



**UNIVERSIDADE DO SUL DE SANTA CATARINA**

**GABRIELA TAVARES GOMES**

**MAX ROMOALDO FERTIG**

**CONSTRUÇÃO DE UMA BASE DE CONHECIMENTO DE DADOS  
GOVERNAMENTAIS ABERTOS BASEADA EM ONTOLOGIA UTILIZANDO  
DADOS CONECTADOS**

Florianópolis

2016

**GABRIELA TAVARES GOMES**

**MAX ROMOALDO FERTIG**

**CONSTRUÇÃO DE UMA BASE DE CONHECIMENTO DE DADOS  
GOVERNAMENTAIS ABERTOS BASEADA EM ONTOLOGIA UTILIZANDO  
DADOS CONECTADOS**

Trabalho de Conclusão de Curso apresentado  
ao Curso de Sistemas de Informação da  
Universidade do Sul de Santa Catarina como  
requisito parcial à obtenção do título de  
Bacharel em Sistemas de Informação.

Orientador: Prof. Flávio Ceci, Dr.

Florianópolis

2016

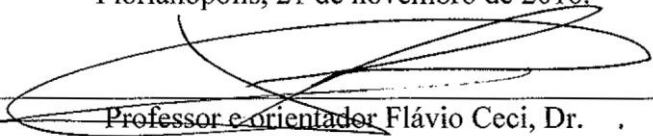
**GABRIELA TAVARES GOMES**

**MAX ROMOALDO FERTIG**

**CONSTRUÇÃO DE UMA BASE DE CONHECIMENTO DE DADOS  
GOVERNAMENTAIS ABERTOS BASEADA EM ONTOLOGIA UTILIZANDO  
DADOS CONECTADOS**

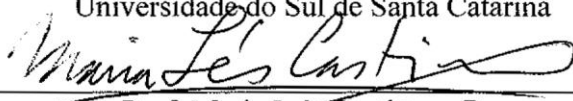
Este Trabalho de Conclusão de Curso foi julgado adequado à obtenção do título de Bacharel em Sistemas de Informação e aprovado em sua forma final pelo Curso de Sistemas de Informação da Universidade do Sul de Santa Catarina.

Florianópolis, 21 de novembro de 2016.



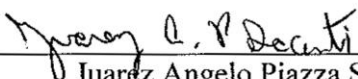
---

Professor e orientador Flávio Ceci, Dr.  
Universidade do Sul de Santa Catarina



---

Prof. Maria Inés Castiñera, Dr.  
Universidade do Sul de Santa Catarina



---

Juárez Angelo Piazza Sacenti, MSc.  
Universidade Federal de Santa Catarina

## **AGRADECIMENTOS**

**Gabriela Tavares Gomes** agradece

Existem muitas pessoas que gostaria de agradecer, sem elas todo o caminho traçado seria muito mais árduo, se não impossível.

Primeiramente agradeço aos meus pais que sempre me deram todo o suporte necessário para que eu pudesse correr atrás dos meus ideais, sempre estando comigo nos momentos necessários, a minha mãe que sempre foi uma guerreira, me ensinou como ser uma pessoa de bem e sem ela nada disso teria acontecido. Ao meu pai que apesar de ter partido antes de poder me ver aqui, tenho a certeza de que estaria muito orgulhoso de saber que cheguei aonde sonhávamos. Amo vocês!

Aos meus padrinhos amados, Ieda e Castelar, que apesar da distância sempre estavam presentes e me dando o suporte necessário nos momentos em que mais precisei.

Ao meu namorado, Bruno, que ao longo de todo o curso me deu um apoio inimaginável, de todas as formas possíveis, tendo sempre muita paciência comigo durante todos esses anos de faculdade, te agradeço infinitamente por ser meu grande parceiro de vida.

Ao meu colega de TCC Max, que além de ter feito este trabalho final do curso comigo, foi um grande companheiro ao longo de toda graduação, com certeza guardarei para sempre a parceria que tivemos nesses cinco anos.

Aos professores da Unisul que sem dúvidas me ensinaram muito, especialmente ao meu orientador de TCC Flavio, que me mostrou como é ser um excelente professor e também um grande amigo, obrigada por tornar as coisas mais leves. A professora Maria Inés que sempre foi uma excelente mestre e também sempre se mostrou companheira e compreensiva com as nossas aflições de orientandos. E todos os demais professores com quem tive contato ao longo do curso.

Aos meus diversos amigos, que estiveram comigo durante essa caminhada, dentre tantos que me aguentaram esse tempo todo, principalmente nesse período complicado, gostaria de agradecer especialmente a Nathália, Roberta, Julia e a Jady. Vocês são especiais demais para mim, obrigada por me ouvirem pacientemente, me acalmarem e me aconselharem durante todos esses anos.

### **Max Romoaldo Fertig agradece**

Primeiramente agradeço minha mãe Celina Ed. Koakoski Fertig, e meu pai Nery Fertig, por hoje estar aqui, e ter concluído mais esta etapa em minha vida, sem a educação passada por eles não seria a pessoa que sou hoje.

Ao Professor José Leomar Todesco e ao projeto MUN.lodKem, da Universidade Federal de Santa Catarina, por disponibilizar os dados de cidades utilizados por este projeto.

Ao orientador, Professor Flávio Ceci, por acreditar em nosso potencial do início ao fim deste trabalho e incentivar ao máximo nossa criatividade.

Ao pesquisador e amigo, Juarez Angelo Piazza Sacenti, da Universidade Federal de Santa Catarina, por dar o apoio inicial e suporte inicial na área de ontologia e dados conectados.

A Lúcia Ceci, minha namorada, amiga e companheira, pelo apoio e compreensão durante o tempo de desenvolvimento, e pelo auxílio na revisão do texto.

A Gabriela Tavares Gomes, pelo companheirismo e confiança na parceria para o desenvolvimento deste projeto e durante esses cinco anos de curso.

Ao professor Mauricio Botelho, e a professora Maria Inés Castinera, pelas leituras e ponderações a respeito deste trabalho.

Enfim, a todos que direta ou indiretamente se envolveram na execução deste projeto, por acreditarem em nosso potencial.



## RESUMO

Apesar do aumento considerável na publicação de dados abertos governamentais, principalmente quando se trata de dados do governo brasileiro, ainda há diversos problemas envolvendo a publicação destes. A falta de padrões, tanto em estrutura como em formatos disponibilizados, criam barreiras para os dados publicados. Dado aberto só é aberto quando é disponível a todos. Desta forma, o grande desafio é desenvolver soluções para a padronização e facilidade no acesso aos dados abertos disponibilizados pelo governo brasileiro, para pessoas e máquinas, garantindo a interoperabilidade entre sistemas. Este trabalho teve como objetivo mostrar uma alternativa para a publicação de dados abertos governamentais, devido ao cenário em que os dados estão publicados hoje dificultando a sua utilização pelas organizações públicas e privadas ou até para uso da sociedade em geral. Para isso foi modelado e implementado uma base de conhecimento baseada em ontologia, utilizando dados abertos conectados, a fim de avaliar as possibilidades que essa solução poderia trazer nos dados publicados pelo governo brasileiro. Os resultados obtidos com a avaliação dos experimentos, realizados através consultas diretas ao banco de dados com a finalidade de responder perguntas específicas acerca do domínio de dados, foram satisfatórios, e demonstraram a simplicidade para recuperar e analisar os dados, através de consultas diretas a um *dataset* de dados conectados.

Palavras-chave: Dados abertos. Dados Conectados. Dados Governamentais Abertos. Ontologia. Triplestore. Base de Conhecimento.

## **ABSTRACT**

Although considerable increase in publications of open government data, mainly when it comes to Brazilian government data, there are still several problems involving the publication of these. The lack of standards, both in structure and in available formats, creates barriers to the use of published data. Open data is only open when it's available to everyone. Therefore, the great challenge is to develop solutions for the standardization and easy access to the open data made available by the Brazilian government, for people and machines, ensuring interoperability between systems. This work aims to show an alternative to the publication of open governmental data, due to the scenario in which the data are published today, making it difficult for public and private organizations or for the use of society in general. For that end, it was modeled and implemented an ontology-based knowledge base, using linked open data, in order to evaluate the possibilities that this solution could bring. The results, obtained from the evaluation of the experiments, made by direct queries to the database in order to answer specific questions about the data domain, were satisfactory and demonstrated the simplicity to retrieve and analyze data through direct queries on a dataset of linked data.

**Keywords:** Open Data, Linked Data, Open Governmental Data, Ontology, Triplestore, Knowledge base.



## LISTA DE ILUSTRAÇÕES

Figura 1 – O processo de transformação de dados em informação.....	20
Figura 2 – Conhecimento segundo o agente gerador e beneficiário.....	21
Figura 3 – Tripla RDF com dois nós (sujeito e objeto) e um predicado os conectando.....	26
Figura 4 – Representação do grafo RDF baseado nos statements do quadro 3.....	27
Figura 5 – Grafo baseado na figura 4 representado com IRIs. ....	28
Figura 6 – Estrutura da OWL 2 .....	33
Figura 7 – Axiomas da OWL 2 .....	34
Figura 8 – Parte do diagrama da nuvem de dados abertos conectados em 2014.....	40
Figura 9 – Fluxograma das atividades do trabalho.....	45
Figura 10 – Conversão dos dados semiestruturados em dados conectados.....	46
Figura 11 – Áreas semânticas da UML .....	48
Figura 12 – Diagrama de atividades do triplicador de dados.....	53
Figura 13 – Diagrama de comunicação do triplicador de dados. ....	54
Figura 14 – Diagrama de Sequência do triplicador de dados.....	55
Figura 15 – Diagrama de Classe do triplicador de dados. ....	56
Figura 16 – Diagrama de Implantação do triplicador de dados. ....	57
Figura 17 – Ferramentas utilizadas.....	58
Figura 18 – Interação entre as APIs do Jena. ....	60
Figura 19 – Passos da metodologia 101. ....	65
Figura 20 – Ontologia empresarial dinâmica. ....	66
Figura 21 – Ontologia de domínio do projeto de solução. ....	67
Figura 22 – Fluxograma de sequência de extração dos dados.....	72
Figura 23 – Grafo gerado com as triplas de município e UF, fatias de tempo e data.....	77
Figura 24 – Nuvem de palavras relativas aos tipos de recursos transferidos. ....	79
Figura 25 – Interface do SPARQL endpoint do projeto de solução.....	80
Figura 26 – Interface HTML para do recurso do município de Florianópolis. ....	81
Figura 27 – Página HTML para a fatia de tempo de Florianópolis em 2010.....	82
Figura 28 – Página HTML do recurso Florianópolis no dataset DBPedia.org. ....	83
Figura 29 – Consulta SPARQL em código Java. ....	84
Figura 30 – Recurso recebido por municípios por faixa de IDH.....	87
Figura 31 – Quantidade de Municípios por faixa de IDH. ....	87
Figura 32 – Recurso recebido por UFs por faixa de IDH.....	89

Figura 33 – Quantidade de UFs por faixa de IDH.....	90
Figura 34 – Arrecadação gerada por UF e sua faixa de IDH. ....	91
Figura 35 – Arrecadação recebida por UF e seu nível de IDH.....	92
Figura 36 – Recurso recebido por faixa de IDH de Educação. ....	93
Figura 37 – Quantidade de UFs por faixa de IDH de Educação. ....	94
Figura 38 – Arrecadação gerada por UF e a faixa de IDH de educação. ....	95
Figura 39 – Recurso recebido x diferentes índices do IDH por UFs.....	98
Figura 40 – Arrecadação gerada x diferentes índices do IDH por UFs.....	99
Figura 41 – Os 20 municípios que mais receberam recursos e seus IDHs.....	101
Figura 42 – Respostas da sexta pergunta.....	103
Figura 43 – Respostas da sétima pergunta. ....	104
Figura 44 – Respostas da oitava pergunta. ....	104
Figura 45 – Respostas da nona pergunta. ....	105
Figura 46 – Respostas da décima pergunta. ....	106
Figura 47 – Respostas da décima primeira pergunta.....	107
Figura 48 – Respostas da décima segunda pergunta. ....	108

## LISTA DE QUADROS

Quadro 1 – Dados, informação e conhecimento.....	20
Quadro 2 – Diferença entre dados tradicionais e dados semiestruturados. ....	22
Quadro 3 – Declarações em dados não estruturados baseadas em locais geográficas .....	27
Quadro 4 – Sintaxe RDF/XML .....	29
Quadro 5 – Conceituação especificada extensionalmente.....	30
Quadro 6 – Requisitos Funcionais.....	51
Quadro 7 – Requisitos Não Funcionais .....	51
Quadro 8 – Características dos relacionamentos das classes. ....	68
Quadro 9 – Consulta em SPARQL para retorno de informações de municípios e estados.....	71
Quadro 10 – Arquivos coletados e suas fontes.....	71
Quadro 11 – URIs base por tipo de elemento.....	73
Quadro 12 – Arquivos coletados e suas fontes.....	74
Quadro 13 – Criação do recurso país usando a API Jena e linguagem de programação Java. ....	75
Quadro 14 – Conectando um município com outros municípios de outros datasets.....	75
Quadro 15 – Exemplo de triplificação de relações da fatia de tempo município.....	76
Quadro 16 – Exemplo de triplificação de dados relacionados a recursos para municípios.....	78
Quadro 17 – Linhas de código para persistência do modelo de domínio no dataset.....	79
Quadro 18 – Consulta recurso recebido por municípios por IDH. ....	86
Quadro 19 – Consulta recurso recebido por estados, por IDH. ....	88
Quadro 20 – Consulta arrecadação gerada por UFs e seu IDH. ....	90
Quadro 21 – Consulta recurso recebido por UFs e seu IDH para educação. ....	92
Quadro 22 – Consulta arrecadação gerada por UFs e seu IDH para educação. ....	95
Quadro 23 – Consulta recursos recebidos por UFs e seu IDH para longevidade.....	96
Quadro 24 – Consulta arrecadação gerada por UFs e seu IDH para longevidade. ....	96
Quadro 25 – Consulta recurso recebido por UFs e seu IDH para renda. ....	97
Quadro 26 – Consulta recurso recebido por UFs e seu IDH para renda. ....	98
Quadro 27 – Consulta recurso recebido por Municípios e seus IDHs.....	100

## LISTA DE SIGLAS

CSV	<i>Comma Separated Value</i>
ENEM	Exame Nacional do Ensino Médio
HTTP	<i>Hypertext Transfer Protocol</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
IDH	Índice de Desenvolvimento Humano
IRI	<i>Internationalized Resource Identifier</i>
OWL	<i>Web Ontology Language</i>
RDF	<i>Resource Description Framework</i>
RDFS	<i>Resource Description Framework Schema</i>
SPARQL	<i>SPA Protocol and RDF Query Language</i>
UML	<i>Unified Modeling Language</i>
URI	<i>Universal Resource Identifier</i>
XML	<i>Extensible Markup Language</i>
W3C	<i>World Wide Web Consortium</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>13</b>
1.1	PROBLEMA DE PESQUISA .....	14
1.2	OBJETIVOS .....	16
1.2.1	Objetivo geral .....	16
1.2.2	Objetivos específicos .....	16
1.3	JUSTIFICATIVA .....	16
1.4	ESTRUTURA DO TRABALHO .....	18
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA.....</b>	<b>19</b>
2.1	DADOS, INFORMAÇÃO E CONHECIMENTO .....	19
2.2	DADOS ESTRUTURADOS, SEMIESTRUTURADOS E NÃO ESTRUTURADOS ..	21
2.3	DADOS ABERTOS.....	23
2.3.1	Dados Governamentais Abertos .....	24
2.4	REPRESENTAÇÃO DO CONHECIMENTO .....	25
2.4.1	Estrutura: Resource Description Framework (RDF) .....	26
2.4.2	Conceito: Ontologia.....	29
2.4.3	Semântica: Web Ontology Language (OWL 2).....	32
2.4.4	Dados Conectados .....	38
2.5	PERSISTÊNCIA E RECUPERAÇÃO DO CONHECIMENTO .....	41
2.6	CONSIDERAÇÕES FINAIS.....	42
<b>3</b>	<b>MÉTODO CIENTÍFICO.....</b>	<b>43</b>
3.1	CARACTERIZAÇÃO DO TIPO DE PESQUISA .....	43
3.2	ATIVIDADES METODOLÓGICAS .....	44
3.3	PROPOSTA DE SOLUÇÃO .....	45
3.4	DELIMITAÇÕES .....	47
<b>4</b>	<b>PROPOSTA DE SOLUÇÃO .....</b>	<b>48</b>
4.1	LINGUAGEM DE MODELAGEM UNIFICADA (UML) .....	48
4.2	MODELAGEM DO TRIPLIFICADOR.....	49
4.2.1	Elicitação de Requisitos .....	50
4.2.1.1	Requisitos Funcionais.....	50
4.2.1.2	Requisitos Não Funcionais .....	51
4.2.2	Diagrama de Atividades.....	52
4.2.3	Diagrama de Comunicação .....	53

<b>4.2.4</b>	<b>Diagrama de Sequência .....</b>	<b>54</b>
<b>4.2.5</b>	<b>Diagrama de Classe .....</b>	<b>56</b>
<b>4.2.6</b>	<b>Diagrama de Implantação .....</b>	<b>57</b>
<b>5</b>	<b>DESENVOLVIMENTO .....</b>	<b>58</b>
<b>5.1</b>	<b>FERRAMENTAS UTILIZADAS .....</b>	<b>58</b>
<b>5.1.1</b>	<b>Enterprise Architect.....</b>	<b>59</b>
<b>5.1.2</b>	<b>Java.....</b>	<b>59</b>
<b>5.1.3</b>	<b>Jena.....</b>	<b>59</b>
5.1.3.1	TDB .....	60
5.1.3.2	Fuseki .....	61
5.1.3.3	Ontology API.....	61
5.1.3.4	Inference API.....	61
<b>5.1.4</b>	<b>Pubby.....</b>	<b>62</b>
<b>5.1.5</b>	<b>Apache Tomcat®.....</b>	<b>62</b>
<b>5.1.6</b>	<b>Protégé.....</b>	<b>62</b>
<b>5.1.7</b>	<b>Git .....</b>	<b>62</b>
<b>5.1.8</b>	<b>GitHub.....</b>	<b>63</b>
<b>5.1.9</b>	<b>Eclipse.....</b>	<b>63</b>
<b>5.2</b>	<b>CENÁRIO.....</b>	<b>63</b>
<b>5.2.1</b>	<b>Ontologia de domínio .....</b>	<b>64</b>
<b>5.2.2</b>	<b>Execução.....</b>	<b>70</b>
5.2.2.1	Coleta de dados.....	70
5.2.2.2	Extração dos dados .....	72
5.2.2.3	Padronização de URIs .....	73
5.2.2.4	Mapeamento e ligação dos recursos .....	74
5.2.2.5	Triplificação dos dados.....	76
5.2.2.6	Acessando os dados .....	80
<b>5.2.3</b>	<b>Consultas ao domínio .....</b>	<b>84</b>
5.2.3.1	Análise dos dados .....	85
<b>5.3</b>	<b>AVALIAÇÃO.....</b>	<b>102</b>
<b>5.4</b>	<b>TRABALHOS CORRELATOS .....</b>	<b>109</b>
<b>6</b>	<b>CONCLUSÕES E TRABALHOS FUTUROS.....</b>	<b>111</b>
<b>6.1</b>	<b>CONCLUSÕES .....</b>	<b>111</b>
<b>6.2</b>	<b>TRABALHOS FUTUROS .....</b>	<b>113</b>

**REFERÊNCIAS .....115**

## 1 INTRODUÇÃO

Com a evolução da *Web* nas últimas décadas e sua crescente utilização, a informação, proveniente dos dados, vem ganhando importância progressiva. Dados que eram disponibilizados somente para uso interno das organizações públicas e privadas começaram a ser disponibilizados para o público em geral, fazendo com que fosse originado o movimento denominado dados abertos. Segundo a Open Knowledge International (2015), os dados são abertos, quando podem ser livremente usados, modificados e compartilhados por qualquer pessoa para qualquer finalidade.

De acordo com Departamento de Assuntos Econômicos e Sociais das Nações Unidas (2013), somente após a posse de Barack Obama, como presidente dos Estados Unidos, o termo governo aberto se tornou popular. A partir da inclusão de políticas para dados abertos no âmbito governamental, introduzidas no memorando: Transparência e Governo Aberto (ESTADO UNIDOS, 2009), o movimento dos dados abertos começou a seguir fortemente a vertente pública, o que serviu como um incentivo para que as outras nações seguissem o mesmo caminho.

Os dados abertos do governo brasileiro, fonte de dados dessa pesquisa, são um exemplo claro das dificuldades para a recuperação de dados em uma *Web* na qual seus documentos são a estrutura principal (BERNERS-LEE, 2009). Disponibilizados em arquivos, alguns em formatos proprietários, armazenados em diferentes bases e sem nenhuma integração, os dados abertos no Brasil se tornam um desafio para a análise de pesquisadores. Mais importante que apenas publicar os dados é publicá-los de forma que os mesmos possuam algum tipo de ligação. (BERNERS-LEE, 2009).

O objetivo central deste estudo de caso é a construção de uma base de conhecimento baseado em um conjunto específico de dados abertos encontrados em portais de dados do governo brasileiro, transformando estes dados no formato de dados conectados, que é um conjunto de boas práticas, criado por Tim Berners-Lee para a publicação de dados na *Web* – utilizando-as para fazer a sua correlação com bases abertas de dados conectados já existentes. Esse trabalho e o método nele desenvolvido devem também servir de referência para futuras publicações.

A análise dos dados abertos extraídos dos portais do governo brasileiro procura identificar e publicar de maneira padronizada o significado e interpretação (semântica) presente



nos documentos, visando reconhecer não somente suas classes (por exemplo, estado e cidade), mas também suas características relevantes para aplicações (por exemplo, coordenadas geográficas, índice de desenvolvimento humano (IDH), arrecadação em impostos, etc.). A amostra de dados extraída de portais brasileiros de dados abertos envolve diversos dados sobre cidades e estados brasileiros, assim como suas características temporais. Entidades e características do conjunto de dados serão definidas através de ontologia, que é a responsável por fornecer o entendimento comum e descrever o conhecimento acerca de um domínio que pode ser comunicado entre pessoas e sistemas (FENSEL, 2001). Deste modo, conhecimento em dados abertos, no contexto deste trabalho, trata de conceitualizar o domínio definido com relação aos dados coletados nos portais do governo brasileiro.

## 1.1 PROBLEMA DE PESQUISA

A maneira de gerar e consumir os dados sofreu mudanças significativas durante nossa história, a *Internet*, uma das maiores responsáveis por isso, mudou radicalmente a forma de compartilhar o conhecimento gerado pelos dados, diminuindo a barreira à publicação e acesso a documentos como parte de um espaço de informação global (BIZER; HEATH; BERNERS-LEE, 2009).

Originados nos mais diversos meios, o grande volume de dados, principalmente os abertos, são um fenômeno que pode desempenhar um papel importante em organizações do setor privado, e também em governos, trazendo assim resultados para a sua população. Por exemplo o *site* americano denominado "*data.gov*" expõe as diversas áreas em que os dados abertos podem atuar, desde negócios locais até manufatura, agricultura entre outros (DATA.GOV, 2016).

Com relação aos dados abertos, analisando somente a esfera de gestão pública, para manter todos os serviços públicos em perfeito funcionamento, muitos dados e informações precisam ser compartilhadas. Citando o exemplo do transporte público urbano, pode haver diversos veículos envolvidos, desde trens, metrô, ônibus entre outros, com a possibilidade de ter rotas municipais, intermunicipais e até interestaduais. Para que seja possível um serviço de transporte público de qualidade é essencial que os dados entre esses serviços sejam comunicáveis entre si aos órgãos competentes desse serviço e, principalmente, para as pessoas que serão afetadas. Os dados abertos possibilitam isso, que o usuário tenha o acesso as

informações que lhe são necessárias, trazendo diversas vantagens para o cotidiano da população, este benefício é estendido para diversas áreas. (ISOTANI; BITTENCOURT, 2015, p.18).

Entretanto, apesar dos benefícios indiscutíveis que os dados podem oferecer, até recentemente eles eram publicados seguindo os princípios da *Web* de documentos, que não são aplicáveis aos dados em seu estado original. (BIZER; HEATH; BERNERS-LEE, 2009). A maior parte dos dados na *Web* está incorporada a documentos (por exemplo. páginas da *Web*, documentos de valores separados por vírgula, documentos de texto), e, adjacente a isso, existe o fato de que parte desses documentos é de formato proprietário, o que se configura uma barreira para a expressividade, recuperação e interoperabilidade dos dados.

Segundo Tauberer (2012), formatos proprietários restringem quem pode utilizar os dados, como ele pode ser utilizado e compartilhado e como ele poderá ser usado no futuro. Se existe um controle exclusivo para determinado formato, têm-se um problema para os dados abertos, porque, os dados não são abertos, se não são abertos a todos.

Em resposta ao reconhecimento pelos direitos dos cidadãos, governos ao redor do mundo têm se engajado no conceito de “Governo Aberto” que inclui três elementos básicos: transparência, participação e colaboração (DEPARTAMENTO DE ASSUNTOS ECONÔMICOS E SOCIAIS DAS NAÇÕES UNIDAS, 2013). O governo brasileiro, em suas esferas e poderes, possui uma grande quantidade de dados abertos disponíveis a todos de forma livre, entretanto, existe uma série de problemas estruturais na disponibilização desses. Apesar de existir um portal específico para os dados abertos, o Portal Brasileiro de Dados Abertos<sup>1</sup>, nem todos os dados originados em esferas federais são publicados nessa ferramenta, há também uma falta de padrão e acessibilidade aos dados, principalmente pela busca ineficiente e parte dos dados estarem disponibilizados apenas em formatos proprietários.

Este trabalho se destina a análise do uso de dados conectados, com a criação de uma base de conhecimento para a publicação de dados abertos do governo brasileiro, como resposta aos problemas de expressividade, recuperação e interoperabilidade dos dados.

---

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

## 1.2 OBJETIVOS

Nesta seção são apresentados o objetivo geral e os objetivos específicos deste trabalho.

### 1.2.1 Objetivo geral

O objetivo geral deste trabalho é a criação de uma base de dados conectados, baseada em diversos dados disponibilizados através do portal de dados abertos de esferas governamentais brasileiras, de forma que estes dados componham uma base de conhecimento.

### 1.2.2 Objetivos específicos

Os objetivos específicos deste trabalho são:

- avaliar as ferramentas e tecnologias disponíveis para conexão e representação dos dados coletados;
- desenvolver um modelo de representação do conhecimento para o domínio de dados coletados;
- fazer a conversão do conjunto de dados em dados conectados, tomando como base a ontologia desenvolvida;
- realizar uma avaliação com base em experimentos a partir do domínio definido;
- identificar qual é o interesse, do público envolvido com a área de tecnologia da informação, a respeito dos dados governamentais abertos, e a aceitação do modelo proposto por este trabalho para a publicação dos mesmos.

## 1.3 JUSTIFICATIVA

O valor dos dados, em uma época em que os mesmos são gerados em escalas sem precedentes, e a falta de estruturação para o melhor aproveitamento desses são desafios determinantes para pesquisas acerca da publicação de dados abertos. Do mesmo modo que

grandes quantidades de dados são geradas a cada dia, grandes quantidades de dados abertos também são disponibilizadas para o acesso de todos. Entretanto, tradicionalmente, os dados publicados na *Web* têm sido disponibilizados como despejos brutos em formatos como CSV ou XML, ou marcados como tabelas HTML, o que sacrifica, e muito, sua estrutura e semântica. (BIZER; HEATH; BERNERS-LEE, 2009).

Bases de conhecimento, estruturadas nos princípios de dados conectados, ou seja, dados que possuem interligação entre si, possibilitam a interoperabilidade dos dados divulgados, assim como a combinação de diferentes fontes para os mesmos. Através do relacionamento e formato explícito dos dados conectados se tem a expressividade necessária para buscas por similaridade semântica nos dados publicados. Por conseguinte, o uso da prática de dados conectados proporciona uma melhor recuperação da informação e usabilidade dos dados em aplicações e humanos. (BERNERS-LEE, 2006)

Dessa forma para este trabalho foi escolhido um conjunto de dados coletados a partir de portais brasileiros de dados abertos como o caso de estudo deste trabalho, esses dados estão relacionados a uma ontologia, determinando a ligação desses dados. Um ponto de dificuldade, na relação entre os dados extraídos dos documentos, no conceito baseado na ontologia, é que o conjunto de dados, formado por diversos documentos de diferentes fins, não possui relações (por exemplo, hierárquicas, associativas) necessariamente explícitas, sendo o nome do município (características comuns da classe município) geralmente a única representação entre os dados extraídos. As relações devem ser explicitadas na conversão do conjunto de dados recuperados de documentos em dados conectados baseados nos conceitos da ontologia, considerando as características da taxonomia no conjunto de dados coletados de portais de dados abertos do governo brasileiro.

Conectar os dados abertos, sejam eles governamentais ou não, gera um grande benefício para a sociedade, tanto para empresas como para a população em geral, pois a conexão dos dados faz com que as informações que antes estavam “soltas” e sem significado possam constituir e interligar diversos outros valores que manualmente seria muito dispendioso. (ISOTANI; BITTENCOURT, 2015, p.22-23).

#### 1.4 ESTRUTURA DO TRABALHO

Este trabalho está estruturado em seis capítulos. O primeiro capítulo apresentou a introdução em que estão definidos o tema, problema de pesquisa, objetivo geral e específicos, justificativa e a presente seção que define a estrutura do trabalho. O segundo capítulo apresenta o referencial teórico acerca dos dados abertos e sua disponibilização no formato de dados conectados em uma base de conhecimento. O terceiro capítulo contempla o método utilizado no desenvolvimento do presente estudo. O quarto capítulo apresentará a modelagem do projeto de solução aplicada no estudo. O quinto capítulo apresentará as ferramentas e técnicas utilizadas na obtenção, estruturação, armazenamento e análise dos dados, o desenvolvimento do experimento e sua avaliação. Por fim, o sexto capítulo apresenta as conclusões do trabalho assim como propostas de trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

O presente capítulo apresenta os fundamentos necessários para o entendimento deste trabalho. Dessa forma, são abordados os seguintes temas: o que são os dados, a informação e o conhecimento, os tipos de dados – estruturados, semiestruturados e não estruturados – conceitos de dados abertos e dados governamentais abertos, representação, persistência e recuperação do conhecimento.

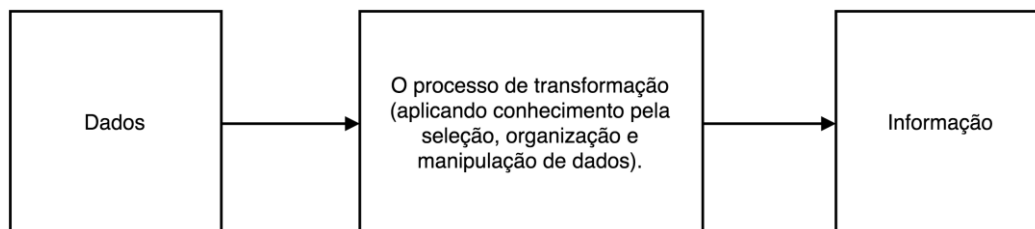
### 2.1 DADOS, INFORMAÇÃO E CONHECIMENTO

Dado na língua inglesa *Data* originada de *Datum* do latim pode possuir diversos significados conforme a área em que está sendo analisada (SORDI, 2008, p.7), para Davenport (1998, p. 19), "dados são observações dos estados do mundo", dessa forma, são definidos os conceitos que seguirão para o entendimento e desenvolvimento do experimento.

O dado trata-se de um fato bruto (por exemplo, número de matrícula, nome de aluno, etc.) ou ainda pode ser a representação de algo, como sons e imagens (AUDY, ANDRADE, CIDRAL, 2005, p. 93). Na definição de Sordi (2008, p.7), os dados são um conjunto de evidências, ou diversas evidências relevantes sobre um fato observado, já, para Setzer (1999, p. 2), o dado é "[...] uma sequência de símbolos quantificados ou quantificáveis. Portanto um texto é um dado."

As aplicações, em geral, hoje estão fundamentadas nos dados que possuem, fornecendo acesso aos mesmos, sejam desde simples visualizações até ferramentas mais sofisticadas, de diversas origens, com seus diversos significados (MASSACHUSETTS INSTITUTE OF TECHNOLOGY, 2013). Portanto, se houver manipulação e organização dos dados de forma que estes ganhem um propósito, o que é denominado de processamento de dados, originam-se as informações (SORDI, 2008, p.9). As informações podem ser extraídas de diversas fontes, sejam elas formais (por exemplo, relatórios das áreas de trabalho da organização), informais (por exemplo, conversas de funcionários), internas ou externas (AUDY, ANDRADE, CIDRAL, 2005, p. 96). A figura 1 exemplifica o processo de transformação de um dado em informação, segundo Stair e Reynolds (1999, p.5).

Figura 1 – O processo de transformação de dados em informação.



Fonte: Stair e Reynolds (1999, p.5).

Segundo Stair e Reynolds (1999, p.5), para que a informação possua valor é necessário que contemple o processo de transformação do dado em informação que é exposto na figura 1. Se a informação é, segundo Sordi (2008, p.10), "[...] a interpretação de um conjunto de dados segundo um propósito relevante e de consenso para o público-alvo (leitor).", conhecimento é, também, segundo Sordi (2008 p.10), "[...] o novo saber, resultante de análises e reflexões de informações segundo valores e modelo mental daquele que o desenvolve, proporcionando a este, melhor capacidade adaptativa às circunstâncias do mundo real".

O conhecimento é a informação de maior valor e de difícil gerenciamento. É originada devido a um contexto, um significado, atribuído por alguém, o conhecimento adquirido pode ser inserido em máquinas, porém é extremamente difícil a sua categorização e manutenção. (DAVENPORT, 1998, p. 19).

A seguir, o quadro 1, apresenta a diferença entre dados, informação e conhecimento.

Quadro 1 – Dados, informação e conhecimento.

Dados	Informação	Conhecimento
<p>Simples observações sobre o estado do mundo</p> <ul style="list-style-type: none"> <li>• Facilmente estruturado</li> <li>• Facilmente obtido por máquinas</li> <li>• Frequentemente quantificador</li> <li>• Facilmente transferível</li> </ul>	<p>Dados dotados de relevância e propósito</p> <ul style="list-style-type: none"> <li>• Requer unidade de análise</li> <li>• Exige consenso em relação ao significado</li> <li>• Exige necessariamente a mediação humana</li> </ul>	<p>Informação valiosa da mente humana Inclui reflexão, síntese, contexto</p> <ul style="list-style-type: none"> <li>• De difícil estruturação</li> <li>• De difícil captura em máquinas</li> <li>• Frequentemente tácito</li> <li>• De difícil transferência</li> </ul>

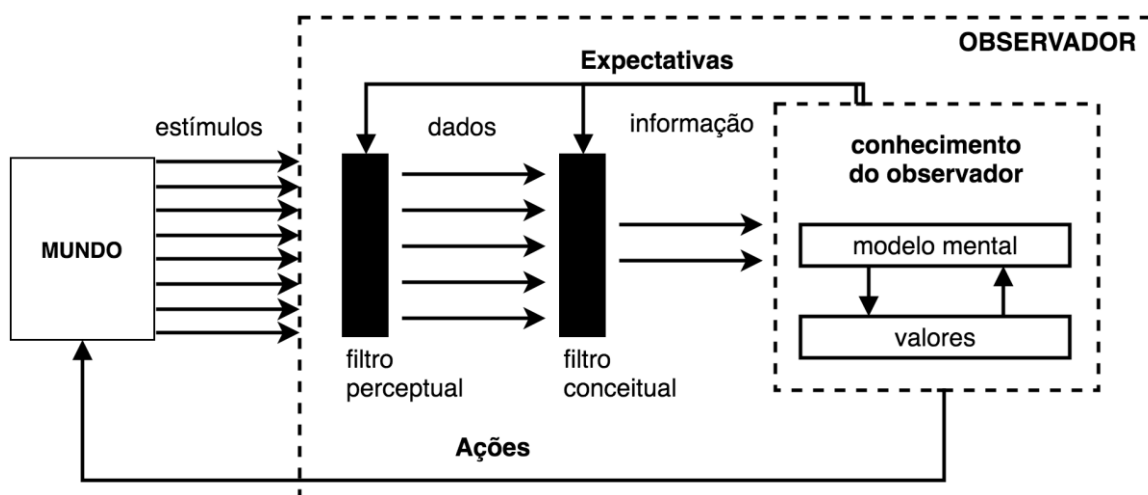
Fonte: (DAVENPORT, 1998, p. 18).

Para Sordi (2008, p.13), enquanto os dados e a informação possuem a veracidade facilmente verificada, pois a coleta dos dados e a geração da informação depende

exclusivamente do algoritmo para consolidação dos dados, não é possível assegurar a veracidade do conhecimento, já que este depende do modelo mental empregado.

A figura 2 demonstra como funciona a geração do conhecimento.

Figura 2 – Conhecimento segundo o agente gerador e beneficiário.



