

# Em direção a uma Ontologia das Eleições Brasileiras

Aíquis R. Gomes<sup>1</sup>, Kele T. Belloze<sup>1</sup>

<sup>1</sup> Programa de Pós-graduação em Ciência da Computação  
Centro Federal de Educação Tecnológica Celso Suckow da Fonseca (CEFET/RJ)  
Rio de Janeiro, RJ – Brasil

aiquis.gomes@eic.cefet-rj.br, kele.belloze@cefet-rj.br

**Abstract.** *This paper presents the process of creating an electoral domain ontology, using open data published by the Superior Electoral Court. In order to conduct this process, we analyzed and adopted activities of different methodologies for the construction of ontologies. The construction of this ontology is part of a work that has the purpose of developing a methodology of support in the process of publication of Open Government Data as Linked Open Data.*

**Resumo.** *Este trabalho apresenta o processo de criação de uma ontologia do domínio eleitoral utilizando dados abertos publicados pelo Tribunal Superior Eleitoral. Para a condução desse trabalho, foram analisadas e adotadas atividades de diferentes metodologias para construção de ontologias. A construção dessa ontologia é parte de um trabalho que tem como propósito o desenvolvimento de uma metodologia de apoio ao processo de publicação de Dados Abertos Governamentais como Dados Abertos Ligados.*

## 1. Introdução

Governos ao redor do mundo têm cada vez mais buscado por ferramentas que permitam maior transparência, que ajudem a combater a corrupção e que contribuam para um governo participativo. Para endereçar esse problema, muitos países estão desenvolvendo iniciativas de Dados Abertos Governamentais (em inglês, OGD - *Open Government Data*), que podem ser definidos como dados produzidos e mantidos por agências governamentais que respeitam os requisitos de Dados Abertos [Kučera et al. 2013]. Essas iniciativas podem ser potencializadas por meio da utilização de Dados Ligados, um conjunto de padrões que buscam ligar dados de modo que sejam facilmente consumidos por máquinas e seu significado seja explicitamente definido [Bizer et al. 2011], criando assim o que é conhecido como Dados Abertos Ligados (em inglês, LOD - *Linked Open Data*).

Apesar de trazer benefícios como controle e responsabilidade com os dados publicados, permitindo um reuso flexível, poucos governos têm utilizado esse padrão por falta de familiaridade, de um guia para implementação e até baixo interesse da comunidade de desenvolvedores [Sheridan and Tennison 2010]. Essas questões expressam baixo grau de maturidade na publicação de dados abertos, o qual pode ser medido por meio do esquema de implementação das 5 estrelas, um modelo de qualidade de 5 etapas proposto por Tim Berners-Lee [Hausenblas and Kim 2015], em que o mais alto nível significa o uso pleno de LOD.

Nesta perspectiva, faz-se importante prover um *framework* que apoie a evolução no modelo de qualidade, detalhando etapas e processos necessários para se chegar no nível

mais alto, de forma a servir como guia para órgãos governamentais. Algumas abordagens para a criação de LOD passam pela elaboração de ontologias [Victorino et al. 2018, Pattuelli et al. 2015, Sérgio et al. 2015]. Nesse sentido, este trabalho tem como objetivo apresentar o processo de construção de uma ontologia do domínio eleitoral, que, dentro do processo maior de construção do *framework*, servirá como apoio à publicação de LOD. Conjuntos de dados abertos do Tribunal Superior Eleitoral (TSE) do Brasil foram utilizados para modelar uma ontologia das eleições brasileiras. O desenvolvimento dessa ontologia apoia-se em metodologias estabelecidas para criação de ontologias com foco em processos que possam ser aplicados ao objetivo de publicação de dados ligados.

Além dessa introdução, esse trabalho está organizado como segue. A seção 2 descreve e analisa as metodologias para construção da ontologia. A seção 3 apresenta o levantamento de requisitos da ontologia. A seção 4 detalha o processo de reutilização de recursos. A seção 5 descreve a conceitualização da ontologia. Por fim, a seção 6 apresenta as considerações finais sobre o trabalho e próximos passos.

