

OGDPub: Ontologia de Domínio para Publicação de Dados Abertos por Municípios Brasileiros

Larissa Mariany Freiberger Pereira¹, Guilherme Bertoni Machado¹, José Leomar Todesco¹, Fernando Ostuni Gauthier¹

¹Programa de Pós-Graduação em Engenharia e Gestão de Conhecimento (PPGEGC)
Universidade Federal de Santa Catarina (UFSC) – Florianópolis – SC – Brasil

{freiberger.lm, gbertonimachado}@gmail.com, jose.todesco@ufsc.br,
gauthier@egc.ufsc.br

***Abstract.** This paper reports the development of a domain ontology for open data publishing by Brazilian municipalities. The developed ontology standardizes the concepts involved in the world of open government data as well as explicit and formalizes the open data publishing process, specifically in the context of Brazilian municipalities.*

***Resumo.** Este artigo relata o desenvolvimento de uma ontologia de domínio para publicação de dados abertos por municípios brasileiros. A ontologia desenvolvida padroniza os conceitos envolvidos no mundo de dados abertos governamentais, bem como explicita e formaliza o processo de publicação destes dados, especificamente no contexto de municípios brasileiros.*

1. Introdução

Como consequência do movimento *Openness*¹, os dados gerados por governos vêm sendo disponibilizados ao público. Este movimento, denominado Dados Abertos Governamentais, diz respeito aos dados produzidos por entes governamentais que “podem ser livremente acessados, utilizados, modificados e compartilhados por qualquer pessoa, para qualquer fim” [OPEN DEFINITION, 2014].

Atualmente, os governos são motivados a abrirem seus dados por três razões principais, sendo elas: promoção da transparência, fomento de uma governança participativa e geração de valor social e comercial. Além disso, os dados produzidos por governos são de grande valor e abrangem muitas áreas de interesse social. Quando os governos optam por tornar seus dados disponíveis à sociedade, acabam também motivando ações inovadoras na criação de novos serviços por pessoas ou agentes que compõem a sociedade [ATTARD et al., 2016].

Apesar disso, no Brasil é possível observar que, em nível municipal, poucos municípios conseguem publicar *datasets* processáveis por máquina. O que efetivamente ocorre é que, em geral, os municípios publicam apenas relatórios de gestão que, além de não agregarem valor aos dados publicados, normalmente estão disponibilizados em uma linguagem de difícil compreensão pelo cidadão. Isso faz com que os dados

¹ Movimento “baseado na ideia de que o conhecimento pode ser disseminado e compartilhado livremente na Internet para o benefício da sociedade” [ZANCANARO, 2015. p. 43]

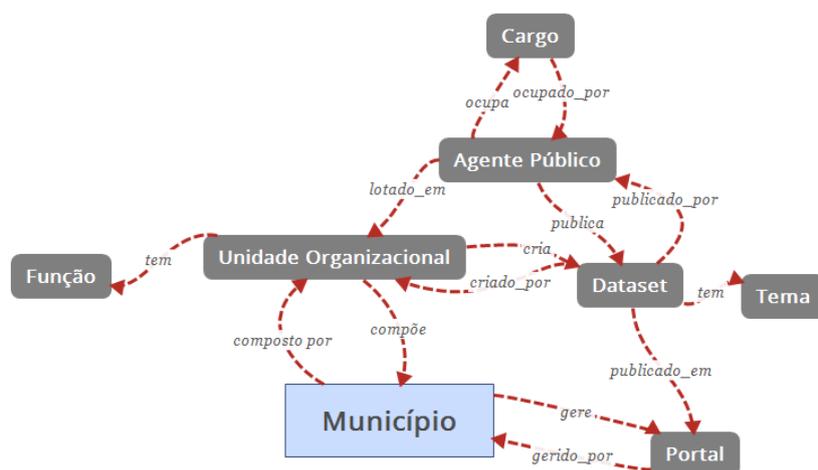
governamentais gerados por municípios acabem sendo subutilizados, isto é, não se aproveita todo o potencial que o dado traz consigo para gerar novas informações que por vezes não estão explícitas, mas implícitas entre os dados.

A questão, neste contexto, é: como ampliar e promover o uso dos dados abertos governamentais municipais no Brasil? Para auxiliar na resolução deste problema, é proposta, neste artigo, uma ontologia de domínio que possibilite que os *datasets* sejam descritos através de metadados, tornando mais simples encontrá-los na *Web* e, além disso, é proposto ainda que a ontologia seja capaz de classificar os *datasets* e apresentá-los a partir de uma “linguagem cidadã”, tornando-os compreensíveis à sociedade em geral.

