  
GFIP (19) UF (19) Equipamentos públicos (17)  
federal (16) INEP (15) GPS (14) IBGE (13)  
Receitas (13) SP (13)

**Publicações mais recentes**

Conjunto de dados	Autor	Quando
Indicadores sobre Cadastro Geral de ...	Gabinete do Ministro - ...	às 11:20
Indicadores sobre Avaliação dos cursos ...	Coordenação de ...	às 11:20
AEROVIAS	Divisão de Operações	10 Mai
AERÓDROMOS DO ROTAER	Divisão de Operações	10 Mai
ESPAÇO AÉREO RESTRITO	Divisão de Operações	10 Mai

**Notícias**

**Ministério do Planejamento divulga Plano de Dados ...** 19/05/2014  
Publicação servirá de base para outros órgãos elaborarem seus planos Brasília, 16/05/2014 – O Ministério do Planejamento (MP) lança nesta ...

**Dados abertos permitem que cidadãos conheçam melhor as ...** 16/04/2014  
Brasília, 15/04/2013 – O Ministério da Justiça coloca para consulta pública, nesta quarta-feira (16), seu Plano Institucional de Dados Abertos e ...

**Prefeitura de São Paulo institui o Catálogo Municipal ...** 29/01/2014  
Brasília, 29/01/14 – A Prefeitura de São Paulo alterou, no dia 23 de janeiro, o decreto que regulamentava a Lei de Acesso à Informação (LAI) no ...

Fonte: dados.gov.br

Na Figura 22 demonstra a página inicial do portal expondo notícias referentes à dados abertos, publicações mais recentes, dados em destaque, navegação nos conjuntos de dados tanto por etiquetas como pelo campo de busca. Também disponibiliza uma relação com os aplicativos desenvolvidos pela comunidade e que utilizam os dados catalogados no sítio de dados abertos do governo federal.

### 3.7 Análise dos Trabalhos Selecionados

Os trabalhos apresentados neste capítulo foram selecionados com a finalidade de colaborar na justificativa e verificação de lacunas no estado da arte, que possam ser utilizadas como motivação para esta pesquisa.

Em linhas gerais, cabe comentar que os trabalhos estudados evidenciam a importância de disponibilizar os dados públicos através da rede mundial de compu-



tadores e a padronização de tecnologias para divulgação dos mesmos. A centralização de dados torna a sua consulta mais fácil pela comunidade e a interligação destes dados entre si torna mais do que útil a sua divulgação. Entretanto no campo de padronização são observadas algumas lacunas e uma baixa adesão ao uso dos padrões que proporcionam maiores possibilidades de integração, como é o caso dos padrões propostos pelo W3C para interligação, tais como o RDF.

O desenvolvimento de aplicações com o intuito de gerar relatórios com a análise desses dados vem se tornando uma realidade. Isso faz com que a tomada de decisões importantes possa ser realizada com mais precisão, confiabilidade e em menor prazo. Também proporciona meios de divulgação aos usuários. Observa-se que as aplicações em sua grande maioria tratam de visualização de conjuntos específicos de dados, o que é um avanço, mas que poderia ser ainda maior, com a criação de aplicações que permitam a integração de mais de um conjunto de dados.

Observando-se os resultados obtidos com esse processo, é possível identificar que a iniciativa das instituições em tornar os dados abertos vem se popularizando bastante, dado o número de portais e o volume de dados disponibilizados. Este fato possui um desdobramento importante pois apoia a criação de novas estratégias para que empresas possam atuar de modo a determinar fontes de problemas com base em análises de dados, além de diminuir gastos desnecessários.

A Tabela 3 resume a análise dos artigos relacionados com base em alguns aspectos acerca dos trabalhos iniciais na área de web semântica e dados abertos conectados. As dimensões analisadas nestes trabalhos, com finalidade de comparação e identificação de lacunas e oportunidades de desenvolvimento de pesquisa, estão relacionadas com a identificação de recursos úteis para aplicações da web semântica e de integração de dados.

Desta forma são analisados os seguintes elementos:

- Uso de RDF: avalia se a aplicação está preparada para utilizar os recursos de representação de dados em formato RDF, o que proporciona ganhos de integração de dados com outras bases e fontes.
- Uso de *triplestore*: indica se a aplicação utiliza recursos de armazenamento com base em *triplestore*, para armazenamento de triplas em formato específico, o que pode facilitar o tratamento de volumes maiores de dados com mais segurança e eficiência.

- **Uso de Ontologia:** explicita o uso de ontologias para apoio nas atividades da aplicação. Considera-se sua existência um indicador de possibilidades maiores de automatização em integrações.
- **Dados Conectados:** indica se a aplicação emprega especificamente dados conectados, ou seja, se utiliza formatos que fomentam a integração de dados. A sua presença atua no sentido de fomentar integrações de modo mais amplo.
- **Aplicação para consumo:** avalia se a aplicação disponibiliza um recurso no formato facilitado de consulta para os dados tratados, tais como o recurso de API para acesso à serviços. A existência desta possibilidade amplia a capacidade de divulgação de dados e de desenvolvimento de aplicações diversas para seu uso.
- **Utilização de dados abertos:** avalia se os dados são disponibilizados no formato aberto, ou seja, sem restrições de acesso, aspecto que também atua como facilitador do compartilhamento.

Tabela 3 - Aspectos dos Artigos Relacionados

Aspectos	(SILVA, 2014)	(CABRAL et al., 2012)	(SPERONI et al., 2014)	(PEREIRA, 2014)	(FONSECA et al., 2014)
Uso de RDF	SIM	SIM	SIM	SIM	NÃO
Uso de <i>triplestore</i>	SIM	SIM	SIM	SIM	NÃO
Uso de Ontologia	SIM	SIM	SIM	SIM	SIM
Dados Conectados	NÃO	NÃO	NÃO	NÃO	NÃO
Aplicação para consumo	NÃO	SIM	SIM	NÃO	NÃO
Utilização de dados abertos	SIM	SIM	SIM	NÃO	SIM

Fonte: próprio autor

A partir da análise realizada, observa-se que os trabalhos estudados empregam de forma consistente os principais elementos identificados como fatores de importância para apoiar a melhor utilização dos dados e sua integração. Apesar de haver uma unanimidade em alguns fatores, como o caso de uso da ontologia, isso não ocorre com os demais elementos. As principais fragilidades ou pontos de melhoria podem ser identificadas na ampliação do uso de dados conectados, que é

uma clara lacuna, além do uso de disponibilização de aplicações para consumo, que aparece somente em parte pequena das situações.

Observa-se como possibilidade de contribuição a integração de aspectos mais amplos em uma única aplicação, tendo em vista que nenhum dos artigos trabalhou com dados abertos e conectados, uma parte não trabalha com dados abertos e existe uma parcela pequena com apoio para aplicações a partir de uma API. Além de integrar todos os aspectos citados na análise, o trabalho desenvolvido também explora a divulgação dos dados em uma aplicação para exemplificar o consumo destes dados após sua integração.

Uma parte complementar desta análise proporcionou a compilação de dados sobre portais governamentais, com o objetivo de averiguar quais os principais formatos utilizados na sua divulgação e também qual o volume de conjuntos de dados disponibilizados. O resultado deste estudo está descrito na tabela 4.

Tabela 4 - Quantidades de *datasets* por Portal

Formatos	Portais Governamentais		Portais Nacionais		
	Reino Unido	Estados Unidos da América	Portal da Transparência Federal	DataPOA	Dados Abertos Federal
CSV	3934	9682	23	35	774
XML	396	23756	0	0	270
JSON	136	8663	0	12	261
HTML	1881	26557	0	0	860
Outros formatos	18543	62.543	0	12	695
Total de <i>datasets</i>	24890	131.201	23	59	1001

Conforme

Tabela 4, é possível identificar a enorme quantidade de conjuntos de dados disponíveis nos portais governamentais de dados abertos. Também é possível identificar a disparidade no número de formatos disponibilizados, entre esta amostra de portais analisados. A existência de um maior número de formatos de disponibilização tende a apoiar a integração de dados, aumentando a possibilidade de uso de ferramentas e técnicas de integração. Entretanto a existência de formatos de dados conectados está associada com o aumento das possibilidades de integração dos mesmos. Até o momento da análise realizada a quantidade de formatos com possibilidade de representação de dados conectados ainda é pequena ou inexistente em certos casos.



## 4 MODELO PROPOSTO

Neste capítulo é apresentado o modelo proposto para esta pesquisa. Tendo como base os estudos de trabalhos relacionados, o modelo proposto busca compor os elementos necessários para a integração de dados de bases diversas, de modo flexível e com possibilidades de disponibilização facilitada dos resultados. A seguir é apresentada a visão geral do modelo e o detalhamento de seus componentes.

### 4.1 Visão geral do modelo

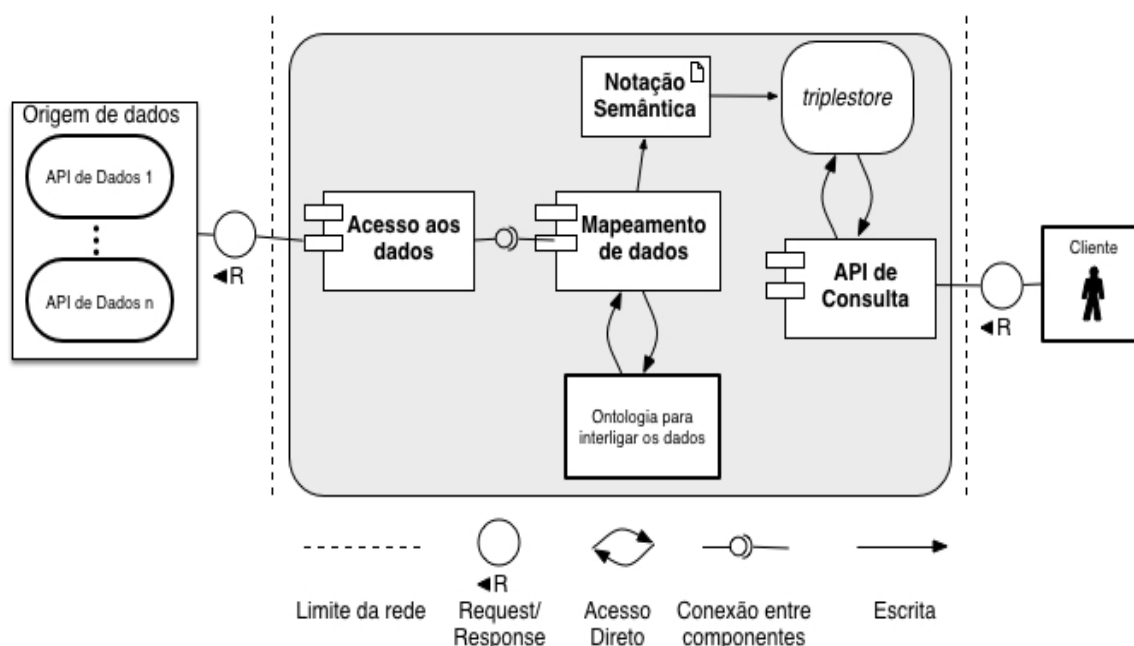
Este modelo tem por contribuição a integração flexível das informações de bases de dados ligadas, com uso de ontologias. Além disso, anseia automatizar o processo de mapeamento dos dados, em parte, baseando-se em nomes de elementos similares nas bases de dados relacionais indicadas.

Por fim, também visa contribuir com o desenvolvimento de aplicações de acesso aos dados, de simples utilização por parte do usuário, para consulta dos dados armazenados em um *triplestore*, com filtros baseados em uma ontologia criada para interligar os dados semanticamente.

A Figura 23 descreve os componentes do modelo proposto, que serão analisados a seguir. Para descrição do modelo foi utilizada a notação TAM/FMC (TAM, 2012). A notação TAM (*Technical Architecture Modeling*) foi definida pela SAP e possui como objetivo facilitar a descrição do nível conceitual das aplicações e também aspectos do seu projeto. Para isso são combinados elementos da descrição FMC (*Fundamental Modeling Concepts*) em conjunto com a notação UML (*Unified Modeling Language*) (KNÖPFEL, 2007).

O modelo proposto é baseado na arquitetura cliente/servidor, considerada a mais adequada para esta finalidade de integração de dados e disponibilização de resultados. Na arquitetura, o componente servidor detém a API disponibilizada para realização das consultas SPARQL e também armazena os dados em formato RDF. O componente cliente do modelo mantém os métodos que implementam as consultas a serem executadas conforme forem solicitadas pelo usuário durante sua utilização.

Figura 23 - Modelo Genérico



Fonte: autor

Na Figura 23 estão representadas as diversas fontes de dados a integrar, no componente denominado “Origem dos dados”. Por sua vez este componente está associado com o componente denominado “Acesso aos dados”, responsável pela realização de atividades que colaboram para as atividades de consulta e acesso, cujo resultado é utilizado pelo componente responsável pelo mapeamento dos dados, denominado “Mapeamento dos dados”. Pode ser observado na Figura 23 que este componente está associado com ontologias cuja finalidade é representar a estrutura ou associações dos dados entre si e ao longo de diversas bases. O resultado do componente de mapeamento de dados é disponibilizado em um conjunto de formatos, sendo representado no modelo como o componente “Notação semântica”, que será empregado como entrada para o componente da triplestore. Este componente tem como objetivo armazenar os dados mapeados e integrados, tendo sido considerado no modelo a possibilidade de flexibilidade quanto a este componente, ou seja, o modelo não é dependente de uma triplestore específica. Os dois componentes finais ilustrados no modelo são o componente “Api de consulta” e o componente “Cliente”, que são os responsáveis por permitir o acesso aos dados integrados.