## 2. Metodologia

O estudo de ontologias é bastante difundido nas áreas de representação do conhecimento e inteligência artificial. Existem várias metodologias para seu desenvolvimento, como a METHONTOLOGY [Fernández-López et al. 1997], Ontology Development 101 [Noy and McGuinness 2001], SABiO [Falbo 2014] e NeOn [Suarez-Figueroa et al. 2012], que são bastante conhecidas entre desenvolvedores de ontologias, e possuem, inclusive, muitos aspectos semelhantes entre si.

A metodologia NeOn é baseada em cenários, a qual apoia diversas situações do desenvolvimento de ontologias, além de ter um foco grande em colaboração e no reuso de ontologias e recursos não-ontológicos [Suarez-Figueroa et al. 2012]. As características em questão motivaram a escolha dessa metodologia como guia primário para o desenvolvimento da ontologia deste trabalho. Contudo, a etapa de conceitualização da ontologia é pouco detalhada pela NeOn. Por este motivo, foi avaliado como a METHONTOLOGY, SABiO e a Ontology Development 101 tratam a etapa referente. Apesar de serem três abordagens diferentes, todas têm como objetivo identificar e explicitar conceitos, atributos e relações. A principal diferença entre elas está nos artefatos que são produzidos e o grau de formalismo derivado desses artefatos.

Após comparação entre as metodologias, optou-se por utilizar a Ontology Development 101 para a conceitualização. Essa escolha se deu pela mesma ter seu foco voltado para a parte prática da implementação, além de ser direcionada para um público com menos experiência em desenvolvimento de ontologias, devido ao número menor de formalismos e artefatos a serem produzidos. Tais características adicionadas à questão de a ontologia construída neste trabalho ser uma ontologia reduzida, com escopo e domínio bem definidos, apoiaram a escolha dessa metodologia entre as analisadas.

## 3. Levantamento de Requisitos

No desenvolvimento de ontologias, uma das principais etapas é a de definição dos requisitos, de modo a garantir que o que for desenvolvido esteja de acordo com as definições prévias, de maneira semelhante à Engenharia de Software. Nesta seção são descritas as

etapas realizadas para o levantamento de requisitos da ontologia das eleições brasileiras, utilizando como base as recomendações da metodologia NeOn. A execução desse processo tem como produto o Documento de Especificação de Requisitos da Ontologia [Suarez-Figueroa et al. 2012].

### 3.1. Definições

Um passo recomendado antes de serem levantados os requisitos funcionais e não-funcionais da ontologia é definir o cenário motivador da construção da ontologia. Além desse, é sugerido também que sejam descritos o escopo, seus usuários e uso pretendido, e a linguagem de desenvolvimento [Grüninger and Fox 1995]. A Tabela 1 apresenta as definições para cada um desses pontos.

**Tabela 1. Definições de cenário, escopo, usuários, uso e linguagem da Ontologia das Eleições Brasileiras**

<b>Característica</b>	<b>Definição</b>
Cenário motivador	Prover um modelo semântico que possa ser usado como base para interligação de diferentes conjuntos de dados disponibilizados pelo TSE
Escopo	Compreende o domínio das eleições no Brasil com definições de relações e regras
Usuários	Pesquisadores e usuários com interesse no aspecto semântico do domínio eleitoral brasileiro
Uso pretendido	Iniciativas para a construção de dados abertos ligados
Linguagem	Web Ontology Language (OWL)

### 3.2. Identificação de requisitos

Nessa etapa busca-se identificar tanto os requisitos funcionais (RF) quanto os não-funcionais (RNF) da ontologia que está sendo desenvolvida. Para a ontologia das eleições brasileiras foi identificado um RNF: a ontologia deve se referir apenas à eleições ocorridas após a redemocratização no país.

