



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA E GESTÃO DO
CONHECIMENTO

Jefferson de Oliveira Chaves

EGCFlow: Uma aplicação de apoio ao ciclo de vida de dados abertos conectados

Florianópolis

2021

Jefferson de Oliveira Chaves

EGCFlow: Uma aplicação de apoio ao ciclo de vida de dados abertos conectados

Dissertação submetida ao Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento da Universidade Federal de Santa Catarina para a obtenção do título de Mestre em Engenharia e Gestão do Conhecimento.

Orientador: Prof. José Leomar Todesco, Dr.

Coorientador: Prof. Alexandre Leopoldo Gonçalves, Dr.

Florianópolis

2021

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Chaves, Jefferson de Oliveira

EGCFlow : Uma aplicação de apoio ao ciclo de vida de dados abertos conectados / Jefferson de Oliveira Chaves ; orientador, José Leomar Todesco, coorientador, Alexandre Leopoldo Gonçalves, 2021.

194 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Florianópolis, 2021.

Inclui referências.

1. Engenharia e Gestão do Conhecimento. 2. Engenharia do Conhecimento. 3. Web Semântica. 4. Dados Abertos Conectados. I. Todesco, José Leomar. II. Gonçalves, Alexandre Leopoldo. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento. IV. Título.

Jefferson de Oliveira Chaves

EGCFlow: Uma aplicação de apoio ao ciclo de vida de dados abertos conectados

O presente trabalho em nível de mestrado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Fernando Ostuni Gauthier, Dr.

Universidade Federal de Santa Catarina (UFSC)

Prof. Denilson Sell, Dr.

Universidade Federal de Santa Catarina (UFSC)

Prof. Sandro Rautenberg, Dr.

Universidade Estadual do Centro-Oeste (UNICENTRO)

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de mestre em Engenharia e Gestão do Conhecimento.

Coordenação do Programa de Pós-Graduação

Prof. José Leomar Todesco, Dr.

Orientador

Florianópolis, 2021.

Dedico esta dissertação a minha companheira de vida Leila, pelo apoio incondicional e constante incentivo.

Dedico também ao meu orientador, prof. Dr. José Leomar Todesco, pela confiança, paciência, incentivo, amizade e excelente orientação.

Sem esse suporte, este trabalho não teria sido realizado. A vocês o meu mais sincero e profundo obrigado.

AGRADECIMENTOS

Cursar um mestrado sempre foi um sonho, que só pode ser realizado com o incentivo e suporte de pessoas que tornaram esse sonho realidade.

Em primeiro lugar, agradeço ao Senhor Deus, por me ouvir, me dar forças e me aliviar quando precisei.

Agradeço ao Instituto Federal Catarinense, minha casa durante bons anos, pelo suporte para a participação no mestrado, e aos companheiros que lá estão.

Existem pessoas que mudam nossa vida e talvez não saibam. Assim, agradeço ao grande amigo Rafael Speroni. Rafa, obrigado!

Agradeço também a meu orientador, prof. Dr. José Leomar Todesco, o prof. Tite. Agradeço muito por cada orientação, sempre com muito conhecimento e serenidade. Evoluí muito com suas aulas e orientação. Prof. muito obrigado!

Agradeço ao prof. Dr. Alexandre Gonçalves, pela orientação e pelo aprimoramento da dissertação por meio de sugestões cirúrgicas. Professor, muito obrigado.

E por fim, agradeço ao amor da minha vida, Leila, pela compreensão e auxílio em todos os momentos dessa caminhada. Obrigado por cada momento ao meu lado, pelas correções, sugestões, dicas e por não me deixar desistir.

“Dados são a alma da tomada de decisões e a matéria-prima para a prestação de contas. É quase impossível a concepção, o acompanhamento e a avaliação de políticas eficazes sem dados de alta qualidade que fornecem as informações corretas sobre as coisas certas no momento certo” (ORGANIZAÇÃO DAS NAÇÕES UNIDAS, 2014).

“É difícil imaginar o poder que você terá quando tantos tipos diferentes de dados estiverem disponíveis” (BERNERS-LEE, 2009).

RESUMO

A *web* se consolidou como principal meio para se publicar, obter e compartilhar informações na internet. Como uma expansão natural da *web*, surgiram os conceitos de Dados Abertos e Dados Abertos Conectados, que podem ser compreendidos como um conjunto de melhores práticas para publicação e conexão de conjuntos de dados de forma estruturada na *web*. Entretanto, o processo de produção e manutenção de conjuntos de dados abertos conectados, constitui-se como uma tarefa complexa e onerosa, mas necessária no cenário de expansão da *web* de dados. Diante disso, utilizando-se o método *Design Science Research*, com vistas a formular uma solução para um problema prático, apresentou-se a modelagem e implementação de uma aplicação para suporte à produção e a manutenção de Dados Abertos Conectados, por meio da automatização dos fluxos de trabalho. A aplicação apresentada, intitulada de EGCFLOW, fundamentou-se na aplicação das Boas Práticas para Publicação de Dados Conectados e das Boas Práticas para Publicação de Dados na *web* com objetivo de atingir o nível máximo no Plano de Dados Abertos 5 Estrelas. Como caso de uso, apresentou-se o conjunto de dados das Eleições de 2018. Foram criados *workflows* para a produção de dados abertos, a partir da reutilização de fluxos já existentes. Como resultado foi possível verificar a aplicabilidade e contribuição da aplicação para a produção de conjuntos de dados abertos conectados, garantindo escala e repetibilidade a esse processo.

Palavras-chave: Dados Abertos Conectados. Automatização. *Web* Semântica. EGCFLOW.

ABSTRACT

The web has established itself as the main means to publish, obtain and share information on the internet. As a natural expansion of the web, the concepts of Open Data and Linked Open Data emerged, which can be understood as a set of best practices for publishing and connecting datasets in a structured way on the web. However, the production and maintenance process of connected open datasets is a complex and costly task, but necessary in the data web expansion scenario. In this sense, using the Design Science Research method, with a view to formulating a solution to a practical problem, the modeling and implementation of a tool to support the production and maintenance of Connected Open Data was presented, through the automation of workflows. The tool presented, entitled EGCFLOW, was based on the application of the Best Practices for Publishing Connected Data and the Best Practices for Publishing Data on the Web, aiming to reach the maximum level in the 5 Star Open Data Plan. As a use case, the 2018 Brazilian Elections dataset was presented. Workflows were created for the production of open data, from the reuse of existing flows. As a result, it was possible to verify the applicability of the tool and its contribution to the production of connected open datasets, ensuring scale and repeatability to this process.

Keywords: Linked Open Data. Automation. Semantic Web. EGCFLOW.

LISTA DE FIGURAS

Figura 1 – <i>Linked Data Lifecycle</i>	17
Figura 2 – Pilha de tecnologias que dão suporte à <i>web</i> semântica	29
Figura 3 – Grafo de RDF.....	34
Figura 4 – Triplas Literais	34
Figura 5 – Ligações RDF	35
Figura 6 – Grafo representando um recurso com triplas literais e ligações RDF	35
Figura 7 – Grafo ilustrando a ligação entre conjuntos de dados RDF	36
Figura 8 – Esquema de estrelas	41
Figura 9 – Diretrizes Metodológicas para Publicação de Dados Abertos.....	50
Figura 10 – Elementos do Ecossistema de Dados Abertos	51
Figura 11 – Ciclo de Vida de Dados Conectados	53
Figura 12 – O Modelo <i>Linked Data Workflow Project Ontology</i>	57
Figura 13 – Visão geral GraphDB OntoRefine	63
Figura 14 – Fluxo simplificado da aplicação	81
Figura 15 – Métodos HTTP e suas operações	84
Figura 16 – Fluxo de requisição do <i>Spring</i>	86
Figura 17 – Interação entre as camadas e as APIs propostas pelo Jena	88
Figura 18 – Representação, em destaque, das camadas implementadas.....	89
Figura 19 – Classes do modelo orientado a objeto	90
Figura 20 – Tabela para persistência dos dados extraídos	93
Figura 21 – Esquema de gerenciamento de <i>Steps</i>	95
Figura 22 – Representação dos passos para interligação com a DBpedia	96
Figura 23 – Seleção dos tipos de um recurso na DBpedia.....	97
Figura 24 – Mapeamento Semântico	99
Figura 25 – Visão geral do processo de geração do Modelo RDF	101
Figura 26 – Portal de Dados Eleitorais do TSE.....	104
Figura 27 – Criação do Projeto.....	106
Figura 28 – Criação de <i>Workflow</i> - estado do Acre	108
Figura 29 – Tabela de dados do <i>Workflow</i> criado	109
Figura 30 – Visualização de alguns <i>Steps</i> disponíveis na aplicação.....	110
Figura 31 – Cabeçalhos renomeados após a execução do Step.....	111
Figura 32 – <i>Step</i> para filtro dos dados da tabela.....	112

Figura 33 – Importação de ontologia ou vocabulário	113
Figura 34 – Mapeamento do conceito de Eleição.....	115
Figura 35 – Formulário de criação de recurso aninhado ao candidato	116
Figura 36 – Visão geral do mapeamento semântico	117
Figura 37 – Seleção da amostra.....	118
Figura 38 – Recursos selecionados pela classe “Town”	119
Figura 39 – Seleção de classe específica.....	119
Figura 40 – Steps executados para o <i>Workflow</i> do Acre.....	120
Figura 41 – Artefatos gerados após a exportação do <i>Workflow</i>	121
Figura 42 – Visão geral do relatório criado.....	122
Figura 43 – Criação de novo <i>Workflow</i> para Alagoas	123
Figura 44 – Importação de <i>Steps</i>	124
Figura 45 – <i>Workflow</i> de Alagoas com os <i>Steps</i> importados	125

LISTA DE QUADROS

Quadro 1 – Trabalhos resultantes da busca com a palavra-chave <i>web</i> semântica.....	22
Quadro 2 – Trabalhos resultantes da busca com o termo Dados Abertos.....	23
Quadro 3 – Boas Práticas para Publicação de Dados na <i>Web</i>	45
Quadro 4 – Benefícios pelo Uso das Boas Práticas	47
Quadro 5 – Boas Práticas para Publicar Dados Conectados	48
Quadro 6 – Ferramentas propostas pela <i>Linked Data Stack</i>	54
Quadro 7 – Etapas da <i>Design Science Research</i> e sua aplicação nesta pesquisa	66
Quadro 8 – <i>String</i> de busca	68
Quadro 9 – Principais contribuições de cada artigo	70
Quadro 10 – Aderência aos princípios.....	72
Quadro 11 – Ciclos de vida apresentados	74
Quadro 12 – <i>Steps</i> da aplicação	100
Quadro 13 – Descrição dos dados.....	104
Quadro 14 – Aderência às Boas Práticas para Publicação de Dados Conectados.....	126
Quadro 15 – Aderência às Boas Práticas para Publicação de Dados na <i>Web</i>	128

LISTA DE TABELAS

Tabela 1 – Filtragem da pesquisa	68
Tabela 2 – Aderência às boas práticas	73

LISTA DE ABREVIATURAS E SIGLAS

API	<i>Application Programming Interface</i>
CSV	<i>Comma Separated Values</i>
DSR	<i>Design Science Research</i>
DWBP	<i>Data on the Web Best Practices</i>
EGCFlow	<i>Easy, General and Creative Flow</i>
EAV	<i>Entity Attribute Value</i>
EC	Engenharia do Conhecimento
XML	<i>Extensible Markup Language</i>
GODI	<i>Global Open Data Index</i>
GREL	<i>Google Refine Expression Language</i>
HTTP	<i>Hypertext Transfer Protocol</i>
JSON	<i>Javascript Object Notation</i>
LDWPO	<i>Linked Data Workflow Project Ontology</i>
LOD	<i>Linked Open Data</i>
REST	<i>Representational State Transfer</i>
RDF	<i>Resource Description Framework</i>
SPARQL	<i>SPARQL Protocol and RDF Query Language</i>
OGI	<i>The Open Government License</i>
TSE	Tribunal Superior Eleitoral
TSV	<i>Tab Separated Values</i>
URI	<i>Uniform Resource Identifier</i>
URI	<i>Universal Resource Identifiers</i>
XML	<i>Extensible Markup Language</i>

SUMÁRIO

1	INTRODUÇÃO	15
1.1	CONTEXTUALIZAÇÃO E PROBLEMA DE PESQUISA	15
1.2	OBJETIVOS.....	18
1.2.1	Objetivo Geral	18
1.2.2	Objetivos Específicos	18
1.3	JUSTIFICATIVA	19
1.4	ESCOPO DA PESQUISA.....	20
1.5	ADERÊNCIA AO EGC.....	21
1.6	METODOLOGIA.....	24
1.7	ESTRUTURA DA DISSERTAÇÃO.....	24
2	DADOS NA WEB: UM PANORAMA.....	26
2.1	DADOS CONECTADOS	26
2.1.1	<i>Uniform Resource Identifier (URI)</i>	29
2.1.2	Ontologias e Modelos de Representação na <i>Web</i> de Dados	30
2.1.3	<i>Resource Description Framework (RDF)</i>	33
2.1.4	Princípios estruturais para Dados Conectados.....	37
2.2	DADOS ABERTOS CONECTADOS.....	38
2.3	PUBLICAÇÃO DE DADOS NA <i>WEB</i>	43
2.3.1	Práticas para Publicação de Dados na <i>Web</i>	43
2.3.1.1	<i>Boas Práticas para Publicação de Dados na Web</i>.....	44
2.3.1.2	<i>Boas Práticas para Publicação de Dados Conectados</i>	47
2.3.2	Ciclo de Vida para Publicação de Dados Abertos Conectados	49
2.3.2.1	<i>Diretrizes Metodológicas para Publicação de Dados Abertos</i>.....	50
2.3.2.2	<i>Dados Abertos da Irlanda</i>.....	50
2.3.2.3	<i>Linked Data Lifecycle</i>	52

2.4	TECNOLOGIAS PARA DESENVOLVIMENTO DE APLICAÇÕES PARA WEB SEMÂNTICA	55
2.4.1	<i>Linked Data Workflow Project Ontology</i>	56
2.5	APLICAÇÕES CORRELATAS	59
2.5.1	LODFlow	59
2.5.2	OpenRefine	61
3	PROCEDIMENTOS METODOLÓGICOS	64
3.1	CONTEXTUALIZAÇÃO METODOLÓGICA: A <i>DESIGN SCIENCE</i> COMO PERSPECTIVA.....	64
3.2	AS ETAPAS DE PESQUISA: A <i>DESIGN SCIENCE RESEARCH</i> COMO MÉTODO DE PESQUISA	65
3.2.1	Etapa 1 - Identificação do Problema	67
3.2.2	Definição dos Resultados Esperados	76
3.2.3	Projeto e Desenvolvimento do Artefato	76
3.2.4	Demonstração, Avaliação e Comunicação	76
4	DESENVOLVIMENTO DA aplicação	77
4.1	CONTEXTUALIZAÇÃO.....	77
4.2	DEFINIÇÃO DO ESCOPO E LEVANTAMENTO DOS REQUISITOS	79
4.3	TECNOLOGIAS E DEFINIÇÃO DA ARQUITETURA	81
4.3.1	Arquitetura REST	82
4.3.2	<i>Spring Project</i>	85
4.3.3	Apache Jena	87
4.3.4	Arquitetura da aplicação	88
4.4	ALGORITMOS E ROTINAS COMPUTACIONAIS.....	91
4.4.1	Seleção, extração e armazenamento do conjunto de dados de entrada	91
4.4.2	<i>Steps</i>	94
4.4.3	Interligação com a DBpedia	95
4.4.4	Mapeamento Semântico	98

4.4.5	Outros Steps	100
5	CASO DE USO DO EGCFLOW: ELEIÇÕES 2018	102
5.1	O SISTEMA ELEITORAL BRASILEIRO	102
5.2	ROTEIRO DE USO DA APLICAÇÃO	103
5.2.1	Seleção da fonte primária	103
5.2.2	Criação de Projeto	106
5.2.3	Criação de <i>Workflow</i>: definição dos fluxos de execução	107
5.2.4	Execução dos Steps	110
5.2.4.1	<i>Renomear cabeçalhos</i>	<i>110</i>
5.2.4.2	<i>Reordenar Conjunto de Dados</i>	<i>111</i>
5.2.4.3	<i>Normalizar caixa alta/caixa baixa</i>	<i>112</i>
5.2.4.4	<i>Filtrar</i>	<i>112</i>
5.2.4.5	<i>Execução do Mapeamento Semântico</i>	<i>113</i>
5.2.4.6	<i>Interligar à DBpedia</i>	<i>117</i>
5.2.5	Exportação de dados	120
5.3	CONSONÂNCIA COM O ESQUEMA 5 ESTRELAS E AS BOAS PRÁTICAS PARA PUBLICAÇÃO DE DADOS	125
5.3.1	Boas Práticas para Publicação de Dados Conectados	126
5.3.2	Boas Práticas para Publicação de Dados na <i>Web</i>	127
5.3.3	Plano de Dados Abertos 5 Estrelas	131
6	CONSIDERAÇÕES FINAIS	133
	REFERÊNCIAS	135
	APÊNDICE A – Ontologia para o Sistema Eleitoral Brasileiro	145
	APÊNDICE B – Relatório gerado pela aplicação após a execução do workflow para o estado do Acre	189

1 INTRODUÇÃO

Os tópicos abordados neste capítulo correspondem à contextualização, apresentação do problema de pesquisa, definição dos objetivos que norteiam o desenvolvimento desta pesquisa, bem como o seu escopo. Além disso, apresentam-se a aderência ao Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento da Universidade Federal de Santa Catarina (PPGEGC-UFSC) e a estrutura da dissertação.

1.1 CONTEXTUALIZAÇÃO E PROBLEMA DE PESQUISA

Atualmente, a *web* passa por um constante processo de evolução, revolucionando a maneira de se produzir, obter e compartilhar informações na internet. Desde sua criação, a *web* teve seus documentos organizados por meio de representações baseadas em hipertexto. Ainda que essa organização tenha facilitado o acesso por usuários humanos, sua representação pouco semântica e pouco expressiva, dificultou a extração de informações destes documentos (CUNHA, LÓSCIO, SOUZA, 2011; BERNERS-LEE, BIZER, HEATH, 2009). Dessa forma, as informações contidas nesses documentos, embora pudessem ser compreendidas por humanos, eram inadequados para o uso por meio de agentes de *software*¹ (BERNERS-LEE; HENDLER; LASSILA, 2001).

Diante desse cenário, Tim Berners-Lee delineou o que ficou conhecida como *web* semântica, com o intuito de trazer estrutura e semântica para páginas da *web*, de forma que os agentes de *software* pudessem realizar operações sobre tais páginas, o que até então não ocorria (BERNERS-LEE; HENDLER; LASSILA, 2001). Para os autores, "(...) a *Web* Semântica é uma extensão da *Web* atual, na qual a informação recebe um significado bem definido, permitindo que computadores e pessoas trabalhem em cooperação" (2001, n.p.).

Essa nova perspectiva para *web*, trouxe como princípio a atribuição de significados para os elementos, dados e expressões, e a interligação com outros conjuntos de dados, de forma a se criar uma relação de significância entre os conteúdos publicados na internet. Dessa forma, agentes de *software* poderiam acessar esses conjuntos estruturados de informações e realizar tarefas que envolvessem raciocínio automatizado desses dados, baseados em um

¹ Compreendem-se por agente de *software* as "(...) entidades de *software* autônomas que podem interagir com seu ambiente. Os agentes de *software* têm a capacidade de reagir a ações de outras entidades, humanos e outros agentes de *software*" (BREITMAN, 2005, p. 155[4]).

conjunto de regras de inferência (CUNHA; LÓSCIO; SOUZA, 2011; BERNERS-LEE; HENDLER; LASSILA, 2001).

No entanto, somente agregar semântica não se demonstrou suficiente para a construção de um espaço global de compartilhamento de dados e descoberta de novas informações. Para tal, era necessário que esses dados estivessem conectados (LAUFER, 2015). Assim, como um componente da evolução da *web* semântica, surgiu o conceito de *Linked Data* ou Dados Conectados², que pode ser entendido como um conjunto de práticas para publicação, conexão e compartilhamento de conjuntos de dados estruturados. Seguindo essas práticas, os dados são interligados em um espaço global conhecido como *web* de dados, em que é possível navegar nos dados de forma semelhante ao que se faz nos documentos da *web* tradicional, além de permitir seu processamento automatizado (BERNERS-LEE, BIZER, HEATH, 2009).

Como uma extensão da publicação de dados de forma conectada, que por vezes apresentava restrições ao seu acesso, relacionadas a licenças e formatos, foram propostos os chamados *Linked Open Data* ou Dados Abertos Conectados. A publicação de dados nesse formato, tinha entre seus principais objetivos a disponibilização de conjuntos de dados sem restrições ou mecanismos de controle, de forma que pudessem ser amplamente explorados.

Entretanto, a publicação de dados, tanto no formato conectado, quanto no formato aberto e conectado, constitui-se um desafio. A própria execução da tarefa de publicação é complexa, uma vez que domínios e contextos distintos exigem infraestrutura, ferramentas, scripts e configurações específicas para cada caso. De acordo com Rautenberg, Burda e Souza:

A criação, a manutenção e o compartilhamento de Dados Abertos Conectados são atividades complexas que requerem o uso substancial de recursos e de tecnologias para preservar os conjuntos de dados na *Web* de Dados. Os esforços relacionados a essas atividades devem ser planejados, possibilitando sua execução sistemática em conformidade às melhores práticas de publicação de Dados Abertos Conectados (2018, p. 112).

Assim, diante da complexidade e a necessidade de planejamento que tais esforços demandam, foram propostos diversos processos metodológicos, compostos por um conjunto de fases ou atividades organizadas de forma sistematizada, denominadas de Ciclo de Vida de

² Na literatura especializada poderão ser encontradas outras nomenclaturas para Dados Conectados, tais como: *Linked Data*, Dados Ligados e Dados Vinculados, entretanto, para fins de padronização optou-se por utilizar a nomenclatura de Dados Conectados.

Esses esforços tais como custo de produção e a complexidade para a publicação de tais dados podem ser atenuadas por meio da automatização do fluxo de trabalho para atividades que compõem o Ciclo de Vida de Publicação e Manutenção de Conjuntos de Dados Abertos Conectados, de forma sistemática.

Com base no exposto, apresenta-se a pergunta de pesquisa que orienta esta dissertação: *Como apoiar e dar escala à produção e manutenção de conjuntos de Dados Abertos Conectados por meio da automatização desse processo?*

1.2 OBJETIVOS

Com base na problemática apresentada foram definidos o objetivo geral e os objetivos específicos, quais sejam:

1.2.1 Objetivo Geral

O objetivo geral deste trabalho, portanto, é o de propor uma aplicação *web* para suporte à automatização da tarefa de produção e manutenção de conjunto de Dados Abertos Conectados (*Linked Open Data*).

1.2.2 Objetivos Específicos

A partir do objetivo geral, definiram-se os seguintes objetivos específicos:

- Identificar as boas práticas para publicação de dados abertos conectados;
- Identificar as tecnologias aplicadas na execução de um fluxo de publicação de dados abertos conectados;
- Especificar os requisitos funcionais para uma aplicação de publicação de dados abertos conectados de forma automatizada;
- Verificar os resultados obtidos pela execução da aplicação quanto a escala, reprodutibilidade e a aplicação de boas práticas.

1.3 JUSTIFICATIVA

A disponibilização de bases de Dados Abertos Conectados, conforme as características descritas acima é essencial para a interoperabilidade entre diferentes conjuntos de dados, permitindo assim, que distintos atores da sociedade construam soluções, sejam aquelas desenvolvidas na esfera governamental ou privada, na acadêmica ou na sociedade civil, cada vez melhores (PIRES, 2015).

Segundo Pires (2015) a clareza sobre a definição de “dados abertos” permite que bases de dados, mesmo que de fontes diferentes, sejam combinadas sem grandes dificuldades técnicas. Isso evita que o governo seja um repositório de bases de “dados fechadas”, passíveis apenas de consultas humanas, e inadequadas para uso em sistemas capazes de proporcionar soluções, visualizações, serviços ou valor para qualquer cidadão ou grupo da sociedade, promovendo a transparência e o controle social.

Buscando avançar nesse sentido, o Brasil, em 2011, participou da fundação da *Open Government Partnership*, uma associação que conta atualmente com a participação de 78 países, além de incluir milhares de organizações da sociedade civil (OPEN GOVERNMENT PARTNERSHIP, 2021). Além disso, outras iniciativas nacionais foram implementadas como o Portal Brasileiro de Dados Abertos, em 2011 e a criação da Lei de Acesso à Informação, também em 2011, que prevê em seu artigo 8º, § 3º, inciso III, que sites de órgãos e entidades públicas brasileiras deverão: “possibilitar o acesso automatizado por sistemas externos em formatos abertos, estruturados e legíveis por máquina” (BRASIL, 2011).

Com a finalidade de medir os avanços na publicação de dados abertos alcançados pelos países, a *Global Open Data Index* (GODI), referência global para medição da publicação de dados governamentais abertos, realiza periodicamente, uma pesquisa sobre a temática. Em seu último relatório, ano-base 2016, a GODI, em um *ranking* sobre publicação de dados abertos de governos nacionais, classificou o Brasil na 8º colocação, em um universo de 94 países. Para tal classificação, a instituição definiu alguns conjuntos de dados, tais como: orçamento público, estatísticas nacionais, leis nacionais e gastos governamentais, de forma a padronizar as variáveis analisadas, buscando conferir maior isonomia nos dados analisados. Cada um desses conjuntos de dados foi avaliado, de acordo com uma série de boas práticas, quais sejam: disponibilização sob licença aberta, em formato aberto e legível por máquina, de forma completa, atualizada e disponível, de forma pública e gratuita (GLOBAL OPEN DATA INDEX, 2017).

Em outro estudo, este realizado pela Fundação Getúlio Vargas, em parceria com a *Open Knowledge Brasil*, e publicado em 2018, foram analisados os índices de dados abertos em oito cidades brasileiras. Os resultados indicam que embora tenham ocorridos pequenos avanços, o quadro apresentado ainda ficou longe do ideal, uma vez que o foco dessas cidades foi direcionado, principalmente, à publicização de dados. O estudo indica que: “(...) é preciso fazer mais: disponibilizar bases em formato aberto, amigável, com metadados adequados, claramente destinados ao domínio público e que facilitem o uso e o entendimento para uma população cada vez maior” (RUEDIGER, MAZZOTTE, 2018, p. 64).

Quando buscadas informações sobre Dados Conectados ou Dados Abertos Conectados, não foram encontrados dados estatísticos que indicassem o atual cenário sobre o assunto. Indicativo dessa dificuldade é que no relatório publicado em 2015, pela GODI, quando analisados os dados do Brasil, em um universo de 13 conjuntos de dados avaliados, apenas um deles estava disponível como Dado Aberto Conectado (GLOBAL OPEN DATA INDEX, 2015).

Diante desse cenário, diversos esforços foram realizados com o intuito de buscar a automatização dos fluxos de trabalho de Dados Conectados, dentre os quais destaca-se o *Linked Data Workflow Project Ontology* (LDWPO), ontologia que serviu de base e inspiração para esta pesquisa.

Cabe ressaltar que, a automatização dos fluxos de trabalho de Dados Conectados traz inúmeros benefícios, do ponto de vista tecnológico, social, para o setor público e privado, dentre os quais podemos citar: geração de valor, *accountability*, transparência, qualidade dos dados, base para tomada de decisões e geração de conhecimento (ENAP, 2017; ORGANIZAÇÃO DAS NAÇÕES UNIDAS, 2014). Com base no exposto, a seguir apresentaremos os objetivos que norteiam esta dissertação.

1.4 ESCOPO DA PESQUISA

Conforme mencionado, esta pesquisa propõe-se a implementar uma aplicação que apoie a produção e manutenção de dados abertos conectados, com o intuito de dar escala a esse processo.

Nesse sentido, destacam-se como elementos de abrangência deste trabalho a realização de uma etapa de conscientização acerca do tema, por meio de uma revisão

sistemática de literatura, e a construção do arcabouço teórico, acerca dos conteúdos que permeiam essa temática, com o intuito de se obter embasamento para modelagem e construção do artefato aqui proposto.

No que se refere a modelagem e desenvolvimento da aplicação, esta pesquisa propõe-se a realizar a criação e descrição de planos de produção de dados abertos conectados, bem como a descrição dos fluxos para produção e manutenção desses dados. Além disso, objetiva-se a execução de passos para enriquecimento do conjunto de dados e para o estabelecimento de relações com outros conjuntos de dados. Também faz parte do escopo o desenvolvimento de meios para reuso dos fluxos de publicação criados.

1.5 ADERÊNCIA AO EGC

Esta dissertação propõe-se a implementar uma aplicação que apoie a produção e manutenção de Dados Abertos Conectados. É, portanto, uma pesquisa de caráter interdisciplinar, que envolve as áreas de Engenharia do Conhecimento, Engenharia de Ontologias, Engenharia de *Software* e Ciência da Computação.

O trabalho está vinculado à área de concentração de Engenharia do Conhecimento (EC) do Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento (PPGEGC), especificamente, vinculado à linha de pesquisa Engenharia do Conhecimento Aplicada às Organizações, que objetiva estudar, desenvolver e a implantar soluções da Engenharia do Conhecimento em organizações públicas e privadas.

Uma das principais atividades da EC é o de transformar dados em conhecimento. Conforme Pacheco (2014), a EC é definida como a “disciplina que se dedica à modelagem de conhecimento e a criação e inserção de sistemas de conhecimento nas organizações”. E ainda, pode ser definida como a disciplina responsável pela definição de metodologias e ferramentas, capazes de contribuir na modelagem e criação de conhecimento, de tal forma que ele possa ser apropriado pelas instituições (ABEL, FIORINI, 2013).

Para alcançar tal propósito, a EC vale-se de ferramentas, tais como a *web* semântica, as ontologias, a inteligência artificial, dentre outras. Nesse sentido, a implementação de uma aplicação que apoie a produção e a manutenção de dados conectados, tal qual está sendo proposto neste trabalho, potencializa a possibilidade de conexão para uma maior geração de conhecimento. Além disso, a aplicação dará escala a disponibilização de dados abertos

conectados, que atualmente não se encontra em bons percentuais quando comparados com outros países, diminuindo assim o compartilhamento de conhecimento.

É necessário destacar também que a automatização da produção de conjuntos de dados como dados abertos conectados permite que os mesmos possam ser lidos por humanos e máquinas, que de forma facilitada podem, por exemplo, visualizar e realizar inferências, ou até mesmo cruzamentos e correlações com outros conjuntos de dados.

Tais possibilidades podem ser um meio para impulsionar a inovação, que a partir do uso de dados colaborativos e compartilhados, permitem mudanças disruptivas alinhadas à economia atual. Além disso, a utilização dessa aplicação para manutenção do fluxo de trabalho permite que, após sua configuração, tal fluxo possa ser reproduzido de forma sistematizada, sem necessidade de conhecimento técnico.

Ainda, ao buscarmos trabalhos no Banco de Teses e Dissertações do PPGE GC utilizando como palavra-chave o termo “*Web Semântica*”, o banco retornou seis trabalhos, dentre os quais cinco são dissertações e um é uma tese, e que estão listados no quadro a seguir:

Quadro 1 – Trabalhos resultantes da busca com a palavra-chave *web semântica*

Tipo	Título	Autor	Ano
D	Aplicação de ontologias para apoiar operações analíticas sobre fontes estruturadas e não estruturadas	Márcio Napoli	2011
D	Uso da <i>Web</i> de Dados como Fonte de Informação no Processo de Inteligência Competitiva Setorial	Leandro Dal Pizzol	2014
D	Aplicação de Ontologias na Organização de Conteúdos para Apoio a Equipes de Desenvolvimento de <i>Software</i>	Maurício Botelho	2015
T	Modelo de Referência Para Indicadores de Inovação Regional Suportado por Dados Ligados	Rafael de Moura Speroni	2016
D	Proposta de Arquitetura para Ecossistema de Inovação em Dados Abertos	Murilo Silveira Gomes	2017
D	Modelo de Repositório Institucional com Suporte a <i>Web Semântica</i> e Rede Social	Renato Cesca	2018

Fonte: Elaborado pelo autor, com base no banco de teses e dissertações do PPGE GC-UFSC.

Napoli (2011) descreveu uma aplicação baseada em ontologias que permitiu o uso de recursos de processamento analítico. Dal Pizzol (2014), por sua vez, propõe alinhar a inteligência competitiva à *web* de dados, por meio da proposição de um modelo. Botelho (2015) definiu uma abordagem para o apoio à organização e localização de conteúdo, de forma orientada ao trabalho de equipes de desenvolvimento de *software* e, para tal, estruturou uma ontologia com base em uma equipe de desenvolvedores de *software* de gestão de recursos humanos na área pública. Já Speroni (2016) propôs um modelo de referência concebido a partir da análise de modelos de indicadores compostos para a mensuração da inovação regional, propondo uma classificação hierárquica para tais indicadores. Gomes (2017) por sua vez, propôs uma modelagem de um ecossistema de inovação baseado em tecnologias semânticas. Por fim, Cesca (2018) propôs a criação de um modelo de repositório institucional com suporte a funcionalidades de redes sociais, de forma que atraísse o interesse dos usuários.

Quando pesquisado o termo “ferramenta” o sistema retornou três teses:

- BERG, Carlos Henrique. Ferramenta Para Identificação de Emoções a Partir de Onomatopeias para Pessoas com Diferentes Habilidades Visuais. Tese, 2017;
- LEONARDI, Juliana. Ferramenta Avaliativa de Relações Dimensionais na Criação de Conhecimento. Tese, 2017;
- OHIRA, Masanao. Ferramenta para análise do estado de evolução do conhecimento em organizações. Tese, 2009.

Por sua vez, quando o termo “Dados Abertos” foi buscado no banco, obteve-se o seguinte resultado:

Quadro 2 – Trabalhos resultantes da busca com o termo Dados Abertos

Tipo	Título	Autor	Ano
D	Uma Proposta de Modelo Conceitual para Uso de Big Data E Open Data para Smart Cities.	Vinicius Barreto Klein	2016
D	Proposta de Arquitetura para Ecossistema de Inovação em Dados Abertos	Murilo Silveira Gomes	2017

D	OGDPub: uma Ontologia para Publicação de Dados Abertos Governamentais	Larissa Mariany Freiberger Pereira	2017
D	Dados Abertos Governamentais: Desafios na Publicação	Paula Assumpção Campos	2018
T	Governança de Dados Abertos Governamentais: Framework Conceitual para as Universidades Federais, Baseado em uma Visão Sistêmica	Júlio César Costa Casaes	2019

Fonte: Elaborado pelo autor, com base no banco de teses e dissertações do PPGE GC-UFSC.

Cabe destacar que, os trabalhos listados acima tratam sobre Dados Abertos. Quando o termo Dados Abertos Conectados foi buscado a plataforma não retornou nenhum trabalho. Dessa forma, esta pesquisa representa um avanço e traz contribuições essenciais para o assunto, uma vez que trata sobre Dados Abertos Conectados.

As próximas seções apresentarão, de forma sucinta, a metodologia da presente pesquisa, bem como a forma como a mesma encontra-se estruturada.

1.6 METODOLOGIA

Num primeiro momento, esta pesquisa caracteriza-se por ser uma pesquisa de cunho bibliográfico. Já em termos estruturais esta pesquisa está orientada pelo *Design Science Research* (DSR).

A Design Science é a ciência que procura desenvolver e projetar soluções para melhorar sistemas existentes, resolver problemas ou, ainda, criar novos artefatos que contribuam para uma melhor atuação humana, seja na sociedade, seja nas organizações. Logo, a natureza deste tipo de pesquisa costuma ser pragmática e orientada à solução (DRESCH, LACERDA, JÚNIOR, 2015, p. 57).

Tal escolha deve-se ao fato que esta pesquisa, além de levantar um problema, busca construir uma solução para o problema, por meio de um artefato, que no caso desta pesquisa é a aplicação que apoiará a publicação e manutenção de Dados Conectados. A metodologia utilizada será abordada com profundidade no capítulo três.

1.7 ESTRUTURA DA DISSERTAÇÃO

Esta dissertação encontra-se estruturada em cinco capítulos, além deste. O capítulo dois apresentará a fundamentação teórica deste trabalho, trazendo um panorama sobre os

Dados Abertos e Dados Conectados; sobre a *Web Semântica* e Ontologias e, por fim, aspectos da Engenharia de *Software*. O capítulo três, por sua vez, apresentará a metodologia deste trabalho, que conforme já mencionado pauta-se no *Design Science Research*. Já o capítulo quatro descreve como ocorreu o desenvolvimento da aplicação. O capítulo cinco abordará os resultados apresentados com a implementação da aplicação. Por fim, o capítulo seis trará as considerações finais desta pesquisa e indicações de possíveis trabalhos futuros.

2 DADOS NA WEB: UM PANORAMA

Desde seu surgimento, no final da década de 1980, a *web* se destacou como um meio importante para publicação de documentos, permitindo o compartilhamento e geração de novas informações, consumidas por usuários em todas as partes do mundo.

Embora o interesse na publicação de dados na *web* não fosse algo novo, nos últimos anos, numerosas iniciativas foram desenvolvidas com o intuito de realizar a publicação de dados de maneira a promover o acesso, o compartilhamento e a reutilização desses dados, tanto por humanos, quanto por agentes de *software*. Contudo, somente assegurar os acessos aos dados não é suficiente. É essencial que os dados publicados sejam disponibilizados de forma que sejam compreendidos e utilizados por consumidores e, que possam ser facilmente processados por aplicações (LÓSCIO *et al.*, 2018).

Entretanto, fatores como a heterogeneidade dos dados e a falta de padrões para descrição e acesso aos conjuntos de dados, tornaram o processo de publicação, compartilhamento e consumo de dados uma tarefa complexa (LÓSCIO *et al.*, 2018).

Atividades comuns ao ciclo de vida para publicação e manutenção de conjunto de dados, revelaram-se um desafio, devido a sua complexidade, o tempo despendido para tal, seu custo e quantidade de recursos envolvidos. Entretanto, essa complexidade pode ser reduzida por meio da automatização do fluxo de trabalho que suporte o ciclo de vida dos conjuntos de Dados Conectados de forma sistemática.

Assim, este capítulo irá apresentar fundamentos relacionados à publicação de dados na *web*, abordando aspectos relevantes, incluindo os conceitos de Dados Conectados, Dados Abertos Conectados, Ciclo de Vida dos Dados na *Web* e as Boas Práticas para Dados na *Web*.

2.1 DADOS CONECTADOS

O conceito de Dados Conectados foi introduzido por Tim Berners-Lee no ano de 2006, e estabeleceu uma nova forma de publicação de dados na *web*. Este conceito surgiu como resultado de outro mais abrangente, definido como um modelo de extensão da *web*, também proposto por Berners-Lee, Hendler e Lassila, no início dos anos 2000, a *Web Semântica*.

Desde o fim da década de 1980, quando Tim Berners-Lee publicou sua proposta para um sistema de compartilhamento de documentos de pesquisa, a *World Wide Web*, ou simplesmente *Web*, a maneira de se compartilhar conhecimento mudou radicalmente, reduzindo a barreira para publicação e acesso a documentos em um espaço de informação global (BERNERS-LEE, BIZER, HEATH, 2009). O surgimento da *web* possibilitou que documentos fossem interligados por meio de *hyperlinks* e disponibilizados na Internet, tornando a interação do usuário com a rede mundial mais amigável, criando assim, um ambiente de compartilhamento de documentos de distintas áreas do conhecimento (BATISTA, LÓSCIO, 2013; CUNHA, LÓSCIO, SOUZA, 2011).

Entretanto, apesar dos benefícios indiscutíveis que a *web* apresentava, a estrutura desses documentos, seguindo apenas regras sintáticas, orientadas pela linguagem de marcação em *HyperText Markup Language* (HTML), com somente objetivos de apresentação e leiaute, não era expressiva o suficientemente para o consumo das informações contidas nesses documentos por agentes de *software*, nem tampouco permitia que elementos individuais fossem identificados e conectados a outros elementos relacionados.

De acordo com Berners-Lee, a *web* se desenvolveu mais rapidamente como meio de disponibilizar documentos para pessoas, em vez de dados e informações que podem ser processados automaticamente por agentes de *software* (BERNERS-LEE, BIZER, HEATH, 2009; BERNERS-LEE, HENDLER, LASSILA, 2001).

Isso fez com que o significado e o formato dos dados disponíveis na *web* não seguissem um padrão, de forma que cada serviço de publicação utilizasse a representação que lhe fosse mais conveniente.

Aplicações que quisessem integrar conjuntos de dados com dados de outras aplicações, precisariam conhecer detalhes desses conjuntos, além de saber como interpretar as diferentes definições a fim de identificar as semelhanças e as diferenças entre os distintos dados retornados. Nesse contexto, para cada nova integração, era necessário compreender a semântica e a maneira como ela foi descrita, de forma que a interoperabilidade entre diferentes serviços e seus dados tivessem que ser realizadas manualmente (LAUFER, 2015).

Essas limitações, tanto na interpretação, quanto na integração das informações, fizeram com que em 2001, fosse proposta o que foi chamada de *Web Semântica*. De acordo com Berners-Lee:

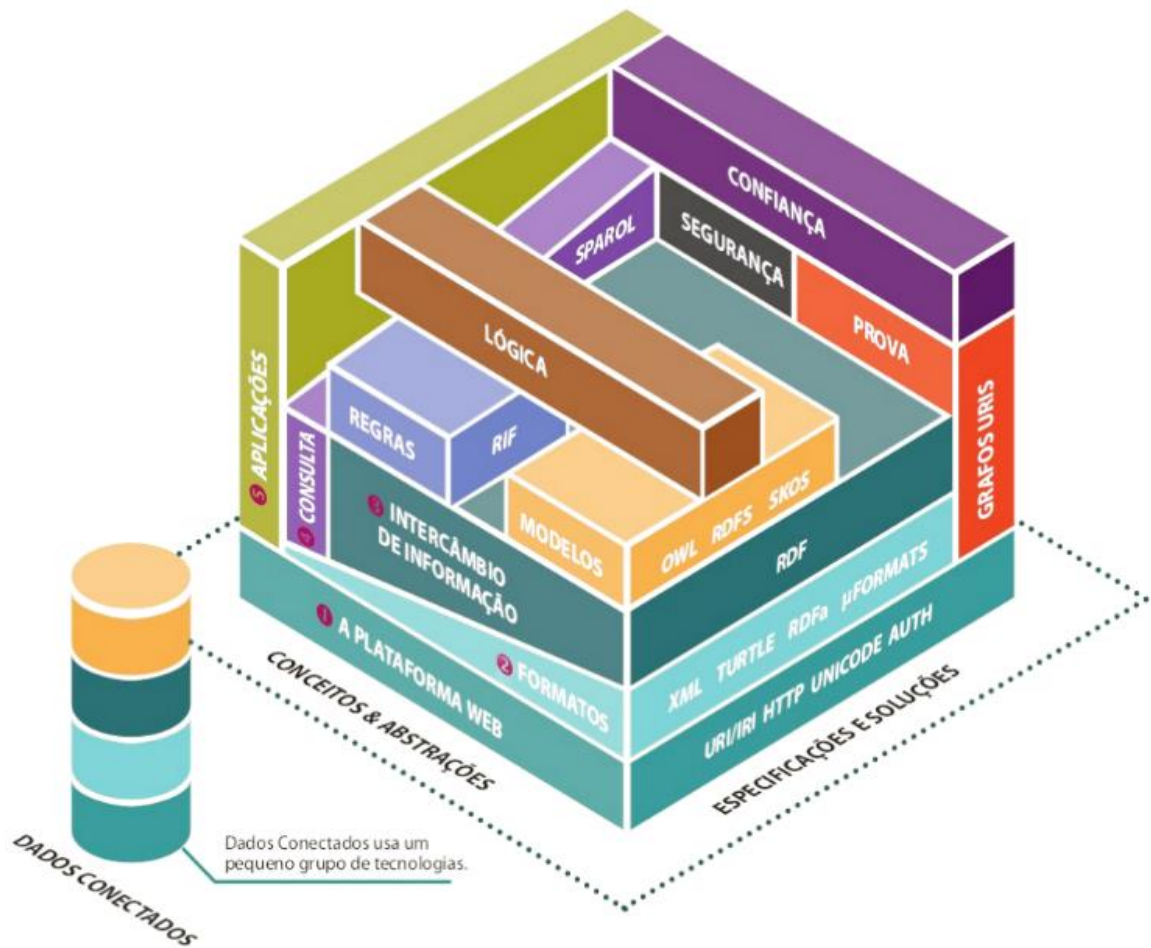
A Web Semântica não é uma Web separada, mas uma extensão da atual, na qual as informações recebem um significado bem definido, permitindo que computadores e pessoas trabalhem em cooperação (2001, p. 2, tradução nossa).