Fonte: (BOISOT; CANALS, 2004, apud SORDI, 2008, p. 13).

Baseando-se na figura 2, é possível concluir, ainda segundo Boisot e Canals (2004), que a informação é uma extração dos dados, que, ao realizar uma transformação, organização e manipulação desses, é capaz de realizar um trabalho de criação de conhecimento, utilizando os modelos mentais e valores pré-existent de um agente.

A geração do conhecimento é composta por diversos fatores, os elementos externos criam os dados, após, a manipulação dos dados, é gerada a informação, só então é que o observador aplica sobre a informação o seu modelo mental e valores também adquiridos de elementos externos para conceber o conhecimento.

Após feitas as distinções entre os conceitos de dados, informação e conhecimento, abordar-se-ão os tipos de estruturas de dados na seguinte seção.

## 2.2 DADOS ESTRUTURADOS, SEMIESTRUTURADOS E NÃO ESTRUTURADOS

Os dados são divididos em dados estruturados, semiestrutturados e não estruturados. Segundo Elmasri e Navathe (2011, p.280), para serem considerados dados estruturados, os



mesmos devem ser armazenados em bancos de dados, nesse tipo de classificação, os dados sempre são representados em um mesmo formato específico. Todos os dados devem seguir a estrutura do esquema em que serão armazenados. O esquema é projetado diligentemente para que todo os dados possam ser incluídos no banco de dados.

Os dados semiestruturados são coletados de forma aleatória, antes mesmo de conhecer a forma em que serão armazenados. Da mesma forma que os dados estruturados, os dados semiestruturados também podem possuir uma estrutura, porém sem a garantia de que todos os dados coletados irão dispor da mesma organização estrutural. Para os dados semiestruturados não existe um esquema pré-definido. Este padrão indefinido acaba caracterizando estes dados como semiestruturados (ELMASRI, NAVATHE, 2011, p.280).

No quadro 2 estão as principais características que diferem os dados semiestruturados para os estruturados:

Quadro 2 – Diferença entre dados tradicionais e dados semiestruturados.

<b>Dados tradicionais</b>	<b>Dados semiestruturados</b>
Esquema predefinido	Nem sempre há um esquema predefinido
Estrutura regular	Estrutura irregular
Estrutura independente dos dados	Estrutura embutida no dado
Estrutura reduzida	Estrutura extensa
Estrutura fracamente evolutiva	Estrutura fortemente evolutiva
Estrutura prescritiva	Estrutura descritiva
Distinção entre estrutura e dado é clara	Distinção entre estrutura e dado não é clara

Fonte: (MELLO et al, 2000).

Os dados tradicionais ou estruturados possuem um esquema pré-definido com uma estrutura homogênea, diferentemente dos dados semiestruturados, que, além de não possuírem um esquema pré-definido, têm introduzida a estrutura na própria descrição dos dados e possuem uma composição bastante heterogênea. Em outras palavras, cada dado semiestruturado pode ter sua própria formação estrutural. (MELLO et al, 2000).

Ao lado disso, há também os dados que não possuem nenhum tipo de estruturação, esses são chamados dados não estruturados. Para Inmon (2006), o mundo dos dados estruturados é muito disciplinado e ordeiro, o mundo dos dados não estruturados é muito indisciplinado e desordenado. Conforme dito por Elmasri e Navathe (2011, p.281), um exemplo de dado não estruturado "[...] é um documento de texto que contém informações incorporadas

a ele." As páginas HTML seguem esse exemplo, pois existe uma *tag* (tabulação), onde os sinais `< ... >` e `</ ... >` representam o início e o fim, respectivamente, essas etiquetas sinalizam para a formatação do documento, porém não é possível entender o significado dos dados que se encontram dentro dessas *tags*. (ELMASRI, NAVATHE, 2011, p.281).

## 2.3 DADOS ABERTOS

A Open Knowledge International é uma rede mundial sem fins lucrativos, que usa advocacia, tecnologia e treinamentos para defender o conhecimento abeto. De acordo com a Open Knowledge International (2015), os dados abertos são dados que devem ser usados livremente, podendo ser reutilizados ou redistribuídos por qualquer entidade ou pessoa, sujeitando-se, no máximo, à atribuição da fonte desses dados.

São pontos importantes, segundo a Open Knowledge International (2015), que definem os dados abertos:

- a) Disponibilidade de acesso: A disponibilização dos dados deve ser completa, não sendo permitido liberar parte dos dados. Para lucro em cima dos dados que serão liberados mediante pagamento, poderá ser cobrado um valor dos dados desde que o custo não seja maior do que o valor para reprodução desses dados. Estes devem preferencialmente disponibilizar a opção para baixar pela *Internet*;
- b) Reutilização e redistribuição: Todos os dados abertos devem permitir a sua redistribuição e reutilização, mesmo que estes sejam utilizados com conjunto de outros dados;
- c) Participação universal: Não deve haver discriminação sob determinadas áreas de atuação, organizações ou pessoas, todos devem possuir o mesmo direito de acesso para usar, reutilizar e/ou redistribuir os dados. Por exemplo, não é permitido a restrição para somente uso comercial, comunidade acadêmica e etc.

Os tópicos informados anteriormente garantem a interoperabilidade entre as organizações, sistemas ou pessoas, em que interoperar é a capacidade de diversos sistemas e organizações trabalharem em conjunto. Nesse caso, trata-se da capacidade de interoperar - ou combinar - diferentes conjuntos de dados. (OPEN KNOWLEDGE INTERNATIONAL, 2015).

Para a Open Knowledge International (2015), existem diversos tipos de dados que podem ser considerados abertos, desde dados culturais, financeiros, clima, transporte, entre

outros. Cada tipo de dado terá uma relevância diferente, esta dependerá da finalidade com que serão utilizados. Os principais propósitos para os dados abertos são a integração entre **valores sociais e comerciais**, para que as empresas consigam ter o entendimento das necessidades sociais e, conseqüentemente, ter um alcance de mercado maior e melhor; a **transparência dos dados**, com a publicação dos dados abertos amplia-se a **participação e engajamento** do público, que consegue ter mais acesso à informação, fazendo com que haja um aumento na qualidade e assertividade sobre a tomada de decisões.

A seção, a seguir, apresenta um segmento específico de dados abertos: os governamentais, objeto de estudo deste trabalho.

### 2.3.1 Dados Governamentais Abertos

“O movimento mundial pela publicação de dados governamentais em formatos abertos teve suas primeiras manifestações práticas a partir do trabalho desenvolvido pela MySociety no Reino Unido e GovTrack nos Estados Unidos”. (PORTAL BRASILEIRO DE DADOS ABERTOS, 2015). Para o Departamento de Assuntos Econômicos e Sociais das Nações Unidas (2013), o início do que é conhecido como dados governamentais abertos ocorreu com o lançamento dos oito princípios de dados governamentais abertos que são introduzidos pelas concepções iniciais dos dados abertos em geral, porém com especificações para a área governamental, segundo Malamud (2007), são eles:

1. Devem ser disponibilizados em sua totalidade, não havendo restrições de privacidade, segurança ou privilégios;
2. Devem estar na sua forma original, sem agregações ou modificações;
3. Os dados devem ser liberados mais rapidamente possíveis de forma a preservar o valor dos dados;
4. É necessário que os dados sejam acessíveis, disponíveis na *internet*, de modo que todos tenham capacidade para extraí-las;
5. Os dados devem possuir uma estruturação razoável para que seja possível a realização de um processamento automatizado;
6. Não deve ser considerado nenhum tipo de discriminação para qualquer grupo ou pessoa;

7. Os dados devem ser concedidos em formatos não proprietários, para que nenhuma entidade tenha o controle exclusivo sobre os mesmos, considerando que os formatos não proprietários podem não atingir um grande público se torna necessário a liberação dos dados em mais de um formato, sendo pelo menos um dos formatos disponíveis não proprietário;
8. Os dados devem possuir uma licença livre, não sendo possível a realização de nenhum tipo de patente, marca registrada e afins, pois os dados governamentais tratam-se de registros públicos, são combinações de diversos dados pessoais.

O que caracteriza os dados governamentais abertos é a livre utilização para qualquer propósito sem restrições associado ao fato de que se trata de dados de setores públicos, gerados junto a documentações por autoridades governamentais. Considera-se que nem todo setor de informação pública gera dados governamentais abertos, pois, para constituir um dado governamental aberto, é necessário a intersecção de três pontos importantes, o governo aberto, o dado aberto e o governo dos dados. (DEPARTAMENTO DE ASSUNTOS ECONÔMICOS E SOCIAIS DAS NAÇÕES UNIDAS, 2013).

Apesar da elaboração dos oito princípios de dados governamentais abertos em 2007, somente em 2009 os dados governamentais abertos ganharam força, após o lançamento de políticas abertas na eleição de Barack Obama, quando foi lançado um memorando com políticas de transparência e governo aberto. (DEPARTAMENTO DE ASSUNTOS ECONÔMICOS E SOCIAIS DAS NAÇÕES UNIDAS, 2013). Entretanto, as políticas de transparência do governo aberto brasileiro ocorreram somente em 2011, quando foi instituído o Plano de Ação Nacional sobre Governo Aberto com o Decretos / nº de 15 de setembro de 2011. (PORTAL BRASILEIRO DE DADOS ABERTOS, 2012).

Segundo Tauberer (2012), a disponibilização dos dados pode trazer diversas melhorias para a sociedade, promovendo avanços na economia e tornando os cidadãos mais conscientes e seguros, pois estes conhecem e possuem a informação disponível.

## 2.4 REPRESENTAÇÃO DO CONHECIMENTO

A representação do conhecimento é uma área de estudo preocupada em usar símbolos formais para representar uma coleção de proposições que possuam legitimidade acreditada por algum agente. (BRACHMAN; LEVESQUE, 2004).

A *Web* semântica é uma extensão *Web* atual, na qual a informação é dada em um sentido bem definido, possibilitando que computadores e pessoas trabalhem melhor em conjunto. (BERNERS-LEE; HENDLER; LASSILA, 2001). O termo “*Web* Semântica” se refere a uma visão da *Web* de dados - de datas, títulos; números, propriedades químicas e qualquer outro dado que se possa conceber. As tecnologias envolvidas nesse modelo permitem que pessoas criem armazenamento de dados na *Web*, construam vocabulários e escrevam regras para o tratamento dos dados. (WORLD WIDE WEB CONSORTIUM, 2015).

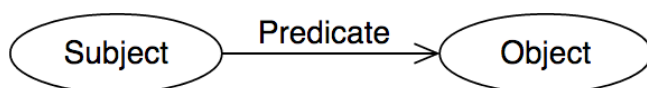
A capacidade de fornecer definições e significados processáveis por máquina é o que basicamente separa a *Web* semântica da *Web* de documentos, que tem foco apenas em apresentar as informações. (HORROCKS; PATEL-SCHNEIDER, 2011).

A presente seção apresenta as ferramentas utilizadas para a representação do conhecimento na *Web* de dados.

#### 2.4.1 Estrutura: *Resource Description Framework (RDF)*

O RDF é uma estrutura para representação de informação na *Web*, onde, basicamente, a informação é representada através de conjuntos de afirmações, denominadas *statements*. Os *statements* são compostos por três elementos: sujeito, predicado e objeto, o que conseqüentemente faz com que sejam referenciados também como triplas (PAN, 2009; CYGANIAK; WOOD; LANTHALER, 2014). A figura 3 demonstra o relacionamento entre sujeito e objeto, que resulta em uma tripla.

Figura 3 – Tripla RDF com dois nós (sujeito e objeto) e um predicado os conectando.



Fonte: Cyganiak, Wood e Lanthaler (2014).

O sujeito da declaração é o que é descrito por ela, e o predicado descreve o relacionamento entre o sujeito e o objeto. O quadro 3 apresenta declarações em dados não estruturados, que podem formar triplas relacionadas a locais geográficos.

### Quadro 3 – Declarações em dados não estruturados baseadas em locais geográficas

```

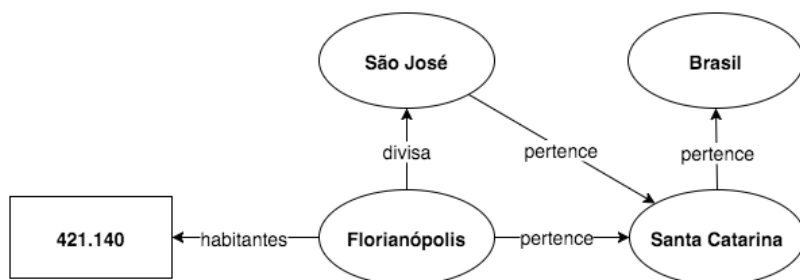
Florianópolis tem 421.240 habitantes
Florianópolis faz divisa com São José
Florianópolis pertence a Santa Catarina
São José pertence a Santa Catarina
Santa Catarina pertence ao Brasil

```

Fonte: Elaboração dos autores, 2016.

Conjuntos de afirmações no formato de triplas formam um grafo RDF. (PAN, 2009; CYGANIAK; WOOD; LANTHALER, 2014). A representação gráfica do conjunto de triplas apresentadas no quadro 3 é demonstrada na figura 4.

Figura 4 – Representação do grafo RDF baseado nos *statements* do quadro 3.



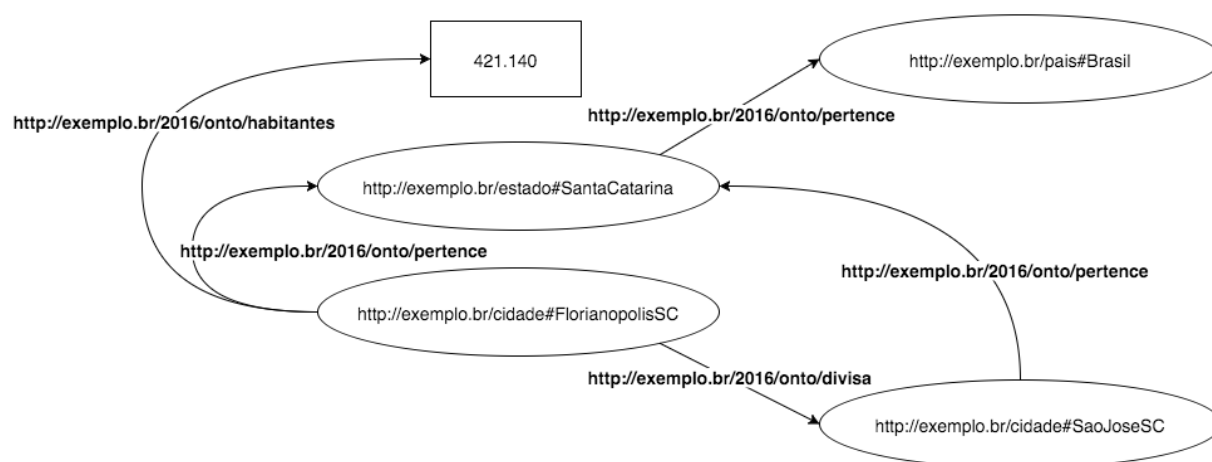
Fonte: Elaboração dos autores, 2016.

Os sujeitos e objetos de cada afirmação representam os nós, e predicados, as arestas. Existem três tipos de nós: recursos, literais e os *blank nodes*, esses tipos são definidos por Cyganiak, Wood e Lanthaler (2014) da seguinte maneira:

- **Literais** representam apenas objetos e são valores de dados concretos, como números ou texto.
- **Blank nodes** servem somente para indicar a existência de algo, sem o uso de uma identificação para qualquer coisa em particular.
- **Recursos** representam todo o restante, podendo ser sujeitos ou objetos. Em RDF, os recursos podem representar qualquer coisa que pode ser nomeada. Um recurso é, na verdade, nada mais que um nome que representa um objeto, ato ou conceito. Nomes de recursos assumem a forma de IRI (*Internationalized Resource Identifiers*).

Os predicados, também conhecidos como propriedades, representam o relacionamento entre os nós (recursos indicados pelo sujeito e objeto). Predicados, além de poder ser sujeitos ou objetos, são propriamente recursos, deste modo, predicados também são representados com IRIs. (CYGANIAK; WOOD; LANTHALER, 2014). IRI é definido pela RFC 3987 (DURST; SUIGNARD, 2005) como uma sequência de caracteres geralmente utilizada para representar palavras da linguagem natural, é um complemento do conhecido URI (*Uniform Resource Identifier*), RFC 3986 (BERNERS-LEE; FIELDING; MASINTER, 2005), uma sequência de caracteres que identificam recursos abstratos e físicos. Enquanto URIs estão limitados a conjuntos de caracteres ASCII, IRIs podem conter conjuntos de caracteres universal (Unicode / ISO 10646). Em grafos RDF, IRIs são utilizados para representar o nome dos recursos. De acordo com o Cyganiak, Wood e Lanthaler (2014), um IRI é um nome que é universalmente exclusivo, sendo assim, permanece válido em qualquer contexto e torna portátil a informação expressa em RDF. A figura 5 representa o grafo da figura 4, utilizando IRIs para identificar cada recurso.

Figura 5 – Grafo baseado na figura 4 representado com IRIs.



Fonte: Elaboração dos autores, 2016.

Para representar os grafos RDF de forma processável por máquinas, o RDF define uma sintaxe de linguagem de marcação extensível (XML) específica, o RDF/XML. (PAN, 2009). A fim de codificar o gráfico em XML, os nós e os predicados têm de ser representados em termos da linguagem – nomes de elementos, nomes de atributos, conteúdo de elementos e atributos. (GANDON; SCHREIBER, 2014). Um exemplo da sintaxe aplicada ao nó

`http://exemplo.br/cidade#FlorianopolisSC` do grafo da figura 5 é representada no quadro 4.

Quadro 4 – Sintaxe RDF/XML

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:onto="http://example.br/onto/">

<rdf:Description rdf:about="http://exemplo.br/cidade#FlorianopolisSC">
<rdf:type rdf:resource="http://www.exemplo.br/cidade"/>
<onto:habitantes rdf:datatype="decimal">421.140</habitantes>
<onto:pertence rdf:resource="http://www.exemplo.br/estado#SantaCatarina"/>
<onto:divisa rdf:resource="http://www.exemplo.br/cidade#SaoJoseSC"/>
</rdf:Description>
</rdf:RDF>
```

Fonte: Elaboração dos autores, 2016.

O uso de declarações RDF apenas não é suficiente para expressar a semântica e explicar o significado dos recursos, a utilização de uma ontologia se faz necessário para esse propósito. Ontologias, abordadas na seguinte seção, fornecem um vocabulário, conceitos, propriedades e suas definições, restrições, pressupostos básicos a respeito do significado pretendido do vocabulário, em um domínio que pode ser comunicado entre pessoas e sistemas heterogêneos e distribuídos (PAN, 2009).

### 2.4.2 Conceito: Ontologia

Durante sua história, o termo ontologia foi bastante restrito a esfera filosófica, em que é um ramo que trata da natureza e estrutura da “realidade” (GUARINO; OBERLE; STAAB, 2009). Entretanto, ontologias têm ganhado um papel importante em áreas da computação, como engenharia do conhecimento, processamento de linguagem natural, sistemas de informação cooperativos, integração inteligente de informação e gestão do conhecimento. (FENSEL, 2001; MCGUINNESS, 2003).

Ontologias fornecem um entendimento comum de um domínio que pode ser comunicado entre pessoas e sistemas heterogêneos amplamente difundidos, ou seja, contemplam apenas a teoria sobre um domínio e não a estrutura do conjunto de dados. (FENSEL, 2001). A definição mais adequada para o contexto deste trabalho é a de Brachman e Levesque (2004), que afirmam que a *Web Semântica* tem como principal característica



providenciar definições para objetos e tipos de objeto, acessíveis e manipuláveis em seu contexto, e que uma coleção desses tipos de definições sobre um determinado domínio é chamada de ontologia.

Uma definição amplamente citada na ciência da computação é a de Gruber. (1993, p.1, tradução nossa) que define ontologia como uma “especificação explícita de uma conceituação”, em que, ainda, segundo Gruber (1993), conceituação é uma visão abstrata e simplificada do mundo que queremos representar para algum propósito.

Em aplicações práticas, assim como na comunicação humana, é preciso usar uma linguagem para se referir aos elementos de uma conceituação: por exemplo, para expressar um mundo onde, A coopera com B, e B reporta-se a C, deve-se especificar símbolos, formalmente: predicados, como *coopera-com* e *reporta-se-a*, para que seja representada uma relação conceitual entre os indivíduos. Desse modo, têm-se uma linguagem que se compromete com uma conceituação. Supondo que na linguagem estabelecida, os elementos não lógicos (assinatura e vocabulário) são elementos do conjunto  $\{A, B, C, \textit{coopera-com}, \textit{reporta-se-a}\}$ , não existe a certeza que tais elementos serão interpretados de acordo com a conceituação proposta. Para evitar problemas de interpretação da conceituação, deve-se especificá-la de forma explícita. (GUARINO; OBERLE; STAAB, 2009).

Uma especificação explícita significa que o modelo deve ser especificado em alguma linguagem inequívoca, tornando-o passível de processamento por máquinas, bem como por seres humanos. (BAADER; HORROCKS; SATTLER, 2009). A princípio, pode-se especificar explicitamente uma conceituação de duas maneiras: *extensionalmente* e *intensionalmente*. Para especificar uma conceituação explicitamente de maneira *extensional*, deve-se listar as extensões de cada relação conceitual. (GUARINO; OBERLE; STAAB, 2009). O quadro 5 apresenta a conceituação, do mundo definido no exemplo anterior, especificada explicitamente de maneira *extensional*.

Quadro 5 – Conceituação especificada *extensionalmente*

A	<i>coopera-com</i>	B
B	<i>coopera-com</i>	A
B	<i>reporta-se-a</i>	C

Fonte: Elaboração dos autores, 2015.

A especificação explícita de modo *extensional* é algo inviável ou no mínimo não prático na maioria dos casos que envolvam muitos relacionamentos. De um modo geral, é mais eficaz explicitar uma conceituação, fixando a linguagem que se pretende usar, restringindo as interpretações dessa linguagem de forma *intensional* por meio de premissas adequadas. Por exemplo, pode-se escrever premissas simples afirmando que *coopera-com* é uma relação simétrica e *reporta-se-a* é uma relação assimétrica. (GUARINO; OBERLE; STAAB, 2009), ou seja, explicitando-se que A *coopera-com* B, entende-se também que B *coopera-com* A, entretanto, se explicitado que B *reporta-se-a* C, fica entendido que C não *reporta-se-a* B.

Ontologias podem ser classificadas de acordo com o nível de generalidade ou especialidade. Guarino (1998) propõe as seguintes classificações:

- **Ontologias de nível superior** (*top-level ontology*): descrevem conceitos genéricos como espaço, tempo, objetos, ações, etc., sendo assim independentes de um problema particular ou domínio.
- **Ontologias de tarefa** (*task ontology*): descrevem o vocabulário relacionado a uma tarefa genérica, como diagnosticar ou vender.
- **Ontologias de domínio** (*domain ontology*): descrevem o vocabulário relacionado a um domínio genérico, como medicina ou família.
- **Ontologias de aplicação** (*application ontology*): descrevem conceitos, que dependem simultaneamente de um domínio particular e tarefas e, muitas vezes, são especializações de ambas as ontologias.

O conhecimento em ontologia pode ser especificado usando cinco tipos de elementos: classes ou conceitos, relações, funções, instâncias e axiomas. (GRUBER, 1993). As definições para esses elementos, baseadas nas publicações de Gruber (1993), Corcho e Gómez-Pérez (2004), são dadas a seguir:

- **Classes ou conceitos** podem representar qualquer coisa abstrata ou concreta como a descrição de uma tarefa, função, ação, etc., geralmente, são organizadas em uma taxonomia. As taxonomias são utilizadas para organizar o conhecimento ontológico definindo classes e o relacionamento entre elas.
- **Relações** representam um tipo de interação entre classes de um domínio.
- **Funções** são consideradas um tipo especial de relação que representam eventos passíveis no contexto da ontologia.

- **Instâncias** representam elementos de um determinado conceito relacionado a uma classe, como, por exemplo, em um domínio composto pela classe número, *zero* pode ser uma instância da mesma.
- **Axiomas** são frases modelo que representam uma verdade absoluta, são usadas na ontologia para restringir suas informações, verificar sua correção ou deduzir novas informações.

Ontologias são utilizadas, em muitos casos, para melhorar o acesso aos dados. Desse modo, os dados existentes devem estar mapeados a conceitos e relacionamentos na ontologia correspondente. Uma abordagem para esse mapeamento entre dados e ontologia é feita através de uma representação RDF dos dados disponíveis, no qual os dados são mapeados a conceitos da ontologia. (HERTEL; BROEKSTRA; STUCKENSCHMIDT, 2009).

No RDFS, recursos pré-definidos como *rdfs:Class*, *rdfs:Resource*, *rdf:Property*, podem ser usados para definir classes, recursos e propriedades, respectivamente. (PAN, 2009, BRICKLEY; GUHA, 2014). Entretanto, a expressividade do RDFS é extremamente limitada, o que fez com que uma linguagem *Web* padronizada de ontologia fosse definida: a OWL. (ANTONIOU; VAN HARMELEN, 2009). A versão atual OWL 2 (W3C OWL WORKING GROUP, 2012), abordada na seguinte seção.

### 2.4.3 Semântica: *Web Ontology Language* (OWL 2)

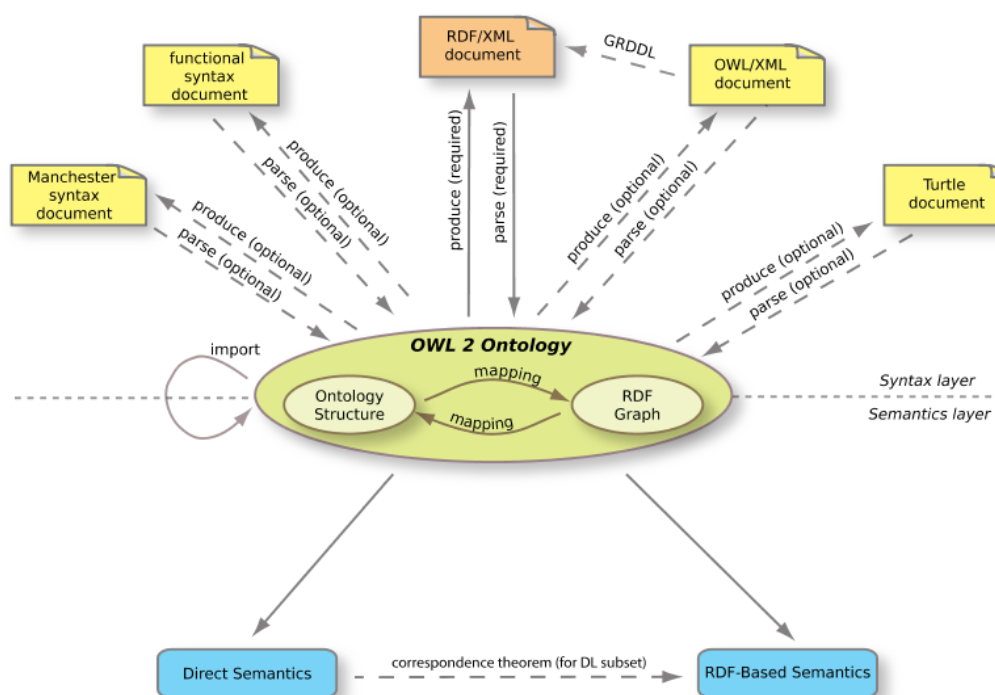
*Web Ontology Language* (OWL 2) é uma linguagem de ontologia projetada para a *Web* Semântica com significado formalmente definido. (W3C OWL WORKINGGROUP, 2012). A OWL 2 é uma extensão e revisão da OWL (MCGUINNESS et al., 2004), baseadas em RDF e RDFS.

Linguagens de ontologia, de acordo com Antoniou e van Harmelen (2009), possibilitam escrever de forma explícita conceituações formais de modelos de domínio, desde que possuam, entre outras características, sintaxe e semântica bem definidas. A sintaxe bem definida se faz necessária para o processamento da informação por máquinas, e a semântica para descrever com precisão o significado do conhecimento, que pode ser interpretado por homens e máquinas.

A visão geral da linguagem OWL 2 é representada na figura 6. A elipse no centro representa a noção de uma ontologia, que pode ser pensada tanto como uma estrutura abstrata

ou como um grafo RDF. No topo, são apresentadas as sintaxes que podem ser usadas para serializar e distribuir ontologias. Na parte inferior, são apresentadas duas especificações semânticas que definem o significado de ontologias OWL 2. (W3C OWL WORKING GROUP, 2012).

Figura 6 – Estrutura da OWL 2



Fonte: W3C OWL Working Group (2012).

Para que ontologias possam ser armazenadas e distribuídas, na prática, é preciso que utilizem uma sintaxe concreta. Como a OWL é baseada em RDF e RDFS, a sintaxe padrão da linguagem é a RDF/XML, essa sintaxe fornece a interoperabilidade entre as ferramentas OWL 2. (HORROCKS; PATEL-SCHNEIDER, 2011). Outras sintaxes também podem ser utilizadas junto a OWL 2, essas são apresentadas pelo W3C OWL Working Group (2012) como:

- **Turtle**<sup>2</sup>: Facilita a leitura e escrita de triplas RDF.
- **OWL 2 XML**<sup>3</sup>: Facilita o processamento, utilizando ferramentas XML.

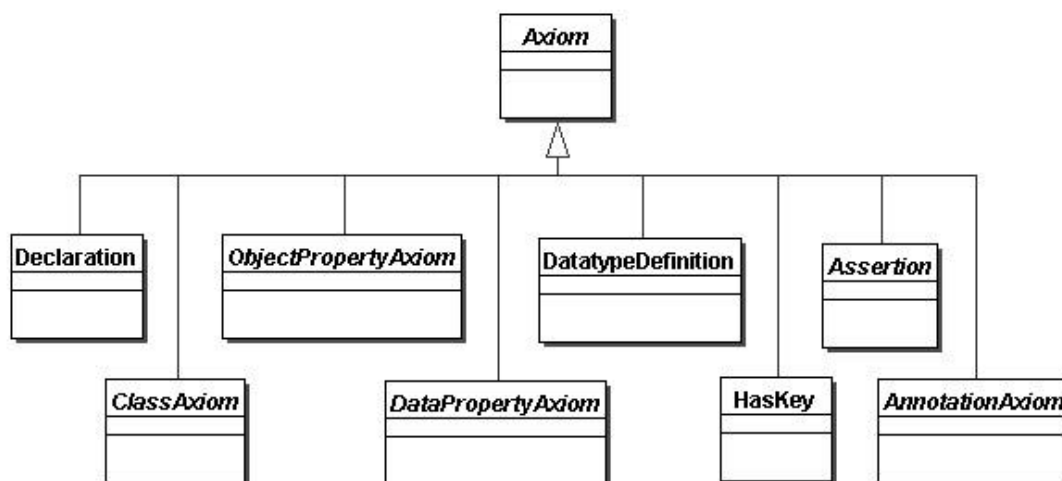
<sup>2</sup> Disponível em: <http://www.w3.org/TR/turtle/>

<sup>3</sup> Disponível em: <http://www.w3.org/TR/owl2-xml-serialization/>

- **OWL 2 Manchester Syntax**<sup>4</sup>: Facilita a leitura e escrita de Ontologias DL.
- **Functional-Style Syntax**<sup>5</sup>: Facilita a visualização da estrutura formal de ontologias.

O principal componente de uma ontologia OWL 2, de acordo com Bock et al. (2012) e Horrocks e Patel-Schneider (2011), é o seu conjunto de axiomas, que são declarações que dizem o que é verdadeiro no domínio. Axiomas em OWL 2, representados na figura 7, podem ser declarações a respeito de: classes, propriedades de objeto e de dados, definições de tipo de dados, chaves, afirmações e anotações (BOCK et al., 2012).

Figura 7 – Axiomas da OWL 2



Fonte: Bock et al. (2012).

Ontologias OWL 2 fornecem classes, propriedades, indivíduos e valores de dados que são armazenados como documentos da *Web Semântica*. (W3C OWL WORKINGGROUP, 2012). Nomes de indivíduos, por exemplo, "John", referem-se a elementos do domínio; classes, por exemplo, "Universidade", descrevem conjuntos de indivíduos com características semelhantes; propriedades, por exemplo, "estudaEm", descrevem as relações entre pares de indivíduos (como "John estudaEm Oxford"); e tipos de dados, por exemplo, "inteiro", descrevem conjuntos de valores de dados, como, "18". (HORROCKS; PATEL-SCHNEIDER,

<sup>4</sup> Disponível em: <http://www.w3.org/TR/owl2-manchester-syntax/>

<sup>5</sup> Disponível em: <http://www.w3.org/TR/owl2-syntax/>

2011). As definições para os elementos da linguagem OWL 2 mencionados anteriormente, com base nas publicações de Bock et al. (2008) e Horrocks, Patel-Schneider (2011), são abordadas a seguir:

- **Classes** são entendidas como conjuntos de indivíduos. A OWL 2 possui duas classes, construídas na própria linguagem e representadas com os seguintes IRIs: *owl:Thing*, e *owl:Nothing*. A classe *Thing*, é uma classe que representa o conjunto de todos os indivíduos, e a classe *Nothing* representa um conjunto vazio. Em um exemplo, as classes *a:Crianca* e *a:Pessoa*, respectivamente, podem ser usadas para representar conjuntos de crianças e pessoas no domínio da aplicação. O axioma **SubClassOf(a:Crianca a:Pessoa)** pode ser usado para definir que toda criança é uma pessoa.
- **Tipos de dados** são entidades que se referem a conjunto de valores de dados, assim são análogos as classes, tendo como principal diferença conter valores de dados como cadeia de caracteres e números, ao contrário das classes que contêm indivíduos. Um exemplo é o tipo de dado *xsd:integer* que denota um conjunto de inteiros. O axioma **DataPropertyRange(a:temIdade xsd:integer)** pode ser usado para definir que o alcance dos valores da propriedade *a:temIdade* é um conjunto de inteiros.
- **Propriedades de objeto** são utilizadas para conectar pares de indivíduos, como por exemplo, a propriedade de objeto *a:paiDe* pode ser usada para representar o relacionamento de paternidade entre indivíduos. O axioma **ObjectPropertyAssertion(a:paiDe a:Peter a:Chris)** pode ser usado para representar que Peter é pai de Chris.
- **Propriedades de dados** são utilizadas para conectar indivíduos a literais, por exemplo, a propriedade *a:temNome* pode ser utilizada para representar o relacionamento de um indivíduo e seu nome. O axioma **DataPropertyAssertion(a:temNome a:Peter "Peter Griffin")** pode ser utilizado para definir que o nome do indivíduo Peter é "Peter Griffin".
- **Propriedades de anotação** são usadas para providenciar anotações a uma ontologia, axioma ou IRI, por exemplo, a propriedade *rdfs:comment* pode ser

utilizada para representar uma anotação sobre algum indivíduo. O axioma **AnnotationAssertion**(*rdfs:comment* *a:Peter* "The fatheroftheGriffinfamilyfromQuahog.") pode ser usado para fornecer para a IRI *a:Peter* a informação de que ele é o pai da família Griffin de Quahog.

- **Indivíduos** são usados para representar objetos reais a partir de um domínio. Pode haver dois tipos de indivíduos, os nomeados e os anônimos. Os indivíduos nomeados possuem um nome explícito que pode ser usado para representá-los em qualquer ontologia, por exemplo, *a:Peter* pode ser usado para representar uma pessoa em particular. O axioma **ClassAssertion**(*a:Persona:Peter*) pode ser usado para definir que Peter é uma pessoa. Entretanto, os indivíduos anônimos, análogos aos *blanknodes* do RDF não possuem um nome global, ficando restritos apenas a ontologia que fazem parte. Indivíduos anônimos podem ser usados, por exemplo, para representar objetos cuja identidade não é relevante, como o endereço de uma pessoa. O axioma **ObjectPropertyAssertion**(*a:livesAta:Peter* *\_:a1*) pode ser usado para definir que Peter vive em algum lugar (desconhecido).
- **Literais** representam valores de dados que tenham como tipo cadeias de caracteres ou números inteiros, por exemplo, "1"^^*xsd:integer* é um literal que representa o inteiro "1" e "Family Guy"^^*rdf:PlainLiteral* é um literal que representa o par de cadeia de caracteres "Family Guy".

A semântica em OWL 2, de acordo com Motik, Patel-Schneider e Grau (2012), é definida por axiomas e ontologias, e é um pré-requisito para apoio ao raciocínio, que permite verificar a consistência da ontologia e do conhecimento, verificar relações não-intencionais entre classes e classificar automaticamente instâncias às classes (ANTONIOU; VAN HARMELEN, 2009).

Segundo Smith, Welt e McGuinness (2009) e também Motik, Patel-Schneider e Parsia (2012) as associações dos objetos podem possuir as seguintes características:

- **Propriedade funcional:** numa propriedade funcional pode-se dizer que, se *P* é uma propriedade funcional então para todos *x*, *y* e *z* são *P(x, y)* e *P(x, z)* então *y = z*. Se a propriedade *P* é funcional, então para cada indivíduo *x* pode haver um indivíduo distinto de *y*;

- **Propriedade inversa funcional:** considerando que agora  $P$  é uma propriedade inversa funcional, e para  $x$ ,  $y$  e  $z$  têm-se a seguinte relação de  $P(y, x)$  e  $P(z, x)$  implica que  $y = z$ . Pensando em elementos no sentido de banco de dados para OWL: Inversa Funcional implica que os elementos da extensão do domínio fornecerão um identificador único;
- **Propriedade reflexiva:** uma propriedade é reflexiva quando essa é usada para ligar indivíduos a eles mesmos;
- **Propriedade irreflexiva:** uma propriedade é irreflexiva quando não é usada para ligar indivíduos a eles mesmos;
- **Propriedade simétrica:** uma propriedade pode ser considerada simétrica quando dois indivíduos estão interligados, ou seja,  $x$  está ligado à  $y$ , como também  $y$  está ligado a  $x$ ; podendo ser representada pela seguinte função  $P(x, y)$  se  $P(y, x)$ ;
- **Propriedade assimétrica:** um objeto é considerado assimétrico quando por exemplo, um indivíduo  $x$  está ligado a  $y$  mas  $y$  não está ligado a  $x$ ;
- **Propriedade transitiva:** Se uma propriedade,  $P$ , é especificada como transitivo, então, para qualquer  $x$ ,  $y$  e  $z$ ,  $P(x, y)$  e  $P(y, z)$  então  $P(x, z)$ , ou seja, um objeto é transitivo quando por exemplo este está ligado de  $x$  à  $y$  e de  $y$  à  $z$ , de forma que  $x$  também está ligado a  $z$ .

É de fundamental importância a definição correta das propriedades citadas anteriormente dentro da ontologia, e esta por muitas vezes acaba sendo negligenciada, toda a forma que os elementos dentro da ontologia serão classificados poderão sofrer impacto caso ocorra alguma definição errada desta etapa. (ISOTANI; BITTENCOURT, 2015, p.143).

Para conhecimento ontológico, Antoniou e van Harmelen (2009) definem que pode raciocinar sobre:

- **Filiação de classes:** Se  $x$  é uma instância da classe  $C$ , e  $C$  é uma subclasse de  $D$ , então pode-se inferir que  $x$  é uma instancia de  $D$ ;
- **Equivalência de classes:** Se a classe  $A$  equivale a classe  $B$ , e a classe  $B$  equivale a classe  $C$ , então  $A$  é equivalente a  $C$ , também;



- **Consistência:** Supondo que  $x$  é declarado como uma instância da classe  $A$ , que é uma subclasse de  $B$ , e  $A$  e  $B$  são declaradas disjuntas. Então, temos uma inconsistência que aponta para um erro provável na ontologia;
- **Classificação:** Se nós declaramos que determinados pares de propriedade-valor são condições suficientes para a filiação a uma classe  $A$ , em seguida, se um indivíduo  $x$  satisfaz essas condições, então  $x$ , deve ser uma instância de  $A$ .

A atribuição de significado a ontologias OWL 2 pode ser dada através de dois modos: *OWL 2 DirectSemantics*<sup>6</sup>, que atribui significado diretamente sobre as estruturas da ontologia e *OWL 2 RDF-BasedSemantics*<sup>7</sup>, que atribui significado diretamente para os grafos RDF e assim indiretamente a ontologias através de mapeamento dos grafos RDF, visto que qualquer ontologia OWL 2 pode ser mapeada para RDF. Essas duas semânticas são utilizadas por *reasoners*, – *softwares* capazes de inferir consequências lógicas através de um conjunto de fatos ou axiomas (ABBURU, 2012), e outras ferramentas, por exemplo, para responder a consistência de classe, subsunção e consultas para a recuperação de instâncias. (W3C OWL WORKINGGROUP, 2012).

*Reasoners*, de acordo com Antoniou e van Harmelen (2009), também podem ser construídos para verificar a consistência de uma ontologia OWL, ou seja, verificar se existem contradições lógicas nos axiomas da ontologia. Além disso, *reasoners* podem ser utilizados para derivar inferências a partir das informações afirmadas, como, por exemplo, inferir se um conceito particular em uma ontologia é um sub-conceito de outra, ou se um indivíduo em particular em uma ontologia pertence a uma classe específica.

#### 2.4.4 Dados Conectados

O termo dados conectados, ou ligados (*linked data*) se refere a um conjunto de boas práticas para a publicação e ligação de dados estruturados na *Web*. (BIZER; HEATH; BERNERS-LEE, 2009).

Existem algumas recomendações técnicas importantes para a publicação dos dados governamentais abertos segundo Beners-Lee (2009):

---

<sup>6</sup> Disponível em: <http://www.w3.org/TR/owl2-direct-semantics/>

<sup>7</sup> Disponível em: <http://www.w3.org/TR/owl2-rdf-based-semantics/>

- Escolher URIs que podem ser facilmente persistidas;
- Utilizar metadados RDF disponibilizando suas licenças;
- Utilizar padrões RDF e SPARQL;
- As páginas devem ser acessíveis para as pessoas;
- Não utilizar arquivos compactados, a menos que possuam também uma versão direta disponível;
- Não inserir dados somente em formatos proprietários;
- Não esperar até possuir um esquema completo ou ontologia para publicar os dados;
- Não substituir os sistemas de dados já existentes.

Para que haja utilidade nos diversos dados disponíveis na *Web*, é preciso que estes possuam algum tipo de sentido, segundo Heath e Bizer (2011) assim como a *World Wide Web* revolucionou a forma de conectar pessoas e consumir documentos, também pode revolucionar a forma como utilizar e conectar os dados. "A *Web* Semântica não é somente colocar dados na web. É fazer ligações, de modo que uma pessoa ou máquina possa explorar a *Web* de dados. Com dados conectados, quando você tem algo disso, você pode encontrar outros dados relacionados.". (BERNERS-LEE, 2006, tradução nossa).

Existem quatro regras para a disponibilização de dados conectados segundo, Beners-Lee (2006):

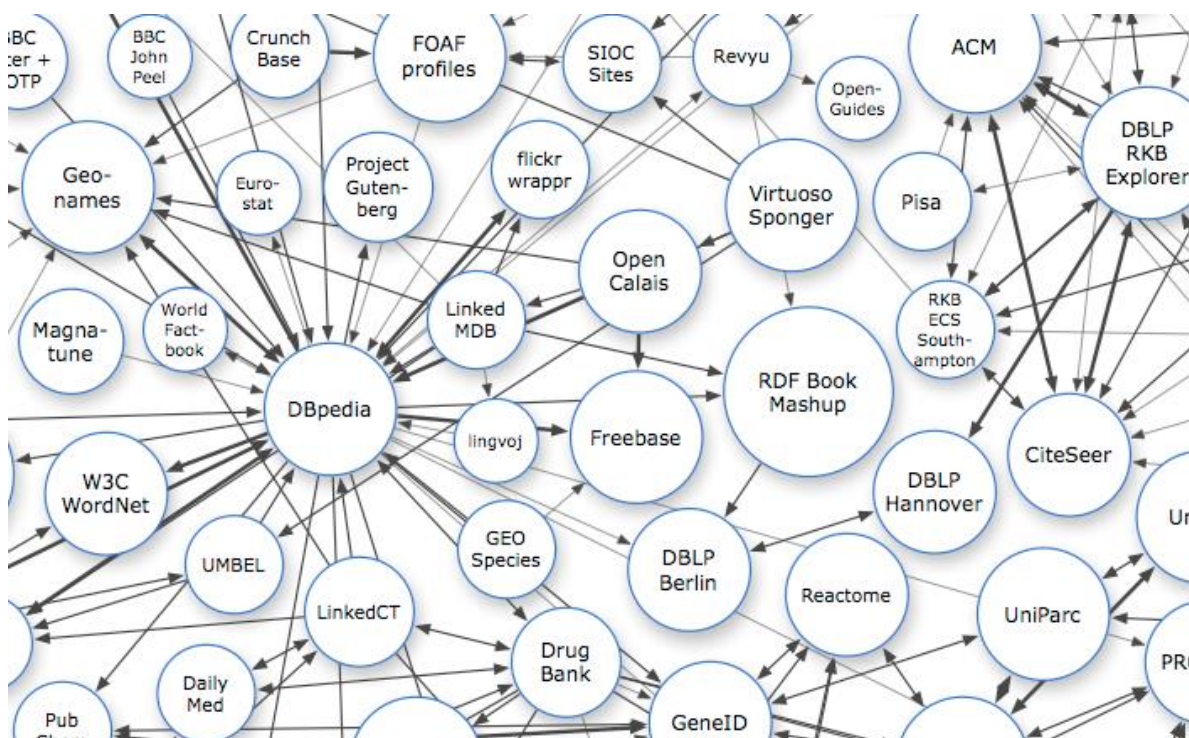
1. Utilizar URIs para identificação dos dados conectados;
2. Usar como padrão HTTP URIs para identificar objetos e conceitos abstratos, para assim possibilitar o acesso e utilização dos dados na *Internet*.
3. Quando alguém busca por um URI, este deve fornecer informações úteis utilizando os padrões RDF e SPARQL.
4. Fazer conexões para outros URIs, é necessário conectar os dados para que assim se possa encontrar mais informações relevantes.

A vinculação dos dados disponibilizados na *Web* requer uma ferramenta padrão que descreve os itens e o significado das conexões dos dados descritos. "O *linked data* constrói diretamente na arquitetura *web* e aplica essa arquitetura para o compartilhamento dos dados em escala global". (HEATH; BIZER, 2011, tradução nossa).

Para Bizer, Heath e Beners-Lee (2009), os princípios e práticas de dados conectados têm sido adotados por diversas organizações, que resultam em bilhões de triplas RDF criadas.

A nuvem de dados conectados, apresentada na figura 8, é uma representação dos conjuntos de dados publicados como dados conectados em agosto de 2014, cada círculo representa um conjunto de dados, esses são divididos por áreas de aplicação e representadas por cores diferentes. O tamanho de cada círculo é relativo ao número de triplas que o conjunto de dados possui, os mais largos chegam a ter aproximadamente um bilhão de triplas, enquanto os menores se aproximam de dez mil triplas. As setas entre os círculos indicam a existência de ligações entre conjunto de dados, essas ligações são triplas RDF, onde o sujeito e objeto são URIs de conjuntos diferentes. (SCHMACHTENBERG et al., 2014).

Figura 8 – Parte do diagrama da nuvem de dados abertos conectados em 2014



Fonte: Schmachtenberg et al. (2014).

Da mesma forma que o advento da *Web* acarretou em um cataclismo na publicação e consumo de documentos, na mesma proporção, os dados conectados têm o potencial de criar uma revolução na maneira em como os dados podem ser consumidos. Os dados conectados

contemplam a visão de evolução dos dados, tornando um bem global, permitindo uma evolução significativa da *web*, atingindo todo o seu potencial.

## 2.5 PERSISTÊNCIA E RECUPERAÇÃO DO CONHECIMENTO

Para que a *Web Semântica* funcione, é preciso que computadores tenham acesso a coleções estruturadas de informação e a um conjunto de regras de inferência que podem ser usadas para conduzir um raciocínio automatizado. (BERNERS-LEE; HENDLER; LASSILA, 2001).

Enquanto pequenos grafos RDF podem ser manipulados de forma eficiente e gerenciados na memória principal de um computador, grafos RDF maiores tornam indispensável o uso de sistemas de armazenamento persistente. *RDF Store* (banco de RDF), também conhecidos como *triplestore* (banco de triplas), são bancos de dados construídos para o armazenamento e recuperação de qualquer tipo de dados expresso em RDF (HASLHOFER, 2011).

O acesso aos dados em um *RDF Store* é realizado através de consultas formuladas em uma linguagem padrão, o SPARQL. (PRUD'HOMMEAUX; SEABORNE, 2008). A linguagem de consulta do SPARQL é baseada na correspondência de padrões em um grafo. O padrão mais simples em um grafo é o padrão tripla, que é como uma tripla RDF, porém, comum a variável no lugar de sujeitos, predicados e objetos, ao invés de termos RDF. A combinação dessas triplas gera um padrão básico no grafo, que é baseado em correspondências exatas. (HERTEL; BROEKSTRA; STUCKENSCHMIDT, 2009). Por exemplo, para obter todas as instâncias da classe “Pessoa” contida no vocabulário `http://xmlns.com/foaf/0.1/`, deve-se usar a seguinte consulta:

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?i
WHERE {
    ?i rdfs:type foaf:Pessoa .
}
```

