

OGDPub: Uma Ontologia para Publicação de Dados Abertos Governamentais

Larissa Mariany Freiberger Pereira¹, José Leomar Todesco¹

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

freiberger.lm@gmail.com, tite@egc.ufsc.br

Abstract. *In Brazil, government entities are required to publish their data in an open format, however most municipalities are unable to comply with this obligation. In this article we present OGD Pub, an ontology to support the publication of open government data by Brazilian municipalities. The proposed ontology provides a metadata framework for describing the datasets, allows the municipality's organizational structure to be represented and proposes a dataset classification scheme in a language understandable to the citizen. OGD Pub was developed based on NeOn Methodology and is represented in OWL language.*

Resumo. *No Brasil, as entidades governamentais são obrigadas a publicar seus dados em formato aberto, todavia a maior parte dos municípios não conseguem cumprir com esta obrigação. Neste artigo é apresentada a OGD Pub, uma ontologia para apoiar a publicação de dados abertos governamentais por municípios brasileiros. A ontologia proposta fornece um arcabouço de metadados para descrição dos datasets, permite que a estrutura organizacional do município seja representada e propõe um esquema de classificação de datasets em linguagem compreensível ao cidadão. A OGD Pub foi desenvolvida a partir da Metodologia NeOn e está representada na linguagem OWL.*

1. Introdução

O volume de dados abertos disponibilizado por agências governamentais na *Web* vem crescendo consideravelmente e os portais de dados abertos governamentais se tornaram um importante canal de comunicação entre governo e cidadãos (Ding et al. 2011; Jetzek et al. 2014).

A publicação de dados abertos por agências governamentais exerce um papel importante no processo de construção de um governo transparente e participativo, além de potencializar a inovação no terceiro setor da sociedade. No que tange à transparência e à participação, a publicação de dados abertos governamentais permite que cidadãos e organizações que compõem a sociedade estejam a par das atividades desenvolvidas por seus governos e se engajem no processo de criação de políticas públicas. Em relação à inovação, alguns exemplos de aplicações desenvolvidas para consumir dados abertos governamentais podem ser citadas: serviços de informação de tráfego em tempo real e localizadores de espaços públicos, como banheiros, escolas, feiras, entre outros (Janssen 2011; Hendler et al. 2012; Jetzek et al. 2014).

No Brasil, a abertura da informação ao cidadão é obrigatória. Todavia, a maior parte dos municípios brasileiros publicam apenas relatórios com dados já processados,

difíceis de serem utilizados pelo cidadão e pela sociedade em geral. Os municípios que possuem um catálogo de dados e publicam *datasets* abertos, os publicam em formatos não-estruturados, dificultando também o uso destes para gerar negócios e impulsionar a inovação. É comum nesses portais a ausência de metadados e informações sobre a proveniência¹ dos dados. O uso de metadados para descrever *datasets*, no contexto governamental, se torna importante para que os dados sejam encontrados, compreendidos e acessados na *Web*, bem como dar proveniência a estes.

Diversos autores concordam sobre o notável valor e potencial que há nos dados produzidos por agências governamentais (Ohemeng e Ofosu-Adarkwa 2015; Shadbolt e O'hara 2013; Shadbolt et al. 2012; Ding et al. 2011; Jetzek et al. 2014). Todavia, para Janssen et al. (2014), só há valor e potencial no dado governamental quando este é consumido, quando passa a ter alguma utilidade. Dessa maneira, entende-se que o valor e potencial não estão no dado em si, mas no uso que se faz deste.

No presente artigo será apresentada uma ontologia de domínio cujo objetivo é apoiar a publicação de dados abertos governamentais por municípios brasileiros. A ontologia garante um arcabouço de metadados capazes de descrever os *datasets* publicados por municípios, tornando-os mais simples de serem encontrados na *Web* e possibilitando que esses *datasets* sejam classificados a partir de uma linguagem compreensível ao cidadão comum. O artigo e a ontologia são frutos da dissertação² apresentada no Programa de Pós-graduação em Engenharia e Gestão do Conhecimento na Universidade Federal de Santa Catarina e está disponível no repositório institucional da universidade.

2. Ontologias

O termo “Ontologia”, oriundo da filosofia, diz respeito ao estudo do ser (Gruber 2009). Se trata de uma parte da metafísica que trata da estrutura de sistemas e, atualmente, está associada à organização e classificação do conhecimento (McComb 2004). No ramo da Engenharia do Conhecimento o termo não tem uma definição universal, sendo cunhado por autores diferentes, a partir de visões diferentes.

Nesta pesquisa, o constructo ONTOLOGIA é compreendido a partir da definição cunhada por Studer et al. (1998, p. 184): uma ontologia é “uma especificação formal e explícita de uma conceitualização compartilhada”. É uma conceitualização porque se refere a um modelo abstrato de algum fenômeno observado no mundo a partir da identificação e definição dos conceitos relevantes no contexto deste fenômeno. É explícita porque os conceitos utilizados e as restrições sobre seu uso são explicitamente definidos. É formal porque a ontologia é processável por máquina, o que exclui a linguagem natural. Por fim, é compartilhada porque a ontologia captura o conhecimento consensual, isto é, não se trata do conhecimento restrito a um indivíduo, mas do conhecimento aceito por um determinado grupo de pessoas.