A *Web Semântica* foi construída com o intuito de possibilitar a expressão de informações e regras sobre os dados e sobre como raciocinar sobre esses dados, criando assim um modelo comum, diminuindo a possibilidade de ambiguidades e facilitando a interoperabilidade entre fontes de dados distintas (BERNERS-LEE, BIZER, HEATH, 2009; LAUFER, 2015).

Essa nova perspectiva, veio acompanhada de uma pilha de tecnologias, apresentadas como um modelo, chamado de “pirâmide da *Web Semântica*”, em que foram descritos recursos e linguagens utilizados na *web semântica*. Esse modelo contava com tecnologias já consolidadas na *web clássica*, tais como os *Universal Resource Identifiers* (URIs), utilizados como meio para identificação de recursos de forma global e o *HyperText Transfer Protocol* (HTTP), mecanismo de acesso a recursos.

Ainda, foram propostas a utilização de outras tecnologias, tais como o *Resource Description Framework* (RDF), como formato para a publicação e conexão entre dados e recursos, o uso de ontologias, taxonomias e regras de inferência, empregadas como meio para formalização dos conceitos utilizados nos conjuntos de dados, além de estabelecer como parte de seu modelo, a SPARQL, como linguagem de consulta e acesso aos dados. (BERNERS-LEE, HENDLER, LASSILA, 2001; ISOTANI, BITTENCOURT, 2015; LAUFER, 2015). A Figura 2 apresenta a pilha de tecnologias que dão suporte à *web semântica*.

Figura 2 – Pilha de tecnologias que dão suporte à *web* semântica



Fonte: Isotani, Bittencourt (2018, p. 30).

Essas tecnologias, segundo Laufer (2015) tiveram como objetivo definir: i) um modelo de dados padrão; ii) um conjunto de vocabulários de referência; e iii) um protocolo padrão de consulta. As seções seguintes apresentam essas tecnologias.

2.1.1 *Uniform Resource Identifier* (URI)

Um *Uniform Resource Identifier* (URI) é uma sequência de caracteres que identifica um recurso físico ou lógico (BERNERS-LEE, FIELDING, MASINTER, 2005). De acordo com Jacobs e Walsh (2004), a utilização de URIs traz inúmeros benefícios, tais como a ligação, a marcação, o *bookmark*, armazenamento em cache e a indexação de dados. Jacobs e Walsh (2004) estabelecem três características para as URIs, quais sejam:

- Uniformidade: refere-se ao emprego de recursos seja no mesmo contexto, seja em contextos distintos;
- Recurso: refere-se a qualquer coisa que possa ser identificada por um URI, tais como vídeos, imagens, serviços ou documentos;
- Identificador: refere-se à informação necessária para que um recurso possa ser identificado e diferenciados dos demais.

Essas características são fundamentais, pois possibilitam a identificação de recursos de forma não ambígua, evitando assim problemas em sua referência, como os que, por exemplo, ocorrem com nomes como Rio de Janeiro e São Paulo, que podem representar tanto o estado, quanto a cidade e, que podem ser evitados com a utilização de URIs, dada sua natureza de unicidade na identificação de recursos.

O emprego de URIs para identificar recursos, o protocolo HTTP como mecanismo de recuperação e o modelo de dados RDF para representar descrições de recursos, são elementares para a concretização de arquitetura de Dados Conectados (JACOBS, WALSH, 2004).

O RDF oferece um modelo de dados genérico, baseado em grafos, na forma de sujeito, predicado, objeto, com o intuito de se estruturar e vincular dados que descrevem coisas no mundo. O sujeito é representado por um URI. O predicado especifica como o sujeito e o objeto estão relacionados e também é representado por um URI. O objeto pode ser representado por um literal de string, ou por um URI, quando representa um recurso.

Cabe ressaltar, porém, que a utilização do modelo RDF e de URIs como identificador de recursos, estão relacionadas à forma como sintaticamente os metadados podem ser agregados às informações (LAUFER, 2015). Para que o significado pretendido pelo publicador dos dados tenha a mesma relação de significância pelo consumidor dos dados é imprescindível o uso de vocabulários com uma semântica bem definida.

2.1.2 Ontologias e Modelos de Representação na *Web* de Dados

A publicação de conjuntos de dados traz inúmeros benefícios, tais como transparência, serendipidade, etc. Entretanto, uma vez que um dado pode ser descrito, representado e interpretado de distintas maneiras, são necessários meios para que o

significado pretendido pelo publicador seja o mesmo interpretado pelo consumidor. Para tanto, a utilização de vocabulários de referência ou ontologias que possuam uma semântica bem definida são fundamentais neste processo.

De acordo com Gruber (1993), uma ontologia como ser definida como “uma especificação explícita de uma conceitualização”. O termo conceitualização pode ser compreendido como a relação de significância dos conceitos e suas relações em determinado domínio. Já a “especificação” pode ser definida como uma representação formal, declarativa e explícita dos mesmos conceitos e das mesmas relações (Isotani e Bittencourt, 2015). Para Swartout *et al.* (1999, p.18), uma ontologia pode ser descrita como uma “estrutura básica ou a couraça em torno da qual uma base de conhecimento pode ser construída”.

As definições apresentadas destacam o quão fundamental é a concordância na conceitualização especificada por uma ontologia. Quando isso não ocorre (se a conceitualização não for comumente aceita) a probabilidade de reuso de uma ontologia é baixa (BORST, 1997). Neste contexto, Borst (1997) estende a definição, afirmando que uma ontologia é “uma especificação formal e explícita de uma conceitualização compartilhada”.

Dessa forma, por oferecer uma estrutura conceitual de entendimento comum, que podem ser utilizadas de base para bases de conhecimento compartilháveis e reutilizáveis, e por viabilizar a interoperabilidade e a combinação de informações, potencializando assim a criação de aplicações computacionais cada vez mais sofisticadas e inteligentes, o uso de ontologias tem se tornado essenciais em aplicações da *web* semântica.

De acordo com Mizoguchi (2004) e Isotani (2009), uma ontologia é constituída por um conjunto de conceitos essenciais e um corpo de conhecimento. Segundo Isotani e Bittencourt (2015, p. 96), o conjunto de conceitos iniciais pode ser definido como um conjunto de conceitos obtidos por meio da “articulação do conhecimento básico presente em um determinado domínio”. Tais conceitos, podem ter sua representação realizada por meio de vocabulários especializados.

Já o corpo de conhecimento, pode ser compreendido como um conjunto de conceitos que definem um domínio. Segundo Isotani e Bittencourt (2015) é composto por: i) uma hierarquia de classes e subclasses, resultado das relações do tipo *is-a* entre os conceitos; ii) um grupo de relações entre conceitos, tipo *part-of*, por exemplo e; iii) axiomas que representem restrições semânticas entre esses conceitos e suas relações.

Os elementos que constituem uma ontologia são essenciais para construção estruturada da representação do conhecimento em um determinado domínio.

Para esta pesquisa os termos vocabulário e ontologia são utilizados de forma intercambiável. De acordo com Laufer (2015), não existe uma separação clara dos dois conceitos, sendo que, em muitos casos, o termo vocabulário é utilizado para o caso de ontologias mais simples.

De acordo com Isotani e Bittencourt (2015), a escolha de um vocabulário adequado, que melhor identifique um conjunto de dados é um desafio e, deve ser feito de forma com que se aumente a expressividade dentro do contexto em que foi criado e se reduzam as ambiguidades em sua interpretação.

Para que o cenário da *web* semântica se concretize, é necessário que sejam definidos vocabulários de referência, como meio de se facilitar a comunicação dos metadados. Novas publicações devem consultar vocabulários já existentes que possam ser utilizados. Segundo Isotani e Bittencourt (2015), caso não existam vocabulários que representem de forma expressa o suficiente as necessidades de expressividade dos metadados, novas ontologias podem ser construídas, observando-se, no entanto, elementos de ontologias já existentes, evitando referências distintas que tratem do mesmo conceito.

Um vocabulário deve ser descrito por meio de um documento apontado por uma URI. Por exemplo, o vocabulário FOAF tem a URI `<http://xmlns.com/foaf/0.1/>`. A URI de referência às classes e às propriedades de cada um dos vocabulários é construída a partir da concatenação da URI do vocabulário com o nome da respectiva classe ou propriedade. Por exemplo, a propriedade “*name*” de FOAF tem a URI `<http://xmlns.com/foaf/0.1/name>`.

Desde a proposta da *web* semântica, a concepção e o uso de ontologias sempre estiveram presentes, estando no centro da arquitetura idealizada por Tim Berners-Lee, mostrando-se uma das tecnologias-chave na criação de aplicações mais inteligentes e adequadas para lidar com grandes quantidades de informações (MCGUINNESS, 2004; HORROCKS, 2008).

Ainda, problemas que cercam a construção de tecnologias que utilizam bases de dados com representação formal - também chamadas de bases de conhecimento - tais como, a dificuldade na reutilização e compartilhamento do conhecimento; a falta de modelos genéricos comuns sobre os quais possamos construir bases de dados e aplicativos de maneira simplificada; e a falta de tecnologias que possibilitem a acumulação incremental dos dados (isto é, estender rapidamente a base de dados) podem ser solucionados por meio do uso de ontologias (MIZOGUCHI, 2004; D’AQUIN, NOY, 2012; DEVEDZIC, 2006).

Assim, para a concretização da *web* de dados, é necessário que a informação e o conhecimento que, por vezes, encontram-se esparsos e fragmentados sejam descritos e publicados, de modo a fornecer um conjunto de conceitos essenciais, representados por meio de um vocabulário especializado, resultantes da articulação do conhecimento básico presente em um determinado domínio. O tópico a seguir, traz um exemplo de utilização de vocabulários como parte de um modelo de dados criado para auxiliar na descrição dos recursos disponíveis na *web*, o RDF.

2.1.3 *Resource Description Framework (RDF)*

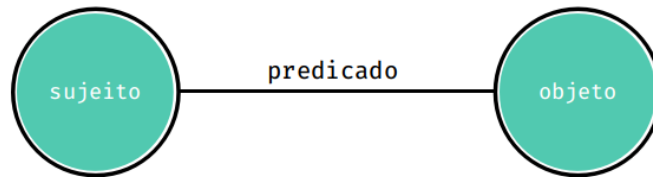
A *web* semântica surgiu como solução ao problema apresentado pela *web* clássica, que apresentava dificuldade em fazer com que os dados publicados fossem processados não somente por humanos, mas também por agentes de *software*. Isso ocorreu principalmente pela falta de meios para se agregar informação semântica aos dados e pelo uso de um formato de representação inadequado (LAUFER, 2015).

A solução proposta foi a utilização de metadados, como forma de descrever os dados publicados, fornecendo descrições sobre a estrutura, o conteúdo e outras informações relacionadas (licença de uso, direitos de propriedade intelectual, proveniência, etc.). Assim, como padrão para descrição de dados e metadados, de forma a possibilitar a utilização uniforme dessas descrições entre distintos sistemas, baseando-se em padrões sintáticos e semânticos dos metadados, foi proposto o modelo RDF.

O RDF é um modelo de dados extremamente simples, adaptado para a arquitetura da *web* (BIZER, CYGANIAK, HEATH, 2007), criado para auxiliar na descrição dos recursos disponíveis na *web* de maneira mais precisa, além de permitir a descrição da relação e dos significados entre diversos recursos (CYGANIAK, WOOD, LANTHALER, 2014). Esse modelo de representação possibilitou que recursos contidos em conjuntos de dados pudessem ser descritos, processados e interligados. A forma de descrição se dá como uma série de afirmações compostas por sujeito, predicado e objeto, chamadas de triplas RDF (LAUFER, 2015; BERNERS-LEE, BIZER, HEATH, 2009).

Uma tripla RDF consiste, portanto, de três componentes, que são convencionalmente escritos na seguinte ordem: i) **sujeito**, que é um URI ou um nó em branco; ii) **predicado**, que é um URI; iii) **objeto**, que é um URI, um literal ou um nó em branco. A representação dessas triplas é chamada de Grafo RDF. Este grafo RDF pode ser visualizado na Figura 3.

Figura 3 – Grafo de RDF

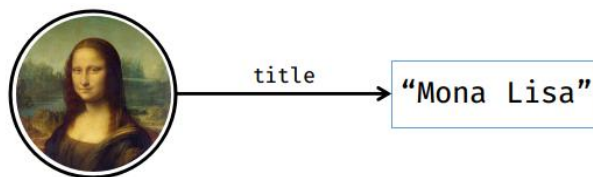


Fonte: Elaborado pelo autor.

O sujeito representa um recurso, que pode ser qualquer coisa, desde que identificada por um URI. O objeto pode também ser um recurso identificado por um URI, relacionado ao sujeito, ou mesmo um valor literal, como um número ou texto qualquer. O predicado estabelece o tipo de relação existente entre sujeito e objeto contidos na tripla. Assim, sujeito e objeto representam dois recursos relacionados, ao passo que o predicado representa a natureza do relacionamento (MANOLA, MILLER, MCBRIDE, 2014).

Bizer, Cyganiak e Heath (2007) classificam as triplas RDF em dois tipos principais: Tripas Literais e Ligações RDF. Tripas Literais são as triplas que possuem literais (um texto, um número, uma data) como objetos e tem como propósito descrever propriedades de recursos, como por exemplo, nome ou data de nascimento de uma pessoa. Um exemplo de tripla literal pode ser visto na Figura 4. Na figura, o sujeito Mona Lisa é descrito por meio do predicado *<title>*, cujo valor é o literal *<Mona Lisa>*.

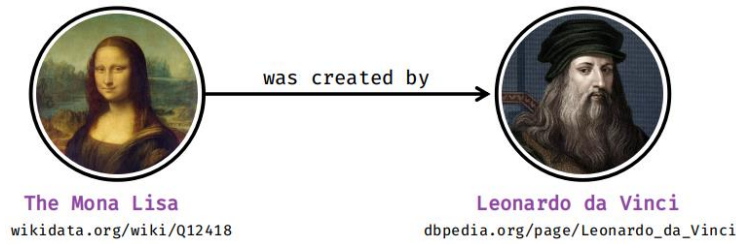
Figura 4 – Tripas Literais



Fonte: Elaborado pelo autor, adaptado de Manola, Miller, Mcbride (2014).

Ligações RDF, por sua vez, são triplas que possuem um recurso como objeto. Essas triplas representam ligações tipificadas entre dois recursos. Uma Ligação RDF se dá pela formação de uma tripla em que sujeito e objeto são recursos interligados por um predicado que representa a relação entre eles e seu intuito é estabelecer relações entre recursos distintos, tais como, a relação entre empregado que trabalha organização, ou uma pessoa que conhece outra pessoa (BIZER, CYGANIAK, HEATH, 2007). A Figura 5 apresenta uma tripla com uma Ligação RDF.

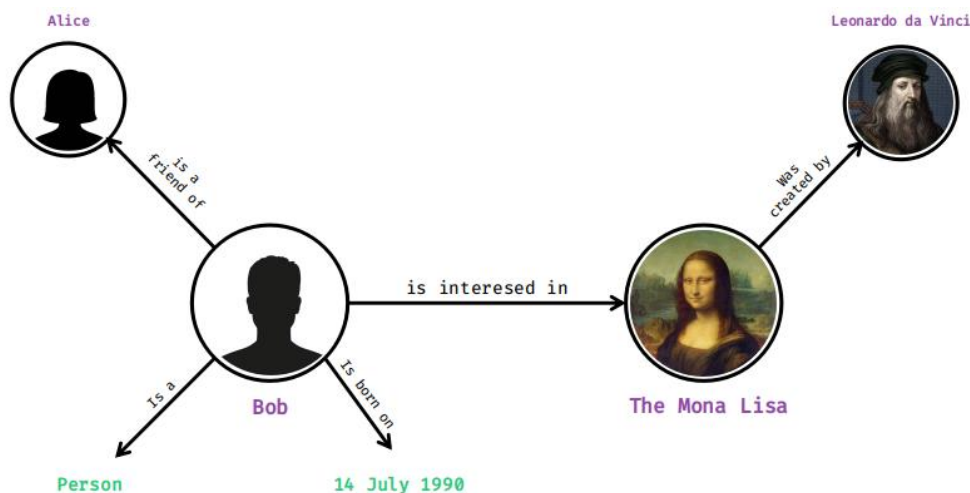
Figura 5 – Ligações RDF



Fonte: Elaborado pelo autor, adaptado de Manola, Miller, Mcbride (2014).

Triplas de ligações RDF são a base da *web* de dados, uma vez que permitem que a partir de um recurso referenciado por um URI, seja possível alcançar diversos outros recursos ligados a este e, assim sucessivamente. Ainda, é possível que um mesmo recurso possa ser referenciado múltiplas vezes em diferentes triplas, possibilitando inclusive que um recurso conste como sujeito em uma tripla e como objeto em outra. Isso permite a navegação na *web* de dados, por navegadores para dados conectados, aplicações automatizadas ou motores de busca (BIZER, CYGANIAK, HEATH, 2007). A Figura 6 apresenta um grafo em que o sujeito <Bob> possui tanto triplas literais, quanto ligações RDF. Ainda, o recurso “*The Mona Lisa*” que pode ser visto como um objeto na tripla <Bob><is interested in><The Mona Lisa>, também é o sujeito da tripla <The Mona Lisa><was created by><Leonardo da Vinci>.

Figura 6 – Grafo representando um recurso com triplas literais e ligações RDF



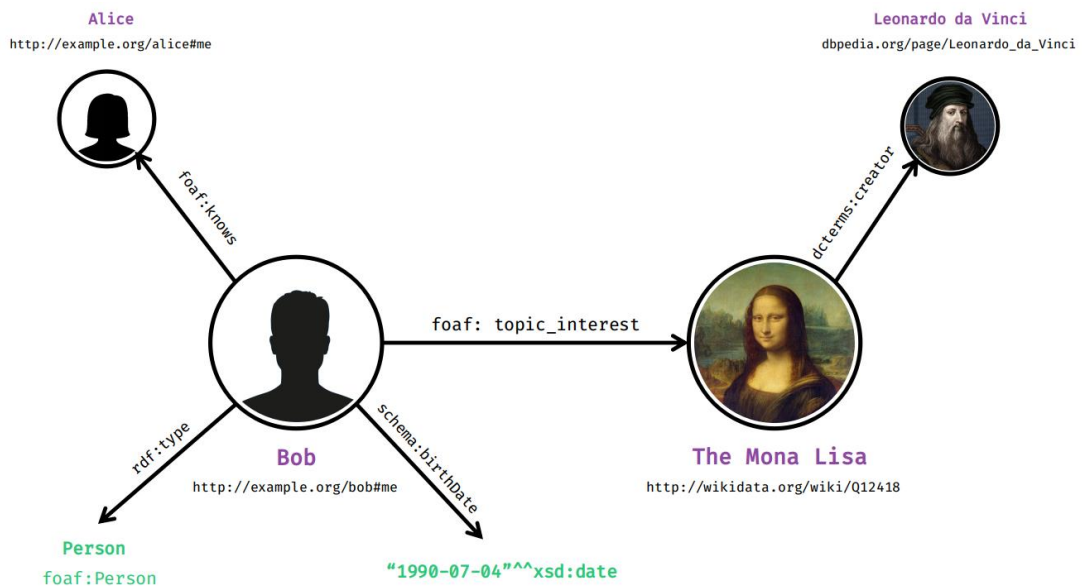
Fonte: Elaborado pelo autor, adaptado de Manola, Miller, Mcbride (2014).

Cabe destacar, conforme supracitado, que recursos que compõem as triplas devem ser representados por URIs HTTP. Desta forma, a mesma representação de tripla,

exemplificada pela Figura 6, pode ser descrita por meio dos URIs dos recursos que a compõem. Predicado e objeto podem ser representados por vocabulários de referências ou conjuntos de URIs, com informações sobre um determinado domínio (LAUFER, 2015). Ainda, o conceito de Múltiplos Grafos permite que ao se criar um documento RDF, outros grafos sejam conectados ao grafo original, proporcionando assim um catálogo de dados conectados (ISOTANI, BITTENCOURT, 2015).

Como resultado do reúso de diversos vocabulários, bem como o uso de Triplas Literais e Ligações, temos o grafo apresentado pela Figura 7.

Figura 7 – Grafo ilustrando a ligação entre conjuntos de dados RDF



Fonte: Elaborado pelo autor, adaptado de Manola, Miller, McBride (2014).

O grafo RDF ilustrado pela da Figura 7, apresenta a ligação entre dois grafos por meio do predicado `<http://xmlns.com/foaf/spec/topic_interest>`. Um navegador ou um agente de *software*, após o recebimento do URI do recurso `<http://exemple.org/bob#me>`, poderia seguir as ligações contidas neste recurso, interpretando os identificadores que compõem a descrição do mesmo. Assim, informações desse novo grafo poderiam ser acessadas e interpretadas de forma a se descobrir novas ligações RDF, permitindo a navegação sucessiva, gerando novos grafos e assim sucessivamente. Dessa forma, a partir da união de dois grafos, forma-se um novo grafo mais abrangente e com novas ligações, possibilitando a obtenção de mais informações a respeito dos recursos envolvidos.

2.1.4 Princípios estruturais para Dados Conectados

A nova perspectiva apresentada pela *Web Semântica*, bem como sua pilha tecnológica, permitiu que a *web* se desenvolvesse e evoluísse de um espaço global em que somente documentos estavam vinculados, para um espaço em que documentos e dados estão conectados, passando assim de uma “*Web* de Documentos” para uma “*Web* de Dados”. O pilar dessa evolução foi um conjunto de práticas para a publicação e a conexão de dados na *web* semântica, conhecidas como Dados Conectados (BIZER, CYGANIAK, HEATH, 2007).

O conceito de Dados Conectados refere-se a um conjunto de melhores práticas para publicação e conexão de conjuntos de dados de forma estruturada na *web*, com a finalidade de expandir a *web* clássica e construir uma “*Web* de Dados” (BERNERS-LEE, BIZER, HEATH, 2009).

O pressuposto básico por trás dessas melhores práticas é de que tanto o valor, quanto a utilidade dos dados aumentam proporcionalmente ao número de conexões estabelecidas com outros conjuntos de dados (BIZER, CYGANIAK, HEATH, 2007), propiciando inúmeras possibilidades para a integração entre conjuntos de dados, motivando o desenvolvimento de novas aplicações e ferramentas, além de possibilitar respostas mais completas à medida que novas fontes de dados surjam na *web* (ISOTANI, BITTENCOURT, 2015; BERNERS-LEE, BIZER, HEATH, 2009).

Estas melhores práticas foram delineadas por um conjunto de regras proposta por Berners-Lee em 2006, em sua nota de proposição dos Dados Conectados, como meio para que os dados passassem a fazer parte de um repositório global. Essas regras são conhecidas como “os quatro princípios” dos Dados Conectados, listados a seguir (BERNERS-LEE, 2006):

1. Use URIs como nomes para os recursos;
2. Use HTTP URIs para que esses nomes possam ser acessados;
3. Ao se acessar uma URI, informações úteis, representadas nos formatos RDF e SPARQL devem ser obtidas;
4. Inclua *links* para outras URIs de forma que outros recursos possam ser descobertos.

Estes princípios fornecem uma fórmula para a publicação e a conexão de dados utilizando a infraestrutura da *web*, de forma distinta ao que é feito na *web* clássica, no qual as relações são representadas por ligações em documentos HTML. Ainda, seguindo os princípios

de Dados Conectados, as ligações devem ser estabelecidas entre entidades, chamadas de recursos, descritas em RDF e identificadas por URIs (BERNERS-LEE, 2006; BERNERS-LEE, BIZER, HEATH, 2009).

De acordo com a terminologia definida no contexto da arquitetura da *web*, itens de um determinado domínio de interesse, tais como documentos, conceitos e objetos do mundo real, cujas propriedades e relacionamentos desejam ser representados, são chamadas de recursos.

Laufer (2015) classifica os recursos como “(...) quaisquer coisas, tanto concretas quanto abstratas. Uma determinada empresa, uma pessoa, uma página *Web* são considerados recursos. Um sentimento, uma cor, também são recursos” (p. 34).

Os recursos são identificados por URIs, mais especificamente, URIs HTTP, como meio de fornecer nomes exclusivos globalmente, sem a necessidade de um gerenciamento centralizado, e também como meio de acessar informações sobre esse recurso na *web* (BERNERS-LEE, 2006).

2.2 DADOS ABERTOS CONECTADOS

As práticas de Dados Conectados possibilitaram a criação de mecanismos necessários para a publicação, divulgação, reutilização e combinação de múltiplos conjuntos de dados na *web*. A publicação de dados seguindo tais princípios, podem levar a criação de novos conhecimentos e descobertas, que por sua vez levam a revelações inesperadas e a campos de aplicação totalmente novos (OPEN KNOWLEDGE, 2012).

Entretanto, o acesso a esses dados, por vezes, apresenta restrições relacionadas ao acesso, licenças, formatos, entre outros. Tais problemas estimularam um movimento global, principalmente, de governos para a criação dos chamados Dados Abertos (ISOTANI, BITTENCOURT, 2015). Nesse sentido, a *Open Knowledge* define Dados Abertos como:

Dados abertos são dados que podem ser livremente usados, reutilizados e redistribuídos por qualquer pessoa - sujeitos, no máximo, à exigência de atribuição da fonte e compartilhamento pelas mesmas licenças (OPEN KNOWLEDGE, 2012, n.p.).

Nesse sentido, um dado pode ser caracterizado como aberto quando sua forma de publicação e difusão *web*, seja realizada em formatos abertos, compreensíveis por agentes de *software*, podendo ser livremente utilizados de forma automatizada pela sociedade (LÓSCIO *et al.*, 2018).

O objetivo da publicação de dados em formato aberto é evitar restrições e mecanismos de controle sobre dados publicados, de forma que qualquer pessoa possa explorar esses dados livremente (ISOTANI, BITTENCOURT, 2015). Para que um dado possa ser considerado aberto é necessário obedecer aos seguintes critérios (OPEN KNOWLEDGE, 2012):

- Disponibilidade e acesso: o dado precisa estar completo, em formato não proprietário e deve ser passível de modificação;
- Reutilização e redistribuição: o dado precisa ser disponibilizado de forma que possa ser reutilizado, redistribuído e combinado com outros dados;
- Participação universal: o uso, reuso e redistribuição do dado deve ser realiza sem restrições de áreas, pessoas ou grupos.

Ainda como forma de classificar um dado como aberto, David Eaves, ativista dos Dados Abertos e palestrante de Políticas Públicas na *Harvard Kennedy School of Government*, propôs um conjunto de verificações que denominou de “três leis de dados abertos”, (EAVES, 2009, n.p., tradução nossa):

1. Se um dado não poder ser localizado e indexado na *web*, então ele não existe;
2. Se não estiver publicado em formato aberto e compreensível por máquina, então não poderá ser reutilizado;
3. Se legalmente não for possível sua replicação, então ele é inútil.

Embora as leis propostas por Eaves fossem direcionadas para Dados Abertos Governamentais, elas se mostraram aplicáveis a quaisquer Dados Abertos. Revelaram também a preocupação com forma de publicação dos dados, de modo que um dado só seria considerado aberto, caso observe-se duas premissas fundamentais: i) ser publicado em um formato aberto e; ii) ser publicado sob uma licença aberta (PORTAL BRASILEIRO DE DADOS ABERTOS, 2020).

A utilização de formatos abertos refere-se à publicação em um formato conveniente e modificável, livremente disponível (ou que ao menos permita o seu processamento por ferramentas de *software* livre/*open-source*) e que não imponha restrições monetárias, tecnológicas, ou de qualquer outro tipo, de forma a caracterizar obstáculos desnecessários para o uso desses dados (SPERONI, 2016; OPEN DEFINITION, 2009). Exemplos de formatos abertos são *Comma Separated Values* (CSV), *Extensible Markup Language* (XML), *Javascript Object Notation* (JSON) e *Resource Description Framework* (RDF). Já o uso de licenças, tem relação com a liberdade de uso, cópia e reutilização, dividindo-se em licenças de *software*, conteúdo e dados.

São muitas as áreas que podem utilizar e se beneficiar da disponibilidade de dados abertos, impactando positivamente em diversos grupos de pessoas e organizações, isto é, a sociedade como um todo, bem como o próprio governo. Ainda que não seja possível prever-se como e quais serão esses benefícios, uma vez que descobrimentos serão feitos ao acaso, já é possível destacar algumas áreas beneficiadas, tais como: transparência; inovação; participação da sociedade civil; melhoria ou criação de produtos e serviços privados; melhoria de serviços públicos; acompanhamento do impacto de políticas públicas; produção novos conhecimentos a partir da combinação de várias fontes de dados e enxergar padrões em grandes volumes de dados (OPEN KNOWLEDGE, 2012).

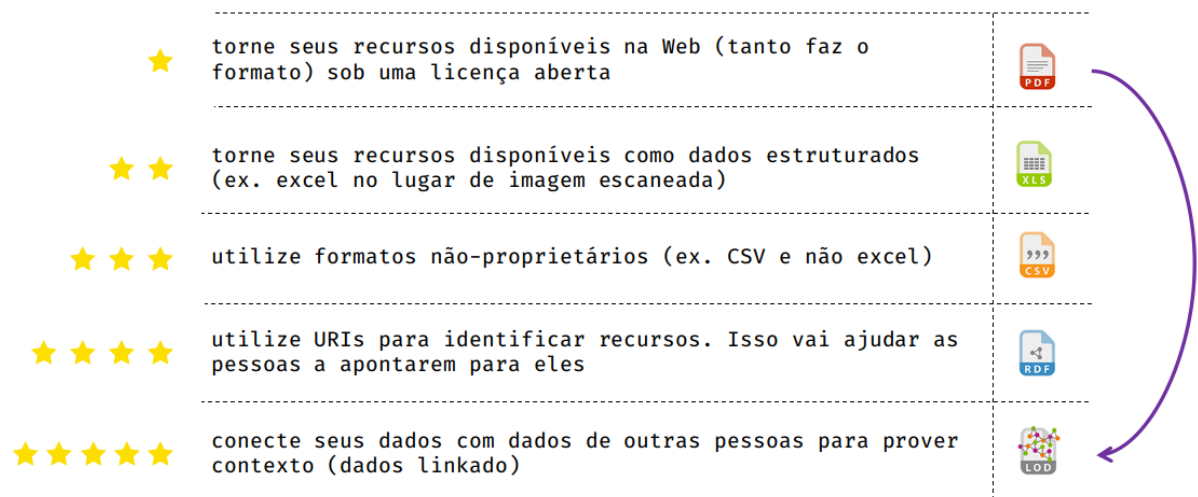
Benefícios como os supracitados, fizeram com que governos de diversos países, impulsionados pela busca de transparência, colaboração e de participação da sociedade/comunidade, passassem a publicar dados e informações sobre suas administrações em formato aberto na *web* em portais chamados Portais de Dados Abertos, os quais oferecem uma interface mais amigável para catalogação e acesso aos dados (GOLDSTEIN; DYSON, 2013). Como exemplos de Portais de Dados Abertos já consolidados, destacam-se o portal dos Estados Unidos da América e o portal do Reino Unido. Países na França, Holanda, Chile e Uruguai, também possuem portais de dados governamentais abertos. O Brasil também conta com o Portal Brasileiro de Dados Abertos, lançado em 2012 pelo Ministério do Planejamento (LÓSCIO *et al.*, 2018).

Diante desse cenário, e como forma de padronizar a publicação de Dados Abertos, seguindo os princípios de Dados Conectados, foi proposto o conceito de *Linked Open Data* ou

Dados Abertos Conectados³. Segundo Berners-Lee (2006), Dados Abertos Conectados referem-se aos Dados Conectados que são disponibilizados sob licença aberta, que não impeçam o seu livre reuso.

Como forma de encorajar a sociedade, principalmente os responsáveis por dados públicos a utilizar as práticas de publicação de Dados Abertos Conectados, Berners-Lee (2006) propôs um esquema de estrelas com o objetivo de classificar o grau de abertura e sua conexão a outros dados. Quanto mais aberto, maior o número de estrelas para os dados e mais facilidade para ser enriquecido e conectado. A Figura 8 apresenta o esquema de estrelas (ISOTANI, BITTENCOURT, 2015; 5STARDATA, 2012).

Figura 8 – Esquema de estrelas



Fonte: 5STARDATA, 2012.

De acordo com a classificação proposta por Berners-Lee, são classificados com 1 Estrela os dados publicados em qualquer formato, tais como imagens, tabelas, arquivos de processadores de textos, ou mesmo arquivos PDF, cuja licença permita sua utilização e reutilização sem qualquer restrição, inclusive comerciais. A publicação de dados em qualquer formato sob uma licença aberta representa um avanço para a abertura de dados. Entretanto, dados com 1 Estrela necessitam ser manipulados manualmente, ou por meio de algoritmos de raspagem de dados, normalmente escritos especificamente para este fim (5STARDATA, 2012; PIRES, 2015).

³ Para este trabalho o termo *Linked Open Data* (LOD) será traduzido para o português como Dados Abertos Conectados.

Dados classificados com 2 Estrelas são publicados em formatos estruturados, que permitem a manipulação de suas linhas e colunas, tais como planilhas Excel (não são admitidos o uso de imagens ou PDF), de forma que possa ser realizado o processamento automático por *software*. Dados com essa classificação, além de todos os benefícios de dados com 1 Estrela, permitem aos usuários o uso de aplicativos proprietários, como o Excel, para visualizar, agregar, combinar, fazer cálculos, e outras operações com os dados disponibilizados. Ainda, é possível exportá-los para outro formato estruturado com relativa facilidade. Entretanto, a dependência de *softwares* proprietários, pode tornar a tarefa de publicação mais complexa ou mesmo inviabilizá-la (5STARDATA, 2012; PIRES, 2015).

Dados com a classificação de 3 Estrelas devem ser publicados em formatos não proprietários, tais como o CSV, um formato estruturado que não depende de programas proprietários para ser manipulado. Dados classificados com duas estrelas, além de fornecer os mesmos benefícios das duas estrelas, permitem que os dados possam ser baixados e manipulados por qualquer um, sem a exigência de aplicativos específicos (5STARDATA, 2012; PIRES, 2015). Contudo, a publicação de dados em formato aberto pode ser onerosa, principalmente, quando o formato de origem do dado se difere do formato definido para sua publicação, sendo, eventualmente, necessárias atividades para conversão dos dados e manutenção da consistência entre o conjunto de dados primários e os dados publicados em formato aberto (LÓSCIO *et al.*, 2018).

De acordo com Pires, “(...) enquanto as três primeiras estrelas dão diretrizes para que os dados possam ser publicados na *Web* na forma de documentos de formato aberto e com licença aberta, a quarta e a quinta estrela apresentam o conceito de dados conectados” (2015, p. 30).

Nesse sentido, dados classificados com 4 Estrelas, são os dados identificados de forma única por meio de URIs. Dados com essa classificação podem ser encontrado na *web* ou mesmo localmente e conectados a outros dados, permitindo assim que seja criada uma rede abrangente de dados em formato abertos e conectado (BERNERS-LEE, BIZER, HEATH, 2009).

Por fim, dados abertos recebem a classificação 5 Estrelas quando conectados com dados já disponíveis na *web*, sendo necessário que dados que representem o mesmo conceito sejam identificados e o *link* entre eles seja estabelecido (LÓSCIO *et al.*, 2018).

O Esquema de Estrelas deixa claro a importância da abertura e da conexão de dados no processo de publicação de conjuntos de dados na *web*, uma vez que promovem condições para a interoperabilidade e o trabalho em conjunto de diferentes bases de dados, por diferentes atores da sociedade. A construção de sistemas e soluções cada vez melhores, sejam aquelas desenvolvidas na esfera governamental, na privada, na acadêmica ou na sociedade civil, dependem da interoperabilidade entre esses conjuntos de dados (PIRES, 2015).

Entretanto, a publicação de dados abertos e de dados abertos conectados, vem acompanhada de desafios técnicos, requerendo a adoção de princípios que possam materializar esses conceitos. Assim, diversas abordagens foram propostas, objetivando a publicação de conjuntos de dados de acordo com diversos aspectos, tais como: sua finalidade, origem dos dados e ferramentas disponíveis. Cada abordagem possui um conjunto de ferramentas associadas e de passos a serem seguidos dentro de um determinado fluxo de trabalho.

2.3 PUBLICAÇÃO DE DADOS NA *WEB*

Conforme discutido anteriormente, a publicação de dados na *web* de forma aberta e conectada é um desafio, diante da complexidade da tarefa. Para dar suporte a essa atividade, diversas tecnologias, metodologias e práticas foram propostas “(...) para incentivar e permitir a expansão continuada da *web* como um meio para o intercâmbio de dados” (LÓSCIO *et al.*, 2018, p. 31).

Os subtópicos a seguir apresentarão as diretrizes para publicação de dados na *web*, utilizadas neste trabalho, quais sejam: boas práticas para publicação de dados na *web*, o ciclo de dados na *web* e as tecnologias para o desenvolvimento de aplicações que dão suporte à publicação de dados na *web*, de forma conectada.

2.3.1 Práticas para Publicação de Dados na *Web*

Com o aumento no interesse na publicação dos dados em formatos abertos e conectados, aumentou também a necessidade de publicação de dados precisos, regularmente atualizados e fáceis de se vincular a outros conjuntos de dados. Contudo, os desafios impostos na publicação de conjuntos de dados seguindo tais características, tornou fundamental a

adoção de boas práticas, com o objetivo de tornar o gerenciamento de conjuntos de dados conectados mais consistente e padronizado.

Os subtópicos seguintes apresentam as Boas Práticas para Publicação de Dados na *Web* propostas por Lóscio, Burle e Calegari (2017) e as Boas Práticas para Publicar Dados Conectados proposta por Hyland, Ateazing e Villazón-Terrazas (2014), ambas, recomendações da W3C.

2.3.1.1 Boas Práticas para Publicação de Dados na Web

As Boas Práticas para Dados na *Web* (do inglês, *Data on The Web Best Practices - DWBP*), foram propostas em 2017, como uma recomendação emitida pela W3C. O documento publicado apresentou trinta e cinco boas práticas projetadas para apoiar a publicação de dados na *web* de forma que pudessem ser detectáveis e compreensíveis tanto por humanos quanto por máquinas.

O objetivo desse conjunto de boas práticas foi o de oferecer orientação técnica, de forma a aprimorar as relações entre provedores e consumidores, incentivando, conseqüentemente, a evolução e a expansão da *web*, como um espaço global de dados (LÓSCIO *et al.*, 2018).

Publicadores de dados desejam compartilhar seus dados, seja abertamente ou com acesso controlado. Os consumidores de dados (que podem ser os próprios produtores), por sua vez, desejam encontrar, usar e conectar esses dados. Para manter esse interesse e encorajá-los, é fundamental que tais dados sejam precisos, atualizados e com garantia de disponibilidade o tempo todo. Portanto, é fundamental um entendimento comum entre editores e consumidores de dados a fim de que os desejos dos consumidores sejam compatíveis com as práticas dos produtores de dados (LÓSCIO *et al.*, 2018).

Nesse contexto, é fulcral que publicadores de dados norteiem suas atividades nas DWBP, buscando aprimorar a forma como os dados são gerenciados, promovendo assim o uso e o reúso de dados, aumentando a confiança nos dados publicados, independentemente de tecnologia, potencializando assim as possibilidades de inovação (LÓSCIO, BURLE, CALEGARI, 2017).

Cabe destacar que, as DWBP englobam um universo de variáveis que vão desde o formato de publicação dos dados, até a descrição desses conjuntos de dados e o fornecimento de licenças de uso. O Quadro 3 lista as DWBP:

Quadro 3 – Boas Práticas para Publicação de Dados na *Web*

Boa Prática	Descrição
BP01	Forneça metadados
BP02	Forneça metadados descritivos
BP03	Forneça metadados estruturais
BP04	Forneça informações sobre a licença dos dados
BP05	Forneça informações de proveniência dos dados
BP06	Forneça informações sobre a qualidade dos dados
BP07	Forneça indicadores de versões
BP08	Forneça informações de histórico de versões
BP09	Use URIs persistentes como identificadores de conjuntos de dados
BP10	Use URIs persistentes como identificadores dentro de conjuntos de dados
BP11	Declare URIs para versões dos conjuntos de dados e séries
BP12	Use formatos de dados padrão legíveis por máquina
BP13	Use representações de dados neutros de localidade
BP14	Forneça dados em múltiplos formatos
BP15	Reúse vocabulários dando preferência aos padronizados
BP16	Escolha o nível de formalização certo
BP17	Forneça <i>download</i> em massa
BP18	Forneça subconjuntos para conjuntos volumosos de dados
BP19	Use negociação de conteúdo para fornecer dados em vários formatos
BP20	Forneça acesso em tempo real
BP21	Forneça dados atualizados
BP22	Forneça uma explicação para dados que não estão disponíveis
BP23	Torne os dados disponíveis usando uma API

BP24	Use padrões da <i>web</i> como base das APIs
BP25	Forneça uma documentação completa para a API
BP26	Evite alterações significativas na API
BP27	Preserve os identificadores
BP28	Avalie a cobertura do conjunto de dados
BP29	Obtenha <i>feedback</i> dos consumidores de dados
BP30	Torne o <i>feedback</i> disponível
BP31	Enriqueça os dados gerando novos dados
BP32	Forneça apresentações complementares
BP33	Forneça <i>feedback</i> para o publicador original
BP34	Siga os termos de licença
BP35	Cite o publicador original

Fonte: Adaptado de Rautenberg *et al.*, 2018, p. 55.

A adoção das DWBP supracitadas, implica em uma série de benefícios, conforme apontado por Lóscio *et al.* (2018):

A fim de incentivar os provedores a adotar as Boas Práticas para dados na *Web*, há uma série de benefícios que podem ser alcançados a partir da aplicação das Boas Práticas, são eles: compreensibilidade; facilidade de processamento; facilidade de descoberta; reuso; confiança; capacidade de conexão de dados; facilidade de acesso; e interoperabilidade (p. 44).

Essa associação entre as DWBP e os benefícios é fundamental, uma vez que permitem que publicadores de dados possam ter uma melhor compreensão das possibilidades, a partir da adoção das DWBP. A adoção de uma ou mais práticas está conectada a pelo menos um desses benefícios. A adoção da DWBP nº 17 - Forneça *download* em massa, por exemplo, orienta que arquivos grandes ou divididos em muitos URIs, devem estar disponíveis por meio de protocolos de transferência de arquivos dedicados. Esta DWBP está associada ao benefício “Facilidade de Acesso” (LÓSCIO *et al.*, 2018). O quadro 4 apresenta a associação entre as Boas Práticas de publicação de Dados na *Web* e seus benefícios.

Quadro 4 – Benefícios pelo Uso das Boas Práticas

Benefícios	Boas Práticas Utilizadas
Compreensibilidade	BP15, BP16 e BP31
Facilidade de Processamento	BP12, BP14, BP15 e BP31
Facilidade de Descoberta	BP10 e BP05
Reúso	Todas as BPs
Confiança	BP15, BP31, BP33, BP34 e BP35
Capacidade de Conexão de Dados	BP10
Facilidade de Acesso	BP17
Interoperabilidade	BP10, BP15, BP16 e BP36

Fonte: Adaptado de Lóscio, Burle, Calegari (2017).

As DWBP para publicação de dados na *web* apresentam práticas abrangentes, independentes de domínio e de aplicação, cobrindo distintos aspectos da publicação e consumo de dados, tais como formatos de dados, acesso a dados, identificadores de dados e metadados. Sendo assim, esse conjunto de boas práticas pode ser estendido ou complementado por outros conjuntos de diretrizes e boas práticas, que cubram contextos mais especializados.

Considerando-se a publicação de Dados Conectados, a W3C recomenda uma série de práticas destinadas a facilitar o desenvolvimento e a entrega de dados como Dados Abertos Conectados, tema do próximo tópico.

2.3.1.2 Boas Práticas para Publicação de Dados Conectados

Com o crescente aumento da publicação de dados, principalmente os governamentais, em formato aberto na *web*, o W3C, por meio de um grupo de trabalho criado para o estudo de Dados Conectados Governamentais, estabeleceu uma série de boas práticas para facilitar o desenvolvimento e fornecimento de Dados Abertos Conectados. De acordo com Hyland, Ateazing e Villazón-Terrazas:

Antes dos padrões internacionais de troca de dados para dados na *Web*, era demorado e difícil construir aplicativos usando técnicas tradicionais de gerenciamento de dados. À medida que mais dados governamentais abertos são publicados na *Web*, as melhores práticas também estão evoluindo (2014).

Nesse cenário, Boas Práticas foram propostas com o objetivo de facilitar o desenvolvimento e a entrega de dados abertos do governo como Dados Abertos Conectados.

Esse compilado de boas práticas apresenta um conjunto de recomendações para o gerenciamento de dados mais relevantes, e para a publicação e uso de dados de alta qualidade publicados por governos em todo o mundo, como Dados Abertos Conectados.