2. OGD Pub

A ontologia, denominada OGD Pub ou “*Open Government Data Publish*”, foi desenvolvida em *Web Ontology Language*² (OWL), padrão definido pela *World Wide Web Consortium*³ (W3C). Em um primeiro momento foi utilizada a ferramenta *XMind*⁴ para construir o mapa mental do processo de publicação de dados abertos governamentais por municípios tal qual deveria ser. Na construção da ontologia, propriamente dita, foi utilizada a ferramenta *NeOn Toolkit*⁵, desenvolvida como parte do projeto *NeOn Project*⁶.

É importante ressaltar que a OGD Pub ainda está em desenvolvimento e a versão que será descrita neste artigo trata-se de um protótipo ou de uma primeira versão da mesma, que será ainda desenvolvida de forma minuciosa, tendo novos elementos agregados. Neste sentido, a OGD Pub começou a ser imaginada e desenhada em um mapa mental, apresentado na Figura 1.



² <https://www.w3.org/OWL/>

³ <https://www.w3.org/>

⁴ <http://www.xmind.net/>

⁵ <http://neon-toolkit.org/>

⁶ <http://www.neon-project.org/>

Figura 1. Mapa mental da OGD Pub

2.1 Concepção da OGD Pub

A partir do problema descrito na Introdução questionava-se de que maneira este poderia ser solucionado e alguns possíveis cenários foram sendo construídos: se os *datasets* publicados pelos municípios fossem descritos por um conjunto de metadados, estes poderiam ser mais fáceis de serem encontrados na *Web*; se esses *datasets* fossem classificados à partir de uma linguagem facilitada para o cidadão, os dados se tornariam mais compreensíveis; se fosse possível descrever todo o processo de publicação de dados por municípios, este processo poderia ser padronizado, sendo também simplificado.

2.2 Metodologia

Para que a OGD Pub pudesse ser implementada, era necessário definir uma ontologia para orientar este processo. Elegeu-se, então, a metodologia *NeOn*. Esta metodologia, baseada em cenários, permite a construção de ontologias baseada no conceito de redes de ontologia: coleção de diversas ontologias que se conectam através de diferentes relações como mapeamento, modularização e versionamento [GÓMEZ-PÉREZ; SUÁREZ-FIGUEROA, 2009].

Os nove cenários descritos na metodologia *NeOn* são: especificação e implementação; reuso e reengenharia de recursos não-ontológicos; reuso de recursos ontológicos; reuso e reengenharia de recursos ontológicos; reuso e fusão de recursos ontológicos; reuso, fusão e reengenharia de recursos ontológicos; reuso de padrões de *design* de ontologias (ODPs); reestruturação de recursos ontológicos e; localização de recursos ontológicos.

A OGD Pub insere-se no cenário 3 (reuso de recursos ontológicos), uma vez que, para sua construção, foram identificados recursos ontológicos já desenvolvidos que poderiam ser reusados (ontologias completas, módulos de ontologias ou relações sujeito-predicado-objeto de uma determinada ontologia) [GÓMEZ-PÉREZ; SUÁREZ-FIGUEROA, 2009]. Após identificadas as ontologias que poderiam ser reusadas, foi possível perceber que não seriam necessárias atividades de reengenharia e/ou fusão, uma vez que os módulos utilizados das ontologias eram aderentes à proposta da OGD Pub, sem que fossem necessárias adaptações.

A OGD Pub ainda foi construída a partir do modelo de ciclo de vida “*Waterfall*” ou “*Cascata*”. Neste modelo de construção de ontologias, uma etapa precisa estar completa para que se inicie a etapa seguinte. O modelo cascata foi escolhido por se tratar de um projeto de curta duração, sem o envolvimento de muitas pessoas em sua execução. A metodologia *NeOn* sugere, no caso da OGD Pub, que sejam seguidas 5 etapas na construção da ontologia, sendo elas: Fase Inicial; Fase de Reuso; Fase de *Design*; Fase de Implementação e; Fase de Manutenção.

2.3 Construção da OGD Pub

Fase Inicial

Na fase inicial foi feito o levantamento de requisitos que a ontologia deveria atender. Parte desses requisitos já haviam sido imaginados no momento da concepção da

OGDPub, todavia nesta fase foi possível formalizá-los, bem como organizá-los em requisitos funcionais e não-funcionais.