Para a definição dos RF foram utilizadas as Questões de Competência (QCs). Elas são parte de um procedimento para desenvolvimento e avaliação de uma ontologia e são utilizadas como requisitos no formato de perguntas que uma ontologia deve responder [Grüninger and Fox 1995]. As QCs foram elaboradas utilizando uma abordagem *middle out*, na qual as questões que são consideradas importantes são escritas independente de sua complexidade e são compostas e decompostas para que se obtenha questões simples e abstratas [Suarez-Figueroa et al. 2012].

Após a finalização da lista inicial de questões, foi realizada uma categorização por temas. Essa categorização serve de apoio para o processo de priorização das QCs. Ela também pode auxiliar em cenários nos quais se deseja dividir ou modularizar uma ontologia em outras menores para que se obtenha uma redução de escopo e complexidade [Suarez-Figueroa et al. 2012]. As QCs desenvolvidas foram classificadas em quatro temas: eleição, candidato, eleitor e partido. Ao todo foram desenvolvidas 26 questões. A lista completa está disponível de maneira aberta em <https://github.com/swrg-cefetrj/ontobras-2019>.

### **3.3. Validação e priorização dos requisitos**

O processo de validação dos requisitos foi realizado em paralelo à elaboração das QCs. Essa tarefa busca refinar os requisitos por meio da utilização de diferentes critérios para análise, os quais tentam identificar problemas como conflitos e contradições entre eles e requisitos faltantes [Suarez-Figueroa et al. 2012]. Os critérios utilizados para validação foram: corretude, consistência, não-ambiguidade e realismo [Suarez-Figueroa et al. 2012]. Na priorização não foi encontrada necessidade de descarte de questões, pois foi entendido que a ontologia deve ser capaz de responder a todas.

### **3.4. Extração de terminologia e frequências**

Essa etapa tem como objetivo a obtenção de um pré-glossário de termos a partir da lista de QCs e suas repostas, no qual sejam identificados elementos no universo de discurso que vão ser transformados em elementos da ontologia em construção. Esses elementos são conceitos (identificados pelos substantivos), atributos (identificados pelos adjetivos), relacionamentos (identificados pelos verbos) e instâncias de objetos. Obter a frequência dos termos ajuda na busca por recursos que possam ser reaproveitados no desenvolvimento da ontologia [Suarez-Figueroa et al. 2012].

Nesse primeiro momento as instâncias não foram classificadas pois não apresentam relevância para o desenvolvimento conceitual da ontologia. O processo de extração foi feito separadamente para as questões e suas respostas pois a classificação dos termos extraídos é diferente entre eles [Suarez-Figueroa et al. 2012]. Foram identificados 91 termos, somando os termos obtidos a partir das perguntas e das respostas. Esses foram ranqueados (maior para menor frequência) para formar o pré-glossário.