Ressalta-se que, embora essas boas práticas tenham sido concebidas inicialmente para o ambiente governamental, é possível estendê-las a outros segmentos. O quadro 5 apresenta essas boas práticas acompanhados de sua descrição:

Quadro 5 – Boas Práticas para Publicar Dados Conectados

Boa prática	Descrição
1) Preparar os <i>stakeholders</i>	Explicar aos <i>stakeholders</i> o processo de criação e manutenção do conjunto de Dados Abertos Conectados.
2) Selecionar um conjunto de dados	Seleção de dados que serão publicados, de forma que possibilitem o uso para distintas finalidades.
3) Modelar os dados	Construir a melhor representação dos dados e a forma como serão utilizados por distintas aplicações, independentemente da origem desses dados.
4) Especificar uma licença apropriada	Definição da licença mais apropriada para as condições de uso, com o intuito de definir termos sobre a origem, propriedade e outras condições de uso dos dados.
5) Construir boas URIs para Dados Conectados	A implementação de URIs deve ser feita considerando o referenciamento dos dados baseados em URIs HTTP. Esse planejamento deve conter: nomes de objetos, o suporte para múltiplos idiomas, o suporte a mudança de dados ao longo do tempo e a estratégia de persistência.
6) Utilizar um vocabulário padrão	Sempre que possível, devem ser usados vocabulários existentes (que podem ser estendidos se necessário) para identificar os objetos. Novos vocabulários podem ser criados, desde que necessários, seguindo as boas práticas;
7) Converter dados para representações de Dados Conectados	Esta etapa objetiva transformar dados em uma representação de Dados Conectados. Essa etapa é tipicamente apoiada por scripts ou <i>softwares</i> que automatizam essa tarefa.
8) Prover acesso automatizado para os dados	Devem ser implementados meios que permitam o acesso automatizado aos dados por motores de busca ou outros mecanismos de processamento e consumo de dados.
9) Anunciar para o público	Essa etapa envolve a tarefa de tornar público o conjunto de dados ligados. Destaca-se que a publicação desse conjunto de dados, implicitamente, gera um efeito de contrato social com o público.

10) Reconheça o contrato social	Esta etapa orienta que seja reconhecida a responsabilidade em manter os dados após sua publicação. É importante certificar-se de que o(s) conjunto(s) de dados permaneçam disponíveis ao longo do tempo.
---------------------------------	--

Fonte: Hyland, Ateazing e Villazón-Terrazas (2014).

Contudo, é importante ressaltar que, ao se construir uma aplicação *web*, não é necessário implementar todas as Boas Práticas. De acordo com Hyland e Villazón-Terrazas (2011), cada melhor prática deve ser considerada como uma medida possível que pode ser implementada com o objetivo de fornecer uma experiência mais rica e dinâmica possível para publicadores e consumidores de dados.

Baseado nesse cenário, este trabalho, que se debruçou também na modelagem e implementação de uma aplicação para apoiar a produção de Dados Abertos Conectados, observou tanto as DWBP, quanto às Boas Práticas para Publicar Dados Conectados. Visando melhor atender a adoção das Boas Práticas para publicação de Dados Abertos Conectados, foi realizada uma revisão sistemática de literatura em que, dentre outras coisas, se buscou identificar nos trabalhos selecionados quais Boas Práticas para publicação têm sido adotadas. Essa revisão sistemática compõe o capítulo 3 desta dissertação.

Além da adoção de boas práticas, é fundamental que, para se transformar dados brutos em Dados Abertos Conectados de alta qualidade e em grande escala, que cada organização estabeleça um conjunto de diretrizes comuns e claras para dar suporte a todo o ciclo de vida de publicação de Dados Abertos Conectados. O tópico a seguir apresenta alguns dos principais ciclos de vida encontrados na literatura.

2.3.2 Ciclo de Vida para Publicação de Dados Abertos Conectados

A publicação de dados na *web* pode ser realizada a partir de diversas metodologias. Cada uma dessas metodologias compreende um conjunto de atividades ou fases para modelagem do ciclo de vida de Dados Conectados.

De acordo com Visintin, Gomes e Todesco (2017), um grande desafio na publicação de dados está justamente em como publicar os dados de forma que, os conjuntos de dados gerados sejam confiáveis e de fácil conexão a outros dados. De acordo com os autores:

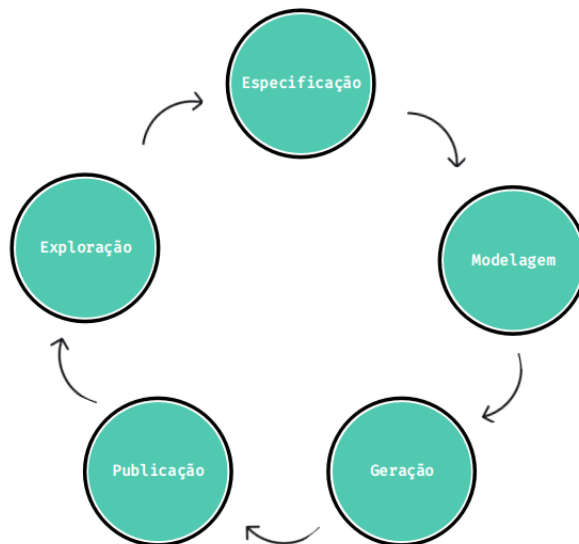
Para isso se faz necessário conhecer quais são as fases ou estágio da publicação de dados como um todo para descrever o desenvolvimento deste fenômeno, assim como para prever os próximos passos a serem realizados quando deseja-se publicar dados (2017, n.p.).

Ressalta-se que os ciclos de vida são diversos e, conforme mencionado, delimitam uma série de fases estabelecidas, que devem ser seguidas para a publicação e consumo de dados na *web*. A seguir apresenta-se alguns exemplos de ciclos de vida mais utilizados, segundo Rautenberg *et al.* (2018).

2.3.2.1 Diretrizes Metodológicas para Publicação de Dados Abertos

Villazón-Terrazas *et al.* (2011) propuseram esse ciclo de vida como resultado da experiência adquirida na produção de dados conectados em distintos contextos governamentais. O conjunto de diretrizes metodológicas para a geração, publicação e exploração consiste nas seguintes atividades: i) Especificar; ii) Modelar; iii) Gerar; iv) Publicar e; v) Explorar. A Figura 9 apresenta o ciclo de vida proposto por Villazón-Terrazas *et al.* (2011).

Figura 9 – Diretrizes Metodológicas para Publicação de Dados Abertos



Fonte: Rautenberg *et al.*, 2018, p. 43.

2.3.2.2 Dados Abertos da Irlanda

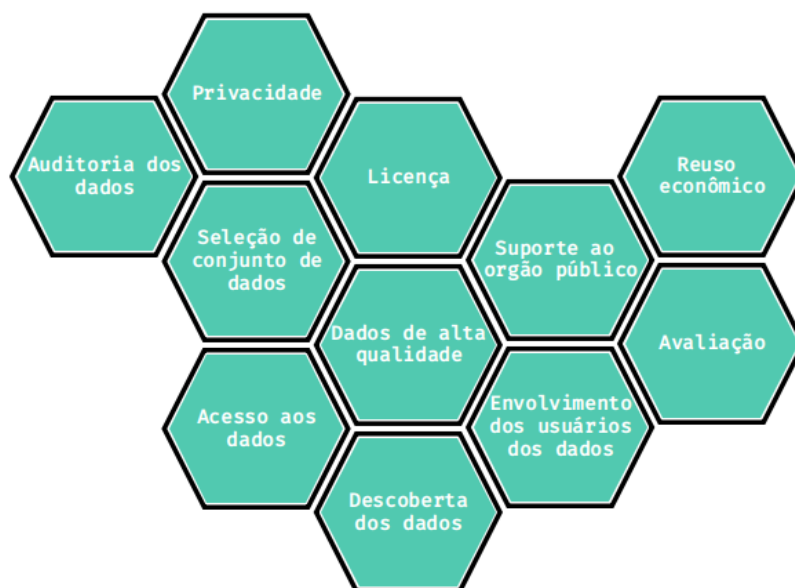
À medida que as iniciativas para abertura de dados foram amadurecendo, surgiram questionamentos acerca de quais são as melhores práticas para selecionar, publicar e apoiar a reutilização de conjuntos de dados (LEE, 2014).

Esse cenário fez com que, com base em sua jornada nacional de dados abertos, e na experiência obtida, fosse definido o *Open Data Ecosystem* (Ecosistema de Dados Abertos).

A estrutura desse ecossistema compreende não apenas aspectos técnicos, mas também aspectos sociais, políticos, econômicos e legislativos. Cada componente do ecossistema proposto representa um elemento chave de uma iniciativa sustentável de dados abertos (LEE, 2014). A metodologia utilizada para criação desse ecossistema se pautou na revisão da literatura, em entrevistas com representantes de órgãos públicos irlandeses, por meio da criação de grupos para discussão do tema com representantes da comunidade de startups, por meio da participação em encontros técnicos e por meio do *feedback* da comunidade.

De acordo com Lee (2014), uma extensa revisão da literatura das melhores práticas internacionais e a ligação com as partes interessadas, fez com que esses componentes evoluíssem. A Figura 10 apresenta os componentes desse ecossistema.

Figura 10 – Elementos do Ecosistema de Dados Abertos



Fonte: Rautenberg *et al.*, 2018, p. 44-45.

Este trabalho resultou em um roteiro concreto e holístico acerca do gerenciamento de Dados Abertos na Irlanda, composto pelas seguintes etapas que, caracterizam-se por elementos relacionados e passíveis de influência interna e externa: i) Auditoria dos dados; ii) Seleção de conjunto de dados; iii) Tratamento e mapeamento dos dados; iv) Privacidade dos

dados; v) Licença dos dados; vi) Publicação dos dados de alta qualidade; vii) Acesso aos dados; viii) Descoberta de dados; ix) Suporte de órgãos públicos; x) Envolvimento dos usuários dos dados; xi) Benefícios econômicos; e xii) Avaliação (RAUTENBERG *et al.*, 2018).

Ainda, salienta-se que, embora o foco do projeto estivesse voltado para a Irlanda, o processo adotado, as lições aprendidas e o ecossistema resultante, poderiam ser adaptados para outras situações (LEE, 2014). Por fim, apresenta-se o *Linked Data Lifecycle*.

2.3.2.3 *Linked Data Lifecycle*

O *Linked Data Lifecycle*, proposto por Auer (2014), apresenta um ciclo de vida cujo propósito é o desenvolvimento e compartilhamento de bases de dados conectados na *web*. De acordo com Rautenberg *et al.* (2018), esse processo foi aplicado com êxito em diversos projetos de publicação de dados conectados da União Europeia, devido, especialmente, ao conjunto de ferramentas computacionais que suportam suas atividades. Ainda, conforme Rautenberg *et al.* (2018), este ciclo de vida compreende as seguintes etapas:

- I. Extração – os dados não estruturados ou estruturados em diferentes formatos ou linguagens (CSV, SQL, respectivamente), ou provenientes de sistemas legados, necessitam ser mapeados para o modelo de dados RDF.
- II. Armazenamento/Consulta – o gerenciamento de dados RDF é realizado por meio do uso de Triple Stores como forma de potencializar as tarefas de publicação e consumo de dados.
- III. Revisão Manual/Autoria - as tarefas de editoração de dados e a disponibilização destes, segundo os formatos permitidos.
- IV. Interligação/Fusão - os dados de uma base são interligados a outros dados, de outras bases, ampliando os contextos de pesquisa e consulta.
- V. Classificação/Enriquecimento – seu objetivo é aumentar a expressividade e a riqueza semântica de uma base de conhecimento em relação a um contexto, representando os dados com ontologias de nível superior ou vocabulários.
- VI. Análise de Qualidade – o tratamento dos aspectos de integridade, precisão, consistência e validade de dados é realizado nesta atividade. De forma geral, verificam-se os quesitos de consistência, concisão, compreensibilidade, disponibilidade e proveniência do modelo de dados.
- VII. Evolução/Reparação - uma vez encontradas inconsistências nos dados disponibilizados ou no modelo de representação perante requisitos estabelecidos (novos ou não), ações podem ser automatizadas para a correção das não-conformidades.
- VIII. Busca/Navegação/Exploração – as técnicas de busca, navegação, exploração ou visualização são usadas para manipular dados conectados em diferentes aplicações (p. 50).

Segundo Rautenberg *et al.* (2018), as atividades apresentadas são executadas de forma incremental, devendo ser tratadas à medida que surjam os desafios na publicação de dados, como por exemplo, o mapeamento de dados legados fazendo uso de vocabulários, ou ainda, a conexão de dados locais com dados disponíveis na *web*.

De acordo com Visintin, Gomes e Todesco (2017), é fundamental, que quando adotado esse ciclo de vida, que processos sejam empregados de modo a garantir que haja uma troca de benefícios entre cada fase, auxiliando assim, na resolução dos desafios apresentados. As atividades propostas podem ser vistas na Figura 11.

Figura 11 – Ciclo de Vida de Dados Conectados



Fonte: Auer, 2014, p. 4.

As atividades do ciclo de vida proposto no *Linked Data Lifecycle* são suportadas por um ferramental tecnológico chamado de *Linked Data Stack*, cujos objetivos principais são o de facilitar tanto a distribuição, quanto a instalação de ferramentas e componentes, além de diminuir a carga de informações entre os componentes, visando assim uma melhor experiência do usuário (RAUTENBERG *et al.*, 2018). O Quadro 6 apresenta algumas das principais ferramentas propostas pela *Linked Data Stack*, organizadas de acordo com a fase do ciclo de vida em que é empregado.

Quadro 6 – Ferramentas propostas pela *Linked Data Stack*

Ferramenta	Objetivo
<i>Atividade de Extração</i>	
<i>D2R</i>	Tem como objetivo a publicação de dados provenientes de banco de dados em um grafo RDF. Possibilita ainda que consultas SPARQL sejam feitas, <i>dumps</i> em RDF sejam realizados.
<i>CSVImport</i>	Tem como objetivo importar dados em lote a partir de um arquivo CSV. Isso se dá por meio do mapeamento dos dados presentes nas colunas do arquivo CSV para RDF.
<i>Atividade de Armazenamento e Consulta</i>	
<i>OpenLink Virtuoso</i>	Trata-se <i>framework</i> usado para armazenar e realizar consultas em dados RDF. Permite o armazenamento e o acesso de grafos RDF ou XML.
<i>Sparqlify</i>	Ferramenta para reescrita de consultas SPARQL/SQL. Tem como objetivo definir visões RDF em bancos relacionais e consultá-las com SPARQL.
<i>Revisão manual e Autoria</i>	
<i>OntoWiki</i>	Trata-se de uma <i>wiki</i> semântica usada para o desenvolvimento colaborativo de ontologias em OWL, por meio de estratégias para estruturar e apresentar a informação em forma de mapas.
<i>RDFauthor</i>	Solução para edição de conteúdo estruturado e distribuído na <i>web</i> . Permite a extração de dados estruturadas na <i>web</i> de forma transparente aos usuários (detalhes do modelo de dados RDF são ocultados).
<i>Interligação e Fusão</i>	
<i>Limes framework</i>	Tem como objetivo a descoberta de <i>links</i> entre recursos na <i>web</i> . Baseado em algoritmos matemáticos, esse <i>framework</i> permite buscar e filtrar em grandes volumes de dados instâncias coincidentes;
<i>Silk framework</i>	<i>Framework</i> para integração de fontes de dados heterogêneas. Tem como objetivos principais realizar a conexão entre itens de dados relacionados, quando armazenados em distintos conjuntos de dados e transformar dados em fontes de dados estruturadas.

Fonte: Elaborado pelo autor, com base em Rautenberg *et al.* (2018).

Dentre as ferramentas da *Linked Data Stack*, destacam-se as ferramentas *CSVImport*, *Limes*, *Silk* e *Sparqlify*. Tais ferramentas tem sido amplamente empregas em distintos projetos

de produção e publicação de dados abertos conectados. Embora não tenham sido utilizados pela aplicação proposta por esta dissertação, as funcionalidades, operações e organização de cada ferramenta inspiraram a modelagem e a implementação dos algoritmos propostos por esta dissertação. Isso se deve ao fato de que a aplicação proposta, tem como um de seus princípios ser independente de domínio, de configurações de ambiente e de orquestração de ferramentas de terceiros. De acordo com Rautenberg *et al.*:

A execução dessas atividades de produção e manutenção de *Linked Data* exige conhecimento profundo de linguagens de script, várias tecnologias de serialização e publicação de RDF, ferramentas como SILK, LIMES para vinculação e padrões como R2RML para mapeamento de dados relacionais. Como resultado, esses processos são pesados, demorados e sujeitos a erros, exigindo habilidades substanciais (2015, n.p., tradução nossa).

Dessa forma, optou-se por uma implementação usando apenas dependências da linguagem Java. Assim, atividades necessárias para execução de tarefas relativas as etapas contempladas no Ciclo de Vida proposto por Auer (2014), apoiada nas Boas Prática para Publicação de Dados Abertos Conectados, serão realizadas por componentes da aplicação desenvolvida.

Todavia, para que aplicações desenvolvidas para *web* semântica possam alcançar 5 estrelas de abertura de dados, atender às diretrizes propostas pelas boas práticas mencionadas anteriormente e seguir os procedimentos metodológicos propostos pelos ciclos de vida, é necessário um ferramental tecnológico, objeto de assunto do próximo tópico.

2.4 TECNOLOGIAS PARA DESENVOLVIMENTO DE APLICAÇÕES PARA *WEB* SEMÂNTICA

Para que uma aplicação para *web* semântica seja implementada, contemplando as etapas previstas nos ciclos de vida de publicação de Dados Abertos Conectados, tais como a extração, transformação, mapeamento e conexão com outros conjuntos de dados, e observando outros aspectos fundamentais, como proveniência, reprodutibilidade e repetibilidade de fluxos de execução, é necessário o suporte de um ferramental tecnológico robusto.

Nesse sentido, nos próximos tópicos serão apresentadas tecnologias para o desenvolvimento de aplicações para *web* semântica.

2.4.1 *Linked Data Workflow Project Ontology*

Com o intuito de facilitar o gerenciamento de dados, descrever e automatizar o Ciclo de Vida de Dados Abertos Conectados, ontologias de fluxo de trabalho e proveniência têm sido empregadas. Nesse quadro, foi proposta por Rautenberg *et al.* (2016) a ontologia denominada *Linked Data Workflow Project Ontology* (LDWPO).

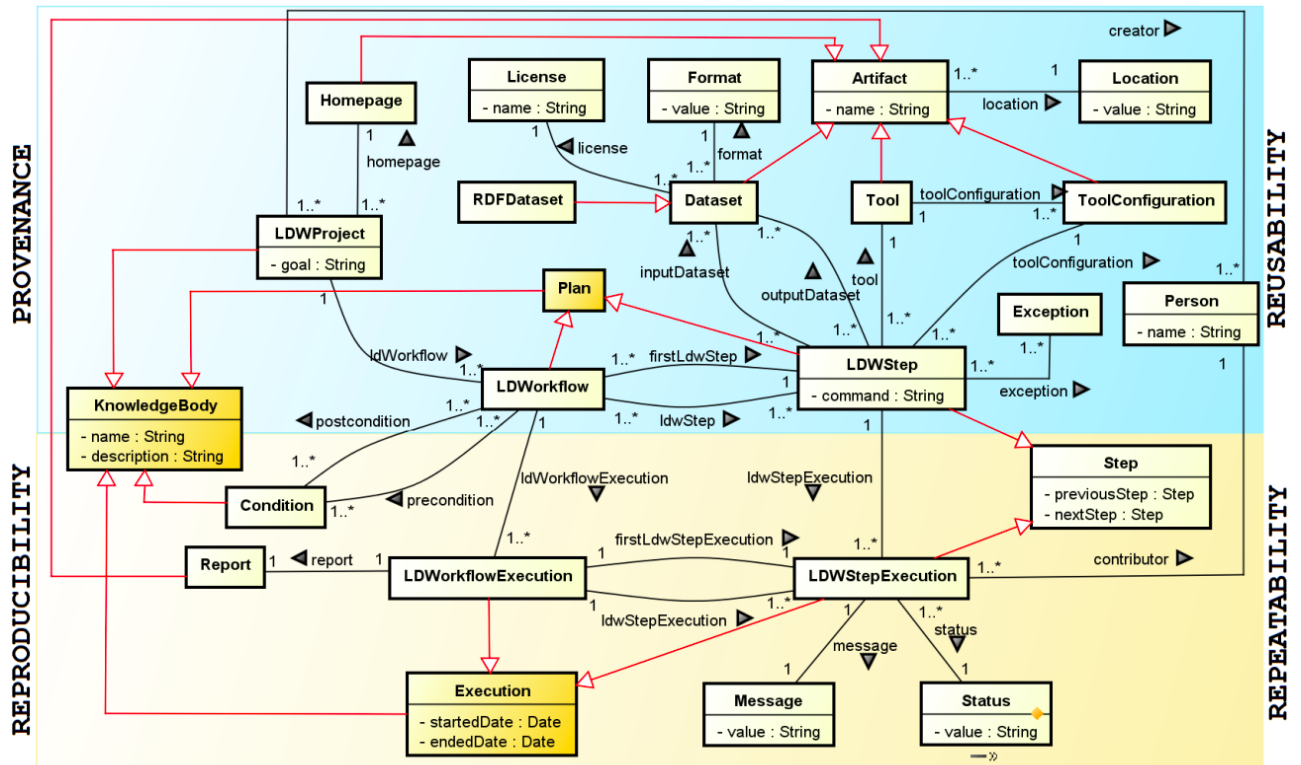
A LDWPO, retratada pela Figura 12, foi apresentada como um modelo de conhecimento para a estrutura de fluxo de trabalho para Dados Conectados, objetivando auxiliar a descrição do processo metodológico, do plano de fluxo de trabalho e das execuções dos fluxos de trabalho para produção e reprodução de Dados Conectados, possibilitando assim, a reprodutibilidade e repetibilidade das etapas de processamento de Dados Conectados (RAUTENBERG *et al.*, 2016).

Inspirada em outras fontes de conhecimento, e diante das limitações encontradas em abordagens de outras ontologias investigadas pelos autores, tais como, a falta de conceitos essenciais para descrição do plano e as perspectivas de execução de um fluxo de trabalho, a ontologia proposta estendeu de outras ontologias e vocabulários, os conceitos para o planejamento metodológico e produção, além de padronizar os conceitos de *Method*, *Plan* e *Execution*, objetivando guiar a produção e preservação de conjuntos de Dados Conectados (RAUTENBERG *et al.*, 2016a; RAUTENBERG *et al.* 2016b). Os requisitos principais da LDWPO são:

1. Descrever os métodos que estabelecem o processo, atividades e tarefas para a produção de planos de conjuntos de dados RDF;
2. Representar os planos como fluxos de trabalho para (re)produzir conjuntos de dados RDF ao longo do tempo. Isso é obtido especificando uma lista de etapas, em que cada etapa corresponde a uma chamada de ferramenta usando uma configuração de ferramenta específica, bem como conjuntos de dados de entrada e saída;
3. Apoiar a reutilização de fluxos de trabalho para orientar a (re)produção de conjuntos de dados RDF ao longo do tempo;
4. Mediar a automação de *workflows*, que envolve um ambiente controlado para a execução de um plano;

5. Preservar *logs* de execução do fluxo de trabalho para verificação da exatidão ou repetibilidade dos resultados; e
6. Gerar relatórios de projetos de conjuntos de dados RDF, seus planos de fluxo de trabalho e execuções em formatos legíveis por humanos.

Figura 12 – O Modelo *Linked Data Workflow Project Ontology*



Fonte: Rautenberg *et al.*, 2016a, p. 62.

De acordo com Rautenberg *et al.* (2016a), o escopo da ontologia está delimitado a projetos que tratem da preservação de conjuntos de Dados Conectados, e que façam uso de conceitos para descrição de planos e execução dos planos para manutenção desse tipo de conjunto de dados. A LDWPO tem como principais conceitos *LDWProject*, *LDWorkflow*, *LDWSteps* e *Task*.

O conceito *LDWProject* permite a descrição do projeto de publicação de Dados Conectados. Um *LDWProject* está associado a um ou mais *LDWorkflow*.

O conceito de *LDWorkflow* por sua vez, descreve o plano necessário para produção e reprodução de conjunto de Dados Conectados, compreendendo um ou mais *LDWSteps*.

Já o *LDWSteps* é um conceito que descreve um conjunto de procedimentos realizados em um *dataset* de entrada, utilizando para tal, uma ferramenta e sua respectiva configuração, com o objetivo de produzir outro *dataset* de saída. Um aspecto importante de um *LDWSteps* é que este pode ser reutilizado, isto é, um mesmo *LDWSteps* pode ser associado a um ou mais *LDWorkflows* dentro de *LDWProjects* existentes. Outro aspecto é que um *LDWSteps* pode ser executado automaticamente em um ambiente computacional, de acordo com a necessidade do usuário (RAUTENBERG *et al.*, 2016a).

Por fim, o conceito *Task*, representa uma prática recomendada coberta por um *LDWSteps*. Quando um *LDWSteps* é planejado, ele pode ser relacionado a um conjunto de *Tasks* necessárias ao longo de uma *LDWStepsExecution*. Exemplos de *Tasks* são: a reutilização de vocabulários, a descrição de conjunto de dados utilizando-se um vocabulário e o estabelecimento de licenças (RAUTENBERG *et al.*, 2016a).

Essa relação entre *Tasks* e *LDWSteps* é importante, uma vez que permite que engenheiros de dados descrevam seus fluxos de trabalho utilizando-se de um nível maior de abstração, tal qual no processo de desenvolvimento de *software* no contexto da Engenharia de *Software*. De acordo com Rautenberg *et al.* (2016b), o uso da LDWPO possibilita:

- a) Que os fluxos de trabalho sejam descritos como planos para produção de bases de Dados Conectados;
- b) O suporte ao processo de automatização da execução desses fluxos de trabalho;
- c) O reúso de planos, permitindo a reprodutibilidade e a repetibilidade de resultados e de bases de Dados Conectados;
- d) A documentação dos planos e de suas execuções, promovendo a proveniência da produção de bases de Dados Conectados.

Alguns trabalhos descreveram a utilização da LDWPO, como modelo de conhecimento de fluxos de trabalho para produção de Dados Conectados, tais como o *LODFlow*, proposto por Rautenberg *et al.* (2015) e *SYSLODFLOW* proposto por Morais e Morais (2016). Devido a sua importância para esta pesquisa, o *LODFlow* será descrito em outro tópico.

O *software* desenvolvido como parte desta dissertação, também se apoiou nos conceitos definidos pela LDWPO, como forma de descrever os fluxos de trabalho para

produção de Dados Conectados. Os conceitos e suas relações ainda foram observados como inspiração para criação do modelo orientado a objetos, implementado nesta dissertação.

2.5 APLICAÇÕES CORRELATAS

Conforme discutido anteriormente, todo o processo de publicação e manutenção de Dados Abertos Conectados, em razão de sua complexidade, muitas vezes torna-se uma atividade demorada e onerosa, que exige uma série de conhecimentos relacionados ao próprio processo de produção de Dados Abertos Conectados, como as boas práticas e o ciclo de vida, bem como o conhecimento profundo de ferramentas de *software* e outras tecnologias para dar suporte a essas atividades.

Em projetos como este é comum a construção de *scripts*, ou a realização de outras atividades técnicas de forma *ad hoc* para resolução de problemas.

Diante desse cenário, para auxílio na produção de Dados Abertos Conectados, algumas ferramentas foram propostas, dentre as quais se destacam a ferramenta *LODFlow*, proposta por Rautenberg e a ferramenta OpenRefine.

2.5.1 LODFlow

O LODFlow é um o sistema de gerenciamento de fluxos de trabalho que provisiona um ambiente que possibilita o planejamento, execução, reutilização e documentação de publicação e manutenção de Dados Conectados (Rautenberg *et al.*, 2015).

Este sistema foi estruturado tendo a LDWPO (descrita no tópico 2.4.1) como modelo de conhecimento, e projetado em conformidade com o *Linked Data Lifecycle* (descrita no tópico 2.3.2.3). Essa estruturação permite a descrição e a execução de fluxos de trabalho de produção e publicação de conjuntos de Dados Conectados, tendo como suporte o ferramental previsto na *Linked Stack* do *Linked Data Lifecycle*. O LODFlow ainda implementa funcionalidades que permitem a geração de relatórios e o tratamento de exceções.

De acordo com Rautenberg *et al.* o LODFlow trata-se de uma “(...) abordagem abrangente para definir, planejar, orquestrar e executar fluxos de trabalho de produção e manutenção de dados conectados” (2015, p. 137, tradução nossa).

O ambiente propiciado pelo LODFlow, toma como entrada os fluxos definidos na LDWPO, executando de forma sistemática, uma série de atividades e etapas definidas pela

ontologia, gerando inclusive relatórios legíveis por humanos e máquinas, que descrevem detalhes da execução, além de eventuais erros. O sistema ainda é capaz de realizar a execução de diversas ferramentas com distintos objetivos, tais como a extração, mapeamento, vinculação de dados e análise de qualidade (RAUTENBERG *et al.*, 2015).

Outra característica do LODFlow, é a preservação da proveniência do conjunto de dados, adicionando uma série de metadados, como por exemplo, versões e configurações das ferramentas executadas para cada instância de um fluxo de trabalho.

Para alcançar esses objetivos, o LODFlow se organiza como um conjunto de componentes com funções bem definidas:

- Modelo de Conhecimento e Base de Conhecimento de Fluxo de Trabalho de Dados Conectados;
- Componente de Manutenção de Fluxo de Trabalho de Dados Conectados;
- Mecanismo de execução de fluxo de trabalho de Dados Conectados;
- Componente de relatório de fluxo de trabalho de Dados Conectados.

Conforme Rautenberg *et al.* (2015), são muitas as vantagens de se utilizar o LODFlow dentre as quais destacam-se: a explicitude, a reutilização, repetibilidade, eficiência e facilidade de uso.

A explicitude está relacionada com a descrição detalhada, com alto nível de abstração, dos fluxos de trabalho para publicação e manutenção de conjuntos de Dados Conectados, tornando assim processos e atividades de fácil compreensão.

Já o benefício da reutilização se dá pela definição detalhada das atividades e fases de um fluxo de trabalho, oportunizando assim a reutilização dos fluxos de trabalho.

O benefício de repetibilidade por sua vez, é caracterizado pela facilidade de execução a qualquer tempo, tornando-os assim repetíveis, facilitando o teste e a confiabilidade.

A execução automática dos fluxos de trabalhos, permite a redução da intervenção humana manual, colaborando assim para uma melhora da eficiência.

Por fim, o benefício da facilidade de uso, também se relaciona com a descrição em alto nível de fluxos de trabalho, além da geração de relatórios das execuções realizadas. Isso faz com que sejam simplificadas a definição, execução e análise de fluxos de trabalho de produção e manutenção de Dados Conectados.

A adoção do LODFlow, permite que engenheiros de Dados Conectados planejem, executem e avaliem sistematicamente os fluxos de trabalho de produção e manutenção de Dados Conectados, melhorando assim a eficiência, facilidade de uso, reprodutibilidade, reutilização e proveniência.

2.5.2 OpenRefine

OpenRefine (anteriormente, Google Refine), é uma ferramenta de código aberto e gratuita, que dentre outras funcionalidades, propicia a organização, filtragem, transformação e limpeza de grandes conjuntos de dados. Sua arquitetura permite que seja utilizado de forma autônoma, sendo executado em um servidor embutido e utilizando um navegador *web* como interface com o usuário (HAM, 2013).

Essa ferramenta permite a exploração de dados de modo geral, possibilitando a limpeza, transformação e reconciliação dos dados com distintos serviços disponíveis na *web*. Entre as funcionalidades do OpenRefine, destacam-se:

- Importação e exportação um *rol* de arquivos em distintos formatos, incluindo CSV, TSV, XLS, XML, entre outros;
- Importação de arquivos tanto armazenados localmente, quanto hospedados em endereços da *web*, incluindo tabelas HTML;
- Realização de filtros e pesquisas, por dados que precisam ser alterados, permitindo a visualização e a execução de operações, apenas nessas entradas;
- Localização de entradas duplicadas, células vazias, inconsistências ou erros para correção e limpeza em massa;
- Análise dos dados contidos no arquivo;
- Desfazer e refazer ações já executadas nos dados, economizando tempo e esforço;
- Utilização da linguagem Google Refine Expression Language⁴ (GREL) como meio para transformar dados existentes ou para criar novos dados;
- Reconciliar, conectar ou estender dados com outras fontes de dados.

⁴ GREL é o Google Refine Expression Language que ajuda a definir transformações complexas.

Por se tratar de uma ferramenta complexa, sua curva de aprendizado é relativamente acentuada, principalmente para o usuário com pouca ou nenhuma experiência com programação ou expressões regulares.

Para não programadores, OpenRefine vem com uma curva de aprendizado íngreme, ao ponto que pode ser inutilizável para qualquer coisa, exceto as ações mais básicas. Para usuários com mais habilidades técnicas (ou acesso a equipe de suporte técnico), a ferramenta tem grande potencial por sua utilidade e funções que economizam tempo (HAM, 2013, p. 233, tradução nossa).

O OpenRefine, por sua capacidade de modificar e transformar muitos registros de uma vez, possibilita que, quando adotado, gere uma economia de tempo e esforço na criação e tratamento de conjuntos de dados. Entretanto, ressalta-se que seu uso depende de um poder de processamento computacional considerável, essencial para trabalhar com grandes quantidades de dados.

Devido a sua capacidade e seu código aberto, diversos projetos foram propostos tendo como base o OpenRefine. Dentre esses projetos destacam-se o LODRefine e o GraphDB OntoRefine, ambos especializações do OpenRefine para se trabalhar com Dados Conectados (OPENREFINE, 2021).

O LodRefine apresenta uma versão do OpenRefine com extensões integradas que tornam a transformação de dados tabulares para dados conectados mais fácil. As extensões utilizadas pelo LodRefine são a *RDF extension*, *DBpedia extension*, *CrowdFlower crowdsourcing extension* e *Stats extension*. Já o GraphDB OntoRefine é uma especialização do OpenRefine, integrada ao GraphDB *Workbench*. Pode ser utilizada para converter dados tabulares em RDF e importá-los para um repositório GraphDB, usando consultas SPARQL simples e um terminal virtual (OPENREFINE, 2021). A Figura 13 apresenta uma visualização da ferramenta OntoRefine.

Figura 13 – Visão geral GraphDB OntoRefine

The screenshot displays the OntoRefine interface for configuring a mapping. At the top, there are tabs for 'Configuration', 'Preview', and 'Both'. Below the tabs, there are buttons for 'Save', 'Download JSON', 'Upload JSON', 'RDF', 'SPARQL', and 'New mapping'. A table of columns is visible, including 'Trcid', 'Title', 'Short ... ption', 'Longdescription', 'Calendarsummary', 'TitleEN', 'Short ... lonEN', 'Longd ... lonEN', 'Calen ... aryEN', 'Types', 'Ids', 'Locatiennaam', 'City', 'Adres', 'Zipcode', 'Latitude', and 'Longitude'. Below this, there are buttons for 'amsterdam', 'geo', 'rdf', 'rdfs', 'schema', 'sf', and 'xsd'. The main area shows a mapping table with columns for source and target. The source column contains 'amsterdam:restaurant/' and 'a'. The target column contains 'schema: Restaurant' and 'a'. The mapping table has several rows, each with a source property (e.g., 'schema: title', 'schema: description', 'schema: latitude', 'amsterdam: zipcode', 'schema: image', 'geo: hasGeometry') and a target property (e.g., 'Title', 'TitleEN', 'Shortde ... ription', 'Zipcode', 'Media', 'amsterdam:geometry/'). The last row is highlighted in green, showing a mapping from 'geo: asWKT' to 'sf: Point' with a GREL expression '*<http: ... + *)' and a target type of 'Literal'.

Fonte: GraphDB, 2021.

O OntoRefine trouxe diversas melhorias para se trabalhar com dados conectados. Além de permitir o mapeamento de qualquer dado estruturado para um esquema RDF armazenado localmente no GraphDB, apresenta uma interface visual otimizada para enriquecimento semântico de dados, facilitando a escolha dos predicados e tipos corretos, definindo o tipo de dados para mapeamentos RDF e implementando transformações complexas usando a linguagem GREL do OpenRefine.

Este projeto se inspirou na forma de apresentação, organização e nas funcionalidades das tecnologias apresentadas, buscando facilidade de uso e produtividade. O próximo capítulo apresenta os procedimentos metodológicos utilizados nesta pesquisa.

3 PROCEDIMENTOS METODOLÓGICOS

Neste capítulo serão apresentados os procedimentos metodológicos que nortearam o desenvolvimento desta dissertação, pautados na perspectiva metodológica da *Design Science* e do método de pesquisa da *Design Science Research*.

3.1 CONTEXTUALIZAÇÃO METODOLÓGICA: A *DESIGN SCIENCE* COMO PERSPECTIVA

Ao se desenvolver uma pesquisa é necessário definir previamente quais serão as abordagens e os métodos adotados na condução do problema. Dessa forma, é preciso ter em mente alguns elementos no momento da escolha de sua abordagem epistemológica, bem como na escolha dos métodos da pesquisa. Dois elementos centrais que devem ser considerados nessa escolha são: a natureza do problema de pesquisa e os seus objetivos.

Assim, para este estudo optou-se pelo paradigma epistemológico da *Design Science*, introduzido por Herbert Simon, em sua obra “As Ciências do Artificial”, na qual o autor demarca as diferenças entre a ciência natural e a ciência do artificial, isto é, a *Design Science*, apontando os limites das chamadas ciências tradicionais na construção de proposições reais para a solução de problemas (DRESCH, LACERDA, JÚNIOR, 2015).

A escolha por tal paradigma justifica-se, uma vez que, quanto ao propósito a *Design Science* busca projetar e/ou produzir sistemas ou modificar os existentes, os chamados artefatos e, quanto ao objetivo ela é orientada a prescrição, isto é, a solução de problemas (DRESCH, LACERDA, JÚNIOR, 2015). Simon (1996) também demarca esse aspecto, ao apontar que o objetivo da *Design Science* é o de “(...) conceber artefatos para atingir objetivos” (p. 114), confluyente, portanto, com o propósito deste trabalho.

Conforme mencionado, a *Design Science* constitui-se no arcabouço epistemológico, e dela derivou-se a *Design Science Research* (DSR) (LACERDA *et. al.*, 2013). Nesse mesmo sentido, Dresch, Lacerda e Júnior (2015) afirmam que:

A Design Science é a base epistemológica quando se trata do estudo do que é artificial. A Design Science Research, por sua vez, é o método que fundamenta e operacionaliza a condução da pesquisa quando o objetivo a ser alcançado é um artefato ou uma prescrição. Como método de pesquisa orientado à solução de problemas, a Design Science Research busca, a partir do entendimento do problema,

construir e avaliar artefatos que permitam transformar situações, alterando suas condições para estados melhores ou desejáveis. Ela é utilizada nas pesquisas como forma de diminuir o distanciamento entre teoria e a prática (p. 67, grifos dos autores).

Assim, a partir de uma perspectiva científica orientada à solução prática do problema e do objetivo, na próxima seção abordaremos a DSR e as etapas aplicadas a esta pesquisa.

3.2 AS ETAPAS DE PESQUISA: A *DESIGN SCIENCE RESEARCH* COMO MÉTODO DE PESQUISA

O desenvolvimento do conhecimento científico e seu avanço são possíveis por meio da pesquisa científica, isto é, a pesquisa que se vale de rigor científico - em termos teóricos e metodológicos. Por outro lado, há outro aspecto essencial que deve ser considerado, a relevância da pesquisa. Estes dois elementos - rigor teórico metodológico e relevância, quando combinados caracterizam uma pesquisa como necessária (DRESCH, LACERDA, JÚNIOR, 2015).

Por pesquisa necessária compreende-se aquela que para além de aliar o rigor necessário à cientificidade, demonstra sua utilidade, isto é, sua aplicação prática na sociedade em termos concretos (DRESCH, LACERDA, JÚNIOR, 2015).

Alcançar o rigor teórico-metodológico só é possível na medida em que o pesquisador consegue fazer com que o problema e o método de pesquisa escolhido estejam confluentes (DRESCH, LACERDA, JÚNIOR, 2015). Ao considerar-se, portanto, o problema de pesquisa: “*Como apoiar e dar escala à produção e manutenção de conjuntos de Dados Abertos Conectados por meio da automatização desse processo?*” e ao considerarmos o objetivo da mesma: “*Propor uma aplicação web para suporte à automatização da tarefa de produção e manutenção de conjunto de Dados Abertos Conectados (Linked Open Data)*”, identificou-se que o problema proposto demanda uma solução prática, e tem como propósito implementar uma aplicação, isto é, um artefato. Dessa forma, é possível observar que problema, objetivo e métodos se encontram alinhados.

Um elemento essencial na *Design Science Research*, portanto, se referem aos artefatos, resultados, soluções práticas da pesquisa e que são definidos como: “(...) organização dos componentes do ambiente interno para atingir objetivos em um determinado ambiente externo” (DRESCH, LACERDA, JÚNIOR, 2015, p. 57). Os artefatos podem ser

classificados em: constructos, modelos, métodos, instanciações ou aprimoramento de teorias (DRESCH, LACERDA, JÚNIOR, 2015). Neste trabalho temos um artefato do tipo instanciação que segundo March e Smith (1995), é o artefato responsável pela operacionalização e implementação dos artefatos dos tipos constructos, modelos e métodos.

Cabe ressaltar, que a DSR possui distintas propostas de operacionalização do método, entretanto, neste trabalho optou-se pela apresentada por Peffers *et al.* (2007), conforme a Quadro 7 a seguir:

Quadro 7 – Etapas da *Design Science Research* e sua aplicação nesta pesquisa

IDENTIFICAÇÃO DO PROBLEMA Capítulos 1, 2 e 3
DEFINIÇÃO DOS RESULTADOS ESPERADOS Capítulo 1
PROJETO E DESENVOLVIMENTO Capítulos 2 e 4
DEMONSTRAÇÃO Capítulo 5
AVALIAÇÃO Capítulos 5, 6
COMUNICAÇÃO

Fonte: Adaptado de Peffers *et al.* (2007).

Conforme Dresch, Lacerda e Júnior (2015), as etapas podem ser definidas da seguinte forma:

- i. **identificação do problema:** justificar a importância da pesquisa, sua relevância e do problema investigado, e da aplicabilidade da solução proposta;
- ii. **definição dos resultados esperados:** estes podem ser quantitativos e/ou qualitativos;
- iii. **projeto e desenvolvimento do artefato:** definição de suas funcionalidades, arquitetura e o desenvolvimento em si;
- iv. **demonstração:** utilização do artefato, por meio de testes;

- v. **avaliação:** comparação dos resultados obtidos com os requisitos definidos na segunda etapa;
- vi. **comunicação:** divulgação da pesquisa.

Com base no exposto, nas próximas seções serão detalhadas a operacionalização das etapas propostas por Peffers *et al.* (2007).

3.2.1 Etapa 1 - Identificação do Problema

Considerando as etapas definidas por Peffers *et al.* (2007), a conscientização e identificação do problema foram realizadas por meio da definição do tema da pesquisa e da definição dos objetivos. Ainda, como meio de se aprofundar mais nas características de aplicações cujo objetivo era produção e publicação de dados abertos conectados foi realizada uma revisão sistemática sobre a temática deste trabalho e que será abordada neste tópico.

Para a construção da revisão sistemática desta pesquisa, foi definida a plataforma Scopus como ferramenta de busca, tomando por base os textos publicados a partir do ano 2010 até 2018.

Considerando-se a pergunta de pesquisa e o objetivo do presente trabalho, foram definidas as seguintes palavras-chave para a busca inicial: (1) *Linked Data*; *Linked Data workflow* e *Linked Data lifecycle*.

A escolha dos termos assinalados foi construída partindo da compreensão de que *Linked Data* seria equivalente ao termo Dados Conectados e que *lifecycle* e *workflow* atenderiam a ideia de uma série de procedimentos que ocorrem em uma determinada ordem e que são necessários para a publicação de Dados Conectados. Cabe destacar que tais termos também satisfazem o conceito de *Linked Open Data*.

Com o intuito de selecionar as publicações com maior relevância para a revisão, estabeleceram-se os seguintes critérios para o refinamento da pesquisa, além das palavras-chave: (2) limitação por área; (3) exclusão de áreas; (4) limitação de outras palavras-chave definidas pelo(s) autor(es), (5) exclusão por meio da leitura dos títulos, resumos e palavras-chave e, por fim, (6) seleção dos 15 artigos mais citados. Assim, os critérios 1, 2, 3 e 4, resultaram na *string* conforme o Quadro 8.