Uma ontologia pode ser expressa por $O = \{C, R, I, A\}$, onde C é um conjunto de classes que representam conceitos de um determinado domínio, R é um conjunto de relações entre as classes definidas, I é um conjunto de instâncias ou

¹ Origem de um objeto de dados ou processo pelo qual o dado chegou em um bando de dados (Buneman et al., 2001).

² Disponível em: <<https://repositorio.ufsc.br/xmlui/handle/123456789/179907>>. Acesso em 19 jun. 2020.

indivíduos do mundo real pertencentes à determinada classe e A é um conjunto de axiomas existentes dentro de um determinado domínio do conhecimento que modelam regras e restrições intrínsecas às instâncias.

2.1. Classificação de Ontologias

Diversos autores buscaram classificar ontologias de diferentes formas. Uma das classificações mais difundidas na literatura é de Guarino (1998). Trata-se de uma classificação simples, porém consistente. Segundo o autor, as ontologias podem ser classificadas em 4 tipos: (1) Ontologias genéricas; (2) ontologias de domínio; (3) ontologias de tarefa; e (4) ontologias de aplicação.

Para o autor, as ontologias genéricas são responsáveis por descrever conceitos gerais, tais como: espaço, tempo, objeto, evento, ação, entre outros. Em suma, esses conceitos genéricos independem de um domínio ou um problema particular. As ontologias de domínio e de tarefa descrevem, respectivamente, os termos relacionados a um domínio específico (medicina ou casa) e a uma tarefa ou atividade genérica (diagnosticar ou construir). Os termos inseridos na ontologia genérica são especializados nas ontologias de domínio e tarefa. Por fim, as ontologias de aplicação descrevem conceitos que dependem tanto de uma ontologia de domínio, quanto de uma ontologia de tarefa. Por vezes a ontologia de aplicação é uma especialização das ontologias de domínio e tarefa, quando relacionadas.

2.2. Representação de Ontologias

As ontologias podem ser representadas, basicamente, de duas formas: graficamente e formalmente. A representação gráfica é utilizada para a compreensão humana. Por outro lado, a representação formal é usada para que a ontologia seja consumida por máquinas. Para representar ontologias graficamente é possível utilizar grafos, estrutura de árvore, UML, entre outras. Já, para representação formal de ontologias, as linguagens mais populares para descrever ontologias de maneira formal são RDF/RDF-S e OWL (Isotani e Bittencourt 2015).

Por ser uma linguagem recomendada pela W3C, ser amplamente utilizada para representar ontologias atualmente e possuir maior expressividade em relação à outras linguagens, a OWL foi adotada para representar a ontologia proposta nesta pesquisa. A OWL (*Web Ontology Language*) é uma linguagem para definir e instanciar ontologias recomendada pela W3C, baseada em lógica descritiva (Wang et al. 2004). Atualmente, OWL “é a linguagem mais utilizada para representar ontologias formalmente, apresenta variantes da linguagem que lidam com a escalabilidade e a expressividade das ontologias e permite que aplicações com diferentes propósitos sejam construídas” (Isotani e Bittencourt 2015, p. 115). A linguagem OWL permite que significado e semântica sejam melhor expressos se comparada à outras linguagens, como XML, RDF e RDF-S. A linguagem OWL descreve um domínio do conhecimento a partir de classes, relações e instâncias e possibilita que seja feita uma rica descrição das características desses objetos (Bechhofer et al. 2009).

3. OGDPub

A literatura apresenta diversas metodologias voltadas para o desenvolvimento de ontologias, como *Methontology* (Fernández-Lopéz et al. 1997; Blázquez et al. 1998), *On-*

To-Knowledge (Staab et al. 2001) e *Diligent* (Pinto et al. 2004). A ontologia *Open Government Data Publish* – OGD_{Pub} – foi desenvolvida a partir da Metodologia NeOn. Esta metodologia propõe um fluxo de trabalho flexível, sugerindo atividades para a construção de redes de ontologias. Este é um grande diferencial da metodologia NeOn em relação às metodologias encontradas na literatura, pois não trata a processo de construção de ontologias de maneira rígida (Suárez-Figueroa 2010). Este aspecto foi consideravelmente relevante para a adoção da metodologia NeOn para orientar o processo de desenvolvimento da ontologia proposta. Além disso, a metodologia NeOn consegue explicar o processo de desenvolvimento de ontologias em um nível de granularidade semelhante às metodologias de desenvolvimento de *software*, dando detalhes em cada uma das etapas descritas. A metodologia ainda considera a atividade de levantamento de requisitos como uma etapa do processo de desenvolvimento de ontologias, dando diretrizes claras nesta dimensão (Suárez-figueroa 2010). Outro aspecto relevante para a adoção da metodologia NeOn nesta pesquisa foi o fato de a mesma possibilitar a construção colaborativa de ontologias através de atividades como reuso e reengenharia de conhecimentos já expressos formalmente (Gómez-Pérez e Suárez-Figueroa 2009).