A seguir são detalhadas algumas das características de cada componente citado.

## 4.2 Origem de dados

Dentro do modelo apresentado, o elemento “Origem dos dados” representa todo o conjunto de possíveis bases de dados abertas e conectadas, ou de bases de dados abertos. Este elemento que representa o ponto de partida no modelo está associado com uma das contribuições do trabalho aqui apresentado, que é justamente a capacidade de tratar a grande diversidade observada nestas bases.

A seguir são comentados alguns aspectos neste sentido, oriundos dos estudos e análises realizadas e apresentadas nos capítulos anteriores.

Observa-se que os diversos níveis de governo apresentam avanços, pois existem instrumentos legais regulando a propagação das informações públicas, já havendo, por conta destas iniciativas, dados na web em um grande volume. Contudo, atualmente, a sociedade enfrenta dificuldades em alcançar os dados em formatos adequados para serem analisados, o que é constatado por entidades da área (COMITÊ GESTOR DA INTERNET NO BRASIL, 2011) e também consta na literatura recente sobre o tema (Anthes, 2015; Isotami e Bittencourt, 2015).

Após uma breve análise nos contextos dos portais de governos exposto no item 3.6, observou-se que mesmo os dados sendo abertos e em sua grande parte, disponibilizados em vários formatos conhecidos, como por exemplo, json, csv, geojson, xml, boa parte destes dados não encontra-se disponibilizada com notação semântica. Muitos destes conjuntos de dados são disponibilizados através de *web services* que consomem bases de dados relacionais. Através de experimentações foi constatado também que os dados nem sempre são atualizados constantemente, e isso fica evidente nos conjuntos de dados que disponibilizam os dados em arquivos compactados.

## 4.3 Acesso aos Dados

Este componente visa interligar os dados na sua origem com o restante do processo no modelo proposto. De acordo a origem dos dados utilizada é determinado um conjunto de aspectos do seu formato, o que gera um impacto no seu tratamento pelo modelo, gerando diferentes possíveis destinos.

Se os dados que se deseja interligar estiverem anotados semanticamente, então estes serão enviados diretamente para o *endpoint* e seu armazenamento na *triplestore* para posterior utilização. Caso contrário, os dados são enviados para o

processo de mapeamento, no qual serão tratados e então mapeados conforme uma ontologia previamente desenvolvida para esta finalidade. Depois do mapeamento ser realizado, os dados resultantes serão armazenados no *endpoint* disponibilizado pela *tripelstore*.

#### 4.4 Mapeamento de Dados

O procedimento principal deste módulo consiste em mapear as tabelas dos bancos de dados relacionais, com base numa ontologia pré-definida, para gerar o arquivo com notação semântica ou disponibilizar um serviço para consultar estes dados em SPARQL, agindo como um conversor de dados simultâneo.

No caso da geração do arquivo final, este será importado para um *triplestore*, podendo as consultas futuras serem feitas neste último e não mais nas bases relacionais. Existem algumas recomendações do W3C para este mapeamento, que são o uso da notação R2RML ou então a aplicação do procedimento de Mapeamento Direto.

De acordo com Arenas et al. (2011), no mapeamento direto o arquivo gerado em RDF, ao final do processo, reflete exatamente o nome dos elementos da base de dados relacional e nada poderá ser alterado. Por outro lado, (DAS; SUNDARA; CYGANIAK, 2012), demonstra que utilizando o R2RML pode-se definir visualizações altamente customizadas sobre os dados relacionais, similar ao obtido com o uso de D2RQ.

Segundo Hert, Reif e Gall (2011), D2RQ é uma linguagem de mapeamento de dados relacionais para RDF, assim como o R2RML, porém este último é um resultado proposto por um grupo de trabalho do W3C para padronizar os mapeamentos de bases de dados relacionais para RDF com o objetivo de definir uma linguagem independente de fornecedor, pública. Apesar do R2RML ser a linguagem de mapeamento sugerida pelo W3C, é preciso uma expertise nessa linguagem para poder utilizá-la de forma eficiente para mapear dados relacionais e finalmente exportá-los em formato RDF.



#### 4.5 Ontologia

O modelo define o uso de ontologias de domínio como apoio no processo de mapeamento, justamente para aproveitar o potencial de representação de conhecimento e sua flexibilidade.

A ontologia visa auxiliar no mapeamento dos dados da base relacional de modo a facilitar a interligação dos dados consumidos, com base em algo comum entre si. Servirá de guia para seu armazenamento em *triplestore* e posteriores consultas neste último. Na Web semântica o conceito central adotado está associado com o objetivo de tornar a informação capaz de ser lida e utilizada da forma adequada automaticamente. Nesse contexto, a ontologia passa a ser o principal instrumento de representação do conhecimento armazenado. A estrutura da Web semântica proporciona que as informações sejam armazenadas de maneira organizada e padronizada e, conseqüentemente, recuperada. (BERNERS-LEE; HENDLER, 2001).

No âmbito da Web semântica, as ontologias caracterizam descrições para conceitos, com semântica clara e bem definida, possivelmente interpretável por máquina, que processam de forma automática as informações encontradas na Web. Para esse fim, utilizam as linguagens compatíveis (*OWL*, *RDF*) com o ambiente digital. Para a organização e recuperação da informação, o conceito utilizado é oriundo da proposta de Bernners-Lee e Hendler (2001), e está associado ao contexto da web semântica.

Na concepção do autor, a ontologia é um documento ou arquivo que define rigorosamente as relações entre limites a serem utilizados nas operações de mapeamento. No capítulo 5, que descreve a implementação de um protótipo deste modelo, é descrita em detalhes a ontologia utilizada.

#### 4.6 Notação Semântica

A utilização de uma notação semântica ajuda a superar a confusão da linguagem natural quando expor noções e sua representação computacional em uma linguagem formal. Informando à um computador como os itens de dados estão relacionados e como estas relações podem ser avaliadas automaticamente, torna-se possível processar filtros complexos e operações de busca (ZHANG et al., 2015).

Nesta etapa do fluxo de funcionamento do modelo, um arquivo em RDF é gerado para que posteriormente seja importado para o *triplestore*, ficando assim os seus dados posteriormente disponíveis para consultas via SPARQL.

Este arquivo é gerado com base no mapeamento realizado da base de dados relacional guiado pela ontologia, utilizando a linguagem de mapeamento definida no item 4.4.

#### 4.7 API de Consulta

Conforme já mencionado anteriormente, após ser gerado o arquivo em formato RDF com os dados anotados, estes são armazenados num *triplestore* para consultas utilizando SPARQL via *endpoint* disponibilizado juntamente com o *triplestore*.

Muitos *triplestores* disponibilizam uma API para consultar os dados armazenados utilizando a linguagem SPARQL, entretanto, nesta etapa do modelo visamos disponibilizar não apenas esta API e a ontologia, mas também desenvolver uma aplicação para visualização dos dados interligados de uma forma, visualmente, agradável. Tal aplicação será desenvolvida com o intuito de facilitar o acesso aos dados interligados, pelo usuário.

#### 4.8 Análise do Modelo

A web possui uma variedade enorme de tipos de dados, mas estes não estão interligados, sendo que a diferença dos modelos dificulta a sua integração. Existem conexões entre as páginas, mas não existem ligações entre os dados. É preciso conhecer as diferentes APIs de acesso e os diferentes modelos de dados para poder combiná-los.

Este modelo propõe uma forma de auxiliar na interligação de dados que ainda não foram anotados semanticamente, propondo uma interligação de dados com algo em comum entre as bases que se deseja interligar. O desenvolvimento de uma ontologia para tal interligação se faz necessário, tendo em vista sua grande utilidade no mapeamento e posteriormente nas consultas SPARQL.

Como diferencial, este modelo apresenta o uso de dados conectados, que segundo Heath e Bizer (2011), nada mais é que um conjunto das melhores práticas

para publicação e interligação de dados estruturados na web, permitindo estabelecer ligações entre itens de fontes de dados variadas para construir um único espaço de dados universal. A implementação deste modelo proporciona facilidades para a integração de dados e para a geração de aplicações de consumo dos mesmos, com os ganhos advindos da integração de fontes diversas.

No capítulo seguinte será descrita em detalhes uma implementação deste modelo realizada com a finalidade de proporcionar sua avaliação.



## 5 PROTÓTIPO IMPLEMENTADO

Neste capítulo são apresentadas as estruturas que ilustram uma implementação do modelo proposto e permitem uma avaliação preliminar do mesmo. O primeiro ponto a ser definido está relacionado com as fontes de dados que se deseja tratar.

Haja vista a enorme quantidade de dados publicados referentes ao assunto das compras governamentais, tal como a forma como são divulgados, nota-se a necessidade de incrementar tais dados com anotações semânticas para poderem ser relacionadas, inferidas e manipuladas por aplicações que auxiliem o cidadão, com uma visualização amigável, à consultar os gastos públicos.

Para encaminhar uma solução para esta questão, esta proposta consiste em consumir a API de Compras Governamentais<sup>20</sup>, interligar com o orçamento federal disponibilizado, anotar os dados com RDF e disponibilizar em formato aberto conforme recomendações da Web Semântica<sup>21</sup>.

Desse modo foi projetada uma aplicação semântica para publicações de dados das compras governamentais, dentro dos padrões estabelecidos pela W3C.

### 5.1 Ontologia de domínio desenvolvida

A estratégia adotada na implementação deste protótipo foi de criar uma ontologia que interligue os dados abertos do orçamento federal com os de compras governamentais, com o objetivo de gerar conhecimento por meio da verificação das ações do governo de forma que possibilite realizar uma avaliação dos resultados alcançados dentro do tempo e, como resultado, revisar o uso para uma melhor destinação dos recursos públicos, podendo inclusive cruzar estes dados com os investimentos, por exemplo, na educação, em tecnologia da informação, entre outras áreas.

A ontologia desenvolvida baseou-se no estudo preliminar dos dados disponibilizados, buscando identificar pontos de associação semântica entre estes dois conjuntos. Sem o uso de especialistas de domínio, seu desenvolvimento foi totalmente ferramental e baseou-se em estudos de casos similares de mapeamento de

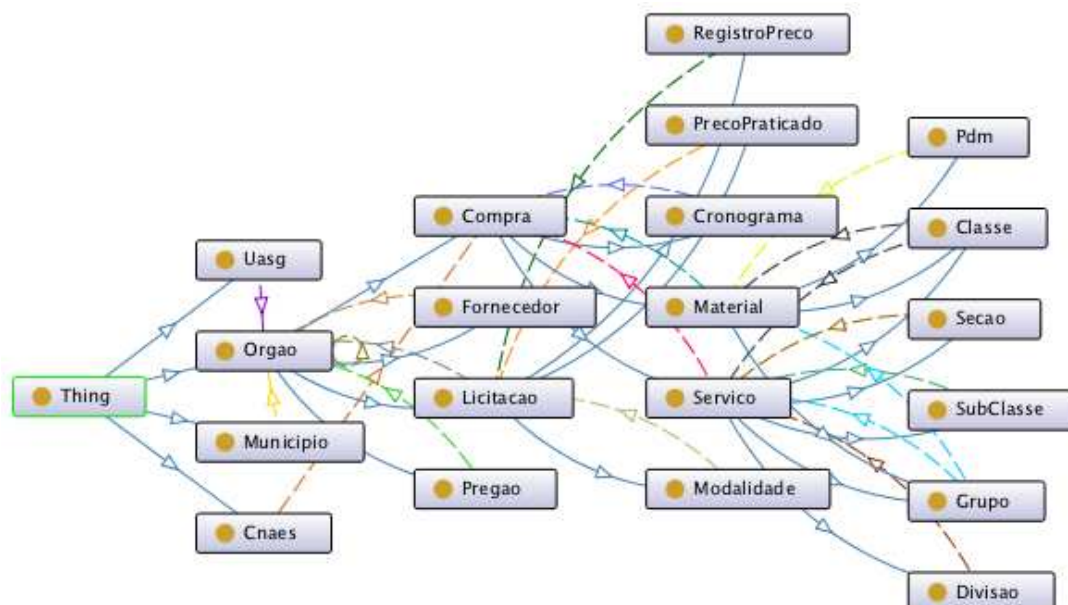
---

<sup>20</sup> <http://compras.dados.gov.br>

<sup>21</sup> <http://www.w3c.br/Padroes/WebSemantica>

dados de bases de dados relacionais. A seguir é descrita a ontologia desenvolvida e seu detalhamento.

Figura 24 - Ontologia de Compras desenvolvida



Fonte: próprio autor

Na Figura 24 os conceitos identificados, nas bases de dados de compras governamentais, são descritos como classes OWL, representadas pelos retângulos, suas interligações com subclasses, pelas setas contínuas e suas propriedades, pelas setas tracejadas.

A descrição das classes da ontologia proposta, são descritas na Tabela 5.