Quadro 8 – String de busca

TITLE-ABS-KEY (“*Linked Data*” AND “*workflow*” OR “*lifecycle*”) AND PUBYEAR > 2009 AND PUBYEAR > 2009 AND (LIMIT-TO (SUBJAREA , “COMP”) OR LIMIT-TO (SUBJAREA, “MATH”) OR EXCLUDE (SUBJAREA, “SOCI”) OR EXCLUDE (SUBJAREA, “ENGI”) OR EXCLUDE (SUBJAREA, “DECI”) OR EXCLUDE (SUBJAREA, “ARTS”) OR EXCLUDE (SUBJAREA, “MEDI”) OR EXCLUDE (SUBJAREA, “BIOC”) OR EXCLUDE (SUBJAREA, “CHEM”) OR EXCLUDE (SUBJAREA, “EART”) OR EXCLUDE (SUBJAREA, “PHYS”) OR EXCLUDE (SUBJAREA, “AGRI”) OR EXCLUDE (SUBJAREA, “BUSI”) OR EXCLUDE (SUBJAREA, “ENVI”) OR EXCLUDE (SUBJAREA, “MATE”)) AND (LIMIT-TO (EXACTKEYWORD, “*Linked Data*”) OR LIMIT-TO (EXACTKEYWORD, “*Life Cycle*”) OR LIMIT-TO (EXACTKEYWORD, “*lifecycle*”) OR LIMIT-TO (EXACTKEYWORD, “*Work-flows*”) OR LIMIT-TO (EXACTKEYWORD, “*Workflow*”))

Fonte: Elaborado pelo autor.

A Tabela 1 traz o número de textos retornados de acordo com os critérios de exclusão aplicados:

Tabela 1 – Filtragem da pesquisa

Ordem	Critério	Resultado
1	Palavras-chave: <i>Linked Data</i> ; <i>workflow</i> e <i>lifecycle</i>	255
2	Limitando as áreas: Computação e Matemática	220
3	Excluindo as áreas	178
4	Limitando as palavras-chave: <i>Linked Data</i> ; <i>life cycle</i> ; <i>lifecycle</i> ; <i>workflow</i> ; <i>workflows</i>	91
5	Leitura do título, resumo e palavras-chave	25
6	Seleção dos mais citados	15

Fonte: Elaborado pelo autor.

Consolidado o quarto critério, chegou-se ao total de 91 trabalhos relacionados pela plataforma. Por fim, a partir da aplicação dos critérios 5 e 6 foi possível consolidar a busca nos quinze trabalhos que versavam sobre a temática.

Assim, uma análise da proveniência da literatura selecionada demonstrou que no período delimitado, os anos de 2014 e 2017 são os que mais tiveram publicações sobre a área com três publicações cada. Seguindo, os anos de 2011, 2013 e 2016 possuem duas publicações cada. Os trabalhos concentram-se, essencialmente, em treze *Conference Papers* e dois artigos em periódicos. Estes por sua vez, estão distribuídos nas seguintes publicações:

- a) Lecture Notes In Computer Science Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics: três publicações;
- b) Ceur Workshop Proceedings: três publicações;
- c) ACM International Conference Proceeding Series: 2 publicações;
- d) Communications In Computer And Information Science: 1 publicação;
- e) Future Generation Computer Systems: 1 publicação;
- f) IEEE International Symposium On Parallel And Distributed Processing Workshops And Phd Forum: 1 publicação;
- g) Journal Of Web Semantics: 1 publicação;
- h) Proceedings International Computer Software And Applications Conference: 1 publicação;
- i) Works 11 Proceedings Of The 6th Workshop On Workflows In Support Of Large Scale Science Co Located With Sc 11: 1 publicação;
- j) WWW 2014 Companion Proceedings Of The 23rd International Conference On World Wide Web: 1 publicação.

A distribuição por autores observa um empate entre Auer, Garijo, Gil, Marshall e Zhao, com duas publicações cada, seguidos por Aoyama, Baierery, Bechhofer, Bizer e Boyce, com uma publicação cada.

Por fim, um ponto a ser destacado é que as publicações se concentram, em termos territoriais, no eixo estadunidense e europeu, países com alto índice de desenvolvimento econômico, com destaque para Alemanha e Estados Unidos com cinco publicações cada, seguidos por Espanha e Reino Unido com três publicações cada, Áustria, Irlanda e Holanda, com 2 publicações e por fim, Bélgica, Brasil e Canadá, com 1 publicação cada.

A partir das informações apresentadas, o quadro a seguir apresenta as principais contribuições dos quinze trabalhos selecionados:

Quadro 9 – Principais contribuições de cada artigo

Trabalho	Contribuição
<i>A New Approach for Publishing Workflows: Abstractions, Standards, and Linked Data</i>	Apresentou uma nova abordagem para a publicação de <i>workflows</i> científicos utilizando Dados Conectados representados por ontologias. Explicitou o método utilizado, permitindo a reutilização do fluxo de publicação.
<i>Emerging practices for mapping and linking life sciences data using RDF - A case series</i>	Representou os resultados de experimentos em um formato padronizado (RDF), o que possibilitou a montagem de consultas que testam hipóteses sobre drogas potencialmente úteis para o tratamento da doença de Alzheimer, com Dados Conectados. Além disso, trouxe um conjunto de boas práticas para criação e publicação de fontes de Dados Conectados.
<i>Semantically Enhanced Quality Assurance in the JURION business use case</i>	Apresentou a arquitetura e o ciclo de vida de uma aplicação que funde e conecta mais de um milhão de dados. Essa arquitetura, apresentada em detalhes, diminuiu a lacuna entre o desenvolvimento de <i>software</i> e o desenvolvimento de dados, por meio da integração de controles de qualidade no conjunto de ferramentas de <i>softwares</i> existentes.
<i>Identifying Web Tables – Supporting a Neglected Type of Content on the Web</i>	Apresentou um <i>framework</i> para extrair e analisar dados de forma automatizada de tabelas HTML. Além disso, apresentou algoritmos de aprendizado para análise da estrutura de uma tabela, bem como a geração automática de ontologias e publicação do conjunto de dados extraídos.
<i>RESTful Open Workflows for Data Provenance and Reuse</i>	Apresentou uma arquitetura que integrou todas as fases de um ciclo de vida típico de trabalho, incluindo a especificação de serviços, sua composição para os trabalhos, bem como a execução dos trabalhos. Por meio do uso de uma ontologia, especificou todos os recursos do ciclo de vida.
<i>Creating and Utilizing Linked Open Statistical Data for the Development of Advanced Analytics Services</i>	Desenvolveu uma ferramenta para facilitar a publicação e a reutilização de cubos de Dados Conectados: <i>OpenCubeToolkit</i> . Essa ferramenta integra, de forma facilitada, componentes separados que lidam com diferentes etapas do ciclo de vida do cubo de Dados Conectados, que permitem ao usuário um conjunto de funcionalidades para trabalhar com dados semânticos estatísticos.
<i>A Linked Data Approach to Sharing Workflows and Workflow Results</i>	Apresentou uma aplicação para gerenciamento de <i>workflows</i> de trabalho para preservação e criação de um equivalente digital, da seção materiais e métodos descritos em artigos científicos na área de Bioinformática.

<i>Abstract, link, publish, exploit: An end to end framework for workflow sharing</i>	Apresentou uma implementação de estrutura para publicação de fluxos de trabalho, com base em padrões, tais como: OWL, RDF e PROV. A abordagem apresentada permitiu, ainda, a publicação de outros sistemas de fluxo de trabalho.
<i>Design Management: a Collaborative Design Solution</i>	Apresentou uma aplicação para gestão de projetos. O armazenamento foi implementado em um repositório central, vinculado a outros repositórios, facilitando a colaboração no desenvolvimento dos conjuntos de dados. Além disso, descreveu o ciclo de vida de toda a geração dos conjuntos de dados.
<i>Designing the Cloud-based DOE Systems Biology Knowledgebase</i>	Apresentou uma aplicação em nuvem que permitiu o enriquecimento semântico por meio de anotações, além da publicação e compartilhamento dos dados. Ainda, trouxe a análise do fluxo de publicação de tais dados.
<i>A Linked Data Lifecycle for Smart Cities in Spain</i>	Descreveu como são aplicados os ciclos de vida de Dados Conectados, da especificação à exploração, dentro dos domínios de Cidade Inteligente. Foram apresentadas algumas abordagens relacionadas às Cidades Inteligentes que seguem o conceito de Dados Conectados.
<i>PROMIS: A Management Platform for Software Supply Networks Based on the Linked Data and OSLC</i>	Contribuiu com a apresentação de uma arquitetura de <i>software</i> (PROMIS), que forneceu uma solução para a troca de dados de gerenciamento de projetos de diferentes domínios.
<i>LinkedPipes ETL in use: Practical publication and consumption of Linked Data</i>	Apresentou o <i>LinkedPipes</i> ETL, uma ferramenta para publicação de Dados Abertos Conectados, que se concentrou, principalmente, no suporte a <i>workflows</i> de publicação de Dados Abertos Conectados, de maneira amigável. Além disso, a ferramenta também facilita o consumo de fontes de dados já existentes.
<i>A Life-cycle Workflow Architecture for Linked Data</i>	Apresentou uma proposta de arquitetura para o ciclo de vida de publicação de Dados Conectados, bem como uma visão geral, sistemática, que descreveu os principais componentes da arquitetura.
<i>LDWPO – A Lightweight Ontology for Linked Data Management</i>	Apresentou um modelo de conhecimento para fluxos de trabalho suportados por uma ontologia. Essa ontologia contempla: o processo metodológico que orienta o ciclo de vida dos conjuntos de dados RDF; o plano completo de produção do conjunto de dados e a documentação das execuções do <i>workflow</i> . Como resultado, a abordagem permitiu a reprodutibilidade e repetibilidade de etapas de processamento de dados conectados ao longo do tempo.

Fonte: Elaborado pelo autor.

Com base na tabela acima é possível observar que: a) onze trabalhos trouxeram abordagens do Ciclo de Vida de publicação de Dados Conectados; b) três trabalhos apresentaram abordagens por meio do desenvolvimento de aplicações; c) quatro trabalhos apresentaram abordagens por meio da construção de uma arquitetura de publicação; d) seis

trabalhos apresentaram abordagens para domínios organizacionais; e) cinco trabalhos apresentaram abordagens para domínios da área da saúde.

A partir da discussão apresentada no capítulo dois desta pesquisa, avaliou-se se os trabalhos aderiram aos princípios estruturais para Dados Conectados, definidos por Berners-Lee (2006), apresentados na seção 2.1.4, deste trabalho. Para fins didáticos optamos por definir três resultados possíveis para análise dos trabalhos: i) sim, quando o trabalho definiu explicitamente a aderência ao princípio; ii) não descrito, quando o trabalho não fez menção ao princípio, e iii) não, quando o trabalho explicitou a não aderência ao princípio. A seguir, apresentamos o quadro 10 em que é possível observar se o trabalho atendeu aos quatro princípios estabelecidos por Berners-Lee.

Quadro 10 – Aderência aos princípios

Trabalho	Princípio 1	Princípio 2	Princípio 3	Princípio 4
<i>A New Approach for Publishing Workflows: Abstractions, Standards, and Linked Data</i>	Sim	Sim	Sim	Sim
<i>Emerging practices for mapping and linking life sciences data using RDF - A case series</i>	Sim	Sim	Sim	Sim
<i>Semantically Enhanced Quality Assurance in the JURION business use case</i>	Sim	Sim	Sim	Sim
<i>Identifying Web Tables – Supporting a Neglected Type of Content on the Web</i>	Não descrito	Não descrito	Sim	Sim
<i>RESTful Open Workflows for Data Provenance and Reuse</i>	Não descrito	Não descrito	Sim	Não
<i>Creating and Utilizing Linked Open Statistical Data for the Development of Advanced Analytics Services</i>	Não descrito	Não descrito	Sim	Sim
<i>A Linked Data Approach to Sharing Workflows and Workflow Results</i>	Sim	Sim	Sim	Sim
<i>Abstract, link, publish, exploit: An end to end framework for workflow sharing</i>	Sim	Sim	Sim	Sim
<i>Design Management: a Collaborative Design Solution</i>	Não descrito	Não descrito	Sim	Sim

<i>Designing the Cloud-based DOE Systems Biology Knowledgebase</i>	Não descrito	Não descrito	Sim	Sim
<i>A Linked Data Lifecycle for Smart Cities in Spain</i>	Não descrito	Não descrito	Sim	Sim
<i>PROMIS: A Management Platform for Software Supply Networks Based on the Linked Data and OSLC</i>	Não descrito	Não descrito	Sim	Sim
<i>LinkedPipes ETL in use: Practical publication and consumption of Linked Data</i>	Sim	Sim	Sim	Sim
<i>A Life-cycle Workflow Architecture for Linked Data</i>	Sim	Sim	Sim	Sim
<i>LDWPO – A Lightweight Ontology for Linked Data Management</i>	Sim	Sim	Sim	Sim

Fonte: Elaborado pelo autor.

A análise dos trabalhos apresenta o seguinte quadro: a) Dos quinze trabalhos, oito aderiram aos quatro princípios; b) Seis trabalhos não mencionaram a aderência aos princípios 1 e 2, contudo, aderiram aos princípios 3 e 4; c) Um trabalho não menciona a aderência ao princípio 1 e 2, adere ao princípio 3, e não adere ao princípio 4.

Para a análise da aderência às boas práticas optamos por definir quatro resultados possíveis para análise dos trabalhos: i) sim, quando o trabalho definiu explicitamente a aderência; ii) não descrito, quando o trabalho não fez menção; iii) não, quando o trabalho explicitou a não aderência, e iv) parcialmente, quando foi evidenciado de que não foi atendido totalmente. Assim, de acordo com a Tabela 2, temos o seguinte cenário:

Tabela 2 – Aderência às boas práticas

Boas práticas	Sim	Não descrito	Não	Parcialmente
Selecionar um conjunto de dados	15	-	-	-
Modelar os dados	15	-	-	-
Especificar uma licença apropriada	4	11	-	-
Construir boas URIs para Dados Conectados	8	-	7	-
Utilizar um vocabulário padrão	9	5	-	1

Converter dados para representações de Dados Conectados	15	-	-	-
Prover acesso automatizado para os dados	15	-	-	-
Anunciar para o público	12	-	3	-

Fonte: Elaborado pelo autor.

De acordo com a análise do quadro 9, onze dos quinze trabalhos analisados explicitaram ou descreveram ciclos de vida para publicação de Dados Conectados. O Quadro 11 descreve os ciclos de vida apresentados por cada um dos trabalhos.

Quadro 11 – Ciclos de vida apresentados

Trabalho	Ciclo de Vida apresentado
<i>A New Approach for Publishing Workflows: Abstractions, Standards, and Linked Data</i>	i) Geração de fluxo de trabalho; ii) Conversão; iii) Publicação; iv) Compartilhamento; v) Reutilização;
<i>Emerging practices for mapping and linking life sciences data using RDF - A case series</i>	i) Seleção das fontes de dados ou partes dos mesmos para ser publicados como RDF; ii) Identificação de URL's persistentes; iii) Personalização manual do mapeamento, se necessário; iv) Fazer a ligação de conceitos no novo mapeamento RDF para conceitos em outras fontes de dados RDF; v) Expor os dados RDF através de um <i>endpoint</i> ou com o SPARQL; vi) Publicação de aplicativos da <i>Web semântica</i> usando os dados publicados.
<i>Semantically Enhanced Quality Assurance in the JURION business use case</i>	i) Extração; ii) Armazenamento; iii) Criação; iv) Interligação; v) Enriquecimento; vi) Análise da qualidade; vii) Reparação; viii) Publicação.
<i>Identifying Web Tables – Supporting a Neglected Type of Content on the Web</i>	i) Busca de páginas da <i>Web</i> relevantes para serem processados; ii) Extração das informações; iii) Determinação da relevância da tabela; iv) Expor a estrutura da informação encontrada; v) Identificação do intervalo de dados da tabela; vi) Mapeamento dos resultados extraídos para vocabulários e ontologias existentes.
<i>A Linked Data Approach to Sharing Workflows and Workflow Results</i>	i) Recuperação; ii) Revisão; iii) Repetição; iv) Reutilização; v) Reposicionamento; vi) Conservação.

<i>Abstract, link, publish, exploit: An end to end framework for workflow sharing</i>	i) Especificação: identificação das fontes de dados e adoção de convenções para nomenclatura de URIs; ii) Modelagem: definição dos vocabulários usados para representar os dados, de acordo com os requisitos e cenários; iii) Geração: processo de transformar os dados de seus formatos heterogêneos para um formato estruturado (RDF); iv) Publicação: o conjunto de dados resultante e seus metadados são disponibilizados por meio de uma loja <i>triplestore</i> ; v) Exploração: os benefícios do conjunto de dados podem ser observados por meio de consultas ou mesmo aplicações que os consomem.
<i>Designing the Cloud-based DOE Systems Biology Knowledgebase</i>	i) Descoberta; ii) Acesso; iii) Integração de dados orientados por semântica.
<i>A Linked Data Lifecycle for Smart Cities in Spain</i>	i) Especificação; ii) Modelagem; iii) Geração; v) Publicação; vi) Exploração.
<i>PROMIS: A Management Platform for Software Supply Networks Based on the Linked Data and OSLC</i>	i) Obtenção; ii) Transformação; iii) Publicação.
<i>A Life-cycle Workflow Architecture for Linked Data</i>	i) Extração de dados de origem; ii) Transformação em RDF; iii) Alinhamento com vocabulários comumente usados; iv) Vinculação a outros conjuntos de dados; v) Publicação na <i>Web</i> ; vi) Carregamento em uma <i>triplestore</i> ; vii) Registro do conjunto de dados em um catálogo de dados como CKAN.
<i>LDWPO – A Lightweight Ontology for Linked Data Management</i>	i) Busca, Pesquisa; ii) Extração; iii) Armazenamento; iv) Revisão Manual; v) Interligação, combinação; vi) Enriquecimento; vii) Avaliação de qualidade; viii) Evolução/reparo.

Fonte: Elaborado pelo autor.

A análise dos 11 trabalhos que apresentaram um ciclo de vida mostra que:

- a) 10 trabalhos descrevem a execução da etapa de Extração ou atividade análoga;
- b) 4 descrevem a execução da etapa de Armazenamento e Consulta ou atividade análoga;
- c) 6 descrevem a execução da etapa de Revisão manual e autoria ou atividade análoga;
- d) 7 descrevem a execução da etapa de Interconexão ou atividade análoga;
- e) 2 descrevem a execução da etapa de Classificação e Enriquecimento ou atividade análoga;
- f) 2 descrevem a execução da etapa de Análise de Qualidade ou atividade análoga;

- g) 2 descrevem a execução da etapa de Evolução e Reparo ou atividade análoga;
- h) 9 descrevem a execução da etapa de Pesquisa/Navegação/Exploração ou atividade análoga.

Além da revisão sistemática aqui apresentada, esta pesquisa também está fundamentada em uma revisão bibliográfica de fontes primárias e secundárias, apresentada no capítulo dois deste trabalho.

3.2.2 Definição dos Resultados Esperados

Foram definidos como resultados esperados da aplicação aqui desenvolvida, os itens a seguir: i) automatização da produção de Dados Abertos Conectados; ii) Produzir dados de acordo com as boas práticas de publicação de dados abertos conectados; e iii) dar escala à produção de Dados Abertos Conectados. Estes resultados serão retomados no capítulo cinco, quando da avaliação da aplicação e dos resultados obtidos a partir dos testes efetuados.

3.2.3 Projeto e Desenvolvimento do Artefato

A etapa projeto e desenvolvimento do artefato será apresentada no próximo capítulo, onde serão demonstradas todas as etapas de desenvolvimento da aplicação.

3.2.4 Demonstração, Avaliação e Comunicação

As etapas de demonstração e avaliação dos resultados serão objeto do capítulo cinco, em que serão abordados os resultados. Já a etapa de comunicação é materializada pela execução da própria dissertação e sua apresentação. O próximo capítulo consistirá na apresentação do desenvolvimento da aplicação.

4 DESENVOLVIMENTO DA APLICAÇÃO

A Engenharia do Conhecimento tem apoiado a implementação de projetos para Gestão de Conhecimento, tais como sistemas de conhecimento e aplicações para processamento e busca de informações (ABEL, FIORINI, 2013). Este capítulo apresenta o processo de planejamento, projeto e implementação de uma aplicação para suporte à produção de dados em formato aberto e conectado, conforme previsto na etapa de Projeto e Desenvolvimento do Artefato na metodologia. O capítulo apresenta também o escopo do projeto, explicitando os aspectos cobertos pelo desenvolvimento desta aplicação, bem como suas restrições e limitações. Serão apresentados os fluxos implementados, além das principais tecnologias e algoritmos utilizados.

4.1 CONTEXTUALIZAÇÃO

Diante do que foi apresentado nos capítulos anteriores evidenciou-se que a produção de Dados Abertos Conectados é uma tarefa complexa. Esse contexto fez com que fossem propostos métodos, técnicas e ferramentas para suporte à execução desta tarefa. Contudo, a orquestração de todo o aparato tecnológico veio acompanhada de grandes desafios, alguns já citados anteriormente, dentre os quais destacam-se:

- i. Aplicações construídas com caráter *ad hoc*, atendendo a produção de conjuntos de dados específicos;
- ii. Exigência de conhecimento técnico e domínio de distintos *softwares*, que suportem as diversas etapas necessárias para publicação do conjunto de dados em formato aberto e conectado;
- iii. Dificuldade na repetibilidade e reprodutibilidade quanto à produção de conjuntos de dados, sem a necessidade de reconfiguração do ambiente de produção.

Esse cenário motivou o desenvolvimento da aplicação, objeto desta dissertação, intitulada de *Easy, General and Creative Flow* - EGCFlow. O EGCFlow teve como objetivo, a sistematização do processo de produção de dados abertos e conectados, por meio da criação de fluxos de trabalho, em que conjuntos de dados primários são carregados, transformados,

semantificados e conectados a outras fontes de dados, automatizando e dando escala à produção de conjuntos de dados no modelo de metadados RDF, subsidiando assim, processos distintos tanto na esfera pública, quanto na privada.

Para alcançar esse objetivo, o processo de desenvolvimento pautou-se nas boas práticas da Engenharia de *Software*. Segundo Pressman e Maxim (2016), ainda que não exista um processo de *software* ideal, os processos existentes podem ser adaptados de acordo com o contexto do *software* a ser implementado. Para esta aplicação, embora o contexto e os requisitos iniciais fossem razoavelmente bem definidos, alguns aspectos não eram bem claros ou mesmo conhecidos. Isso levou a adoção de um modelo de processo iterativo e incremental de desenvolvimento. Esse modelo foi composto das seguintes etapas metodológicas: definição do escopo, levantamento de requisitos, modelagem e implementação.

Cada etapa metodológica foi realizada por meio da execução de atividades fundamentadas nos princípios de Engenharia de *Software*, nos princípios estruturais e nas Boas Práticas para publicação de Dados Abertos Conectados.

Todo o processo de desenvolvimento preocupou-se ainda, em proporcionar ao usuário final, fácil instalação e uso. A abordagem adotada permitiu o mapeamento semântico orientado pelo usuário, possibilitando que grandes quantidades de dados tabulares pudessem ser convertidas para o modelo de dados RDF, até mesmo por usuários casuais.

A interface implementada, guarda semelhanças comuns a muitos programas de gerenciamento de planilhas, além de não ser necessária a configuração de diversas ferramentas ou arquivos externos, tendo como únicas dependências: o *Java Runtime Environment*⁵, em sua versão 8 ou superior; o sistema gerenciador de banco de dados *MySQL*⁶, em sua versão 5.7 ou superior; e um navegador *web*.

Cabe destacar que, a modelagem aqui apresentada foi inspirada e realizada: i) observando-se as etapas necessárias à produção de dados em um formato aberto e conectado a outros recursos externos, previstas no Ciclo de Vida para Publicação de Dados Abertos Conectados, descrito no tópico 2.3.2; ii) observando-se as classes e suas relações, propostas na ontologia LDWPO, descrita no tópico 2.4.1; e, iii) observando-se e realizando-se a experimentação da definição de fluxos de trabalho, bem como as atividades e etapas

⁵ <https://www.oracle.com/br/java/technologies/javase-jre8-downloads.html>

⁶ <https://dev.mysql.com/downloads/mysql/5.7.html>

realizadas pelo LODFlow, descrito no tópico 2.5.1. O tópico seguinte apresenta o processo de definição de escopo da aplicação e o levantamento dos requisitos.

4.2 DEFINIÇÃO DO ESCOPO E LEVANTAMENTO DOS REQUISITOS

O sucesso de um projeto de *software* está intimamente ligado à adoção de práticas de gerenciamento de *software*. Tais práticas são realizadas por um conjunto de atividades chamadas de “planejamento de projeto” (PRESSMAN, MAXIM, 2016). Uma das etapas do planejamento de projeto segundo os autores é a definição do escopo. Essa etapa é prevista pela metodologia DSR, como uma das atividades da fase de Projeto e Desenvolvimento do Artefato.

O escopo de um projeto de *software* está relacionado com a definição da cobertura e das funcionalidades que serão implementadas, indicando o que e como serão atendidas. A definição e delimitação do escopo é fundamental para se alcançar os objetivos do projeto, de acordo com o planejado.

Como parte do processo de definição do escopo foi estabelecido um conjunto de requisitos que nortearam o desenvolvimento da aplicação, permitindo sua compreensão de forma abrangente. Assim, atendendo ao objetivo específico de “definir os requisitos funcionais para uma aplicação de publicação de dados abertos conectados”, foram elicitados os seguintes requisitos:

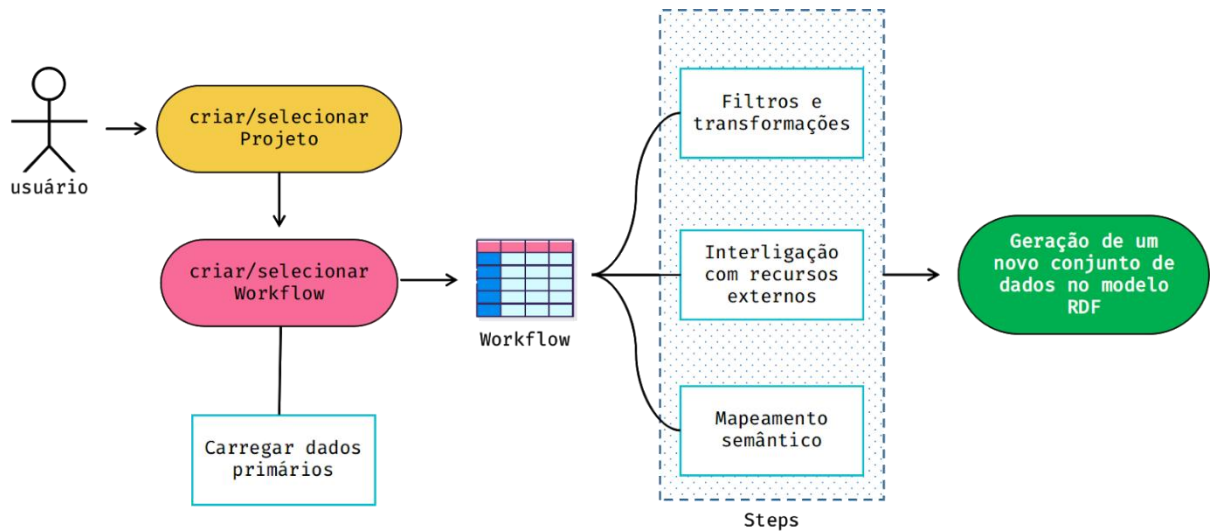
- Manter Projeto — A aplicação deve ser capaz de permitir a criação e manutenção de um Projeto para publicação de conjuntos de Dados Abertos Conectados. O Projeto terá como base a classe LDWProject, da ontologia LDWPO;
- Manter *Workflow* (fluxo de execução de atividade) — A aplicação deve permitir a criação e manutenção de um plano que possibilite a execução de um fluxo de trabalho para publicação de um conjunto de Dados Abertos Conectados, baseado na classe LDWorkflow, da ontologia LDWPO;
- Carregar *Dataset* — A aplicação deve permitir que um conjunto de dados semiestruturados seja carregado no *Workflow*. O *Dataset* será baseado na classe LDWorkflow na ontologia LDWPO;

- Manter *Steps* (atividade sobre o conjunto de dados) — A aplicação deve permitir a criação e manutenção de um conjunto de passos como parte de um fluxo de trabalho para publicação de conjunto de Dados Abertos Conectados, baseado na classe LDWSteps, da ontologia LDWPO;
- Executar *Steps* — A aplicação deve possibilitar a criação de um conjunto de passos de execução de um determinado *Workflow*, baseado na classe LDWStepsExecution da ontologia LDWPO;
- Importar *Steps* — A aplicação deve possibilitar que um *Workflow* importe os *Steps* já realizados em outros *Workflows*;
- Executar *Workflow* — A aplicação deve permitir a execução de um determinado fluxo de trabalho (*Workflow*), baseado na classe LDWorkflowExecution, da ontologia LDWPO;
- Conectar dados — A aplicação deve permitir que os recursos presentes no conjunto de dados sejam interligados com outras fontes de dados, por meio do uso de ontologias;
- Semantificar dados — A aplicação deve possibilitar que os recursos presentes no conjunto de dados sejam enriquecidos semanticamente, por meio do uso de vocabulários e ontologias;
- Exportar — A aplicação deve possibilitar que o conjunto de dados de origem seja exportado de acordo com os *Steps* executados. Como resultado, um conjunto de dados em formato aberto e conectado deve ser gerado. Esse novo conjunto de dados deve estar serializado no formato RDF.

Assim como na LDWPO, um projeto é associado a um ou mais *Workflows*. Cada *Workflow* pode possuir um conjunto de *Steps*, de modo que sua execução de forma sistemática produza um novo conjunto de Dados Abertos Conectados. Um *Step*, por sua vez, representa uma operação sobre o conjunto de dados, realizando alterações no conjunto de dados ou no procedimento de exportação para o formato RDF.

A definição do escopo possibilitou que fosse delineado o fluxo de uso da aplicação. Esse fluxo, simplificado, pode ser observado na Figura 17.

Figura 14 – Fluxo simplificado da aplicação



Fonte: Elaborado pelo autor.

Na abordagem apresentada, o usuário deve criar ou selecionar um Projeto existente. O objetivo do Projeto é manter um conjunto de dados conectados, a partir do carregamento de dados primários. Um Projeto pode estar associado a um ou mais *Workflows*.

Um *Workflow* por sua vez, está associado, principalmente, a um conjunto de dados primários e a um conjunto de *Steps*. Um *Step*, assim como na LDWPO, representa uma operação sobre o conjunto de dados primários, com o intuito de transformar, limpar, conectar a outros recursos e semantificar os recursos presentes nesse conjunto de dados.

Após a execução de um ou mais *Steps*, o usuário poderá exportar o conjunto de dados resultante para um novo conjunto de dados de acordo com o modelo RDF.

A definição dos requisitos e delimitação do escopo, permitiram que fosse realizada a etapa de definição da arquitetura da aplicação, tema do próximo tópico.

4.3 TECNOLOGIAS E DEFINIÇÃO DA ARQUITETURA

Como uma das atividades da etapa de Projeto e Desenvolvimento de *Software*, está contemplada a definição da arquitetura da aplicação. A escolha da arquitetura de *software* é uma tarefa capital e deve ser orientada ao domínio do problema. A construção de ferramentas flexíveis, de qualidade e que atendam aos requisitos propostos no tópico anterior, dependem da escolha de um modelo arquitetural adequado.

Bass, Clements e Kazman (2012) descrevem arquitetura de *software* como uma coleção de estruturas necessárias para um sistema de *software*, compreendendo seus componentes, as propriedades desses componentes e os relacionamentos entre eles. Rotem-Gal-Oz (2012), por sua vez, caracteriza a arquitetura de *software* como a descrição dos componentes, atributos e suas interações, com o objetivo de produzir *software* de qualidade.

Assim, a escolha da arquitetura de *software* deve ser realizada considerando os componentes do *software*, suas interações e limites, tendo em conta que tal escolha orientará a implementação desses componentes durante o ciclo de vida de desenvolvimento. As próximas seções descrevem as principais componentes da arquitetura utilizada.

4.3.1 Arquitetura REST

A arquitetura de *software* tem um papel fundamental na qualidade do *software*, afetando o desempenho, robustez, capacidade de distribuição e de manutenção de um sistema (BOSCH, 2000 *apud* SOMMERVILLE, 2011).

O REST é um estilo arquitetural de *software* que estabelece um conjunto de regras e restrições para a criação de Serviços *Web* (FIELDING, 2000). Esse modelo arquitetural tem ênfase nas interações entre componentes com alta escalabilidade, na independência entre componentes, na redução da latência, no reforço à segurança e no encapsulamento desses componentes, orientando-se pelos conceitos do protocolo HTTP, tais como, o uso adequado dos métodos HTTP e de URL's (FIELDING, 2000; SAUDATE, 2013).

O REST foi desenvolvido tendo como pilares um conjunto de restrições arquiteturais que podem ser compreendidas como um conjunto de regras criadas para estabelecer as características da arquitetura. Cada restrição é uma decisão de projeto que pode ter impactos positivos e negativos para o design da arquitetura. A intenção é que os aspectos positivos de cada restrição equilibrem os aspectos negativos. Dentre elas destacam-se (FIELDING, 2000):

- Cliente-Servidor: a troca de informações ocorre entre os agentes em rede, cliente e servidor. O cliente realiza requisições que devem ser respondidas pelo servidor. Essa separação permite que os componentes da aplicação evoluam de forma independente;

- Ausência de Estado (*stateless*): cada requisição deve conter todos os parâmetros necessários para ser atendida. O servidor não deve guardar nenhum tipo de estado entre requisições (anteriores ou futuras);
- Cache: implementação de estratégias de cache para otimização do desempenho (redução da latência média entre interações) da aplicação REST. Cada resposta deve informar clientes ou elementos intermediários, tais como *proxies*, *gateways* e balanceadores de carga, objetivando estratégias de cache mais adequadas. Caso uma resposta possa ser armazenada em cache, o cliente poderá reutilizar os dados de resposta para solicitações posteriores que sejam equivalentes;
- Interface Uniforme: cliente e servidor devem realizar troca de mensagens por meio de interfaces uniformes. Quando usado com o HTTP, por exemplo, é possível utilizar os verbos HTTP (*post*, *get*, *put*, *delete*, etc.) para requisições, e utilizar os códigos HTTP como meio para que os clientes possam lidar com respostas de forma padronizada;
- Sistema em camadas (*Layered*): a adoção do estilo de sistemas em camadas permite a organização hierárquica dos componentes em camadas, restringindo o acesso do componente a camada imediata com a qual está interagindo. Camadas podem ser adicionadas para melhorar o desempenho do sistema (por exemplo, implementando *caching*).

O REST tem sido amplamente adotado como o método para construção de APIs⁷ nos últimos anos, permitindo que desenvolvedores criem aplicações de todos os tipos para a *web*. Isso se deve, em grande parte, pela relação do REST com o protocolo HTTP. De acordo com as diretrizes estabelecidas pelo REST, a troca de mensagens entre aplicação e servidor, deve ser realizada utilizando-se os princípios estabelecidos pelo HTTP (VATWANI, 2020). Para exemplificar, a Figura 15 apresenta os quatro métodos HTTP e suas operações equivalentes para obter os recursos necessários do servidor.

⁷ API é um conjunto de definições e protocolos, utilizado no desenvolvimento e na integração de *software* de aplicações. API é um acrônimo em inglês que significa interface de programação de aplicações (REDHAT, 2021).

Figura 15 – Métodos HTTP e suas operações

Método HTTP	Operação	Propósito
GET	Buscar ou obter	Buscar um ou mais registros
POST	Criar	Criar um registro com um novo identificador
PUT	Atualizar	Atualizar um recurso atribuindo a ele novos valores
DELETE	Apagar	Apagar um registro

Fonte: Adaptado de Vatwani (2020).

As informações manipuladas pela abordagem REST são organizadas em recursos *web*, que constituem um conjunto coeso e mínimo de dados (FIELDING *et al.*, 2017).

Um recurso pode ser definido como um objeto que possui um tipo, dados associados, relacionamentos com outros recursos e um conjunto de métodos correspondentes aos padrões do protocolo HTTP, que fazem operações sobre ele (FIELDING *et al.*, 2017). O acesso a um recurso ocorre por meio de outro elemento fundamental do protocolo HTTP, o URI, descrito no item 2.1.1. Uma interface que descreve uma coleção de recursos acessíveis pela rede é denominada de REST API (FERREIRA *et al.*, 2017).

Entre as razões para adoção de uma REST APIs, destaca-se a integração de dados. Seu uso possibilita que diferentes aplicações troquem informações com um nível mais baixo de acoplamento, se comparado ao compartilhamento de dados por meio de base de dados, por exemplo, que tem sua aplicação restrita a uma determinada organização. O uso de REST APIs, por sua vez, permite que sejam compartilhadas informações em escala global na *web*, por meio de seus recursos.

Como forma de representação de recursos, o JSON tem sido amplamente aplicado (RICHARDSON *et al.*, 2013). Trata-se de uma forma de representação de dados estruturados em uma coleção de pares no formato de *chave:valor* (CROCKFORD, 2006). O JSON ganhou popularidade para intercâmbio de informações, por sua natureza independente de linguagem de programação, de fácil criação, manipulação e análise (JIANG; QIU; ZHAO, 2020).

A utilização da arquitetura apresenta diversos benefícios aos desenvolvedores, como por exemplo, escalabilidade, suporte a diversos tipos de representação, flexibilidade,

independência entre componentes, além de ser altamente adaptável às plataformas, ambientes e linguagens de programação. Tais características, endossaram o emprego de uma REST API como parte do modelo arquitetural da aplicação proposta. Para possibilitar o uso deste modelo arquitetural, foi utilizado o *Spring Project*, um conjunto de módulos que dão suporte ao desenvolvimento de APIs REST. O Spring é apresentado no próximo tópico.

4.3.2 *Spring Project*

O *Spring Project*, ou somente *Spring* é um conjunto de projetos *open source* que podem atuar de forma orquestrada, para resoluções de problemas comuns no desenvolvimento de aplicações, apoiando a implementação de soluções com simplicidade e flexibilidade. Foi projetado para que o foco do desenvolvimento de aplicações estivesse concentrado mais nas regras de negócio e menos na infraestrutura (JOHNSON *et al.*, 2018).

Os projetos do ecossistema do *Spring* cobrem diversas áreas de desenvolvimento de aplicações, tais como: o gerenciamento de operações com bancos de dados; o gerenciamento de aspectos de segurança; e o gerenciamento de aplicações distribuídas, *mobile* e *web*. Sua flexibilidade permite que as aplicações sejam implementadas em distintas arquiteturas, de acordo com suas necessidades.

O principal projeto do *Spring* é o *Spring Framework*, base para todos os outros projetos do ecossistema. Trata-se de um *framework* organizado em módulos, apoiado nos princípios de inversão de controle e injeção de dependências⁸. O fato de ser modular, possibilita que cada aplicação utilize apenas os módulos que forem convenientes (SPRING, 2021; WEISSMANN, 2014). Dentre os módulos do *Spring Framework* destacam-se o Spring MVC e o *Spring Data*.

O Spring MVC é um *framework* para criação de aplicações *web* e serviços REST. Foi projetado em torno do padrão arquitetural *FrontController*, em que as requisições são concentradas em um *Servlet* central, o *DispatcherServlet*, que então as roteia para os outros componentes da aplicação.

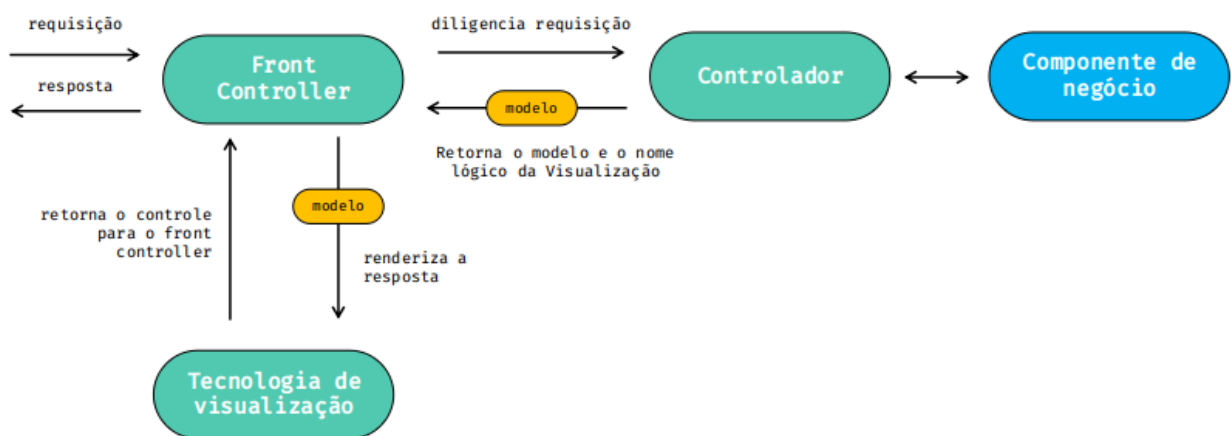
Dessa forma, o Spring MVC é capaz de, entre outras coisas, atender as requisições HTTP, delegar responsabilidades de processamento de dados para outros componentes e

⁸ Na injeção de dependências, não é a classe cliente a responsável por definir quais serão suas dependências. Esta responsabilidade é delegada a um container de injeção de dependências.

retornar a resposta adequada. O fluxo de trabalho de processamento de solicitação do *Spring Web MVC* é ilustrado pela Figura 16.

A utilização do *Spring MVC* proporciona inúmeros benefícios para a aplicação, tais como, a separação clara de funções e a configuração direta de classes por meio de anotações (JOHNSON *et al.*, 2018).

Figura 16 – Fluxo de requisição do *Spring*



Fonte: Adaptado de Johnson *et al.*, 2018, p. 421.

Outro módulo suportado pelo *Spring Framework* é o *Spring Data*. Seu objetivo principal é unificar e facilitar o acesso a diferentes tipos de armazenamento de persistência, sejam eles relacionais ou não. Além disso, o *Spring Data* oferece um conjunto de interfaces genéricas que facilitam a realização de operações triviais na maioria das aplicações, tais como a inclusão ou exclusão de um registro em uma base de dados (JOHNSON *et al.*, 2018).

A utilização do *Spring Project*, por sua estrutura robusta e flexível, confere diversos benefícios. Quando implementado, possibilita que aspectos não funcionais sejam responsabilidade do *framework*. Ainda, facilita a observação das boas práticas de programação, permitindo o desenvolvimento de aplicações complexas de forma simples (JOHNSON *et al.*, 2018). Além disso, sua estrutura modular possibilita que sejam utilizados de acordo com a necessidade específica (JOHNSON *et al.*, 2018).

O projeto desenvolvido nesta dissertação fez uso do *framework Spring*. Além dos benefícios supracitados, tais como sua arquitetura modular e seu conjunto de componentes, que permitem o foco no problema a ser desenvolvido, a opção pelo *framework* também se deu

em razão de possibilitar o desenvolvimento utilizando-se a linguagem Java. Esse foi um requisito basilar, uma vez que o *framework* necessário para a manipulação de triplas RDF, o Apache Jena, também é implementado nessa mesma linguagem.

Como meio para dar suporte ao gerenciamento do uso de ontologias, vocabulários e outros aspectos de uma aplicação para *web* semântica, foi utilizado o *framework* Jena, apresentado no próximo tópico.

4.3.3 Apache Jena

Ao se construir sistemas baseados em ontologias, é fundamental o uso de ferramentas que viabilizem a manipulação de triplas RDF e a conexão com bancos de dados RDF. De acordo com Isotani e Bittencourt, essas ferramentas “(...) funcionam como *middlewares* entre a aplicação e a ferramenta *triplestore* (2015, p. 150)”. Nesse cenário, o *framework* Jena tem se destacado como um dos mais populares (ISOTANI, BITTENCOURT, 2015).

O Apache Jena, ou Jena, é um *framework* de código aberto, desenvolvido na linguagem de programação Java, voltado para o desenvolvimento de aplicações para *web* semântica e dados conectados (APACHE JENA, 2021). De acordo com Isotani e Bittencourt:

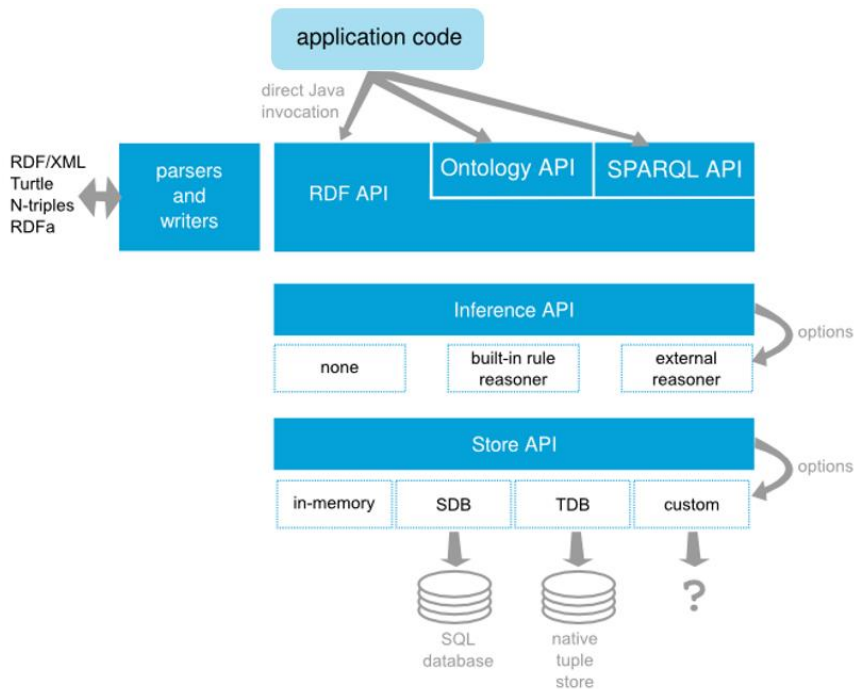
Jena também é uma coleção de ferramentas e bibliotecas Java com o objetivo de suportar o desenvolvimento de sistemas baseados na *Web Semântica*. A ferramenta inclui: uma API para leitura e escrita de dados RDF em arquivos; uma API para manipulação de ontologias em OWL e RDFS; um motor de inferência baseado em regras para raciocínio; mecanismos de armazenamento de grandes volumes de dados em triplas RDF; e um motor de consulta conforme a nova especificação do SPARQL (2015, p. 151).