A ontologia OGD_{Pub} foi desenvolvida a partir do cenário 4 proposto pela metodologia NeOn: reuso e reengenharia de recursos ontológicos. Por definição, o reuso de recursos ontológicos trata do uso de uma ontologia (ou parte desta) para a solução de um problema específico. A Metodologia NeOn ainda define a reengenharia de recursos ontológicos como o processo de recuperação e transformação de um modelo conceitual implementado através de uma ontologia em um novo modelo conceitual, adaptado à solução de um problema específico, resultando em uma nova ontologia. Em relação à escolha do modelo de ciclo de vida utilizada para o desenvolvimento da OGD_{Pub}, optou-se pelo Modelo Iterativo-Incremental, visto que foi necessário retornar à fase de reuso cada vez que uma ontologia foi importada para fornecer recursos ontológicos à OGD_{Pub}. Além disso, tratou-se de um projeto de curta duração, sem o envolvimento de muitas pessoas na equipe de desenvolvimento. Neste modelo de ciclo de vida, o desenvolvimento da ontologia se dá a partir de fases bem definidas, organizadas em uma espécie de cascata, isto é, uma fase deve ser completada para que se possa seguir para a fase seguinte (Suárez-Figueroa e Gómez-Pérez 2008).

O desenvolvimento da OGD_{Pub} se deu em cinco etapas, sendo elas: (1) Fase Inicial; (2) Fase de Reuso; (3) Fase de Reengenharia; (4) Fase de Design; e (5) Fase de Implementação. A Metodologia NeOn propõe, ainda, que, ao serem identificados erros ou falhas na ontologia, esta seja submetida à uma fase de manutenção. A OGD_{Pub} não passou por essa fase, uma vez que no processo de verificação da ontologia por meio de cenários de uso não foram encontrados erros. Todavia, trata-se de uma fase fundamental especialmente quando a ontologia for adaptada para uso por cada município.

Durante o processo de desenvolvimento da OGD_{Pub}, buscou-se dar à ontologia algumas propriedades consideradas importantes por Gruber (1993), sendo elas: clareza, coerência, extensibilidade, viés de codificação mínima e mínimo compromisso ontológico. Em relação à clareza, uma ontologia deve descrever exatamente o significado dos termos definidos de maneira objetiva, ou seja, os termos devem ter definições claras, sem ambiguidades, compreensíveis. Essas definições devem estar descritas na documentação da ontologia em linguagem natural. No que tange à coerência, uma ontologia deve fazer inferências que estejam de acordo com as definições atribuídas aos termos que a compõem. Aqui é possível verificar quão importante é a lógica para o

processo de construção de uma ontologia. Sobre a extensibilidade, é necessário que o vocabulário específico de uma ontologia possa ser estendido caso se inclua termos mais específicos de determinado domínio do conhecimento sem que seja necessário realizar a revisão das definições já existentes. A quarta propriedade, viés de codificação mínima, assegura que as definições dos termos que compõem a ontologia sejam autoexplicativas, isto é, estas definições não devem depender de uma codificação em nível de símbolos ou em nível de outras definições. Por fim, em relação ao critério de mínimo compromisso ontológico, é necessário que o compromisso ontológico assumido no processo de construção de uma ontologia se limite à representação do conhecimento desejado.

3.1. Fase Inicial

Nesta fase foram levantados os requisitos que deverão ser atendidos pela ontologia, bem como a viabilidade do projeto. Ainda nesta fase a metodologia NeOn sugere que a especificação de requisitos da ontologia seja descrita através do Documento de Especificação de Requisitos ou *Ontology Requirements Specification Document (ORS)*, apresentado na Tabela 1.

Tabela 1. *Ontology Requirements Specification Document*

Documento de Especificação de Requisitos	
1	Propósito Contribuir para a ampliação e promoção da utilização de dados abertos governamentais publicados por municípios brasileira
2	Escopo A ontologia deve atender à realidade dos municípios brasileiros
3	Linguagem de Implementação A ontologia formal será expressa em linguagem OWL
4	Usuários Pretendidos Municípios brasileiros comprometidos com a publicação de dados abertos (de modo específico, a equipe responsável pelo processo de publicação dos dados abertos governamentais que deverão trabalhar diretamente com a OGDPub)
5	Usos Pretendidos Ser a base para a extensão da plataforma CKAN para desenvolver um catálogo de dados para municípios brasileiros
6	Requisitos a. Não-funcionais Descrever a estrutura hierárquica organizacional do município Fornecer um conjunto rico de metadados para descrever os <i>datasets</i> Propor um esquema de classificação dos <i>datasets</i> publicados em uma linguagem compreensível ao cidadão
	b. Funcionais: Grupo das Questões de Competência Quanto <i>datasets</i> estão disponíveis no Catálogo de Dados? / Qual o órgão gerador do <i>dataset</i> ? / Quantas distribuições estão relacionadas a determinado <i>dataset</i> ? / Em que formato está a distribuição? / Qual o tamanho da distribuição? / Quem é o responsável pela publicação do <i>dataset</i> ? / Qual o contato do responsável pela publicação do <i>dataset</i> ? / De qual tema trata-se determinado <i>dataset</i> ? / Quando o <i>dataset</i> foi publicado? / Do que se trata determinado <i>dataset</i> ?
7	Pré-glossário de Termos a. Termos das Questões de Competência <i>Dataset</i> (8) / Catálogo de Dados (1) / Órgão (1) / Distribuições (3) / Formato (1) / Tamanho (1) / Responsável (1) / Contato (1) / Tema (1) / Quando [data] (1) b. Termos das Respostas PF, CSV, JSON, Educação, Saúde, Turismo, Mobilidade Urbana [...] c. Objetos Pessoa, Data, Descrição ...