Tabela 5 - Relação de Classes e seus Significados

<b>Classe</b>	<b>Descrição</b>
Uasg	Os elementos dessa classe referem-se às Unidades Administrativas de Serviços gerais ou Unidades Cadastradoras.
Órgão	Os Elementos dessa classe referem-se ao Órgão responsável pela compra.
Município	Os elementos dessa classe referem-se aos dados de todos os municípios cadastrados.
Cnaes	Os elementos dessa classe referem-se aos dados da Classificação Nacional de Atividade Econômica.
Compra	Os Elementos dessa classe correspondem às compras reali-

	zadas.
Fornecedor	Os elementos dessa classe referem-se ao ganhador da licitação que originou a compra.
Licitação	Os elementos dessa classe referem-se às licitações realizadas pelo governo federal.
Pregão	Os elementos dessa classe referem-se aos pregões realizados pelo governo federal que resultam nas licitações.
Cronograma	Os elementos dessa classe referem-se ao cronograma cadastrado para cada compra realizada de uma licitação.
Material	Os elementos dessa classe referem-se ao material entregue pelo fornecedor.
Serviço	Os elementos dessa classe referem-se ao serviço entregue pelo fornecedor.
Registro de Preço	Os elementos dessa classe referem-se aos registros de preços realizados pelo governo federal.
Preço Praticado	Os elementos dessa classe referem-se aos preços praticados nas licitações realizadas pelo governo federal.
Modalidade	Os elementos dessa classe referem-se às modalidades de licitação.
Classe	Os elementos dessa classe referem-se às classes de materiais e serviços cadastrados no catálogo de materiais.
Seção	Os elementos dessa classe referem-se às seções informadas dos serviços.
SubClasse	Os elementos dessa classe referem-se às subclasses de serviços cadastrados no catálogo de materiais.
Grupo	Os elementos dessa classe referem-se aos grupos de materiais e serviços cadastrados no catálogo de materiais.
Divisão	Os elementos dessa classe referem-se às divisões informadas dos serviços.

Fonte: próprio autor

Já com relação às propriedades dos objetos, os seus detalhes estão descritos na Tabela 6.

Tabela 6 - Propriedades de Objetos

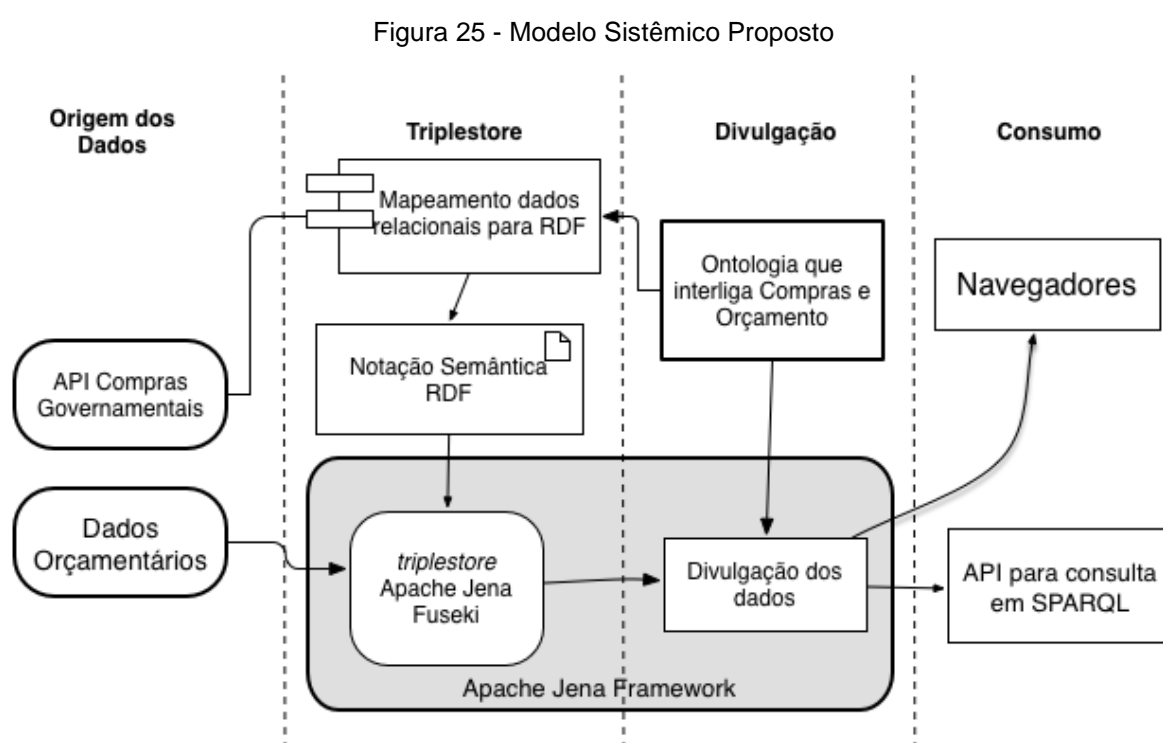
<b>Propriedade</b>	<b>Descrição</b>
TemOrcamento	Relaciona a classe "Compra" com "Orcamento".
TemOrgao	Relaciona a classe "Compra" com "Orgao".
TemFornecedor	Relaciona a classe "Compra" com "Fornecedor".
TemMaterial	Relaciona a classe "Compra" com "Material".
TemServico	Relaciona a classe "Compra" com "Servico".
TemClasse	Relaciona as classes "Servico, Pdm, SubClasse e Material" com "Classe"
TemCnae	Relaciona a classe "Fornecedor" com "Cnae"
temDivisao	Relaciona as classes "Grupo e Servico" com "Divisao"
temGrupo	Relaciona as classes "Classe, Material e Servico" com "Grupo"
temLicitacao	Relaciona as classes "RegistroPreco, PrecoPraticado e Compra" com "Licitacao"
temModalidade	Relaciona as classes "Licitacao, RegistroPreco, Cronograma, PrecoPraticado e Compra" com "Modalidade"
temMunicipio	Relaciona as classes "Fornecedor e Orgao" com "Municipio"
temPdm	Relaciona a classe "Material" com "Pdm"
temSecao	Relaciona as classes "Divisao e Servico" com "Secao"
temSubClasse	Relaciona a classe "Servico" com "SubClasse"
temUasg	Relaciona as classes "Pregao, Licitacao, Fornecedor, RegistroPreco, Compra, PrecoPraticado e Cronograma" com "Uasg"

Fonte: próprio autor



## 5.2 Detalhamento do modelo neste protótipo

O modelo proposto no capítulo 4 pode ser ajustado de acordo com alguns de seus componentes, tais como a quantidade e variedade de fontes de dados, o tipo de *triplestore* utilizada, o conjunto de ontologias empregado. Desta forma, pode-se considerar que a implementação de um protótipo consiste em uma instanciação deste modelo, que neste caso é descrita na Figura 25. Para a descrição do protótipo, foi utilizada a notação TAM/FMC (SAP AG, 2007) e o mesmo pode ser observado na figura 25.



Fonte: próprio autor

Na Figura 25 é possível visualizar a origem dos dados como sendo composta por dois elementos, descritos anteriormente, ou seja, o conjunto de dados de Compras governamentais e o conjunto de dados orçamentários. O primeiro está sendo acessado com base em uma API de acesso aos dados de Compras Governamentais e o segundo não dispõe deste recurso, no caso dos Dados Orçamentários. Este aspecto do protótipo visa exemplificar as possibilidades de utilização com bases de dados diversas.

Ainda com relação à figura 25, observa-se o processo de anotação semântica, através do qual os dados das Compras Governamentais e os dados orçamentários são armazenados em um *triplestore*, para posteriormente serem divulgados com base em uma ontologia, tal como proposta no item 5.1, para poderem ser consumidos por sistemas clientes ou por navegadores web.

O *triplestore* selecionado para este protótipo foi o Fuseki, que também disponibiliza uma ferramenta para consulta em SPARQL.

A API de compras governamentais foi escolhida como fonte de dados para este trabalho pelo fato de não possuir uma ontologia associada com a descrição de seus dados e também por não disponibilizar os dados anotados semanticamente.

O conjunto de dados contendo os dados orçamentários do governo federal já se encontram anotados semanticamente e possuem uma ontologia própria, sendo que por esse fato ela foi escolhida como fonte de dados para efetuar a interligação de dados com as compras governamentais.

Pelo fato dos dados do orçamento já estarem anotados, neste trabalho os mesmos já foram importados diretamente para o *triplestore*, enquanto os dados das compras governamentais passaram por um mapeamento com base numa ontologia também desenvolvida neste trabalho. Tal ontologia contém classes e propriedades que refletem os dados das compras governamentais e também funciona como guia para consultas de dados no *endpoint* utilizando SPARQL.

Para divulgação dos dados, o *triplestore* dispõe de um *endpoint* para executar consultas SPARQL e também foi desenvolvida uma aplicação para visualização dos dados utilizando um *framework* para comunicação com o *endpoint*, acessível por qualquer navegador no computador e até mesmo nos *smartphones*.

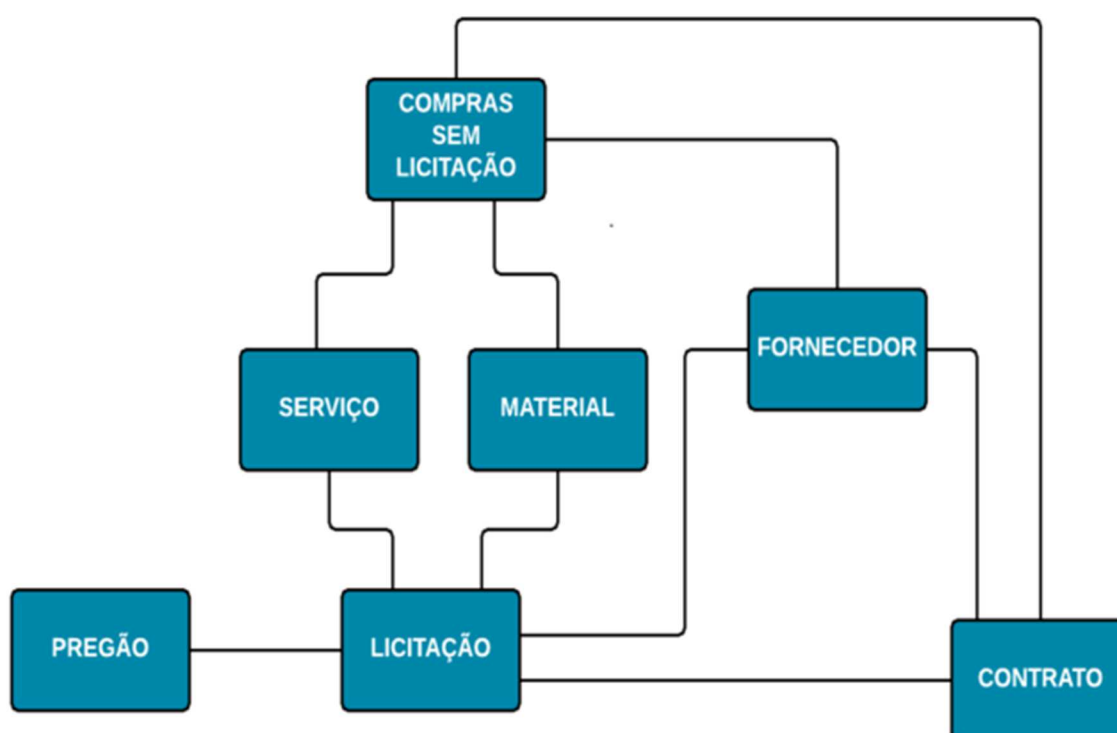
A seguir são apresentados mais detalhes das fontes de dados e dos processos de mapeamento, armazenamento e consulta.

### 5.3 API Compras Governamentais

Disponibilizada pelo governo federal para consulta pública, a base de dados com os dados de Compras governamentais contém os dados do Sistema Integrado de Administração e Serviços Geras – SIASG. Este sistema é composto por vários módulos que contemplam o catálogo de materiais e serviços, as intenções e divulgações de licitações, as intenções de registros de preços, cadastramento de forne-

cedores, a realização e resultado das licitações, o empenhos de pagamentos e o registro e gestão dos contratos, todos estes módulos resultam na operacionalização das compras do Governo Federal. Apesar de serem bastante completos neste aspecto, estes dados não estão anotados semanticamente. Atualmente estes dados estão disponíveis de forma aberta, a partir da seguinte URL: <http://compras.dados.gov.br/docs/home.html>. Um modelo de dados disponibilizado e maiores detalhes de sua utilização no protótipo são descritos a seguir.

Figura 26 - Modelo de Dados Compras Governamentais



Fonte: (“API Compras Governamentais”, 2014)

Pode-se verificar o modelo de dados utilizado pela API na Figura 26, onde constam as interligações entre os dados das compras e contratações firmadas pelo poder executivo federal mantidos pelo Sistema Integrado de Administração e Serviços Gerais – SIASG. Segue a lista de dados disponibilizados pela API:

- Dados dos fornecedores;
- Dados do catálogo de serviços;
- Dados do catálogo de materiais;
- Dados de licitações;
- Dados de contratos, que são as compras propriamente ditas;

- Dados de compras sem licitação;

Os dados são disponibilizados sob a licença *Open Database License* - OBDL<sup>22</sup>, todos os direitos dos conteúdos individuais estão licenciados sob *Database Content License* - DBCL<sup>23</sup>. Atualmente os dados são divulgados como sendo atualizados mensalmente, entretanto sua última atualização disponibilizada até o momento da divulgação deste trabalho data de março de 2014.