Para tornar as consultas menores e mais fáceis de ler, de acordo com Hertel, Broekstra e Stuckenschmidt (2009), o SPARQL permite a criação de prefixos, como o **foaf**, demonstrado no exemplo acima, utilizado para encurtar o endereço do recurso.

O conhecimento de base necessário para especificar de forma clara o significado dos dados representados em RDF pode ser adquirido usando a linguagem de ontologia padrão OWL 2. Quando uma ontologia é usada para aumentar a semântica dos dados RDF, as respostas de consultas precisam considerar triplas adicionais, cuja existência é implicada pela combinação da ontologia com os dados. Subconjuntos da OWL 2, como o perfil “RL”, destinado a aplicações que exigem raciocínio escalável sem sacrificar muito poder expressivo (MOTIK et al., 2012), são utilizados para lidar com essa situação. O uso de *reasoners* RL providencia suporte robusto e escalável para respostas a consultas SPARQL baseadas em ontologias OWL 2 RL e conjunto de dados RDF (ZHOU et al., 2013).

## 2.6 CONSIDERAÇÕES FINAIS

O presente trabalho tem como foco principal a criação de uma base de conhecimento baseada em dados abertos, sendo que, dentre as diversas áreas de aplicação desses, optou-se pelos dados governamentais do Brasil.

Desenvolveu-se uma revisão da literatura buscando chegar ao estado da arte. Tem-se como objetivo utilizar as teorias apresentadas para auxiliar na criação da base de conhecimento.

Identificou-se que a informação deverá ser representada com sentido bem definido, de modo a permitir o seu processamento por pessoas e máquinas. Uma ontologia será necessária para especificar de forma clara o significado dos dados. E o uso de um RDF *Store* se fará necessário para a persistência do conhecimento.

Este capítulo, além das definições de dados, informação e conhecimento, também conceituou dados abertos e dados governamentais abertos, apresentou conceitos e ferramentas necessários para conceituar e estruturar os dados no ambiente da *Web* de dados.

### 3 MÉTODO CIENTÍFICO

Para que um conhecimento seja considerado científico, é necessário que este seja verificado, ou seja, que se prove a veracidade dos fatos. Essa verificação é feita através de procedimentos intelectuais e técnicos. O conjunto desses procedimentos é o método científico (GIL, 2008).

Este capítulo define o tipo de pesquisa de acordo com os pontos de vista abordados. São definidos os processos para atender os objetivos propostos inicialmente, bem como o cronograma para estes, que está junto ao apêndice deste trabalho. É apresentada também a proposta de solução para o problema abordado, bem como as delimitações deste trabalho.

#### 3.1 CARACTERIZAÇÃO DO TIPO DE PESQUISA

A pesquisa, de acordo com Gil (2008, p.42), é um “processo formal e sistemático de desenvolvimento do método científico. O objetivo fundamental da pesquisa é descobrir respostas para problemas mediante o emprego de procedimentos científicos”. Já, para Menezes e Silva (2005, p.20), “pesquisa é um conjunto de ações, propostas para encontrar a solução para um problema, que tem por base procedimentos racionais e sistemáticos”.

No cunho científico, a pesquisa é a busca de solução para um problema relevante a alguém. A ciência é produzida através da pesquisa, desse modo, pesquisa é o caminho para se chegar à ciência, ao conhecimento. (KAUARK, MANHÃES e MEDEIROS, 2010).

O uso de pesquisa é necessário quando não se possui informações suficientes para solucionar os problemas encontrados ou, quando se tem a informação desordenada, dificultando a sua análise (GIL, 2002).

Existem várias formas de classificar pesquisas de acordo com diferentes pontos de vista. A intenção de proporcionar maior familiaridade com o problema de modo a torná-lo explícito, de acordo com Gil (2012), faz com que este trabalho, no ponto de vista dos seus objetivos, faça uso da pesquisa em sua forma exploratória.

Do ponto de vista dos procedimentos técnicos, o presente trabalho utiliza a pesquisa bibliográfica, por basear-se em materiais já publicados (GIL, 2002).

Este trabalho objetiva gerar conhecimento a aplicações práticas e dirigir soluções de problemas específicos, além de envolver verdades e interesses locais, desse modo, de acordo

com Menezes e Silva (2005), partindo do ponto de vista da natureza, é utilizado a pesquisa em sua forma aplicada.

Considerando a forma de abordagem do problema, este trabalho utiliza a pesquisa em sua forma qualitativa, em que há uma relação dinâmica entre o mundo real e o sujeito, não requer métodos e técnicas estatísticas e o processo e seu significado são os focos principais de abordagem. (MENEZES; SILVA, 2005).

### 3.2 ATIVIDADES METODOLÓGICAS

Afim de contemplar os objetivos propostos neste trabalho, o processo de pesquisa e elaboração deste trabalho é organizado nas atividades dispostas a seguir:

1. Identificação do problema.
2. Definição da proposta de solução.
3. Definição do método de pesquisa.
4. Definição do domínio de dados abordado.
5. Revisão técnico científica.
6. Modelagem do domínio de conhecimento.
7. Triplificação dos dados do domínio.
8. Persistência dos dados triplificados.
9. Definição de experimentos com base no domínio definido.
10. Avaliação dos experimentos.
11. Apresentação dos resultados obtidos.
12. Conclusões acerca dos resultados obtidos.

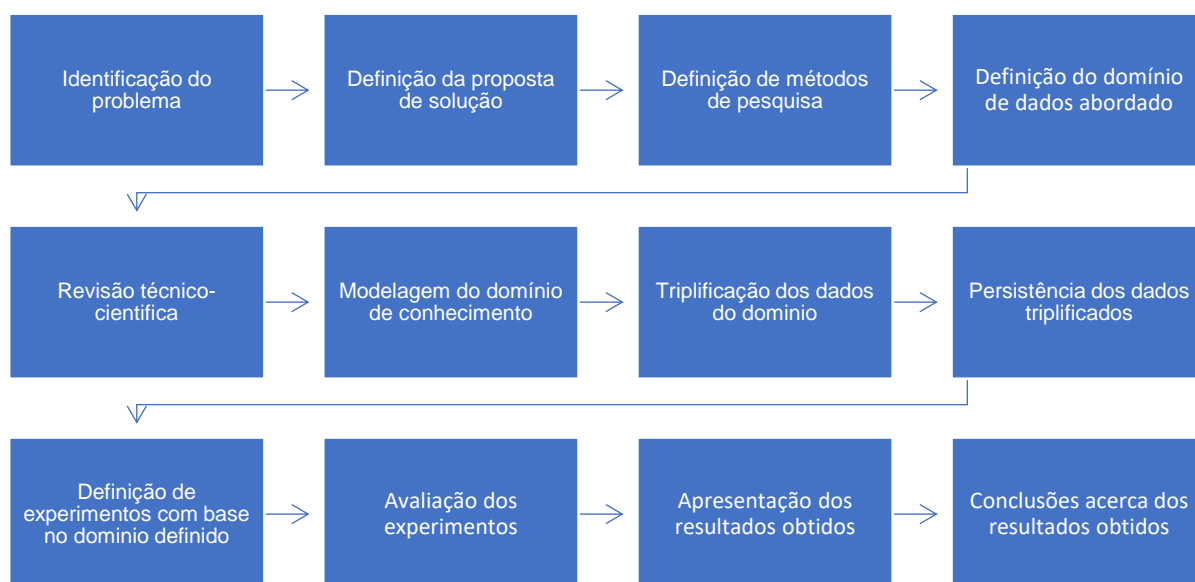
Identificado o problema, a proposta de solução e o método de pesquisa a ser aplicado em um domínio de dados definido, a seguir é apresentado a modelagem do projeto proposto, nesta, é contemplado o embasamento necessário para o desenvolvimento da proposta de solução.

A etapa de modelagem do domínio de conhecimento refere-se à elaboração do triplificador de dados, nesta são identificadas as classes, relacionamentos e verdades acerca do domínio de dados coletado.

A triplificação dos dados do domínio consiste em converter para triplas RDF, os dados semiestruturados encontrados em arquivos coletados, em que cada recurso deve ser alinhado ao seu conceito na ontologia, e identificado propriamente através de um URI. Após essa etapa, é possível realizar a persistência das triplas RDF geradas em um *RDF store*, deste modo tem-se uma base de conhecimento.

A definição e avaliação dos experimentos é responsável por analisar a resolução do problema e verificar as possibilidades alcançadas com a solução proposta. Por último, os resultados são documentados e conclusões acerca desse são apresentadas.

Figura 9 – Fluxograma das atividades do trabalho.



Fonte: Elaboração dos autores, 2016.

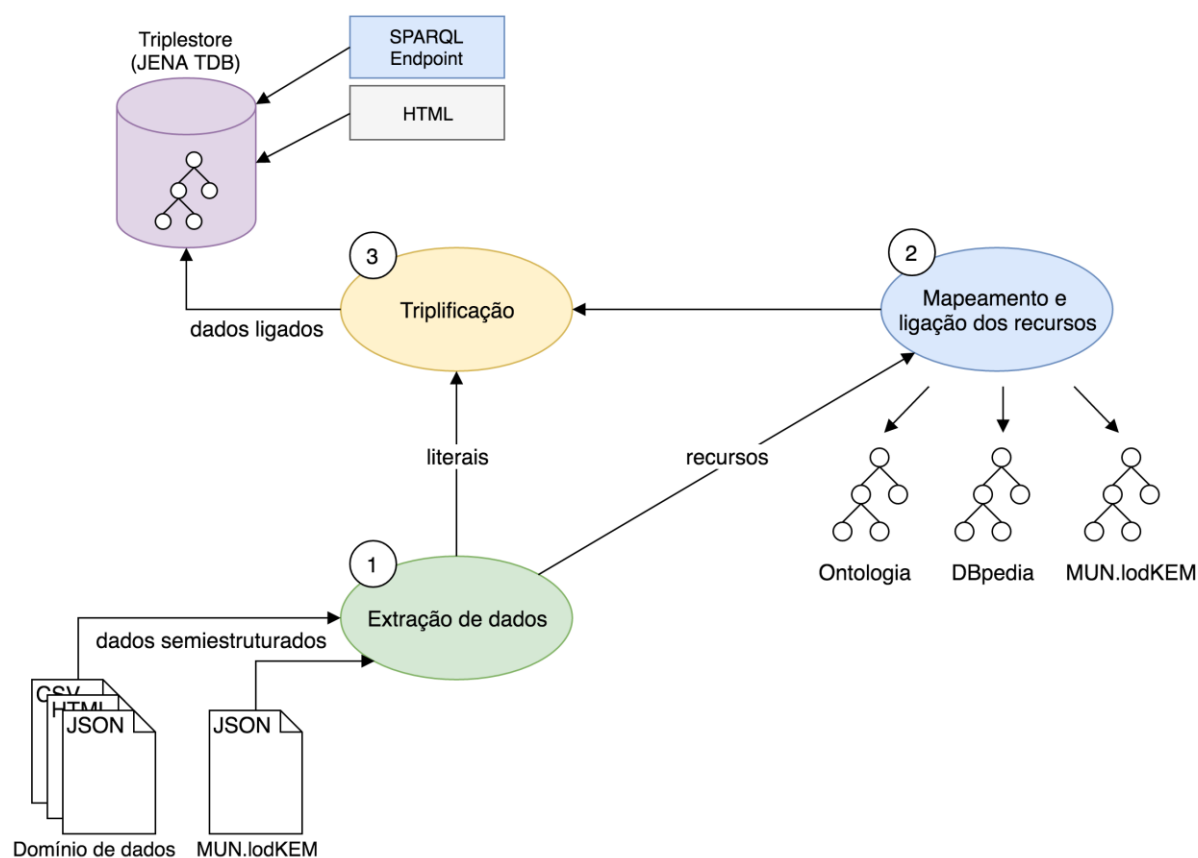
As atividades metodológicas, representadas no fluxograma da figura 9, delimitam e organizam o presente trabalho. O cronograma com os prazos para cada atividade é apresentado no apêndice A deste trabalho.

### 3.3 PROPOSTA DE SOLUÇÃO

Considerando os objetivos pretendidos, com base no domínio de dados definido e as tecnologias levantadas, é proposta a solução para o problema, apresentada na figura 10.



Figura 10 – Conversão dos dados semiestruturados em dados conectados.



Fonte: Elaboração dos autores, 2016.

Os dados semiestruturados, coletados em portais de dados abertos do governo brasileiro, e dados conectados, extraídos via JSON através do SPARQL *endpoint*, do projeto MUN.lodKem, passam pelo processo de extração no qual recursos e literais são identificados.

Recursos (por exemplo, Santa Catarina, Florianópolis, *parteDeUf*, *idhLongevidade*), ao contrário de literais (por exemplo, números e textos), devem passar pelo processo de mapeamento com a ontologia e ligação com *datasets* externos. Nesse processo, os recursos são relacionados com conceitos da ontologia desenvolvida e recebem identificadores adequados.

Na terceira e última etapa, os dados são convertidos para o formato RDF, em que são geradas as triplas derivadas dos relacionamentos entre recurso–recurso e recurso–literal. Nesse processo, também, alguns dos recursos mapeados a ontologia desenvolvida são

conectados para os respectivos recursos de *datasets* externos (por exemplo, DBpedia<sup>8</sup> e MUN.lodKEM<sup>9</sup>), que possuem um vocabulário semelhante ao da ontologia desenvolvida.

A persistência e a publicação dos dados podem então ser feitas. Desse modo, os dados podem ser divulgados através de um SPARQL *endpoint* e interface HTML, que permite consultas integradas na base de conhecimento e nas ontologias externas, e estão prontos para a realização dos experimentos propostos.

### 3.4 DELIMITAÇÕES

O presente trabalho tem como proposta a conversão de dados abertos semiestruturados em dados conectados e a persistência desses em uma base de conhecimento, liberando essa para consulta direta via SPARQL e interface HTML, como uma alternativa para disponibilização de dados abertos na *Web*.

Apesar de propor uma solução específica para a publicação de dados abertos na *Web*, este trabalho não aborda a disponibilização em si desses, limitando-se apenas na criação e avaliação da solução com base em ferramentas já existentes. A divulgação de dados abertos conectados (*linked open data*) na *Web* envolve uma série de requisitos como: uso de uma licença aberta, infraestrutura adequada, confiança, qualidade e relevância dos dados, etc. A solução proposta é executada em servidores locais e utiliza dados coletados em um domínio específico a fim de exemplo, portanto, este trabalho não aborda os meios para publicação de *linked open data*.

Por se tratar de uma pesquisa exploratória e qualitativa, a avaliação da solução tem como único propósito a identificação das possibilidades alcançadas pelo modo de publicação proposto, deste modo não é avaliado nenhum indicador estatístico.

---

<sup>8</sup>Disponível em: <http://wiki.dbpedia.org/>

<sup>9</sup>Disponível em: <http://lodkem.led.ufsc.br/municipios/>

## 4 PROPOSTA DE SOLUÇÃO

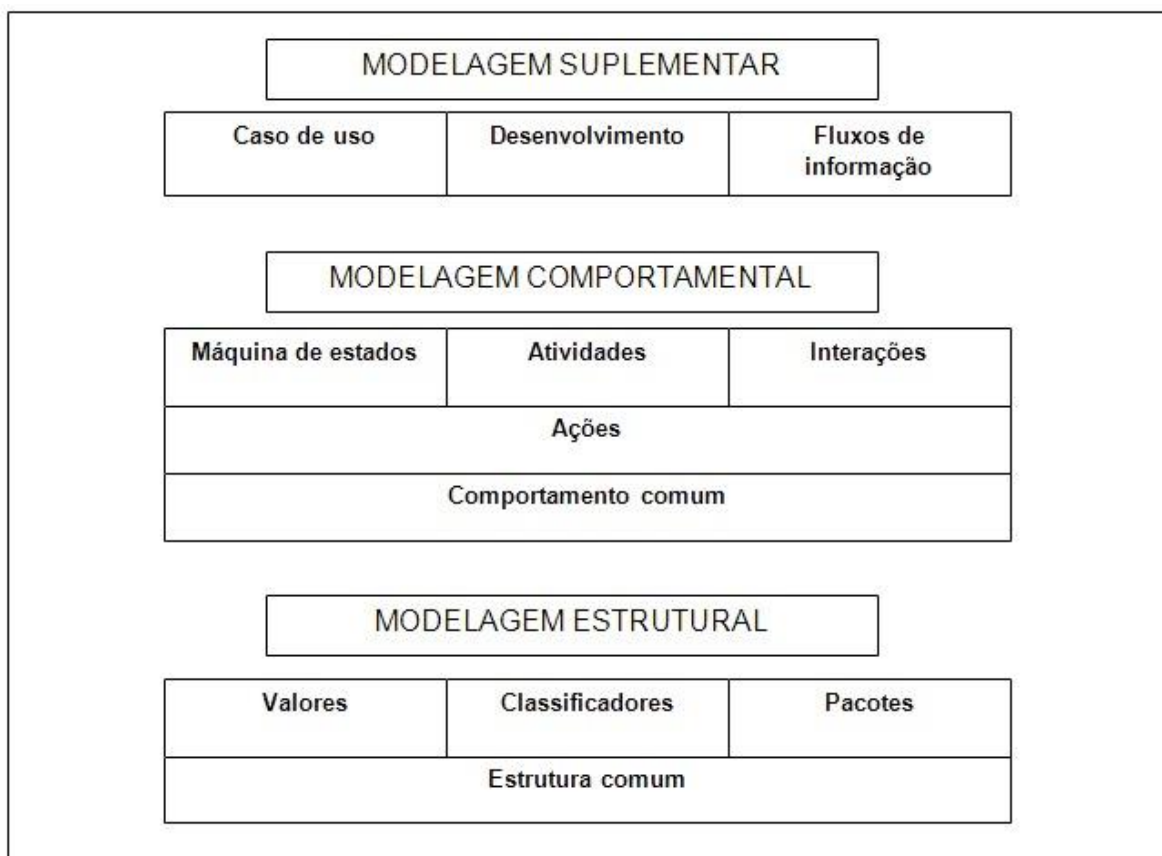
Neste capítulo apresentar-se-á a modelagem do triplicador dos dados da ontologia de domínio. Primeiramente será abordado a metodologia aplicada, para então exibir a modelagem em questão.

### 4.1 LINGUAGEM DE MODELAGEM UNIFICADA (UML)

A UML é uma linguagem gráfica criada para especificação, documentação e concepção de *softwares*, criando um padrão para a produção, realização e engenharia de projetos de sistemas. (BOOCH, RUMBAUGH, JACOBSON, 2006, p. 13) .

Na figura 11 é evidenciado todas as áreas de conhecimento que a linguagem contempla.

Figura 11 – Áreas semânticas da UML



Fonte: OMG (Adaptada pelo autor), 2016.

Segundo Lima (2011, p. 30) é possível tirar a seguinte conclusão sobre a linguagem de modelagem unificada:

Além de flexível, a UML é extensível e independente de processos ou linguagens de programação, o que garante a liberdade para o desenvolvedor adotar qualquer processo, metodologia ou linguagem de programação sem deixar de expressar-se claramente para os seus usuários e outros desenvolvedores, pois utiliza uma notação padrão, comum a todos os ambientes e empresas. Por ser um conjunto de melhores práticas, ela tem suporte de alcance mundial.

A UML, como é possível ver na imagem anterior, é uma linguagem bastante completa e com diversos recursos que abrangem todas as áreas, assim sendo cada item exposto na imagem pode possuir um ou mais diagramas, dessa forma a linguagem permite que seja desenvolvido diversos conceitos e artefatos do sistema de uma forma intuitiva e objetiva. (LIMA, 2011, p.30).

Na seção a seguir será abordada as etapas de modelagem do triplificador deste trabalho.

## 4.2 MODELAGEM DO TRIPLIFICADOR

Nesta seção é apresentado a modelagem da solução proposta. A modelagem do triplificador de dados para este trabalho está relacionada ao domínio de dados que será persistido no banco, dessa forma a maneira como os dados estão segmentados impactam diretamente na forma como o triplificador será modelado, o cenário dos dados e as suas representações serão abordadas em seguida no capítulo 5, seção 5.2.

Na modelagem do triplificador de domínio serão utilizados os seguintes recursos da UML: Requisitos funcionais, requisitos não funcionais, diagrama de atividades, diagrama de comunicação, diagrama de sequência, diagrama de classe e por fim o diagrama de implantação, todos serão conceituados em seus respectivos tópicos.

Para a modelagem do triplificador de dados deste trabalho foi utilizado a versão 2.3 da linguagem padrão unificada (UML).

### 4.2.1 Elicitação de Requisitos

A etapa de elicitação de requisitos é tão importante quanto a aplicação das técnicas da UML, pois é na fase de análise dos requisitos que a ideia do sistema e de seus componentes são maturados, é importante abordar todos os processos, recursos e artefatos que estarão englobados nos requisitos. (FOWLER, SCOTT, 2000, p.32-33).

Segundo Engholm (2010, p.70) o estágio de elicitação de requisitos requer alguns passos, são eles:

- Elicitar requisitos em alto nível, com o cliente, ou dono do negócio;
- Elicitar requisitos funcionais e não funcionais;
- Identificar riscos e restrições;
- Refinar requisitos;
- Aprovar requisitos;
- Criar matriz de responsabilidade de requisitos;
- Criar diagramas.

As etapas anteriores são sugestões que devem ser avaliadas e modeladas para suprir as expectativas e necessidade de cada projeto atendendo suas individualidades. À visto disso a elicitação de requisitos, levantamento ou análise dos requisitos é a fase inicial de qualquer projeto de software, portanto os requisitos funcionais e não funcionais definem as necessidades do software. (ENGHOLM, 2010, p.69).

#### 4.2.1.1 Requisitos Funcionais

Os requisitos funcionais são responsáveis por expressar as funcionalidades do projeto de solução, segundo Pádua (2000 p.186) pode-se ter o seguinte entendimento sobre os requisitos funcionais:

Os requisitos funcionais definem as ações fundamentais através das quais o produto aceita e processa as entradas especificadas, gerando as respectivas saídas. Nesta seção é feito o detalhamento destes requisitos, a nível suficiente para o desenho do produto, de seus testes de aceitação e de seu manual de usuário.

Para este trabalho, no quadro 6, foram levantados os seguintes requisitos funcionais, tendo em vista o domínio de aplicação do cenário de avaliação:

Quadro 6 – Requisitos Funcionais

<b>REQUISITOS FUNCIONAIS</b>
RFN001. Sistema deverá importar informações de cidades
RFN002. Sistema deverá importar informações de estados
RFN003. Sistema deverá importar informações de IDH de cidades
RFN004. Sistema deverá importar informações de IDH de estados
RFN005. Sistema deverá importar informações de Receitas Geradas Por Estados
RFN006. Sistema deverá importar informações de Recursos Recebidos Por Cidades
RFN007. Sistema deverá importar informações de Recursos Recebidos Por Estados
RFN008. Sistema deve interligar os dados importados
RFN009. Sistema deve persistir os dados interligados em um banco de dados triplestore

Fonte: Elaboração dos autores, 2016.

Com a apresentação dos requisitos funcionais do triplicador a seguir será apresentado os requisitos não funcionais da proposta de solução deste trabalho.

#### 4.2.1.2 Requisitos Não Funcionais

Os requisitos não funcionais ao contrário dos requisitos funcionais não apresentam determinada funcionalidade do software, mas deve trazer comportamentos e restrições que o sistema precisa atender, estes podem ser atributos de desempenho, segurança, qualidade entre outros. (PRESSMAN, MAXIM. 2016. p.141). A seguir, no quadro 7, são apresentados os requisitos não funcionais do sistema:

Quadro 7 – Requisitos Não Funcionais

<b>REQUISITOS NÃO FUNCIONAIS</b>
<b>RNF001. O sistema deverá ser cliente/servidor</b> - O sistema deverá funcionar em um computador local e poderá persistir os dados em um servidor.

<b>RNF002. O sistema deverá ser escrito usando orientação a objetos</b> - O sistema será desenvolvido em linguagem e estrutura voltada para orientação a objetos
<b>RNF003. A linguagem do sistema deverá ser JAVA</b> - O sistema será implementado na linguagem de programação JAVA
<b>RNF004. O banco de dados Triplestore deverá ser o TDB</b> - O banco em que será persistido os dados triplificados pelo sistema desenvolvido será o TDB
<b>RNF005. O servidor SPARQL utilizado deverá ser o FUSEKI</b> – O <i>endpoint</i> para consultas SPARQL será providenciado com o uso do FUSEKI.
<b>RNF006. A frontend de dados conectados deverá ser o Pubby</b> - A interface gráfica para os recursos obtidos via servidor SPARQL será disponibilizada usando o Pubby.

Fonte: Elaboração dos autores, 2016.