Desde a concepção da ontologia OGDPub, identificou-se três requisitos que deveriam ser atendidos para atingir o seu objetivo, a saber, contribuir para a ampliação e promoção da utilização de dados abertos governamentais publicados por municípios brasileiros: (1) possibilidade da descrição da estrutura organizacional do município, para que fosse possível identificar de onde provém o dado, bem como o indivíduo responsável por sua publicação; (2) proposta de um conjunto de metadados que descrevesse o *dataset*, e (3) proposta de um esquema para classificação dos *datasets* publicados em uma linguagem compreensível ao cidadão leigo.

Para definir uma classificação coerente com a realidade da publicação de dados no contexto dos municípios brasileiros, visitou-se o Portal de Dados Abertos ou Catálogo de Dados Abertos das 9 capitais brasileiras que o possuem, a saber: Fortaleza, Goiânia,

Belo Horizonte, Curitiba, Recife, Rio de Janeiro, Natal, Porto Alegre e São Paulo, bem como da capital do Distrito Federal, Brasília. Foram visitados também os Portais de Dados Abertos de cidades de outros países que trabalham com a publicação de dados abertos em catálogo de dados, a saber: Madri (Espanha), Washington (Estados Unidos), Nova Iorque (Estados Unidos), Toronto (Canada), Londres (Inglaterra), Roma (Itália), Berlim (Alemanha), Paris (França), Estocolmo (Suécia), Helsínquia (Finlândia), Amsterdã (Holanda), Buenos Aires (Argentina), Ottawa (Canadá) e Seul (Coréia do Sul). Além disso, visitou-se dos portais de dados abertos governamentais que são referência no contexto da publicação de dados abertos governamentais: o data.gov (Portal de Dados Abertos dos Estados Unidos) e o data.gov.uk (Portal de Dados Abertos do Reino Unido). Ambos portais utilizam a plataforma CKAN³. Esta plataforma permite gerenciar o catálogo de dados, além de fornecer uma API para acesso automático aos *datasets*. Esses portais foram visitados e foram analisadas as classificações propostas por cada um deles para, posteriormente, ser proposto um esquema de classificação de *datasets* com uma linguagem próxima do cidadão, facilitando que este possa encontrar e compreender os dados publicados.

Ao final deste processo, a classificação proposta foi composta pelos seguintes temas: (1) Administração Pública, (2) Agricultura, (3) Ciência e Tecnologia, (4) Cultura, Esporte e Lazer, (5) Demografia, (6) Economia, (7) Educação, (8) Finanças, (9) Indústria, Comércio e Serviços, (10) Infraestrutura e Planejamento Urbano, (11) Legislação, (12) Recursos Naturais e Meio Ambiente, (13) Saúde, (14) Segurança Pública, (15) Sociedade e Bem-estar, (16) Transporte e Mobilidade e (17) Turismo. A composição desta classificação se deu a partir de temas globais utilizados pelos portais analisados, isto é, temas comuns encontrados em todos ou grande parte dos portais analisados (embora em alguns a nomenclatura utilizada possa ser diferente). É importante ressaltar que a OGD_{Pub} pode ser estendida, isto é, novos temas podem ser implementados a partir das necessidades específicas de cada município.

3.2. Fase de Reuso

Nesta etapa foram identificados os recursos ontológicos que deveriam ser reutilizados na construção da ontologia OGD_{Pub}. Foram identificados recursos ontológicos provenientes de três ontologias principais, sendo elas: (1) *The Organization Ontology*⁴ (Ontologia das Organizações) e (2) *Data Catalog Vocabulary*⁵ (DCAT); e (3) *Simple Knowledge Organization System*⁶ (SKOS).

A Ontologia das Organizações forneceu recursos ontológicos que permitiram a descrição da estrutura hierárquica de um município. Esta ontologia é capaz de descrever todo e qualquer tipo de organização, sendo esta generalista. Para a construção da OGD_{Pub}, foram considerados alguns dos recursos ontológicos presentes nesta ontologia. É importante ressaltar também que a própria Ontologia das Organizações reusa alguns recursos ontológicos da ontologia *Friend of a Friend*⁷ (FOAF). Por consequência, esta foi reusada também na ontologia OGD_{Pub}.