Cada API que compõem o *framework* realiza uma tarefa bem definida, que interagem entre si, colaborando assim, para o processamento de conjuntos de dados RDF. Uma das principais contribuições do Jena é a manipulação de grafos RDF. O *framework* possui um conjunto de ferramentas que permitem realizar a análise de dados representados em RDF, consultas, e o armazenamento de grafos RDF na memória ou em armazenamentos persistentes em forma de triplas. A Figura 17 apresenta a interação entre essas diferentes camadas propostas pelo Jena e suas API's.

O *software* desenvolvido como parte desta pesquisa, fez uso principalmente das API's RDF e de Ontologia. Essas API's fornecem classes que possibilitam a representação de componentes de ontologias e vocabulários tais como, recursos, propriedades e literais, além

do próprio grafo RDF. Também é possível realizar a criação, leitura e escrita de grafos RDF com suporte para distintos formatos, tais como, RDF/XML e *Turtle* (APACHE JENA, 2021).

Figura 17 – Interação entre as camadas e as APIs propostas pelo Jena



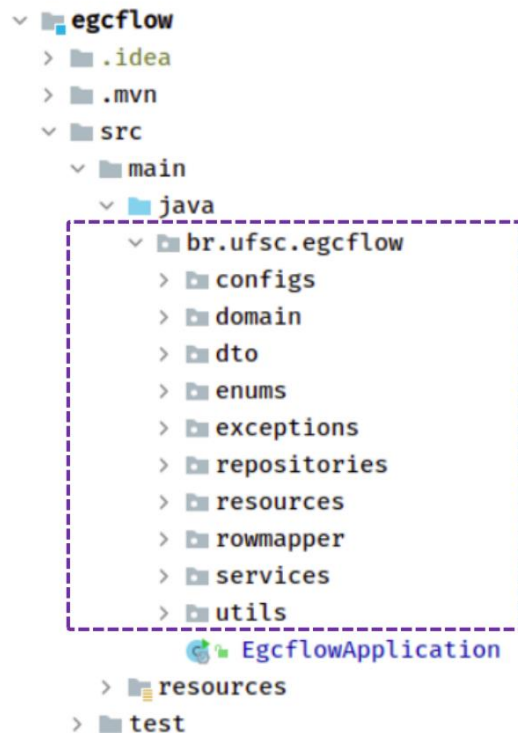
Fonte: Apache Jena, 2021.

A API de Ontologia fornece os recursos necessários para a manipulação de ontologias, como ontologias em RDF, RDFS e OWL, realizando a abstração do modelo ontológico para um conjunto de classes muito abrangente, que no caso deste trabalho, permite a interação com a ontologia LDWPO, por exemplo.

4.3.4 Arquitetura da aplicação

Objetivando melhorias na escalabilidade e modularização, a arquitetura proposta foi organizada em camadas lógicas, cada uma com uma função específica, de acordo com o princípio da separação de interesses (SHVETS, 2019). A Figura 18 apresenta, em destaque, as camadas utilizadas.

Figura 18 – Representação, em destaque, das camadas implementadas



Fonte: Elaborado pelo autor.

Essas camadas foram organizadas de modo a concentrar classes que possuem comportamento em comum, isto é, que executam funcionalidades com responsabilidades similares para a composição da ferramenta. A seguir é apresentada uma descrição sintetizada das responsabilidades das classes contidas nas camadas, com maior relevância para a aplicação:

1. *br.ufsc.egcflow.domain*: esta camada agrupa classes que representam no modelo de orientação a objetos, as classes contidas na ontologia LDWPO que foram utilizadas na ferramenta, além de outras classes de interesse para o domínio do problema;
2. *br.ufsc.egcflow.dto*: esta camada agrupa classes que representam abstrações de entidades de domínio, com o objetivo de melhor representar os recursos;
3. *br.ufsc.egcflow.repositories*: esta camada agrupa classes que tenham como responsabilidade a manipulação de objetos armazenados em um banco de dados;

4. *br.ufsc.egcflow.resources*: esta camada agrupa classes que representam as interfaces para acesso aos recursos da REST API. É responsabilidade destas classes receber uma requisição HTTP, interagir com as regras de negócio da ferramenta e devolver uma resposta;
5. *br.ufsc.egcflow.service*: esta camada agrupa classes que implementam as regras de negócio, tais como a importação e exportação de dados, regras de mapeamento semântico, manipulação de ontologias e de triplas RDF e a execução dos *Steps*;
6. *br.ufsc.egcflow.util.utils*: esta camada agrupa classes utilitárias para realização de operações internas da ferramenta, tais como a criação de pastas e a normalização de *strings*.

As camadas *br.ufsc.egcflow.domain* e *br.ufsc.egcflow.service* apresentam classes que no modelo orientado a objetos representam conceitos descritos pela ontologia LDWPO. Tais classes foram nomeadas com o prefixo “LDW”. Os atributos presentes nas classes do modelo de conhecimento da LDWPO foram observados e respeitados, sempre que possível. A Figura 19 exibe as classes implementadas.

Figura 19 – Classes do modelo orientado a objeto



Fonte: Elaborado pelo autor.

É importante destacar que algumas classes foram excluídas da Figura 19, a fim de facilitar sua visualização.

Além da arquitetura descrita, esta aplicação utilizou-se de algoritmos e rotinas computacionais para alcançar seu objetivo. Essas abordagens serão descritas no tópico seguinte.

4.4 ALGORITMOS E ROTINAS COMPUTACIONAIS

Com o intuito de lidar com o conjunto de atividades necessárias para a transformação do conjunto de dados primários em um conjunto de dados abertos conectados, foram adotadas abordagens utilizando-se de algoritmos, estratégias e tecnologias para cada fim. As seções a seguir detalham as abordagens utilizadas.

4.4.1 Seleção, extração e armazenamento do conjunto de dados de entrada

O ciclo de vida para produção e manutenção de conjuntos de dados abertos conectados compreende um conjunto de etapas. Uma das primeiras etapas é a seleção do conjunto de dados primários, cujo objetivo é a extração de dados primários em diferentes formatos e linguagens ou provenientes de sistemas legados, que serão mapeados para o modelo de dados RDF (RAUTENBERG *et al.*, 2018).

Para a aplicação desenvolvida nesta dissertação, o procedimento adotado foi extrair as informações desses conjuntos de dados e armazená-las em um banco de dados relacional MySQL. De acordo com Rautenberg, Burda e Souza (2018, p. 117), realizar essa operação “(...) facilita os processos futuros de manipulação e conversão da representação dos dados em outros formatos”. Entretanto, a execução dessa tarefa apontou dois importantes desafios: o primeiro deles relaciona-se com a extração de dados, em razão de sua estrutura.

É possível classificar os dados quanto a sua organização e apresentação como estruturados, não-estruturados e semiestruturados. São classificados como dados estruturados, dados organizados e representados por meio de estruturas rígidas, previstas previamente (esquema) para armazená-los, por exemplo, um banco de dados. Já os dados não-estruturados, são aqueles sem um formato específico, apresentando uma estrutura flexível e dinâmica ou mesmo sem qualquer estrutura, como por exemplo, um documento de texto, um livro, um arquivo de áudio ou uma imagem. Por fim, dados semiestruturados são dados que apresentam

uma representação estrutural heterogênea, não sendo nem completamente não-estruturados, nem estritamente tipados. Exemplos de dados semiestruturados são os arquivos CSV, JSON e XML.

Devido à complexa organização e formatação dos dados não-estruturados e a alta disponibilidade de dados semiestruturados na *web* (MAHMUD *et al.*, 2018), delimitou-se, para esta aplicação, a importação de dados semiestruturados, disponibilizados nos formatos CSV, TSV e JSON. Cabe ressaltar que, a abordagem utilizada para extração dos dados de entrada foi modelada para ser flexível quanto aos formatos de entrada, possibilitando a implementação e suporte para outros formatos de dados em versões futuras. Dessa forma, à medida que novas implementações para algoritmos de importação forem realizadas para outros formatos de entrada, a própria aplicação define a melhor estratégia para atender esses tipos específicos durante sua execução.

O segundo desafio, por sua vez, relaciona-se com a organização e armazenamento de dados altamente diversos e heterogêneos. Em esquemas convencionais, conceitos e atributos que representam um recurso qualquer são previamente conhecidos, podendo ser modelados como uma tabela no modelo relacional. Por exemplo, ao se realizar a modelagem de um banco de dados hipotético, cujo domínio fosse o sistema eleitoral, seria possível a modelagem de uma tabela *<candidato>* que tivesse como atributos, o nome do candidato, o cargo para o qual concorre, seu número e seu nome de urna.

Entretanto, essa premissa não é verdadeira para esta aplicação, uma vez que conjuntos de dados primários de distintos domínios, com dezenas ou até centenas de atributos descritivos aplicáveis, podem ser carregados, demandando facilidade no acesso. Tais requisitos resultaram em desafios significativos de modelagem e implementação.

Uma solução para esse problema é realizar a modelagem dos dados usando a abordagem conhecida como *Entity Attribute Value* (EAV) (CHEN *et al.*, 2000). A abordagem EAV oferece muitas vantagens, incluindo sua flexibilidade e capacidade de armazenar dados heterogêneos em um formato simples e de fácil manutenção. De acordo com Divu, Nadkarni e Brandt (2006), o modelo EAV pode ser entendido conceitualmente como uma tabela de banco de dados com três colunas, em que: uma das colunas representasse a identificação da entidade ou objeto; a segunda coluna representasse um atributo ou um ponteiro para uma tabela separada de atributos; e, por fim, uma terceira coluna, contendo o valor correspondente à entidade e atributo em particular. Contudo, essa abordagem apresenta diversos

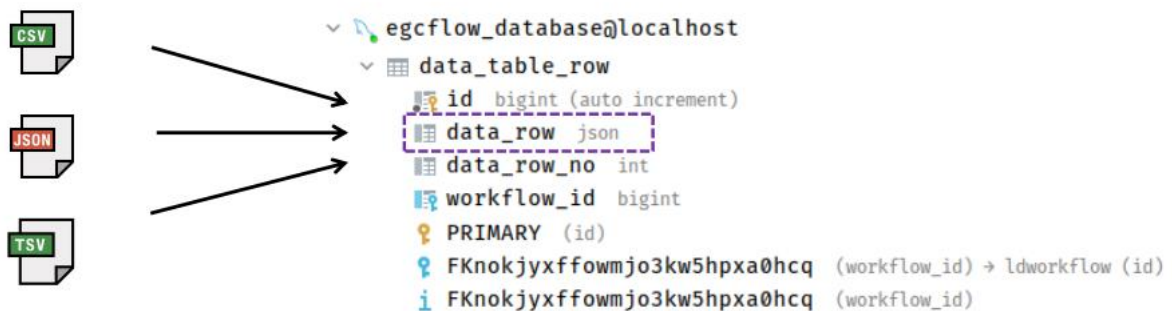
inconvenientes, sendo o principal deles, a baixa eficiência em consultas quando comparada com esquemas de banco de dados convencionais.

Outra abordagem possível é a utilização de colunas dinâmicas implementadas nos principais bancos de dados atualmente disponíveis. No MySQL, o tipo de dados JSON, disponível desde sua versão 5.7, pode ser utilizado para esse fim. O uso de colunas dinâmicas permite que a aplicação incorpore novos atributos sem que o esquema de banco de dados tenha que ser alterado (MySQL, 2021).

Essa abordagem possibilita, dentre outros benefícios, o armazenamento de diversos dados em uma única (ou poucas) tabelas, simplificando o esquema físico do banco de dados, além de fornecer uma maneira eficiente de acessar todos os dados relacionados a um objeto específico, isto é, consultas centradas na entidade, por meio de um conjunto específico de funções. Esta foi a abordagem adotada neste trabalho.

Para operacionalizar esta atividade, as informações extraídas dos dados de entrada foram organizadas como objetos que representam uma linha da tabela apresentada ao usuário em cada *Workflow*. Cada objeto é então armazenado em uma tabela que, dentre outras colunas possui uma do tipo JSON. A Figura 20 apresenta a tabela `<data_table_row>`, que armazena dados extraídos dos dados de entrada.

Figura 20 – Tabela para persistência dos dados extraídos



Fonte: Elaborado pelo autor.

A coluna `<data_row>`, destacada na Figura 20, organiza e armazena, no formato JSON, as linhas ou objetos extraídos dos conjuntos de dados de entrada. Os dados são descritos por um conjunto *chave:valor*, em que a chave representa o cabeçalho de uma coluna e o valor, o dado dessa coluna.

A abordagem utilizada realizou o processamento dos arquivos de entrada em lotes, possibilitando que o carregamento de conjuntos de dados grandes (maiores que 100MB), pudessem ser realizados sem grandes dificuldades técnicas.

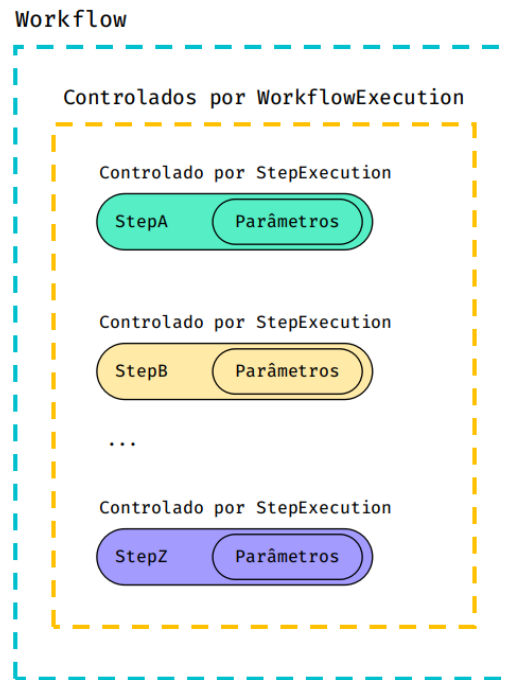
Essa forma de persistência possibilitou também que qualquer que fosse o conjunto dos dados de entrada, operações sobre esse conjunto de dados, tais como, o processo de mapeamento e exportação para modelo de dados RDF, pudessem ser realizadas de forma eficiente. Essas operações, aqui denominadas como *Steps*, são apresentadas no tópico a seguir.

4.4.2 *Steps*

A criação de um *Step*, baseado no conceito LDWSteps, da LDWPO tem como objetivo a execução de uma ou mais operações sobre o conjunto de dados de entrada, utilizando-se algoritmos em Java, consultas a banco de dados de triplas ou relacionais, além de outras operações a fim de produzir um conjunto de dados seguindo o modelo RDF.

A implementação dos *Steps* foi fundamentada em um padrão arquitetural que permitiu o encapsulamento de parâmetros necessários para realizar a execução, a reversão ou o agendamento de uma ação para execução automatizada.

Os parâmetros incluem toda informação necessária para execução da ação, como o objeto que possui o método a ser executado e os valores para os parâmetros do método. Essa estratégia possibilitou que a execução, inclusão e reaproveitamento de *Steps* fossem simplificadas. Isso significa que, ao se realizar um *Step* em um *Workflow*, esse mesmo *Step* pode, eventualmente, ser executado mais de uma vez ou mesmo ser utilizado por outros *Workflows*. A abordagem adotada, ainda permitiu que praticamente qualquer operação pudesse ser implementada como um *Step*, de forma a usufruir os benefícios mencionados acima. Uma visão geral dessa abordagem pode ser visualizada na Figura 21.

Figura 21 – Esquema de gerenciamento de *Steps*

Fonte: Elaborado pelo autor.

A Figura 21 exibe o esquema para Execução do Conjunto de *Steps*. Um *Step* tem sua execução controlada por *StepsExecution*, que sabe lidar com seus parâmetros e execução. Um *WorkflowExecution*, por sua vez, orquestra a execução dos *StepsExecution*. Os tópicos seguintes apresentam alguns *Steps* implementados.

4.4.3 Interligação com a DBpedia

O ciclo de vida proposto por Auer (2014), tem como uma de suas etapas a Interligação ou Fusão. Nesta etapa, um conjunto de dados é interligado a outro, com o intuito de ampliar os contextos de pesquisa e consulta.

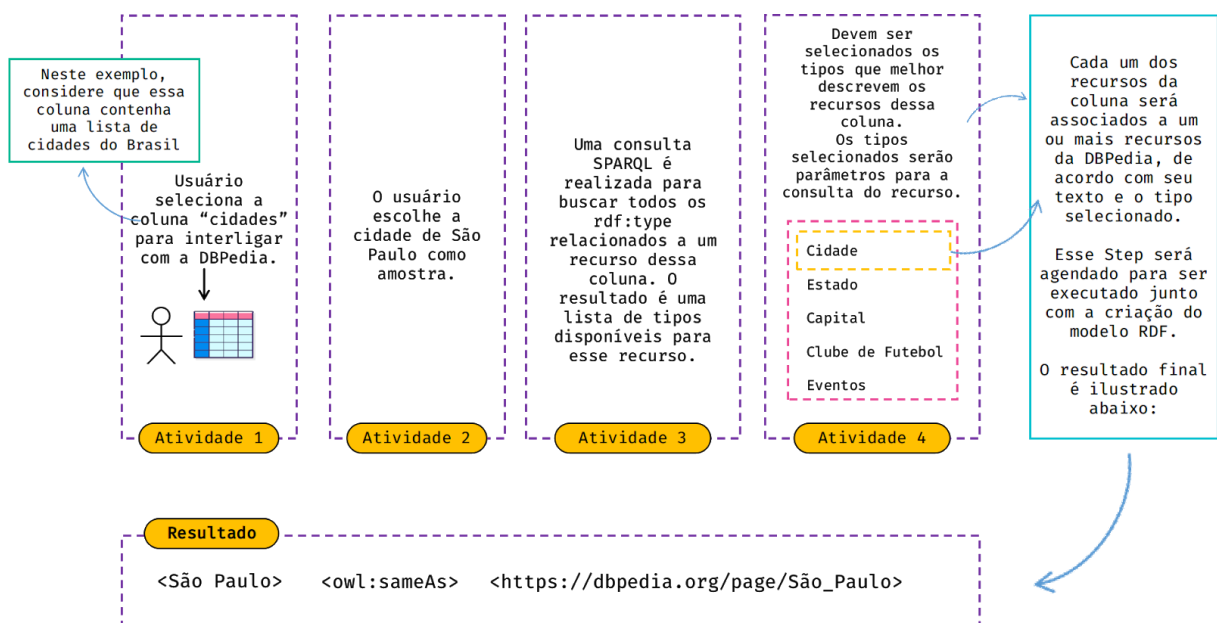
Assim, este *Step* teve como objetivo vincular os recursos presentes no conjunto de dados primários a outros conjuntos de dados (*mashup* semântico), nesse caso a DBpedia⁹, de modo que a união entre esses conjuntos de dados contribuísse para criação de dados mais completos. Além da interligação entre conceitos ser um dos princípios da *web* semântica e

⁹ A DBpedia é um dos principais nós da *web* de dados. Foi utilizada para este projeto como ponto de conexão para dados abertos conectados.

viabilizar a chamada “*Web de Dados Conectados*”, a interligação faz com que o conjunto de dados resultado alcance cinco estrelas, de acordo com o modelo proposto por Berners-Lee, conforme citado no item 2.2.

A execução deste *Step* é composta por um conjunto de atividades bem definidas, ilustradas pela Figura 22. Neste exemplo, o propósito é interligar dados de uma coluna contendo nomes de cidades do Brasil aos respectivos recursos presentes na DBpedia.

Figura 22 – Representação dos passos para interligação com a DBpedia



Fonte: Elaborado pelo autor.

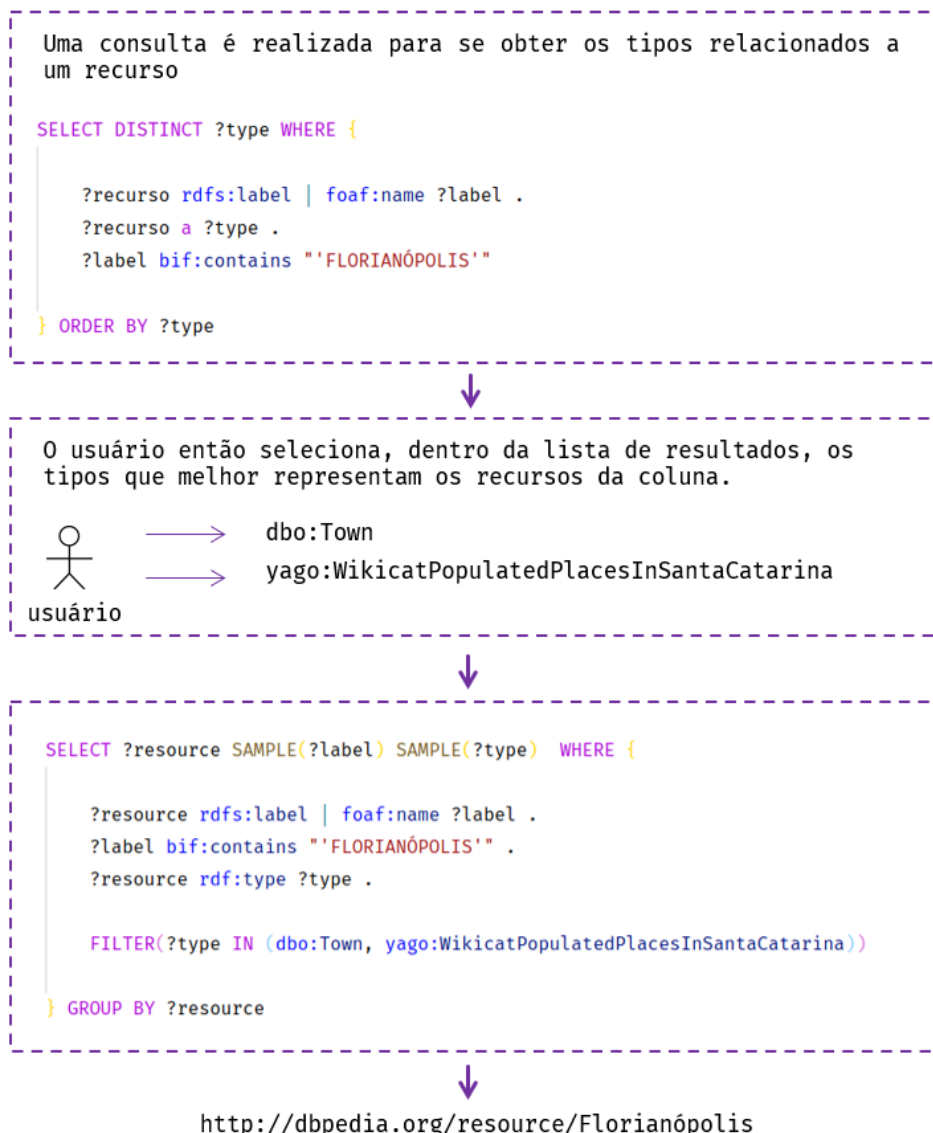
Durante a execução da atividade 1, o usuário deve apenas selecionar a coluna cujos dados deseja mapear, nesse caso a coluna com a lista de cidades. Em seguida, na atividade 2, deve selecionar uma ou mais cidades para servir de amostra para obtenção das classes, identificadas pela propriedade *<rdf:type>*, do recurso na DBpedia, relacionadas a essa cidade.

Selecionada a cidade de amostra, uma lista de tipos relacionados será exibida ao usuário. Ao selecionar como amostra a cidade de São Paulo, por exemplo, os tipos Cidade, Estado, Capital, Clube de Futebol e Eventos podem figurar entre os resultados possíveis. Para obtenção desses tipos, uma consulta utilizando a linguagem SPARQL¹⁰, é realizada no *endpoint* da DBpedia, utilizando o texto contendo o nome da cidade para compor a consulta. Essas operações são executadas pela atividade 3.

¹⁰ SPARQL *Protocol and RDF Query Language* (SPARQL), linguagem para consultas em grafos RDF.

Por fim, na atividade 4, o usuário seleciona um ou mais tipos resultantes dessa consulta, para servir de parâmetro para seleção dos recursos a serem interligados. Uma nova consulta SPARQL é realizada, agora incluindo em seus parâmetros além do nome da capital, os tipos que melhor representam os recursos dessa coluna. A Figura 23 ilustra o conjunto de consultas realizadas para obtenção de recursos relacionados a cidade de Florianópolis.

Figura 23 – Seleção dos tipos de um recurso na DBpedia



Fonte: Elaborado pelo autor.

A realização desses *Step* faz com que cada linha da coluna selecionada seja associada a um recurso presente na DBpedia, quando existir, desde que a informação presente nessa

linha contenha o mesmo nome ou identificador desse recurso, e que os tipos selecionados pelo usuário estejam presentes nesse mesmo recurso.

Os resultados encontrados irão compor as triplas do modelo RDF por meio do predicado `<owl:sameAs>`, que significa que dois URIs diferentes representam o mesmo recurso do mundo real, ligando os recursos do conjunto de dados com os da DBpedia.

Ao ser salvo, esse *Step* será agendado para ser executado no momento de exportação do conjunto de dados RDF.

4.4.4 Mapeamento Semântico

O objetivo deste *Step* é criar um modelo semântico para definição dos recursos existentes no domínio, por meio da atribuição de propriedades e relacionamentos que podem ser usados para descrevê-los, apoiando-se pelo uso de vocabulários e ontologias de referência. Esse modelo deve ser projetado pelo usuário, de modo a selecionar os dados que irão compor as triplas que farão parte do modelo RDF resultante.

A abordagem utilizada consistiu na criação de um modelo baseado em triplas RDF, para o mapeamento das informações presentes na representação tabular, de modo que as colunas, identificadas por seus cabeçalhos, fossem descritas como parte de uma tripla RDF, na forma `<sujeito><predicado><objeto>`.

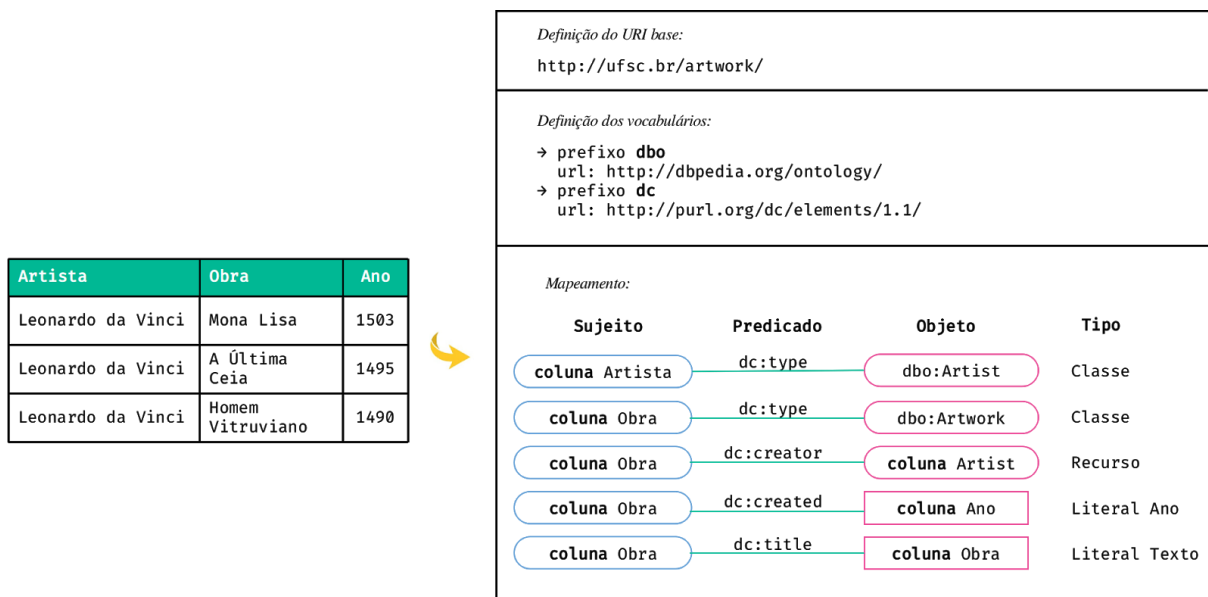
Os sujeitos são definidos por meio da seleção da coluna da representação tabular, ao qual se deseja atribuir predicados. Os sujeitos devem ser recursos identificados por URIs.

Os predicados especificam como os sujeitos e os objetos estão relacionados e devem ser definidos por meio do uso de vocabulários de referência que melhor identifiquem o conjunto de dados, aumentando a expressividade e reduzindo ambiguidades. Este *Step* oferece suporte a reutilização de ontologias e vocabulários existentes. Dessa forma, o vocabulário deve ser importado, informando-se um prefixo e o *link* para o vocabulário, ou realizando o *upload* do arquivo, nos formatos RDF e OWL, com o vocabulário. Ao se importar um vocabulário, são importadas suas propriedades (propriedades de objeto e de tipo de dados) e suas classes, que podem ser usadas para representar um objeto na tripla.

Por fim, os objetos podem ser representados por literais definidos pelo XML *Schema DataType*, por classes da ontologia importada ou por outros recursos, selecionando-se as colunas ao qual se deseja conectar.

Quando um recurso é mapeado, é necessário que esteja devidamente identificado por URIs únicos, uma vez que representa um recurso específico na *web* de dados. Nesse sentido, este *Step* gerencia a escrita automatizada de URIs de acordo com o modelo $\{workflow_uri\}/\{column\}/\{data\}$. Nesse modelo, $\{workflow_uri\}$ está relacionada com um URI válida, definida no momento de criação do *Workflow*. O $\{column\}$, por sua vez, refere-se à coluna em que o recurso está disposto e $\{data\}$ descreve à própria coisa a qual se refere o recurso. A Figura 24 ilustra o procedimento para realizar o mapeamento semântico.

Figura 24 – Mapeamento Semântico



Fonte: Elaborado pelo autor.

A Figura 24 apresenta a tabela de interesse, que contém os dados, sobre artistas e suas obras, a serem descritos conforme o modelo RDF. É necessário inicialmente a verificação do URI, definido no momento de criação do *Workflow*, que será usado como base para as demais URIs, bem como a importação de vocabulários para suporte à descrição dos recursos. No exemplo, são importados os vocabulários *DBpedia Ontology* para descrição das obras de arte, identificado pelo prefixo $\langle dbo \rangle$ e o vocabulário *Dublin Core*, identificado pelo prefixo $\langle dc \rangle$, para descrição das propriedades e classes de um recurso.

Inicialmente a coluna *Artista* da tabela, é definida como sendo do tipo $\langle dbo:Artist \rangle$ e a coluna *Obra* é definida como sendo do tipo $\langle dbo:Artwork \rangle$. Em seguida é criada uma relação entre a coluna *Obra* e a coluna *Artista*, por meio da propriedade $\langle dc:creator \rangle$. Por

fim, para a coluna Obra são definidas as propriedades <dc:created> e <dc:title>, relacionando-as com os literais presentes nas colunas Ano e Obra, respectivamente.

É importante ressaltar que neste *Step*, o modelo deve ser generalizado, ou seja, as colunas devem ser mapeadas de acordo com seus aspectos gerais, excluindo mapeamentos sobre indivíduos específicos do domínio.

Quando salvo este *Step*, o mapeamento realizado servirá como estrutura ontológica para criação da base de conhecimento, sendo então, agendado para executar no momento de criação do modelo RDF. Cabe ressaltar que, apenas os dados mapeados serão exportados para o modelo RDF.

4.4.5 Outros Steps

Além dos *Steps* de interligação com a DBpedia e de mapeamento semântico, outros *Steps* foram implementados, com o intuito de realizar transformações no conjunto de dados primários. Esses *Steps* são apresentados no Quadro 12:

Quadro 12 – Steps da aplicação

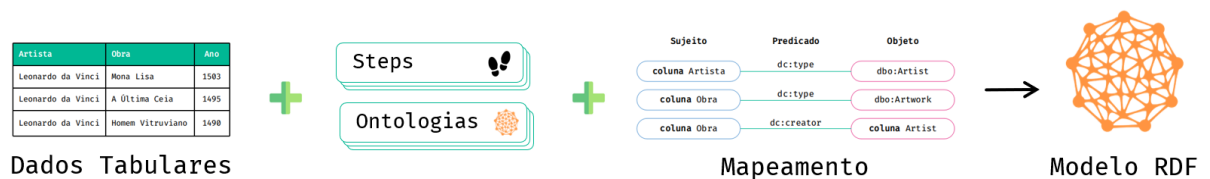
Escopo	Steps	Descrição
Conjunto de dados	Reimportar	Reimporta o conjunto de dados, com base no conjunto de dados original.
Conjunto de dados	Reordenar	Reordena o conjunto de dados, de acordo com uma coluna especificada.
Linha	Adicionar linha	Cria uma nova linha para o conjunto de dados.
Linha	Remover linha	Remove uma linha existente do conjunto de dados.
Colunas	Renomear colunas	Renomeia o cabeçalho das colunas.
Colunas	Mover colunas	Move as colunas do conjunto de dados.
Colunas	Desativar colunas	Ocultas as colunas do conjunto de dados.
Colunas	Maiúsculas	Converte o texto da coluna para letras maiúsculas.
Colunas	Minúsculas	Converte o texto da coluna para letras minúsculas.
Colunas	Substituir	Substitui a ocorrência de uma palavra por outra em uma determinada coluna.

Fonte: Elaborado pelo autor.

É importante salientar que, os *Steps* executados em um *Workflow*, podem ser reaproveitados em outros *Workflows*.

Para concretizar a criação da base de conhecimento ainda é necessária a execução de uma última etapa pelo usuário, chamada aqui de exportação. É nesta etapa que os dados dispostos no modelo tabular são traduzidos para o modelo RDF. A Figura 25 representa um esquema desta etapa.

Figura 25 – Visão geral do processo de geração do Modelo RDF



Fonte: Elaborado pelo autor.

A execução desta etapa é gerenciada por um *WorkflowExecution*. O *WorkflowExecution* é responsável pela criação da base de conhecimento, por meio da combinação do *Step* de Mapeamento Semântico e dos dados no modelo tabular. O *WorkflowExecution* ainda é responsável por verificar os *Steps* agendados para antes ou após a criação da base de conhecimento e executá-los.

O capítulo seguinte apresenta um caso de uso prático da aplicação, como meio de demonstrar seu funcionamento.

5 CASO DE USO DO EGCFLOW: ELEIÇÕES 2018

Considerando o exposto no capítulo anterior, este capítulo objetiva apresentar a aplicação EGCFlow, por meio de um caso de uso, com o intuito de demonstrar suas principais características, funcionalidades e os benefícios para a geração e manutenção de conjuntos de dados abertos conectados, conforme previsto na etapa de Demonstração da metodologia. A fim de demonstrar seu uso, optou-se pela utilização de dados primários do mundo real, referentes às eleições no Brasil.

Além disso, o capítulo apresenta os resultados da execução da aplicação, isto é, a etapa de Avaliação prevista pela metodologia adotada.

5.1 O SISTEMA ELEITORAL BRASILEIRO

Uma eleição é um processo de escolha coletiva em que são eleitos os representantes da população para os Poderes Executivo e Legislativo, de acordo com o sistema eleitoral. Este, por sua vez, consiste num conjunto “estruturado e funcional de regras, instrumentos e mecanismos para conferir, de forma legítima, mandato político aos representantes do povo (...)” (AZEVEDO, 2014, p. 187). A escolha dessa representação é definida pelo voto secreto e direto, no qual o cidadão manifesta sua escolha, de forma individual, no ato de votação.

No Brasil, as eleições são realizadas empregando-se dois sistemas eleitorais distintos: o majoritário, utilizado nas eleições para o Executivo e o Senado Federal, e o proporcional, utilizado para as eleições do Legislativo (TORRES, 2014). A estrutura eleitoral ainda é composta por uma série de componentes e variáveis, tais como a existência de partidos e coligações; a própria adoção de dois sistemas eleitorais, que por sua vez pode ser dividido em subsistemas; a presença do cálculo de quociente eleitoral para as eleições de Legislativo e as configurações de seções e zonas eleitorais.

A complexidade inerente ao processo eleitoral, associado ao desafio gerado pelo fato de que os dados relacionados às eleições, muitas vezes, não estão adequados ao processamento computacional, acaba por gerar inconvenientes em sua coleta, modelagem, padronização, consumo e combinação com outros conjuntos de dados, impactando no uso dos mesmos e, conseqüentemente, na produção de conhecimento acerca deles. De acordo com Isotani e Bittencourt (2015), isso demonstra a importância que, tanto a representação e a

estruturação, quanto à conexão entre conjuntos de dados distintos possuem no processo de recuperação e produção de novos conhecimentos. Assim, as próximas seções apresentarão o passo a passo adotado no uso da aplicação, para o enriquecimento semântico e a interligação com outras fontes de dados, para o conjunto de dados das eleições brasileiras.

5.2 ROTEIRO DE USO DA APLICAÇÃO

Conforme mencionado, para a demonstração de uso da aplicação optou-se pela utilização de um conjunto de dados referentes às eleições do Brasil. Para tal, o Tribunal Superior Eleitoral (TSE)¹¹ foi consultado, de forma a se obter esse conjunto de dados primários.

5.2.1 Seleção da fonte primária

A primeira atividade realizada, conforme mencionado, foi a seleção da fonte primária, que no caso em questão, foram os dados referentes às eleições de 2018, extraídos do portal do Tribunal Superior Eleitoral. Para tal, foi acessado o Repositório de Dados Eleitorais, do TSE. Em seguida, foi selecionada a opção “Resultados” do ano de 2018, seguida da seleção do *link* “Boletim de urna - Primeiro Turno”.

¹¹ Dados extraídos do Tribunal Superior Eleitoral. Disponível em: <<https://www.tse.jus.br/eleicoes/estatisticas/repositorio-de-dados-eleitorais-1/repositorio-de-dados-eleitorais>>.

Figura 26 – Portal de Dados Eleitorais do TSE

Fonte: Tribunal Superior Eleitoral, 2021.

O *link* em questão continha os dados referentes ao boletim de urna do primeiro turno das Eleições Federais e Estaduais em 2018, com arquivos individuais por estado, em formato CSV. Cada arquivo apresentava dados da votação por candidato, para os cargos de deputado estadual, deputado federal, senador, governador e presidente. As informações contidas nos conjuntos de dados apresentavam as variáveis listadas abaixo, conforme arquivo “leia-me”, disponibilizado pelo próprio TSE, que pode ser observado no Quadro 13:

Quadro 13 – Descrição dos dados

Variável	Descrição
DT_GERACAO	Data de geração do arquivo (data da extração dos dados)
HH_GERACAO	Hora de geração do arquivo (hora da extração dos dados) - Horário de Brasília
CD_PLEITO	Código do pleito da eleição
ANO_ELEICAO	Ano da eleição (referente ao ano eleitoral de pesquisa)
DT_PLEITO	Data do pleito da eleição.

NR_TURNO	Número do turno da eleição.
CD_ELEICAO	Código da eleição.
DS_ELEICAO	Descrição da eleição.
DT_ELEICAO	Data em que ocorreu a eleição.
SG_UF	Sigla da Unidade da Federação em que ocorreu a eleição.
CD_MUNICIPIO	Código TSE do município onde ocorreu a eleição.
NM_MUNICIPIO	Nome do município onde ocorreu a eleição.
NR_ZONA	Número da Zona Eleitoral em que ocorreu a eleição.
NR_SECAO	Número da Seção Eleitoral em que ocorreu a eleição.
NR_LOCAL_VOTACAO	Número do local de votação referente ao boletim de urna.
CD_CARGO_PERGUNTA	Código do cargo do candidato ou pergunta, no caso de plebiscito.
DS_CARGO_PERGUNTA	Descrição do cargo do candidato ou pergunta, no caso de plebiscito.
NR_PARTIDO	Número de identificação do partido.
SG_PARTIDO	Sigla do partido.
NM_PARTIDO	Nome do partido.
QT_APTOS	Quantidade de eleitores aptos a votar naquele município, zona e seção.
QT_COMPARECIMENTO	Quantidade de eleitores que compareceram às eleições naquele município, zona e seção.
QT_ABSTENCOES	Quantidade de eleitores que não compareceram às eleições naquele município, zona e seção
CD_TIPO_URNA	Código de identificação do tipo de urna.
DS_TIPO_URNA	Descrição de identificação do tipo de urna.
CD_TIPO_VOTAVEL	Código do Tipo de votável.
DS_TIPO_VOTAVEL	Descrição do tipo votável.
NR_VOTAVEL	Pode assumir os valores: Número do candidato (quando voto nominal); Número do partido (quando voto em legenda); Número 95 (quando voto em branco); Número 96 (quando voto nulo); Número 97 (quando voto anulado e apurado em separado) e Número 98 (quando voto anulado).

Fonte: Tribunal Superior Eleitoral, 2018.

Após a seleção dos dados, a próxima atividade consistiu na criação do projeto, objeto do próximo item.

5.2.2 Criação de Projeto

A partir da fonte primária de dados, o primeiro passo para utilização da aplicação foi a criação de um Projeto. Para isso, foi necessário cadastrar os seguintes dados de entrada: i) nome do projeto; ii) *homepage*; iii) autor; iv) palavras-chave; v) idioma; vi) propósito e, por fim, vii) descrição. Todos estes dados de entrada, podem ser observados na Figura 27, a seguir.

Figura 27 – Criação do Projeto

The screenshot shows the 'EGCFLOW' application interface. A modal window titled 'NOVO PROJETO' is open, displaying a form with the following fields and values:

- Projeto:** Eleições 2018
- Homepage:** <http://ufsc.edu.br/eleicoes>
- Autor:** Jefferson Chaves
- Palavras Chaves:** Eleições, Eleições Federais, Eleições 2018
- Idioma:** pt-br
- Propósito:** Transformar e conectar dados das eleições de 2018
- Descrição:** Número de votos por candidato das eleições de 2018

At the bottom of the modal, there are two buttons: 'CANCELAR' and 'SALVAR'. In the background, a table of projects is visible, and a '+ PROJETO' button is present in the top right corner.

Fonte: Elaborado pelo autor.

Nesse sentido, os dados registrados foram: i) nome do projeto: “Eleições 2018”; ii) *homepage*: <<http://ufsc.edu.br/eleicoes>>; iii) autor: Jefferson Chaves; iv) palavras-chave: Eleições, Eleições Federais, Eleições 2018; v) idioma: pt-br; vi) propósito: transformar e conectar dados das eleições de 2018; v) descrição: Número de votos por candidato das eleições de 2018.

É importante destacar que após a criação do projeto, automaticamente, foi criada uma pasta no computador do usuário, em que foram salvos os artefatos, separados por *Workflow*, referentes a este Projeto.

Concluída a criação do Projeto, a próxima atividade consistiu na criação de fluxos de execução, tema do próximo tópico.

5.2.3 Criação de *Workflow*: definição dos fluxos de execução

Um Projeto é o ponto de partida para criação de *Workflows*. Cada *Workflow* de um Projeto possui um conjunto de dados primários relacionados a um mesmo conteúdo ou tema, mas que apresenta distinções quanto a sua categoria, grupo, licença, cobertura geográfica ou tempo. Neste exemplo, os *Workflows* foram criados baseados na unidade federativa em que os dados foram produzidos, ou seja, por cobertura geográfica. A seleção dos estados ocorreu por ordem alfabética, de forma que o estado do Acre foi o primeiro selecionado.

Para a criação de um *Workflow*, é necessário informar: i) um nome; ii) um URI, que será utilizado como base para criação dos recursos; iii) fonte primária; iv) cobertura espacial; v) período temporal; vi) o padrão de codificação do documento; vii) o caractere separador, para o caso de arquivos CSV ou TSV; viii) uma licença; e por fim, ix) o próprio arquivo contendo os dados primários. A Figura 28 apresenta a tela de criação de um *Workflow*.