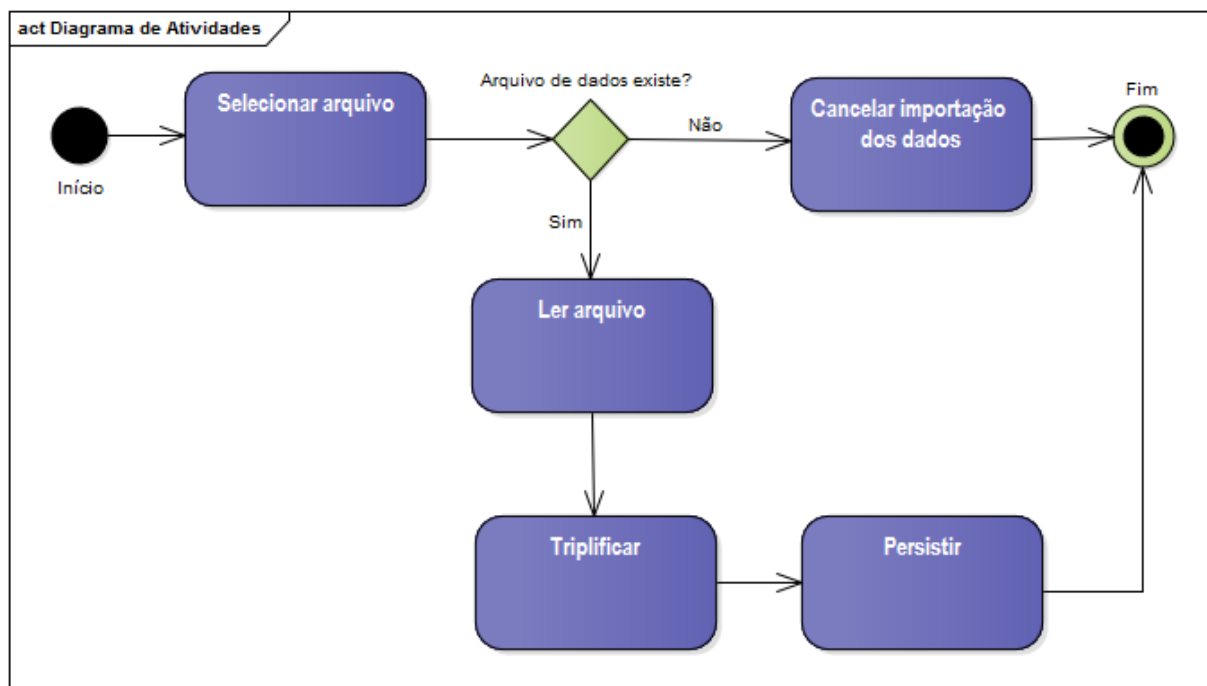
Com os requisitos funcionais e não funcionais levantados é possível ter o conhecimento necessário do que se pretende realizar no sistema. Na seção que se segue são apresentados o conceito e o modelo do digrama de atividades do projeto.

#### 4.2.2 Diagrama de Atividades

O diagrama de atividades é o modelo responsável por descrever a sequência que as atividades devem ser realizadas no sistema proposto, esse padrão de modelagem suporta a descrição de atividades quando estas ocorrem em tempos divergentes, ou também podendo ocorrer em paralelo, em outras palavras o diagrama de atividades determina as regras que devem ser seguidas e suas respectivas ordens de execução. (FOWLER, SCOTT. 2000. p 121-123).

Este diagrama é muito utilizado quando pretende-se descrever um comportamento e como ocorrem suas interações, este aproxima-se da lógica que o algoritmo deve executar, exibindo suas lógicas de sequência de execução, repetição e decisão. (REZENDE, 2005, p.211). A seguir, na figura 12 é apresentado o diagrama de atividades do triplificador do projeto de solução proposto:

Figura 12 – Diagrama de atividades do triplicador de dados



Fonte: Elaboração dos autores, 2016.

A figura 12 demonstra o modelo do diagrama de atividades do triplicador de dados do projeto de solução. Com o devido entendimento de como ocorrem as atividades do sistema proposto é apresentado a seguir o diagrama de comunicação.

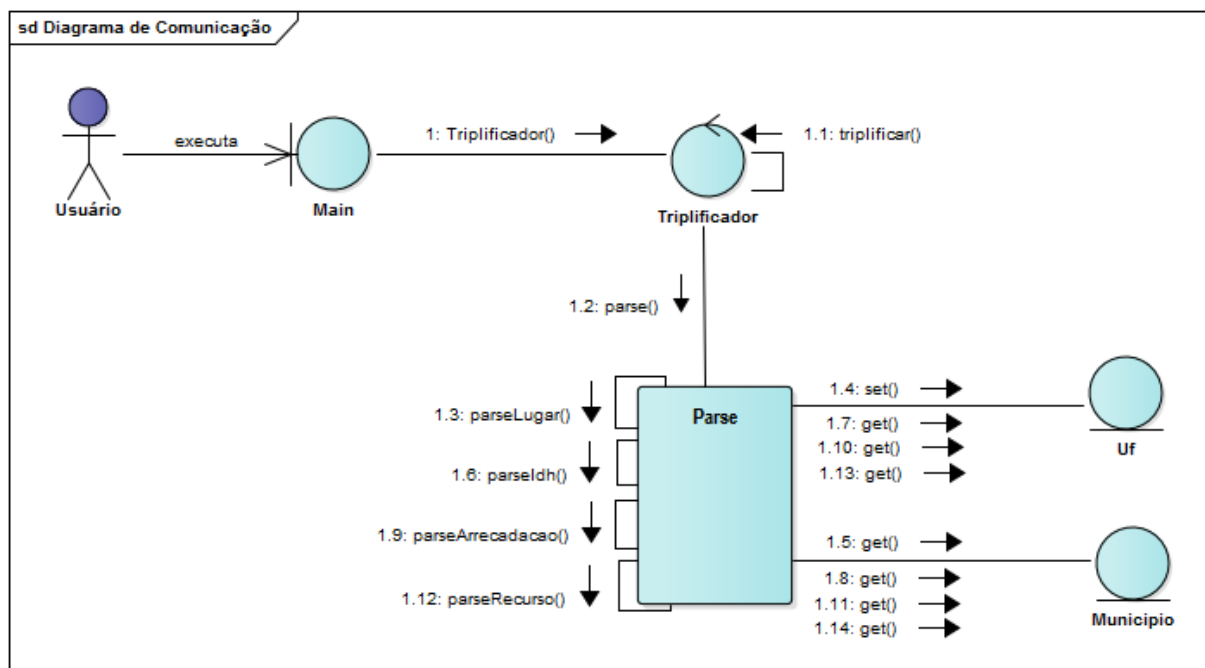
#### 4.2.3 Diagrama de Comunicação

O diagrama de comunicação é equivalente à um diagrama de sequência simplificado, dessa forma a disposição dos objetos não é linear e sim estrutural, normalmente é utilizado quando é preciso exemplificar a interação dos objetos, dessa forma ele não traz a sequência das interações, somente a comunicação entre eles e de que forma ocorrem. (LIMA. 2011. p 280-283).

Na figura 13 é apresentado o diagrama de comunicação do triplicador, é possível verificar como os objetos do sistema se comunicam.



Figura 13 – Diagrama de comunicação do triplificador de dados.



Fonte: Elaboração dos autores, 2016.

O diagrama de comunicação do projeto de solução anterior, expõe a maneira em que os objetos se comunicam compartilham e executam seus processos.

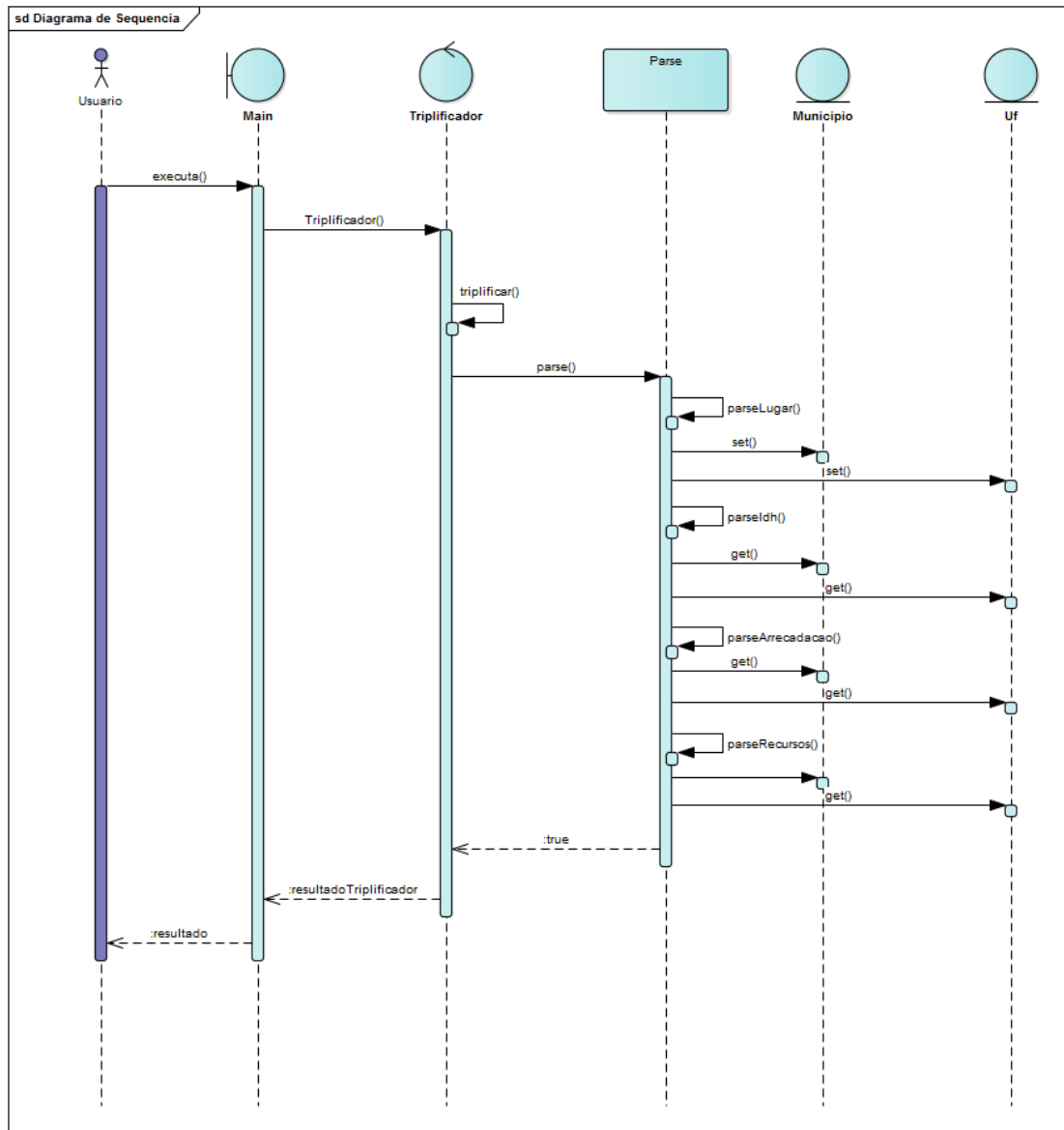
Em continuidade a modelagem do sistema proposto a seguir é apresentado o diagrama de sequência do projeto de solução.

#### 4.2.4 Diagrama de Sequência

O diagrama de sequência enfatiza a ordenação temporal das mensagens trocadas entre os objetos do sistema. Segundo a organização Object Management Group OMG (2015. p 593) o diagrama de sequência é o tipo mais comum encontrado nos diagramas de interação, este modelo é focado na sequência de trocas de mensagens em sua linha do tempo.

Um diagrama de sequência ilustra a interação dos objetos baseado na linha do tempo em que as mensagens são trocadas, ou seja, como as mensagens são enviadas e recebidas entre os objetos do sistema. (REZENDE. 2005. p 208). A figura 14 a seguir demonstra o diagrama de sequência do triplificador do projeto:

Figura 14 – Diagrama de Sequência do triplicador de dados.



Fonte: Elaboração dos autores, 2016.

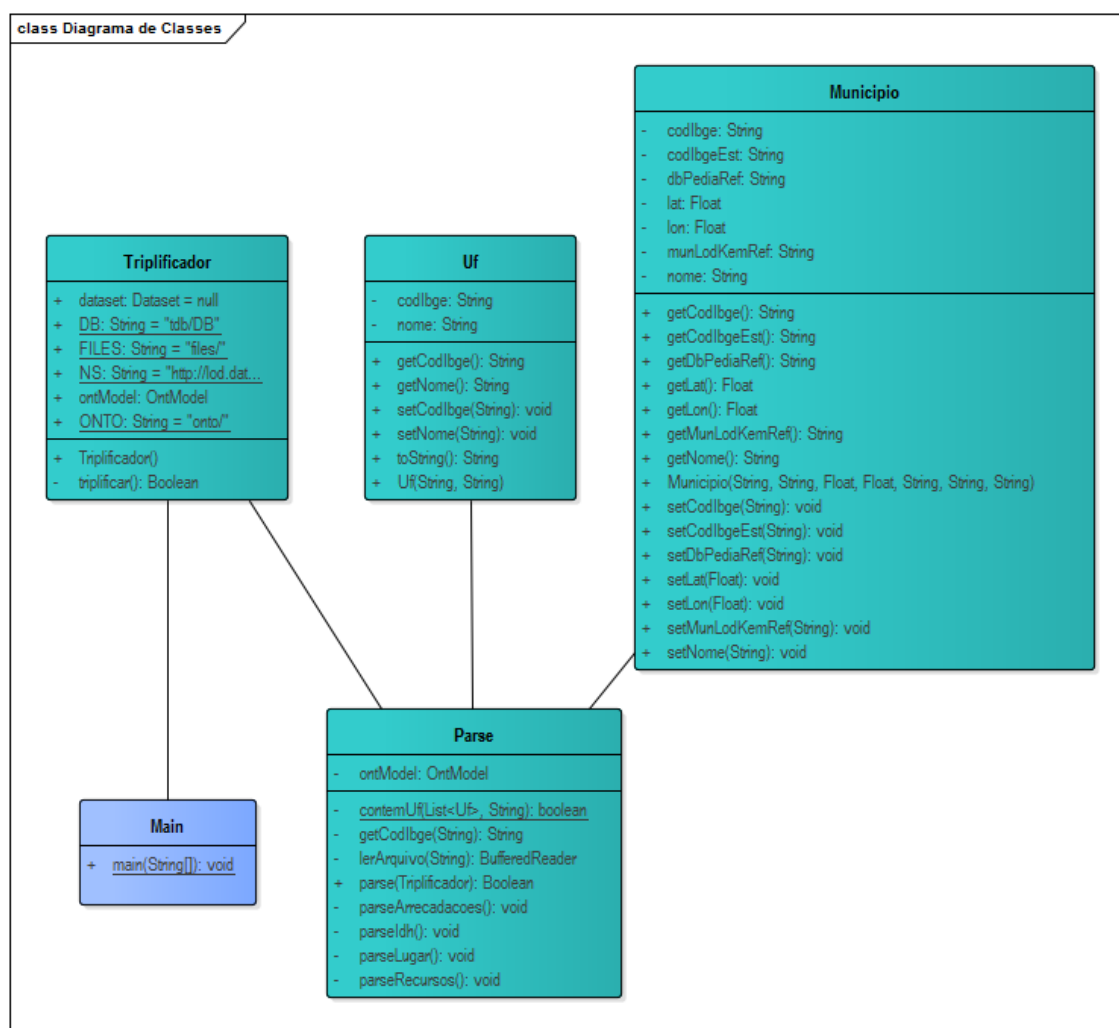
Na figura 14 é descrito como ocorre a interação entre os objetos do sistema e a sequência em que seus processos ocorrem.

Com a compreensão de todos os diagramas realizados até então, é possível conceber o diagrama de classes do sistema, este será apresentado a seguir.

#### 4.2.5 Diagrama de Classe

O diagrama de classe do projeto demonstra as especificações das classes e interfaces do sistema. O diagrama de classe trabalha basicamente com classes, associações, atributos, interfaces, métodos e dependências. (REZENDE, 2005, p.209). Na figura 15 é apresentado o diagrama de classe do triplificador, nele é possível identificar as classes da aplicação.

Figura 15 – Diagrama de Classe do triplificador de dados.



Fonte: Elaboração dos autores, 2016.

Com o entendimento dos modelos de implementação expostos acima é possível dar andamento no diagrama de implantação do sistema realizado na seção seguinte.

#### 4.2.6 Diagrama de Implantação

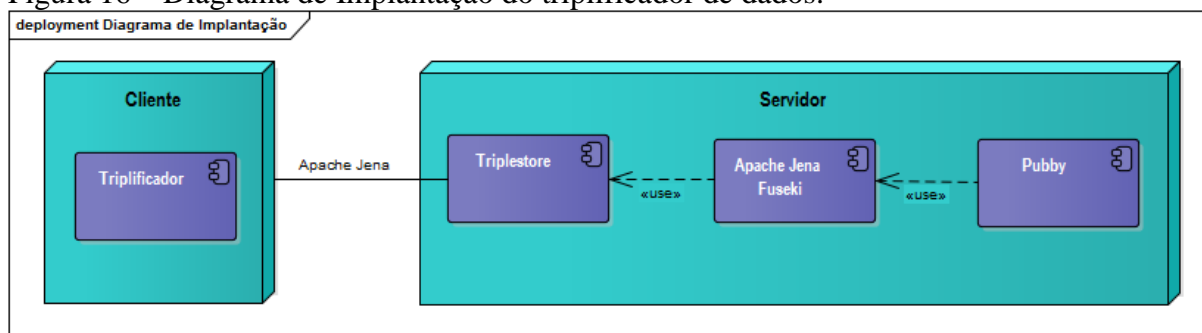
O diagrama de implantação expõe a estrutura do sistema interligando a parte física dos componentes de software com as configurações de hardware. (REZENDE. 2005. p 212).

Para Rezende (2005. p. 212) pode-se tirar a seguinte conclusão sobre o diagrama de implantação:

Trata-se de um gráfico de nós conectados por associações de comunicação, que podem conter instâncias de componentes, classes, bibliotecas ou executáveis, inclusive os desenhos dos sistemas de telecomunicações e os sistemas operacionais.

Na figura 16, consta o diagrama de implantação da aplicação deste trabalho.

Figura 16 – Diagrama de Implantação do triplificador de dados.



Fonte: Elaboração dos autores, 2016.

Com a apresentação do diagrama de implantação encerra-se o Capítulo 4. Foi possível visualizar a modelagem do sistema de solução proposto, todos os recursos que serão utilizados para iniciar o experimento, a análise e desenvolvimento da maneira que será realizada a solução foram obtidos. A etapa do desenvolvimento é descrita no capítulo seguinte.

## 5 DESENVOLVIMENTO

No presente capítulo são apresentadas as ferramentas e tecnologias utilizadas, para o desenvolvimento da proposta de solução, o histórico com a progressão do experimento e o cenário, além de expor as consultas ao domínio, a avaliação e a análise dos resultados obtidos.

### 5.1 FERRAMENTAS UTILIZADAS

Nesta seção são apresentadas as ferramentas e tecnologias utilizadas para o desenvolvimento da proposta de solução. A figura 17 a seguir exhibe as ferramentas utilizadas.

Figura 17 – Ferramentas utilizadas.



Fonte: Elaboração dos autores, 2016.

As ferramentas da imagem anterior foram utilizadas para modelagem, desenvolvimento e a análise da proposta de solução deste trabalho. A seguir será descrito cada ferramenta, suas principais funcionalidades e qual foi a motivação para a utilização das mesmas.

### 5.1.1 Enterprise Architect

A ferramenta *Enterprise Architect* é utilizada para realizar a modelagem de sistemas e processos, possui diversos recursos e linguagens disponíveis para utilização. Segundo descrição no site da ferramenta, *Enterprise Architect* é uma plataforma visual para elaboração e construção de *softwares*, utilizado para modelagem de processos de negócios e os mais diversos projetos de modelagem. (SPARX SYSTEMS, 2016).

O motivo principal para a escolha dessa ferramenta, foi a sua facilidade de uso, conhecimento a respeito da ferramenta, a sua disponibilidade de uso e também por ser uma das opções mais completas disponíveis.

### 5.1.2 Java

Java é uma linguagem de programação orientada à objetos e também uma plataforma computacional lançada pela primeira vez pela Sun Microsystems em 1995. "O Java é a base para praticamente todos os tipos de aplicações na rede, sendo o padrão global para o desenvolvimento e distribuição de aplicações móveis e incorporadas, jogos, conteúdo baseado na *web* e *softwares* corporativos." (ORACLE, 2016).

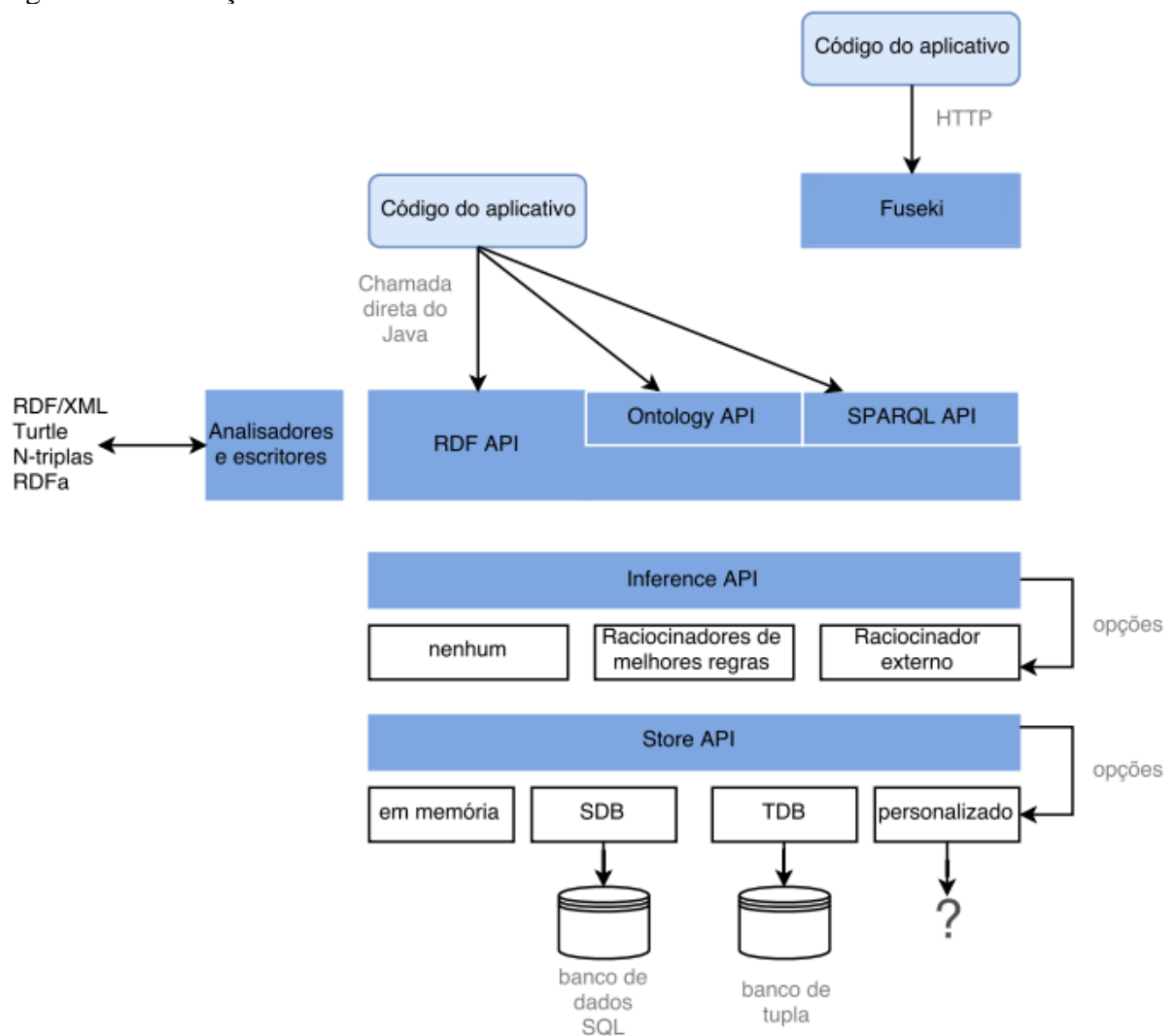
Seus números de utilização vão desde 97% dos Desktops Corporativos, 9 milhões de desenvolvedores no mundo utilizando Java em todo o mundo (ORACLE, 2016), a escolha desta linguagem de programação para este trabalho deu-se por esses motivos citados anteriormente, como também devido a familiaridade dos autores com esta linguagem.

### 5.1.3 Jena

A ferramenta Apache Jena, ou somente Jena é um *framework* aberto desenvolvido em Java para a construção de web semântica e aplicações de dados conectados. Toda a estrutura do Jena é composta por diferentes APIs que interagem em conjunto para o processamento dos dados RDF. (THE APACHE SOFTWARE FOUNDATION, 2016a).

A figura 18 a seguir mostra como é a interação entre as APIs do Jena:

Figura 18 – Interação entre as APIs do Jena.



Fonte: The Apache Software Foundation (2016a, tradução nossa).

A ferramenta Jena foi escolhida para o desenvolvimento deste trabalho devido a sua gama de APIs para tratamento de dados na área semântica. Na sequência serão abordados as APIs do Jena que serão utilizados para o desenvolvimento deste trabalho.

### 5.1.3.1 TDB

O TDB é um componente do Jena que é o responsável pelo armazenamento e consultas em triplas RDF, possui suporte para todas as APIs do Jena. Uma biblioteca TDB pode ser acessada e gerenciada por linhas de comando, ou também pela API do Jena. Quando

acessada usando transações o TDB é protegido contra processos inesperados, falhas de sistemas e etc. (THE APACHE SOFTWARE FOUNDATION, 2016e).

#### 5.1.3.2 Fuseki

O Apache Jena Fuseki é um servidor SPARQL, que pode ser executado de diversas maneiras, como um serviço do sistema operacional, ou como uma aplicação web, ou também como um servidor independente. Esse componente é utilizado para o compartilhamento de dados TDB entre múltiplas aplicações, como ele fornece um servidor SPARQL pode usar o TDB para o armazenamento e persistência dos dados. Também pode ser utilizado para fornecer o motor de busca para outros sistemas de consultas e de armazenamento em RDF. (THE APACHE SOFTWARE FOUNDATION, 2016b).

#### 5.1.3.3 *Ontology* API

A API *Ontology* do Jena tem como objetivo fornecer uma interface de programação consistente para o desenvolvimento de aplicações de ontologia, independentemente de qual é a linguagem de ontologia que está sendo utilizado no sistema. (THE APACHE SOFTWARE FOUNDATION, 2016c).

#### 5.1.3.4 *Inference* API

A biblioteca *Inference* API é uma biblioteca que foi projetado para realizar inferências com raciocinadores e motores de regras dentro do Jena. A biblioteca de inferência do Jena foi desenvolvida para permitir uma gama de motores de inferência ou raciocinadores sendo conectados ao Jena. Tais motores são usados para derivar afirmações RDF adicionais que são decorrentes de alguma base de RDF bem como qualquer informação opcional da ontologia, axiomas e regras associadas ao raciocinador. A principal utilização deste mecanismo é apoiar o uso de linguagens como RDFS e OWL que permitem fatos adicionais possam ser inferidas a partir de dados de instância e descrições de classe. (THE APACHE SOFTWARE FOUNDATION, 2016d).



#### 5.1.4 Pubby

O Pubby é usado para adicionar interfaces de dados conectados a SPARQL *endpoints*. (CYGANIAK, BIZER, 2011). Hoje boa parte da Web Semântica vive dentro de bancos de dados de triplas, e só pode ser acessada através de um SPARQL Endpoint. O Pubby providencia uma interface para *datasets* de dados conectados locais e remotos, através de uma interface HTML simples que mostra os dados disponíveis em cada recurso.

#### 5.1.5 Apache Tomcat®

O Tomcat é um servidor web Java, mais especificamente, um container de *servlets*. O Apache Tomcat® é utilizado por softwares em grande escala, aplicações web de missão crítica através de uma variada gama de indústrias e organizações. (THE APACHE SOFTWARE FOUNDATION, 2016f).

#### 5.1.6 Protégé

O sistema Protégé é um *software* livre e distribuído, foi desenvolvido pelo centro de biomedicina da universidade de Stanford (BMIR - *Biomedical Informatics Research*) com o intuito de avançar o estudo em tecnologias semânticas, bioestatísticas e a modelagem de sistema biomédicos para beneficiar a pesquisa clínica, como também ao atendimento ao cliente na área da saúde.

Foi criado para realizar a manutenção e criação de ontologias, com suporte total para o OWL 2, entre várias utilidades do Protégé, estão o suporte para a criação e edição de uma ou mais ontologias em uma única área de trabalho, possui uma interface de usuário totalmente personalizável, com ferramentas de visualização que permitem a navegação interativa entre as relações da ontologia. (PROTÉGÉ, 2016).

#### 5.1.7 Git

Git é um sistema livre de código aberto distribuído para controle de versão de projetos de software, tornando o controle das alterações do projeto de software muito mais fácil para os desenvolvedores, além de trazer o histórico das modificações. (GIT, 2016).

### 5.1.8 GitHub

O *GitHub* é um sistema de compartilhamento de código, utilizada por desenvolvedores de *software*, é uma comunidade com mais de 15 milhões de pessoas utilizando a ferramenta para aprender, compartilhar e trabalhar junto para construção de sistemas, possui mais de 38 milhões de projetos de *software*. (GITHUB, 2016).

Devido a facilidade para compartilhar o projeto de solução deste trabalho, a ferramenta foi utilizada para armazenar o desenvolvimento de uma forma segura, como também para o compartilhamento das atualizações entre os participantes em tempo real.

### 5.1.9 Eclipse

Eclipse é uma ferramenta IDE (*Integrated Development Environment*) que possui diversos pacotes com variadas linguagens para desenvolvimento de *software*, a sua principal utilização é na linguagem JAVA, porém a plataforma também contempla as linguagens C / C++ IDE e PHP IDE entre outros. Dentre todas as vantagens oferecidas nessa ferramenta a possibilidade de combinar diferentes linguagens nessa IDE torna o seu uso muito mais eficaz. (ECLIPSE, 2016).

## 5.2 CENÁRIO

O cenário deste trabalho contempla dados geopolíticos temporais. Os dados são representados por três entidades principais: município, unidade federativa, e país. Os dados de nomes de municípios e estados, e de coordenadas geográficas de municípios foram obtidos via consulta SPARQL no *endpoint*, do projeto MUN.lodKEM<sup>10</sup>.

Os municípios, unidades federativas e país possuem dados temporais como: recebimento de recursos e arrecadação de impostos no ano de 2010, IDH geral, IDH de longevidade, IDH de renda e IDH de educação nos anos de 1991, 2000 e 2010. Estes dados foram extraídos de arquivos retirados de esferas governamentais que disponibilizam os dados

---

<sup>10</sup> Disponível em: <http://lodkem.egc.ufsc.br/>

de forma aberta: o Portal Brasileiro de Dados Abertos<sup>11</sup>, Portal da Transparência<sup>12</sup> e Receita Federal.

### 5.2.1 Ontologia de domínio

Uma ontologia de domínio consegue definir e caracterizar o domínio no qual as tarefas ocorrem, ou seja, esse tipo de ontologia representa o conhecimento sobre um determinado tópico, permitindo a criação de sistemas mais flexíveis e com inteligência, capacitando a sua aplicação prática em diversos domínios. (ISOTANI; BITTENCOURT, 2015, p.102).

Existem diversas metodologias para criação de ontologias, metodologias inspiradas na inteligência artificial, metodologias inspiradas na interação humano-computador e por fim metodologias inspiradas na engenharia de software, para o desenvolvimento da ontologia deste trabalho será utilizado a última metodologia citada, por sua vez esta metodologia foi inspirada em modelos tradicionais da engenharia de software, e também em metodologias ágeis. (ISOTANI; BITTENCOURT, 2015, p.133).

Existem diversas metodologias inspiradas na engenharia de software as principais são: *METHODOLOGY*, *Ontology Development 101* e *RapidOWL*. Para a construção da ontologia da proposta de solução deste trabalho será utilizada a metodologia *101*.

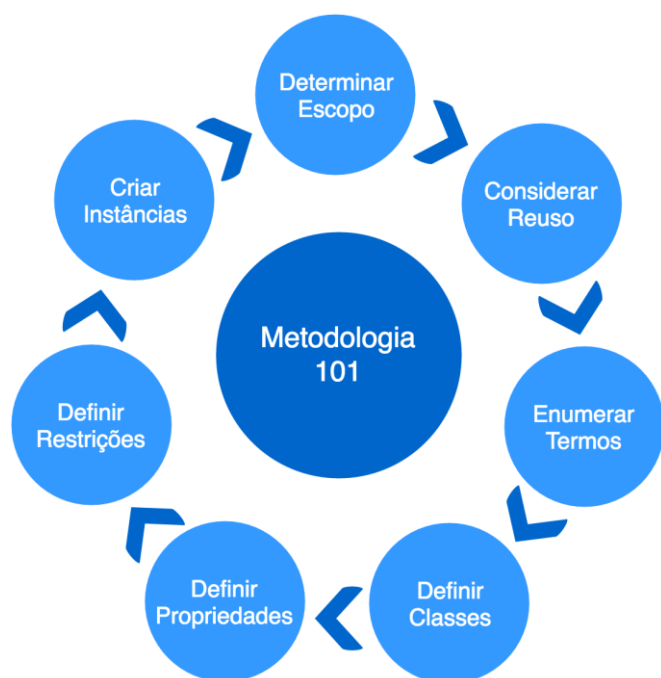
A metodologia 101 foi criada por Noy e McGuinness (2001), pesquisadoras da universidade de Stanford, e segundo guia disponibilizado a metodologia é dividida em sete passos que estão descritos na figura 19:

---

<sup>11</sup> Disponível em: <http://dados.gov.br/>

<sup>12</sup> Disponível em: <http://portaldatransparencia.gov.br/>

Figura 19 – Passos da metodologia 101.



Fonte: (Noy et al apud ISOTANI; BITTENCOURT, 2015, p.136).

Determinar o escopo é a etapa inicial para o desenvolvimento da ontologia, é preciso compreender quais serão os propósitos e os cenários envolvidos. Algumas perguntas são fundamentais nessa etapa, qual será o objetivo da ontologia? Para qual finalidade será seu uso? Quem irá utilizar? Definir e responder as questões acima amparam a realização e a definição do escopo. (NOY; MCGUINNESS, 2001).

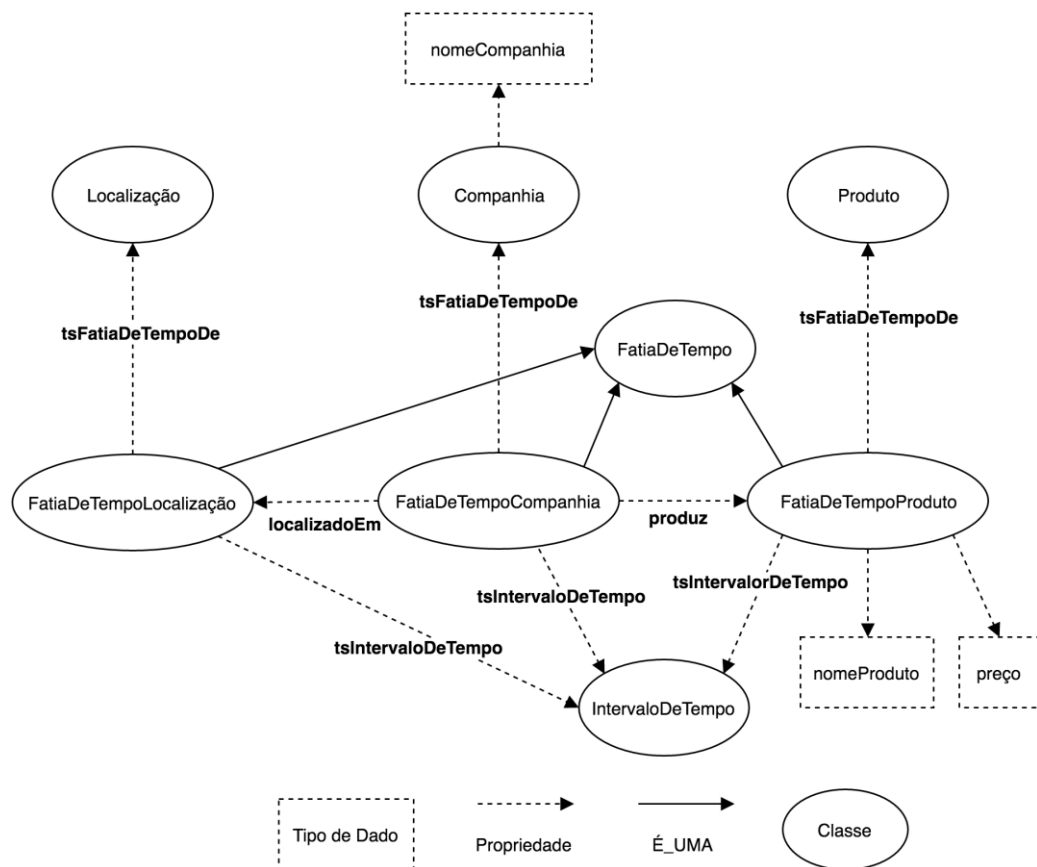
Considerar o reuso de ontologias já existentes se essas complementarem e forem adequadas ao desenvolvimento da ontologia proposta. Enumerar os termos importante dentro da ontologia ajudam a elucidar os conceitos envolvidos da ontologia, os passos acima descritos precisam ser realizados para que a ontologia possa ser corretamente planejada. (NOY; MCGUINNESS, 2001).