Figura 27 - Dados do Órgão Instituto Federal de Roraima

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<resource count="5" offset="0">
  <_links>
    <link href="http://compras.dados.gov.br/licitacoes/v1/uasgs.xml?id_orgao=26437"
      rel="self" title="Lista de uasgs filtrados por id_orgao=26437."/>
    <link href="/licitacoes/v1/uasgs.xml?id_orgao=26437" rel="first" title="
      Primeira página"/>
  </_links>
  <_embedded>
    <resource>
      <_links>
        <link href="/licitacoes/id/uasg/158152" rel="self" title="UASG 158152:
          INST.FED.DE EDUC.CIENC.E TEC.DE RORAIMA"/>
        <link href="/fornecedores/v1/fornecedores?id_unidade_cadastradora=158152
          " rel="fornecedores" title="Fornecedores cadastrado nesta UASG"/>
        <link href="/licitacoes/v1/licitacoes?uasg=158152" rel="licitacoes"
          title="Licitações realizadas por esta UASG"/>
        <link href="/licitacoes/v1/irps?uasg=158152" rel="irps" title="IRPs
          realizadas por esta UASG"/>
        <link href="/contratos/v1/contratos?uasg=158152" rel="contratos" title="
          Contratos feitas por esta UASG"/>
        <link href="
          /fornecedores/v1/ocorrencias_fornecedores?id_unidade_cadastradora=158152
          " rel="ocorrencias_fornecedores" title="Ocorrências cadastradas por
          esta UASG"/>
        <link href="/licitacoes/id/orgao/26437" rel="orgao" title="Orgão 26437:
          INST.FED.DE EDUC.,CIENC.E TEC. DE RORAIMA"/>
        <link href="/fornecedores/id/municipio/3018" rel="municipio" title="
          Municipio 3018: Boa Vista"/>
      </_links>
      <id>158152</id>
      <nome>INST.FED.DE EDUC.CIENC.E TEC.DE RORAIMA</nome>
      <id_orgao>26437</id_orgao>
      <id_municipio>3018</id_municipio>
      <cep>69303220</cep>
      <total_fornecedores_cadastrados>504</total_fornecedores_cadastrados>
      <total_fornecedores_recadastrados>206</total_fornecedores_recadastrados>
      <unidade_cadastradora>true</unidade_cadastradora>
      <ativo>true</ativo>
    </_embedded>
  </resource>

```

Fonte: [http://compras.dados.gov.br/licitacoes/v1/uasgs.xml?id\\_orgao=26437](http://compras.dados.gov.br/licitacoes/v1/uasgs.xml?id_orgao=26437)

<sup>22</sup> <http://opendatacommons.org/licenses/odbl/1.0/>

<sup>23</sup> <http://opendatacommons.org/licenses/dbcl/1.0/>

O acesso aos dados é feito através de URLs, e os resultados das consultas são disponibilizados em formato XML, JSON ou CSV, sendo também possibilitada uma visualização em formato HTML. Em cada consulta é possível especificar uma série de parâmetros que agem como filtros, que devem compor a URL.

Um exemplo de como os dados são disponibilizados é visto na Figura 27, onde são exibidos os dados de um órgão específico. Na figura 27 é de fácil identificação os campos *nome*, *id\_orgao*, *id\_municipio*, *cep*, *total\_fornecedores\_cadastrados*, *total\_fornecedores\_recadastrados*, *unidade\_cadastradora*, *ativo*, que estão disponíveis para acesso. Estes campos poderão ser mapeados para se tornarem propriedades de uma classe da ontologia que interligará estes dados com os do Orçamento Federal.

Para essa consulta exemplificada com o resultado da figura 27, foi utilizada a seguinte URL como ponto de partida:

*http://compras.dados.gov.br/licitacoes/v1/uasgs.{formato}?{parametro1=valor1}&{parametro2=valor2}&{parametroN=valorN}*

A análise dos campos desta URL permite identificar os seguintes elementos:

- *uasgs* é o nome do módulo;
- *formato* é o tipo de saída da resposta, neste caso XML ;
- *parametro1* utilizado como o atributo *id\_orgao=26437* que corresponde ao Instituto Federal de Roraima.

Por se tratar de uma API, a resposta das consultas é personalizada com links para acessar demais dados correspondentes a este órgão, sendo que vale lembrar que estes são links fixos acrescidos apenas do *id\_orgao* como filtro. No caso do link ser modificado, será necessária uma atualização para acesso.

#### 5.4 API Dados Orçamentários

Esta API foi disponibilizada com o intuito de dar maior transparência e acesso para que qualquer cidadão, com conhecimento em informática, interessado em conhecer melhor os dados do orçamento federal possam fazer pesquisas rápidas e eficientes.

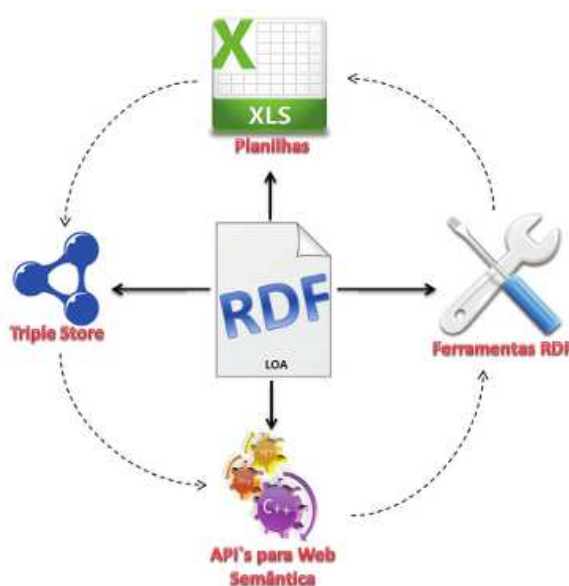
A partir desta definição surgiu então o primeiro produto, resultado do projeto “Governo Federal em Formato Aberto”. Este consiste na publicação dos dados da despesa do Orçamento Geral da União, no padrão de dados abertos definido pelo

W3C. Conta com uma ontologia própria, uma página para consulta em SPARQL, além de disponibilizar os dados para *download* em arquivo para consultas locais (“API Orçamento Federal”, 2012).

O orçamento da União para o exercício de 2012 possui um volume de 2.645 páginas em PDF, o que torna uma pesquisa rápida inviável. A divulgação dos dados possibilita que um cidadão com conhecimentos técnicos possa desenvolver aplicações para fazer análises e até cruzamento de dados de forma veloz e eficiente.

A possibilidade mais comum é o uso de *triplestores* para persistência de dados em formato RDF similar ao armazenamento de bases de dados relacionais. A partir do *triplestore* e de um *endpoint*, uma consulta SPARQL pode ser realizada na base e o resultado recuperado em uma planilha, em um formato XML, em formato JSON, no próprio formato RDF, entre outros. A Figura 28 demonstra as possibilidades de uso dos dados abertos.

Figura 28 - Possibilidades de Uso do RDF



Fonte: (“API Orçamento Federal”, 2012)

A Figura 29 demonstra um exemplo de consulta na API onde são informados os itens de despesa para o ano de 2015 como resultado. Esta consulta pode ser executada na API de consulta SPARQL do Orçamento Federal<sup>24</sup>.

<sup>24</sup> <http://orcamento.dados.gov.br/sparql/>

Figura 29 - Listagem de Itens de Despesa

```
PREFIX loa: <http://vocab.e.gov.br/2013/09/loa#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT ?item
WHERE {
?item loa:temExercicio [loa:identificador 2015]. } OFFSET 0 LIMIT 6
```

Fonte: <http://orcamento.dados.gov.br/sparql>

Como resultado desta consulta, são exibidos 6 itens de despesa do ano de 2015, demonstrados na Tabela 7, resultado obtido na consulta diretamente no *endpoint* que a API disponibiliza.

Tabela 7 - Resultado da Pesquisa Item de Despesa

ITEM
<a href="http://orcamento.dados.gov.br/id/2015/ItemDespesa/305423">http://orcamento.dados.gov.br/id/2015/ItemDespesa/305423</a>
<a href="http://orcamento.dados.gov.br/id/2015/ItemDespesa/305424">http://orcamento.dados.gov.br/id/2015/ItemDespesa/305424</a>
<a href="http://orcamento.dados.gov.br/id/2015/ItemDespesa/305425">http://orcamento.dados.gov.br/id/2015/ItemDespesa/305425</a>
<a href="http://orcamento.dados.gov.br/id/2015/ItemDespesa/305426">http://orcamento.dados.gov.br/id/2015/ItemDespesa/305426</a>
<a href="http://orcamento.dados.gov.br/id/2015/ItemDespesa/305427">http://orcamento.dados.gov.br/id/2015/ItemDespesa/305427</a>
<a href="http://orcamento.dados.gov.br/id/2015/ItemDespesa/305428">http://orcamento.dados.gov.br/id/2015/ItemDespesa/305428</a>

Fonte: próprio autor

Este resultado também pode ser obtido através da ferramenta disponibilizada pela API de Orçamento<sup>25</sup>, para fácil navegabilidade pelos dados. Pode-se visualizar na Figura 30, de uma forma mais amigável, todos os dados da consulta anterior, mas com uma navegabilidade empregada e demais dados relacionados aos itens de despesa.

Esta página conta com um formulário para consulta de qualquer dado relacionado ao item, links para ordenação do resultado conforme as propriedades da classe a qual se referencia o item na ontologia<sup>26</sup> do orçamento.

<sup>25</sup> <http://orcamento.dados.gov.br/doc/2015/ItemDespesa>

<sup>26</sup> <http://vocab.e.gov.br/2013/09/loa.owl>

Figura 30 - Ferramenta de Visualização do Orçamento

orcamento.dados.gov.br

html | json | rdf | text | ttl | xml

### Resultados da Busca

Exibir Formulário de Busca

**305423**

valor dotacao inicial	0
valor empenhado	389.53
valor lei mais credito	0
valor liquidado	389.53
valor pago	389.53
valor projeto lei	0

tem acao: Pagamento de Pessoal Ativo da União

codigo: 20TP

type: Acao, Atividade

tem categoria economica: Despesas Correntes

codigo: 3

tem elemento despesa: Contribuição a Entidades Fechadas de Previdência

codigo: 07

tem esfera: Orçamento da Seguridade Social

codigo: 20

tem exercicio: 2015

tem fonte recursos: Recursos Destinados à Manutenção e Desenvolvimento do Ensino

codigo: 112

### Nesta Página

- > 305423
- > 305424
- > 305425
- > 305426
- > 305427
- > 305428
- > próxima >

### Ordenar por

- valor dotacao inicial
- valor empenhado
- valor lei mais credito
- valor liquidado
- valor pago
- valor projeto lei
- tem acao > codigo
- tem acao > label
- tem acao > type
- tem categoria economica > codigo
- tem categoria economica > label
- tem elemento despesa > codigo
- tem elemento despesa > label
- tem esfera > codigo
- tem esfera > label
- tem exercicio
- tem fonte recursos > codigo
- tem fonte recursos > label
- tem funcao > codigo
- tem funcao > label
- tem GND > codigo
- tem GND > label
- tem identificador uso > codigo
- tem identificador uso > label
- tem modalidade aplicacao >

Fonte: <http://orcamento.dados.gov.br/doc/2015/ItemDespesa>

## 5.5 Notação RDF

Neste processo foi realizado o mapeamento dos campos do modelo relacional da API de Compras Governamentais, que vieram a ser as propriedades da classe no arquivo RDF final.

A Figura 31 reflete um pequeno trecho do arquivo de mapeamento e é possível visualizar que o mesmo já utiliza a linguagem D2RQ e também faz uso da ontologia criada no item 5.1.

Na figura 31, o uso da ontologia é destacado na linha 9. Outro elemento importante é a informação da origem dos dados, feita nas linhas 19 à 26, trecho onde também é personalizado um conjunto de campos (ID e DESCRIÇÃO) da tabela CNAES com base na classe Cnaes e propriedades (Codigo e Nome) da ontologia.



Dessa forma, todos os dados encontrados na base de dados relacional serão exportados, de forma personalizada, para um arquivo em formato RDF baseado na ontologia.

Figura 31 – Trecho do Mapeamento de Dados

```

1  @prefix map: <#> .
2  @prefix db: <> .
3  @prefix vocab: <vocab/> .
4  @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
5  @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
6  @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
7  @prefix d2rq: <http://www.wiwiss.fu-berlin.de/suhl/bizer/D2RQ/0.1#> .
8  @prefix jdbc: <http://d2rq.org/terms/jdbc/> .
9  @prefix : <http://sistemas.ifrr.edu.br/fuseki/compras.owl#> .
10 @prefix d2r: <http://sites.wiwiss.fu-berlin.de/suhl/bizer/d2r-server/config.rdf#> .
11
12 <> a d2r:Server;
13   d2r:baseURI <http://localhost:2020/>;
14   d2r:port 2020;
15   d2r:limitPerClassMap false;
16   d2r:limitPerPropertyBridge false .
17
18
19 map:database a d2rq:Database;
20   d2rq:jdbcDriver "com.mysql.jdbc.Driver";
21   d2rq:jdbcDSN "jdbc:mysql://localhost:3306/prototipo";
22   d2rq:username "root";
23   d2rq:password "";
24   jdbc:autoReconnect "true";
25   jdbc:zeroDateTimeBehavior "convertToNull";
26   .
27
28 # Table cnaes
29 map:cnaes a d2rq:ClassMap;
30   d2rq:dataStorage map:database;
31   d2rq:uriPattern "Cnaes/@cnaes.id|urlify@";
32   d2rq:class :Cnaes;
33   d2rq:classDefinitionLabel "cnaes";
34   .
35 map:cnaes__label a d2rq:PropertyBridge;
36   d2rq:belongsToClassMap map:cnaes;
37   d2rq:property rdfs:label;
38   d2rq:pattern "Cnaes #@cnaes.id@";
39   .
40 map:cnaes_id a d2rq:PropertyBridge;
41   d2rq:belongsToClassMap map:cnaes;
42   d2rq:property :Codigo;
43   d2rq:propertyDefinitionLabel "cnaes id";
44   d2rq:column "cnaes.id";
45   .
46 map:cnaes_descricao a d2rq:PropertyBridge;
47   d2rq:belongsToClassMap map:cnaes;
48   d2rq:property :Nome;
49   d2rq:propertyDefinitionLabel "cnaes descricao";
50   d2rq:column "cnaes.descricao";
51   .

```

Ontologia

Origem dos Dados

Mapeamento de uma tabela

Fonte: próprio autor

Ao final dessa etapa, um arquivo em formato RDF foi gerado contendo todos os dados, da base relacional, anotados semanticamente. Na Figura 32 é apresentado um trecho do arquivo RDF resultado do mapeamento dos dados das Compras Governamentais, onde na linha 7 é identificada a ontologia utilizada na etapa do mapeamento dos dados e nas demais linhas são identificados os dados reais com sua correspondência nas classes e propriedades da ontologia de Compras.

Figura 32 - Trecho do arquivo RDF