³ <http://ckan.org/instances/#>

⁴ <https://www.w3.org/TR/vocab-org/#org:Organization>

⁵ <https://www.w3.org/TR/vocab-dcat/>

⁶ <https://www.w3.org/2004/02/skos/>

⁷ <http://xmlns.com/foaf/spec/>

A ontologia DCAT forneceu o arcabouço de metadados para descrever os *datasets* publicados por municípios brasileiros. Esta ontologia reusa recursos ontológicos da ontologia *DCMI Metadata Terms*⁸ e, como citado anteriormente, conseqüentemente a ontologia OGDPub reusou também alguns desses recursos ontológicos. Embora existam outras ontologias que fornecem um arcabouço de metadados consideravelmente completo para descrever qualquer tipo de recurso, optou-se por utilizar a ontologia DCAT, uma vez que esta foi desenvolvida, de forma particular, para descrever catálogos de dados que são publicados na *Web*, sendo este também, de maneira generalista, um dos propósitos da ontologia OGDPub.

Por fim, a terceira ontologia elencada para reuso de recursos ontológicos, SKOS, proporcionou a construção do esquema de classificação dos *datasets*. A ontologia SKOS é própria para representar esquemas de conhecimento, sejam eles, tesouros, esquemas de classificação, entre outros.

3.3. Fase de Reengenharia

Nesta fase, parte dos recursos ontológicos identificados na fase anterior foram transformados e adaptados à ontologia OGDPub. Aqui, a principal atividade de reengenharia se deu em alguns recursos ontológicos provenientes da ontologia *The Organization Ontology*. Isto porque, originalmente, a ontologia concebe as unidades organizacionais como pequenas organizações dentro de uma organização maior. No contexto da ontologia OGDPub as organizações são os municípios e estes são compostos com unidades organizacionais que, em suma, não se enquadraram como pequenas organizações. Por sua vez, estas unidades organizacionais podem ser compostas por subunidades organizacionais.

A atividade de reengenharia, de maneira concreta, resulta em uma mudança no modelo conceitual do recurso ontológico ou da ontologia a ser reusada e, de maneira específica, nesta pesquisa foi de suma importância para atender aos requisitos inicialmente levantados para o desenvolvimento da ontologia OGDPub.

3.4. Fase de Design

Nesta fase foi desenvolvido um modelo informal⁹ da ontologia OGDPub a partir dos requisitos levantados na fase inicial, bem como dos recursos ontológicos reutilizados e transformados nas fases de Reuso e Reengenharia. Este modelo informal está ilustrado referida dissertação.

3.5. Fase de Implementação

Nesta fase, o modelo informal concebido na etapa anterior foi implementado em uma linguagem formal. Neste projeto, a linguagem escolhida para desenvolver a ontologia OGDPub foi a *Web Ontology Language* ou OWL, padrão da W3C, conforme citado anteriormente. A ontologia OGDPub foi desenvolvida através da ferramenta *Protégé*¹⁰, desenvolvida pelo Centro de Informática Biomédica, na Universidade de Stanford. A

⁸ <http://dublincore.org/documents/dcmi-terms/>

⁹ Disponível em: < <https://ibb.co/7NRD7nN>>. Acesso em: 9 mar. 2020.

¹⁰ <http://protege.stanford.edu/>

Figura 1(a) apresenta todas as classes que compõem da OGDPub implementadas por meio da ferramenta *Protégé*.