Após o planejamento da ontologia, inicia-se a criação das classes, hierarquias, propriedades, restrições das propriedades, que normalmente é uma etapa extremamente significativa, devido ao grau de granularidade fino, em que todas as restrições e exceções deverão ser observadas para que o conceito da ontologia não seja perdido, por fim é realizada a criação das instâncias. (ISOTANI; BITTENCOURT, 2015, p.139).

As classes principais da ontologia do projeto de solução, que fornecem a base ao domínio de dados são: município, unidade federativa e país. Estas classes receberão informações, porém como estas representam coisas reais passíveis de mudança com o tempo, as informações serão temporais.

Para resolver o fator tempo em uma ontologia de domínio, foi tomado como base, e adaptado para o domínio de dados do projeto de solução, um modelo de ontologia proposto por Batsakis e Petrakis (2010), que utiliza classes do tipo *FatiaDeTempo* e *IntervaloDeTempo* com propriedades *fatiaDeTempoDe* e *tempoEm* como proposta para representar partes temporais. A adaptação feita conta com a classe *FatiaDeTempo* e a propriedade *fatiaDeTempoDe*, entretanto se utiliza da classe *Data* e da propriedade *aconteceEm* em substituição a classe *IntervaloDeTempo* e da propriedade *tempoEm*, na figura 20 se tem a ontologia um exemplo de uma ontologia usando as classes e propriedades propostas por Batsakis e Petrakis (2010).

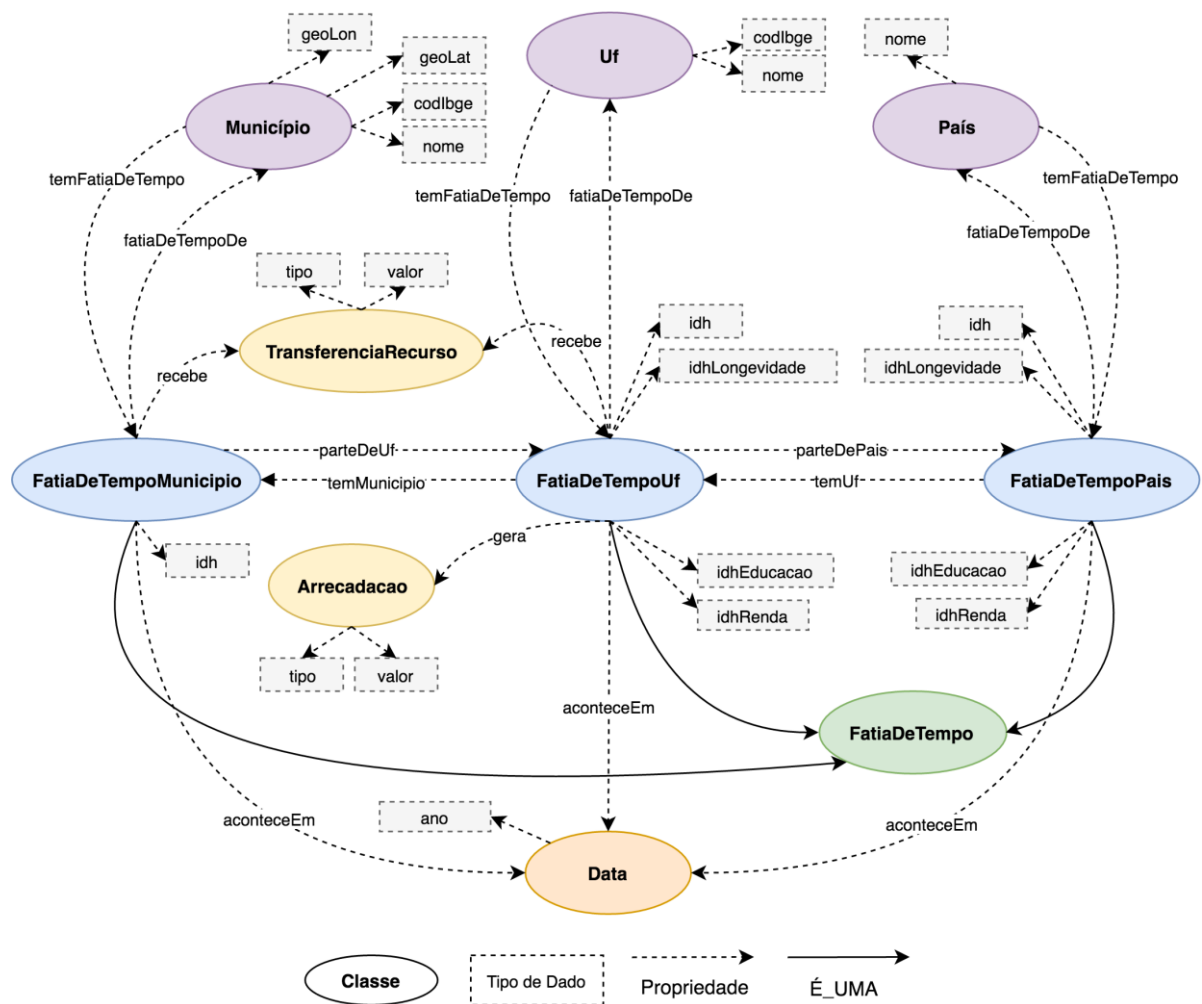
Figura 20 – Ontologia empresarial dinâmica.



Fonte: Batsakis e Petrakis (2010, p.3, tradução nossa).

O modelo proposto por Batsakis e Petrakis (2010) que trabalha com intervalos de tempo para cada evento, é mais aconselhado em caso de grandes volumes de dados, pois evita a criação desnecessária de triplas que reduziriam o desempenho em consultas. A substituição do intervalo de tempo pela data se deve ao fato da restrição dos dados de tempo no conjunto de dados coletados, representados geralmente por ano, pelo tamanho reduzido do conjunto de dados coletados, e pelo tempo limitado para o desenvolvimento do projeto de solução, a figura 21 apresenta a ontologia proposta para o projeto de solução.

Figura 21 – Ontologia de domínio do projeto de solução.



Fonte: Elaboração dos autores, 2016.

Os municípios, estados e país, possuem informações estáticas como coordenadas geográficas, nome e código IBGE para municípios, nome e código IBGE para unidades federativas, e nome para país, essas informações são representadas por propriedades de dados nas respectivas classes de cada entidade. As propriedades e recursos dinâmicos como informações de IDH, transferências de recursos e arrecadação são representadas por classes *FatiaDeTempo* relacionadas a uma unidade de tempo. As classes de *FatiaDeTempo* estão ligadas a suas respectivas entidades através da propriedade de objeto *fatiaDeTempoDe* e estão ligadas a classe *Data* que representa a unidade de tempo, através da propriedade de objeto *aconteceEm*.

As propriedades de objeto *parteDeUf*, *parteDePais*, *contemMunicipio* e *contemUf* são propriedades fluentes, ou seja, podem mudar com o tempo, por isso também estão associadas com as classes *FatiaDeTempo* que representam municípios, unidades federativas e país. No quadro 8 é apresentada as propriedades dos relacionamentos das classes da ontologia:

Quadro 8 – Características dos relacionamentos das classes.

Propriedade	Funcional	Inversa Funcional	Transitiva	Simétrica	Assimétrica	Reflexiva	Irreflexiva
<i>aconteceEm</i>	Sim	-	-	-	Sim	-	Sim
<i>fatiaDeTempoDe</i>	Sim	-	-	-	Sim	-	Sim
<i>gera</i>	-	Sim	-	-	Sim	-	Sim
<i>parteDePais</i>	Sim	-	Sim	-	Sim	-	Sim
<i>parteDeUf</i>	Sim	-	Sim	-	Sim	-	Sim
<i>recebe</i>	-	Sim	-	-	Sim	-	Sim
<i>temFatiaDeTempo</i>	-	Sim	-	-	Sim	-	Sim
<i>temMunicipio</i>	-	Sim	-	-	Sim	-	Sim
<i>temUf</i>	-	Sim	-	-	Sim	-	Sim

Fonte: Elaboração dos autores, 2016.

As propriedades dos relacionamentos foram definidas conforme descrição e orientação de Smith, Welt e McGuinness (2009) e também de Motik, Patel-Schneider e Parsia (2012):

- a propriedade de relacionamento ***aconteceEm*** é funcional pois cada indivíduo das classes ***FatiaDeTempoMunicipio***, ***FatiaDeTempoUf*** e ***FatiaDeTempoPais*** fará relação com apenas um indivíduo da classe ***Ano***, é assimétrica pois o sentido da relação é único, e por fim é irreflexiva pois não é

usada para que indivíduos das classes **FatiaDeTempoMunicipio**, **FatiaDeTempoUf** e **FatiaDeTempoPais** se relacionem entre si;

- a propriedade **fatiaDeTempoDe** é funcional pois cada indivíduo das classes **FatiaDeTempoMunicipio**, **FatiaDeTempoUf** e **FatiaDeTempoPais** fará relação com apenas um indivíduo de suas classes de destino **Municipio**, **Uf** e **Pais**, é uma propriedade assimétrica pois o sentido da relação é único, e por fim é irreflexiva pois não é usada para que os indivíduos das classes **FatiaDeTempo** se relacionem entre si;
- a propriedade **temFatiaDeTempo**, inversa de **fatiaDeTempoDe**, é uma propriedade inversa funcional pois cada indivíduo das classes **FatiaDeTempoMunicipio**, **FatiaDeTempoUf** e **FatiaDeTempoPais**, podem fazer ligação com apenas um indivíduo das suas respectivas classes originais **Municipio**, **Uf** e **Pais**, trata-se também de propriedades assimétricas, já que o sentido da relação é único, e finalmente é considerada irreflexiva pois as não são utilizadas para relacionar indivíduos de mesma classe;
- a propriedade **gera** é inversa funcional pois indivíduo da classe **Arrecadacao** pode fazer relação com apenas um indivíduo da classe **FatiaDeTempoUf**, é uma propriedade assimétrica pois o sentido da relação é único, e por fim é irreflexiva pois não é usada para que os indivíduos da classe **FatiaDeTempoUf** se relacionam entre si;
- as propriedades **parteDeUf** e **partePais**, inversas de **contemMunicipio** e **contemUf**, são propriedades funcionais, pois cada indivíduo da classe **FatiaDeTempoMunicipio** pode se relacionar através da propriedade **parteDeUf** com apenas um indivíduo da classe **FatiaDeTempoPais**, e um indivíduo da classe **FatiaDeTempoUf** pode fazer relação com apenas um indivíduo da classe **FatiaDeTempoPais** usando a propriedade **parteDePais**, é uma propriedade transitiva já que se uma **FatiaDeTempoMunicipio** faz **parteDeUf** **FatiaDeTempoUf** então **FatiaDeTempoMunicipio** também faz parte de **FatiaDeTempoPais**, trata-se também de propriedades assimétricas, já que o sentido da relação é único, e finalmente é considerada irreflexiva pois as propriedades não são usadas para relacionar indivíduos da mesma classe;



- as propriedades **temMunicipio** e **temUf**, inversas de **parteDeUf** e **parteDePais**, são propriedades inversa funcionais pois cada indivíduo das classes **FatiaDeTempoMunicipio**, pode fazer ligação com apenas um indivíduo da classe **FatiaDeTempoUf**, e cada indivíduo da classe **FatiaDeTempoUf** pode fazer relação com apenas um indivíduo da classe **FatiaDeTempoPais**, trata-se também de propriedades assimétricas, já que o sentido da relação é único, e finalmente é considerada irreflexiva pois as não são utilizadas para relacionar indivíduos de mesma classe;
- por último a propriedade **recebe**, semelhante a propriedade gera, a propriedade recebe é inversa funcional pois cada indivíduo da classe **Arrecadacao** pode fazer relação com apenas um indivíduo da classe **FatiaDeTempoUf** ou da classe **FatiaDeTempoMunicipio**, é uma propriedade assimétrica pois o sentido da relação é único, e por fim é irreflexiva pois não é usada para que os indivíduos da classe **FatiaDeTempoUf** ou **FatiaDeTempoMunicipio** se relacionam entre si.

Com a devida apresentação da elaboração da ontologia de domínio, suas propriedades e de toda a estrutura envolvida para a representação dos dados, a seguinte seção apresenta a execução dos passos para o desenvolvimento da proposta de solução.

### 5.2.2 Execução

Esta seção demonstra a execução dos passos da proposta de solução. A estrutura desenvolvida é dividida em coleta de dados, extração dos dados, alinhamento e/ou mapeamento ontológico, triplificação, persistência de dados e consulta de dados.

#### 5.2.2.1 Coleta de dados

Após definido o domínio de dados verificou-se que a coleta automática desses demandaria tempo de implementação, como os objetivos desse trabalho não consideram uma coleta automática de dados, foi decidido por fazer esta de maneira manual.

Primeiramente foi feito a coleta de dados de municípios, com código IBGE, nome, latitude, longitude, nome do estado, código IBGE do estado, URI do recurso no projeto MUN.lodKem, e URI do recurso no *dataset* DBPedia.org<sup>13</sup>. Esses dados foram obtidos através de uma consulta SPARQL direta no *endpoint* do projeto MUN.lodKem, o *endpoint* está em: <http://lodkem.led.ufsc.br:8890/sparql>, na versão: **geopoliticabr**, na ontologia: <http://150.162.114.62/ontologias/geopoliticabr2.owl>, o quadro 9 apresenta a consulta realizada para obter os dados mencionados.

Quadro 9 – Consulta em SPARQL para retorno de informações de municípios e estados.

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX geopol-owl: <http://lodkem.ufsc.br/onto/geopoliticabr#>
SELECT DISTINCT *
WHERE {
    ?municipio rdf:type geopol-owl:Municipio.
    ?municipio geopol-owl:temPontoCentralMun ?geo.
    ?municipio geopol-owl:pertenceEstado ?uf.
    ?municipio geopol-owl:temCodIbgeMun ?cod_ibge.
    ?municipio geopol-owl:temNomeMun ?nome.
    ?geo geopol-owl:lat ?lat.
    ?geo geopol-owl:lon ?lon.
    ?uf geopol-owl:temNomeEst ?nome_est.
    ?uf geopol-owl:temCodIbgeEst ?cod_ibge_est.
    ?municipio owl:sameAs ?dbpedia.
}
```

Fonte: Elaboração dos autores, 2016.

Também foram coletados dados temporais referentes aos municípios, estados e país, como IDH geral de município, estado e país, IDH de longevidade de estados e país, IDH de educação de estados e país, IDH de renda de estados e país, arrecadações, com valor e tipo, geradas por estados, recursos, com valor e tipo, recebidos por estados e municípios. Esses dados foram obtidos de maneira manual através dos arquivos identificados no quadro 10.

Quadro 10 – Arquivos coletados e suas fontes.

Conteúdo do Arquivo	Fonte do Arquivo
Municípios - IDH geral	<a href="http://api.pgi.gov.br/api/1/serie/2427.json">http://api.pgi.gov.br/api/1/serie/2427.json</a>
País - IDH de educação	<a href="http://api.pgi.gov.br/api/1/serie/2461.json">http://api.pgi.gov.br/api/1/serie/2461.json</a>
País - IDH de longevidade	<a href="http://api.pgi.gov.br/api/1/serie/2460.json">http://api.pgi.gov.br/api/1/serie/2460.json</a>
País - IDH de renda	<a href="http://api.pgi.gov.br/api/1/serie/2459.json">http://api.pgi.gov.br/api/1/serie/2459.json</a>
País - IDH geral	<a href="http://api.pgi.gov.br/api/1/serie/2458.json">http://api.pgi.gov.br/api/1/serie/2458.json</a>

<sup>13</sup> Disponível em: <http://dbpedia.org/>

Ufs - IDH de educação	<a href="http://api.pgi.gov.br/api/1/serie/2457.json">http://api.pgi.gov.br/api/1/serie/2457.json</a>
Ufs - IDH de longevidade	<a href="http://api.pgi.gov.br/api/1/serie/2456.json">http://api.pgi.gov.br/api/1/serie/2456.json</a>
Ufs - IDH de renda	<a href="http://api.pgi.gov.br/api/1/serie/2455.json">http://api.pgi.gov.br/api/1/serie/2455.json</a>
Ufs - IDH geral	<a href="http://api.pgi.gov.br/api/1/serie/2454.json">http://api.pgi.gov.br/api/1/serie/2454.json</a>
Ufs – Arrecadações em 2010	<a href="http://www.receita.fazenda.gov.br/publico/arre/2010/PorEstado/ArrecadacaoUFJan-Dez10.xls">http://www.receita.fazenda.gov.br/publico/arre/2010/PorEstado/ArrecadacaoUFJan-Dez10.xls</a>
Ufs – Recebimento de recursos em 2010	<a href="http://www.portaltransparencia.gov.br/PortalTransparenciaListaUFs.asp?Exercicio=2010&amp;Pagina=1">http://www.portaltransparencia.gov.br/PortalTransparenciaListaUFs.asp?Exercicio=2010&amp;Pagina=1</a>
Municípios – Recebimento de recursos em 2010	<a href="http://arquivos.portaldatransparencia.gov.br/PortalTransparenciaEscolheTipoDePlanilha.asp?origem=TransferenciasEstMun&amp;Planilha=2010">http://arquivos.portaldatransparencia.gov.br/PortalTransparenciaEscolheTipoDePlanilha.asp?origem=TransferenciasEstMun&amp;Planilha=2010</a>

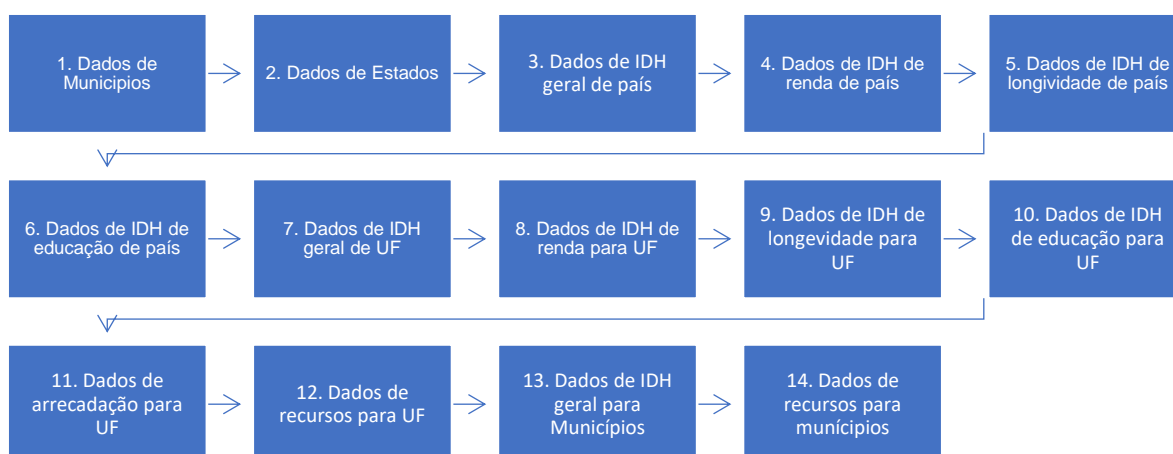
Fonte: Elaboração dos autores, 2016.

Tendo conhecimento do cenário e arquivos coletados para a contemplação do mesmo, a seguinte seção apresenta o processo de extração dos dados.

#### 5.2.2.2 Extração dos dados

Os dados foram extraídos de maneira sequencial para facilitar o mapeamento e alinhamento dos mesmos através do triplificador desenvolvido. A extração dos dados se dá por meio da leitura dos arquivos JSON, arquivos que não estavam nesse formato foram convertidos pelo triplificador antes da extração. A figura 22 demonstra a sequência em que os dados foram extraídos.

Figura 22 – Fluxograma de sequência de extração dos dados.



Fonte: Elaboração dos autores, 2016.

Dados de municípios e estados são extraídos do arquivo gerado pela consulta realizada no *dataset* MUN.lodKem, esses dados são armazenados em objetos de classe *Município* e *Uf*. Os demais dados são extraídos na sequência, entretanto não são armazenados em objetos, pois ao serem extraídos já são utilizados para a criação dos recursos.

### 5.2.2.3 Padronização de URIs

A utilização de URIs para nomear as coisas é o primeiro princípio dos dados conectados, como URIs são responsáveis por identificar os recursos por meio de um endereço, é fundamental que exista um padrão para a escolha desses.

Os recursos, neste trabalho, utilizam um endereço base fictício para suas URIs, **`http://lod.data.gov.br/`**, criado com o propósito de uma melhor ilustração nas consultas e acessos realizados no decorrer do trabalho. O quadro 11 apresenta a base de URI para elementos da ontologia e recursos.

Quadro 11 – URIs base por tipo de elemento.

Tipo	URI
Ontologia	<code>http://lod.data.gov.br/onto/</code>
Recursos	<code>http://lod.data.gov.br/resource/</code>

Fonte: Elaboração dos autores, 2016.

As classes e propriedade de ontologia recebem seus URIs previamente na sua criação no *Protégé*, como exemplos de URIs de elementos de ontologia, sabendo que a URI base para estes é **`http://lod.data.gov.br/onto/`**, temos, respectivamente, as URIs da classe município e da propriedade *fatiaDeTempoDe*:

**`http://lod.data.gov.br/onto/Municipio`**  
**`http://lod.data.gov.br/onto/fatiaDeTempoDe`**

Os URIs de elementos de ontologia são usados posteriormente para mapeamento de recursos junto a ontologia e consultas via SPARQL.

Os recursos, neste trabalho, são identificados seguindo a URI base **`http://lod.data.gov.br/resource/`**, como apresentado anteriormente no quadro 11. O quadro 12 apresenta um detalhamento do padrão seguido por tipo de recurso.

Quadro 12 – Arquivos coletados e suas fontes.

Tipo de Recurso	URI
Municípios	<code>http://lod.data.gov.br/resource/municipio(código IBGE)</code>
FatiaDeTempoMunicipio	<code>http://lod.data.gov.br/resource/municipio(código IBGE)Em(ano)</code>
Uf	<code>http://lod.data.gov.br/resource/uf(código IBGE)</code>
FatiaDeTempoUf	<code>http://lod.data.gov.br/resource/uf(código IBGE)Em(ano)</code>
País	<code>http://lod.data.gov.br/resource/(nome país)(código IBGE)</code>
FatiaDeTempoPaís	<code>http://lod.data.gov.br/resource/(nome país)Em(ano)</code>
Data	<code>http://lod.data.gov.br/resource/(ano)</code>
TransferenciaRecurso	<code>http://lod.data.gov.br/resource/transferencia_recurso(id)</code>
Arrecadacao	<code>http://lod.data.gov.br/resource/arrecadacao_(id)</code>

Fonte: Elaboração dos autores, 2016.

Os URIs apresentados no quadro 12, apresentam valores entre parênteses, esses valores podem ser encarados como variáveis dependendo dos recursos criados, como por exemplo, os URIs para o município de Florianópolis e fatia de tempo do mesmo:

**`http://lod.data.gov.br/resource/municipio4205407`**

**`http://lod.data.gov.br/resource/municipio4205407Em2010`**

Com o conhecimento sobre os padrões utilizados nas URIs de identificação, a seguinte seção apresenta o mapeamento e ligação dos recursos com bases externas.

#### 5.2.2.4 Mapeamento e ligação dos recursos

Os recursos criados, precisam ser mapeados com a ontologia de domínio e alguns destes conectados a recursos externos de outros *datasets*. Na criação cada recurso já é mapeado com a ontologia, como por exemplo, o recurso de País é o primeiro a ser criado, no quadro 13 é detalhada a criação desse recurso usando a API Jena e a linguagem de programação Java.

Quadro 13 – Criação do recurso país usando a API Jena e linguagem de programação Java.

```
OntClass pais = ontModel.getOntClass(URI_BASE_ONT + "/País");
Individual brasil = ontModel.createIndividual(URI_BASE_REC + "/brasil", pais);
```

Fonte: Elaboração dos autores, 2016.

Na primeira linha do quadro 13, referenciamos a classe País da ontologia através da variável `pais`, na segunda linha, criamos o indivíduo “brasil”, esse indivíduo é criado passando a URI que ele assumirá, como definido anteriormente, o URI que identifica um indivíduo de País é composto pela URI base de recursos acrescentando o nome do país em seguida, o URI deste código então ficaria: `http://lod.data.gov.br/resource/brasil`, e por fim é informada a classe da ontologia na qual esse indivíduo será mapeado, no exemplo a variável `pais` representando a classe país da ontologia.

Recursos de Município e Uf são criados posteriormente e reunidos em uma lista de instâncias de suas respectivas classes em Java, para uso posterior na criação de outros recursos. Recursos de Município além do mapeamento com a classe da ontologia, seguindo o quarto princípio dos dados conectados, também são conectados com indivíduos de outros *datasets*, como o DBPedia.org e o MUN.lodKEM, o quadro 14 demonstra esse mapeamento e ligação em código Java com uso da API Jena.

Quadro 14 – Conectando um município com outros municípios de outros *datasets*.

```
municipioCriado.addSameAs(municipioLaco.getDbPediaRef);
municipioCriado.addSameAs(municipioLaco.getMunLodKemRef);
```

Fonte: Elaboração dos autores, 2016.

Na primeira linha de código do quadro 14, o indivíduo é conectado com o indivíduo semelhante no *dataset* DBPedia.org e na segunda linha é conectado com o indivíduo semelhante no *dataset* MUN.lodKem, para isso é utilizada uma propriedade denominada *sameAs* que é responsável por inferir que um recurso é igual a outro.

Outros recursos como, as fatias de tempo, arrecadação e transferência de recursos, são criados diretamente após a extração dos dados dos arquivos, já fazendo o mapeamento devido com a ontologia.

### 5.2.2.5 Triplificação dos dados

A triplificação consiste na transformação dos dados em informação, através da criação de triplas com (**sujeito** predicado **objeto**). As triplas são a base do nosso banco de dados, a ontologia é formada em triplas, o mapeamento dos recursos com a ontologia é em forma de triplas, assim como as relações entre os recursos, que é abordada nessa seção.

Propriedades de IDH são acrescentadas as fatias de tempo de município, estado e país. Recursos de arrecadações, com valores e tipos, são relacionados fatias de tempo de estados e transferências de recursos, com valores e tipos, são relacionadas a fatias de tempo de estados e municípios. Como os indivíduos das classes de fatias de tempo são temporais, eles também serão relacionados a indivíduos da classe Data.

As triplas para fatias de tempo de país são adicionadas diretamente ao recurso do país Brasil, fatias de tempo de município e fatias de tempo de estado são criadas dentro de um laço que percorre a lista das respectivas classes, o quadro 15 apresenta um exemplo em código da triplificação de relações da fatia de tempo de município relacionada ao ano 2010.

Quadro 15 – Exemplo de triplificação de relações da fatia de tempo município.

```
for( Municipio munLaco : listaMunicipios ) {

    Individual municipioOriginal = ontModel.getIndividual(URI_BASE_REC + " /municipio" +
munLaco.getCodIbge());
    Individual municipioEm2010 = ontModel.getIndividual(URI_BASE_REC + " /municipio" +
munLaco.getCodIbge() + " /Em2010");

    municipioEm2010.addProperty(parteDeUf, ontModel.getIndividual(URI_BASE_REC + "/uf" +
munLaco.getCodIbgeEst() + "Em2010"));
    municipioEm2010.addProperty(aconteceEm, a2010);
    municipioEm2010.addProperty(fatiaDeTempoDe, municipioOriginal);

}
```

Fonte: Elaboração dos autores, 2016.

As duas primeiras linhas de código dentro do laço *for* são responsáveis por recuperar os indivíduos no modelo. A terceira linha de código define que o município do laço faz parte de seu UF, a quarta linha define que a fatia de tempo do município acontece no ano de 2010, e a quinta e última linha define que a fatia de tempo é uma fatia de tempo de seu município. Ao definir as propriedades *parteDeUf* e *fatiaDeTempoDe* as propriedades

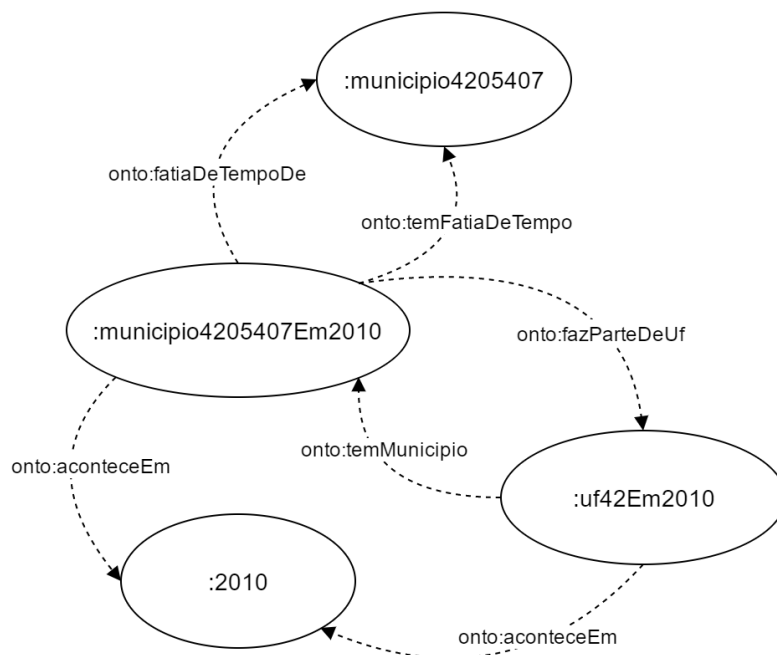
*temMunicípio* e *temFatiaDeTempo* não precisam ser definidas pois ao ser o inverso de *parteDeUf* e *fatiaDeTempoDe* podem ser automaticamente inferidas.

Considerando que (x) é o código IBGE do município e (y) é o código IBGE da UF, as últimas três linhas de código no quadro 15 criam as triplas:

<b>Município (x) Em2010</b>	<b>parteDeUf</b>	<b>Uf (y) Em2010</b>
<b>Município (x) Em2010</b>	<b>aconteceEm</b>	<b>2010</b>
<b>Município (x) Em2010</b>	<b>fatiaDeTempoDe</b>	<b>Município</b>

As triplas expressadas acima são representadas em um grafo apresentado na figura 23, o grafo é montado com base na cidade de Florianópolis, que tem código do IBGE: 4205407 e no estado de Santa Catarina com código IBGE: 42.

Figura 23 – Grafo gerado com as triplas de município e UF, fatias de tempo e data.



Fonte: Elaboração dos autores, 2016.

As relações expressadas explicitamente são: *aconteceEm*, *fazParteDeUf* e *fatiaDeTempoDe*. Relações como: *temMunicípio*, *temFatiaDeTempo* são inferidas automaticamente, por serem o inverso de *fazParteDeUf* e *fatiaDeTempoDe*.



As fatias de tempo recebem conteúdos temporais como dados de IDH, transferências de recursos e arrecadações, esses dados são relacionados a fatia de tempo de acordo com o tempo em que a mesma acontece. O quadro 16 por exemplo, apresenta o código em Java, onde as transferências de recursos são relacionadas às fatias de tempo de município.

Quadro 16 – Exemplo de triplificação de dados relacionados a recursos para municípios.

```
OntClass recurso = ontModel.getOntClass("http://lod.data.gov.br/onto/TranferenciaRecurso");

for( Uf uflaco : listaUfs ) {

    Object jsonObj = parser.parse(this.lerArquivo("trs_"+uflaco.getCodIbge()+".json"));
    JSONArray jArray = (JSONArray) jsonObj;
    Iterator<JSONObject> lacoObjRecursoMunValores = jArray.iterator();

    while (lacoObjRecursoMunValores.hasNext()) {

        JSONObject recObj = lacoObjRecursoMunValores.next();

        Individual rec = ontModel.createIndividual(URI_BASE_REC +
            "/transferencia_recurso_"+UUID.randomUUID().toString(), recurso);

        rec.addProperty(tipo, recObj.get("NomeFuncao").toString()).addLiteral(valor,
            Double.parseDouble(recObj.get("Valor").toString()));

        ontModel.getIndividual(URI_BASE_REC +
            "/municipio"+(recObj.get("CodigoMunicipio").toString())+"Em2010").addProperty(recebe, rec);

    }
}
```

Fonte: Elaboração dos autores, 2016.

A leitura dos arquivos de transferência de recursos acontece dentro de um laço que percorre os estados, pois cada arquivo está com o código da UF respectivo. Em cada volta, o laço faz a leitura de um arquivo referente a uma UF, e gera outro laço para a leitura dos itens dentro do documento, aí então cada item é associado a fatia de tempo do respectivo município no ano de 2010, pois o arquivo de recursos é somente do ano de 2010. O objeto de transferência é criado com um ID randômico e único para representar seu URI, logo após são acrescentados o tipo e o valor relativo ao recurso, na figura 24 é apresentado a nuvem de palavras relativas aos tipos de recursos transferidos a municípios.

Figura 24 – Nuvem de palavras relativas aos tipos de recursos transferidos.



Fonte: Elaboração dos autores, 2016.

Podemos verificar que na figura 24 que as transferências relativas a encargos especiais se sobressaem em relação a outras, seguida de educação.

Após a triplificação, os dados podem ser persistidos no *dataset*, no caso deste trabalho, o Jena TDB, o quadro 17 demonstra as linhas de código para persistir o modelo de domínio no *dataset*.

Quadro 17 – Linhas de código para persistência do modelo de domínio no *dataset*.