```

1 <?xml version="1.0"?>
2 <rdf:RDF
3   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
4   xmlns:vocab="http://sistemas.ifrr.edu.br/compras/vocab/"
5   xmlns:map="http://sistemas.ifrr.edu.br/compras/"
6   xmlns:owl="http://www.w3.org/2002/07/owl#"
7   xmlns="http://sistemas.ifrr.edu.br/fuseki/compras.owl#"
8   xmlns:d2r="http://sites.wiwiiss.fu-berlin.de/suhl/bizer/d2r-server/config.rdf#"
9   xmlns:db="http://sistemas.ifrr.edu.br/compras/"
10  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
11  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
12  xml:base="http://sistemas.ifrr.edu.br/compras/" >
13  <rdf:Description rdf:about="PrecoPraticado/456">
14    <Codigo rdf:resource="Licitacao_modalidade/5"/>
15    <temLicitacao rdf:resource="Licitacao/15408005000022005"/>
16    <Status>Publicado</Status>
17    <Nome>Material de Consumo</Nome>
18    <NumeroAviso>22005</NumeroAviso>
19    <rdfs:label>PrecoPraticado #456</rdfs:label>
20    <Codigo>456</Codigo>
21    <ValorTotal>255095</ValorTotal>
22    <rdf:type rdf:resource="/fuseki/compras.owl#PrecoPraticado"/>
23    <temUasg rdf:resource="Uasg/154080"/>
24  </rdf:Description>
25  <rdf:Description rdf:about="Pdm/16344">
26    <Codigo rdf:resource="Classe/6505"/>
27    <rdfs:label>Pdm #16344</rdfs:label>
28    <Nome>ÓXIDO DE ZINCO</Nome>
29    <Codigo>16344</Codigo>
30    <rdf:type rdf:resource="/fuseki/compras.owl#Pdm"/>
31  </rdf:Description>

```

Fonte: próprio autor

## 5.6 Aplicação de consulta desenvolvida

Utilizando o Framework Jena, uma aplicação web com páginas dinâmicas em linguagem java J2EE, para ser utilizada através de um navegador localmente ou em uma rede local de computadores, foi desenvolvida com o fim de visualizar os dados interligados das compras governamentais com o orçamento federal, ambos sendo consultados no *endpoint* Fuseki via consultas SPARQL.

A Figura 33 demonstra a tela inicial da aplicação onde existem, em uma primeira ocasião, os dados de todos os órgãos federais disponíveis, com filtros no ID UASG, nome do órgão, nome do município e na unidade federativa, dessa forma o usuário poderá, facilmente, encontrar o órgão sobre o qual que deseja filtrar as licitações.

Figura 33 - Listagem de Órgãos

Compras Governamentais + Orçamento Federal

Órgãos

ID Uasg	Uasg	ID Órgão	Órgão	Município	UF	
1581						
158142	INST.FED.DE EDUC.,CIENC.E TEC.DO AMAZONAS	26403	INST.FED.DE EDUC., CIENC E TEC.DO AMAZONAS	Manaus	AMAZONAS	<a href="#">Licitações</a>
158142	INST.FED.DE EDUC.,CIENC.E TEC.DO AMAZONAS	Federal	INST.FED.DE EDUC., CIENC E TEC.DO AMAZONAS	Manaus	AMAZONAS	<a href="#">Licitações</a>
158152	INST.FED.DE EDUC.CIENC.E TEC.DE RORAIMA	26437	INST.FED.DE EDUC.,CIENC.E TEC. DE RORAIMA	Boa Vista	RORAIMA	<a href="#">Licitações</a>
158152	INST.FED.DE EDUC.CIENC.E TEC.DE RORAIMA	Federal	INST.FED.DE EDUC.,CIENC.E TEC. DE RORAIMA	Boa Vista	RORAIMA	<a href="#">Licitações</a>

Licitações +

Compras +

Compras com Orçamento +

Fonte: próprio autor

Ao clicar no botão “Licitações” na listagem de órgãos, o sistema executa uma consulta SPARQL usando o atributo id uasg do órgão para filtrar as licitações disponíveis no *endpoint*. Em seguida, todas as licitações são listadas, como apresenta a Figura 34.

Figura 34 - Listagem de Licitações de um Órgão

Compras Governamentais + Orçamento Federal

Órgãos +

Licitações -

Nome	Identificador	Responsável	Processo	Modalidade	
Pregão Eletrônico O objeto da presente licitação é o registro de preços para eventual contratação para os serviços de reprografia, com fornecimento de cópias, encadernações e plastificação.	15815205000102014	LIDYOMARA ALVES SILVA BARBOSA	23231000421201345	PREGÃO	<a href="#">Compras</a>

Compras +

Compras com Orçamento +

Fonte: próprio autor

A listagem exibida com todas as licitações do órgão selecionado é apresentada para o usuário com filtros em seu nome e no responsável por ela, nesse caso, um servidor do órgão, como é ilustrado na Figura 34. Ao filtrar e escolher uma licitação, é possível visualizar todas as compras realizadas referentes àquela licitação.

A Figura 35 demonstra claramente o resultado com o valor de cada compra e no rodapé, um somatório de todas as compras da licitação escolhida pelo usuário.

Figura 35 - Relação de compras de uma licitação

Compras Governamentais + Orçamento Federal								
Órgãos +								
Licitações +								
Compras -								
Número	CNPJ	Identificador	Processo	Valor Inicial	Data Assinatura	ID Uasg	ID Órgão	
0	07217926000182	15815250000192014	23231000421201345	R\$ 48.251,00	2014-05-07	158152	26437	Compras
102014	07217926000182	15815250000192014	23231000421201345	R\$ 48.251,00	2014-05-07	158152	26437	Compras
192014	07217926000182	15815250000192014	23231000421201345	R\$ 48.251,00	2014-05-07	158152	26437	Compras
0	07217926000182	15815250000192014	23231000421201345	R\$ 48.251,00	2014-05-07	158152	Federal	Compras
102014	07217926000182	15815250000192014	23231000421201345	R\$ 48.251,00	2014-05-07	158152	Federal	Compras

<< < 1 2 > >>

Valor Total de Compras desta Licitação: R\$ 289.506,00

Fonte: próprio autor

Ao selecionar uma das compras listadas, serão filtrados todos os dados referentes à essa compra, juntamente com o orçamento disponibilizado para o grupo de natureza da despesa número 4 – investimentos, como podemos ver na Figura 36.

Figura 36 - Dados da compra interligada com Orçamento Federal

## Compras Governamentais + Orçamento Federal

Órgãos								+
Licitações								+
Compras								+
Compras com Orçamento								-
Nome	CNPJ	Identificador	Licitação Ident.	Processo	Valor Inicial	ID Uasg	Orçamento Anual	
O objeto do presente Termo de Contrato e a aquisição de material de consumo para atender a Coordenacao de Comunicacao Social CCS.	01259682000114	158152500002320	158152050010720	232310003762013	R\$ 3.608,00	158152	R\$ 18.710.367,00	

Fonte: próprio autor

Ao final de todo o procedimento na aplicação, o usuário consegue visualizar todos os dados da compra selecionada juntamente com o orçamento total para investimentos utilizado pelo órgão no ano da compra. Também é possível notar que a compra mostrada na Figura 36 foi realizada no ano de 2014, enquanto sua licitação data de 2013, como nota-se nos últimos 4 dígitos do identificador da licitação selecionada na Figura 34. Nesse caso, o orçamento mostrado é o do ano da compra, ou seja, 2014.



## 6 AVALIAÇÃO

A avaliação do modelo foi realizada a partir de três experimentações. O objetivo das atividades de avaliação foi demonstrar a viabilidade das operações de integração de dados em bases diferentes, com base no protótipo implementado.

A primeira atividade consistiu em uma avaliação baseada em estudos de cenários, com o intuito de demonstrar o funcionamento adequado segundo as características do modelo proposto. A segunda tem a função de verificar a percepção de facilidade de uso pelos usuários e sua percepção de utilidade. Esta avaliação tem como foco a aplicação desenvolvida e funcionalidades propostas pelo modelo. Por último, foi realizada uma avaliação da ontologia desenvolvida para publicação dos dados das Compras Governamentais e posterior utilização na interligação com os dados do Orçamento Federal.

### 6.1 Avaliação por Cenários

Cenário 1 – Visualização dos dados de uma compra de uma certa licitação do Instituto Federal de Roraima – IFRR.

Este cenário visa demonstrar a utilização do protótipo em uma situação onde um usuário navega pelos dados através de filtros até encontrar a licitação desejada e listar suas compras.

Descrição do cenário: *“Um usuário qualquer deseja listar todas as compras de uma licitação do ano de 2014 no órgão Instituto Federal de Roraima – IFRR para saber quando elas foram efetivamente realizadas e visualizar o somatório de seus valores pagos.”*

Para cumprir essa tarefa, o usuário deve abrir o navegador no computador e acessar o endereço eletrônico da aplicação desenvolvida junto do protótipo. Após isso, será exibida uma lista com todos os órgãos federais disponíveis no *triplestore*, assim como exibido na Figura 33.

Realizando o filtro pelo nome do órgão, município ou Estado, o usuário poderá localizar o órgão que deseja e poderá clicar no botão “Licitações”, o qual irá executar um método que irá listar todas as licitações disponíveis. Na Figura 34, pode-se visualizar este resultado e um exemplo de filtragem para o ano de 2014.

Ao localizar a licitação desejada, o usuário deverá clicar no botão “Compras”, que irá executar um método que efetua uma consulta SPARQL no *endpoint* e retorna uma lista com todas as compras da licitação selecionadas anteriormente, como pode-se ver na Figura 35 onde, inclusive, o somatório de todas as compras é informado no rodapé da listagem.

As consultas, em SPARQL, utilizadas pela aplicação para listar as compras da licitação selecionada e executar o somatório são exibidas na Figura 37 e Figura 38, respectivamente.

Figura 37 - Consulta de Compras de uma Licitação

```
PREFIX c: <http://sistemas.ifrr.edu.br/fuseki/compras.owl#>

SELECT ?nome ?numero ?cnpj ?identificador ?processo ?valorInicial
       ?fundamento ?dataAss ?idUasg ?idOrgao
WHERE {
  ?compra c:Nome ?nome . ?compra c:Numero ?numero . ?compra c:CNPJ ?cnpj .
  ?compra c:Identificador ?identificador . ?compra c:Processo ?processo .
  ?compra c:temLicitacao ?licitacao . ?compra c:ValorInicial ?valorInicial .
  ?compra c:FundamentoLegal ?fundamento . ?compra c:DataAssinatura ?dataAss .
  ?compra c:temUasg ?uasg . ?uasg c:Codigo ?idUasg .
  ?uasg c:temOrgao ?orgao . ?orgao c:Codigo ?idOrgao .
  ?licitacao c:Identificador ?idLicitacao .
  FILTER (?idLicitacao = '15815205000102014')
} ORDER BY ?identificador
```

Fonte: próprio autor

Figura 38 - Somatório do valor das compras de uma licitação

```
PREFIX c: <http://sistemas.ifrr.edu.br/fuseki/compras.owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT (SUM(xsd:decimal(?valorInicial)) AS ?valorTotal)
WHERE {
  ?compra c:Nome ?nome . ?compra c:Numero ?numero . ?compra c:CNPJ ?cnpj .
  ?compra c:Identificador ?identificador . ?compra c:Processo ?processo .
  ?compra c:temLicitacao ?licitacao . ?compra c:ValorInicial ?valorInicial .
  ?compra c:FundamentoLegal ?fundamento . ?compra c:DataAssinatura ?dataAss .
  ?compra c:temUasg ?uasg . ?uasg c:Codigo ?idUasg .
  ?uasg c:temOrgao ?orgao . ?orgao c:Codigo ?idOrgao .
  ?licitacao c:Identificador ?idLicitacao .
  FILTER (?idLicitacao = '15815205000102014')
}
```

Fonte: próprio autor



Cenário 2 – Visualização do valor do orçamento destinado para uma determinada compra.

Descrição do cenário: “Um usuário deseja saber qual o valor do orçamento que foi destinado para uma determinada compra realizada no ano de 2014 pelo Instituto Federal de Roraima – IFRR.”

Para realizar tal tarefa, o usuário deverá acessar a aplicação através de um navegador, filtrar o órgão que deseja, em seguida listar as licitações e clicar no botão “Compras” para listar todas as compras da licitação selecionada.

Por fim, o usuário deverá escolher a compra que tem interesse e clicar no botão “Compras” para que a aplicação execute uma consulta SPARQL e liste os dados dessa compra, juntamente com o orçamento federal disponível para o grupo de natureza de despesa número 4 – Investimentos.

Tal consulta SPARQL pode ser visualizada na Figura 39 e seu resultado é observado na Figura 36. Nota-se que a consulta foi executada utilizando tanto a ontologia do modelo proposto quanto a ontologia do orçamento federal, resultando em dados ligados, com o órgão como item comum.

Figura 39 - Detalhes da Compra com Orçamento

```

PREFIX c: <http://sistemas.ifrr.edu.br/fuseki/compras.owl#>
PREFIX loa: <http://vocab.e.gov.br/2013/09/loa#>

SELECT ?nome ?cnpj ?identificador ?licitaIdent ?compra ?processo ?licitacao
       ?valorInicial ?uasg ?idUasg (SUM(?dotacaoInicial) AS ?somaDotacaoInicial)
WHERE {
  ?compra c:Nome ?nome . ?compra c:Numero '0' . ?compra c:CNPJ ?cnpj .
  ?compra c:Identificador ?identificador . ?compra c:Processo ?processo .
  ?compra c:temLicitacao ?licitacao . ?licitacao c:Identificador ?licitaIdent .
  ?compra c:ValorInicial ?valorInicial .
  ?compra c:Identificador '15815250000232014' .
  ?compra c:temUasg ?uasg . ?uasg c:Codigo ?idUasg .