Destaca-se, nesta etapa, como foram elencados os temas para compor a OGDPub. O objetivo era elencar temas que abrangessem todos os tipos de dados produzidos por municípios brasileiros e que fossem de fácil entendimento para o cidadão. Para isso, foram feitas buscas em portais de dados abertos de 19 cidades em todo o mundo que se destacam no universo de publicação de dados abertos e nas principais capitais de estados brasileiros que atualmente publicam seus dados governamentais para verificar de que maneira seus dados eram classificados. As cidades cujos portais foram verificados foram: Madri, Washington, New York City, Toronto, Londres, Roma, Berlim, Paris, Lisboa, São Paulo, Rio de Janeiro, Porto Alegre, Curitiba, João Pessoa, Recife, Estocolmo, Helsinque, Copenhague e Seul.

A partir desta busca, foram elencados 18 temas que englobam os dados produzidos por municípios brasileiros: Agricultura; Cultura, Esporte e Lazer; Demografia; Economia; Educação; Eleições; Dados Geográficos; Saúde; Indústria, Comércio e Serviços; Infraestrutura e Urbanismo; Legislação; Recursos Naturais e Meio Ambiente; Administração Pública; Segurança; Ciência e Tecnologia; Sociedade e Bem-estar; Turismo; e Transporte e Mobilidade. Ainda nesta etapa foram organizadas as atividades que deveriam ser desempenhadas no processo de construção da OGDPub, ajustando com o tempo disponível para finalizá-la.

Fase de Reuso

Nesta fase foram identificados os recursos ontológicos que poderiam ser reusados na construção da OGDPub [SUÁREZ-FIGUEROA, 2008]. Após buscas e análises em ontologias já existentes, foi definido o reuso de 4 ontologias que seriam conectadas à OGDPub:

- *Dublin Core Metadata Element Set*⁷ – vocabulário composto por 15 termos utilizados para descrição de recursos;
- *The Organization Ontology*⁸ – ontologia utilizada para descrever estruturas organizacionais;
- *Friend of a Friend*⁹ (FOAF) – ontologia para descrição de pessoas, suas atividades e seus relacionamentos com outras pessoas e objetos;
- *Simple Knowledge Organization System*¹⁰ (SKOS) – ontologia utilizada para representar a estrutura básica e conteúdo de esquemas conceituais, tais como listas, tesouros, taxonomias, entre outros.

A *Dublin Core Metadata Element Set* proporciona à OGDPub o conjunto de metadados para descrição dos *datasets* publicados por municípios brasileiros, tais como:

⁷ <http://dublincore.org/documents/dces/>

⁸ <https://www.w3.org/TR/vocab-org/>

⁹ <http://xmlns.com/foaf/spec/>

¹⁰ <https://www.w3.org/TR/2008/WD-skos-reference-20080829/skos.html>

criador, formato, língua, publicador, título, tema, descrição, identificador, entre outros. Esses metadados, quando descrevem o *dataset*, permitem que este seja mais facilmente encontrado na *Web*. Com a Ontologia Organizacional (*The Organization Ontology*) é possível descrever toda a estrutura organizacional de um município, com suas secretarias, coordenações, departamentos, entre outros.

Esta ontologia também permite relacionar as pessoas que são colaboradoras em um município e possibilita também a representação de portal de dados abertos do município. A ontologia FOAF representa os indivíduos que estão vinculados ao município e que estão inseridas no processo de publicação de dados abertos. Por fim, a ontologia SKOS, permite representar o universo de publicação de dados abertos por município, como um todo.

Fase de *Design*

Nesta fase o objetivo principal foi tornar a OGD_{Pub} uma ontologia multilíngue através do plug-in *LabelTranslator* [DZBOR, 2009]. Inicialmente, as classes foram implementadas em português. Na fase de reuso de ontologia, as ontologias importadas forma construídas a partir do idioma inglês. Dessa maneira, foram utilizados *labels* (etiquetas) que adaptavam o nome das classes aos dois idiomas, inglês e português.

É importante ressaltar que a OGD_{Pub} foi pensada, em sua concepção, para atender ao domínio dos municípios brasileiros. Todavia, a estrutura do Estado Brasileiro não é exclusividade do país. Outros países possuem estrutura igual ou semelhante, composta pela União, Distrito-Federal, Estados e Municípios. Dessa forma a OGD_{Pub} pode ser utilizada também nestes contextos.

