



Universidade Federal de Pernambuco

Centro de Informática - Cin

Graduação em Engenharia da Computação

EcOGD: Uma ontologia para descrição de Ecossistemas de Dados Abertos Governamentais

Glória de Fátima Andrade Barros Lima

TRABALHO DE GRADUAÇÃO

Recife, Dezembro de 2016



Universidade Federal de Pernambuco

Centro de Informática - Cin

Graduação em Engenharia da Computação

EcOGD: Uma ontologia para descrição de Ecossistemas de Dados Abertos Governamentais

Trabalho apresentado ao Programa de Graduação em Engenharia da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Engenharia da Computação.

Aluno: Glória de Fátima Andrade Barros Lima

Orientadora: Bernadette Farias Lóscio

Dedico este trabalho a Deus e a minha família.

Agradecimentos

Agradeço ao meu pai, Henrique, que mesmo não estando presente nessa etapa da minha vida, é meu guia e força para encarar os desafios da vida. Agradeço à minha mãe, Gorete, por ser meu exemplo de força, persistência e dedicação. Agradeço imensamente pelo amor que recebi deles e pelo esforço em me proporcionar uma educação de qualidade, pois como meu pai dizia: Conhecimento ninguém tira de você. Orgulho descreve o que eu sinto dos meus pais.

Agradeço aos colegas e amigos que estiveram presentes durante os anos de graduação, ao CIn que foi minha segunda casa e à professora Bernadette, por ter me acolhido como aluna de IC quando eu ainda não tinha experiência. Obrigada pela atenção, carinho e esforço em me ensinar e me fazer uma estudante melhor.

Resumo

Nos últimos anos com a publicação em massa de dados abertos governamentais muitas soluções têm sido desenvolvidas com o intuito de suprir as demandas dos usuários consumidores de dados abertos. Iniciativas privadas e governamentais estimulam o desenvolvimento de pesquisas e soluções que utilizem os dados abertos, mas tais aplicações acabam não sendo propriamente utilizadas ou terminam dispersas por falta de comunicação entre as partes interessadas. Este trabalho de graduação visa especificar e criar uma ontologia para o ecossistema de dados abertos governamentais através do mapeamento dos atores que compõem o ecossistema, os papéis que cada um exerce e os relacionamentos existentes no processo de produção de soluções. Além disso, uma base de conhecimento será gerada disponibilizando informações dos membros dos ecossistemas, resultados das parcerias entre eles, áreas de interesse, demandas, inferências, recomendações e novos dados sobre ecossistemas de dados abertos governamentais.

Palavras-chaves: Dados Abertos Governamentais, Ecossistema, Ontologia, Aplicações.

Abstract

In recent years with the mass publication of government open data many solutions have been developed to meet the demands of open data consumers. Private and government initiatives stimulate the development of research and solutions that use open data, but such applications end up not being properly used or end up dispersed due to lack of communication between the interested parties. This work aims to specify and create an ontology for the open government data ecosystem through the mapping of the actors that compose the ecosystem, the roles that each one plays and the relationships existing in the process of producing solutions. In addition, a knowledge base will be generated providing information from ecosystem members, results of partnerships among them, areas of interest, demands, data inferences, recommendations and new data about open government data ecosystems.

Keywords: Open Government Data, Ecosystem, Ontology, Applications.

Lista de Figuras

1	Diagrama dos Atores e Papéis	21
2	Ciclo da Metodologia 101	25
3	Tela inicial da ferramenta Protègè	35
4	Tela de criação das classes da ferramenta Protègè	35
5	Tela de criação das <i>datatype properties</i> da ferramenta Protègè	36
6	Tela de criação das <i>object properties</i> da ferramenta Protègè	36
7	Hierarquia de classes	37
8	Interface do Protègè para instanciação	38
9	Interface do Protègè com todos os indivíduos instanciados	39
10	Interface do Protègè para consultas em SPARQL	39

Lista de Tabelas

1	Atores e Papéis do Ecossistema	22
2	Termos Reusados para Classes	28
3	Termos Reusados para <i>Datatype Properties</i>	28
4	Termos Criados para Classes	31
5	Termos Criados para <i>Object Properties</i>	31
6	Termos Criados para <i>Datatype Properties</i>	32
7	Prefixos dos vocabulários utilizados nas consultas SPARQL	40
8	Resultado da consulta 1	41
9	Resultado da consulta 2	42

Sumário

1	Introdução	9
1.1	Motivação	9
1.2	Objetivos	10
1.3	Estrutura do Documento	11
2	Contextualização	11
2.1	Dados Abertos e Dados Abertos Governamentais	11
2.2	Ecosistemas de Dados Abertos	13
2.3	Web Semântica	14
2.3.1	Vocabulários e Ontologias	14
2.3.2	RDF, RDFS e OWL	16
2.3.3	SPARQL	18
3	Criação da Ontologia de Ecosistemas de Dados Abertos Governamentais	19
3.1	Metodologia	19
3.2	Mapeamento de atores, papéis e relacionamentos	20
3.3	Criação da ontologia EcOGD	24
3.3.1	Reuso de vocabulários	27
3.3.2	Criação de novos termos	30
3.3.3	Ferramenta Protègè	34
4	Criação da base de conhecimentos	37
5	Conclusão	42

1 Introdução

Nesta seção será apresentada a motivação para o desenvolvimento deste trabalho de graduação, os objetivos propostos e por fim será apresentada a estrutura do documento.