  ?itemBlankNode loa:temExercicio ?exercicoURI .
  ?exercicoURI loa:identificador 2014 .
  ?itemBlankNode loa:valorDotacaoInicial ?dotacaoInicial .
  ?itemBlankNode loa:temUnidadeOrcamentaria [loa:codigo '26437'] .
}
GROUP BY ?nome ?cnpj ?identificador ?licitaIdent ?compra ?processo
         ?licitacao ?valorInicial ?uasg ?idUasg

```

Fonte: próprio autor

## 6.2 Avaliação da Usabilidade

Utilizando o Modelo de Aceitação de Tecnologia (TAM) proposto por (DAVIS, 1989) e posteriormente ampliado por (YOON; KIM, 2007), usando uma escala de Likert (LIKERT, 1932), foi realizada uma avaliação de usabilidade.

Segundo (DAVIS, 1989), estudos mostram que a percepção de utilidade, ou seja, a medida em que o usuário avalia que a aplicação pode melhorar sua experiência, é a variável mais importante a considerar dentre os vários fatores que as pessoas consideram para aceitar ou rejeitar uma aplicação. De acordo com pesquisas os benefícios da aplicação devem superar o seu esforço para usá-la (DAVIS, 1989). Considerando estes aspectos, foi utilizada uma segunda variável, também muito importante, chamada de facilidade de utilização, a qual é definida como o grau em que uma pessoa acredita que a utilização de um sistema está livre de estresse, esta variável complementa a primeira. A metodologia TAM foi utilizada para realizar a avaliação neste estudo para medir a aceitação do modelo.

A avaliação de cunho não probabilístico, utilizou uma amostra reduzida por comodidade do autor, foi realizada com cinco voluntários apresentando o seguinte perfil: todos com idade acima de 24 anos, servidores do Instituto Federal de Roraima – IFRR, ocupantes do cargo de analista de tecnologia da informação que avaliaram o protótipo com o objetivo de averiguar a percepção de utilidade e facilidade de uso do modelo proposto. Como instrumento de coleta, foi utilizado um questionário com dez questões utilizando a escala de Likert, ao responderem um questionário baseado nesta escala, os voluntários especificam seu nível de concordância com uma afirmação.

Em seguida os voluntários foram convidados à usar a aplicação e executar algumas tarefas que foram, previamente demonstradas: (i) listar as licitações de um órgão qualquer, (ii) localizar o valor total de compras de uma licitação qualquer do ano de 2014 de órgão IFRR, (iii) listar os dados de uma compra do IFRR no ano de 2010. Depois de realizar essas tarefas, os voluntários responderam um questionário com 10 questões.

Quanto à quantidade de pessoas a serem entrevistadas (UNGER; CHANDLER, 2009) indicam que esta é uma demanda ligada a abordagem da pesquisa: qualitativa ou quantitativa. Alguns autores, como Nielsen (2000), argumentam que um número pequeno de usuários pode proporcionar uma boa relação entre

o custo das entrevistas e o benefício de obtenção de informações relevantes sobre os sistemas. Segundo Nielsen e Landauer (1993),

O número de problemas de usabilidade encontrados em um teste de usabilidade com  $n$  usuários é  $N(1-(1-L)^n)$ , onde  $N$  é o número total de problemas de utilização na criação e  $L$  é a proporção de problemas de usabilidade descobertos durante o teste de um único usuário. O valor  $L$  é de 31%, em média, valor obtido através de um grande número de projetos estudados.

Em Nielsen (2012), a estimativa é que este número pode variar entre 5 e 8 usuários para um único teste. Entretanto (UNGER; CHANDLER, 2009) indicam que entre 5 e 8 usuários por grupo em cada rodada de pesquisa é comumente suficiente.

A Tabela 8 apresenta o resumo das respostas em relação à facilidade de uso. A primeira coluna apresenta a afirmativa e as seguintes 5 colunas mostram os resultados obtidos em cada proporção, em graus absolutos e relativos. Pode-se observar que a maioria achou a aplicação parcialmente fácil de entender e usar.

Tabela 8 - Resultado Avaliação Facilidade de Uso

Pergunta	Discordo Totalmente	Discordo Parcialmente	Indiferente	Concordo Parcialmente	Concordo Totalmente
01 - A aplicação é fácil de entender.	0% (0)	20% (1)	20% (1)	60% (3)	0% (0)
02 - A aplicação é fácil de usar.	0% (0)	40% (2)	0% (0)	60% (3)	0% (0)
03 - Com pequeno esforço posso encontrar dados de um órgão federal.	0% (0)	0% (0)	20% (1)	60% (3)	20% (1)
04 - Com um pequeno esforço posso encontrar as licitações do ano de 2014 do IFRR.	0% (0)	0% (0)	0% (0)	60% (3)	40% (2)
05 - A visualização dos dados é clara e de fácil entendimento.	0% (0)	60% (3)	0% (0)	20% (1)	20% (1)

Fonte: Elaborada pelo autor

Analisando o resultado da percepção de utilidade apresentado na Tabela 9, pode-se observar que a maior taxa foi relacionada à navegabilidade, mas precisa ser melhorada (pergunta 01). Em relação à busca e visualização as taxas foram observadas acima da média (perguntas 3 e 4).

Tabela 9 - Avaliação Percepção de Utilidade

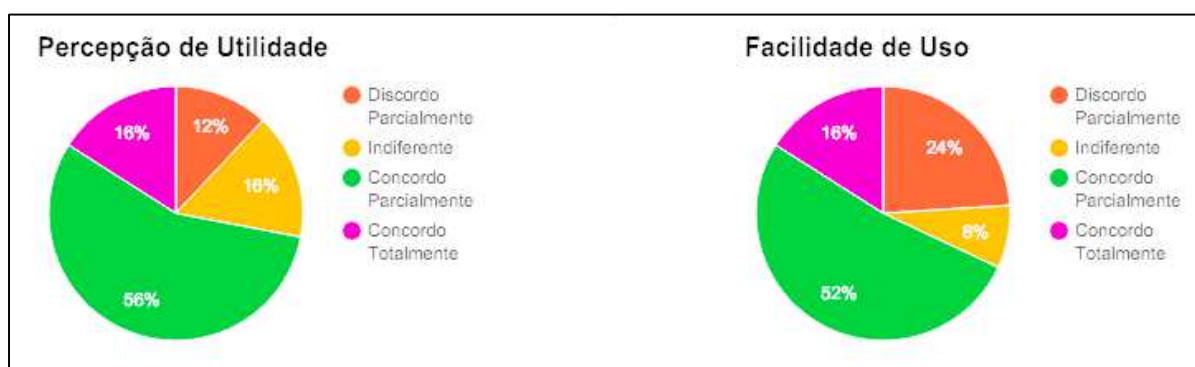
Pergunta	Discordo Totalmente	Discordo Parcialmente	Indiferente	Concordo Parcialmente	Concordo Totalmente
01 - As opções de navegabilidade foram relevantes.	0% (0)	0% (0)	20% (1)	80% (4)	0% (0)
02 - A aplicação facilita a visualização dos dados.	0% (0)	40% (2)	0% (0)	40% (2)	20% (1)
03 - A aplicação facilita encontrar as compras de cada órgão federal.	0% (0)	0% (0)	20% (1)	60% (3)	20% (1)
04 - A aplicação facilita a visualização de compras detalhadas.	0% (0)	20% (1)	0% (0)	60% (3)	20% (1)
05 - Se eu precisasse visualizar dados de uma compra interligada com dados do orçamento federal, eu usaria a aplicação.	0% (0)	0% (0)	40% (2)	40% (2)	20% (1)

Fonte: Elaborada pelo autor

O resumo dos resultados obtidos é exposto na Figura 40. A maioria dos usuários concordou parcialmente que o protótipo foi fácil de usar e também bastante útil. Nenhum usuário, considerado neste estudo, discordou da afirmação colocada.

Os resultados são incentivadores, mostram que o protótipo tem potencial de utilização de forma satisfatória, apesar de precisar de algumas melhorias visuais, sendo que pode ser expandido para demais áreas que precisam de seus dados interligados para melhor visualização.

Figura 40 - Resumo da Avaliação de usabilidade



Fonte: Elaborada pelo autor

### 6.3 Avaliação da Ontologia

Segundo Poveda-Villalón et al. (2012), a necessidade de validar a ontologia, ou seja, avaliar a sua qualidade e precisão, se dá pelo fato de que sua qualidade pode ser afetada pelos obstáculos envolvidos na sua modelagem, podendo acarretar em falhas na ontologia. De acordo com Hartmann et al. (2005), tentativas de estudar as diferentes abordagens e ferramentas de avaliação de ontologias foram realizadas com base no crescimento da quantidade de ontologias disponíveis.

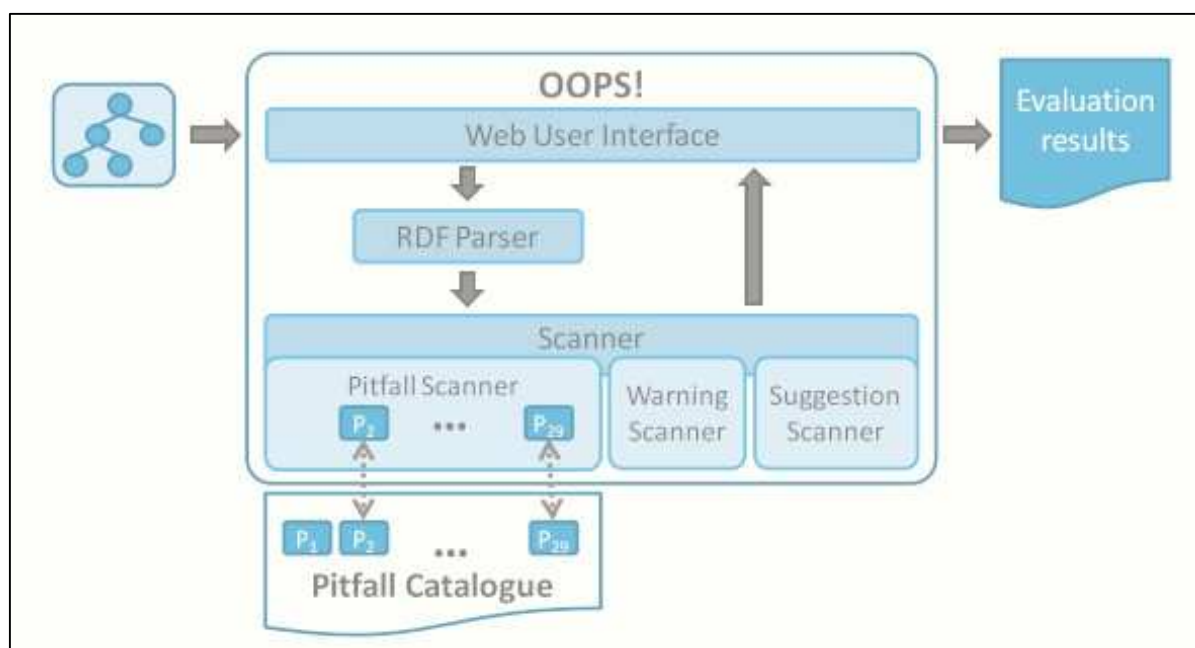
A seguir, é exposto em resumo algumas abordagens para avaliação e validação de ontologias.

- Avaliação baseada na Evolução: Tal abordagem parte do advento característico à mudança do conhecimento ao longo do tempo. É sempre necessário que sejam adicionadas às ontologias, os domínios em contínuo crescimento. Esta abordagem persegue mudanças na mesma ontologia em diferentes versões para atingir uma indicação da qualidade da ontologia. Podemos citar as avaliações de (HAASE et al., 2005) e (PLESSERS; DE TROYER, 2005) como exemplo.
- Avaliação Lógica (baseada em regras): Esta abordagem é baseada em regras, tem por objetivo avaliar a ontologia e a qualidade das regras de uso que estão embutidas nas linguagens de ontologia e identificar conflitos entre estas. Exemplo desta abordagem é definida por (ARPINAR; GIRILOGANATHAN, 2012).
- Avaliação em Métricas (características): Técnicas baseadas em métricas tem por objetivo oferecer um ponto de vista quantitativo da qualidade da ontologia. Tais técnicas utilizam diversas estatísticas sobre o conhecimento exposto na ontologia, ou solicitam ao usuário a inserção de alguma informação que não está inserida na ontologia. As técnicas de (CORCHO et al., 2004), (MOSTOWFI; FOTOUHI, 2006) e (SUPEKAR; PATEL; LEE, 2004) exemplificam esta abordagem.

Para a avaliação da ontologia do modelo implementado, foi utilizada uma técnica baseada em métricas, definida por (POVEDA-VILLALÓN; SUÁREZ-FIGUEROA; GÓMEZ-PÉREZ, 2012) e chamada de Ontology Pitfall Scanner – OOPS!. Baseia-se na análise da ontologia para saber se está em conformidade com as melhores práticas de modelagem de ontologias, isto é, verificar se as onto-

logias apresentam anomalias ou *pitfalls* (armadilhas). A escolha desta metodologia foi baseada no benefício da aplicação do conjunto de melhores práticas reconhecidas e compiladas pelo autor neste trabalho. OOPS! é uma ferramenta baseada na web, independente de qualquer ambiente de desenvolvimento de ontologias, para localização de prováveis armadilhas que possam levar à erros de modelagem, sua arquitetura é apresentada na Figura 41.

Figura 41 - Arquitetura OOPS!



Fonte: (POVEDA-VILLALÓN; SUÁREZ-FIGUEROA; GÓMEZ-PÉREZ, 2012)

OOPS! é uma aplicação web que utiliza as tecnologias Java EE, HTML, JQuery, JSP e CSS. Sua interface consiste em uma visão singular, onde o usuário fornece uma URI ou o documento em formato RDF para que a ontologia seja analisada. Durante sua análise, uma busca pelos principais problemas, relacionados na Tabela 10, ocorre. Durante esse processo de varredura, os elementos da ontologia comprometidos em potenciais erros são detectados, ademais, os avisos sobre sintaxe RDF e sugestões de modelagem são gerados. Por fim, os resultados são apresentados na interface web mostrando uma lista de armadilhas, caso haja, e os elementos afetados da ontologia, assim como explicações adicionais.

Tabela 10 - Principais Problemas

<b>Entendimento Humano</b>	<b>Questões de Modelagem</b>
P1. Criação de elementos polissêmicos	P2. Criando sinônimos como classes
P2. Criação de sinônimos como classes	P3. Criação de relação “is a” no lugar de “rdfs:subClassOf”, “rdf:type” ou “owl:sameAs”
P7. Mesclando conceitos diferentes na mesma classe	P4. Criação de elementos desconexos ontologia
P8. Falta de anotações	P5. Definição de relações inversas equivocadas
P11. Falta de domínio (domain) ou intervalo(range) em propriedades	P6. Inclusão de ciclos na hierarquia
P12. Falta de propriedades equivalentes	P7. Mesclando conceitos diferentes na mesma classe
P13. Falta de relações inversas	P10. Falta de disjunção
P19. Troca entre intersecção e união	P17. Hierarquia muito específica
P20. Mau uso de anotações na ontologia	P11. Falta de domínio(domain) ou intervalo(range) nas propriedades
P22. Usando diferentes critérios de nomenclatura na ontologia	P12. falta de propriedades equivalentes
Consistência Lógica	P13. Falta de relações inversas
P5. Definição errada de relações inversas	P14. Mau uso de “owl:allValuesFrom”
P6. Inclusão de ciclos na hierarquia	P15. O mau uso de “not some” e “some not”
P14. Mau uso de “owl:allValuesFrom”	P18. Domínio ou intervalo muito específico
P15. Mau uso de “not some” e “some not”	P19. Troca entre intersecção e união
P18. Domínio ou intervalo muito específico	P21. Usando uma classe modificada
P19. Troca entre intersecção e união	P23. Usando elementos da ontologia incorretamente
P27. Definição errada de relações equivalentes	P24. Usando definição recursiva
P28. Definição errada de relações simétricas	P25. Definição de uma relação inversa a ela mesma
P29. Definição errada de relaciones transitivas	P26. Definição de uma relação inversa para uma simétrica
<b>Representação Mundo Rela</b>	P27. Definição de relações equivalentes equivocadas
P9. Falta de informação básica	P28. Definição de relações simétricas equivocadas
P10. Falta de disjunção	P29. Definição de relações transitivas equivocadas

Fonte: (POVEDA-VILLALÓN; SUÁREZ-FIGUEROA; GÓMEZ-PÉREZ, 2012)

A aplicação *Pitfall Scanner*, Figura 42, realiza a descoberta automática de um subconjunto de 21 erros classificados como erros de compreensão humana, consistência lógica, representação do mundo real e questões de modelagem. Para verificar os resultados produzidos, bastou inserir o código fonte OWL da ontologia

Compras no website do projeto para poder utilizar o modelo de avaliação OOPS!, foram obtidos os seguintes resultados, resumidos na Figura 42.

- P8 – Falta de anotações: 93 casos.
- P10 – Falta de disjunção.
- P13 – Falta de relacionamentos inversos: 19 casos.
- P19 – Definido múltiplos domínios ou intervalos nas propriedades: 26 casos.
- P22 – Uso de diferentes convenções de nomenclatura na ontologia.
- P41 – Licença não declarada.
- Sugestão – Propriedade de objetivo simétrica ou transitiva.

Figura 42 - Avaliação da Ontologia

**OOPS! (Ontology Pitfall Scanner!)** helps you to detect some of the most common pitfalls appearing when developing ontologies. To try it, enter a URI or paste an OWL document into the text field above. A list of pitfalls and the elements of your ontology where they appear will be displayed.

Scanner by URI:  Scanner by URI  
 Example: [http://data.semanticweb.org/ns/swc/swc\\_2009-05-09.rdf](http://data.semanticweb.org/ns/swc/swc_2009-05-09.rdf)

Scanner by direct input: 