Figura 28 – Criação de *Workflow* - estado do Acre

EGCFLOW

projeto ≡ Eleições 2018

Homepage
http://ufsc.edu.br/eleicoes

Propósito
Transformar e conectar dados das eleições de 2018

Descrição
Número de votos por candidato das eleições de 2018

LDWorkflows

#id Workflow

Ações

por página 10

NOVO WORKFLOW

nome
Votação Acre

URI
http://ufsc.br/eleicoes/votacao_acre/2018/v1/

Fonte primária
https://cdn.tse.jus.br/estatistica/sead/eleicoes/eleicoes2018/buweb/BWEB_1t_1938.zip

Cobertura espacial
Estado do Acre

Período temporal
2018

codificação
ISO_8859_1 ;

licença
Creative Commons License - CC

bweb_1t_AC...181938.csv

CANCELAR SALVAR

Fonte: Elaborado pelo autor.

O *Workflow* em questão foi intitulado de “Votação Acre”. Seu URI foi definido como $\langle \text{http://ufsc.br/eleicoes/votacao_acre/2018/v1} \rangle$. Esse URI permite que seja identificada a versão desse conjunto de dados (v1), além de compor os URIs dos recursos contidos no conjunto de dados. A fonte primária foi definida como $\langle \text{https://cdn.tse.jus.br/estatistica/sead/eleicoes/eleicoes2018/buweb/BWEB_1t_AC_101020181938.zip} \rangle$. A cobertura espacial, por sua vez, foi definida como o estado do Acre. Já o período temporal foi definido como 2018. A licença foi selecionada, com base na Portaria nº 93, de 12 de fevereiro de 2021, do TSE¹², que define os dados em questão como dados abertos e, que são definidos como:

(...) dados acessíveis ao público, representados em meio digital, estruturados em formato aberto, não proprietário, processáveis por máquina, referenciados na internet e disponibilizados sob licença aberta que permita sua livre utilização, consumo ou cruzamento, limitando-se a creditar a autoria ou a fonte (TRIBUNAL SUPERIOR ELEITORAL, 2021, n.p.).

¹² Disponível em: $\langle \text{www.tse.jus.br/legislacao/compilada/prt/2021/portaria-no-93-de-12-de-fevereiro-de-2021} \rangle$. Acesso em: 18 de Jun. de 2018.

Dessa forma, a licença que melhor atendeu a essas características foi a *The Open Government License* (OGL), dentro de um rol de 10 (dez) licenças disponibilizadas pela aplicação. O padrão de codificação utilizado, segundo informado pelo TSE no arquivo “leia-me”, é o Latin 1, também conhecido como “ISO_8859_1”, sendo, portanto, este padrão o selecionado para o *Workflow*. O caractere separador “ponto e vírgula” foi selecionado para este conjunto de dados e, por fim, o arquivo contendo os dados do Acre, no formato CSV, foi selecionado. A Figura 29 apresenta uma visão geral do *Workflow* criado.

Figura 29 – Tabela de dados do *Workflow* criado

≡ Votação Acre

ORDENAR EXECUTAR WORKFLOW RESETAR SEMANTIFICAR EXPORTAR

linha	DT_GERACAO	HH_GERACAO	ANO_ELEICAO	CD_PLEITO	DT_PLEITO	NR_TURNO	CD_ELEICAO	DS_ELEICAO
2	10/10/2018	19:38:23	2018	228	07/10/2018	1	295	Eleição Geral F 2018
3	10/10/2018	19:38:23	2018	228	07/10/2018	1	295	Eleição Geral F 2018
4	10/10/2018	19:38:23	2018	228	07/10/2018	1	295	Eleição Geral F 2018
5	10/10/2018	19:38:23	2018	228	07/10/2018	1	295	Eleição Geral F 2018
6	10/10/2018	19:38:23	2018	228	07/10/2018	1	295	Eleição Geral F 2018
7	10/10/2018	19:38:23	2018	228	07/10/2018	1	295	Eleição Geral F 2018
8	10/10/2018	19:38:23	2018	228	07/10/2018	1	295	Eleição Geral F 2018
9	10/10/2018	19:38:23	2018	228	07/10/2018	1	295	Eleição Geral F 2018
10	10/10/2018	19:38:23	2018	228	07/10/2018	1	295	Eleição Geral F 2018
11	10/10/2018	19:38:23	2018	228	07/10/2018	1	295	Eleição Geral F 2018

itens por página 10 1-10 de 279230

Fonte: Elaborado pelo autor.

Ao ser criado um *Workflow*, uma pasta com seu nome é criada dentro da pasta do Projeto e o conjunto de dados primários é armazenado nesta pasta.

Concluída a criação do *Workflow*, é possível a execução dos *Steps* para transformar, semantificar e gerar os dados no modelo RDF, etapa descrita no próximo item.

5.2.4 Execução dos *Steps*

A aplicação desenvolvida possibilitou que os *Steps*, identificados por um ícone de pegadas, fossem executados zero ou muitas vezes, independente de ordem. A Figura 30 exibe a disposição de alguns *Steps* na aplicação.

Figura 30 – Visualização de alguns *Steps* disponíveis na aplicação

☰ Votação Acre

ORDENAR
EXECUTAR WORKFLOW
RESETAR
SEMANTIFICAR
EXPORTAR

linha	DT_GERACAO	HH_GERACAO	ANO_ELEICAO	CD_PLEITO	DT_PLEITO	NR_TURNO	CD_ELEICAO	DS_ELEICAO
2	10/10/2018	19:38:23	2018	228		1	295	Eleição Geral F 2018
3	10/10/2018	19:38:23	2018	228		1	295	Eleição Geral F 2018
4	10/10/2018	19:38:23	2018	228		1	295	Eleição Geral F 2018
5	10/10/2018	19:38:23	2018	228		1	295	Eleição Geral F 2018
6	10/10/2018	19:38:23	2018	228		1	295	Eleição Geral F 2018
7	10/10/2018	19:38:23	2018	228		1	295	Eleição Geral F 2018
8	10/10/2018	19:38:23	2018	228	07/10/2018	1	295	Eleição Geral F 2018
9	10/10/2018	19:38:23	2018	228	07/10/2018	1	295	Eleição Geral F 2018
10	10/10/2018	19:38:23	2018	228	07/10/2018	1	295	Eleição Geral F 2018
11	10/10/2018	19:38:23	2018	228	07/10/2018	1	295	Eleição Geral F 2018

ordenar A-Z
ordenar Z-A
Substituir
Filtrar
Maiúsculas
Minúsculas
interligar DBPedia

itens por página 10 1-10 de 279230

Fonte: Elaborado pelo autor.

O roteiro estabelecido para a execução dos *Steps*, bem como o resultado em razão de sua execução, será descrito nos tópicos seguintes.

5.2.4.1 Renomear cabeçalhos

Neste *Step*, os cabeçalhos das colunas da tabela de dados foram renomeados. Colunas com abreviações foram renomeadas com seus nomes por extenso, a fim de melhor descrever os recursos associados. Assim, as abreviações foram substituídas da seguinte forma: “DT” por “DATA”; “HH” por “HORA”; “CD” por “CÓDIGO”; “NR” por “NÚMERO”;

“DS” por “DESCRIÇÃO”; “NM” por “NOME”, “SG” por “SIGLA”; “QT” por “QUANTIDADE” e “NH” por “NÃO HABILITADOS”.

Também foram alterados os títulos “CD_CARGO_PERGUNTA” e “DS_CARGO_PERGUNTA” para “CÓDIGO CARGO” e “DESCRIÇÃO CARGO”, respectivamente, uma vez que o termo “PERGUNTA” só é aplicado para plebiscitos, o que não é o caso. O título “CD_FLASCARD_URNA_EFETIVADA” foi alterado para “CÓDIGO FLASHCARD URNA EFETIVADA”, para correção da palavra em inglês “FLASHCARD”. As alterações mencionadas podem ser visualizadas de forma geral, observando-se os cabeçalhos da tabela de dados na Figura 31.

Figura 31 – Cabeçalhos renomeados após a execução do Step

≡ Votação Acre

ORDENAR EXECUTAR WORKFLOW RESETAR SEMANTIFICAR EXPORTAR

LINHA	DATA GERAÇÃO	HORA GERAÇÃO	ANO ELEIÇÃO	CÓDIGO PLEITO	DATA PLEITO	NÚMERO TURNO	CÓDIGO ELEIÇÃO	DESCRIÇÃO ELI
2	10/10/2018	19:38:23	2018	228	07/10/2018	1	295	Eleição Geral 2018
3	10/10/2018	19:38:23	2018	228	07/10/2018	1	295	Eleição Geral 2018
4	10/10/2018	19:38:23	2018	228	07/10/2018	1	295	Eleição Geral 2018
5	10/10/2018	19:38:23	2018	228	07/10/2018	1	295	Eleição Geral 2018
6	10/10/2018	19:38:23	2018	228	07/10/2018	1	295	Eleição Geral 2018
7	10/10/2018	19:38:23	2018	228	07/10/2018	1	295	Eleição Geral 2018
8	10/10/2018	19:38:23	2018	228	07/10/2018	1	295	Eleição Geral 2018
9	10/10/2018	19:38:23	2018	228	07/10/2018	1	295	Eleição Geral 2018
10	10/10/2018	19:38:23	2018	228	07/10/2018	1	295	Eleição Geral 2018
11	10/10/2018	19:38:23	2018	228	07/10/2018	1	295	Eleição Geral

Itens por página 10 1-10 de 279230

Fonte: Elaborado pelo autor.

A seguir será abordado o *Step* de reordenamento de conjunto de dados.

5.2.4.2 Reordenar Conjunto de Dados

Neste *Step*, o conjunto de dados foi reordenado, a partir da execução do *Step* “Reordenar Conjunto de Dados”, aplicado à coluna “MUNICÍPIO”. Cabe destacar que, a

reordenação, ainda que tenha sido aplicada à coluna supracitada, se estende às demais colunas do conjunto de dados. Dessa forma, a tabela de dados foi disposta de maneira organizada, em ordem alfabética crescente.

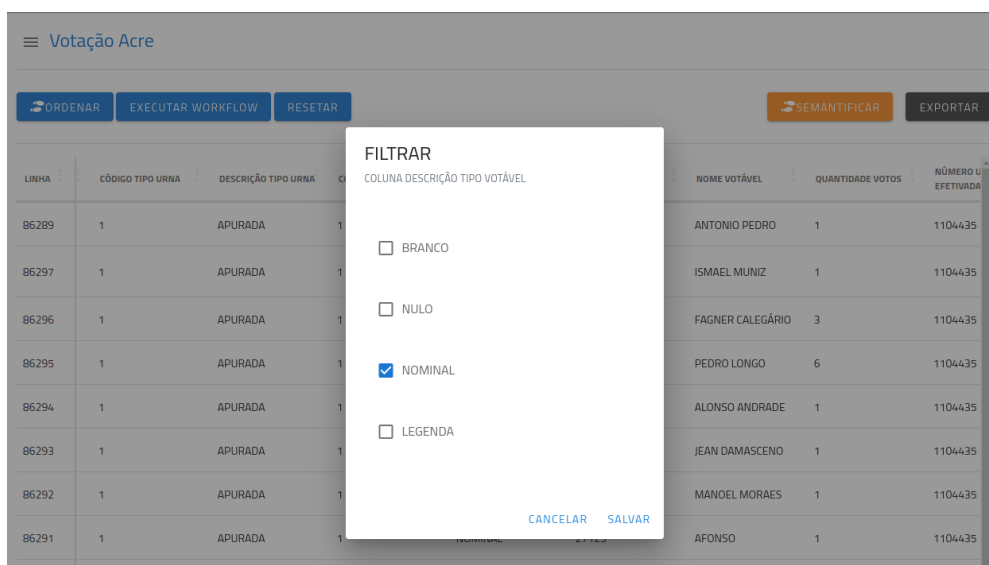
5.2.4.3 Normalizar caixa alta/caixa baixa

Os dados primários apresentavam, ora dados em caixa baixa, ora dados em caixa alta. Assim optou-se, para fins de padronização, a transformação das colunas com textos em caixa baixa para caixa alta. Para tanto, o *Step* intitulado “Maiúsculas” foi aplicado para as colunas “DESCRIÇÃO ELEIÇÃO”, “DESCRIÇÃO CARGO”, “NOME PARTIDO”, “DESCRIÇÃO TIPO URNA” e “DESCRIÇÃO TIPO VOTÁVEL”.

5.2.4.4 Filtrar

O *Step* de Filtro, permite que linhas sejam filtradas de acordo com o valor de seus campos. Neste caso, foi aplicado o *Step* de Filtro à coluna “DESCRIÇÃO TIPO VOTÁVEL”, com o objetivo de manter no conjunto de dados, apenas os dados referentes a votos nominais, excluindo linhas cujos valores dessa coluna fossem iguais a “LEGENDA”, “BRANCO” ou “NULO”, conforme pode ser observado na Figura 32.

Figura 32 – *Step* para filtro dos dados da tabela



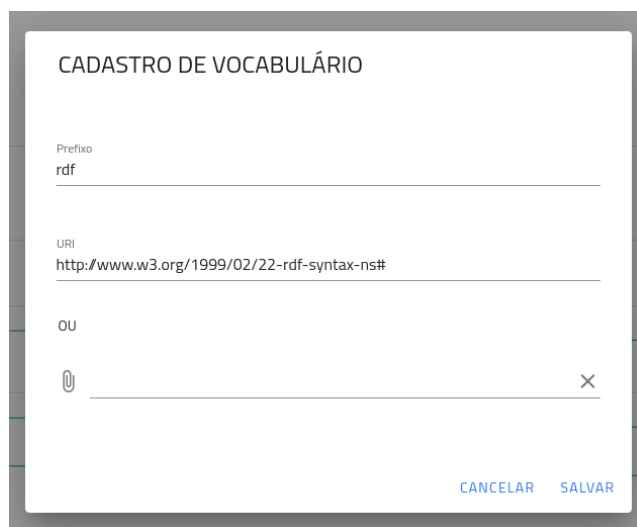
Fonte: Elaborado pelo autor.

Como parte dessa transformação, o título do cabeçalho “NOME VOTÁVEL” foi alterado para CANDIDATO, já que valores que descrevem partidos, ou votos em branco foram filtrados na etapa descrita acima.

5.2.4.5 Execução do Mapeamento Semântico

Neste *Step*, detalhado no item 4.4.4 deste trabalho, foi necessário realizar manualmente um conjunto de tarefas para o mapeamento. A primeira delas consistiu na importação das ontologias e vocabulários para definição das classes e propriedades, que iriam compor o mapeamento. Deste modo, foram importadas a ontologia de “Eleições”¹³, também disponível no Apêndice A, identificada pelo prefixo <ele>, o vocabulário RDF, identificado pelo prefixo <rdf> e o vocabulário *Dublin Core*, identificado pelo prefixo <dc>. Essa importação pode ser realizada por meio do *link* do vocabulário ou ontologia na *web*, ou ainda, por meio do *upload* de um arquivo, conforme pode ser observado na Figura 33.

Figura 33 – Importação de ontologia ou vocabulário





CADASTRO DE VOCABULÁRIO

Prefixo
rdf

URI
http://www.w3.org/1999/02/22-rdf-syntax-ns#

OU

CANCELAR SALVAR

Fonte: Elaborado pelo autor.

Cabe ressaltar que, após a importação dos vocabulários, estes são adicionados à pasta contendo os artefatos gerados pelo *Workflow*. Além disso, uma vez importados vocabulários e ontologias, propriedades de dados e objetos, bem como classes dessas ontologias, poderão ser utilizadas para compor o mapeamento.

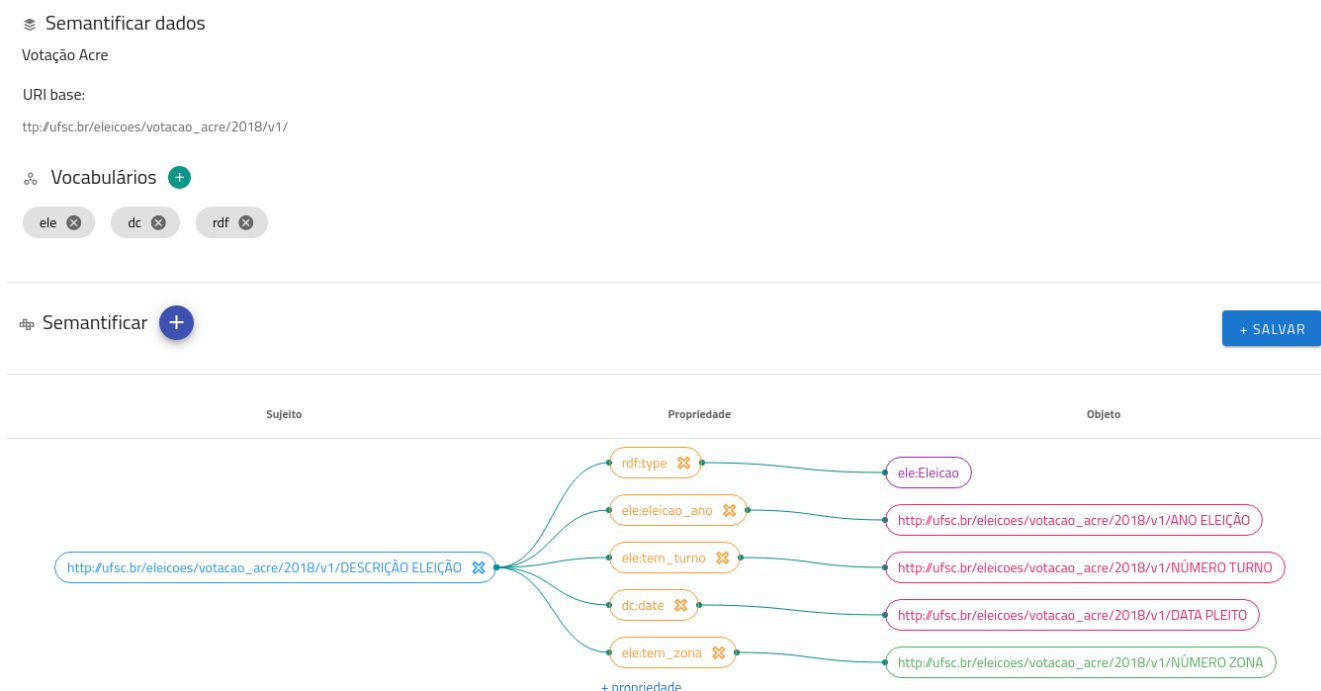
¹³ Ontologia elaborada pelo autor. Disponível em: <<http://ceur-ws.org/Vol-2519/short4.pdf>>.

Após a configuração das ontologias e vocabulários, é necessário realizar o mapeamento semântico. O mapeamento segue o modelo proposto pelas triplas RDF, estabelecendo uma relação binária entre dois recursos, ou um recurso e um literal, conectados por meio de um predicado.

Inicialmente, optou-se pelo mapeamento do conceito de “Eleição”. Foi identificada a coluna “Descrição Eleição”, como sendo a que melhor representa os recursos de uma eleição. A esse recurso foram atribuídos: i) seu tipo, por meio da propriedade *<rdf:type>*, do vocabulário RDF; ii) data do pleito, por meio da propriedade *<dc:date>*, presente no vocabulário Dublin Core, além das iii) propriedades ano da eleição, turno da eleição e zona eleitoral, por meio das propriedades *<ele:eleicao_ano>*, *<ele:tem_turno>* e *<ele:tem_zona>*, respectivamente, presentes na ontologia “Eleição”.

Isto posto, o recurso foi descrito por meio das seguintes relações: a propriedade *<rdf:type>* relacionou o recurso ao tipo *<ele:Eleicao>*; a propriedade *<ele:eleicao_ano>* relacionou o recurso a um valor literal que representa o ano, presente na coluna “ANO ELEIÇÃO”; a propriedade *<ele:tem_turno>* ao literal que representa o turno presente na coluna “NÚMERO TURNO”; a propriedade *<dc:date>* ao literal que representa a data do pleito definido na coluna “ANO ELEIÇÃO” e, por fim, a propriedade *<ele:tem_zona>* foi relacionada à coluna “NÚMERO ZONA”, definida aqui como um recurso. A Figura 34 exibe o resultado do mapeamento realizado para esses recursos.

Figura 34 – Mapeamento do conceito de Eleição



Fonte: Elaborado pelo autor.

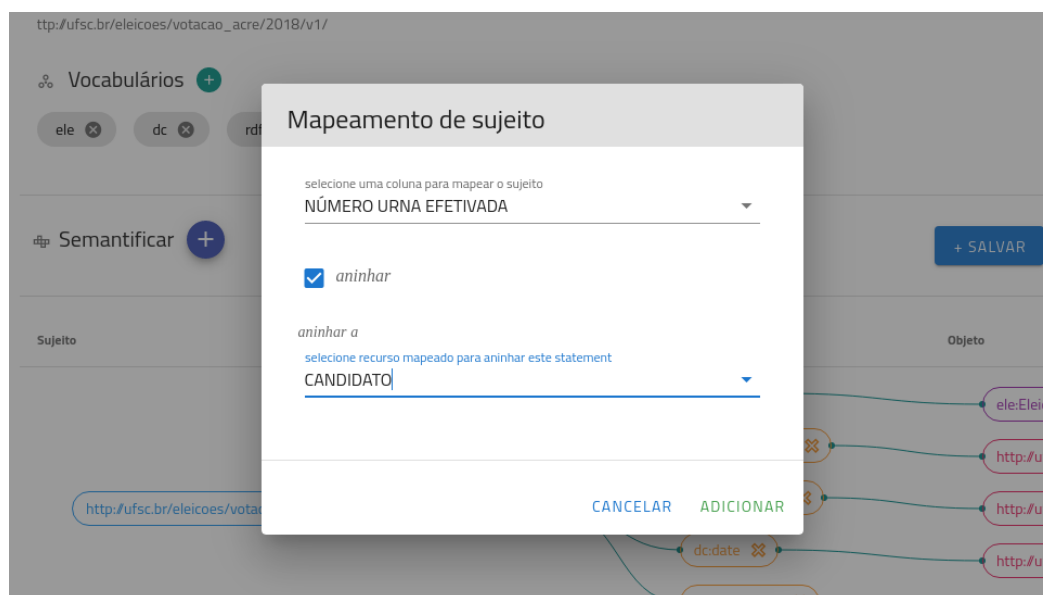
Num segundo momento, foi realizado o mapeamento semântico para o conceito de “Candidato”, análogo ao realizado para o conceito de “Eleição”. Assim, para os recursos que representam candidatos, foram atribuídos: i) o tipo, por meio da propriedade *<rdf:type>*, do vocabulário RDF; ii) o nome do candidato, o cargo ao qual concorre, seu número de candidato, o partido ao qual pertence, e o número de votos, por meio das propriedades *<ele:candidato_nome>*, *<ele:concorre_ao_cargo_de>*, *<ele:candidato_numero>*, *<ele:faz_parte_de>* e *<ele:tem_voto>*, respectivamente, da ontologia “Eleição”.

Dessa forma, o recurso foi descrito por meio das seguintes relações: a propriedade *<rdf:type>* relacionou o recurso ao tipo *<ele:Candidato>*; a propriedade *<ele:candidato_nome>* relacionou o recurso ao valor literal que representa o apelido do candidato na urna, presente na coluna “CANDIDATO”; a propriedade *<ele:concorre_ao_cargo_de>* relacionou o recurso ao valor literal que representa o nome do cargo pretendido pelo candidato, presente na coluna “DESCRIÇÃO CARGO”; a propriedade *<ele:candidato_numero>* relacionou o recurso ao valor literal que representa o número votável do candidato; a propriedade *<ele:faz_parte_de>* foi relacionada à coluna “NOME

PARTIDO”, definida aqui como um recurso e, por fim, a propriedade `<ele:tem_voto>` associada à coluna “NÚMERO URNA EFETIVADA”.

Essa última relação apresenta uma característica importante, uma vez que a quantidade de votos de um candidato está associada à urna em que ocorreu essa votação e não propriamente ao candidato. Para atender a este caso, criou-se também um mapeamento para a coluna “NÚMERO URNA EFETIVADA”, mas definindo-a como um mapeamento aninhado ao recurso que representa o candidato. Dessa forma, a propriedade `<ele:tem_voto>` do candidato foi associada a um *blank node*, descrito por meio das propriedades `<ele:urna_numero>` e `<ele:numero_de_votos>`. A Figura 35 apresenta o formulário de criação do mapeamento da quantidade de votos em uma urna para o candidato.

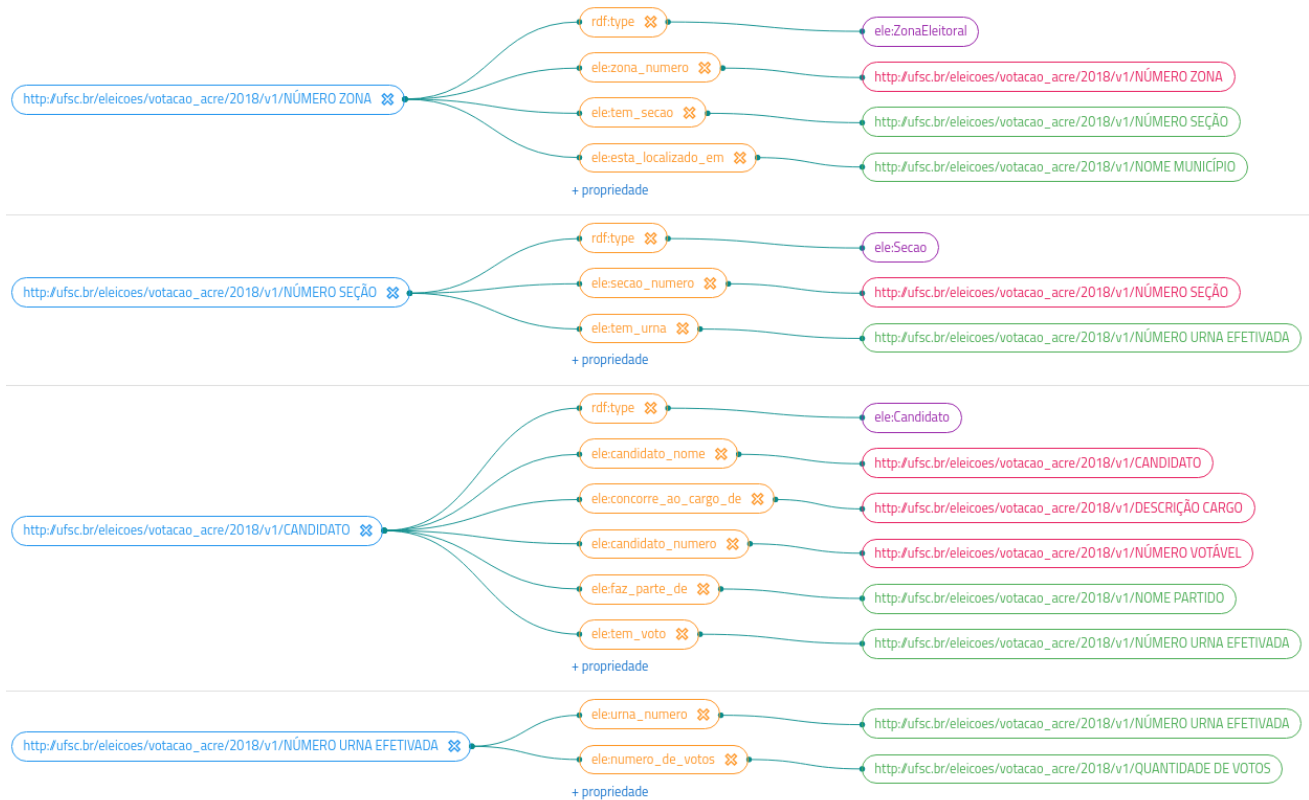
Figura 35 – Formulário de criação de recurso aninhado ao candidato



Fonte: Elaborado pelo autor.

De forma semelhante foi realizado o mapeamento para recursos que representam os conceitos de Zona, Seção Eleitoral e Partido, conforme pode ser observado na Figura 36.

Figura 36 – Visão geral do mapeamento semântico



Fonte: Elaborado pelo autor.

Na figura acima, os recursos são representados por elipses na cor azul. As elipses na cor laranja, por sua vez, representam os predicados. Já os objetos, quando representados pelas elipses com bordas lilás, correspondem a classes de ontologias; com bordas em rosa, correspondem a valores literais, e com bordas na cor verde correspondem a outros recursos.

Realizadas o mapeamento, o *Step* deve ser salvo. Neste momento, sua execução será agendada para ser realizada na etapa de exportação dos dados para o modelo RDF.

5.2.4.6 Interligar à DBpedia

Neste *Step*, cujo método foi detalhado no tópico 4.4.3, os recursos da coluna “MUNICÍPIO” foram associados a recursos presentes na DBpedia, quando existentes. Para execução deste *Step* foi necessário selecionar as classes que melhor descreviam os recursos dessa coluna.

As classes listadas foram obtidas a partir de uma amostra, neste caso, uma cidade presente na coluna. A cidade selecionada foi Rio Branco, como apresentado na Figura 37.

Figura 37 – Seleção da amostra

Interligar os recursos da coluna NOME MUNICÍPIO

The screenshot shows a web interface with a dropdown menu at the top left containing the text "a amostra escolhida é RIO BRANCO". Below this, there is a section titled "Marque as classes que melhor definem estes recursos" with a list of checkboxes and class URIs. To the right, there is a list of resources under the heading "resource".

Class	Resource
<input type="checkbox"/> http://dbpedia.org/class/yago/Abolitionist	http://dbpedia.org/resource/Rio_Branco
<input type="checkbox"/> http://dbpedia.org/class/yago/Abstraction	http://dbpedia.org/resource/Rio_Branco,_Rio_Grande_do_Sul
<input type="checkbox"/> http://dbpedia.org/class/yago/Act	http://dbpedia.org/resource/Rio_Branco_antbird
<input type="checkbox"/> http://dbpedia.org/class/yago/Activity	http://dbpedia.org/resource/Order_of_Rio_Branco
<input type="checkbox"/> http://dbpedia.org/class/yago/AdministrativeDistrict	http://dbpedia.org/resource/Branco_River_(Jamari_River)
<input type="checkbox"/> http://dbpedia.org/class/yago/Administrator	http://dbpedia.org/resource/Branco_River_(Roosevelt_River)
<input type="checkbox"/> http://dbpedia.org/class/yago/Adult	http://dbpedia.org/resource/Branco_River_(Iaciparaná_River)
<input type="checkbox"/> http://dbpedia.org/class/yago/Airfield	http://dbpedia.org/resource/Branco_River_(Guaporé_River)
	http://dbpedia.org/resource/Rio_Branco_EC

Fonte: Elaborado pelo autor.

Como resultado, foram retornados 65 (sessenta e cinco) recursos, cujo rótulo ou nome continham o texto “Rio Branco”. Entre esses recursos, figuravam recursos relacionados a rios, times de futebol, lugares, pessoas e cidades.

Assim, com o objetivo de filtrar tais resultados foi selecionada, dentre as classes disponibilizadas, a classe $\langle \text{http://dbpedia.org/ontology/Town} \rangle$, que representa o conceito de cidade. Como resultado foram retornados quatro recursos, quais sejam: Rio Branco do Ivaí; Rio Branco, Mato Grosso; Rio Branco, Acre; e Rio Branco do Sul, conforme pode ser observado na Figura 38.

Figura 38 – Recursos selecionados pela classe “Town”

a amostra escolhida é
RIO BRANCO

Town

☰ Marque as classes que melhor definem estes recursos

<http://dbpedia.org/ontology/Stream>

<http://dbpedia.org/ontology/Town>

<http://dbpedia.org/ontology/University>

<http://schema.org/AdministrativeArea>

<http://schema.org/Airport>

resource

http://dbpedia.org/resource/Rio_Branco_do_Ivaí

http://dbpedia.org/resource/Rio_Branco_do_Sul

http://dbpedia.org/resource/Rio_Branco_Acre

http://dbpedia.org/resource/Rio_Branco_Mato_Grosso

Fonte: Elaborado pelo autor.

Como a classe selecionada se aplicava a outros recursos, que não somente a cidade Rio Branco para o Estado do Acre, optou-se pela seleção de uma classe mais específica, a classe *<http://dbpedia.org/class/yago/WikicatPopulatedPlacesInAcre(state)>*, que retornou de forma correta os recursos relacionados à cidade do estado do Acre, conforme a Figura 39.

Figura 39 – Seleção de classe específica

a amostra escolhida é
RIO BRANCO

WikicatPopulatedPlacesInAcre(state)

☰ Marque as classes que melhor definem estes recursos

<http://dbpedia.org/class/yago/WikicatPeopleFromSalvador,Bahia>

<http://dbpedia.org/class/yago/WikicatPopulatedPlacesEstablishedIn>

[http://dbpedia.org/class/yago/WikicatPopulatedPlacesInAcre\(state\)](http://dbpedia.org/class/yago/WikicatPopulatedPlacesInAcre(state))

<http://dbpedia.org/class/yago/WikicatPopulatedPlacesInMatoGrosso>

<http://dbpedia.org/class/yago/WikicatPopulatedPlacesInMinasGerais>

[http://dbpedia.org/class/yago/WikicatPopulatedPlacesInParaná\(state\)](http://dbpedia.org/class/yago/WikicatPopulatedPlacesInParaná(state))

<http://dbpedia.org/class/yago/WikicatPopulatedPlacesInTheCerroLargoDep>

resource

http://dbpedia.org/resource/Rio_Branco

Fonte: Elaborado pelo autor.

Foi realizada uma verificação manual de cada uma das 22 (vinte e duas) cidades do estado, a fim de averiguar a consistência do *Step*. Realizada a seleção da classe para interligação com a DBpedia, foi necessário salvar o *Step*.

A Figura 40, apresenta alguns dos os *Steps* executados para o *Workflow*, com os dados do estado do Acre.

Figura 40 – Steps executados para o *Workflow* do Acre

The screenshot shows the LDWSteps interface. On the left, a sidebar lists several steps, each with a red 'X' icon indicating an error or failure:

- GERENCIAR COLUNAS: Gerenciamento de colunas realizado.
- ORDER BY STEP: Tabela ordenada por NOME MUNICIPIO.
- MAIÚSCULAS STEP: Texto do campo DESCRIÇÃO ELEIÇÃO alterado para maiúsculas.
- MAIÚSCULAS STEP: Texto do campo DESCRIÇÃO CARGO alterado para maiúsculas.
- MAIÚSCULAS STEP: Texto do campo NOME PARTIDO alterado para maiúsculas.
- MAIÚSCULAS STEP: Texto do campo DESCRIÇÃO TIPO URNA alterado para maiúsculas.
- MAIÚSCULAS STEP: Texto do campo DESCRIÇÃO TIPO VOTÁVEL alterado para maiúsculas.
- FILTRAR STEP: Tabela filtrada pela coluna DESCRIÇÃO.

The main area displays the 'Votação Acre' workflow with buttons for 'ORDENAR', 'EXECUTAR WORKFLOW', and 'RESETAR'. Below these buttons is a table with the following columns: LINHA, DATA GERAÇÃO, HORA GERAÇÃO, ANO ELEIÇÃO, CÓDIGO PLEITO, DATA PLEITO, NÚMERO TURNO, and CÓDIGO ELEIÇÃO. The table contains 10 rows of data, all generated on 10/10/2018 at 19:38:23 for the 2018 election year, with a pleito code of 228, a pleito date of 07/10/2018, and a turn number of 1. The election codes range from 297 to 303.

LINHA	DATA GERAÇÃO	HORA GERAÇÃO	ANO ELEIÇÃO	CÓDIGO PLEITO	DATA PLEITO	NÚMERO TURNO	CÓDIGO ELEIÇÃO
86289	10/10/2018	19:38:23	2018	228	07/10/2018	1	297
86297	10/10/2018	19:38:23	2018	228	07/10/2018	1	297
86296	10/10/2018	19:38:23	2018	228	07/10/2018	1	297
86295	10/10/2018	19:38:23	2018	228	07/10/2018	1	297
86294	10/10/2018	19:38:23	2018	228	07/10/2018	1	297
86293	10/10/2018	19:38:23	2018	228	07/10/2018	1	297
86292	10/10/2018	19:38:23	2018	228	07/10/2018	1	297
86291	10/10/2018	19:38:23	2018	228	07/10/2018	1	297
86290	10/10/2018	19:38:23	2018	228	07/10/2018	1	297
86287	10/10/2018	19:38:23	2018	228	07/10/2018	1	303

Fonte: Elaborado pelo autor.

Após a execução dos *Steps*, a próxima atividade consistiu na exportação dos dados, objeto do próximo item.

5.2.5 Exportação de dados

Realizados os *Steps* desejados, foi necessário realizar a exportação dos dados do *Workflow*. Suportado pelo mapeamento semântico, a exportação dos dados criou os modelos RDF serializados nos formatos Turtle, JSON-LD e RDF/XML, de acordo com a Figura 41.

Figura 41 – Artefatos gerados após a exportação do *Workflow*

Fonte: Elaborado pelo autor.

A fim de fornecer informações e proveniência acerca do conjunto de dados gerado, nesta etapa foi criado um relatório em formato RDF. Neste relatório foram adicionados metadados descritivos e estruturais, com o objetivo de enriquecer o entendimento, tanto do conjunto de dados, quanto do seu processo de criação. O relatório em questão pode ser visualizado no Apêndice B.

Além disso, foram adicionadas informações acerca da proveniência da fonte primária de dados. Adicionar tais informações é essencial. Os metadados fornecem um meio para que humanos e agentes de *software* possam encontrar, compreender e utilizar o conjunto de dados gerado de maneira adequada (RAUTENBERG *et al.*, 2018). Já as informações sobre proveniência apresentam elementos sobre a origem dos dados primários, de forma a garantir sua integridade e credibilidade, para os consumidores de dados (RAUTENBERG *et al.*, 2018).

Nesse sentido, o relatório criado apresenta as informações de título, autor, origem, palavras-chave, data de criação, cobertura espacial, período temporal, licença, objetivo, descrição, bem como a lista de *Steps* executados para transformar, interligar e gerar o conjunto de dados em RDF. A Figura 42 apresenta uma visão geral do relatório criado.

Figura 42 – Visão geral do relatório criado

```

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/terms/"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:ldwpo="http://ldwpo.aksw.org/terms/1.0/"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#">

  <rdf:Property rdf:about="http://ldwpo.aksw.org/terms/1.0/creator" />
  <rdf:Property rdf:about="http://ldwpo.aksw.org/terms/1.0/goal" />
  <rdf:Property rdf:about="http://ldwpo.aksw.org/terms/1.0/license" />
  <rdf:Property rdf:about="http://ldwpo.aksw.org/terms/1.0/ldwStep" />
  <rdf:Description rdf:about="http://ufsc.br/eleicoes/votacao_acre/2018/v1/">

    <dc:title>WORKFLOW - Votação Acre</dc:title>
    <ldwpo:creator>Jefferson Chaves</ldwpo:creator>
    <dc:source>https://cdn.tse.jus.br/estatistica/sead/eleicoes/eleicoes2018/buweb/BWEB_1t_AC_101020181938.zip</dc:source>
    <dc:subject>Eleições, Eleições Federais, Eleições 2018</dc:subject>
    <dc:created>15/08/2021</dc:created>
    <dc:language>pt-br</dc:language>
    <dc:temporal>2018</dc:temporal>
    <dc:spatial>Estado do Acre</dc:spatial>
    <ldwpo:license>Creative Commons License - CC</ldwpo:license>
    <ldwpo:goal>Transformar e conectar dados das eleições de 2018</ldwpo:goal>
    <dc:description>Número de votos por candidato das eleições de 2018</dc:description>
    <ldwpo:ldwStep>
      <ldwpo:Step>
        <rdfs:comment>Texto do campo DS_TIPO_VOTAVEL alterado para maiúsculas.</rdfs:comment>
        <dc:title>toUpperCaseStep</dc:title>
      </ldwpo:Step>
    </ldwpo:ldwStep>
    <ldwpo:ldwStep>
      <ldwpo:Step>
        <rdfs:comment>A relação semântica entre os dados e vocabulários foram estabelecidas.</rdfs:comment>

```

Fonte: Elaborado pelo autor.

Uma vez criado um *Workflow*, o conjunto de *Steps* executados pode ser reutilizado em outros *Workflows*. Para tanto, basta importar os *Steps* do *Workflow* desejado e realizar sua execução. Nesta demonstração, foi criado um novo *Workflow* para transformar os dados do Estado de Alagoas. Após esta atividade, foram importados os *Steps* executados no *Workflow* para o Estado do Acre. Para concluir esta atividade, foi necessário refazer o *Step* de interligação com a DBpedia. Isto ocorreu considerando que a classe escolhida para representar as cidades do Estado do Acre, não representava as cidades de Alagoas. Assim, o *Step* em questão foi excluído e um novo *Step* foi criado.

A execução do processo mencionado pode ser observada nas figuras a seguir. Na Figura 43 é observada a criação de um novo *Workflow* para o estado de Alagoas.

Figura 43 – Criação de novo *Workflow* para Alagoas

EGCFLOW

projeto **Eleições 2018**

Homepage
http://ufsc.edu.br/eleicoes

Propósito
Transformar e conectar dados das eleições de 20

Descrição
Número de votos por candidato das eleições de 2

LDWorkflows

#id	Workflow
1	Votação Acre

NOVO WORKFLOW

nome
Votação Alagoas

URI
http://ufsc.br/eleicoes/votacao_alagoas/2018/v1/

Fonte primária
https://cdn.tse.jus.br/estatistica/sead/eleicoes/eleicoes2018/buweb/BWEB_1t_

Cobertura espacial
Estado de Alagoas

Período temporal
2018

codificação
ISO_8859_1 ;

licença
Creative Commons License - CC

bweb_1t_AL...181938.csv

CANCELAR SALVAR

Fonte: Elaborado pelo autor.

Concluída a criação do *Workflow*, o passo seguinte foi realizar a importação dos *Steps* executados para o *Workflow* de votação do estado do Acre, que pode ser visualizado na Figura 44.

Figura 44 – Importação de *Steps*

The screenshot displays a web application interface for workflow management. On the left, a dropdown menu is open, showing 'selecione um workflow co...' and 'Votação Acre'. Below it is the 'LDWSteps' logo. The main area is titled 'Votação Alagoas' and contains several buttons: 'ORDENAR', 'EXECUTAR WORKFLOW', 'RESETAR', 'SEMANTIFICAR', and 'EXPORTAR'. A table below these buttons lists execution steps with the following columns: 'linha', 'DT_GERACAO', 'HH_GERACAO', 'ANO_ELEICAO', 'CD_PLEITO', 'DT_PLEITO', and 'NR_1'. The table contains 15 rows of data, all with the same values: DT_GERACAO: 10/10/2018, HH_GERACAO: 19:38:54, ANO_ELEICAO: 2018, CD_PLEITO: 228, DT_PLEITO: 07/10/2018, and NR_1: 1. At the bottom right, there is a pagination control showing 'itens por página 10' and '1-10 de 757957'.

linha	DT_GERACAO	HH_GERACAO	ANO_ELEICAO	CD_PLEITO	DT_PLEITO	NR_1
4	10/10/2018	19:38:54	2018	228	07/10/2018	1
5	10/10/2018	19:38:54	2018	228	07/10/2018	1
6	10/10/2018	19:38:54	2018	228	07/10/2018	1
7	10/10/2018	19:38:54	2018	228	07/10/2018	1
8	10/10/2018	19:38:54	2018	228	07/10/2018	1
9	10/10/2018	19:38:54	2018	228	07/10/2018	1
10	10/10/2018	19:38:54	2018	228	07/10/2018	1
11	10/10/2018	19:38:54	2018	228	07/10/2018	1
12	10/10/2018	19:38:54	2018	228	07/10/2018	1
15	10/10/2018	19:38:54	2018	228	07/10/2018	1

Fonte: Elaborado pelo autor.

Realizada a importação dos *Steps*, foi necessário excluir o *Step* de interligar à DBpedia, conforme mencionado anteriormente, e criar um novo *Step* para o mapeamento dos municípios com a DBpedia. Após estas ações, a execução do *Workflow* foi realizada e seu resultado pode ser verificado pela Figura 45.

Figura 45 – *Workflow* de Alagoas com os *Steps* importados

FILTRO STEP
Tabela filtrada pela coluna DESCRIÇÃO
TIPO VOTÁVEL

SEMANTIFICAR
A relação semântica entre os dados e vocabulários foram estabelecidas.

GERENCIAR COLUNAS
Gerenciamento de colunas realizado.

INTERLIGAR À DBPEDIA
Os recursos da coluna NOME MUNICÍPIO foram conectados com a DBPEDIA.

SEMANTIFICAR
A relação semântica entre os dados e vocabulários foram estabelecidas.

SEMANTIFICAR
A relação semântica entre os dados e vocabulários foram estabelecidas.

SEMANTIFICAR
A relação semântica entre os dados e vocabulários foram estabelecidas.

SEMANTIFICAR
A relação semântica entre os dados e vocabulários foram estabelecidas.