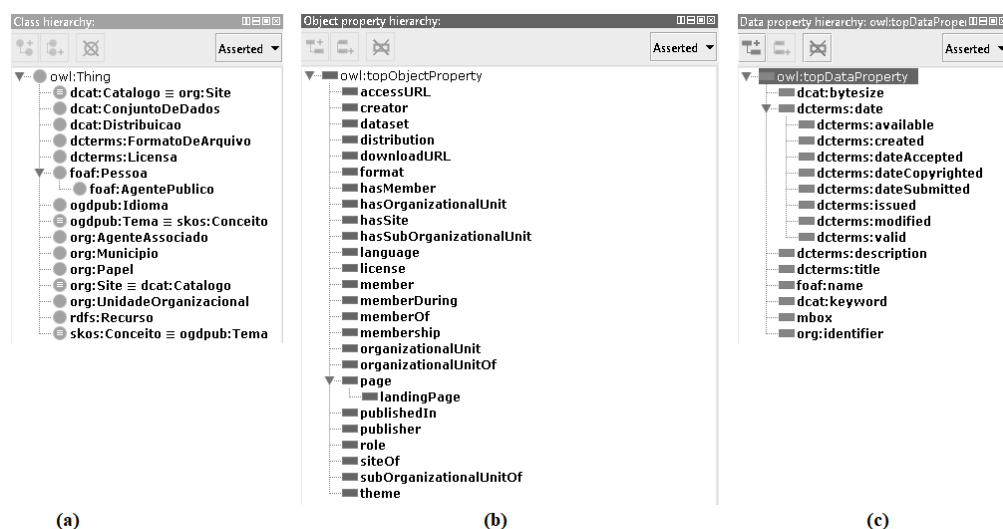


Figura 1. Classes da OGDPub

É possível observar na Figura 1(a) que a classe *Tema*, nativa da própria ontologia OGDPub, é equivalente à classe *Conceito*, classe oriunda da ontologia SKOS. O mesmo acontece com a classe *Catalogo*, oriunda da ontologia DCAT, e a classe *Site*, proveniente da ontologia *The Organization Ontology*. Na ontologia OGDPub a classe *AgentePublico* é implementada como uma subclasse de *Pessoa*, para descrever que todo agente público é, primeiramente, uma pessoa. Esta hierarquia, por exemplo, permite que, posteriormente, caso seja necessário, seja implementada uma classe para representar o conceito de Cidadão e este conceito herdaria todas as características de *Pessoa*. Este exemplo é possível de ser imaginado caso se deseje que esta ontologia descreva o processo de prestação de serviços do município. A hierarquia de classes permite também que uma pessoa que é agente pública possa, eventualmente, assumir o papel de cidadão e necessite de um serviço prestado do município.

A Figura 1(b) apresenta as *Object Properties* ou propriedades entre as classes que compõem a ontologia OGDPub. As *object properties* “*publishedIn*”, “*hasSubOrganizationalUnit*” e “*subOrganizationalUnitOf*” são nativas da ontologia OGDPub, as demais são propriedades provenientes das ontologias DCAT, *Dublin Core Terms* e *The Organization Ontology*, algumas sendo apenas reusadas e outras sendo adaptadas à OGDPub, passando pela tarefa de reengenharia. É possível observar também a hierarquia de classes entre as propriedades *Page* e *landingPage*, que são provenientes da ontologia FOAF. Ainda na Figura 1(b) é possível notar o uso do conceito de propriedades inversas, como é o caso das propriedades *hasSite* e *siteOf*, *hasMember* e *memberOf*, *hasOrganizationalUnit* e *organizationalUnitOf*, *hasSubOrganizationalUnit* e *subOrganizationalUnitOf* e *member* e *membership*.

A Figura 1(c) apresenta as *Data Properties* ou Propriedades de Dados que compõem a ontologia OGDPub. A hierarquia de classes observada entre a propriedade *date* e as demais possibilita que o município defina qual o grau de descrição, em termos de data, que deseja dar aos *datasets*. Neste sentido, o município pode utilizar a propriedade *date* para identificar a data de publicação do *dataset* e ignorar as demais propriedades da hierarquia, ou descrever o *dataset* com propriedades referentes à data de

publicação, data da última modificação, data de criação, data de disponibilização, entre outros.

A Figura 2, por fim, apresenta os indivíduos instanciados na classe *Tema*. Esta é, portanto, a classificação proposta para *datasets* e foi definida com instâncias para que cada município possa adaptá-la conforme suas necessidades. Por exemplo, uma cidade litorânea gera dados que uma cidade que não está localizada no litoral não gera ou, ainda, uma cidade como Rio de Janeiro gera dados sobre turismo que uma pequena cidade do interior do Acre não produz.



Figura 2. Classificação para *Datasets*

4. Verificação da OGD Pub

A ontologia OGD Pub foi verificada em duas etapas: (1) Instanciação de *Datasets* de um município brasileiro e (2) Pesquisas SPARQL. Na primeira etapa foram instanciados 5 *datasets* reais do município de Rio de Janeiro, conforme apresentado na Figura 3(a). O município do Rio de Janeiro foi escolhido por ser um dos principais entes municipais no contexto da disponibilização de dados aberto através de seu portal data.rio. Atualmente o catálogo de dados abertos do município conta com 1.012 *datasets*, disponibilizados em 12 formatos diferentes. Os cinco *datasets* instanciados na ontologia OGD Pub são disponibilizados no Catálogo de Dados do Rio de Janeiro em doze distribuições distintas, ilustrados na Figura 3(b). Já a Figura 3(c) mostra todas as relações sobre uma distribuição instanciada na ontologia.

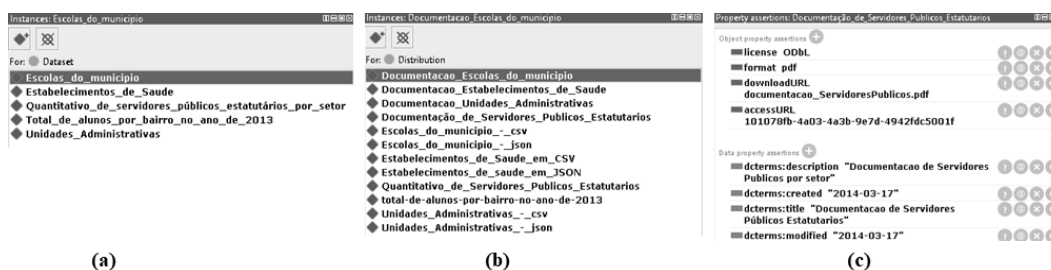


Figura 3. *Datasets* Instanciados na OGD Pub

Na segunda etapa da verificação da ontologia OGD Pub, foram realizadas consultas SPARQL na ontologia, simulando consultas possíveis de serem realizadas por cidadãos, baseadas nas questões de competência levantadas inicialmente. A primeira consulta (Figura 4a) teve como objetivo responder à seguinte questão: Quais são os

datasets publicados relacionados ao tema “Educação”? Como resultado, foram obtidos dois *datasets* que foram instanciados na ontologia OGDPub.