```
triplificador.dataset.getDefaultModel().add(ontModel);
TDB.sync(triplificador.dataset);
```

Fonte: Elaboração dos autores, 2016.

Neste caso o modelo de domínio usado para a adição das triplas é adicionado como o modelo padrão do *dataset*, após isso o *dataset* é sincronizado com o TDB e está pronto para receber consultas e acessos via HTML nos recursos. A seguinte seção apresenta as possibilidades de acesso aos dados.

### 5.2.2.6 Acessando os dados

Contemplando o segundo e terceiro princípio dos dados conectados, os dados tipificados podem ser acessados através de consultas no SPARQL *endpoint* ou através do uso da URI dos recursos em qualquer navegador. Na figura 25 temos a interface do SPARQL *endpoint* do projeto de solução, que no cenário do projeto é acessado pelo endereço **`http://lod.data.gov.br:3030/`**.

Figura 25 – Interface do SPARQL *endpoint* do projeto de solução.

The screenshot displays the SPARQL endpoint interface. At the top, the 'SPARQL ENDPOINT' field contains the URL 'http://lod.data.gov.br:3030/GeoPolitica/query'. The 'CONTENT TYPE (SELECT)' dropdown is set to 'JSON', and the 'CONTENT TYPE (GRAPH)' dropdown is set to 'Turtle'. Below these, a query editor shows the following SPARQL query:

```

1 PREFIX : <http://lod.data.gov.br/GeoPolitica#>
2 SELECT *
3 WHERE {
4   ?municipio a :Municipio;
5               :codIbge "4205407";
6               :nome ?nome.
7 }
8 LIMIT 10
9

```

Below the query editor, the 'QUERY RESULTS' section shows a table view. The table has two columns: 'municipio' and 'nome'. The first entry is:

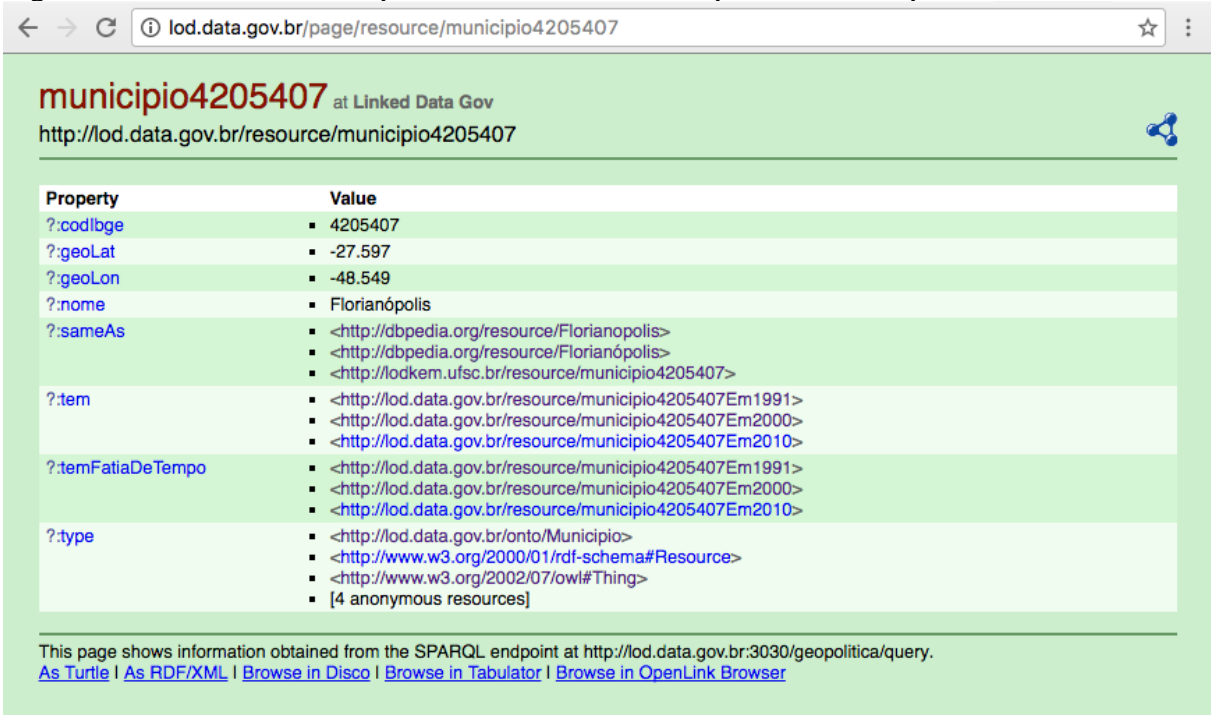
	municipio	nome
1	<a href="http://lod.data.gov.br/GeoPolitica/resource/municipio4205407">http://lod.data.gov.br/GeoPolitica/resource/municipio4205407</a>	"Florianópolis"

The interface also includes a search bar, a 'Show 50 entries' dropdown, and a 'Raw Response' button.

Fonte: Elaboração dos autores, 2016.

Na figura 25, é realizada uma consulta que listará o URI e nome do município que tenha o código de IBGE igual a 4205407. Ao clicar no URI do recurso é possível ver as informações detalhadas do mesmo em um navegador, como ilustrado na figura 26.

Figura 26 – Interface HTML para do recurso do município de Florianópolis.



← → ↻ lod.data.gov.br/page/resource/municipio4205407 ☆ ⋮

**municipio4205407** at Linked Data Gov  
<http://lod.data.gov.br/resource/municipio4205407>

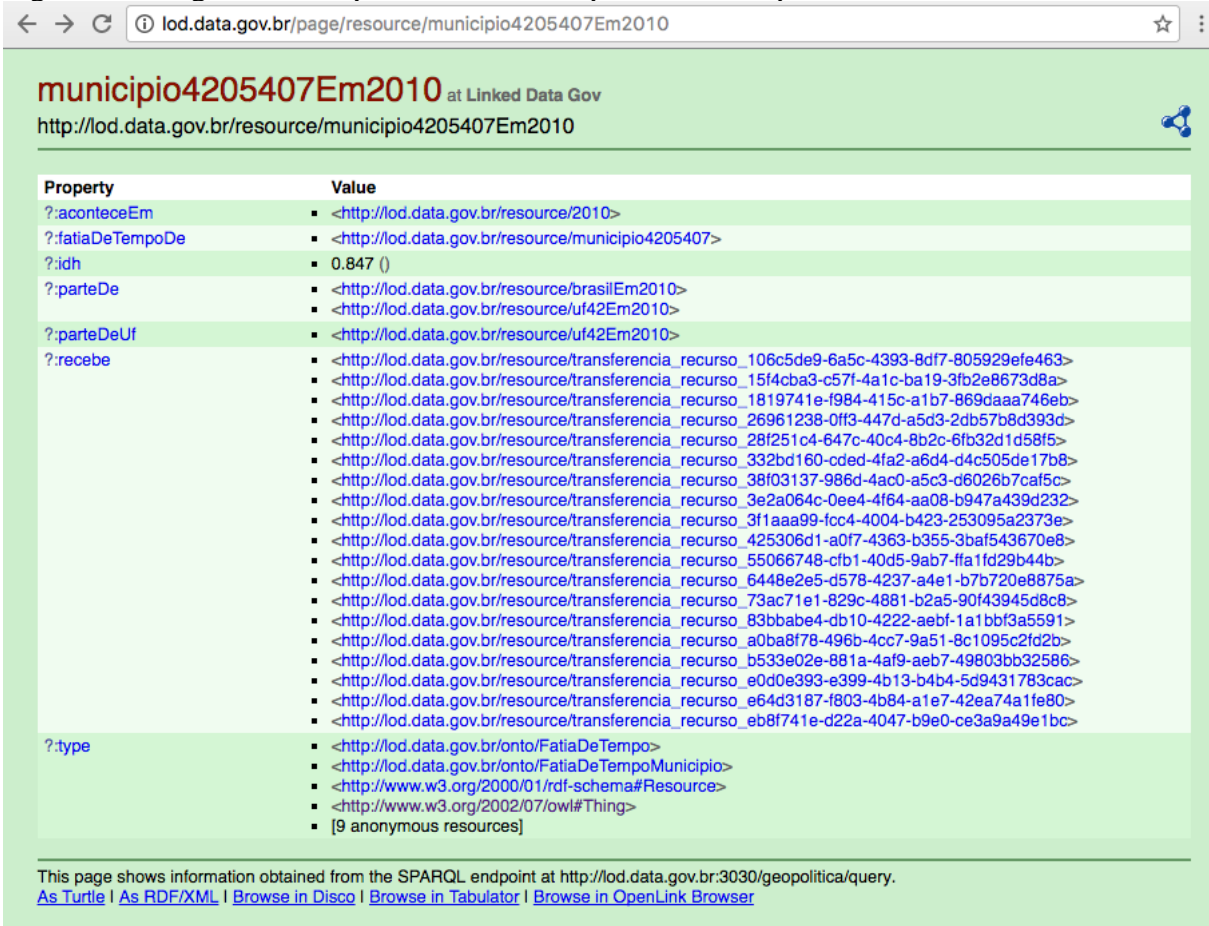
Property	Value
?:codibge	▪ 4205407
?:geoLat	▪ -27.597
?:geoLon	▪ -48.549
?:nome	▪ Florianópolis
?:sameAs	▪ < <a href="http://dbpedia.org/resource/Florianopolis">http://dbpedia.org/resource/Florianopolis</a> > ▪ < <a href="http://dbpedia.org/resource/Florianópolis">http://dbpedia.org/resource/Florianópolis</a> > ▪ < <a href="http://lodkem.ufsc.br/resource/municipio4205407">http://lodkem.ufsc.br/resource/municipio4205407</a> >
?:tem	▪ < <a href="http://lod.data.gov.br/resource/municipio4205407Em1991">http://lod.data.gov.br/resource/municipio4205407Em1991</a> > ▪ < <a href="http://lod.data.gov.br/resource/municipio4205407Em2000">http://lod.data.gov.br/resource/municipio4205407Em2000</a> > ▪ < <a href="http://lod.data.gov.br/resource/municipio4205407Em2010">http://lod.data.gov.br/resource/municipio4205407Em2010</a> >
?:temFatiaDeTempo	▪ < <a href="http://lod.data.gov.br/resource/municipio4205407Em1991">http://lod.data.gov.br/resource/municipio4205407Em1991</a> > ▪ < <a href="http://lod.data.gov.br/resource/municipio4205407Em2000">http://lod.data.gov.br/resource/municipio4205407Em2000</a> > ▪ < <a href="http://lod.data.gov.br/resource/municipio4205407Em2010">http://lod.data.gov.br/resource/municipio4205407Em2010</a> >
?:type	▪ < <a href="http://lod.data.gov.br/onto/Municipio">http://lod.data.gov.br/onto/Municipio</a> > ▪ < <a href="http://www.w3.org/2000/01/rdf-schema#Resource">http://www.w3.org/2000/01/rdf-schema#Resource</a> > ▪ < <a href="http://www.w3.org/2002/07/owl#Thing">http://www.w3.org/2002/07/owl#Thing</a> > ▪ [4 anonymous resources]

This page shows information obtained from the SPARQL endpoint at <http://lod.data.gov.br:3030/geopolitica/query>.  
[As Turtle](#) | [As RDF/XML](#) | [Browse in Disco](#) | [Browse in Tabulator](#) | [Browse in OpenLink Browser](#)

Fonte: Elaboração dos autores, 2016.

Podemos ver que ao acessar o município por seu URI, são exibidas as informações detalhadas do mesmo, bem como a referência deste em outros *datasets*, podemos conferir também as fatias de tempo conectadas ao município, e navegar por elas, ao clicar na fatia de tempo de 2010 é aberto o recurso referente a esta, como ilustrado na figura 27.

Figura 27 – Página HTML para a fatia de tempo de Florianópolis em 2010.



The screenshot shows a web browser window with the address bar displaying `lod.data.gov.br/page/resource/municipio4205407Em2010`. The page title is **municipio4205407Em2010** at Linked Data Gov. Below the title, the URL `http://lod.data.gov.br/resource/municipio4205407Em2010` is shown. The main content is a table with two columns: **Property** and **Value**.

Property	Value
?aconteceEm	<http://lod.data.gov.br/resource/2010>
?fatiaDeTempoDe	<http://lod.data.gov.br/resource/municipio4205407>
?idh	0.847 ()
?parteDe	<http://lod.data.gov.br/resource/brasilEm2010> <http://lod.data.gov.br/resource/uf42Em2010>
?parteDeUf	<http://lod.data.gov.br/resource/uf42Em2010>
?recebe	<http://lod.data.gov.br/resource/transferencia_recurso_106c5de9-6a5c-4393-8df7-805929efe463> <http://lod.data.gov.br/resource/transferencia_recurso_15f4cba3-c571-4a1c-ba19-3fb2e8673d8a> <http://lod.data.gov.br/resource/transferencia_recurso_1819741e-f984-415c-a1b7-869daaa746eb> <http://lod.data.gov.br/resource/transferencia_recurso_26961238-0ff3-447d-a5d3-2db57b8d393d> <http://lod.data.gov.br/resource/transferencia_recurso_28f251c4-647c-40c4-8b2c-6fb32d1d58f5> <http://lod.data.gov.br/resource/transferencia_recurso_332bd160-cded-4fa2-a6d4-d4c505de17b8> <http://lod.data.gov.br/resource/transferencia_recurso_38f03137-986d-4ac0-a5c3-d6026b7caf5c> <http://lod.data.gov.br/resource/transferencia_recurso_3e2a064c-0ee4-4f64-aa08-b947a439d232> <http://lod.data.gov.br/resource/transferencia_recurso_3f1aaa99-fcc4-4004-b423-253095a2373e> <http://lod.data.gov.br/resource/transferencia_recurso_425306d1-a0f7-4363-b355-3ba5f543670e8> <http://lod.data.gov.br/resource/transferencia_recurso_55066748-cfb1-40d5-9ab7-ffa1fd29b44b> <http://lod.data.gov.br/resource/transferencia_recurso_6448e2e5-d578-4237-a4e1-b7b720e8875a> <http://lod.data.gov.br/resource/transferencia_recurso_73ac71e1-829c-4881-b2a5-90f43945d8c8> <http://lod.data.gov.br/resource/transferencia_recurso_83bbabe4-db10-4222-aebf-1a1bbf3a5591> <http://lod.data.gov.br/resource/transferencia_recurso_a0ba8f78-496b-4cc7-9a51-8c1095c2fd2b> <http://lod.data.gov.br/resource/transferencia_recurso_b533e02e-881a-4af9-aeb7-49803bb32586> <http://lod.data.gov.br/resource/transferencia_recurso_e0d0e393-e399-4b13-b4b4-5d9431783cac> <http://lod.data.gov.br/resource/transferencia_recurso_e64d3187-f803-4b84-a1e7-42ea74a1fe80> <http://lod.data.gov.br/resource/transferencia_recurso_eb8f741e-d22a-4047-b9e0-ce3a9a49e1bc>
?type	<http://lod.data.gov.br/onto/FatiaDeTempo> <http://lod.data.gov.br/onto/FatiaDeTempoMunicipio> <http://www.w3.org/2000/01/rdf-schema#Resource> <http://www.w3.org/2002/07/owl#Thing> [9 anonymous resources]

At the bottom of the page, there is a note: "This page shows information obtained from the SPARQL endpoint at `http://lod.data.gov.br:3030/geopolitica/query`." followed by links: [As Turtle](#), [As RDF/XML](#), [Browse in Disco](#), [Browse in Tabulator](#), and [Browse in OpenLink Browser](#).

Fonte: Elaboração dos autores, 2016.

Os detalhes da fatia de tempo de 2010 do município de Florianópolis apresentam as transferências de recursos recebidas e as relações que esta fatia de tempo ocorre com outros recursos. Ao acessar as ligações externas do município é possível visualizar a página referente ao indivíduo similar em outro *dataset*, como mostra a figura 28, que apresenta o recurso referenciado do *dataset* DBPedia.org, onde podemos ter acesso a informações mais detalhadas do município.

Figura 28 – Página HTML do recurso Florianópolis no *dataset* DBPedia.org.



The screenshot shows a web browser window with the address bar displaying 'dbpedia.org/page/Florianópolis'. The page features the DBpedia logo and navigation links: 'Browse using', 'Formats', 'Faceted Browser', and 'Sparql Endpoint'. The main heading is 'About: Florianópolis', followed by a sub-header: 'An Entity of Type : settlement, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org'. A paragraph describes Florianópolis as the capital of Santa Catarina, Brazil, located on the main island and several smaller islands, with a population of 469,690 in 2015. Below this is a table with two columns: 'Property' and 'Value'.

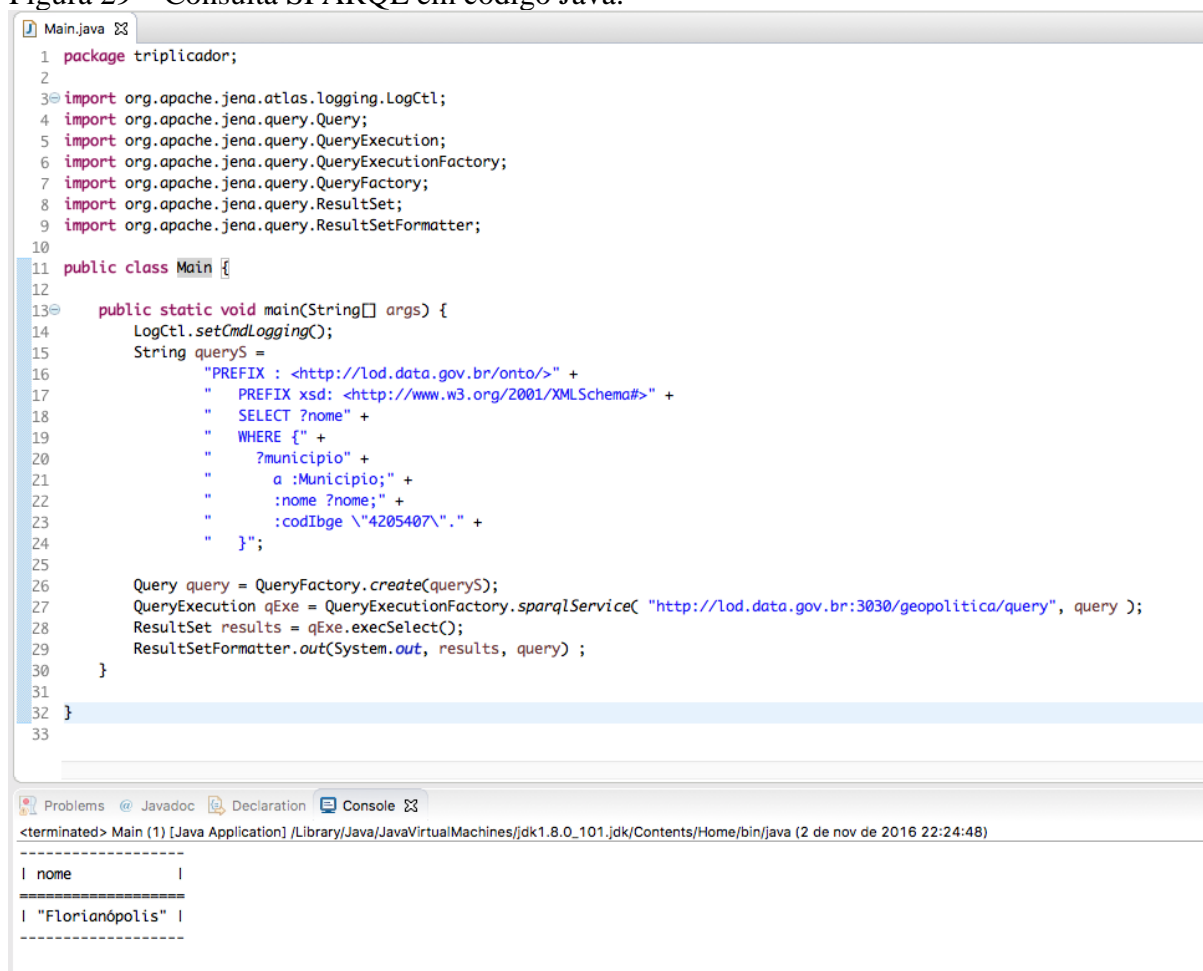
Property	Value
<code>dbo:PopulatedPlace/area</code>	▪ 1.0E-6
<code>dbo:PopulatedPlace/areaTotal</code>	▪ 675.409 ▪ 675406.7394609808
<code>dbo:abstract</code>	▪ Florianópolis (Portuguese pronunciation: [flori.aˈnɔpɔlis]) is the capital city and second largest city of the state of Santa Catarina, in the South region of Brazil. It is composed of one main island, the Island of Santa Catarina (Ilha de Santa Catarina), and several smaller islands.

Fonte: Elaboração dos autores, 2016.

O uso de páginas HTML para exibir os recursos persistidos faz com que seja mais fácil a visualização de propriedades e conexões do mesmo, o que auxiliará tanto em consultas internas quanto externas.

O SPARQL *endpoint* fica aberto para consultas tanto manuais como automatizadas por máquina, o mesmo pode ser usado para consultas automáticas originadas em sistemas. A figura 29 apresenta uma consulta gerada via linguagem de programação Java e seu resultado logo abaixo.

Figura 29 – Consulta SPARQL em código Java.



```

1 package triplicador;
2
3 import org.apache.jena.atlas.logging.LogCtl;
4 import org.apache.jena.query.Query;
5 import org.apache.jena.query.QueryExecution;
6 import org.apache.jena.query.QueryExecutionFactory;
7 import org.apache.jena.query.QueryFactory;
8 import org.apache.jena.query.ResultSet;
9 import org.apache.jena.query.ResultSetFormatter;
10
11 public class Main {
12
13     public static void main(String[] args) {
14         LogCtl.setCmdLogging();
15         String queryS =
16             "PREFIX : <http://lod.data.gov.br/onto/>" +
17             "PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>" +
18             "SELECT ?nome" +
19             "WHERE {" +
20             "    ?municipio" +
21             "    a :Municipio;" +
22             "    :nome ?nome;" +
23             "    :codIbge \"4205407\"." +
24             "}";
25
26         Query query = QueryFactory.create(queryS);
27         QueryExecution qExe = QueryExecutionFactory.sparqlService( "http://lod.data.gov.br:3030/geopolitica/query", query );
28         ResultSet results = qExe.execSelect();
29         ResultSetFormatter.out(System.out, results, query);
30     }
31 }
32
33

```

Problems Javadoc Declaration Console

<terminated> Main (1) [Java Application] /Library/Java/JavaVirtualMachines/jdk1.8.0\_101.jdk/Contents/Home/bin/java (2 de nov de 2016 22:24:48)

```

-----
| nome |
=====
| "Florianópolis" |
-----

```

Fonte: Elaboração dos autores, 2016.

Apresentado os modos no qual os dados podem ser acessados, a próxima seção aborda as consultas realizadas ao domínio como uma forma de análise do projeto de solução.

### 5.2.3 Consultas ao domínio

Para realizar a avaliação do experimento e concluir a proposta de solução deste trabalho, para que os dados persistidos possam ser de fato analisados, a fase de consulta aos dados é importante pois permite a análise dos dados persistidos.

Na definição das consultas ao domínio foram considerados os seguintes fatores, o conjunto de dados levantados, a data em que os dados mais se relacionam, a ontologia de domínio definida e as suas relações, considerando a relevância que a pergunta teria para uma análise qualitativa dos dados.

Com base no ano de 2010, para realizar a avaliação dos dados persistidos, foram definidas as perguntas a seguir:

1. Qual é a relação do recurso recebido pelos municípios e seu nível de IDH geral?
2. Qual é a relação entre recursos recebidos pelos estados e seu nível de IDH geral?
3. Qual é a relação entre arrecadações geradas pelos estados e seu nível de IDH geral?
4. Qual é a relação entre recursos recebidos pelos estados e seu nível de IDH de educação?
5. Qual é a relação entre arrecadações geradas pelos estados e seu nível de IDH de educação?
6. Qual é a relação entre recursos recebidos pelos estados e seu nível de IDH de longevidade?
7. Qual é a relação entre arrecadações geradas pelos estados e seu nível de IDH de longevidade?
8. Qual é a relação entre recursos recebidos pelos estados e seu nível de IDH de renda?
9. Qual é a relação entre arrecadações geradas pelos estados e seu nível de IDH de renda?
10. Qual o município que mais recebeu recurso e qual seu IDH?

Com as perguntas devidamente elaboradas, a seção seguinte terá como objetivo a análise dos dados obtidos com a realização dessas consultas no banco.

#### 5.2.3.1 Análise dos dados

Após a definição das perguntas anteriores de forma a realizar a análise dos dados levantados e persistidos, para que estas fossem respondidas foram elaboradas consultas no formato de SPARQL para os dados do domínio. O SPARQL foi elaborado pelo W3C RDF Data Access Working Group, e nada mais é que uma linguagem de consulta e um protocolo para acesso a RDF. (THE APACHE SOFTWARE FOUNDATION, 2016g).



A seguir são apresentadas as perguntas elaboradas, como também as consultas no banco e a análise dos resultados obtidos.

A primeira pergunta visava saber qual é a relação do recurso recebido pelos municípios e seu nível de IDH geral. No quadro 18 é apresentada a consulta ao banco e em continuidade desta constam as análises relacionadas aos dados extraídos da consulta.

#### Quadro 18 – Consulta recurso recebido por municípios por IDH.

```

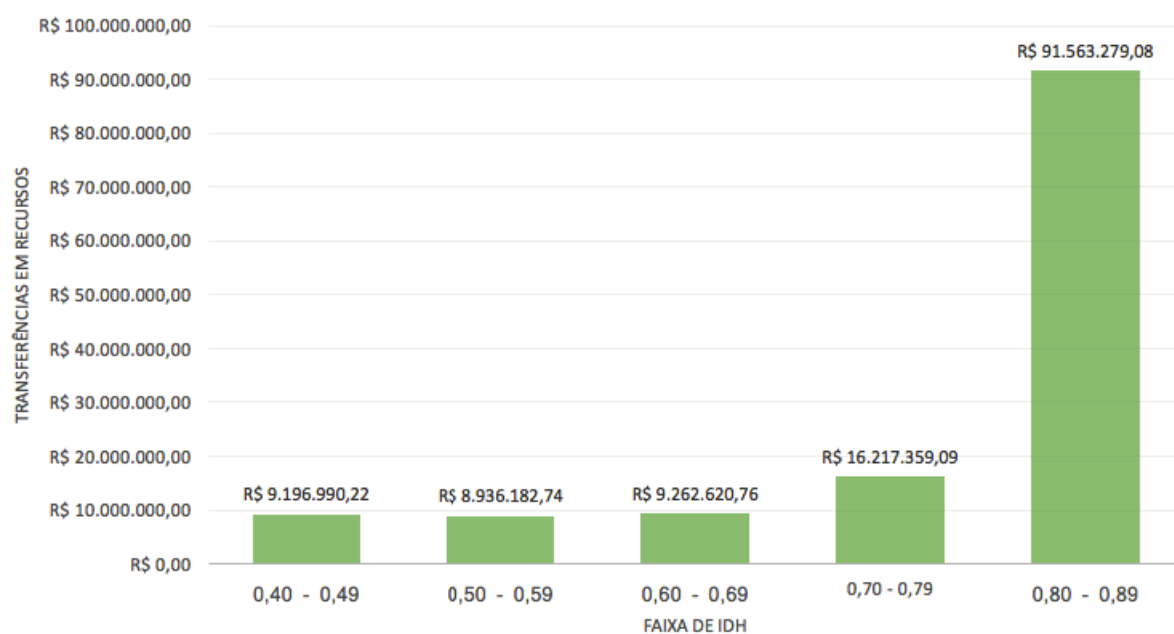
PREFIX : <http://lod.data.gov.br/onto/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT (SAMPLE(xsd:decimal(?idh)) AS ?idhMun) (SUM(xsd:decimal(?valor))
AS ?valorRecebido)
WHERE {
  ?municipioEm2010
    a :FatiaDeTempoMunicipio;
    :aconteceEm ?ano;
    :fatiaDeTempoDe ?municipio;
    :idh ?idh;
    :recebe ?recurso.
  ?ano
    :ano "2010".
  ?municipio
    :codIbge ?codIbge.
  ?recurso
    :valor ?valor.
}
GROUP BY ?codIbge

```

Fonte: Elaboração dos autores, 2016.

Com a consulta ao banco realizada anteriormente, os dados obtidos foram analisados, tendo a relação do recurso recebido e o IDH por município, chegamos no resultado exposto na figura 30 exposta a seguir:

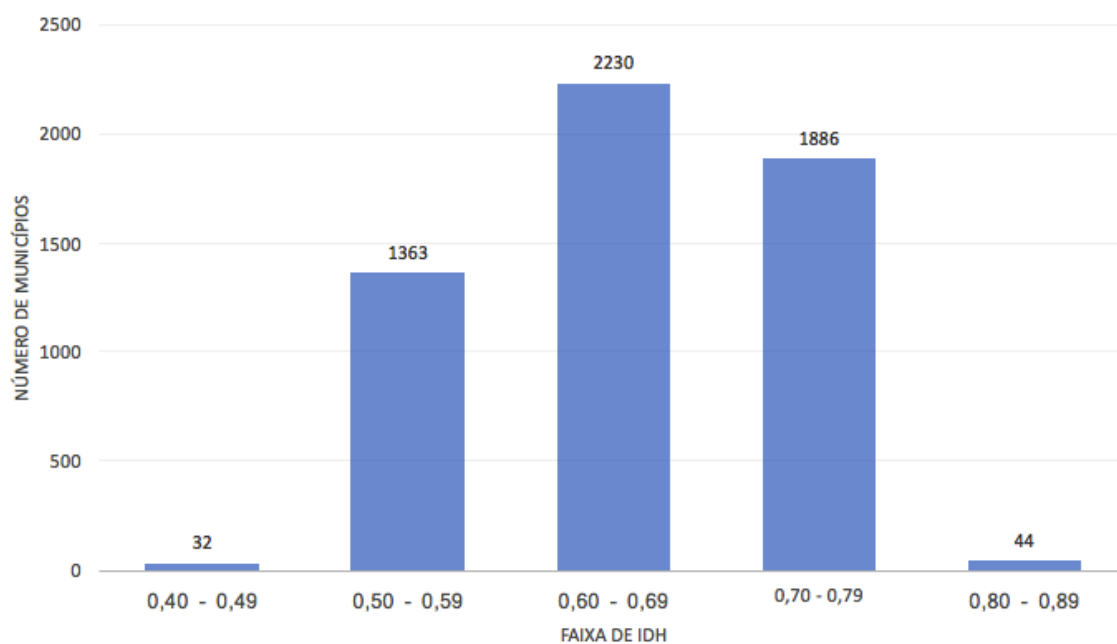
Figura 30 – Recurso recebido por municípios por faixa de IDH.



Fonte: Elaborado pelos autores, 2016.

Além da análise anterior, foi necessário realizar também a observação entre a quantidade de municípios que constam por faixa de IDH, que é apresentada na figura 31.

Figura 31 – Quantidade de Municípios por faixa de IDH.



Fonte: Elaborado pelos autores, 2016.

Verificamos que apesar da faixa de IDH nos níveis entre 0,80 e 0,89 terem recebido o maior valor de recurso, é também uma das faixas de IDH que possui a menor quantidade de municípios (44 municípios), ficando atrás somente da faixa de IDH entre 0,40 e 0,49 (32 municípios). Foi possível identificar com a análise dessas informações que os municípios que receberam a maior quantidade de recurso constam grandes capitais brasileiras, como São Paulo, Curitiba, Belo Horizonte, Porto Alegre, entre outros, que naturalmente possuem uma quantidade maior na sua população, porém foi possível verificar também que entre essa faixa com alto nível de IDH, possuem municípios com menos de 100 mil habitantes, entre eles um que teve bastante destaque entre os dados analisados, foi o município Águas de São Pedro que possui uma população de somente 2.707 habitantes, porém com alto recurso recebido, e também alto nível de IDH (R\$4.274.658,13 e 0,854).

Em continuidade a análise dos dados obtidos através das perguntas, a seguir constam as análises relacionadas a segunda pergunta, ao qual pretende obter qual é a relação entre recursos recebidos pelos estados e seu nível de IDH geral. O quadro 19, apresenta a consulta ao banco:

Quadro 19 – Consulta recurso recebido por estados, por IDH.

```

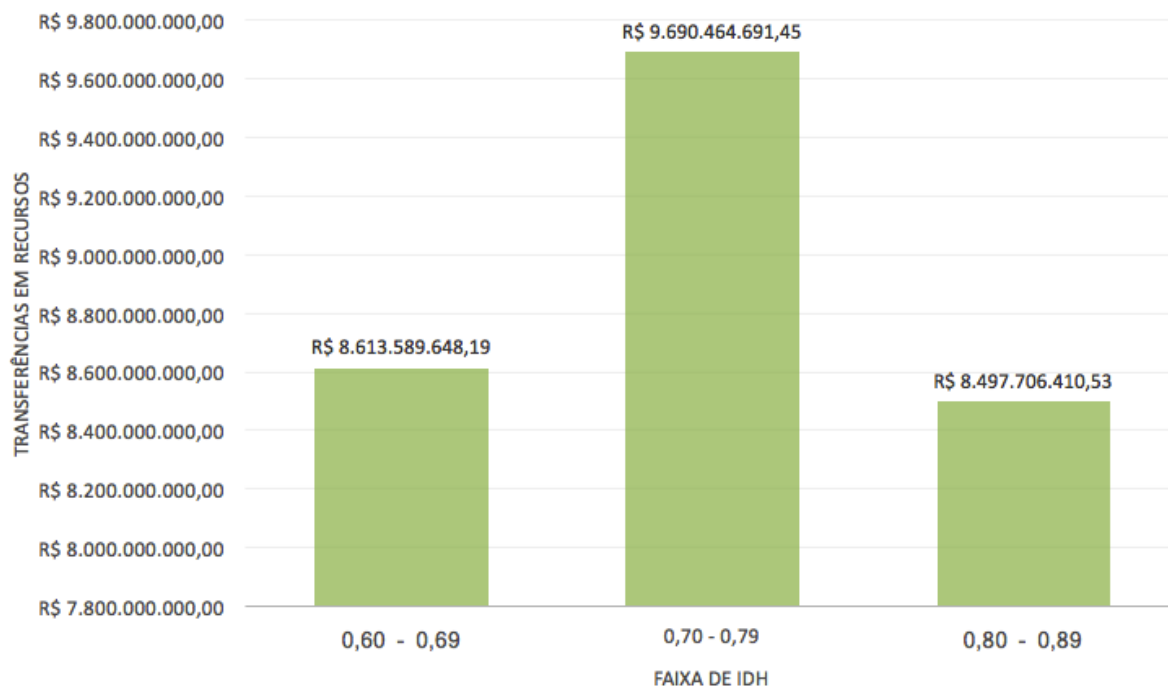
PREFIX : <http://lod.data.gov.br/onto/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT (SAMPLE(xsd:decimal(?idh)) AS ?idhUf) (SUM(xsd:decimal(?valor))
AS ?valorRecebido)
WHERE {
  ?ufEm2010
    a :FatiaDeTempoUf;
    :aconteceEm ?ano;
    :fatiaDeTempoDe ?uf;
    :idh ?idh;
    :recebe ?recurso.
  ?ano
    :ano "2010".
  ?uf
    :codIbge ?codIbge.
  ?recurso
    :valor ?valor.
}
GROUP BY ?codIbge

```

Fonte: Elaboração dos autores, 2016.

Com a consulta realizada anteriormente, os dados obtidos foram analisados a relação do recurso recebido e o IDH por UFs, chegando no resultado exposto na figura 32 exposta a seguir:

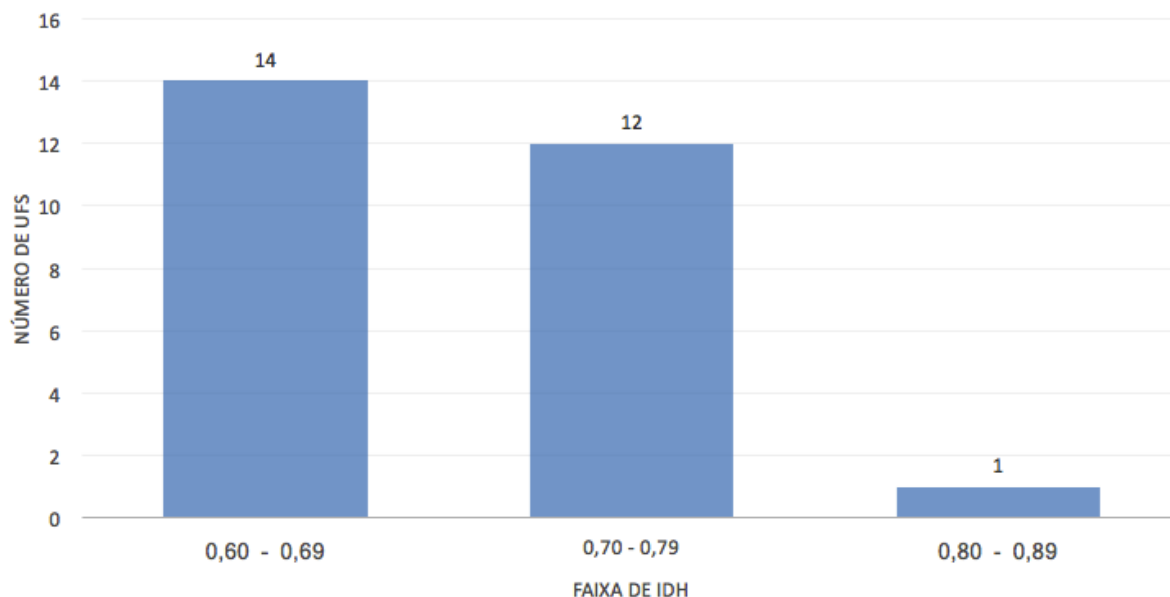
Figura 32 – Recurso recebido por UFs por faixa de IDH.



Fonte: Elaborado pelos autores, 2016.

Verifica-se que grande parte das unidades federativas brasileiras se encontram na faixa de IDH entre 0,70 e 0,79 recebendo o montante de R\$ 9.690.464.691,45 de recurso. Na figura 33 a seguir é possível verificar quantos estados encontram-se por faixa de IDH.

Figura 33 – Quantidade de UFs por faixa de IDH.



Fonte: Elaborado pelos autores, 2016.

Apesar da pouca diferença de recurso recebido entre as faixas de IDH por unidades federativas, é possível verificar que somente um estado recebeu um valor de recurso alto se comparado aos demais estados, em que no seu somatório tem praticamente o mesmo que uma única unidade federativa recebeu. A única unidade federativa na faixa de IDH 0,80 e 0,89, que neste caso trata-se do Distrito Federal recebeu praticamente o mesmo recurso de 14 estados que estão na faixa de IDH de 0,60 e 0,69.

A seguir constam as análises relacionadas a terceira pergunta elaborada, "Qual é a relação entre arrecadações geradas pelos estados e seu nível de IDH geral?". O quadro 20, abaixo, exhibe a consulta ao banco relacionada a pergunta citada:

Quadro 20 – Consulta arrecadação gerada por UFs e seu IDH.