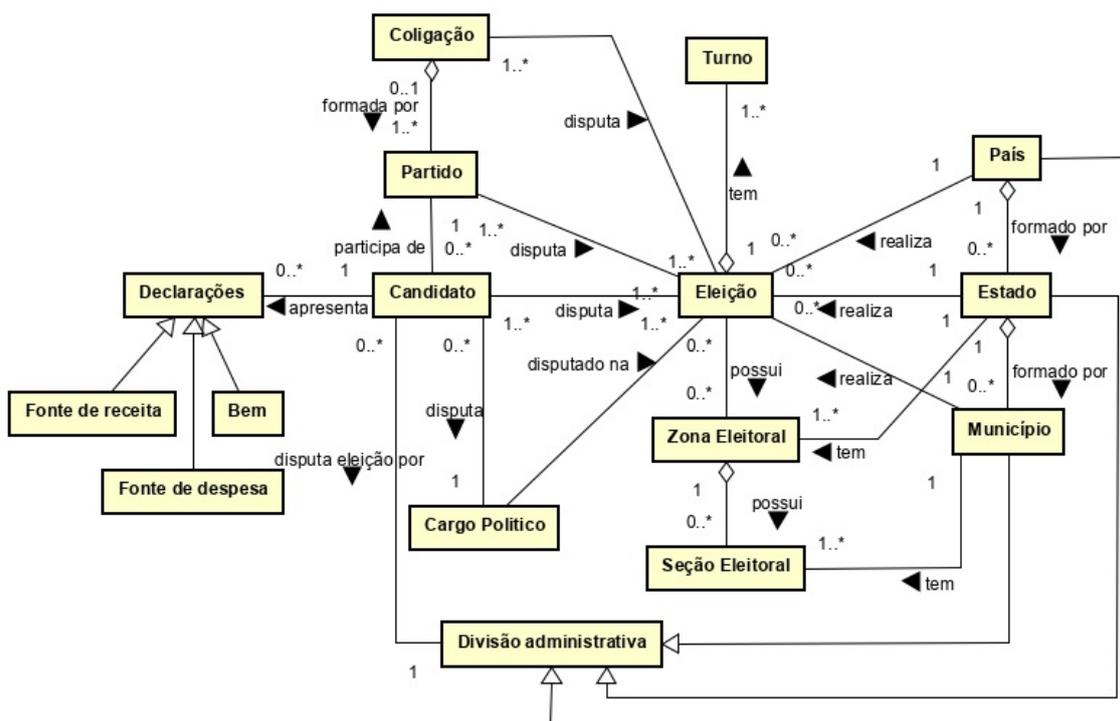
## **4. Reutilização de recursos ontológicos e não-ontológicos**

Uma busca e avaliação de recursos ontológicos (ROs) e não-ontológicos (RNO) destacou o Tesouro da Justiça Eleitoral [Tribunal Superior Eleitoral 2010] e o Glossário Eleitoral [Tribunal Superior Eleitoral 2019]. Esses recursos podem ser utilizados em trabalhos futuros para criar uma ontologia mais completa do domínio ou até mesmo uma rede de ontologias que representem diferentes aspectos do domínio das eleições brasileiras.

## **5. Conceitualização da Ontologia**

Para o processo de conceitualização, foi utilizada a metodologia *Ontology Development 101*, como descrito na seção 2. A primeira etapa proposta pela metodologia é a listagem de termos importantes da ontologia. Para isso foi utilizado o pré-glossário de termos produzido. Ele apresenta informações adicionais, como a classificação de cada termo (conceito, atributo e relacionamento). Devido a essa classificação, as etapas seguintes foram facilitadas, pois englobavam a identificação das classes, suas hierarquias, e propriedades. A metodologia foi seguida como uma maneira de reavaliar o que havia sido mapeado no pré-glossário, o que fez com que alguns termos tivessem sua classificação alterada de maneira a refletirem melhor o conhecimento do domínio. A Figura 1 apresenta as classes e relacionamentos propostos para a ontologia das eleições brasileiras.

Como classe principal da ontologia há a Eleição, que representa um processo eleitoral realizado por um País, Estado ou Município. Um País é formado por Estados que por sua vez é formado por Municípios. Um País e um



**Figura 1. Classes e relacionamentos da Ontologia das Eleições Brasileiras**

Estado possuem várias Zonas eleitorais e os Municípios possuem Seções eleitorais que são parte das zonas. Uma Eleição no Brasil pode ter um ou dois Turnos. Uma Eleição é disputada por vários Candidatos que disputam Cargos políticos por um Estado ou Município e que precisam ser associados a um Partido, o qual geralmente (mas nem sempre) forma Coligação para disputar a Eleição. Para fazer parte da disputa, um candidato precisa declarar os Bens que possui. Além disso, ele também precisa declarar as Fontes de receita e Fontes de despesa da sua campanha, ou seja, qual a origem e destino do dinheiro gasto.