Fase de Implementação

Nesta fase, os recursos ontológicos identificados anteriormente para reuso foram integrados à OGD_{Pub}. Aqui, concretamente, através da ferramenta *NeOn Toolkit*, os recursos ontológicos identificados, avaliados e escolhidos na Fase de Reuso foram agregados à OGD_{Pub}. Nenhuma das ontologias foram reusadas por completo, uma vez que não eram todos os termos e relações aderentes à OGD_{Pub}, mas foram selecionados recursos que se faziam necessários para que a ontologia construída fosse consistente.

Fase de Manutenção

A manutenção da ontologia se dá a partir da detecção de erros ou falta de conhecimento implementado. Nesta fase, após detectado algum desses aspectos, a equipe de desenvolvimento deve retornar para a fase de concepção da ontologia para atualizá-la [SUÁREZ-FIGUEROA; GOMEZ-PEREZ; FERNANDEZ-LOPEZ, 2012]. Na OGD_{Pub} ainda não foi realizada nenhuma manutenção, uma vez que não foi detectado nenhum erro ou conhecimento não explicitado na ontologia.

3. Considerações Finais

Neste artigo foi descrito o processo de desenvolvimento da OGD_{Pub}, ontologia para ampliar e promover o uso de dados abertos governamentais municipais no Brasil. A OGD_{Pub} possui alguns aspectos que conseguem cumprir com estes objetivos. O primeiro deles é o conjunto de metadados proveniente da ontologia *Dublin Core*, que

descreve o *dataset* e facilita com que este seja encontrado na *Web* através de ferramentas de busca.

Outro aspecto é a classificação que a OGD_{Pub} propõe para que os *datasets* publicados por municípios sejam organizados em temas. Esta classificação foi concebida a partir de uma linguagem cidadã, com o objetivo de tornar simples a compreensão dos dados pela sociedade em geral.

Em âmbito organizacional, a OGD_{Pub} possibilita a descrição de toda estrutura organizacional de um município. Além disso, a ontologia se volta para a responsabilização do cidadão e do município, sendo necessária a indicação do agente público responsável pela publicação do *dataset*. Acreditamos que a responsabilização contribui na garantia da autenticidade dos dados.

De maneira geral, a OGD_{Pub}, ampliando e promovendo o uso de dados abertos governamentais municipais, fomenta também a geração de negócios e serviços inovadores para a sociedade em geral.

Para trabalhos futuros, a OGD_{Pub} será verificada através de cenários de uso. Em suma, serão elencados um conjunto de *datasets* de algum município brasileiro, possivelmente Rio de Janeiro (prefeitura que disponibiliza a maior quantidade de *datasets* no país), e a ontologia será utilizada para instanciá-los. Por fim, deseja-se implementar efetivamente a ontologia nos municípios brasileiros.

Referências

- ATTARD, Judie et al. Value Creation on Open Government Data. In: **2016 49th Hawaii International Conference on System Sciences (HICSS)**. IEEE, 2016. p. 2605-2614.
- DZBOR, Martin. D5. 7.1 Plan for evaluating support for the NeOn methodology in the NeOn Toolkit. 2009.
- GÓMEZ-PÉREZ, Asunción; SUÁREZ-FIGUEROA, Mari Carmen. Scenarios for building ontology networks within the NeOn methodology. In: **Proceedings of the fifth international conference on Knowledge capture**. ACM, 2009. p. 183-184.
- OPEN DEFINITION. Definição de conhecimento aberto [Internet]. [atualizado em 2014 set 14; citado em 2016 abr 05]. Disponível em: <<http://opendefinition.org/>>
- SUÁREZ-FIGUEROA, Mari Carmen. D5. 4.1. NeOn Methodology for Building Contextualized Ontology Networks. 2008.
- SUÁREZ-FIGUEROA, Mari Carmen; GOMEZ-PEREZ, Asuncion; FERNANDEZ-LOPEZ, Mariano. The NeOn methodology for ontology engineering. In: **Ontology engineering in a networked world**. Springer Berlin Heidelberg, 2012. p. 9-34.
- ZANCANARO, Airton. Produção de recursos educacionais abertos com foco na disseminação do conhecimento: uma proposta de framework. 2015. 192 f. Tese (Doutorado em Engenharia e Gestão do Conhecimento) – Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis. 2015.