1.1 Motivação

Nos últimos anos com a publicação em massa de dados abertos governamentais muitas soluções tem sido desenvolvidas com o intuito de divulgar os dados e torná-los úteis aos cidadãos. Iniciativas privadas e governamentais estimulam o desenvolvimento de pesquisas e aplicações que utilizam os dados abertos para que sejam gerados serviços que incitem a transparência. Dawes et al. (2016) em seu artigo sobre planejamento e criação de programas de dados abertos governamentais, fala da prática recente dos governos de tornar seus conjuntos de dados de diversos âmbitos disponíveis em formatos que sejam lidos por máquina, sejam disponíveis, de fácil acesso e que possam ser acessados através de portais online sem custo para o usuário.

Em 2011 surgiu o *Open Government Partnership* (OGP)¹ que é uma parceria entre 8 nações ao redor do mundo (incluindo Estados Unidos da América, Brasil, Indonésia, México, Noruega, Filipinas, África do Sul e Reino Unido) que desejam combater a corrupção e aumentar a prestação de contas. Além dos países, fazem parte do comitê organizações da sociedade civil. Para se juntar à OGP é necessário que o país obedeça a critérios mínimos de elegibilidade focados em práticas existentes relacionadas com transparência fiscal, acesso à informação, engajamento dos cidadãos e abertura de contas dos governantes eleitos e dos funcionários públicos. Cada país deve ter um plano de governo aberto detalhando as práticas que estão sendo executadas, em progresso e que serão implantadas. Atualmente, a OGP é composta por 70 países que assumiram o compromisso de aumentar a transparência dos seus governos por meio da abertura de informações sobre as atividades e decisões governamentais seguindo os padrões básicos de publicação de dados abertos (Dietrich et al., 2009)

Com a abertura e publicação dos dados abertos governamentais, surge o incentivo ao desenvolvimento de serviços e aplicações que ajudem a sociedade a tomar decisões, a se informar sobre as contas públicas e a cobrar por melhorias. Nesse processo de produção

¹<http://www.opengovpartnership.org>

de soluções que utilizam os dados abertos governamentais como base de informação, existem atores que exercem diversos papéis, incluindo publicadores e consumidores de dados e soluções que fazem uso desses dados, bem como investidores, os quais viabilizam tais iniciativas. Esses atores exercem seus papéis de forma independente mas precisam trabalhar de forma conjunta para que o ecossistema de dados abertos governamentais tenha como resultado soluções úteis e sustentáveis para a sociedade.

Para que os cidadãos se empoderem das informações é preciso que elas estejam dispostas de forma clara, com uma interface intuitiva e que os dados sejam relevantes para o público que irá consumir. Portanto, para o sucesso de um ecossistema dessa natureza é preciso a análise minuciosa dos atores que o compõem, dos relacionamentos existentes entre eles e do contexto que as soluções são produzidas. Com a análise de portais como o portal Brasileiro de Dados Abertos² e o portal de Dados Abertos da Prefeitura do Recife³, pode-se perceber que há lacunas nos processos de produção e consumo dos dados. Os dados abertos publicados nesses portais de dados abertos governamentais e as soluções produzidas acabam não sendo propriamente utilizados ou terminam dispersos por falta de divulgação e comunicação entre as partes interessadas (Gama and Lóscio, 2014).

A motivação deste trabalho é a depreciação que as soluções produzidas com os dados abertos acabam tendo com o tempo e a falta de um canal de comunicação que divulgue informações e estreite relações entre os diferentes atores que compõem o ecossistema de dados abertos governamentais. Além disso, outra deficiência é a ausência de informações sobre o funcionamento desse tipo de ecossistema, quais os principais atores, seus papéis e os relacionamentos existentes. Por isso, a criação da ontologia EcOGD, além de explorar e revelar novos fluxos de comunicação entre os atores, torna possível a troca de informações que sejam relevantes entre os atores e ajuda no surgimento de novas parcerias, investimentos, clientes e soluções.

1.2 Objetivos

Este trabalho tem como objetivo principal a especificação de uma ontologia para descrição dos principais atores dos ecossistemas de dados abertos governamentais, bem como seus papéis e relacionamentos.