Votação Alagoas

ORDENAR EXECUTAR WORKFLOW RESETAR SEMANTIFICAR EXPORTAR

linha	DT_GERACAO	HH_GERACAO	ANO_ELEICAO	CD_PLEITO	DT_PLEITO	NR_1
4	10/10/2018	19:38:54	2018	228	07/10/2018	1
5	10/10/2018	19:38:54	2018	228	07/10/2018	1
6	10/10/2018	19:38:54	2018	228	07/10/2018	1
7	10/10/2018	19:38:54	2018	228	07/10/2018	1
8	10/10/2018	19:38:54	2018	228	07/10/2018	1
9	10/10/2018	19:38:54	2018	228	07/10/2018	1
10	10/10/2018	19:38:54	2018	228	07/10/2018	1
11	10/10/2018	19:38:54	2018	228	07/10/2018	1
12	10/10/2018	19:38:54	2018	228	07/10/2018	1
15	10/10/2018	19:38:54	2018	228	07/10/2018	1

itens por página 10 1-10 de 757957

Fonte: Elaborado pelo autor.

Processo análogo ao aqui apresentado, foi realizado para alguns outros estados, como exemplo. Nesse sentido, é possível observar que o objetivo de automatizar e fornecer escala à produção e manutenção de Dados Abertos Conectados foi alcançado. Os artefatos criados, podem ser encontrados no endereço <<http://bit.ly/3iPs8Wg>>.

Com o intuito de se aferir a consonância da aplicação com as boas práticas, bem como avaliar a aderência aos níveis propostos pelo Plano de Dados Abertos 5 Estrelas, o próximo tópico apresenta uma relação entre tais indicadores, com os recursos presentes na aplicação.

5.3 CONSONÂNCIA COM O ESQUEMA 5 ESTRELAS E AS BOAS PRÁTICAS PARA PUBLICAÇÃO DE DADOS

O projeto e implementação da aplicação proposta nesta dissertação, se sustentou nas boas práticas para publicação de dados para alcançar o nível de 5 estrelas. Isto posto, este tópico tem por objetivo apresentar a relação entre as boas práticas para publicação dos dados

na *web*, boas práticas para publicação de dados conectados e a conformidade com o padrão de 5 estrelas.

5.3.1 Boas Práticas para Publicação de Dados Conectados

A adoção das Boas Práticas para Publicação de Dados Conectados, estabelecida como uma recomendação da W3C, tem como objetivo facilitar o desenvolvimento e a entrega de dados estruturados como dados abertos conectados. Seguir tais recomendações aumenta a confiança nos dados produzidos, ampliando seu potencial para inovação.

Nesse sentido, serão retomadas as Boas Práticas para Publicação de Dados Conectados, vistas no tópico 2.3.1.2, apresentando sua aderência ou não pela aplicação, bem como o expediente usado para atendê-lo, conforme o Quadro 14.

Quadro 14 – Aderência às Boas Práticas para Publicação de Dados Conectados

Boa Prática		Atendida	Descrição
1	Preparar os <i>stakeholders</i>	Sim	Entende-se que o arquivo “ <i>report</i> ” gerado pela aplicação, provê uma explicação do processo de criação e manutenção do conjunto de Dados Abertos Conectados, incluindo o autor, descrição, fonte primária e os <i>Steps</i> utilizados para transformar o conjunto de dados.
2	Selecionar um conjunto de dados	Sim	O conjunto de dados primários pode ser filtrado, mapeado e interligado, por meio do uso de <i>Steps</i> , quando necessário.
3	Modelar os dados	Sim	O objetivo geral da aplicação é a conversão de uma representação tabular dos dados em uma representação baseada em grafo. Esse processo de conversão é realizado utilizando um <i>Step</i> de mapeamento, por meio de ontologias e vocabulários para criar relações entre os objetos.
4	Especificar uma licença apropriada	Sim	Durante a criação do <i>Workflow</i> é possível a seleção da licença apropriada.
5	Construir bons URIs para Dados Conectados	Sim	Os recursos são mapeados utilizando-se URIs HTTP, contendo o nome dos objetos e a sua versão.

6	Utilizar um vocabulário padrão	Sim	A aplicação permite que vocabulários, desde que nos formatos RDF e OWL, possam ser utilizados durante o mapeamento semântico dos dados, para a descrição dos recursos e seus relacionamentos.
7	Converter dados para representações de Dados Conectados	Sim	A atualização do RDF para representação dos recursos permite que os dados sejam conectados a outras fontes de dados. O <i>Step</i> de Interligação com a DBpedia possibilita que os recursos do conjunto de dados sejam associados a recursos da DBpedia, por meio do predicado <code><owl:sameAs></code> .
8	Prover acesso automatizado para os dados	Não	Embora a aplicação realize com sucesso a produção de conjuntos de dados, não foram implementados recursos que permitam o acesso automatizado dos dados.
9	Anunciar para o público	Não	O processo de publicização dos dados não foi automatizado.
10	Reconhecer o contrato social	Não se aplica	Entende-se que essa atividade está ligada à administração do conjunto de dados, junto à entidade que o produziu.

Fonte: Elaborado pelo autor.

De acordo com o Quadro 14, é possível observar que foram atendidas as boas práticas 1, 2, 3, 4, 5, 6 e 7, e que cada uma delas contribuiu para o enriquecimento do conjunto de dados, proporcionando uma experiência mais rica e dinâmica, tanto para publicadores, quanto para consumidores desses conjuntos de dados.

O tópico seguinte apresenta a aderência da aplicação com relação às Boas Práticas para Publicação e Consumo de Dados na *Web*.

5.3.2 Boas Práticas para Publicação de Dados na *Web*

Outra recomendação da W3C, as Boas Práticas para Publicação de Dados na *web*, apresentam um conjunto de práticas abrangentes, não apenas para publicação de dados abertos conectados, mas para dados de forma geral. Cada boa prática apresenta uma melhoria, que quando implementada, tende a aperfeiçoar a maneira com que os dados publicados, serão consumidos tanto por humanos, quanto por agentes de *software*.

Nesse sentido, de forma análoga ao tópico anterior, serão retomadas as Boas Práticas para Publicação de Dados na *web*, vistas no tópico 2.3.1.1, apresentando sua aderência ou não, bem como o expediente usado para atendê-lo, conforme Quadro 15.

Quadro 15 – Aderência às Boas Práticas para Publicação de Dados na *Web*

Boa Prática		Atendida	Descrição
BP01	Forneça metadados	Sim	Entende-se que o arquivo “ <i>report</i> ” gerado pela aplicação, apresenta dados, tais como título, URI, data de publicação, autor, fonte primária, bem como provê uma explicação do processo de criação e manutenção do conjunto de Dados Abertos Conectados, por meio da descrição dos <i>Steps</i> utilizados para transformar o conjunto de dados. Ainda, são apresentadas informações estruturais e descritivas para o conjunto de dados.
BP02	Forneça metadados descritivos	Sim	
BP03	Forneça metadados estruturais	Sim	
BP04	Forneça informações sobre a licença dos dados	Sim	Durante a criação do <i>Workflow</i> é possível selecionar a licença apropriada., que poderá ser consultada no “ <i>report</i> ”.
BP05	Forneça informações de proveniência dos dados	Sim	A proveniência é atendida por meio da descrição de fonte original, criador, editor e data de publicação do conteúdo no arquivo “ <i>report</i> ”.
BP06	Forneça informações sobre a qualidade dos dados	Não	A aplicação não apresenta recursos quanto à qualidade dos dados.
BP07	Forneça indicadores de versões	Sim	É possível informar a versão do conjunto de dados no momento de criação do <i>Workflow</i> .
BP08	Forneça informações de histórico de versões	Não	A aplicação não apresenta recursos quanto à descrição do histórico de versões. Contudo, é possível a criação de versões para um <i>Workflow</i> , criando-se um novo <i>Workflow</i> , e importando-se os <i>Steps</i> realizados em <i>Workflows</i> anteriores, realizando as alterações necessárias.
BP09	Use URIs persistentes como identificadores de conjuntos de dados	Sim	É possível definir um URI HTTP para o conjunto de dados contidos em um <i>Workflow</i> . Os recursos são mapeados utilizando-se esse URI como base, seguida do nome do conceito e do objeto que a define.
BP10	Use URIs persistentes como identificadores dentro de conjuntos de dados	Sim	
BP11	Declare URIs para versões dos conjuntos de dados e séries	Sim	É possível definir um URI para o conjunto de dados, de acordo com sua versão.
BP12	Use formatos de dados padrão legíveis por máquina	Sim	Os dados gerados pela aplicação são escritos de acordo com o modelo de dados RDF e serializados nos formatos recomendados pela W3C: RDF/XML, JSON-LD e Turtle.

BP13	Use representações de dados neutros de localidade	Sim	Além de ser possível a descrição do idioma por um projeto, o <i>Step</i> de mapeamento possibilita que o usuário da aplicação selecione vocabulários adequados para a representação de dados com parâmetros de localidade, quando for o caso.
BP14	Forneça dados em múltiplos formatos	Sim	Os dados gerados pela aplicação são escritos de acordo com o modelo de dados RDF e serializados nos formatos recomendados pela W3C: RDF/XML, JSON-LD e Turtle.
BP15	Reúse vocabulários dando preferência aos padronizados	Sim	A aplicação permite que qualquer vocabulário, desde que nos formatos RDF e OWL, possam ser usados para realizar o mapeamento semântico dos dados. Assim, o nível de formalização pode ser adequado de acordo com cada contexto.
BP16	Escolha o nível de formalização certo	Sim	
BP17	Forneça <i>download</i> em massa	Não	Essa boa prática, apresenta melhoria do ponto de vista do consumidor dos dados. Assim, não se aplica para a aplicação. Do ponto de vista do produtor do conjunto de dados, cada <i>Workflow</i> contém o conjunto de <i>Steps</i> necessário para sua produção, sendo preciso que sua execução seja realizada de forma individual.
BP18	Forneça subconjuntos para conjuntos volumosos de dados	Sim	O objetivo geral da aplicação é a conversão de uma representação tabular dos dados em uma representação baseada em grafo. Esse processo de conversão é realizado por um <i>Step</i> de mapeamento, utilizando ontologias e vocabulários para criar relações entre os objetos. Antes de ser criado, o conjunto de dados primários pode ser filtrado, mapeado e interligado, por meio do uso de <i>Steps</i> , quando necessário. Isso permite que os conjuntos de dados sejam subdivididos.
BP19	Use negociação de conteúdo para fornecer dados em vários formatos	Não	A aplicação não apresenta recursos quanto à negociação de conteúdo.
BP20	Forneça acesso em tempo real	Não	A aplicação não apresenta recursos quanto ao acesso em tempo real dos dados.
BP21	Forneça dados atualizados	Não	A aplicação não apresenta recursos quanto à frequência de atualização dos dados.
BP22	Forneça uma explicação para dados que não estão disponíveis	Não	A aplicação não apresenta recursos quanto à explicação para dados que não estão disponíveis.

BP23	Torne os dados disponíveis usando uma API	Não	A aplicação não apresenta recursos para a disponibilização de dados por meio de APIs.
BP24	Use padrões da <i>web</i> como base das APIs	Não	
BP25	Forneça uma documentação completa para a API	Não	
BP26	Evite alterações significativas na API	Não	
BP27	Preserve os identificadores	Não	Como a aplicação não tem por objetivo a publicação dos dados, não foram implementados recursos para atender a essa boa prática.
BP28	Avalie a cobertura do conjunto de dados	Sim	Embora não seja possível determinar a quantidade de anos que recursos permanecem disponíveis, tais como os recursos conectados a DBpedia, os dados de origem, bem como vocabulários e ontologias utilizados são preservadas na pasta criada para cada Projeto.
BP29	Obtenha <i>feedback</i> dos consumidores de dados	Não	A aplicação não apresenta recursos para a obtenção de <i>feedbacks</i> .
BP30	Torne o <i>feedback</i> disponível	Não	
BP31	Enriqueça os dados gerando novos dados	Sim	É possível realizar o enriquecimento dos dados por meio do uso dos <i>Steps</i> . O <i>Step</i> de mapeamento, por exemplo, permite a descrição das relações entre os objetos. O <i>Step</i> de interligação com a DBpedia, por sua vez, permite que conjuntos de dados sejam conectados criando um conjunto de dados mais rico e abrangente.
BP32	Forneça apresentações complementares	Não	A aplicação não apresenta recursos para apresentações complementares.
BP33	Forneça <i>feedback</i> para o publicador original	Não	A aplicação não apresenta recursos para <i>feedback</i> ao publicador original.
BP34	Siga os termos de licença	Sim	A aplicação apresenta meios para o atendimento à licença.
BP35	Cite o publicador original	Sim	A fonte primária de dados pode ser encontrada no documento de “ <i>report</i> ”, fornecido conjuntamente com o conjunto de dados.

Fonte: Elaborado pelo autor.

Como apresentado pelo Quadro 15, foram atendidas as Boas Práticas para Publicação de Dados na *Web* 1,2,3,4,5,7,9,10,11,12,13,14,15,16,18,28,31,33 e 34. É válido destacar que as boas práticas no quadro, além de melhores práticas para publicação, também contemplam práticas para o consumo de dados na *web*. Como o objetivo da aplicação aqui apresentada era o de produção de dados para posterior publicação, boas as práticas de publicação como as descritas nos itens dezenove à vinte e sete, por exemplo, não foram atendidas.

A adoção das boas práticas apresentadas neste e no item anterior contribui para alcançar um nível mais alto no esquema de estrelas, objeto do próximo item.

5.3.3 Plano de Dados Abertos 5 Estrelas

O Plano de Dados Abertos 5 Estrelas proposto por Berners-Lee (2006), descrito no tópico 2.2, teve como objetivo incentivar a adoção de práticas formais para publicação de Dados Abertos Conectados.

No esquema proposto, quanto mais aberto, maior o número de estrelas para os dados e, por conseguinte, maior facilidade para enriquecimento e conexão dos dados. Desse modo, a seguir apresenta-se a lista, que indica quais estrelas foram atingidas:

★ Torne seus recursos disponíveis na *web* sob uma licença aberta: ao se cadastrar um *Workflow*, é possível selecionar, entre um rol de 10 licenças disponíveis, qual licença melhor representa os usos possíveis do conjunto de dados;

★★ Torne seus recursos disponíveis como dados estruturados: como a aplicação aceita apenas dados estruturados, nos formatos CSV, TSV e JSON, o nível de duas estrelas já está estabelecido;

★★★ Utilize formatos não proprietários: os conjuntos de dados gerados, como resultado da execução dos *Steps* e da aplicação, são estruturados seguindo o modelo RDF e serializados nos formatos não proprietários RDF/XML, JSON-LD e Turtle, atendendo assim ao nível de três estrelas;

★★★★ Utilize URIs para identificar recursos. Isso vai ajudar as pessoas a apontar para eles: é possível definir um URI HTTP para o conjunto de dados. Os recursos são mapeados utilizando-se esse URI como base, seguido do nome do conceito e do objeto que o define;

★★★★ Conecte seus dados com dados de outras pessoas para prover contexto: o uso do *Step* de interligação com a DBpedia, possibilita que os recursos do *Workflow*, sejam conectados a outros recursos, presentes na DBpedia.

Assim, conforme pode ser observado, a aplicação atingiu as 5 estrelas o que, portanto, indica mais facilidade para enriquecimento dos dados.

Considerando o exposto neste capítulo, foi possível observar que no rol das dez Boas Práticas para Publicação de Dados Conectados, sete foram atendidas. No que diz respeito às Boas Práticas para Publicação de Dados na *web*, das trinta e cinco, dezanove foram atendidas, isto porque, no universo dessa recomendação, há uma série de boas práticas direcionadas ao consumo de dados na *web*, o que não era a proposta deste trabalho. Por fim, ao tratarmos do esquema de 5 Estrelas, voltado para dados abertos, a aplicação alcançou o nível máximo proposto por Berners-Lee (2006).

Desta forma, o presente capítulo explicitou que a aplicação proposta e implementada neste trabalho, atingiu o objetivo de automatizar a produção e a manutenção de Dados Abertos Conectados, além de conferir escalabilidade e repetibilidade a esse processo apoiada nas recomendações da W3C para publicação de dados da *web*.

6 CONSIDERAÇÕES FINAIS

Esta dissertação vincula-se no âmbito do PPGEGC à área de Engenharia do Conhecimento, e à linha de Engenharia do Conhecimento Aplicada às Organizações, que dentre seus principais objetivos, busca implantar soluções nas organizações, sejam elas de caráter público ou privado.

Assim, este trabalho se propôs a pensar soluções orientadas para sanar o problema de apoio e escalabilidade à produção e manutenção de Dados Abertos Conectados, por meio da automatização de fluxos de trabalho.

Conforme se observou ao longo dos capítulos que compõem esta pesquisa, é necessário que os dados sejam publicados seguindo práticas, que permitam com que os dados possam ser compreendidos e utilizados por consumidores, além de ser passíveis de processamento por agentes de *software*.

Observou-se também que a produção e manutenção de Dados Abertos Conectados, por vezes é realizada como um processo *ad hoc*, executado de acordo com a *expertise* do usuário ou grupo que projeta a produção desses conjuntos de dados, dependendo de seu conhecimento na escrita de *scripts* e conhecimento de ferramentas proprietárias ou não, muitas vezes não portáteis, dificultando a aplicabilidade em distintos contextos.

A utilização do EGCFLOW por meio do emprego em cenário real das eleições brasileiras, demonstrou a aplicabilidade da aplicação e o cumprimento do objetivo proposto para o trabalho: apoiar e dar escala à produção e manutenção de conjunto de Dados Abertos Conectados, por meio da automatização de fluxos de trabalho.

A aplicação desenvolvida resultou em uma série de benefícios de ordem técnica, dentre os quais destacam-se: i) a produção de conjuntos de dados abertos conectados, sem a dependência de distintas ferramentas ou conhecimentos técnicos aprofundados; ii) a flexibilidade de utilização, possibilitando o uso para distintos casos, uma vez que o do sistema eleitoral brasileiro utilizado aqui, foi apenas um exemplo; iii) a criação e reutilização dos fluxos de trabalho, otimizando, dentre outros fatores, o tempo; e iv) a interface semiautomatizada para execução dos fluxos de trabalho.

De forma indireta a aplicação tem potencial para proporcionar outros benefícios quando adotada. Ao se analisar o que é possível realizar, para além da execução da aplicação, mas observando-se os benefícios de se utilizar os conjuntos de dados por ela produzidos, é possível destacar diversos benefícios relacionados à gestão para organizações de caráter tanto

público quanto privado, dentre os quais destacam-se: i) aumento do uso e reúso de dados; ii) aumento na transparência; iii) geração de valor; iv) *accountability*; v) aumento na qualidade de dados; vi) apoio à tomada de decisões; e vii) geração de conhecimento. Todos estes elementos contribuem, direta ou indiretamente, nos processos de gestão do conhecimento em distintas organizações.

Ainda assim, é importante destacar alguns limites da aplicação, como por exemplo, a necessidade de prover suporte a outros tipos de entrada além de CSV, TSV e JSON, tais como, XLS, XLSX e RDF. Além disso, a impossibilidade de lidar com: i) arquivos que não apresentem cabeçalhos, ou cabeçalhos repetidos; ii) um único cabeçalho para múltiplas colunas; iii) formatos tabulares com células mescladas; e iv) células vazias.

Nesse sentido, como forma de ampliar as contribuições que a aplicação apresenta e buscando aprimorar esta pesquisa, algumas variáveis apresentam-se como possibilidades de trabalhos futuros, quais sejam: i) a disponibilização de um *endpoint* para consulta SPARQL; ii) a integração com plataformas de publicação de dados tais como CKAN; iii) a interligação com outras bases de dados, além da DBpedia; iv) a utilização de conectivos lógicos para seleção dos recursos; v) a conexão de recursos que apresentem conceitos, ou propriedades; vi) a verificação de *Range* e *Domain* no *Step* de mapeamento; vii) a possibilidade de desfazer ou refazer o comando dos *Steps*; e por fim, viii) a ampliação da aderência às Boas Práticas.

Assim, considerando-se as múltiplas iniciativas de caráter internacional, como a *Open Government Partnership*; a previsão do direito de acesso à informação na Declaração Universal dos Direitos Humanos, da qual o Brasil é signatário, bem como na própria Constituição Federal de 1988; a Lei de Acesso à Informação e o Decreto que institui a Política de Dados Abertos do Poder Executivo Federal, é possível observar que há cada vez mais um movimento na direção para a expansão de uma *Web* de Dados.

Por fim, iniciativas como as elencadas ao longo deste trabalho e a própria aplicação aqui desenvolvida, caminham para o aprimoramento de uma *web* de dados, orientada para o uso, reúso, combinação e compartilhamento democrático dos dados produzidos na *web*.

REFERÊNCIAS

- ABEL, Mara; FIORINI, Sandro R. UMA REVISÃO DA ENGENHARIA DO CONHECIMENTO: EVOLUÇÃO, PARADIGMAS E APLICAÇÕES. **International Journal Knowledge Engineering Management**, Florianópolis, v. 2, n. 2, p. 1-35, mar./maio, 2013. Disponível em: <https://www.researchgate.net/publication/264156426_Uma_revisao_da_Engenharia_do_Conhecimento_Evolucao_Paradigmas_e_Aplicacoes> . Acesso em: 11 de Mar. de 2020.
- AOYAMA, M. *et al.* **PROMIS: A management platform for software supply networks based on the linked data and OSLC**. Proceedings - International Computer Software and Applications Conference. 2013.
- AUER, Sören. **Introduction to LOD2**. Chapter Linked Open Data - Creating Knowledge Out of Interlinked Data. Volume 8661 of the series Lecture Notes in Computer Science. p. 1-17, 2014.
- AZEVEDO, R. F. de B. C. **Um modelo ontológico do sistema eleitoral brasileiro**. Faculdade de Ciência da Informação, Universidade de Brasília, 2014. 208 p. Dissertação (Mestrado em Ciência da Informação). Disponível em:<https://repositorio.unb.br/bitstream/10482/17303/1/2014_RafaelFernandesdeBarrosCostaAzevedo.pdf>. Acesso em: 29 out. 2020.
- BASS, L.; CLEMENTS, P.; KAZMAN, R. **Software Architecture in Practice**. [S.l.: s.n.], 2012.
- BATISTA, Mateus G. R; LÓSCIO, Bernadette F. **OpenSBBD: Usando Linked Data para Publicação de Dados Abertos sobre o SBBD**. 2013. Disponível em: <https://sbbd2013.cin.ufpe.br/Proceedings/artigos/pdfs/sbbd_shp_10.pdf>. Acesso em: 25 de Mar. de 2020.
- BERG, Carlos H. **Ferramenta para Identificação de Emoções a partir de Onomatopéias para Pessoas com Diferentes Habilidades Visuais**. Tese (Doutorado em Engenharia e Gestão do Conhecimento) - Universidade Federal de Santa Catarina - UFSC. Florianópolis, p. 120. 2017. Disponível em: <<http://btd.egc.ufsc.br/wp-content/uploads/2017/10/Carlos-Henrique-Berg.pdf>>. Acesso em: 15 de Mar. de 2020.
- BERNERS-LEE, Tim. **Linked Data**. 2006. Disponível em: <<http://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: 20 de Mar. de 2020.
- BERNERS-LEE, T.; BIZER, C.; HEATH, T. Linked Data - The Story So Far. **International Journal on Semantic Web and Information Systems**, vol. 5, no. 3, 2009, pp. 1–22. Disponível em: <<https://eprints.soton.ac.uk/271285/>>. Acesso em: 09 de Mar. de 2020.
- BERNERS-LEE, Tim; FIELDING, R.; MASINTER, L. **Uniform Resource Identifier (URI): Generic Syntax**. 2005. Disponível em: <<https://www.rfc-editor.org/rfc/pdf/rfc3986.txt.pdf>>. Acesso em: 20 de Jun. de 2020.

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The Semantic Web A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American: The Semantic Web**, 2001. Disponível em: <https://www.researchgate.net/publication/225070375_The_Semantic_Web_A_New_Form_of_Web_Content_That_is_Meaningful_to_Computers_Will_Unleash_a_Revolution_of_New_Possibilities>. Acesso em: 09 de Mar. de 2020.

BIZER, Chris; CYGANIAK, Richard, HEATH, Tom. **How to publish linked data on the web**. 2007. Disponível em: <<http://wifo5-03.informatik.uni-mannheim.de/bizer/pub/LinkedDataTutorial/>>. Acesso em: 30 de Mar. de 2020.

BORST, Willem Nico. **Construction of Engineering Ontologies for Knowledge Sharing and Reuse**. 1997. 243 f. Tese. CIPDATA KONINKLIJKE BIBLIOTHEEK, Enschede, Netherlands, 1997.

BOTELHO, Maurício. **Aplicação de Ontologias na Organização de Conteúdos para Apoio a Equipes De Desenvolvimento De Software**. Dissertação (Mestrado em Engenharia e Gestão do Conhecimento) - Universidade Federal de Santa Catarina - UFSC. Florianópolis, p. 133. 2015. Disponível em: <<http://btd.egc.ufsc.br/wp-content/uploads/2015/11/Maur%C3%ADcio-Botelho.pdf>>. Acesso em: 15 de Mar. de 2020.

BRASIL. **LEI Nº 12.527, DE 18 DE NOVEMBRO DE 2011**. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/112527.htm>. Acesso em: 26 de Mar. de 2021.

BREITMAN, Karin K. **Web Semântica - A Internet do Futuro**. Rio de Janeiro: LTC, 2005.

BROEK, tijs van den; VEENSTRA, Anne Fleur; FOLMER, Erwin. **Walking the extra byte: A Lifecycle Model for Linked Open. Pilot Linked Open Data Nederland**. Linked Open Data, 2014.

CESCA, Renato. **Modelo de Repositório Institucional com Suporte a Web Semântica e Rede Social**. Dissertação (Mestrado em Engenharia e Gestão do Conhecimento) - Universidade Federal de Santa Catarina - UFSC. Florianópolis, p. 156. 2018. Disponível em: <<http://btd.egc.ufsc.br/wp-content/uploads/2018/05/Renato-Cesca.pdf>>. Acesso em: 15 de Mar. de 2020.

CHEN, Roland S. Exploring Performance Issues for a Clinical Database Organized Using an Entity-Attribute-Value Representation. **Journal of the American Medical Informatics Association**, v. 7, Issue 5, 2000, pp. 475-487. Disponível em: <<https://academic.oup.com/jamia/article/7/5/475/759278?login=true>>. Acesso em: 22 de Jun. de 2021.

CROCKFORD, Douglas. **The application/json Media Type for JavaScript Object Notation (JSON)**. 2006. Disponível em: <<https://datatracker.ietf.org/doc/html/rfc4627>>. Acesso em: 19 de Jun. de 2021.

CUNHA, Danusa R. B.; LÓSCIO, Bernadette F.; SOUZA, Damires. Linked Data: da Web de Documentos para a Web de dados. In: **Livro Texto dos Minicursos ERCEMAPI**, A. M. Santana *et al.*, SBC: Teresina, BR, p. 79-99, 2011. Disponível em: <https://www.cin.ufpe.br/~daise/arquivos/publications/2011/2011_Cap%204%20Linked%20Data%20-%20MC2.pdf>. Acesso em: 09 de Mar. de 2020.

CYGANIAK, R.; WOOD, D.; LANTHALER, M. **RDF 1.1 Concepts and Abstract Syntax**. W3C, 2014. Disponível em: <<http://www.w3.org/TR/rdf11-concepts/>>.

D'AQUIN, M.; NOY, N. Where to publish and find ontologies? A survey of ontology libraries. *Web Semantics: Science, Services and Agents on the World Wide Web*, v. 11, p. 96-111, 2012. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3293483/>>. Acesso em: 09 de Jul. de 2020.

DEVEDZIC, V. **Semantic Web and Education**. Berlin: Springer, 2006.

DIVU, Valentin; NADKARNI, Prakash; BRANDT, Cynthia. Pivoting approaches for bulk extraction of Entity–Attribute–Value data. **Computer Methods and Programs in Biomedicine**, v. 82, Issue 1, 2006, pp. 38-43. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/S0169260706000289>>. Acesso em: 23 de Jun. de 2021.

DRESCH, Aline; LACERDA, Daniel Pacheco; JUNIOR, Jose Antonio Valle Antunes. **Design Science Research: Método de Pesquisa para Avanço da Ciência e Tecnologia**. Porto Alegre: Bookman, 2015.

EAVES, David. **The Three Laws of Open Government Data**. 2009. Disponível em: <<https://eaves.ca/2009/09/30/three-law-of-open-government-data/>>. Acesso em: 15 de Ago. de 2020.

ECKERT, K. et al. **RESTful open workflows for data provenance and reuse**. WWW 2014 Companion - Proceedings of the 23rd International Conference on World Wide Web. 2014.

ELAASAR, M.; CONALLEN, J. **Design management: A collaborative design solution**. [s.l: s.n.]. v. 7949 LNCS. 2013.

ESCOLA NACIONAL DE ADMINISTRAÇÃO PÚBLICA - ENAP. **Elaboração de Plano de Dados Abertos - Módulo 1 - Conceitos de Dados Abertos**. Brasília: ENAP, 2017.

5STARDATA. **5 ★ DOS DADOS ABERTOS**. 2012. Disponível em: <<https://5stardata.info/pt-BR/>>. Acesso em: 20 de Ago. de 2020.

FERREIRA, Fábio *et al.* **Especificação de Interfaces Aplicacionais REST**. Actas do 9º Encontro Nacional de Informática, INFORUM, 2017. Disponível em:

<http://rss.di.fc.ul.pt/wp-content/uploads/2017/12/headrest_inforum.pdf>. Acesso em: 19 de Jun. de 2021.

FIELDING, Roy Thomas. **Architectural Styles and the Design of Network-based Software Architectures**. Tese (Doutorado em Ciência da Computação e Informação) - University of California. Irvine, p. 180. 2000. Disponível em: <https://www.ics.uci.edu/~fielding/pubs/dissertation/fielding_dissertation.pdf>. Acesso em: 09 de Abr. de 2021.

FIELDING, Roy Thomas *et al.* **Reflections on the REST Architectural Style and “Principled Design of the Modern Web Architecture”** (Impact Paper Award). ESEC/FSE’17, September 4–8, 2017, Paderborn, Germany. Disponível em: <<https://dl.acm.org/doi/pdf/10.1145/3106237.3121282>>. Acesso em: 19 de Jun. de 2021.

GALKIN, M.; MOUROMTSEV, D.; AUER, S. **Identifying web tables: Supporting a neglected type of content on the web**. [s.l: s.n.]. v. 518. 2015.

GARIJO, D.; GIL, Y. **A new approach for publishing workflows: Abstractions, standards, and linked data**. WORKS’11 - Proceedings of the 6th Workshop on Workflows in Support of Large-Scale Science, Co-located with SC’11. 2011.

GARIJO, D.; GIL, Y.; CORCHO, O. Abstract, link, publish, exploit: An end to end framework for workflow sharing. **Future Generation Computer Systems**, v. 75, p. 271–283, 2017.

GLOBAL OPEN DATA INDEX. **Brazil**. 2017. Disponível em: <<https://index.okfn.org/place/br/>>. Acesso em: 26 de Mar. de 2021.

_____. **Brazil**. 2015. Disponível em: <<http://2015.index.okfn.org/place/brazil/>>. Acesso em: 26 de Mar. de 2021.

GOLDSTEIN, Brett; DYSON, Lauren (Ed.). **Beyond Transparency: Open Data and the Future of Civic Innovation**. San Francisco: Code For America Press, 2013.

GOMES, Murilo S. **Proposta de Arquitetura para Ecosistema de Inovação em Dados Abertos**. Dissertação (Mestrado em Engenharia e Gestão do Conhecimento) - Universidade Federal de Santa Catarina - UFSC. Florianópolis, p. 104. 2017. Disponível em: <<http://btd.egc.ufsc.br/wp-content/uploads/2017/04/Murilo-Gomes.pdf>>. Acesso em: 15 de Mar. de 2020.

GONZÁLEZ, A.; VILLAZÓN-TERRAZAS, B.; GÓMEZ, J. M. **A linked data lifecycle for smart cities in Spain**. CEUR Workshop Proceedings. 2014.

GRAPHDB. **Loading data using OntoRefine**. 2021. Disponível em: <<https://graphdb.ontotext.com/documentation/free/loading-data-using-ontorefine.html#create-a-project>>. Acesso em: 09 de Abr. 2021.

GRUBER, T. R. **A Translation Approach to Portable Ontology Specifications**. Knowledge Acquisition. 5, v. 2, p. 199-220, 1993.

HAM, Kelli. OpenRefine (version 2.5). Free, open-source tool for cleaning and transforming data. **J Med Lib Assoc** 101(3), Julho 2013. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3738091/>>. Acesso em: 09 de Abr. de 2021.

HORROCKS, I. Ontologies and the Semantic Web. **Communication of the ACM**. 2008, v. 51, 12, p. 58-67. Disponível em: <https://dl.acm.org/doi/pdf/10.1145/1409360.1409377?casa_token=IFy-rc42KawAAAAA:ShpTx2irhBSRAPNx8tvyuwjblmxthwY0Ux8yKkEIVF7KYuXKeC5zdZl3g1lJQBmiiS2I7yLrTrL8278>. Acesso em: 25 de Jun. de 2020.

HYLAND, Bernadette; ATEMEZING, Ghislain; VILLAZÓN-TERRAZAS, Boris. **Best Practices for Publishing Linked Data**. W3C Working Group Note 09 January 2014. Disponível em: <<https://www.w3.org/TR/ld-bp/>>. Acesso em: 09 de Abr. de 2021.

HYLAND, Bernadette; VILLAZÓN-TERRAZAS, Boris. **Linked Data Cookbook**. 2011. Disponível em: <https://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook>. Acesso em: 09 de Abr. de 2021.

ISOTANI, Seiji; BITTENCOURT, Ig Ibert. **DADOS ABERTOS CONECTADOS**. São Paulo: NOVATEC Editora, 2015.

JACOBS, Ian; WALSH, Norman. **Architecture of the World Wide Web, Volume One**. W3C Recommendation 15 December 2004. Disponível em: <<https://www.w3.org/TR/webarch/>>. Acesso em: 20 de Jun. de 2020.

JOHNSON, Rod *et al.* **Spring Framework Reference Documentation**. 2018. Disponível em: <<https://docs.spring.io/spring-framework/docs/3.2.19.BUILD-SNAPSHOT/spring-framework-reference/pdf/spring-framework-reference.pdf>>. Acesso em: 09 de Abr. de 2021.

KALAMPOKIS, E. et al. **Creating and utilizing linked open statistical data for the development of advanced analytics services**. CEUR Workshop Proceedings. 2014.

KLÍMEK, J.; ŠKODA, P. **LinkedPipes ETL in use: Practical publication and consumption of Linked Data**. ACM International Conference Proceeding Series. 2017.

KONTOKOSTAS, D. *et al.* **Semantically enhanced quality assurance in the JURION business use case**. [s.l: s.n.]. v. 9678. 2016.

JIANG, Lin; QIU, Junqiao; ZHAO, Zhijia. **Scalable Structural Index Construction for JSON Analytics**. Proceedings of the VLDB Endowment, Volume 14, Issue 4, December 2020, pp. 694–707. 2020. Disponível em: <<https://dl.acm.org/doi/pdf/10.14778/3436905.3436926>>. Acesso em: 19 de Jun. de 2021.

LACERDA, Daniel Pacheco *et al.* Design Science Research: método de pesquisa para a engenharia de produção. **Gestão e Produção**, v. 20, p. 741–761, 2013. Disponível em: <

<https://www.scielo.br/j/gp/a/3CZmL4JJxLmxCv6b3pnQ8pq/?lang=pt>>. Acesso em: 20 de Jul. de 2021.

LANSING, C. *et al.* **Designing the cloud-based DOE systems biology knowledgebase.** IEEE International Symposium on Parallel and Distributed Processing Workshops and Phd Forum. 2011.

LAUFER, Carlos. **Guia de Web Semântica.** São Paulo: SPUK, 2015.

LEE, Dreirde. **Building an Open Data Ecosystem – An Irish Experience.** ICEGOV '14: Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance. 2014. Disponível em: <<https://dl.acm.org/doi/abs/10.1145/2691195.2691258>>. Acesso em: 09 de Abr. de 2021.

LEE, Y. **A life-cycle workflow architecture for Linked Data.** ACM International Conference Proceeding Series. 2017.

LEONARDI, Juliana. **Modelo para Avaliação de Relações Dimensionais na Criação de Conhecimento Organizacional.** Tese (Doutorado em Engenharia e Gestão do Conhecimento) - Universidade Federal de Santa Catarina - UFSC. Florianópolis, p. 169. 2016. Disponível em: <<http://btd.egc.ufsc.br/wp-content/uploads/2017/02/Juliana-Leonardi.pdf>>. Acesso em: 15 de Mar. de 2020.

LÓSCIO, Bernadette; BURLE, Caroline; CALEGARI, Newton. **Data on the Web Best Practices.** W3C Recommendation 31 January 2017. Disponível em: <<https://www.w3.org/TR/dwbp/#bib-LD-BP>>. Acesso em: 09 de Abr. de 2021.

LÓSCIO, Bernadette *et al.* **Fundamentos para publicação de dados na web.** São Paulo: Comitê Gestor da Internet no Brasil, 2018.

MAHMUD, Hasan *et. al.* CSV2RDF: GENERATING RDF DATA FROM CSV FILE USING SEMANTIC WEB TECHNOLOGIES. **Journal of Theoretical and Applied Information Technology**, v. 96, n. 20, 2018. Disponível em: <<http://www.jatit.org/volumes/Vol96No20/19Vol96No20.pdf>>. Acesso em: 20 de Jun. de 2021.

MANOLA, Frank; MILLER, Eric; MCBRIDE, Brian. **RDF 1.1 Primer.** W3C Recommendation 2014. Disponível em: <<https://www.w3.org/TR/rdf11-primer/>>. Acesso em: 12 de Ago. de 2020.

MARCH, S. T.; SMITH, G. F. Design and Natural Science Research on Information Technology. **Decision Support Systems**, v. 15, p. 251-266, 1995.

MARSHALL, M. S. *et al.* Emerging practices for mapping and linking life sciences data using RDF - A case series. **Journal of Web Semantics**, v. 14, p. 2–13, 2012.

MCGUINNESS, D. L. **Question Answering on the Semantic Web.** IEEE Intelligent Systems, v. 19, n. 1, p. 82-85, 2004.

MIZOGUCHI, R. Tutorial on ontological engineering: part 3: Advanced course of ontological engineering. **New Generation Computing**, v. 22, n. 2, p. 198-220, 2004. Disponível em: <https://www.researchgate.net/publication/220618894_Tutorial_on_Ontological_Engineering_Part_3_Advanced_Course_of_Ontological_Engineering>. Acesso em: 09 de Jul. de 2020.

MORAIS, Jean Carlos de; MORAIS, Jhonatan Carlos de. **SYSLODFLOW – UMA FERRAMENTA DE APOIO A AUTOMAÇÃO DA PUBLICAÇÃO E MANUTENÇÃO DE LINKED DATA UTILIZANDO O LODFLOW**. Monografia (Graduação em Sistemas de Informação) - Universidade Federal de Santa Catarina - UFSC. Florianópolis, p. 103. 2016.

MySQL. **The JSON Data Type**. 2021. Disponível em: <<https://dev.mysql.com/doc/refman/8.0/en/json.html>>. Acesso em: 30 de Jun. de 2021.

NAPOLI, Marcio. **Aplicação de ontologias para apoiar operações analíticas sobre fontes estruturadas e não estruturada**. Dissertação (Mestrado em Engenharia e Gestão do Conhecimento) - Universidade Federal de Santa Catarina - UFSC. Florianópolis, p. 133. 2011. Disponível em: <<https://repositorio.ufsc.br/bitstream/handle/123456789/95978/291294.pdf?sequence=1&isAllowed=y>>. Acesso em: 15 de Mar. de 2020.

NUFFELEN, Bert Van *et al.* Supporting the Linked Data Life Cycle Using an Integrated Tool Stack. In: AUER, Sören; BRYL, Volha; TRAMP, Sebastian (eds.). **Linked Open Data – Creating Knowledge Out of Interlinked Data** - Results of the LOD2 Project. New York: Springer, 2014.

OHIRA, Masanao. **Ferramenta para análise do estado de evolução do conhecimento em organizações**. Tese (Doutorado em Engenharia e Gestão do Conhecimento) - Universidade Federal de Santa Catarina - UFSC. Florianópolis, p. 118. 2009. Disponível em: <<http://btd.egc.ufsc.br/wp-content/uploads/2010/06/Masanao-Ohira.pdf>>. Acesso em: 15 de Mar. de 2020.

OPEN DEFINITION. **Open Definition - version 1.1**. 2009. Disponível em: <<https://opendefinition.org/od/1.1/pt/>>. Acesso em: 15 de Ago. de 2020.

OPEN GOVERNMENT PARTNERSHIP. **Quiénes somos**. Disponível em: <<https://www.opengovpartnership.org/es/>>. Acesso em: 26 de Mar. de 2021.

OPEN KNOWLEDGE. **Open Data Handbook**. 2012. Disponível em: <<http://opendatahandbook.org/>>. Acesso em: 01 out. 2020.

OPENREFINE. **Download**. 2021. Disponível em: <<https://openrefine.org/download.html>>. Acesso em: 09 de Abr. de 2021.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS - ONU. **A World that Counts Mobilising the Data Revolution for Sustainable Development**. 2014. Disponível em: <<https://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf>>. Acesso em: 10 de Mar. de 2020.

PACHECO, Roberto Carlos dos S. **Dados e Governo Abertos na Sociedade do Conhecimento**. LOD BRASIL - Linked Open Data - Brasil. Florianópolis, 2014. Disponível em: <<http://www.inf.ufsc.br/~jose.todesco/LODBrasil/Abertura/DadosEGovernoAbertoNaSocConh.pdf>>. Acesso em: 11 de Mar. de 2020.

PEFFERS, Ken *et al.* A design science research methodology for information systems research. **Journal of Management Information Systems**, v. 24, n. 3, p. 45-77, 2007. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.535.7773&rep=rep1&type=pdf>>. Acesso em: 06 de Out. de 2020.

PIRES, Marco Túlio. **Guia de dados abertos**. São Paulo: SPUK, 2015.

PIZZOL, Leandro Dal. **Uso da Web de Dados como Fonte de Informação no Processo de Inteligência Competitiva Setorial**. Dissertação (Mestrado em Engenharia e Gestão do Conhecimento) - Universidade Federal de Santa Catarina - UFSC. Florianópolis, p. 138. 2014. Disponível em: <<http://btd.egc.ufsc.br/wp-content/uploads/2015/04/Leandro-Dal-Pizzol.pdf>>. Acesso em: 15 de Mar. de 2020.

PRESSMAN, Roger S., MAXIM, Bruce R. **Engenharia de software: uma abordagem profissional**. 8. ed. Porto Alegre: AMGH, 2016.

PORTAL BRASILEIRO DE DADOS ABERTOS. **O que são dados abertos?** 2020. Disponível em: <<https://dados.gov.br/pagina/dados-abertos>>. Acesso em: 15 de Ago. de 2020.

RAUTENBERG, S. *et al.* **Guia Prático para Publicação de Dados Abertos Conectados na Web**. Curitiba: Appris, 2018.

RAUTENBERG, S. *et al.* **LDWPO - A lightweight ontology for linked data management**. CEUR Workshop Proceedings. 2016a.

RAUTENBERG, S. *et al.* **Linked Data Workflow Project Ontology: Uma Ontologia de Domínio para Publicação e Preservação de Dados Conectados**. XVII Encontro Nacional de Pesquisa em Ciência da Informação (XVII ENANCIB). 2016b.

RAUTENBERG, S. *et al.* **LODFlow: a workflow management system for linked data processing**. Proceedings of the 11th International Conference on Semantic Systems, 2015.

RAUTENBERG, Sandro; BURDA, Alessandra C.; SOUZA, Lucélia de. Um workflow para compartilhamento de dados científicos primários como Dados Abertos Conectados. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 23, n. 53, p.110-123, set./dez., 2018. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2018v23n53p110/37293>>. Acesso em: 09 de Mar. de 2020.

REDHAT. **Interface de Programação de Aplicações - O que é API?** 2021. Disponível em: <<https://www.redhat.com/pt-br/topics/api/what-are-application-programming-interfaces>>. Acesso em: 09 de Abr. de 2021.

RICHARDSON, Leonard; *et al.* **RESTful Web API**. O'Reilly, 2013.

ROSS, Marco. **A Linked Data Approach to Sharing Workflows and Workflow Results**. International Symposium On Leveraging Applications of Formal Methods, Verification and Validation ISoLA, 2010. Disponível em: <https://link.springer.com/chapter/10.1007/978-3-642-16558-0_29>. Acesso em: 18 de Ago. de 2020.

ROTEM-GAL-OZ, A. **SOA Patterns**. [S.l.: s.n.], 2012.

RUEDIGER, Marco Aurélio.; MAZZOTTE, Natália (coordenadores). **Índice de dados abertos para cidades - 2018**. Rio de Janeiro: FGV, DAPP, 2018. Disponível em: <<https://www.ok.org.br/wp-content/uploads/2020/04/WEB-I%CC%81ndice-de-dados-abertos-1.pdf>>. Acesso em: 26 de Mar. de 2021.

SAUDATE, Alexandre. **REST - Construa API's inteligentes de maneira simples**. São Paulo: Casa do Código, 2013. 1. ed.