```
<?xml version="1.0"?>
<!DOCTYPE rdf:RDF [
<ENTITY owl "http://www.w3.org/2002/07/owl#" >
<ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
<ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
<ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
]>
```

 Scanner by RDF

Uncheck this checkbox if you don't want us to keep a copy of your ontology. [Go to advanced evaluation](#)

### Evaluation results

It is obvious that not all the pitfalls are equally important; their impact in the ontology will depend on multiple factors. For this reason, each pitfall has an importance level attached indicating how important it is. We have identified three levels:

- **Critical** 🚨 : It is crucial to correct the pitfall. Otherwise, it could affect the ontology consistency, reasoning, applicability, etc.
- **Important** 🟡 : Though not critical for ontology function, it is important to correct this type of pitfall.
- **Minor** 🟢 : It is not really a problem, but by correcting it we will make the ontology nicer.

[Expand All] | [Collapse All]

Results for P08: Missing annotations.	93 cases   Minor 🟢
Results for P10: Missing disjointness.	ontology*   Important 🟡
Results for P13: Inverse relationships not explicitly declared.	19 cases   Minor 🟢
Results for P19: Defining multiple domains or ranges in properties.	26 cases   Critical 🚨
Results for P22: Using different naming conventions in the ontology.	ontology*   Minor 🟢
Results for P41: No license declared.	ontology*   Important 🟡
SUGGESTION: symmetric or transitive object properties.	1 case

According to the highest importance level of pitfall found in your ontology the conformance badge suggested is "Critical pitfalls" (see below). You can use the following HTML code to insert the badge within your ontology documentation:

Want to help?

- Suggest new pitfalls
- Provide feedback

Documentation:

- Pitfall catalogue
- User guide
- Technical report

Related papers:

- IJSWIS 2014
- EKAW 2012
- ESWC 2012 Demo
- Ontoqual 2010
- CAEPIA 2009

Fonte: <http://oops.linkeddata.es/response.jsp>



Posteriormente a execução da OOPS!, constatou-se que ainda existiam algumas inconsistências simples no modelo. Uma vez que a ferramenta informa com exatidão qual classe, indivíduo ou propriedade está com erro, sua manutenção fica simples. A ferramenta Protégé foi utilizada para corrigir os erros indicados na ontologia Compras.



## 7 CONCLUSÕES

Este trabalho apresentou um modelo para integração de informações de bases de dados abertas e conectadas, com uso de ontologias, que visa contribuir com a área de web semântica envolvendo o suporte para aplicações com uso de dados abertos e conectados.

No transcorrer deste trabalho, evidencia-se a importância da divulgação dos dados, principalmente os governamentais, mas também a necessidade de padronizar tal processo, tendo em vista o futuro da web, onde espera-se que todo e qualquer dado esteja disponível e interligado. Nesse sentido, foram estudados os trabalhos relacionados e as iniciativas propostas para colaborar na solução desta questão. Pode-se destacar as recomendações do W3C para divulgação de dados e aquelas destinadas ao mapeamento de dados oriundos de bases de dados relacionais, além da parceria criada entre os governos internacionais para divulgação de seus dados.

Estudando os trabalhos relacionados constatou-se que a divulgação de dados governamentais com notação semântica no Brasil ainda é pequena. Um exemplo claro disso é exposto no capítulo 3, na seção que trata dos dados do Orçamento Federal Brasileiro. Observa-se também que, no portal de Dados Abertos, apesar dos dados estarem disponibilizados em formato abertos, estes não são recentes, sendo que muitos são datados de anos anteriores ao atual.

Com esse estudo não foi identificada uma iniciativa de consumo desses dados diretamente a partir de suas origens, para se manterem atualizados, nem de interligar tais dados entre si através de ontologias, motivando elaboração deste trabalho que propõe o desenvolvimento de um modelo genérico para consumo de dados em bases relacionais, mapeando-os, inserindo notação semântica e facilitando a sua divulgação.

Para um estudo detalhado, a Tabela 3 apresenta a comparação realizada entre os trabalhos científicos relacionados. Em relação ao trabalho desenvolvido, as características exclusivas são o desenvolvimento de uma ontologia específica para os dados que foram mapeados e a interligação de bases distintas.

A implementação do protótipo com base no modelo proposto foi realizada especificamente com os dados das compras governamentais, proporcionando a sua interligação com os dados do orçamento federal. Neste protótipo, foram desenvolvi-

das as funcionalidades propostas no modelo, tais como o desenvolvimento de uma ontologia, mapeamento de dados, utilização de um *triplestore* para armazenamento dos dados já em formato RDF e o desenvolvimento de uma aplicação para divulgação dos dados já interligados entre si para livre utilização do usuário.

Foi realizada uma avaliação com base em três metodologias: avaliação de cenário, de usabilidade e de ontologia. Na ontologia foi realizada uma avaliação definida por (POVEDA-VILLALÓN; SUÁREZ-FIGUEROA; GÓMEZ-PÉREZ, 2012), chamada Ontology PitFall Scanner – OOPS!. Executada com o objetivo de analisar se a ontologia Compras estava em conformidade com as melhores práticas de modelagem de ontologia, isto é, para verificar se a ontologia apresenta anomalias ou *pitfalls*(armadilhas). Os resultados mostraram que existiam alguns erros simples de definição da ontologia, que foram a falta de anotações, disjunções e ausência de relacionamentos inversos, prontamente corrigidas. Uma segunda avaliação foi realizada utilizando a descrição de cenários, a partir da definição de alguns casos de testes possíveis para avaliar a utilização do protótipo, os quais avaliaram as principais características que o modelo tem por objetivo atender. Finalmente, foi realizada uma avaliação de usabilidade utilizando a metodologia TAM com o objetivo de investigar a percepção de utilidade e facilidade de uso do modelo proposto, com envolvimento de voluntários que avaliaram o protótipo implementado. Os resultados obtidos demonstraram que houve uma aceitação boa em relação a percepção de utilidade e a facilidade de uso, entretanto foram identificadas necessidades de melhorias para facilitar ainda mais o uso da aplicação.

Em suma, com os resultados obtidos e com o trabalho desenvolvido, é possível concluir que as principais contribuições do modelo proposto estão relacionadas com a utilização de uma ontologia específica para integração de bases de dados abertas, com a definição e implementação de mecanismos de interligação de dados abertos e, por fim, com a modelagem e implementação de facilidades para a implementação de ferramentas para a divulgação de dados abertos e ligados.

Como parte de possibilidades de trabalhos futuros está prevista a implementação de uma área administrativa para a aplicação com o intuito de gerenciar melhor as configurações de consultas, dados apresentados, endereço do *endpoint* que será usado e com base nas avaliações, proporcionando uma melhoria da aplicação em si. Além disso, pretende-se aplicar o modelo na interligação de dados de outras áreas fins, com novas instanciações do protótipo para avaliar a generalidade conti-

da no modelo proposto. Por fim, considera-se que os testes apresentados são limitados em suas possibilidades de gerar informações consistentes sobre o protótipo, tendo em vista que foram delimitados um número pequeno de casos de uso e também foram realizados experimentos com um número pequeno de participantes. Desta forma, pretende-se ampliar as informações de avaliação, com a realização de novos experimentos contando com um número maior de casos e de participantes.



## 8 BIBLIOGRAFIA

ANTHES, G. Estonia. **Communications of the ACM**, v. 58, n. 6, p. 18–20, 2015.

**API Compras Governamentais**. Disponível em:  
<<http://compras.dados.gov.br/docs/home.html>>. Acesso em: 24 maio. 2015.

**API Orçamento Federal**. Disponível em: <<http://vocab.e.gov.br/2013/09/loa>>. Acesso em: 15 maio. 2015.

ARENAS, M. et al. A Direct Mapping of Relational Data to RDF. **W3C Working Draft 20 September 2011**, n. September 2011, p. 1–11, 2011.

ARPINAR, I.; GIRILOGANATHAN, K. Ontology Quality by Detection of Conflicts in Metadata. **Proceedings of the 4th**, 1922.

BERNERS-LEE, B. T.; HENDLER, J. The Semantic Web. **Scientific American**, v. 21, p. 29–37, 2001.

BERNERS-LEE, T. **Tim Berners-Lee**. Disponível em:  
<[http://en.wikipedia.org/wiki/Tim\\_Berners-Lee](http://en.wikipedia.org/wiki/Tim_Berners-Lee)>. Acesso em: 12 maio. 2015.

BEX, F. et al. ArguBlogging: An application for the Argument Web. **Journal of Web Semantics**, v. 25, p. 9–15, 2014.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked data—the story so far. **International journal on Semantic Web and Information Systems**, v. 5, p. 1–22, 2009.

BIZER, CHRIS; CYGANIAK, RICHARD; HEATH, T. **How to Publish Linked Data on the Web**. Disponível em: <<http://wifo5-03.informatik.uni-mannheim.de/bizer/pub/LinkedDataTutorial/>>. Acesso em: 16 maio. 2015.

CABRAL, S. P. et al. Aplicando Linked Data na publicação de dados do ENEM. **CEUR Workshop Proceedings**, v. 938, p. 176–181, 2012.

CALDERÓN, C.; LORENZO, S. **Open Government. Gobierno Abierto**. [s.l: s.n.].

COMITÊ GESTOR DA INTERNET NO BRASIL. **Manual dos Dados Abertos: Desenvolvedores**. [s.l: s.n.].

CORCHO, Ó. et al. ODEval: a tool for evaluating RDF (S), DAML+ OIL, and OWL concept taxonomies. **Artificial Intelligence Applications and Innovations**, p. 369–382, 2004.