```

PREFIX : <http://lod.data.gov.br/onto/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT (SAMPLE(xsd:decimal(?idh)) AS ?idhUf) (SUM(xsd:decimal(?valor)) AS
?valorArrecadado)
WHERE {
  ?ufEm2010
    a :FatiaDeTempoUf;
    :aconteceEm ?ano;
    :fatiaDeTempoDe ?uf;
    :idh ?idh;

```

```

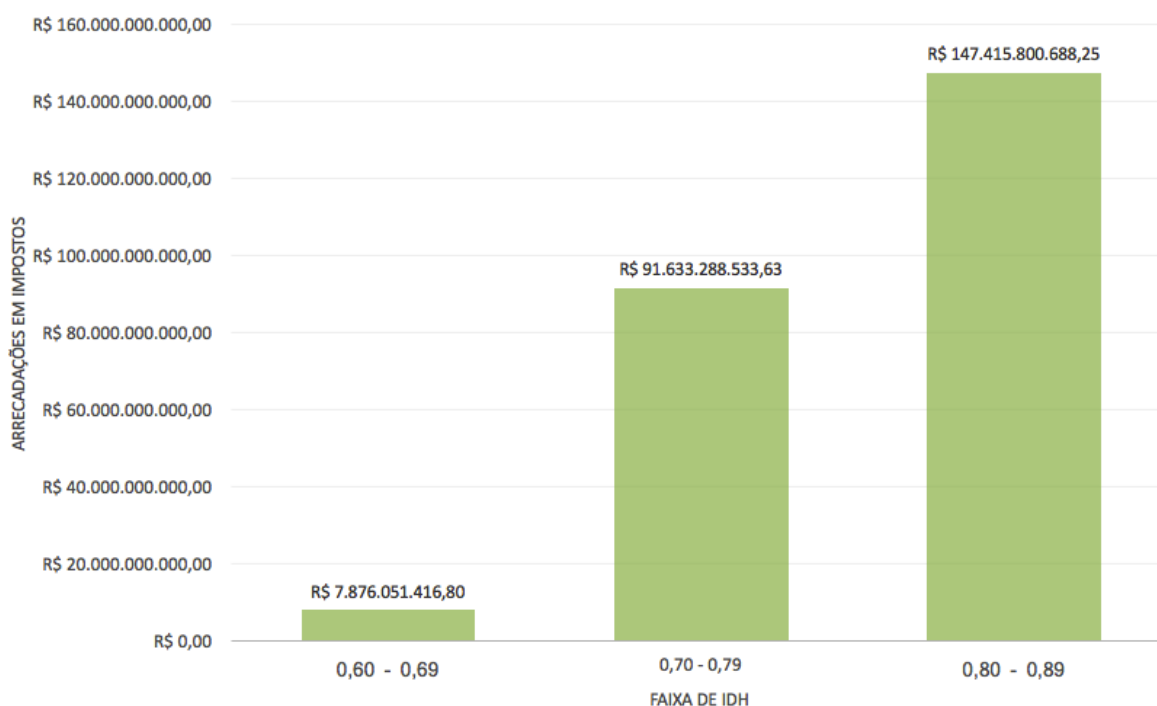
:gera ?arrecadacao.
?ano
:ano "2010".
?uf
:codIbge ?codIbge.
?arrecadacao
:valor ?valor.
}
GROUP BY ?codIbge

```

Fonte: Elaboração dos autores, 2016.

Após a realização da consulta ao banco exposta anteriormente no quadro 17, foi possível realizar análise dos dados extraídos, com os dados foi possível gerar os seguintes gráficos abaixo. A figura 34 apresenta a arrecadação gerada por UF e a sua respectiva faixa de IDH.

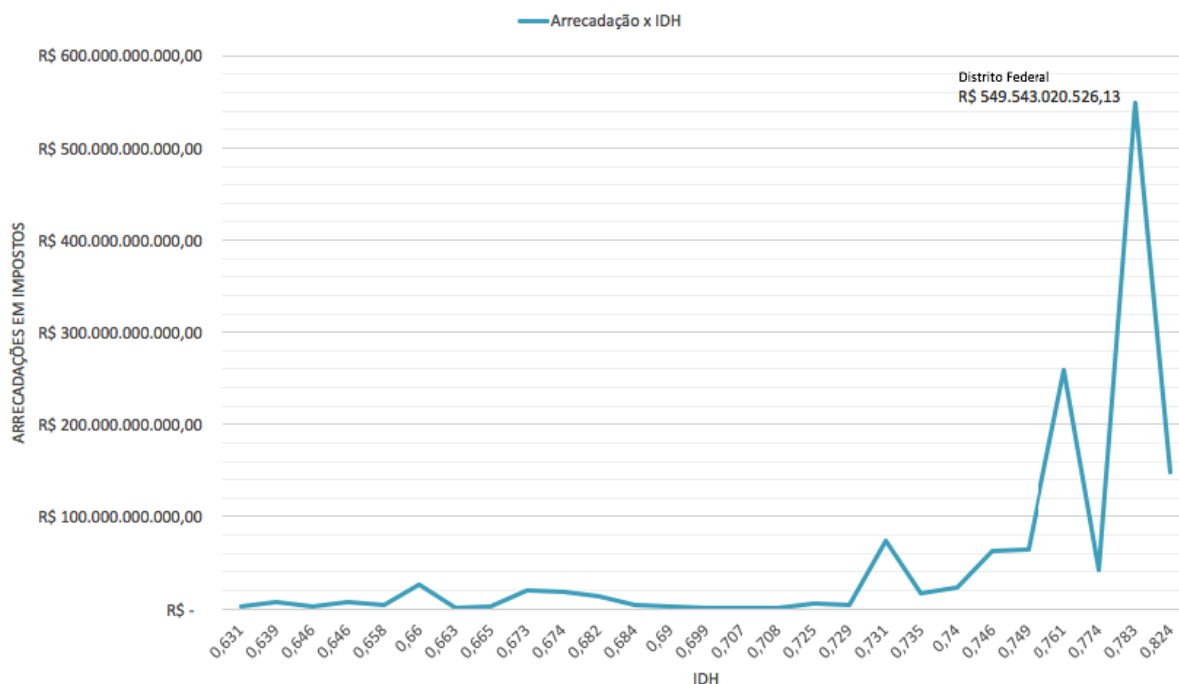
Figura 34 – Arrecadação gerada por UF e sua faixa de IDH.



Fonte: Elaborado pelos autores, 2016.

No gráfico, exibido na figura 35 a seguir é possível verificar a disparidade de valores arrecadados do Distrito Federal em relação aos demais estados.

Figura 35 – Arrecadação recebida por UF e seu nível de IDH.



Fonte: Elaborado pelos autores, 2016.

Além de ser a unidade federativa brasileira que possui o maior valor arrecadado, também é a que possui o maior valor de IDH, com o índice de 0,824, única a entrar na faixa de 0,80 e 0,89.

Na quarta questão elaborada, que tinha o objetivo de buscar qual é a relação entre recursos recebidos pelos estados e seu nível de IDH de educação, foi elaborada uma consulta SPARQL, o quadro 21 apresenta como realizamos a consulta ao banco para trazer os dados com relação a essa pergunta.

Quadro 21 – Consulta recurso recebido por UFs e seu IDH para educação.

```

PREFIX : <http://lod.data.gov.br/onto/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT (SAMPLE(xsd:decimal(?idhEducacao)) AS ?idhEducacaoUf)
(SUM(xsd:decimal(?valor)) AS ?valorRecebido)
WHERE {
  ?ufEm2010
    a :FatiaDeTempoUf;
    :aconteceEm ?ano;
    :fatiaDeTempoDe ?uf;
    :idhEducacao ?idhEducacao;
    :recebe ?recurso.

```

```

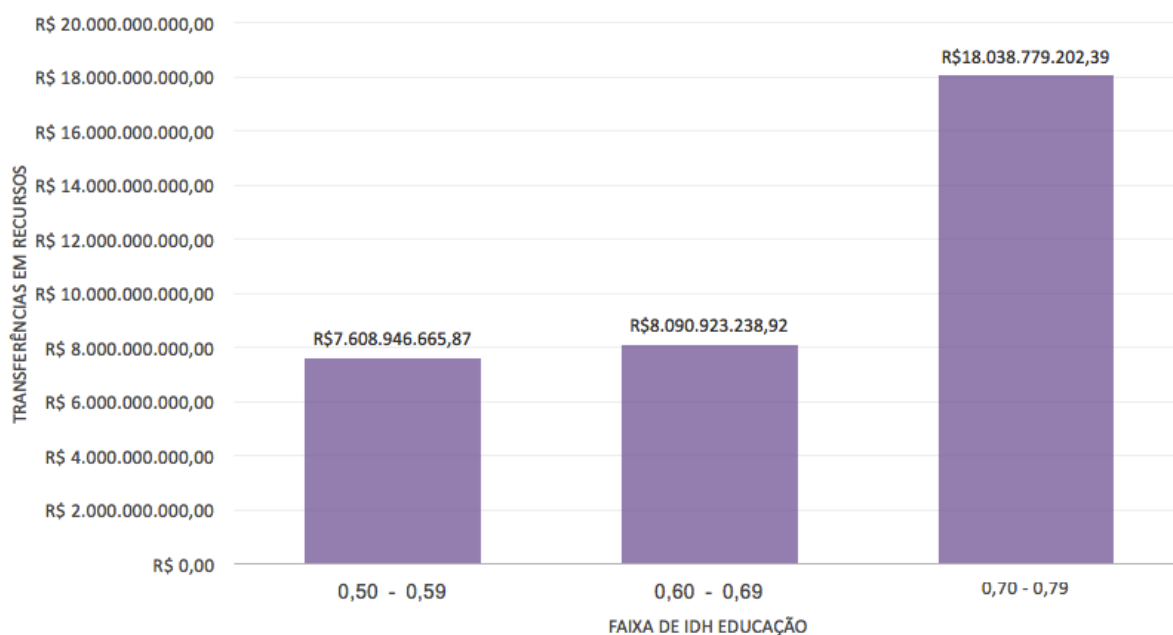
?ano
  :ano "2010".
?uf
  :codIbge ?codIbge.
?recurso
  :valor ?valor.
}
GROUP BY ?codIbge

```

Fonte: Elaboração dos autores, 2016.

Com o resultado obtido a partir da consulta executada foi possível extrair os dados necessários para gerar os gráficos apresentados a seguir, e realizar a análise sobre os dados, o gráfico da figura 36 apresenta o levantamento entre o recurso recebido pelas UFs e suas respectivas faixas de índice de desenvolvimento para educação.

Figura 36 – Recurso recebido por faixa de IDH de Educação.

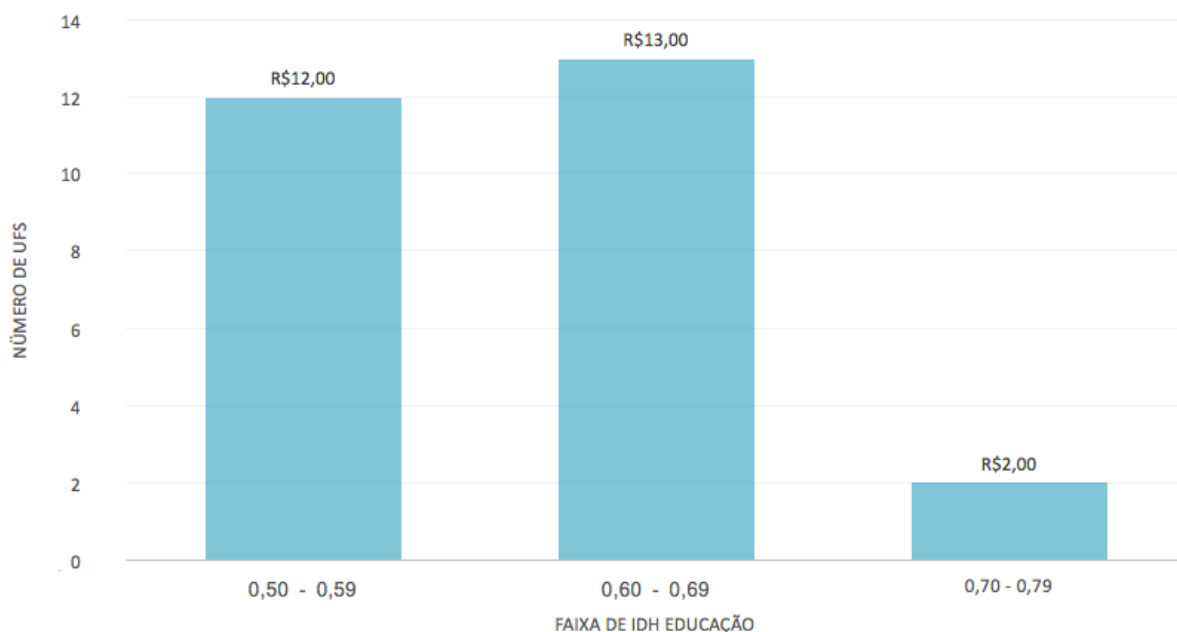


Fonte: Elaborado pelos autores, 2016.

A figura 36 expõe os índices das UFs para o IDH de educação e o seu recurso recebido, o maior índice no caso da educação fica na faixa entre 0,70 e 0,79 e possui o maior valor de recurso recebido. Para tirarmos alguma conclusão com relação as faixas e valores exibidos, precisamos entender quantos estados estão em cada faixa. A figura 37 mostra essa informação:



Figura 37 – Quantidade de UFs por faixa de IDH de Educação.



Fonte: Elaborado pelos autores, 2016.

Apenas duas unidades federativas estão na faixa de 0,70 e 0,79, enquanto os demais estados estão entre 0,50 e 0,69 quando se trata do IDH para o índice de educação, as unidades federativas que estão no último nível apresentado São Paulo (IDH 0,719 e recurso recebido R\$27.579.851.994,24) e o Distrito Federal (IDH 0,742 e recurso recebido R\$ 8.497.706.410,53), existem diversos outros estados que recebem mais recurso que o Distrito Federal, dentre eles o Rio de Janeiro (IDH educação 0,675 e recurso recebido de R\$ 20.679.553.070,17), mesmo o Rio de Janeiro recebendo o segundo maior valor de recurso encontra-se na faixa de 0,60 a 0,69.

A quinta pergunta elaborada, "Qual é a relação entre arrecadações geradas pelos estados e seu nível de IDH de educação?" Essa questão aborda em vez do recurso recebido, como apresentado anteriormente, mostra a arrecadação gerada pelas unidades federativas e seu nível de IDH. A seguir, no quadro 22, é apresentada a consulta ao banco para responder a pergunta:

Quadro 22 – Consulta arrecadação gerada por UFs e seu IDH para educação.

```

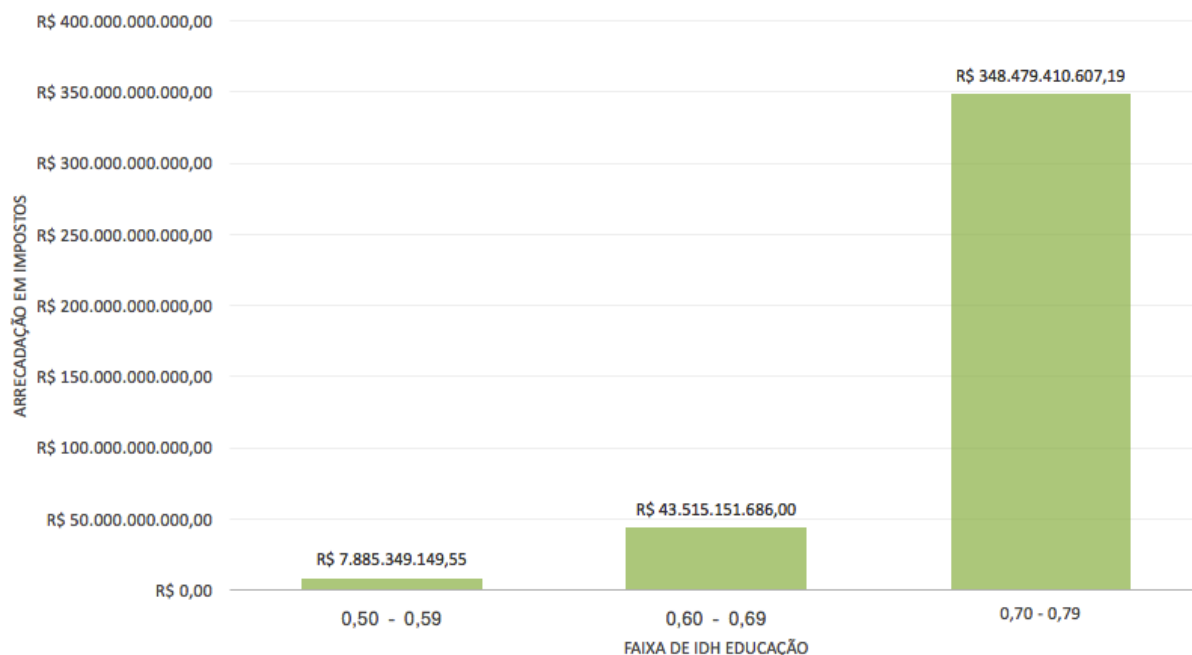
PREFIX : <http://lod.data.gov.br/onto/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT (SAMPLE(xsd:decimal(?idhEducacao)) AS ?idhEducacaoUf)
(SUM(xsd:decimal(?valor)) AS ?valorArrecadado)
WHERE {
  ?ufEm2010
    a :FatiaDeTempoUf;
    :aconteceEm ?ano;
    :fatiaDeTempoDe ?uf;
    :idhEducacao ?idhEducacao;
    :gera ?arrecadacao.
  ?ano
    :ano "2010".
  ?uf
    :codIbge ?codIbge.
  ?arrecadacao
    :valor ?valor.
}
GROUP BY ?codIbge

```

Fonte: Elaboração dos autores, 2016.

Após realizada a consulta SPARQL anteriormente, foi possível efetuar a geração com base nos dados extraídos, a figura 38 mostra a arrecadação gerada por UF somada por faixa de IDH considerando somente o índice de educação.

Figura 38 – Arrecadação gerada por UF e a faixa de IDH de educação.



Fonte: Elaborado pelos autores, 2016.

A sexta pergunta elaborada é tem a intenção de saber qual é a relação entre recursos recebidos pelos estados e seu nível de IDH de longevidade. Para responde-la geramos a consulta SPARQL representada no quadro 23.

**Quadro 23 – Consulta recursos recebidos por UFs e seu IDH para longevidade.**

```
PREFIX : <http://lod.data.gov.br/onto/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT (SAMPLE(xsd:decimal(?idhLongevidade)) AS ?idhLongevidadeUf)
(SUM(xsd:decimal(?valor)) AS ?valorRecebido)
WHERE {
  ?ufEm2010
    a :FatiaDeTempoUf;
    :aconteceEm ?ano;
    :fatiaDeTempoDe ?uf;
    :idhLongevidade ?idhLongevidade;
    :recebe ?recurso.
  ?ano
    :ano "2010".
  ?uf
    :codIbge ?codIbge.
  ?recurso
    :valor ?valor.
}
GROUP BY ?codIbge
```

Fonte: Elaboração dos autores, 2016.

A análise relacionada aos dados obtidos nessa questão serão analisados mais à frente num gráfico mais abrangente com outras perguntas.

A sétima questão elaborada: "Qual é a relação entre arrecadações geradas pelos estados e seu nível de IDH de longevidade?". Essa pergunta visa trazer as informações com relação as arrecadações e o nível de IDH de longevidade, para isto foi elaborada a consulta SPARQL representada no quadro 24.

**Quadro 24 – Consulta arrecadação gerada por UFs e seu IDH para longevidade.**

```
PREFIX : <http://lod.data.gov.br/onto/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT (SAMPLE(xsd:decimal(?idhLongevidade)) AS ?idhLongevidadeUf)
(SUM(xsd:decimal(?valor)) AS ?valorArrecadado)
WHERE {
  ?ufEm2010
    a :FatiaDeTempoUf;
    :aconteceEm ?ano;
    :fatiaDeTempoDe ?uf;
```

```

:ldhLongevidade ?ldhLongevidade;
:gera ?arrecadacao.
?ano
:ano "2010".
?uf
:codIbge ?codIbge.
?arrecadacao
:valor ?valor.
}
GROUP BY ?codIbge

```

Fonte: Elaboração dos autores, 2016.

A análise relacionada aos dados obtidos nessa questão serão analisados mais à frente num gráfico mais abrangente juntamente com os dados de outras perguntas.

A oitava pergunta elaborada busca qual é a relação entre os recursos recebidos pelos estados e seu nível de IDH de renda, para responder essa consulta foi elaborada a consulta SPARQL, representada no quadro 25, em cima da ontologia de domínio.

Quadro 25 – Consulta recurso recebido por UFs e seu IDH para renda.

```

PREFIX : <http://lod.data.gov.br/onto/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT (SAMPLE(xsd:decimal(?ldhRenda)) AS ?ldhRendaUf)
(SUM(xsd:decimal(?valor)) AS ?valorRecebido)
WHERE {
  ?ufEm2010
    a :FatiaDeTempoUf;
    :aconteceEm ?ano;
    :fatiaDeTempoDe ?uf;
    :ldhRenda ?ldhRenda;
    :recebe ?recurso.
  ?ano
    :ano "2010".
  ?uf
    :codIbge ?codIbge.
  ?recurso
    :valor ?valor.
}
GROUP BY ?codIbge

```

Fonte: Elaboração dos autores, 2016.

A análise relacionada aos dados extraídos, com relação a oitava pergunta serão analisados mais à frente num gráfico mais abrangente junto com demais perguntas.

A nona questão elaborada busca encontrar a relação entre as arrecadações geradas pelas unidades federativas e seus níveis de IDH de renda. Para consultar esses dados elaboramos a seguinte consulta no banco exposta no quadro 26.

Quadro 26 – Consulta recurso recebido por UFs e seu IDH para renda.

```

PREFIX : <http://lod.data.gov.br/onto/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT (SAMPLE(xsd:decimal(?idhRenda)) AS ?idRendaUf)
(SUM(xsd:decimal(?valor)) AS ?valorArrecadado)
WHERE {
  ?ufEm2010
    a :FatiaDeTempoUf;
    :aconteceEm ?ano;
    :fatiaDeTempoDe ?uf;
    :idhRenda ?idhRenda;
    :gera ?arrecadacao.
  ?ano
    :ano "2010".
  ?uf
    :codIbge ?codIbge.
  ?arrecadacao
    :valor ?valor.
}
GROUP BY ?codIbge

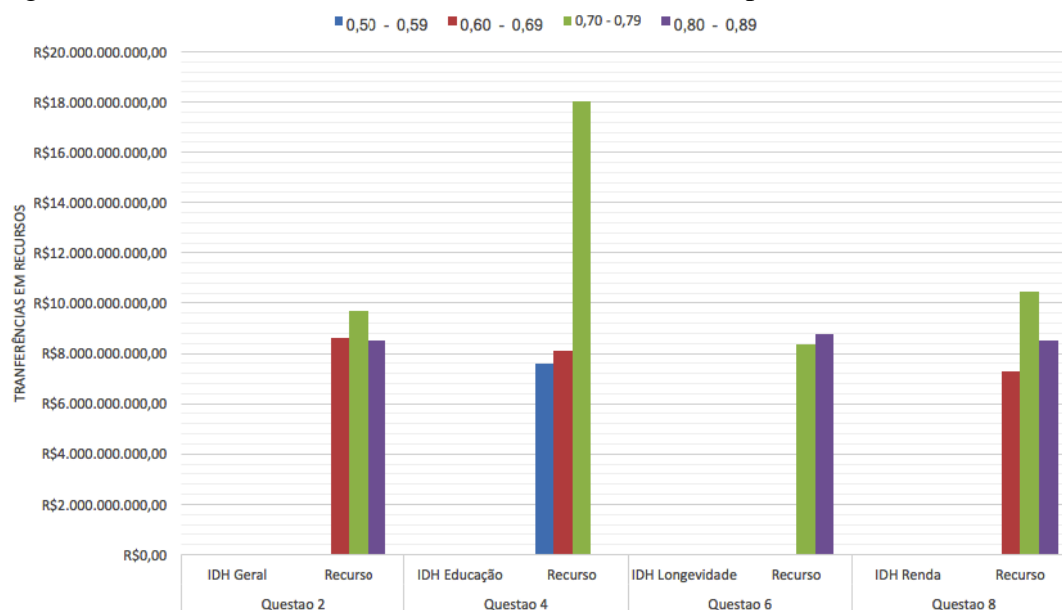
```

Fonte: Elaboração dos autores, 2016.

As observações com relação a pergunta anterior serão realizadas mais à frente, juntamente com a análise de outras perguntas.

Para uma análise comparativa dos dados extraídos as questões 2, 4, 6 e 8 das perguntas elaboradas serão exibidas num único gráfico na figura 39 a seguir.

Figura 39 – Recurso recebido x diferentes índices do IDH por UFs.



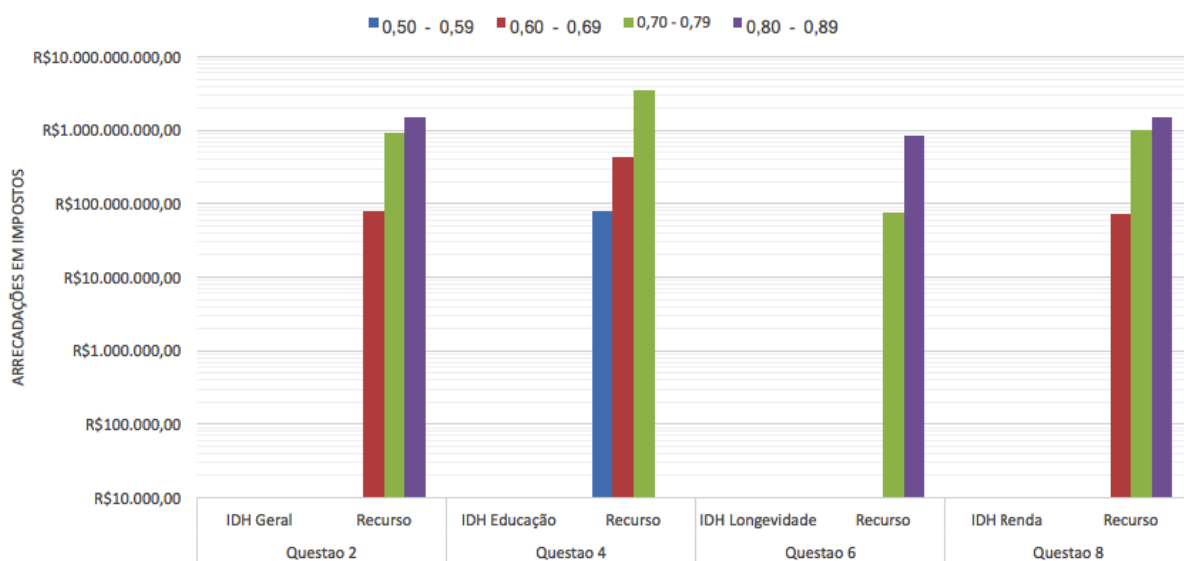
Fonte: Elaborado pelos autores, 2016.

Com a realização da análise em cima dos dados extraídos das questões 2, 4, 6 e 8, é possível verificar que, o índice que possui uma maior incidência em todos os IDHs analisados é a faixa de 0,70 a 0,79, além disso para os diversos índices de IDH analisados, o índice que possui a maior diferença de recurso recebido está no nível de IDH de educação, também é o único índice onde a faixa de IDH chega no nível de 0,50 a 0,59, todos os demais índices analisados não atingem essa faixa.

É possível verificar também que o IDH de longevidade possui a menor diferença entre as unidades federativas, considerando que possui a menor quantidade de faixas de IDH se comparado as demais faixas.

Por fim será realizado análise em cima dos dados extraídos a partir das questões 3, 5, 7 e 9 ao qual as consultas realizadas no banco para extrair os dados necessários já foram realizadas. A comparação entre os dados obtidos em todas as consultas é exibida na figura 40 a seguir.

Figura 40 – Arrecadação gerada x diferentes índices do IDH por UFs.



Fonte: Elaborado pelos autores, 2016.

No gráfico apresentado anteriormente, na figura 40, é exibida a relação das arrecadações geradas pelas unidades federativas e as faixas dos níveis de IDH. Em todos os níveis de IDH é possível verificar que as menores faixas são as que no seu somatório possuem

a menor arrecadação. Como também as maiores faixas de cada índice é que possuem a maior arrecadação registrada no somatório das unidades federativas.

A décima pergunta elaborada visa analisar qual eram os 20 municípios que mais receberam recursos em 2010 e seus respectivos IDHs. A seguir no quadro 27 é apresentado a consulta realizada para obter os dados necessários para a análise.

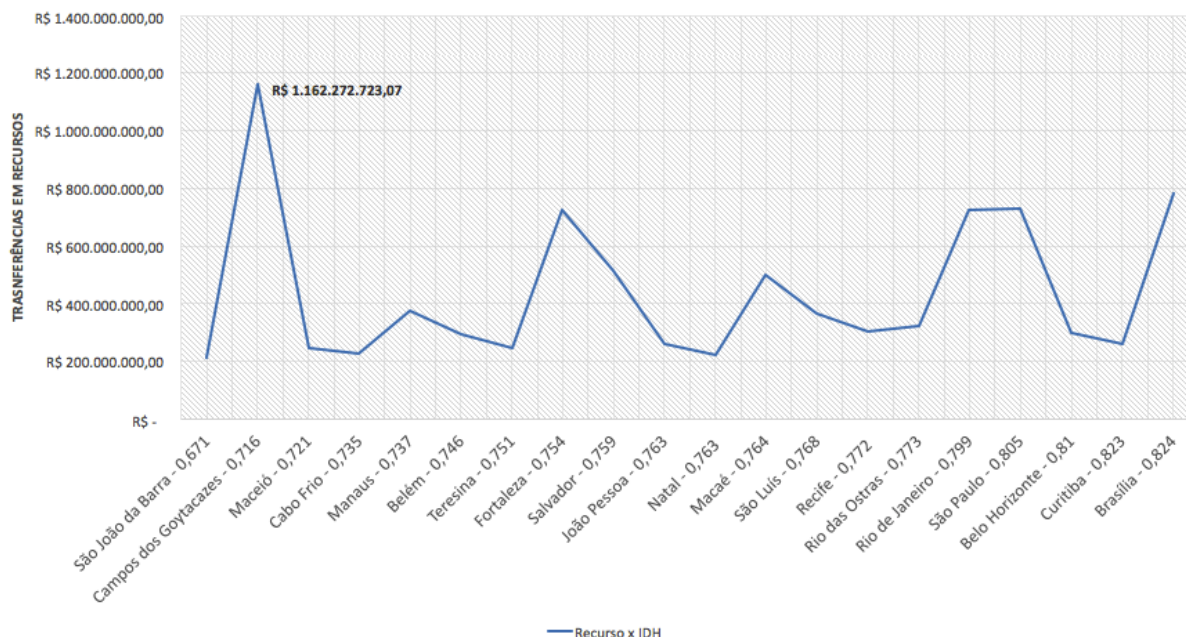
**Quadro 27 – Consulta recurso recebido por Municípios e seus IDHs.**