## 6. Considerações finais

Nesse trabalho foi apresentado o processo de criação de uma ontologia para apoiar a publicação de Dados Abertos Governamentais como Dados Abertos Ligados. Para esse trabalho, o escopo limitou-se ao domínio eleitoral brasileiro com base nos conjuntos de dados disponibilizados pelo TSE, dando origem a uma ontologia das eleições brasileiras.

Para o desenvolvimento da ontologia, a metodologia NeOn foi utilizada primariamente. Por se apresentar bastante focada na reutilização de recursos, essa metodologia tem uma aderência grande com iniciativas de Dados Ligados. O desenvolvimento da ontologia seguiu muitas etapas indicadas pela metodologia, iniciando pelo levantamento de requisitos da ontologia, atividade conduzida por meio das Questões de Competência, e passando pela identificação e análise de recursos passíveis de reutilização. A modelagem conceitual apoiou-se na metodologia *Ontology Development 101*.

Como principal contribuição deste trabalho destaca-se a construção, até a etapa de conceitualização, de uma ontologia sobre as eleições brasileiras. Os próximos passos

do trabalho englobam: i) implementação e validação da ontologia utilizando as QCs; ii) criação de uma instância de uma eleição para a realização do processo de validação e; iii) publicação em RDF da ontologia. Como trabalho futuro destacamos a criação de uma metodologia e posterior *framework* para apoiar governos e desenvolvedores dispostos a trabalhar em iniciativas de LOD, tornando mais facilitado esse processo.

## Referências

- Bizer, C., Heath, T., and Berners-Lee, T. (2011). Linked data: The story so far. In *Semantic services, interoperability and web applications: emerging concepts*, pages 205–227. IGI Global.
- Falbo, R. D. A. (2014). SABiO : Systematic Approach for Building Ontologies. *1st Joint Workshop Onto.Com/ODISE on Ontologies in Conceptual Modeling and Information Systems Engineering*.
- Fernández-López, M., Gómez-Pérez, A., and Juristo, N. (1997). Methontology: from ontological art towards ontological engineering.
- Grüniger, M. and Fox, M. S. (1995). Methodology for the Design and Evaluation of Ontologies.
- Hausenblas, M. and Kim, J. (2015). 5-star open data.
- Kučera, J., Chlapek, D., and Nečaský, M. (2013). Open government data catalogs: Current approaches and quality perspective. In Kő, A., Leitner, C., Leitold, H., and Prosser, A., editors, *Technology-Enabled Innovation for Democracy, Government and Governance*, pages 152–166, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Noy, N. F. and McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology. pages 1–25.
- Pattueli, M. C., Provo, A., and Thorsen, H. (2015). Ontology Building for Linked Open Data: A Pragmatic Perspective. *Journal of Library Metadata*, 15(3-4):265–294.
- Sérgio, L., Araújo, O., Santos, M. T., and Silva, D. A. (2015). The Brazilian Federal Budget Ontology – A Semantic Web Case of Public Open Data. *The 7th International Conference on Management of Digital EcoSystems*, (October):85–89.
- Sheridan, J. and Tennison, J. (2010). Linking UK Government Data. *Ldow 2010*, pages 1–4.
- Suarez-Figueroa, M. C., Gomez-Perez, A., Motta, E., and Gangemi, A. (2012). Introduction: Ontology engineering in a networked world. In *Ontology Engineering in a Networked World*, pages 1–6. Springer.
- Tribunal Superior Eleitoral (2010). Tesouro da Justiça Eleitoral. Acessado em 20/02/2019.
- Tribunal Superior Eleitoral (2019). Glossário Eleitoral. Acessado em 20/02/2019.
- Victorino, M., de Holanda, M. T., Ishikawa, E., Oliveira, E. C., and Chhetri, S. (2018). Transforming Open Data to Linked Open Data Using Ontologies for Information Organization in Big Data Environments of the Brazilian Government: the Brazilian Database Government Open Linked Data – DBgoldbr. *Knowledge Organization*, 45(6):443–466.