Posteriormente foi realizada uma consulta (Figura 4b) para obter o nome e o *e-mail* da pessoa responsável pela publicação do *dataset* “Escolas do município”. A terceira consulta (Figura 4c) buscou identificar quais as distribuições referentes ao *dataset* “Escolas do município”. Por fim, foi realizada uma consulta (Figura 4d) SPARQL para identificar a data de criação e a data da última modificação da distribuição “Escolas do Município” em CSV (Figura 4d). As consultas realizadas buscaram exemplificar possibilidades de consultas SPARQL que podem ser realizadas na ontologia OGDPub e, por isso, não esgotam essas possibilidades. Outras consultas podem ser realizadas para obter a estrutura organizacional do município, bem como o setor de onde provém os *datasets*, entre outros.

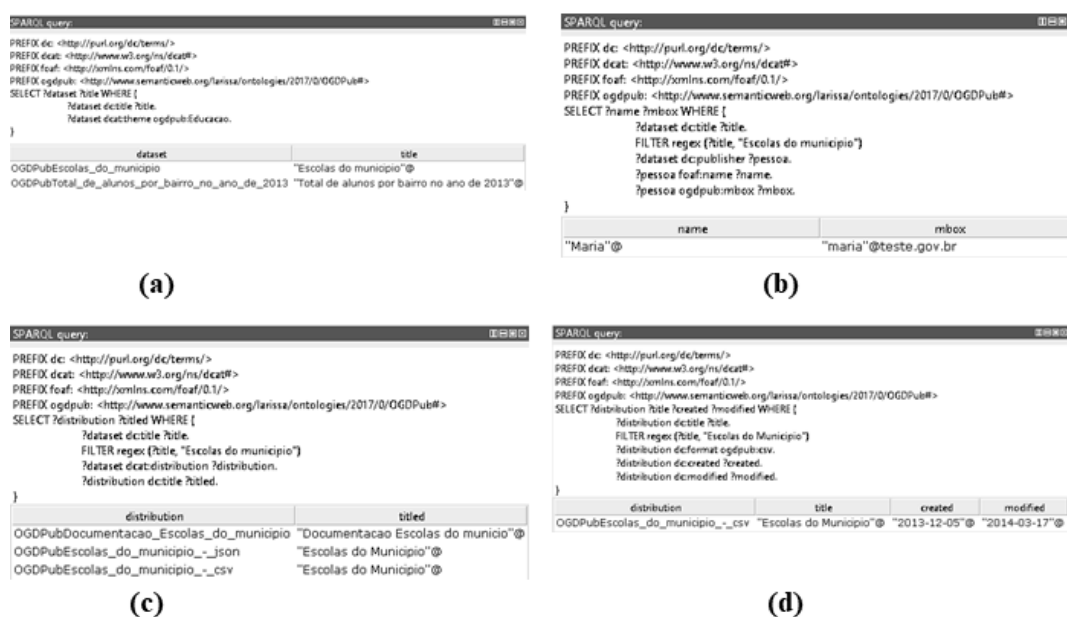


Figura 4. Consultas em SPARQL

5. Considerações Finais

A ontologia OGDPub foi desenvolvida a partir do estudo das experiências de municípios de outros países na publicação de dados abertos governamentais municipais. A ontologia foi representada por meio da linguagem OWL e foi verificada por meio de cenários de uso. Observa-se, portanto, a primeira limitação da ontologia.

Observou-se na literatura que as ontologias governamentais existentes tratavam, de modo geral, da estrutura hierárquica das organizações e nenhuma delas considerava o contexto da publicação de dados abertos. Poucas ontologias tratavam da questão da publicação de dados abertos, mas, por outro lado, não implementavam a estrutura hierárquica das organizações, tampouco propunham um esquema de classificação de *datasets* a partir da visão do usuário, no caso, o cidadão leigo. Para atingir o seu objetivo (a saber, apoiar municípios brasileiros no processo de publicação de dados abertos governamentais), a OGDPub contemplou três aspectos considerados relevantes no contexto: (1) descrição da estrutura hierárquica do município para garantir a proveniência do dado; (2) arcabouço de metadados para descrever o *dataset*, facilitando sua busca na

Web; e (3) proposta de um esquema de classificação para os *datasets* em uma linguagem cidadã, ou seja, compreensível ao cidadão leigo, que é flexível o suficiente para ser adaptada à realidade específica de cada município.

Como trabalhos futuros, sugere-se que a ontologia OGD Pub seja submetida a um processo formal de validação, uma vez que, neste contexto, foram dados os primeiros passos: a verificação da ontologia por meio de cenários de uso. Sugere-se também que a ontologia OGD Pub seja utilizada como base de uma extensão para a plataforma CKAN. Desde modo, os municípios brasileiros teriam à sua disposição uma plataforma para publicação de dados abertos adaptadas à sua realidade, tornando mais simples as tarefas de publicação, compartilhamento, busca e uso dos dados. Além disso, desenvolvedores teriam à sua disposição uma API para acesso direto e consumo dos dados tornados públicos.