```
PREFIX : <http://lod.data.gov.br/onto/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
SELECT (SAMPLE(?nome) AS ?nomeMun) (SAMPLE(?nomeUf) AS ?nomeUfMun)
(SAMPLE(xsd:decimal(?idh)) AS ?idhMun) (SUM(xsd:decimal(?valor)) AS
?valorRecebido)
WHERE {
  ?municipioEm2010
    a :FatiaDeTempoMunicipio;
    :aconteceEm ?ano;
    :fatiaDeTempoDe ?municipio;
    :parteDeUf ?ufEm2010;
    :idh ?idh;
    :recebe ?recurso.
  ?ano
    :ano "2010".
  ?municipio
    :codIbge ?codIbge;
    :nome ?nome.
  ?ufEm2010
    :fatiaDeTempoDe ?uf.
  ?uf
    :nome ?nomeUf.
  ?recurso
    :valor ?valor.
}
GROUP BY ?codIbge
ORDER BY DESC(?valorRecebido)
LIMIT 20
```

Fonte: Elaboração dos autores, 2016.

Após a realização da consulta SPARQL realizada no quadro 27, obtemos os dados necessários para gerar o gráfico da figura 41:

Figura 41 – Os 20 municípios que mais receberam recursos e seus IDHs.



Fonte: Elaborado pelos autores, 2016.

É possível verificar com a figura 40 que nenhum dos 20 municípios que mais receberam recurso fica com o IDH abaixo do nível 0,671 (município São João da Barra - RJ), sendo que do total de 5.556 municípios brasileiros mais da metade dos municípios brasileiros não atingem esse nível de IDH.

Dentre os 20 municípios levantados, verificamos que 30% dos municípios que mais receberam recursos são de estado do Rio de Janeiro, sendo que o município que mais recebeu recurso foi Campos dos Goytacazes - RJ, que apesar de ter uma quantidade bem menor na sua população (356.608 de habitantes) se comparado a grandes cidades como São Paulo (11.032.000 de habitantes) e, recebeu a maior quantidade de recurso no ano de 2010 de R\$ 1.162.272.723,07.

Na seção seguinte será mostrado a avaliação realizada para a validação da proposta de solução deste trabalho.



### 5.3 AVALIAÇÃO

Para avaliação da proposta de solução deste trabalho, foi elaborado um questionário via ferramenta do Google Forms, as questões abordavam desde o conhecimento do entrevistado com relação aos dados abertos, até a avaliação dos resultados da estrutura utilizada no projeto deste trabalho, para isso foram elaboradas as seguintes perguntas:

- Qual seu nome?
- Qual sua idade?
- Você é professor de algum curso de tecnologia da informação? (Ex. Sistemas de Informação, Ciências da Computação e afins)
- Qual é o seu cargo atual?
- Há quanto tempo você trabalha com tecnologia?
- Você é a favor da disponibilização dos dados abertos do governo?
- Você já precisou acessar algum tipo de dado governamental aberto?
- No seu ponto de vista qual é o grau de utilidade que a disponibilização dos dados abertos pode trazer para a sociedade?
- Você considera que a estrutura semântica que foi desenvolvida neste trabalho, pode auxiliar no consumo e integração das informações?
- Na sua opinião, seria importante que todos os dados abertos fossem entregues também em padrões similares ao proposto neste trabalho?
- A possibilidade de consultas baseadas em SPARQL podem auxiliar nas inferências necessárias sobre os dados do domínio?
- Qual é a sua opinião sobre a ligação e disponibilização dos dados abertos governamentais?

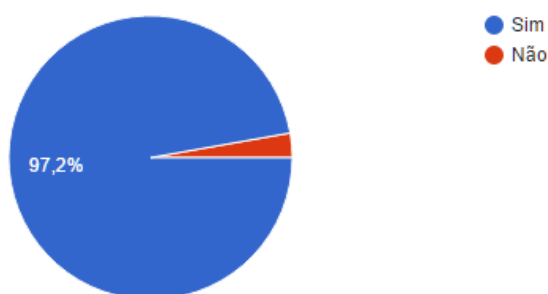
O questionário possui no total 12 perguntas, sendo que as cinco primeiras perguntas são do âmbito pessoal do entrevistado, para mapearmos o perfil do público do questionário.

Foram no total 36 entrevistados, todos eles possuem algum tipo de relação com a área de tecnologia, seja por experiência profissional, formados na área, ou por estarem estudando em cursos de tecnologia, dentre estes, 11,1% dos entrevistados são professores da área da tecnologia. A faixa etária dos entrevistados ficou na média de 30,19 anos.

A sexta questão abordava um dos temas deste trabalho, perguntando a opinião do entrevistado sobre a disponibilização dos dados abertos governamentais, o resultado é ilustrado na figura 42.

Figura 42 – Respostas da sexta pergunta.

Você é a favor da disponibilização dos dados abertos do governo? (36 respostas)



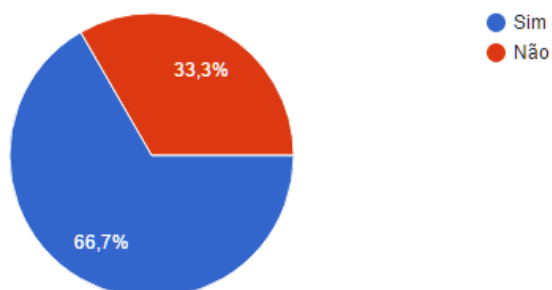
Fonte: Elaborado pelos autores, 2016.

Como é possível verificar a grande maioria respondeu que é a favor da disponibilização dos dados abertos governamentais, apenas 2,8% dos entrevistados se mostrou contrário a disponibilização de dados abertos do governo.

Na sétima questão a intenção era identificar se o entrevistado já havia tido alguma experiência com dados abertos do âmbito governamental, necessitando da sua utilização, o resultado é ilustrado na figura 43.

Figura 43 – Respostas da sétima pergunta.

Você já precisou acessar algum tipo de dado governamental aberto?  
(36 respostas)



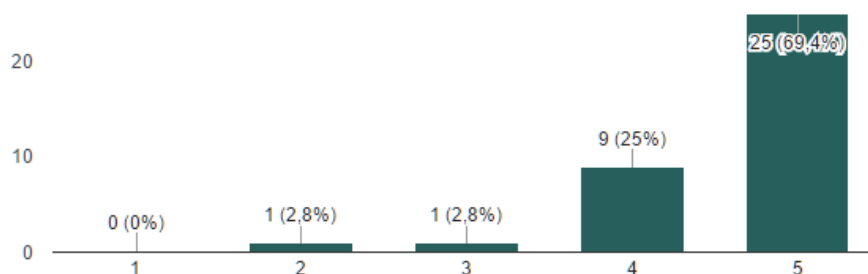
Fonte: Elaborado pelos autores, 2016.

Conforme imagem anterior, é possível verificar que dentre as respostas 66,7% dos entrevistados já necessitou do uso de algum dado governamental aberto e 33,3% dos entrevistados disseram nunca precisar acessar algum tipo de dados governamental.

A oitava pergunta, apresentada na figura 44, visa conhecer qual é a opinião do entrevistado com relação a utilidade da disponibilização dos dados abertos.

Figura 44 – Respostas da oitava pergunta.

No seu ponto de vista qual é o grau de utilidade que a disponibilização dos dados abertos pode trazer para a sociedade?  
(36 respostas)



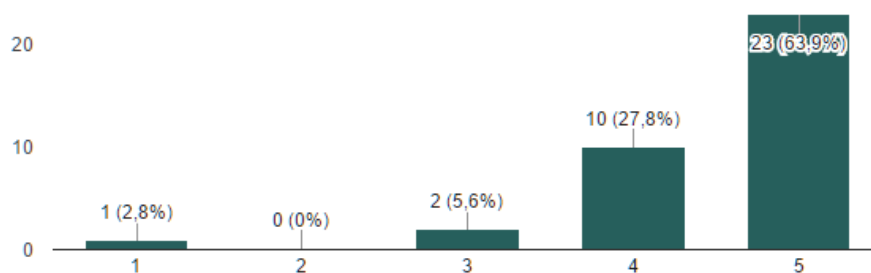
Fonte: Elaborado pelos autores, 2016.

De acordo com a figura 44 a maioria dos entrevistados afirmam que a disponibilização dos dados para a sociedade é de útil 25% (classificação 4) e 69,4% dos entrevistados consideram a disponibilização dos dados muito útil para a sociedade (classificação 5).

A nona questão era relacionada a proposta de solução apresentada neste trabalho, após a apresentação do tema ao entrevistado, o mesmo deveria responder se a estrutura desenvolvida utilizando os dados abertos do governo brasileiro traria uma melhoria no consumo e integração dessas informações, o resultado é ilustrado na figura 45.

Figura 45 – Respostas da nona pergunta.

Você considera que a estrutura semântica que foi desenvolvida neste trabalho, pode auxiliar no consumo e integração das informações?  
(36 respostas)



Fonte: Elaborado pelos autores, 2016.

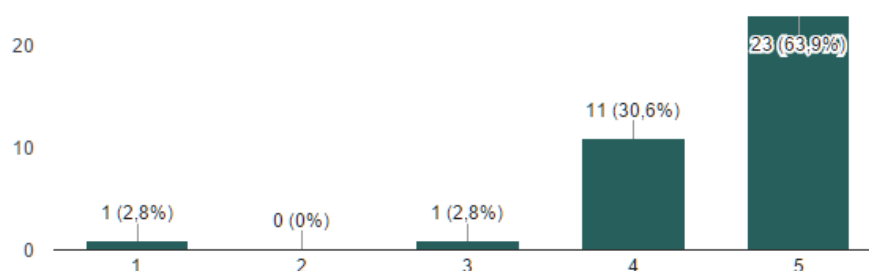
Segundo a resposta apresentada na figura 45, 91,7% dos entrevistados responderam que concordam parcialmente (27,8%), ou plenamente (63,9%), considerando então que a semântica deste trabalho pode auxiliar no consumo e integração das informações coletadas.

Para a décima pergunta a intenção era saber se o padrão proposto para os dados abertos era satisfatório para o entrevistado, considerando a importância de todos os dados abertos serem disponibilizados nesse formato, o resultado é ilustrado na figura 46.

Figura 46 – Respostas da décima pergunta.

Na sua opinião, seria importante que todos os dados abertos fossem entregues também em padrões similares ao proposto neste trabalho?

(36 respostas)



Fonte: Elaborado pelos autores, 2016.

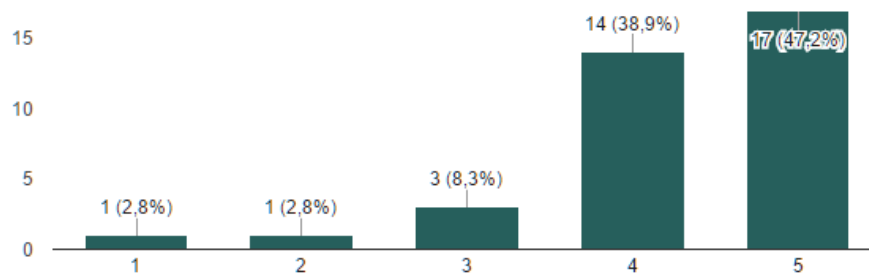
Para 63,9% dos entrevistados concordam plenamente sobre a importância de os dados abertos serem entregues nos padrões similares a proposta deste trabalho. Já 30,6% concordam parcialmente com relação a importância de os dados serem publicados no formato de dados conectados.

A décima primeira questão abordava sobre as consultas SPARQL que foram mostradas neste trabalho, questionando se elas poderiam auxiliar nas inferências necessárias para os dados obtidos, o resultado é ilustrado na figura 47.

Figura 47 – Respostas da décima primeira pergunta.

A possibilidade de consultas baseadas em SPARQL podem auxiliar nas inferências necessárias sobre os dados do domínio?

(36 respostas)



Fonte: Elaborado pelos autores, 2016.

É possível verificar na imagem anterior que a maioria dos entrevistados, 86,1% (entre notas 4 e 5) considera que a possibilidade das consultas em SPARQL podem auxiliar nas inferências dos dados utilizados neste trabalho.

Por fim a décima segunda, e última pergunta do questionário era aberta para os entrevistados darem a sua opinião a respeito da ligação e disponibilização dos dados abertos governamentais, as respostas são ilustradas na figura 48.

Figura 48 – Respostas da décima segunda pergunta.

Os dados abertos podem auxiliar muito em alguns aspectos. Tem alguns riscos tbm da utilização desses dados por pessoas mal intencionadas.
Extremamente útil para a sociedade, pois os dados coletados seriam reais e não apenas informações sensacionalistas que geralmente são massificadas pela mídia.
Maior transparência sobre a forma que os políticos administram nossos gastos públicos.
Acredito que a possibilidade de armazenamento dos dados governamentais em uma base única facilitaria a análise quantitativa e qualitativa dos dados, possibilitando a geração e explicitação de conhecimento.
Nao deveria
Dados abertos é sinônimo de transparência
Muito bom.
Quanto mais fácil o acesso às informações e mais simples a organização das mesmas, melhor será a compreensão pelas pessoas comuns. Isso é muito importante para que todos os brasileiros analisem, através de dados reais e de fácil compreensão, onde são usados nossos recursos.
Seria muito útil para dar a transparência que muitas vezes não fica evidente nos meios de comuniação
É muito importante não só porque facilita a análise, mas pricipalmente por possibilitar o acesso de maneira mais objetiva. Assim, quanto mais números se tem sobre um problema, mais soluções podem ser apresentadas. A análise sobre uma leitura precisa das informações do governo, pode nos dizer muito sobre a realidade da gestão dos recursos e a política de aplicação desses recursos. Isso remete a uma reflexão sobre todo o sistema e o papel dos poderes que influenciam, o modo que estes são geridos. O primeiro passo, é conhecer a realidade dos fatos, posteriormente mudar tudo o que nao t
Agilizaria a pesquisa de informações entre órgãos.
Interessante
Importante, para uma maior accountability e planejamento da gestão pública.
Penso que é de extrema importância porque pode fornecer muito auxilio para novos negócios que estão surgindo no mercado e que podem utilizar estas informações
Acho que a informação é um ativo a ser explorado com muito mais atenção e o governo possui várias informações que impactam a vida das pessoas é os negócios a serem criados. Dados abertos governamentais são de fato uma forma de democratização da informação pública
Concordo que deve ser disponibilizado para estudos e estratégias.
Competitividade entre fornecedores, transparência para comunidade e comparações entre serviços realizados.
Garantindo uma boa interoperabilidade entre os dados governamentais e sistemas de gestão, essa ligação e disponibilização pode trazer inúmeros benefícios para o cidadão.
Maior transparência para a população, permitindo que qualquer um fiscalize ações que possam interferir na sociedade.
Acredito que o aprofundamento na estruturação e disponibilização desses dados pode contribuir com a construção bases de conhecimento importantes tanto para estudos da área acadêmica quanto para a inferência de novas informações relevantes até mesmo para gestão pública mais eficaz.

Fonte: Elaborado pelos autores, 2016.

Como é mostrado na figura 48, e nas demais respostas do questionário, a maioria das pessoas que colocaram suas opiniões a respeito da disponibilização dos dados abertos governamentais conectados expuseram comentários concordando com essa maneira de publicação dos dados, encontrando diversas vantagens e maneiras de utilização, para a população, para o próprio governo e até para administração de empresas.

Com a aplicação do questionário foi possível concluir que, para a maioria dos entrevistados a conexão dos dados abertos governamentais se faz útil e necessário para utilização em diversas esferas organizacionais.

#### 5.4 TRABALHOS CORRELATOS

Os principais trabalhos relacionados encontrados são o projeto MUN.LodKEM<sup>14</sup>, o projeto Um modelo para integração de informações de bases de dados abertos, com uso de ontologias (TOSIN, 2016), o projeto Aplicando Linked Data na Publicação de Dados do ENEM (CABRAL et al., 2012) e o projeto Aumentando a Transparência do Governo por Meio da Transformação de Dados Governamentais Abertos em Dados Ligados (ARAÚJO; SOUZA, 2011).

Este trabalho teve como propósito a obtenção de dados abertos governamentais e transformação desses em dados conectados, de forma a facilitar a análise e consulta por pesquisadores e máquinas.

Um projeto semelhante ao proposto que envolve a publicação de dados de municípios no formato de dados conectados, é o MUN.LodKEM. Este projeto foi utilizado para coletar dados relativos a cidades e suas ligações com *datasets* externos, os recursos de cidades presentes neste trabalho são conectados com o *dataset* do projeto MUN.LodKEM.

Outros trabalhos como o de Cabral et al. (2012), propõe a publicação e uso de dados conectados do senso escolar brasileiro, utilizando dados do Exame Nacional do Ensino Médio (ENEM), já o projeto de Araújo e Souza (2011), propõe práticas de dados conectados para criar um conjunto de dados abertos de políticos brasileiros com informações coletadas de diferentes fontes.

---

<sup>14</sup> Disponível em: <http://lodkem.led.ufsc.br/municipios/>



O trabalho Tosin (2016), apresenta o desenvolvimento de recursos para inferir e relacionar informações no contexto de aplicações Web voltadas para a integração de bases de dados abertos e conectados, utilizando dados das compras governamentais que estavam armazenados em bases de dados relacionais, com o uso de uma ontologia, desenvolvida especificamente para o domínio dos dados.

A ontologia desenvolvida neste trabalho foi inspirada na ontologia empresarial dinâmica apresentada no trabalho de Batsakis e Petrakis (2010), que aborda temas como representação espaço-temporal, raciocínio e consultas na *Web Semântica*.

## 6 CONCLUSÕES E TRABALHOS FUTUROS

Neste capítulo são apresentadas, a partir dos resultados obtidos e experiências adquiridas, as conclusões acerca do projeto de solução realizado ao longo deste trabalho, também será apresentado as propostas para trabalhos futuros que poderão ser executados em continuidade deste projeto.

### 6.1 CONCLUSÕES

Através deste trabalho de conclusão de curso foram identificados diversos problemas com relação aos dados aberto governamentais brasileiros, problemas de publicação dos dados, falta de formatação e padronização nos tipos de formatos disponibilizados (*csv*, *html*, *json* e outros).

Como uma alternativa para resolver este problema e trazer um maior significado para os dados governamentais abertos, foi realizado a construção de uma base de dados conectados, baseada em diversos dados disponibilizados através de portais de dados abertos de esferas governamentais brasileiras, de forma que estes dados pudessem então gerar algum tipo informação.

A forma em que os dados são publicados, grande parte sem relação alguma com outros dados, sem identificadores únicos para elementos como município, e o fato dos portais de dados abertos governamentais brasileiros não disponibilizarem os dados com ligações entre esses, acaba trazendo uma grande dificuldade tanto para busca manual quanto por máquinas, dificultando a análise que envolva diferentes domínios de dados, pois a recuperação, avaliação e aproveitamentos dos dados torna-se mais demorada e maçante.

Com os diversos problemas elencados anteriormente, o objetivo geral deste trabalho é a criação de uma base de conhecimento baseada em ontologia, com dados abertos conectados, utilizando diversos dados disponibilizados através de portais de dados abertos de esferas governamentais brasileiras.

Para que o objetivo geral ser alcançado foram estabelecidos alguns objetivos específicos. O primeiro objetivo era a obtenção dos dados nos portais brasileiros de dados abertos, este foi realizado coletando arquivos com dados de diversos portais, dados de

municípios, estados, dados de IDH, suas receitas geradas e recursos recebidos, entre outros dados do ano de 2010, 2000 e 1991.

O segundo objetivo específico era desenvolver uma ontologia para o domínio dos dados coletados, após a coleta dos dados, foi realizada uma análise a partir de cada arquivo coletado, de forma a rastrear todas as características necessárias desses arquivos para o desenvolvimento da ontologia, que foi desenvolvida seguindo a metodologia escolhida, a 101.

Concluindo a ontologia de domínio, o terceiro objetivo específico visava a conversão do conjunto de dados em dados conectados. Tomando como base a ontologia desenvolvida, os dados foram convertidos, persistidos, e disponibilizados localmente, seguindo os quatro princípios dos dados conectados: uso de URI para identificar recursos, uso de URIs via HTTP para identificar os recursos na *Internet*, disponibilização de informações úteis dos recursos seguindo padrões RDF e SPARQL, ligação dos dados a outras bases de dados conectados.

O quarto objetivo específico era a realização de uma avaliação com base em experimentos a partir do domínio definido. Para a realização dessa avaliação foram elaboradas algumas perguntas com a finalidade de consultá-las no banco. Com a realização das consultas foi feita uma análise qualitativa em cima dos dados levantados, de forma que estes dados, agora conectados, gerassem o conhecimento que outrora, não seria possível de uma maneira simples, considerando o formato original em que estes dados se encontravam.

A segunda parte da avaliação, e último objetivo específico deste trabalho, foi a realização de um questionário avaliativo com a comunidade. O trabalho foi apresentado para pessoas da área da tecnologia (estudantes, profissionais de T.I e professores) para que estes avaliassem qual era a relevância que este trabalho proporcionava. Houve um posicionamento muito positivo com as respostas dos entrevistados, e comentários bem produtivos com relação ao benefício que a conexão dos dados abertos pode trazer para a sociedade.

Para os autores, toda a arquitetura realizada nos dados coletados trouxe um resultado muito positivo e perceptivo, a ontologia representou bem o domínio de dados coletados, e ajudou na persistência e recuperação dos recursos e relações entre eles.

A disponibilização de um servidor para consulta aos dados, facilitou o acesso a análise tanto para humanos quanto para máquinas, garantindo assim a interoperabilidade entre sistemas. Não há mais a necessidade de baixar arquivos de dados e relacioná-los, o que era uma

tarefa difícil, pois esses são disponibilizados em um formato que dificulta qualquer tipo de automação.

Após analisar o problema de pesquisa apresentado neste trabalho, a proposta de solução, resultados obtidos através da avaliação dos experimentos com base no domínio de dados, e a avaliação feita com pessoas da área de tecnologia da informação, conclui-se que a proposta de utilização de uma base de conhecimento baseada em ontologia, utilizando dados abertos conectados é válida, podendo também ser utilizada em outros cenários e não apenas no apresentado neste trabalho.

A elaboração deste trabalho proporcionou uma oportunidade única de desenvolvimento técnico e pessoal aos autores do mesmo. O aprendizado obtido com os novos conhecimentos adquiridos, possibilitou uma nova visão dos dados em relação ao novo mundo, movido pelo consumo dos dados em massa.

## 6.2 TRABALHOS FUTUROS

Como trata-se de um projeto de solução é possível expandir esse projeto em diversas formas. Durante o desenvolvimento deste trabalho houveram diversas dificuldades em torno dos dados utilizados e da transformação e persistência dos dados em formato de dados conectados. É notável que o processo de triplificação deve ser automático, deste modo podemos propor como trabalho futuro a automação na conversão dos dados e/ou uma ferramenta triplificadora para organizações governamentais responsáveis por disponibilizar os dados abertos na *Internet*.

Pode-se e deve-se trabalhar com uma gama maior de diferentes áreas dos dados governamentais abertos, uma vez que este trabalho focou principalmente na parte dos dados de administração de municípios, estados e país, concentrando a análise em alguns atributos específicos como IDH, recurso e arrecadação, uma vez expandido esse domínio de dados, poderá ser adquirido um conhecimento muito maior a respeito dos dados abertos governamentais disponibilizados pelos portais de dados do governo brasileiro.

Outra possibilidade de trabalho futuro é a expansão da ontologia de domínio e seus dados relacionados, poderá também ser abordado juntamente com os conceitos de big data e suas tecnologias envolvidas.

Este trabalho não teve como objetivo realizar uma análise aprofundada nos dados obtidos, outra opção de continuidade seria realizar um aprofundamento dos estudos sociais e seus impactos com relação aos dados coletados.

## REFERÊNCIAS

- ABBURU, Sunitha. A survey on ontology reasoners and comparison. **International Journal of Computer Applications**, v. 57, n. 17, p. 33-39, 2012. Disponível em: <<http://research.ijcaonline.org/volume57/number17/pxc3883748.pdf>>. Acesso em: 22 out. 2015.
- ANTONIOU, Grigoris; VAN HARMELEN, Frank. Web ontology language: Owl. In: **Handbook on ontologies**. Springer Berlin Heidelberg, 2004. p. 67-92.
- ARAÚJO, Lucas de Ramos; SOUZA, Jairo Francisco de. Aumentando a Transparência do Governo por Meio da Transformação de Dados Governamentais Abertos em Dados Ligados. **Revista Eletrônica de Sistemas de Informação**, v. 10, n. 1, 2011.
- AUDY, Jorge Luis Nicolas; ANDRADE, Gilberto Keller de; CIDRAL, Alexandre. **Fundamentos de sistemas de informação**. São Paulo, Artmed, 2005.
- BAADER, Franz; HORROCKS, Ian; SATTler, Ulrike. Description logics. In: **Handbook on ontologies**. Springer Berlin Heidelberg, 2009. p. 21-43.
- BATSAKIS, Sotiris; PETRAKIS, Euripides GM. SOWL: spatio-temporal representation, reasoning and querying over the semantic web. In: **Proceedings of the 6th International Conference on Semantic Systems**. ACM, 2010. p. 15.
- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web: Scientific American. **Scientific American**, maio 2001. Disponível em: <<http://www.scientificamerican.com/article/the-semantic-web>>. Acesso em: 22 out. 2015.
- BERNERS-LEE, Tim; FIELDING, R.; MASINTER, Larry. RFC 3986. **Uniform resource identifier (URI): Generic syntax**, 2005. Disponível em: <<http://www.ietf.org/rfc/rfc3986.txt>>. Acesso em: 22 out. 2015.
- BERNERS-LEE, Tim. **Linked data**. 2006. Disponível em: <<http://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: 08 set. 2015.
- \_\_\_\_\_. **Putting Government Data online**. 2009. Disponível em: <<http://www.w3.org/DesignIssues/GovData.html>>. Acesso em: 05 set. 2015.
- BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked data – the story so far. **International Journal on Semantic Web and Information Systems**, v. 5, n. 3, p. 1–22, 2009. Disponível em: <<http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>>. Acesso em: 22 out. 2015.
- BOCK, Conrad et al. OWL 2 web ontology language structural specification and functional-syntax. **W3C Recommendation 11 December 2012**, 2012. Disponível em: <<http://www.w3.org/TR/2012/REC-owl2-syntax-20121211/>>. Acesso em: 22 out. 2015.

BOISOT, Max; CANALS, Agustí. Data, information and knowledge: have we got it right? **Journal of Evolutionary Economics**, v. 14, n. 1, p. 43-67, 2004. Disponível em: <<http://link.springer.com/article/10.1007%2Fs00191-003-0181-9>>. Acesso em: 22 out. 2015.

BOOCH, Grady; RUMBAUGH, James; JACOBSON, Ivar. **UML: Guia do usuário**. 6. ed. Rio de Janeiro, Elsevier, 2005. p 13.

BRACHMAN, Ronald; LEVESQUE, Hector. **Knowledge representation and reasoning**. Elsevier, 2004. Disponível em: <<http://store.elsevier.com/Knowledge-Representation-and-Reasoning/Ronald-Brachman/isbn-9781558609327/>>. Acesso em: 22 out. 2015.

BRICKLEY, Dan; GUHA, R. V. RDF Schema 1.1. **W3C Recommendation 25 February 2014**, 2014. Disponível em: <<http://www.w3.org/TR/2014/REC-rdf-schema-20140225/>>. Acesso em: 22 out. 2015.

CABRAL, Samuel Pierri et al. Aplicando Linked Data na Publicação de Dados do ENEM. **ONTOBRAS-MOST**, p. 176-181, 2012.

CORCHO, Oscar; GÓMEZ-PÉREZ, Asunción. A road map to ontology specification languages. In: **Knowledge Engineering and Knowledge Management Methods, Models, and Tools**. Springer Berlin Heidelberg, 2000. p. 80-96. Disponível em: <<http://link.springer.com/book/10.1007%2F3-540-39967-4>>. Acesso em: 22 out. 2015.

CYGANIAK, Richard; BIZER, Chris. **Pubby: A Linked Data Frontend for SPARQL Endpoints**. 2011. Disponível em: <<http://wifo5-03.informatik.uni-mannheim.de/pubby/>>. Acesso em: 22 out. 2016.

CYGANIAK, Richard; WOOD, David; LANTHALER, Markus. RDF 1.1 Concepts and abstract syntax. **W3C Recommendation 25 February 2014**, 2014. Disponível em: <<http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>>. Acesso em: 22 out. 2015.

DATA.GOV. **Data.Gov**. Disponível em: <<https://www.data.gov/>>. Acesso em: 14 ago. 2016.

DAVENPORT, Thomas H. **Ecologia da Informação: Porque só a tecnologia não basta para o sucesso na era da informação**. 6. ed. São Paulo, Futura, 1998.

DEPARTAMENTO DE ASSUNTOS ECONÔMICOS E SOCIAIS DAS NAÇÕES UNIDAS. **Guidelines on Open Government Data for Citizen Engagement**, United Nations Department of Economic and Social Affairs, 2013. Disponível em: <<http://workspace.unpan.org/sites/Internet/Documents/Guidenlines%20on%20OGDCE%20May17%202013.pdf>>. Acesso em: 14 out. 2015.

DURST, M.; SUIGNARD, M. RFC 3987. **Internationalized Resource Identifiers (IRIs)**, 2005. Disponível em: <<http://www.ietf.org/rfc/rfc3987.txt>>. Acesso em: 22 out. 2015.

ECLIPSE. **Desktop IDEs**. 2016. Disponível em: <<https://eclipse.org/ide/>>. Acesso em: 23 out. 2016.

ELMASRI, Ramez; NAVATHE, Shamkant B. **Sistemas de banco de dados**. 6. ed. São Paulo: Pearson, 2011.

ENGHOLM, Hélio. **Engenharia de Software na Prática**. São Paulo: Novatec, 2010. p. 69-70.

ESTADOS UNIDOS. BARACK OBAMA. **Transparency and open government**. 2009. Disponível em: <<https://www.whitehouse.gov/the-press-office/transparency-and-open-government>>. Acesso em: 05 set. 2015.

FENSEL, Dieter. **Ontologies: silver bullet for knowledge management and electronic commerce**. Berlin: Springer-Verlag, 2001. Disponível em: <<http://www.springer.com/us/book/9783540003021>>. Acesso em: 22 out. 2015.

FOWLER, Martin; SCOTT, Kendall. **UML Essencial: Um breve guia para a linguagem-padrão de modelagem de objetos**. 2 ed. Porto Alegre: Bookman, 2000.

GANDON, F.; SCHREIBER, G. RDF 1.1 XML syntax. **W3C Recommendation 25 February 2014**, 2014. Disponível em: <<http://www.w3.org/TR/2014/REC-rdf-syntax-grammar-20140225/>>. Acesso em: 22 out. 2015.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

\_\_\_\_\_. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GIT. **Git, local branching on the cheap**. 2016. Disponível em: <<https://git-scm.com/>>. Acesso em: 23 out. 2016.

GITHUB. **GitHub is How people build software**. 2016. Disponível em: <<https://github.com/about>>. Acesso em: 23 out. 2016.

GRUBER, Thomas R. A translation approach to portable ontology specifications. **Knowledge acquisition**, v. 5, n. 2, p. 199-220, 1993. Disponível em: <<http://tomgruber.org/writing/ontologia-kaj-1993.pdf>>. Acesso em: 22 out. 2015.

GUARINO, Nicola; OBERLE, Daniel; STAAB, Steffen. What is an Ontology?. In: **Handbook on ontologies**. Springer Berlin Heidelberg, 2009. p. 1-17.

GUARINO, Nicola. Formal ontology in information systems. **Proceedings of the first international conference (FOIS'98)**, Junho 6-8, Trento, Itália. IOS press, 1998.

HASLHOFER, B. et al. **Europeana RDF store report: The results of qualitative and quantitative study of existing RDF stores in the context of Europeana**. Technical report, 2011. Disponível em: <[http://eprints.cs.univie.ac.at/2833/1/europeana\\_ts\\_report.pdf](http://eprints.cs.univie.ac.at/2833/1/europeana_ts_report.pdf)>. Acesso em: 29 out. 2015.

HEATH, Tom; BIZER, Christian. Linked data: Evolving the web into a global data space. **Synthesis lectures on the semantic web: theory and technology**, v. 1, n. 1, p. 1-136, 2011. Disponível em:



<<http://www.morganclaypool.com/doi/abs/10.2200/S00334ED1V01Y201102WBE001>>. Acesso em: 22 out. 2015.

HERTEL, Alice; BROEKSTRA, Jeen; STUCKENSCHMIDT, Heiner. RDF storage and retrieval systems. In: **Handbook on ontologies**. Springer Berlin Heidelberg, 2009. p. 489-508.

HORROCKS, Ian; PATEL-SCHNEIDER, Peter F. KR and reasoning on the semantic web: OWL. In: **Handbook of Semantic Web Technologies**. Springer Berlin Heidelberg, 2011. p. 365-398.

INMON, Bill. **Matching unstructured data and structured data**. 2006. Disponível em: <<http://tdan.com/matching-unstructured-data-and-structured-data/5009>>. Acesso em: 12 out. 2015.

ISOTANI, Seiji; BITTENCOURT, Ig Ibert. **Dados Abertos Conectados**. 1 ed. São Paulo: Novatec Editora, 2015. p. 18, 22-23, 102, 133,136,139. Disponível em: <<http://www.pgcl.uenf.br/2016/download/DadosAbertosConectados.pdf>>. Acesso em: 19 set. 2016.

KAUARK, Fabiana da Silva; MANHÃES, Fernanda Castro; MEDEIROS, Carlos Henrique. **Metodologia da Pesquisa: Um guia prático**. Itabuna: Via Litterarum. 86 p.

LIMA, Adilson da Silva. **UML 2.3 do Requisito à Solução**. 1. ed. São Paulo: Érica, 2011. p. 30, 281-283.

MALAMUD, Carl et al. **Open Government Data Principles**. 2007. Disponível em: <[https://public.resource.org/8\\_principles.html](https://public.resource.org/8_principles.html)>. Acesso em: 26 out. 2015.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY et al. **W3C Data Activity - Building the Web of data**. 2013. Disponível em:<<http://www.w3.org/2013/data/>>. Acesso em: 04 out. 2015.

MCGUINNESS, Deborah L. et al. OWL Web Ontology Language overview. **W3C Recommendation 10 February 2004**, 2004. Disponível em: <<http://www.w3.org/TR/2004/REC-owl-features-20040210/>>. Acesso em: 23 out. 2015.

MCGUINNESS, Deborah L. Spinning the semantic web: bringing the world wide web to its full potential. *Ontologies Come of Age*, **MIT Press, Cambridge, MA**, p. 171-195, 2003. Disponível em: <[http://www-ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mit-press-\(with-citation\).htm](http://www-ksl.stanford.edu/people/dlm/papers/ontologies-come-of-age-mit-press-(with-citation).htm)>. Acesso em: 23 out. 2015.

MELLO, Ronaldo. et al. Dados Semi-Estruturados. **Simpósio brasileiro de banco de dados**, 15. 2000, João Pessoa, PB. Anais... João Pessoa: [s.n.], 2000.

MENEZES, E.M.; SILVA, E.L.da. **Metodologia da Pesquisa e elaboração de Dissertação**. 4a Ed. UFSC, Florianópolis, 2005. Disponível em: [https://projetos.inf.ufsc.br/arquivos/Metodologia\\_de\\_pesquisa\\_e\\_elaboracao\\_de\\_teses\\_e\\_dissertacoes\\_4ed.pdf](https://projetos.inf.ufsc.br/arquivos/Metodologia_de_pesquisa_e_elaboracao_de_teses_e_dissertacoes_4ed.pdf). Acesso em: 23 out. 2015.

MOTIK, Boris et al. OWL 2 Web ontology language profiles, **W3C Recommendation 11 December 2012**. 2012. Disponível em: <<http://www.w3.org/TR/2012/REC-owl2-profiles-20121211/>>. Acesso em: 23 out. 2015.

MOTIK, Boris; PATEL-SCHNEIDER, Peter F.; GRAU, Bernardo Cuenca. OWL 2 web ontology language direct semantics. **W3C Recommendation 11 December 2012**, 2012. Disponível em: <<http://www.w3.org/TR/2012/REC-owl2-direct-semantics-20121211/>>. Acesso em: 23 out. 2015.

MOTIK, Boris; PATEL-SCHNEIDER, Peter F.; PARSIA, Bijan. OWL 2 web ontology language. **Structural Specification and Functional-Style Syntax**, 2012. Disponível em: <[http://www.w3.org/TR/owl2-syntax/#Transitive\\_Object\\_Properties](http://www.w3.org/TR/owl2-syntax/#Transitive_Object_Properties)> Acesso em: 17/10/2016.

NOY, Natalya F.; MCGUINNESS, Deborah L. **Ontology Development 101: A Guide to Creating Your First Ontology**. 2001. Disponível em: <[http://protege.stanford.edu/publications/ontology\\_development/ontology101.pdf](http://protege.stanford.edu/publications/ontology_development/ontology101.pdf)> Acesso em: 12 out. 2016.

OMG. **OMG Unified Modeling Language**. 2015. Disponível em: <<http://www.omg.org/spec/UML/2.5/PDF>>. Acesso em: 27 ago. 2016.

OPEN KNOWLEDGE INTERNATIONAL. **Open Data Handbook**. 2015. Disponível em: <[http://opendatahandbook.org/guide/pt\\_BR](http://opendatahandbook.org/guide/pt_BR)>. Acesso em: 14 out. 2015.

\_\_\_\_\_. **The Open Definition**. Disponível em: <<http://opendefinition.org/>>. Acesso em: 05 set. 2015.

ORACLE. **Obtenha informações sobre a Tecnologia Java**. 2016. Disponível em: <[https://www.java.com/pt\\_BR/about/](https://www.java.com/pt_BR/about/)>. Acesso em: 23 out. 2016.

PÁDUA, Wilson. **Engenharia de Software: Fundamentos, métodos e padrões**. 2000. Disponível em: <[http://aulasprof.6te.net/Arquivos\\_Aulas/07-Proces\\_Desen\\_Soft/Livro\\_Eng\\_Soft\\_Fund\\_Met\\_Padrees.pdf](http://aulasprof.6te.net/Arquivos_Aulas/07-Proces_Desen_Soft/Livro_Eng_Soft_Fund_Met_Padrees.pdf)> Acesso em: 07 set. 2016.

PAN, Jeff Z. Resource description framework. In: **Handbook on Ontologies**. Springer Berlin Heidelberg, 2009. p. 71-90.

PORTAL BRASILEIRO DE DADOS ABERTOS. **Instrução Normativa N°4**. 2012. Disponível em: <<http://dados.gov.br/instrucao-normativa-da-inda/>> Acesso em: 26 out. 2015.

\_\_\_\_\_. **O processo de participação social da INDA**. 2015. Disponível em: <<http://dados.gov.br/processo-de-participacao-social-da-inda/#linha-do-tempo>> Acesso em: 27 out. 2015.

PRESSMAN, Roger; MAXIM, Bruce. **Engenharia de Software-8ª Edição**. McGraw Hill Brasil, 2016.

PROTÉGÉ. **Protégé**. 2016. Disponível em: <<http://protege.stanford.edu/products.php>> Acesso em: 23 out. 2016.

PRUD'HOMMEAUX, Eric; SEABORNE, Andy. SPARQL Query Language for RDF. **W3C Recommendation 15 January 2008**. 2008. Disponível em: <<http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/>>. Acesso em: 14 out. 2015.

REZENDE, Denis Alcides. **Engenharia de software e sistemas de informação**. Brasport, 2005. p. 208-212. Disponível em: <[https://books.google.com.br/books?hl=pt-BR&lr=&id=rtBvL\\_L-1mcC&oi=fnd&pg=PT23&dq=engenharia+de+software&ots=9zhl1K-s0q&sig=OnMww\\_8nY\\_nNyPGSKV2fIJDL1BE#v=onepage&q=diagrama%20de%20classe&f=false](https://books.google.com.br/books?hl=pt-BR&lr=&id=rtBvL_L-1mcC&oi=fnd&pg=PT23&dq=engenharia+de+software&ots=9zhl1K-s0q&sig=OnMww_8nY_nNyPGSKV2fIJDL1BE#v=onepage&q=diagrama%20de%20classe&f=false)> Acesso em: 08 set. 2016.

SCHMACHTENBERG, Max; BIZER, Christian; JENTZSCH, Anja; CYGANIAK, Richard. **Linking Open Data cloud diagram 2014**. 2014. Disponível em: <<http://lod-cloud.net/>>. Acesso em: 05 out. 2015.

SETZER, Valdemar W. Dado, Informação, Conhecimento e Competência. **Data Grama Zero - Revista da Ciência da Informação**. São Paulo, n.6. 1999. Disponível em: <[http://www.dgz.org.br/dez99/Art\\_01.htm](http://www.dgz.org.br/dez99/Art_01.htm)>. Acesso em: 22 out. 2015.

SMITH, Michael K.; WELTY, Chris; MCGUINNESS, Deborah L. **OWL Web Ontology Language Guide**. 2009. Disponível em: <<https://www.w3.org/TR/owl-guide/>>. Acesso em: 15 out. 2016.

SORDI, José Osvaldo de. **Administração da informação: Fundamentos e práticas para uma nova gestão do conhecimento**. São Paulo, Saraiva, 2008.

SPARX SYSTEMS. **Enterprise Architect**. 2016. Disponível em: <<http://www.sparxsystems.com.au/enterprise-architect/>>. Acesso em: 22 out. 2016.

STAIR, Ralph M.; REINOLDS, George W. **Sistemas de Informação: Uma abordagem gerencial**. 4. ed. Rio de Janeiro: Ltc, 2002.

TAUBERER, Joshua. **Open Government Data: The Book**. 2012. Disponível em: <<https://opengovdata.io/>>. Acesso em: 08 set. 2015.

THE APACHE SOFTWARE FOUNDATION. **Apache Jena Fuseki**. 2016b. Disponível em: <<https://jena.apache.org/documentation/fuseki2/index.html>>. Acesso em: 22 out. 2016.

\_\_\_\_\_. **Apache Jena**. 2016a. Disponível em: <<https://jena.apache.org/getting-started/>>. Acesso em: 22 out. 2016.

\_\_\_\_\_. **Apache Tomcat®**. 2016f. Disponível em: <<http://tomcat.apache.org/>>. Acesso em: 22 out. 2016.

\_\_\_\_\_. **Jena Ontology API**. 2016c. Disponível em: <<https://jena.apache.org/documentation/ontology/>>. Acesso em: 22 out. 2016.

\_\_\_\_\_. **Reasoners and Rule Engines: Jena Inference support**. 2016d. Disponível em: <<https://jena.apache.org/documentation/inference/index.html>>. Acesso em: 23 out. 2016.

\_\_\_\_\_. **TDB**. 2016e. Disponível em: <<https://jena.apache.org/documentation/tdb/index.html>>. Acesso em: 22 out. 2016.

\_\_\_\_\_. **Tutorial SPARQL**. 2016g. Disponível em: <[https://jena.apache.org/tutorials/sparql\\_pt.html](https://jena.apache.org/tutorials/sparql_pt.html)> Acesso em: 30 out. 2016.

TOSIN, Thyago de Melo. Um modelo para integração de informações de bases de dados abertos, com uso de ontologias. 2016.

W3C OWL Working Group. OWL 2 Web Ontology Language document overview. **W3C Recommendation 11 December 2012**, 2012. Disponível em: <<http://www.w3.org/TR/2012/REC-owl2-overview-20121211/>>. Acesso em: 14 out. 2015.

ZHOU, Yujiao et al. Making the most of your triple store: query answering in OWL 2 using an RL reasoner. In: **Proceedings of the 22nd international conference on World Wide Web**. International World Wide Web Conferences Steering Committee, 2013. p. 1569-1580. Disponível em: <<http://dl.acm.org/citation.cfm?id=2488525/>>. Acesso em: 14 out. 2015.