DAS, S.; SUNDARA, S.; CYGANIAK, R. R2RML: RDB to RDF Mapping Language. **W3C Recommendation**, 2012.

DAVIS, B. F. D. Information Technology Perceived Usefulness and Perceived Ease of Use. **MIS Quarterly**, n. September, p. 319–339, 1989.

DEAN, M.; SCHREIBER, G. **OWL Web Ontology Language Reference**. Disponível em:  
<<http://www.w3.org/TR/owl-ref/>>. Acesso em: 15 maio. 2015.

EAVES, D. **David Eaves - About**. Disponível em: <<http://eaves.ca/about/>>. Acesso em: 14 maio. 2015a.

EAVES, D. **The Three Laws of Open Government Data**. Disponível em: <<http://eaves.ca/2009/09/30/three-law-of-open-government-data/>>. Acesso em: 14 maio. 2015b.

FONSECA, LUCAS B.R., AZEVEDO, CARLOS L. B., ALMEIDA, J. P. A. Mapeando Dados Governamentais com uma Ontologia de Organizações. **LOD Brasil, Linked Open Data**, p. 27–42, 2014.

GERBER, A J.; BARNARD, A.; VAN DER MERWE, A J. TOWARDS A SEMANTIC WEB LAYERED ARCHITECTURE. **The International Journal on Very Large Data Bases**, p. 353–362, 2007.

GRUBER, T. R. A translation approach to portable ontology specifications. **Knowledge Acquisition**, v. 5, n. April, p. 199–220, 1993.

GUARINO, N. Formal Ontology and Information Systems. **Fois'98**, v. 46, n. June, p. 3–15, 1998.

HAASE, P. et al. A framework for handling inconsistency in changing ontologies. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 3729 LNCS, p. 353–367, 2005.

HARTMANN, J. et al. Methods for Ontology Evaluation. **Knowledge Web Deliverable D**, p. 5–49, 2005.

HEATH, T.; BIZER, C. **Linked Data: Evolving the Web into a Global Data Space**. [s.l.: s.n.], v. 1

HERT, M.; REIF, G.; GALL, H. A comparison of RDB-to-RDF mapping languages. **Proceedings of the 7th International Conference on Semantic Systems - I-Semantics '11**, p. 25–32, 2011.

ISOTANI, S. B. I. I. **Dados Abertos Conectados : em Busca da Web do Conhecimento**. São Paulo: Novatec, 2015.

**Jena**. Disponível em: <[https://jena.apache.org/about\\_jena/architecture.html](https://jena.apache.org/about_jena/architecture.html)>. Acesso em: 26 maio. 2015.

KNÖPFEL, A. **FMC Quick Introduction**FMC Consortium, , jun. 2007.

KÖCHE, J. C. **Fundamentos de metodologia científica : teoria da ciência e iniciação à pesquisa**. Digital ed. Petrópolis, RJ: Vozes, 2011.

**LAI - Mapa**. Disponível em: <<http://www.acaoainformacao.gov.br/assuntos/conheca-seu-direito/a-lei-de-acesso-a-informacao/mapa-da-lai>>. Acesso em: 15 maio. 2015.

**LAI - Principais Aspectos**. Disponível em: <<http://www.acaoainformacao.gov.br/assuntos/conheca-seu-direito/principais-aspectos/principais-aspectos>>. Acesso em: 15 maio. 2015.



LIKERT, R. A technique for the measurement of attitudes. **Archives of Psychology**, v. 22, n. 140, p. 1–55, 1932.

MCGUINNESS, D. L.; HARMELEN, F. VAN. OWL Web Ontology Language Overview. <http://www.w3.org/TR/owl-features/>, 2003.

MOSTOWFI, F.; FOTOUHI, F. Improving Quality of Ontology: An Ontology Transformation Approach. **Data Engineering Workshops, 2006. Proceedings. 22nd International Conference on**, p. 61, 2006.

NIELSEN, J. **Why You Only Need to Test with 5 Users**. Disponível em: <<http://www.useit.com/alertbox/20000319.html>>.

NIELSEN, J. **How many test users in a usability study**. Disponível em: <<http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:How+Many+Test+Users+in+a+Usability+Study?#8>>.

NIELSEN, J.; LANDAUER, T. K. A mathematical model of the finding of usability problems. **Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '93**, p. 206–213, 1993.

**OGP - Iniciativa**. Disponível em: <<http://www.governoaberto.cgu.gov.br/a-ogp/o-que-e-a-iniciativa>>. Acesso em: 15 maio. 2015.

**Open Definition**. Disponível em: <<http://opendefinition.org/od/>>. Acesso em: 15 maio. 2015.

**Open Government Data Principles**. Disponível em: <<http://opengovdata.org>>. Acesso em: 14 maio. 2015.

PEREIRA, C. M. Uma Abordagem para Publicação de Dados Ligados Obtidos a partir de bases de dados relacionais. p. 141–156, 2014.

PLESSERS, P.; DE TROYER, O. Ontology change detection using a version log. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 3729 LNCS, p. 578–592, 2005.

POVEDA-VILLALÓN, M.; SUÁREZ-FIGUEROA, M. C.; GÓMEZ-PÉREZ, A. Validating Ontologies with OOPS ! State of the Art. **Knowledge Engineering and Knowledge Management**, p. 267–281, 2012.

**Protégé**. Disponível em: <[http://protegewiki.stanford.edu/wiki/Pr4\\_UG](http://protegewiki.stanford.edu/wiki/Pr4_UG)>. Acesso em: 23 maio. 2015.

RAMALHO, R. D. S.; VIDOTTI, S.; FUJITA, M. **Web semântica: uma investigação sob o olhar da Ciência da Informação**. Disponível em: <[http://www.dgz.org.br/dez07/Art\\_04.htm](http://www.dgz.org.br/dez07/Art_04.htm)>.

SAP AG. **Standardized Technical Architecture Modeling**. [s.l.: s.n.].

SILVA, D. A. **Aplicação de Inferência Semântica para Exploração Conceitual de Bases de Dados Relacionais**. [s.l.] Dissertação de Mestrado - Universidade de Brasília, Faculdade de Tecnologia. Departamento de Engenharia Elétrica., 2014.

**SPARQL by Example**. Disponível em: <<http://www.w3.org/2009/Talks/0615-qbe/>>. Acesso em: 16 maio. 2015.

SPERONI, R. D. M. et al. Avaliação do ensino superior público no Brasil : protótipo de aplicação linked data. p. 157–170, 2014.

SUPEKAR, K.; PATEL, C.; LEE, Y. Characterizing Quality of Knowledge on Semantic Web. **7th International Florida Artificial Intelligence Research Society Conference**, p. 220–228, 2004.

TAM. **Technology Acceptance Model**. Disponível em: <<http://www.fmc-modeling.org/fmc-and-tam>>. Acesso em: 13 ago. 2015.

UNGER, R.; CHANDLER, C. **O Guia Para Projetar UX : A Experiência do Usuário (UX) para Projetistas de Conteúdo Digital, Aplicações e Web Sites**. Rio de Janeiro: Alta Books, 2009.

**W3C Semantic Web Activity**. Disponível em: <<http://www.w3.org/2001/12/semweb-fin/w3csw>>. Acesso em: 13 maio. 2015.

WAZLAWICK, R. **Metodologia de Pesquisa para Ciência da Computação**. 2<sup>a</sup>. ed. [s.l.] Elsevier Brasil, 2009.

YOON, C.; KIM, S. Convenience and TAM in a ubiquitous computing environment: The case of wireless LAN. **Electronic Commerce Research and Applications**, v. 6, n. 1, p. 102–112, 2007.

ZHANG, L. et al. Exploiting Knowledge Bases for Multilingual and Cross-lingual Semantic Annotation and Search. p. 1–8, 2015.

## APÊNDICE A – QUESTIONÁRIO PARA A AVALIAÇÃO DA USABILIDADE

Preencha o questionário abaixo com a sua opinião sobre as afirmativas a respeito *da aplicação de Compras Governamentais*.

Sobre a facilidade de uso

1. A aplicação é fácil de entender.				
<input type="checkbox"/> Discordo totalmente	<input type="checkbox"/> Discordo parcialmente	<input type="checkbox"/> Não concordo nem discordo	<input type="checkbox"/> Concordo parcialmente	<input type="checkbox"/> Concordo totalmente
2. A aplicação é fácil de usar.				
<input type="checkbox"/> Discordo totalmente	<input type="checkbox"/> Discordo parcialmente	<input type="checkbox"/> Não concordo nem discordo	<input type="checkbox"/> Concordo parcialmente	<input type="checkbox"/> Concordo totalmente
3. Com um pequeno esforço posso encontrar dados de um órgão federal específico.				
<input type="checkbox"/> Discordo totalmente	<input type="checkbox"/> Discordo parcialmente	<input type="checkbox"/> Não concordo nem discordo	<input type="checkbox"/> Concordo parcialmente	<input type="checkbox"/> Concordo totalmente
4. Com um pequeno esforço posso encontrar as licitações do ano de 2014 do IFRR.				
<input type="checkbox"/> Discordo totalmente	<input type="checkbox"/> Discordo parcialmente	<input type="checkbox"/> Não concordo nem discordo	<input type="checkbox"/> Concordo parcialmente	<input type="checkbox"/> Concordo totalmente
5. A visualização dos dados é clara e de fácil entendimento.				
<input type="checkbox"/> Discordo totalmente	<input type="checkbox"/> Discordo parcialmente	<input type="checkbox"/> Não concordo nem discordo	<input type="checkbox"/> Concordo parcialmente	<input type="checkbox"/> Concordo totalmente

Sobre a sua Percepção de Utilidade

6. As opções de navegabilidade foram relevantes.				
<input type="checkbox"/> Discordo totalmente	<input type="checkbox"/> Discordo parcialmente	<input type="checkbox"/> Não concordo nem discordo	<input type="checkbox"/> Concordo parcialmente	<input type="checkbox"/> Concordo totalmente
7. A aplicação facilita a visualização dos dados.				
<input type="checkbox"/> Discordo totalmente	<input type="checkbox"/> Discordo parcialmente	<input type="checkbox"/> Não concordo nem discordo	<input type="checkbox"/> Concordo parcialmente	<input type="checkbox"/> Concordo totalmente
8. A aplicação facilita encontrar as compras de cada órgão federal.				
<input type="checkbox"/> Discordo totalmente	<input type="checkbox"/> Discordo parcialmente	<input type="checkbox"/> Não concordo nem discordo	<input type="checkbox"/> Concordo parcialmente	<input type="checkbox"/> Concordo totalmente
9. A aplicação facilita a visualização de compras detalhadas.				
<input type="checkbox"/> Discordo totalmente	<input type="checkbox"/> Discordo parcialmente	<input type="checkbox"/> Não concordo nem discordo	<input type="checkbox"/> Concordo parcialmente	<input type="checkbox"/> Concordo totalmente
10. Se eu precisasse visualizar dados de uma compra interligada com dados do orçamento federal, eu usaria a aplicação.				
<input type="checkbox"/> Discordo totalmente	<input type="checkbox"/> Discordo parcialmente	<input type="checkbox"/> Não concordo nem discordo	<input type="checkbox"/> Concordo parcialmente	<input type="checkbox"/> Concordo totalmente