SHVETS, Alexander. **Dive Into Design Patterns**. Sem local: Refactoring.Guru, 2019.

SIMON, Herbert A. **The Sciences of the Artificial**. Cambridge: MIT Press, 1996.

SOMMERVILLE, Ian. **Engenharia de Software**. São Paulo: Pearson Prentice Hall, 2011. 9. ed.

SPERONI, Rafael de M. **Modelo de Referência para Indicadores de Inovação Regional Suportado por Dados Ligados**. Tese (Doutorado em Engenharia e Gestão do Conhecimento) - Universidade Federal de Santa Catarina - UFSC. Florianópolis, p. 233. 2016. Disponível em: <<http://btd.egc.ufsc.br/wp-content/uploads/2016/08/Rafael-de-Moura-Speroni-1.pdf>>. Acesso em: 15 de Mar. de 2020.

SPRING. **Spring Framework Overview**. 2021. Disponível em: <<https://docs.spring.io/spring-framework/docs/current/reference/html/overview.html#overview-philosophy>>. Acesso em: 09 de Abr. de 2021.

SWARTOUT, W.; TATE, A. **Ontologies**. IEEE Intelligent Systems, v. 1, n. 14, p. 18-19, 1999.

TORRES, D. Sistemas eleitorais brasileiros. **Revista Eletrônica da EJE**, ano 4, n. 4 (jun./jul. 2014), Tribunal Superior Eleitoral, 2014. Disponível em: <<https://www.tse.jus.br/o-tse/escola-judiciaria-eleitoral/publicacoes/revistas-da-eje/artigos/revista-eletronica-eje-n.-4-ano-4/sistemas-eleitorais-brasileiros>>. Acesso em: 29 out. 2020.

TRIBUNAL SUPERIOR ELEITORAL. **Repositório de Dados Eleitorais**. 2021. Disponível em: <<https://www.tse.jus.br/eleicoes/estatisticas/repositorio-de-dados-eleitorais-1/repositorio-de-dados-eleitorais>>. Acesso em: 09 de Jun. de 2021.

_____. **Leia-me – Boletim Urna Web**. 2018.

VATWANI, Ritesh. **What is REST API?** 2020. Disponível em: <<https://www.zestard.com/blog/rest-api-benefits/>>. Acesso em: 09 de Abr. de 2021.

VILLAZÓN-TERRAZAS, Boris *et al.* **Methodological Guidelines for Publishing Government Linked Data.** Linking Government Data. WOOD, David (Ed.), p. 27-49, 2012.

VISINTIN, Lidiane; GOMES, Murilo Silveira; TODESCO, José Leomar. **UM ESTUDO DOS CICLOS DE VIDA DE DADOS ABERTOS CONECTADOS.** I Workshop de Informação, Dados e Tecnologia, UFSC, Brasil, 2017. Disponível em: <<http://www.widat2017.ufsc.br/wp-content/uploads/2017/09/ST1.4.pdf>>. Acesso em: 09 de Abr. de 2021.

WEISSMANN, Henrique Lobo. **Vire o jogo com Spring Framework.** São Paulo: Casa do Código, 2014.

APÊNDICE A – Ontologia para o Sistema Eleitoral Brasileiro

```

<?xml version="1.0"?>
<rdf:RDF
  xml:base="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#"
  xmlns="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#"
  xmlns:eleicoes="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#"

  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xml="http://www.w3.org/XML/1998/namespace"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:foaf="http://xmlns.com/foaf/spec/"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
<owl:Ontology rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes"/>

<!--
////////////////////////////////////
//
// Object Properties
//
////////////////////////////////////
-->
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#concorre_ao_cargo_de -->

<owl:ObjectProperty
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#concorre_ao_cargo_de">
  <owl:inverseOf
    rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#eh_concorrido_por"/>
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Candidato"/>
  <rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Cargo"/>
</owl:ObjectProperty>

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#eh_candidato_em -->

<owl:ObjectProperty
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#eh_candidato_em">
  <owl:inverseOf
    rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#tem_candidato"/>

```

```

<rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Candidato"/>
<rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleicao"/>
<rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Turno"/>
</owl:ObjectProperty>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#eh_candidato_no_turno -->

```

```

<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#eh_candidato_no_turno">
  <owl:inverseOf
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#tem_candidato_turno"/>
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Candidato"/>
  <rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Turno"/>
</owl:ObjectProperty>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#eh_composta_por -->

```

```

<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#eh_composta_por">
  <owl:inverseOf
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#faz_parte_de"/>
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#TransitiveProperty"/>
</owl:ObjectProperty>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#eh_concorrido_por -->

```

```

<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#eh_concorrido_por">
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Cargo"/>
  <rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Candidato"/>
</owl:ObjectProperty>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#eh_endereco_de -->

```

```

<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#eh_endereco_de">
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#FunctionalProperty"/>
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Endereco"/>

```

```

<rdfs:range
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#LocalDeVotacao"/>
<rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#ZonaEleitoral"/>
</owl:ObjectProperty>

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#eh_turno_de -->

<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#eh_turno_de">
<owl:inverseOf
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#tem_turno"/>
<rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Turno"/>
<rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleicao"/>
</owl:ObjectProperty>

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#eh_zona_de -->

<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#eh_zona_de">
<owl:inverseOf
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#tem_zona"/>
<rdfs:domain
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#ZonaEleitoral"/>
<rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleicao"/>
</owl:ObjectProperty>

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#esta_localizado_em -->

<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#esta_localizado_em">
<owl:inverseOf rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#possui"/>
<rdf:type rdf:resource="http://www.w3.org/2002/07/owl#TransitiveProperty"/>
<rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Municipio"/>
<rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Secao"/>
<rdfs:domain
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#ZonaEleitoral"/>
<rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Endereco"/>
<rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Estado"/>

```

```

<rdfs:range
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#LocalDeVotacao"/>
<rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Municipio"/>
<rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#ZonaEleitoral"/>
</owl:ObjectProperty>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#faz_parte_de -->

```

```

<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#faz_parte_de">
<rdf:type rdf:resource="http://www.w3.org/2002/07/owl#TransitiveProperty"/>
</owl:ObjectProperty>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#possui -->

```

```

<owl:ObjectProperty rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#possui">
<rdf:type rdf:resource="http://www.w3.org/2002/07/owl#TransitiveProperty"/>
<rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Estado"/>
<rdfs:domain
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#LocalDeVotacao"/>
<rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Municipio"/>
<rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Endereco"/>
<rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Estado"/>
<rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Municipio"/>
<rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Secao"/>
<rdfs:comment>Relacao de localizacao</rdfs:comment>
</owl:ObjectProperty>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#realiza_voto -->

```

```

<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#realiza_voto">
<owl:inverseOf
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#voto_realizado_por"/>
<rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleitor"/>

```

```
<rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Voto"/>
</owl:ObjectProperty>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#tem_candidato -->
```

```
<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#tem_candidato">
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleicao"/>
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Turno"/>
  <rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Candidato"/>
</owl:ObjectProperty>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#tem_candidato_turno -->
```

```
<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#tem_candidato_turno">
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Turno"/>
  <rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Candidato"/>
</owl:ObjectProperty>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#tem_eleitor -->
```

```
<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#tem_eleitor">
  <owl:inverseOf
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#vota_em_secao"/>
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Secao"/>
  <rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleitor"/>
</owl:ObjectProperty>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#tem_turno -->
```



```

<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#tem_turno">
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleicao"/>
  <rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Turno"/>
</owl:ObjectProperty>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#tem_zona -->

```

```

<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#tem_zona">
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleicao"/>
  <rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#ZonaEleitoral"/>
</owl:ObjectProperty>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#tem_secao -->

```

```

<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#tem_secao">
  <rdfs:domain
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#ZonaEleitoral"/>
  <rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Secao"/>
</owl:ObjectProperty>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#tem_urna -->

```

```

<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#tem_urna">
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Secao"/>
  <rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Urna"/>
</owl:ObjectProperty>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#vota_em_secao -->

```

```

<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#vota_em_secao">

```

```

<rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleitor"/>
<rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Secao"/>
</owl:ObjectProperty>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#voto_realizado_por -->

```

```

<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#voto_realizado_por">
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Voto"/>
  <rdfs:range rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleitor"/>
</owl:ObjectProperty>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#tem_voto -->

```

```

<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#tem_voto">
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Candidato"/>
</owl:ObjectProperty>

```

```

<!--
////////////////////////////////////
//
// Data properties
//
////////////////////////////////////
-->

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_apelido -->

```

```

<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_apelido">
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Candidato"/>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>

```

```
</owl:DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_nome -->
```

```
<owl:DatatypeProperty
```

```
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_nome">
```

```
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Candidato"/>
```

```
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
```

```
</owl:DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#abrangencia -->
```

```
<owl:DatatypeProperty
```

```
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#abrangencia">
```

```
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleicao"/>
```

```
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
```

```
</owl:DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_numero -->
```

```
<owl:DatatypeProperty
```

```
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_numero">
```

```
  <rdfs:type rdf:resource="http://www.w3.org/2002/07/owl#FunctionalProperty"/>
```

```
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Candidato"/>
```

```
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#integer"/>
```

```
</owl:DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#zona_numero -->
```

```
<owl:DatatypeProperty
```

```
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#zona_numero">
```

```
  <rdfs:type rdf:resource="http://www.w3.org/2002/07/owl#FunctionalProperty"/>
```

```
  <rdfs:domain
```

```
  rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#ZonaEleitoral"/>
```

```
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#integer"/>
```

```
</owl:DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#urna_numero -->
```

```
<owl:DatatypeProperty
```

```
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#urna_numero">
```

```
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#FunctionalProperty"/>
```

```
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Urna"/>
```

```
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#integer"/>
```

```
</owl:DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#secao_numero -->
```

```
<owl:DatatypeProperty
```

```
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#secao_numero">
```

```
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#FunctionalProperty"/>
```

```
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Secao"/>
```

```
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#integer"/>
```

```
</owl:DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_nome -->
```

```
<owl:DatatypeProperty
```

```
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_nome">
```

```
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Coligacao"/>
```

```
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
```

```
</owl:DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#eleicao_ano -->
```

```
<owl:DatatypeProperty
```

```
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#eleicao_ano">
```

```
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#FunctionalProperty"/>
```

```
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleicao"/>
```

```
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#integer"/>
```

```
</owl:DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#eleitor_nome -->
```

```
<owl:DatatypeProperty
```

```
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#eleitor_nome">
```

```
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleitor"/>
```

```
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
```

```
</owl:DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#numero_de_eleitores -->
```

```
<owl:DatatypeProperty
```

```
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#numero_de_eleitores">
```

```
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleicao"/>
```

```
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Estado"/>
```

```
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Municipio"/>
```

```
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Secao"/>
```

```
  <rdfs:domain
```

```
  rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#ZonaEleitoral"/>
```

```
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
```

```
</owl:DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#numero_de_votos -->
```

```
<owl:DatatypeProperty
```

```
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#numero_de_votos">
```

```
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Estado"/>
```

```
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Municipio"/>
```

```
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Secao"/>
```

```
  <rdfs:domain
```

```
  rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#ZonaEleitoral"/>
```

```
</owl:DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_legenda -->
```

```
<owl:DatatypeProperty
```

```
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_legenda">
```

```
  <owl:equivalentProperty
```

```
    rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_sigla"/>
```

```
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Partido"/>
```

```
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
```

```
</owl:DatatypeProperty>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_nome -->
```

```
<owl:DatatypeProperty
```

```
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_nome"/>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_sigla -->
```

```
<owl:DatatypeProperty
```

```
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_sigla">
```

```
  <rdfs:domain rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Partido"/>
```

```
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
```

```
</owl:DatatypeProperty>
```

```
<!--
```

```
////////////////////////////////////
```

```
//
```

```
// Classes
```

```
//
```

```
////////////////////////////////////
```

```
-->
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Candidato -->
```

```
<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Candidato">
  <rdfs:subClassOf rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
  <rdfs:subClassOf>
  <owl:Restriction>
    <owl:onProperty
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#faz_parte_de"/>
      <owl:someValuesFrom
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Partido"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
  <owl:Restriction>
    <owl:onProperty
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#concorre_ao_cargo_de"/>
      <owl:allValuesFrom
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Cargo"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Cargo -->
```

```
<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Cargo"/>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Ambito -->
```

```
<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Ambito"/>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Urna -->
```

```
<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Urna"/>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Coligacao -->
```

```
<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Coligacao">
  <rdfs:subClassOf rdf:resource="http://xmlns.com/foaf/0.1/Group"/>
</owl:Class>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#DeputadoEstadual -->
```

```
<owl:Class
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#DeputadoEstadual">
  <rdfs:subClassOf rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Cargo"/>
  <rdfs:comment>O mesmo que legenda</rdfs:comment>
</owl:Class>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#DeputadoFederal -->
```

```
<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#DeputadoFederal">
  <rdfs:subClassOf rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Cargo"/>
</owl:Class>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleicao -->
```

```
<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleicao">
  <rdfs:subClassOf>
  <owl:Restriction>
    <owl:onProperty
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#tem_candidato"/>
    <owl:someValuesFrom
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Cargo"/>
  </owl:Restriction>
</rdfs:subClassOf>
<rdfs:subClassOf>
```



```

<owl:Restriction>
  <owl:onProperty
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#tem_turno"/>
  <owl:minQualifiedCardinality
rdf:datatype="http://www.w3.org/2001/XMLSchema#nonNegativeInteger">1</owl:minQualifiedCardinality>
  <owl:onClass rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Turno"/>
</owl:Restriction>
</rdfs:subClassOf>
</owl:Class>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleitor -->

```

```

<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleitor">
  <rdfs:subClassOf rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
  <rdfs:subClassOf>
  <owl:Restriction>
    <owl:onProperty
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#vota_em_secao"/>
    <owl:someValuesFrom
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Secao"/>
  </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Endereco -->

```

```

<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Endereco">
  <rdfs:subClassOf>
  <owl:Restriction>
    <owl:onProperty
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#eh_endereco_de"/>
    <owl:someValuesFrom
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#LocalDeVotacao"/>
  </owl:Restriction>
  </rdfs:subClassOf>

```

```

<rdfs:subClassOf>
<owl:Restriction>
  <owl:onProperty
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#eh_endereco_de"/>
  <owl:someValuesFrom
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#ZonaEleitoral"/>
  </owl:Restriction>
</rdfs:subClassOf>
</owl:Class>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Estado -->

```

```

<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Estado">
<rdfs:subClassOf>
<owl:Restriction>
  <owl:onProperty rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#possui"/>
  <owl:someValuesFrom
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Municipio"/>
  </owl:Restriction>
</rdfs:subClassOf>
</owl:Class>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Estadual -->

```

```

<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Estadual">
<rdfs:subClassOf rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Ambito"/>
<owl:disjointWith
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Federal"/>
</owl:Class>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Executivo -->

```

```

<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Executivo">

```

```

<rdfs:subClassOf rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Poder"/>
<owl:disjointWith
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Legislativo"/>
</owl:Class>

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Federal -->

<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Federal">
<rdfs:subClassOf rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Ambito"/>
</owl:Class>

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Governador -->

<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Governador">
<rdfs:subClassOf rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Cargo"/>
</owl:Class>

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Legislativo -->

<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Legislativo">
<rdfs:subClassOf rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Poder"/>
</owl:Class>

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#LocalDeVotacao -->

<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#LocalDeVotacao">
<rdfs:subClassOf>
<owl:Restriction>
<owl:onProperty
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#esta_localizado_em"/>

```

```

    <owl:someValuesFrom
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Endereco"/>
    </owl:Restriction>
</rdfs:subClassOf>
<rdfs:subClassOf>
<owl:Restriction>
    <owl:onProperty
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#esta_localizado_em"/>
    <owl:someValuesFrom
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Municipio"/>
    </owl:Restriction>
</rdfs:subClassOf>
<rdfs:subClassOf>
<owl:Restriction>
    <owl:onProperty
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#faz_parte_de"/>
    <owl:someValuesFrom
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#ZonaEleitoral"/>
    </owl:Restriction>
</rdfs:subClassOf>
<rdfs:subClassOf>
<owl:Restriction>
    <owl:onProperty rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#possui"/>
    <owl:someValuesFrom
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Secao"/>
    </owl:Restriction>
</rdfs:subClassOf>
</owl:Class>

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Municipio -->

<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Municipio">
    <rdfs:subClassOf>
    <owl:Restriction>
    <owl:onProperty
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#esta_localizado_em"/>

```

```

    <owl:someValuesFrom
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Estado"/>
  </owl:Restriction>
</rdfs:subClassOf>
</owl:Class>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Partido -->

```

```

<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Partido">
  <rdfs:subClassOf rdf:resource="http://xmlns.com/foaf/0.1/Organization"/>
  <rdfs:subClassOf>
  <owl:Restriction>
    <owl:onProperty
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#faz_parte_de"/>
      <owl:someValuesFrom
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Coligacao"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Poder -->

```

```

<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Poder"/>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Presidente -->

```

```

<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Presidente">
  <rdfs:subClassOf rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Cargo"/>
</owl:Class>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Secao -->

```

```

<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Secao">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#esta_localizado_em"/>
        <owl:someValuesFrom
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#LocalDeVotacao"/>
      </owl:Restriction>
    </rdfs:subClassOf>
  </owl:Class>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Senador -->

```

```

<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Senador">
  <rdfs:subClassOf rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Cargo"/>
</owl:Class>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Turno -->

```

```

<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Turno"/>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Voto -->

```

```

<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Voto">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#voto_realizado_por"/>
        <owl:someValuesFrom
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleitor"/>
      </owl:Restriction>
    </rdfs:subClassOf>

```

```
</owl:Class>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#VotoPorCandidato -->
```

```
<owl:Class
```

```
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#VotoPorCandidato">
```

```
  <rdfs:subClassOf rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Voto"/>
```

```
</owl:Class>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#VotoPorSecao -->
```

```
<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#VotoPorSecao">
```

```
  <rdfs:subClassOf rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Voto"/>
```

```
</owl:Class>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#VotosAnulados -->
```

```
<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#VotosAnulados">
```

```
  <rdfs:subClassOf
```

```
  rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#VotoPorSecao"/>
```

```
</owl:Class>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#VotosBranco -->
```

```
<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#VotosBranco">
```

```
  <rdfs:subClassOf
```

```
  rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#VotoPorSecao"/>
```

```
</owl:Class>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#VotosNulos -->
```

```
<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#VotosNulos">
  <rdfs:subClassOf
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#VotoPorSecao"/>
</owl:Class>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#VotosVálidos -->
```

```
<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#VotosVálidos">
  <rdfs:subClassOf
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#VotoPorCandidato"/>
</owl:Class>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#ZonaEleitoral -->
```

```
<owl:Class rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#ZonaEleitoral">
  <rdfs:subClassOf>
  <owl:Restriction>
    <owl:onProperty rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#possui"/>
    <owl:someValuesFrom
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#LocalDeVotacao"/>
  </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
```

```
<!-- http://xmlns.com/foaf/0.1/Agent -->
```

```
<owl:Class rdf:about="http://xmlns.com/foaf/0.1/Agent"/>
```

```
<!-- http://xmlns.com/foaf/0.1/Group -->
```



```
<owl:Class rdf:about="http://xmlns.com/foaf/0.1/Group">
  <rdfs:subClassOf rdf:resource="http://xmlns.com/foaf/0.1/Agent"/>
</owl:Class>
```

```
<!-- http://xmlns.com/foaf/0.1/Organization -->
```

```
<owl:Class rdf:about="http://xmlns.com/foaf/0.1/Organization">
  <rdfs:subClassOf rdf:resource="http://xmlns.com/foaf/0.1/Agent"/>
</owl:Class>
```

```
<!-- http://xmlns.com/foaf/0.1/Person -->
```

```
<owl:Class rdf:about="http://xmlns.com/foaf/0.1/Person">
  <rdfs:subClassOf rdf:resource="http://xmlns.com/foaf/0.1/Agent"/>
</owl:Class>
```

```
<!--
////////////////////////////////////
//
// Individuals
//
////////////////////////////////////
-->
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleicoes2018 -->
```

```
<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleicoes2018">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleicao"/>
```

```
</owl:NamedIndividual>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Presidente -->
```

```
<owl:NamedIndividual
```

```
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Presidente">
```

```
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Presidente"/>
```

```
</owl:NamedIndividual>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_Alvaro_Dias -->
```

```
<owl:NamedIndividual
```

```
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_Alvaro_Dias">
```

```
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Candidato"/>
```

```
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleitor"/>
```

```
  <eleicoes:concorre_ao_cargo_de
```

```
    rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Presidente"/>
```

```
  <eleicoes:eh_candidato_em
```

```
    rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleicoes2018"/>
```

```
  <eleicoes:eh_candidato_no_turno
```

```
    rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#primeiro_turno"/>
```

```
  <eleicoes:faz_parte_de
```

```
    rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_podemos"/>
```

```
</owl:NamedIndividual>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_Cabo_Daciolo -->
```

```
<owl:NamedIndividual
```

```
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_Cabo_Daciolo">
```

```
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Candidato"/>
```

```
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleitor"/>
```

```
  <eleicoes:concorre_ao_cargo_de
```

```
    rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Presidente"/>
```

```

<eleicoes:faz_parte_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_patriota"/>
<eleicoes:candidato_apelido
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Cabo
Daciolo</eleicoes:candidato_apelido>
<eleicoes:candidato_nome
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Benevenuto Daciolo
Fonseca dos Santos</eleicoes:candidato_nome>
</owl:NamedIndividual>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_Ciro_Gomes -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_Ciro_Gomes">
<rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Candidato"/>
<rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleitor"/>
<eleicoes:concorre_ao_cargo_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Presidente"/>
<eleicoes:faz_parte_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_democratico_trabalhi
sta_PDT"/>
<eleicoes:candidato_apelido
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Ciro</eleicoes:candidato_apelido>
<eleicoes:candidato_nome
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Ciro
Gomes</eleicoes:candidato_nome>
</owl:NamedIndividual>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_Fernando_Haddad -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_Fernando_Haddad">
<rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Candidato"/>
<rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleitor"/>
<eleicoes:concorre_ao_cargo_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Presidente"/>
<eleicoes:eh_candidato_em
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleicoes2018"/>

```

```

<eleicoes:eh_candidato_no_turno
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#primeiro_turno"/>
<eleicoes:eh_candidato_no_turno
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#segundo_turno"/>
<eleicoes:faz_parte_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_dos_trabalhadores_PT"/>
<eleicoes:candidato_apelido
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Hadad</eleicoes:candidato_apelido>
<eleicoes:candidato_nome          rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Fernando
Hadad</eleicoes:candidato_nome>
</owl:NamedIndividual>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_Geraldo_Alckmin -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_Geraldo_Alckmin">
<rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Candidato"/>
<rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleitor"/>
<eleicoes:concorre_ao_cargo_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Presidente"/>
<eleicoes:faz_parte_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_da_social_democracia_brasileira_PSDB"/>
</owl:NamedIndividual>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_Guilherme_Boulos -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_Guilherme_Boulos">
<rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Candidato"/>
<rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleitor"/>
<eleicoes:concorre_ao_cargo_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Presidente"/>

```

```

<eleicoes:faz_parte_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_socialismo_e_liberda
de_PSOL"/>
</owl:NamedIndividual>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_Henrique_Meirelles -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_Henrique_Meirelles"
>
<rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Candidato"/>
<rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleitor"/>
<eleicoes:concorre_ao_cargo_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Presidente"/>
<eleicoes:faz_parte_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_movimento_democrat
ico_brasileiro_MDB"/>
</owl:NamedIndividual>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_Jair_Bolsonaro -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_Jair_Bolsonaro">
<rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Candidato"/>
<rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleitor"/>
<eleicoes:concorre_ao_cargo_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Presidente"/>
<eleicoes:eh_candidato_em
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleicoes2018"/>
<eleicoes:eh_candidato_no_turno
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#primeiro_turno"/>
<eleicoes:eh_candidato_no_turno
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#segundo_turno"/>

```

```

<eleicoes:faz_parte_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_social_liberal_PSL"/
>
<eleicoes:candidato_apelido
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Bolsonaro</eleicoes:candidato_apelido>
<eleicoes:candidato_nome rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Jair Messias
Bolsonaro</eleicoes:candidato_nome>
</owl:NamedIndividual>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_Joao_Amoedo -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_Joao_Amoedo">
<rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Candidato"/>
<rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleitor"/>
<eleicoes:concorre_ao_cargo_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Presidente"/>
<eleicoes:faz_parte_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_novo"/>
</owl:NamedIndividual>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_Marina_Silva -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#candidato_Marina_Silva">
<rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Candidato"/>
<rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleitor"/>
<eleicoes:concorre_ao_cargo_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Presidente"/>
<eleicoes:faz_parte_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_rede_sustentabilidade
"/>
</owl:NamedIndividual>

```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#cidade_Florianopolis -->
```

```
<owl:NamedIndividual
```

```
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#cidade_Florianopolis">
```

```
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Municipio"/>
```

```
  <eleicoes:esta_localizado_em
```

```
  rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#estado_SantaCatarina"/>
```

```
  <eleicoes:numero_de_eleitores
```

```
  rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">343645</eleicoes:numero_de_eleitores>
```

```
</owl:NamedIndividual>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#cidade_Joinville -->
```

```
<owl:NamedIndividual
```

```
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#cidade_Joinville">
```

```
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Municipio"/>
```

```
  <eleicoes:esta_localizado_em
```

```
  rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#estado_SantaCatarina"/>
```

```
  <eleicoes:tem_zona rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#019"/>
```

```
  <eleicoes:tem_zona rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#076"/>
```

```
  <eleicoes:tem_zona rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#095"/>
```

```
  <eleicoes:tem_zona rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#096"/>
```

```
  <eleicoes:tem_zona rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#105"/>
```

```
  <eleicoes:numero_de_eleitores
```

```
  rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">396719</eleicoes:numero_de_eleitores>
```

```
</owl:NamedIndividual>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_Brasil_Acima_de_Tudo -->
```

```
<owl:NamedIndividual
```

```
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_Brasil_Acima_de_Tudo">
```

```
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Coligacao"/>
```

```
  <rdfs:comment>Jair Bolsonaro
```

```
PSL / PRTB</rdfs:comment>
</owl:NamedIndividual>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_Brasil_Soberano -->
```

```
<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_Brasil_Soberano">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Coligacao"/>
  <rdfs:comment>Ciro Gomes
PDT / AVANTE</rdfs:comment>
</owl:NamedIndividual>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_Essa_e_a_Solucao -->
```

```
<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_Essa_e_a_Solucao">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Coligacao"/>
  <rdfs:comment>Henrique Meirelles
MDB / PHS</rdfs:comment>
</owl:NamedIndividual>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_Mudanca_de_Verdade -->
```

```
<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_Mudanca_de_Verdade">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Coligacao"/>
  <rdfs:comment>Alvaro Dias
Podemos, PSC, PRP, PTC</rdfs:comment>
</owl:NamedIndividual>
```



```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_NOVO -->
```

```
<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_NOVO">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Coligacao"/>
  <rdfs:comment>Amoedo
NOVO</rdfs:comment>
</owl:NamedIndividual>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_O_Povo_Feliz_de_Novo -->
```

```
<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_O_Povo_Feliz_de_Novo">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Coligacao"/>
  <rdfs:comment>Fernando Haddad
PT / PC do B / PROS</rdfs:comment>
</owl:NamedIndividual>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_Para_Unir_o_Brasil -->
```

```
<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_Para_Unir_o_Brasil">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Coligacao"/>
  <rdfs:comment>Alckmin
PSDB / PTB / PP / PR / DEM / SOLIDARIEDADE / PPS / PRB / PSD</rdfs:comment>
</owl:NamedIndividual>
```

```
<!--
```

```
http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_Unidos_para_Transformar_o_Brasil -->
```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_Unidos_para_Transfo
rmar_o_Brasil">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Coligacao"/>
  <rdfs:comment>Marina Silva
REDE / PV</rdfs:comment>
</owl:NamedIndividual>

```

```

<!--
http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_Vamos_Sem_Medo_de_Mudar_o
_Brasil -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_Vamos_Sem_Medo_d
e_Mudar_o_Brasil">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Coligacao"/>
  <rdfs:comment>Guilherme Boulos
PSOL / PCB</rdfs:comment>
</owl:NamedIndividual>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_patri -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_patri">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Coligacao"/>
  <rdfs:comment>Daciolo
Patriota</rdfs:comment>
</owl:NamedIndividual>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#eleitor_Jefferson_Chaves -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#eleitor_Jefferson_Chaves">

```

```

<rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleitor"/>
<eleicoes:vota_em_secao
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#secao_500"/>
<eleicoes:eleitor_nome rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Jefferson de Oliveira
Chaves</eleicoes:eleitor_nome>
</owl:NamedIndividual>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#estado_SantaCatarina -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#estado_SantaCatarina">
<rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Estado"/>
</owl:NamedIndividual>

```

```

<!--

```

```

http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_CENTRO_COMUNITARIO_COSTA
_E_SILVA -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_CENTRO_COMUNITAR
IO_COSTA_E_SILVA">
<rdf:type
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#LocalDeVotacao"/>
<eleicoes:esta_localizado_em
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#cidade_Joinville"/>
<eleicoes:numero_de_eleitores
rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">3128</eleicoes:numero_de_eleitores>
</owl:NamedIndividual>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_COLEGIO_BOM_JESUS_
_SAGUACU -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_COLEGIO_BOM_JESUS
_-SAGUACU">
  <rdf:type
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#LocalDeVotacao"/>
  <eleicoes:esta_localizado_em
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#cidade_Joinville"/>
  <eleicoes:numero_de_eleitores
rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">5157</eleicoes:numero_de_eleitores>
</owl:NamedIndividual>

```

```

<!--
http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_COLEGIO_ESTADUAL_GERMAN
O -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_COLEGIO_ESTADUAL_
GERMANO">
  <rdf:type
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#LocalDeVotacao"/>
  <eleicoes:esta_localizado_em
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#cidade_Joinville"/>
  <eleicoes:numero_de_eleitores
rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">3069</eleicoes:numero_de_eleitores>
</owl:NamedIndividual>

```

```

<!--
http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_ESCOLA_MUNICIPAL_GOVERNA
DOR_PEDRO_IVO_CAMPOS -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_ESCOLA_MUNICIPAL_
GOVERNADOR_PEDRO_IVO_CAMPOS">
  <rdf:type
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#LocalDeVotacao"/>

```

```

<eleicoes:esta_localizado_em
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#cidade_Joinville"/>
<eleicoes:numero_de_eleitores
rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">2380</eleicoes:numero_de_eleitores>
</owl:NamedIndividual>

```

```

<!--
http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_ESCOLA_PLACIDO_OLIMPIO_DE_OLIVEIRA -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_ESCOLA_PLACIDO_OLIMPIO_DE_OLIVEIRA">
<rdf:type
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#LocalDeVotacao"/>
<eleicoes:esta_localizado_em
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#cidade_Joinville"/>
<eleicoes:numero_de_eleitores
rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">5523</eleicoes:numero_de_eleitores>
</owl:NamedIndividual>

```

```

<!--
http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_INSTITUTO_FEDERAL_DE_SANTA_CATARINA_-_IFSC -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_INSTITUTO_FEDERAL_DE_SANTA_CATARINA_-_IFSC">
<rdf:type
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#LocalDeVotacao"/>
<eleicoes:esta_localizado_em
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#cidade_Joinville"/>
<eleicoes:numero_de_eleitores
rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">2321</eleicoes:numero_de_eleitores>
</owl:NamedIndividual>

```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_UNOPAR_-
_POLO_JOINVILLE -->
```

```
<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_UNOPAR_-
_POLO_JOINVILLE">
  <rdf:type
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#LocalDeVotacao"/>
  <eleicoes:esta_localizado_em
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#cidade_Joinville"/>
  <eleicoes:numero_de_eleitores
rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">3605</eleicoes:numero_de_eleitores>
</owl:NamedIndividual>
```

```
<!--
http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_da_social_democracia_brasileira_P
SDB -->
```

```
<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_da_social_democracia_b
rasileira_PSDB">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Partido"/>
</owl:NamedIndividual>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_democratico_trabalhista_PDT
-->
```

```
<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_democratico_trabalhista
_PDT">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Partido"/>
```

```

<eleicoes:faz_parte_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_Brasil_Soberano"/
>
<eleicoes:partido_legenda
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">PDT</eleicoes:partido_legenda>
<eleicoes:partido_nome rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Partido Democrático
Trabalhista</eleicoes:partido_nome>
</owl:NamedIndividual>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_dos_trabalhadores_PT -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_dos_trabalhadores_PT">
<rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Partido"/>
<eleicoes:faz_parte_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_O_Povo_Feliz_de_
Novo"/>
<eleicoes:partido_nome rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Partido dos
Trabalhadores</eleicoes:partido_nome>
<eleicoes:partido_sigla
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">PT</eleicoes:partido_sigla>
</owl:NamedIndividual>

```

```

<!--

```

```

http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_movimento_democratico_brasileiro_
_MDB -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_movimento_democratico_brasileiro_MDB">
<rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Partido"/>
</owl:NamedIndividual>

```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_novo -->
```

```
<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_novo">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Partido"/>
  <eleicoes:partido_legenda
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">NOVO</eleicoes:partido_legenda>
  <eleicoes:partido_nome
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">NOVO</eleicoes:partido_nome>
</owl:NamedIndividual>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_patriota -->
```

```
<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_patriota">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Partido"/>
  <eleicoes:faz_parte_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_patri"/>
  <eleicoes:partido_nome
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Patriota</eleicoes:partido_nome>
</owl:NamedIndividual>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_podemos -->
```

```
<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_podemos">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Partido"/>
  <eleicoes:faz_parte_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_Mudanca_de_Verdade"/>
</owl:NamedIndividual>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_rede_sustentabilidade -->
```



```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_rede_sustentabilidade">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Partido"/>
</owl:NamedIndividual>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_social_liberal_PSL -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_social_liberal_PSL">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Partido"/>
  <eleicoes:faz_parte_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#coligacao_Brasil_Acima_de_
Tudo"/>
  <eleicoes:partido_legenda    rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Partido    Social
Liberal</eleicoes:partido_legenda>
  <eleicoes:partido_nome      rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Partido    Social
Liberal</eleicoes:partido_nome>
  <eleicoes:partido_sigla
rdf:datatype="http://www.w3.org/2001/XMLSchema#string">PSL</eleicoes:partido_sigla>
</owl:NamedIndividual>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_socialismo_e_liberdade_PSOL
-->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#partido_socialismo_e_liberdade_
PSOL">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Partido"/>
</owl:NamedIndividual>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#primeiro_turno -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#primeiro_turno">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Turno"/>
  <eleicoes:eh_turno_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleicoes2018"/>
</owl:NamedIndividual>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#secao_114 -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#secao_114">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Secao"/>
  <eleicoes:esta_localizado_em
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_INSTITUTO_FEDERA
L_DE_SANTA_CATARINA_-_IFSC"/>
  <eleicoes:numero_de_eleitores
rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">330</eleicoes:numero_de_eleitores>
</owl:NamedIndividual>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#secao_115 -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#secao_115">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Secao"/>
  <eleicoes:esta_localizado_em
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_INSTITUTO_FEDERA
L_DE_SANTA_CATARINA_-_IFSC"/>
  <eleicoes:numero_de_eleitores
rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">333</eleicoes:numero_de_eleitores>
</owl:NamedIndividual>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#secao_116 -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#secao_116">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Secao"/>
  <eleicoes:esta_localizado_em
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_INSTITUTO_FEDERA
L_DE_SANTA_CATARINA_-_IFSC"/>
  <eleicoes:numero_de_eleitores
rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">332</eleicoes:numero_de_eleitores>
</owl:NamedIndividual>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#secao_117 -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#secao_117">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Secao"/>
  <eleicoes:esta_localizado_em
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_INSTITUTO_FEDERA
L_DE_SANTA_CATARINA_-_IFSC"/>
  <eleicoes:numero_de_eleitores
rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">328</eleicoes:numero_de_eleitores>
</owl:NamedIndividual>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#secao_123 -->

```

```

<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#secao_123">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Secao"/>
  <eleicoes:esta_localizado_em
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_INSTITUTO_FEDERA
L_DE_SANTA_CATARINA_-_IFSC"/>
  <eleicoes:numero_de_eleitores
rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">333</eleicoes:numero_de_eleitores>
</owl:NamedIndividual>

```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#secao_500 -->
```

```
<owl:NamedIndividual
```

```
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#secao_500">
```

```
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Secao"/>
```

```
  <eleicoes:esta_localizado_em
```

```
  rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_INSTITUTO_FEDERAL_DE_SANTA_CATARINA_-_IFSC"/>
```

```
  <eleicoes:numero_de_eleitores
```

```
  rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">331</eleicoes:numero_de_eleitores>
```

```
</owl:NamedIndividual>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#secao_603 -->
```

```
<owl:NamedIndividual
```

```
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#secao_603">
```

```
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Secao"/>
```

```
  <eleicoes:esta_localizado_em
```

```
  rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_INSTITUTO_FEDERAL_DE_SANTA_CATARINA_-_IFSC"/>
```

```
  <eleicoes:numero_de_eleitores
```

```
  rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">334</eleicoes:numero_de_eleitores>
```

```
</owl:NamedIndividual>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#segundo_turno -->
```

```
<owl:NamedIndividual
```

```
  rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#segundo_turno">
```

```
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Turno"/>
```

```
  <eleicoes:eh_turno_de
```

```
  rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleicoes2018"/>
```

```
</owl:NamedIndividual>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#019 -->
```

```
<owl:NamedIndividual rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#019">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#ZonaEleitoral"/>
  <eleicoes:eh_zona_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleicoes2018"/>
  <eleicoes:possui
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_CENTRO_COMUNIT
ARIO_COSTA_E_SILVA"/>
  <eleicoes:possui
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_COLEGIO_BOM_JES
US_-_SAGUACU"/>
  <eleicoes:possui
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_COLEGIO_ESTADUA
L_GERMANO"/>
  <eleicoes:possui
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_ESCOLA_MUNICIPA
L_GOVERNADOR_PEDRO_IVO_CAMPOS"/>
  <eleicoes:possui
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_ESCOLA_PLACIDO_
OLIMPIO_DE_OLIVEIRA"/>
  <eleicoes:possui
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_INSTITUTO_FEDERA
L_DE_SANTA_CATARINA_-_IFSC"/>
  <eleicoes:possui
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#local_UNOPAR_-
_POLO_JOINVILLE"/>
</owl:NamedIndividual>
```

```
<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#076 -->
```

```
<owl:NamedIndividual rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#076">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#ZonaEleitoral"/>
  <eleicoes:eh_zona_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleicoes2018"/>
```

```

<rdfs:comment>http://www.tse.jus.br/eleitor/servicos/cartorios-e-zonas-eleitorais/pesquisa-a-zonas-
eleitorais</rdfs:comment>
</owl:NamedIndividual>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#095 -->

```

```

<owl:NamedIndividual rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#095">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#ZonaEleitoral"/>
  <eleicoes:eh_zona_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleicoes2018"/>
</owl:NamedIndividual>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#096 -->

```

```

<owl:NamedIndividual rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#096">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#ZonaEleitoral"/>
  <eleicoes:eh_zona_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleicoes2018"/>
</owl:NamedIndividual>

```

```

<!-- http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#105 -->

```

```

<owl:NamedIndividual rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#105">
  <rdf:type rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#ZonaEleitoral"/>
  <eleicoes:eh_zona_de
rdf:resource="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Eleicoes2018"/>
</owl:NamedIndividual>

```

```

<!--

```

```

////////////////////////////////////

```

```

//

```

```

// General axioms
//
////////////////////////////////////
-->

<rdf:Description>
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#AllDisjointClasses"/>
  <owl:members rdf:parseType="Collection">
    <rdf:Description
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#DeputadoEstadual"/>
      <rdf:Description
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#DeputadoFederal"/>
        <rdf:Description
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Governador"/>
          <rdf:Description rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Presidente"/>
            <rdf:Description rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#Senador"/>
              </owl:members>
            </rdf:Description>
          <rdf:Description>
            <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#AllDisjointClasses"/>
            <owl:members rdf:parseType="Collection">
              <rdf:Description
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#VotosAnulados"/>
                <rdf:Description
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#VotosBrancos"/>
                  <rdf:Description
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#VotosNulos"/>
                    <rdf:Description
rdf:about="http://www.semanticweb.org/jefferson/ontologies/2018/10/eleicoes#VotosVálidos"/>
                      </owl:members>
                    </rdf:Description>
                  </rdf:RDF>
                <!-- Generated by the OWL API (version 4.5.9.2019-02-01T07:24:44Z) https://github.com/owlcs/owlapi -->

```

APÊNDICE B – Relatório gerado pela aplicação após a execução do workflow para o estado do Acre

```

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/terms/"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:ldwpo="http://ldwpo.aksw.org/terms/1.0/"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#">
  <rdf:Property rdf:about="http://ldwpo.aksw.org/terms/1.0/creator"/>
  <rdf:Property rdf:about="http://ldwpo.aksw.org/terms/1.0/goal"/>
  <rdf:Property rdf:about="http://ldwpo.aksw.org/terms/1.0/license"/>
  <rdf:Property rdf:about="http://ldwpo.aksw.org/terms/1.0/ldwStep"/>
  <rdf:Description rdf:about="http://ufsc.br/eleicoes/votacao_acre/2018/v1/">
    <ldwpo:ldwStep>
      <ldwpo:Step>
        <rdfs:comment>Tabela ordenada por NOME MUNICÍPIO.</rdfs:comment>
        <dc:title>datasetOrderByStep</dc:title>
      </ldwpo:Step>
    </ldwpo:ldwStep>
    <ldwpo:goal>Transformar e conectar dados das eleições de 2018</ldwpo:goal>
    <ldwpo:ldwStep>
      <ldwpo:Step>
        <rdfs:comment>A relação semântica entre os dados e vocabulários foram estabelecidas.</rdfs:comment>
        <dc:title>semanticMappingStep</dc:title>
      </ldwpo:Step>
    </ldwpo:ldwStep>
    <dc:spatial>Estado do Acre</dc:spatial>
    <dc:license>Creative Commons License - CC</dc:license>
  </rdf:Description>
</rdf:RDF>

```



```

<dc:source>https://cdn.tse.jus.br/estatistica/sead/eleicoes/eleicoes2018/buweb/BWEB_1t_AC
_101020181938.zip</dc:source>
  <dc:created>15/08/2021</dc:created>
  <dc:subject>Eleições, Eleições Federais, Eleições 2018</dc:subject>
  <dc:title>WORKFLOW - Votação Acre</dc:title>
  <dc:title>PROJECT - Eleições 2018</dc:title>
  <ldwpo:ldwStep>
    <ldwpo:Step>
      <rdfs:comment>Gerenciamento de colunas realizado.</rdfs:comment>
      <dc:title>manageColumnsStep</dc:title>
    </ldwpo:Step>
  </ldwpo:ldwStep>
  <ldwpo:creator>Jefferson Chaves</ldwpo:creator>
  <dc:description>Número de votos por candidato das eleições de
2018</dc:description>
  <dc:creator>Jefferson Chaves</dc:creator>
  <ldwpo:ldwStep>
    <ldwpo:Step>
      <rdfs:comment>Texto do campo DESCRIÇÃO ELEIÇÃO alterado para
maiúsculas.</rdfs:comment>
      <dc:title>toUpperCaseStep</dc:title>
    </ldwpo:Step>
  </ldwpo:ldwStep>
  <ldwpo:ldwStep>
    <ldwpo:Step>
      <rdfs:comment>A relação semântica entre os dados e vocabulários foram
estabelecidas.</rdfs:comment>
      <dc:title>semanticMappingStep</dc:title>
    </ldwpo:Step>
  </ldwpo:ldwStep>
  <ldwpo:ldwStep>

```

```

<ldwpo:Step>
  <rdfs:comment>Os recursos da coluna NOME MUNICÍPIO foram conectados
com a DBPEDIA.</rdfs:comment>
  <dc:title>DBPediaReconciliationStep</dc:title>
</ldwpo:Step>
</ldwpo:ldwStep>
<ldwpo:ldwStep>
  <ldwpo:Step>
    <rdfs:comment>NOMINAL foi selecionado como filtro para a coluna
CANDIDATO.</rdfs:comment>
    <dc:title>filterStep</dc:title>
  </ldwpo:Step>
</ldwpo:ldwStep>
<ldwpo:license>Creative Commons License - CC</ldwpo:license>
<dc:language>pt-br</dc:language>
<ldwpo:ldwStep>
  <ldwpo:Step>
    <rdfs:comment>Gerenciamento de colunas realizado.</rdfs:comment>
    <dc:title>manageColumnsStep</dc:title>
  </ldwpo:Step>
</ldwpo:ldwStep>
<ldwpo:ldwStep>
  <ldwpo:Step>
    <rdfs:comment>Gerenciamento de colunas realizado.</rdfs:comment>
    <dc:title>manageColumnsStep</dc:title>
  </ldwpo:Step>
</ldwpo:ldwStep>
<dc:temporal>2018</dc:temporal>
</rdf:Description>
</rdf:RDF>

```