Sugere-se, ainda, que a plataforma CKAN estendida com a ontologia OGD Pub seja implementada nos municípios brasileiros, inserindo-os neste movimento de dados abertos, uma vez que, no Brasil, a menor parte dos municípios aderiram ao mesmo. O objetivo é que, com a implementação da plataforma CKAN estendida com a ontologia OGD Pub, os municípios tenham mais facilidade em publicar seus dados, levando em consideração a qualidade da informação e, por outro lado, que os cidadãos e a sociedade em geral possam consumi-los de maneira mais simples. Uma vez que a abertura dos dados é um dos pilares de um governo aberto, espera-se, ainda, com a implementação da plataforma CKAN estendida pela ontologia OGD Pub que os municípios brasileiros tenham uma governança mais aberta e transparente e que os cidadãos sejam atores proativos neste novo modelo de governo participativo e colaborativo.

Agradecimentos

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Referências

- Bechhofer, S., Van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D. L., Patel-Schneider, P. F., & Stein, L. A. (2004). OWL web ontology language reference. *W3C recommendation*, 10(02).
- Blázquez, M., Fernández, M., García-Pinar, J. M., & Gómez-Pérez, A. (1998). Building ontologies at the knowledge level using the ontology design environment. In *Proc. of the 11th Int. Workshop on Knowledge Acquisition, Modeling and Mangement (KAW'98)*. Banff, Canada.
- Buneman, P., Chapman, A., Cheney, J., & Vansummeren, S. (2006, May). A provenance model for manually curated data. In *International Provenance and Annotation Workshop* (pp. 162-170). Springer, Berlin, Heidelberg.
- Ding, L., Lebo, T., Erickson, J. S., Difranzo, D., Williams, G. T., Li, X., ... & Flores, J. (2011). TWC LOGD: A portal for linked open government data ecosystems. *Journal of Web Semantics*, 9(3), 325-333.
- Fernández-López, M., Gómez-Pérez, A., & Juristo, N. (1997). Methontology: from ontological art towards ontological engineering.
- Gómez-Pérez, A., & Suárez-Figueroa, M. C. (2009). NeOn methodology for building ontology networks: a scenario-based methodology. In *Proceedings of the International Conference on Software, Services & Semantic Technologies*. Sofia.

- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2), 199-221.
- Gruber, T., Ontology, I. L. L., & Özsu, M. T. (2009). Encyclopedia of database systems. *Ontology*.
- Guarino, N. (Ed.). (1998). *Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy* (Vol. 46). IOS press.
- Hendler, J., Holm, J., Musialek, C., & Thomas, G. (2012). US government linked open data: semantic. data. gov. *IEEE Intelligent Systems*, (3), 25-31.
- Isotani, S., & Bittencourt, I. I. (2015). *Dados Abertos Conectados: Em busca da Web do Conhecimento*. Novatec Editora.
- Janssen, K. (2011). The influence of the PSI directive on open government data: An overview of recent developments. *Government Information Quarterly*, 28(4), 446-456.
- Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information systems management*, 29(4), 258-268.
- Jetzek, T., Avital, M., & Bjorn-Andersen, N. (2014). Data-driven innovation through open government data. *Journal of theoretical and applied electronic commerce research*, 9(2), 100-120.
- McComb, D. (2004). *Semantics in business systems: The savvy manager's guide*. Morgan Kaufmann.
- Ohemeng, F. L., & Ofosu-Adarkwa, K. (2015). One way traffic: The open data initiative project and the need for an effective demand side initiative in Ghana. *Government Information Quarterly*, 32(4), 419-428.
- Pinto, H. S., Staab, S., & Tempich, C. (2004). DILIGENT: Towards a fine-grained methodology for DIstributed, Loosely-controlled and evolvInG Engineering of oNTologies. In *Proceedings of the 16th European Conference on Artificial Intelligence* (pp. 393-397). IOS Press.
- Shadbolt, N., O'Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H., & Hall, W. (2012). Linked open government data: Lessons from data. gov. uk. *IEEE Intelligent Systems*, 27(3), 16-24.
- Shadbolt, N., & O'Hara, K. (2013). Linked data in government. *IEEE Internet Computing*, 17(4), 72-77.
- Staab, S., Studer, R., Schnurr, H. P., & Sure, Y. (2001). Knowledge processes and ontologies. *IEEE Intelligent systems*, 16(1), 26-34.
- Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: principles and methods. *Data & knowledge engineering*, 25(1-2), 161-197.
- Suárez-Figueroa, M. C. (2010). *NeOn Methodology for building ontology networks: specification, scheduling and reuse* (Doctoral dissertation, Informatica).
- Suárez-Figueroa, M. C., & Gómez-Pérez, A. (2008). Building ontology networks: How to obtain a particular ontology network life cycle. In *Proceedings of the third International Conference on Semantic Systems (I-Semantics)*.
- Wang, X. H., Zhang, D. Q., Gu, T., & Pung, H. K. (2004). Ontology based context modeling and reasoning using OWL. In *IEEE annual conference on pervasive computing and communications workshops, 2004. Proceedings of the second* (pp. 18-22). IEEE.