²<http://www.dados.gov.br/>

³<http://www.dados.recife.pe.gov.br/>

Como objetivos específicos destacam-se:

- Mapeamento dos atores dos ecossistemas de dados abertos governamentais, bem como seus papéis e relacionamentos entre eles.
- Especificação e criação de uma ontologia para descrição dos principais conceitos associados aos ecossistemas de dados abertos governamentais.
- Geração da base de conhecimento para extração de informações que estreitem o canal de comunicação entre os atores, facilitando assim o desenvolvimento de produtos e serviços baseados nos dados abertos governamentais.

1.3 Estrutura do Documento

Os capítulos seguintes estão estruturados da seguinte forma: o capítulo 2 faz a contextualização deste trabalho, apresentando os conceitos de dados abertos e dados abertos governamentais, ecossistemas de dados abertos e conceitos e técnicas utilizadas na Web Semântica. O capítulo 3, por sua vez, descreve todo o trabalho realizado, desde a metodologia utilizada, o mapeamento dos atores, papéis e relacionamentos, até os detalhes para a criação da ontologia EcoGD. No capítulo 4, é descrita a criação da base de conhecimentos e quais tipos de informações podem ser extraídas dela. Por fim, o capítulo 5 expõe a conclusão deste trabalho e sugestões para projetos futuros.

2 Contextualização

Neste capítulo serão abordados conceitos essenciais para o entendimento deste trabalho. A seção 2.1 descreve o que são dados abertos e dados abertos governamentais. Na seção 2.2 é descrito o cenário atual dos ecossistemas de dados abertos. Na seção 2.3 são apresentados o conceito de Web Semântica, o uso de vocabulários e ontologias e as principais tecnologias da Web Semântica, como RDF, RDFS, OWL e SPARQL.

2.1 Dados Abertos e Dados Abertos Governamentais

O grande volume de dados publicados em formato aberto na Web tem estimulado fortemente o interesse em soluções que viabilizam a publicação e o consumo de dados na

Web (Lóscio, 2015). O termo Dados Abertos corresponde à ideia de que certos dados devem estar disponíveis para que todos usem e publiquem, sem restrições de direitos autorais e patentes (Dietrich et al., 2009). Nesse cenário, tanto a publicação quanto o consumo dos dados são tarefas fundamentais, pois não basta apenas publicar os dados, é importante que os dados sejam publicados de maneira que sejam úteis para seus potenciais consumidores.

Segundo a *Open Knowledge Foundation*⁴, dados abertos são dados que podem ser livremente utilizados, reutilizados e redistribuídos por qualquer pessoa - sujeitos, no máximo, à exigência de atribuição à fonte original e compartilhamento pelas mesmas licenças em que as informações foram apresentadas. Um dado é considerado aberto quando apresenta as seguintes características (Dietrich et al., 2009):

1. Disponibilidade e acesso: o dado precisa estar disponível por inteiro. Deve estar num formato conveniente e modificável;
2. Reuso e redistribuição: o dado precisa ser fornecido em condições de reuso e redistribuição podendo ser combinado com outros;
3. Participação universal: todos podem usar, reusar e redistribuir o dado sem restrições de áreas, pessoas ou grupos.

O conceito de publicação de dados governamentais abertos é definido como disponibilizar, através da Internet, informações e dados governamentais de domínio público para a livre utilização pela sociedade (Agune and Bolliger, 2010). A W3C⁵ define dados abertos governamentais como: "a publicação e disseminação das informações do setor público na web, compartilhados em formato bruto e aberto, compreensíveis logicamente, de modo a permitir sua reutilização em aplicações digitais desenvolvidas pela sociedade". O uso e aplicação dos dados abertos vem se expandindo e se popularizando ao longo dos anos. A abertura dos dados governamentais vem acontecendo em períodos diferentes para cada país e com isso contribui para o aumento progressivo da transparência. Além disso, há a possibilidade de criação de novas informações e serviços que podem se originar a partir da interação entre o governo e a sociedade através da exploração dos dados abertos governamentais. A colaboração entre governo e instituições privadas também é possível,

⁴<https://okfn.org/>

⁵<http://www.w3c.br>

uma vez que estas criem conteúdo a partir da reutilização dos dados. (Vaz et al., 2010) Pesquisas apontam os benefícios que os dados abertos podem trazer para a economia e para o desenvolvimento social, pois, como analisado por Manyika et al. (2013), os setores de educação, transporte, comércio, eletricidade, óleo e gás, saúde e finanças poderiam ter aumentado o potencial econômico se explorassem o uso e compartilhamento de dados abertos.

2.2 Ecossistemas de Dados Abertos

No ecossistema de dados abertos os consumidores dependem dos dados que são publicados pelos provedores de dados abertos para que possam fazer uso deles e os provedores precisam do feedback dos consumidores sobre a qualidade e utilização desses dados abertos. Com isso, os consumidores e provedores de dados abertos interagem nesse fluxo (Zuiderwijk et al., 2014). Com essa interação entre os consumidores e provedores de dados abertos Pollock (2011) afirma que o ecossistema de dados abertos contém ciclos com loops de feedback, compartilhamento de dados de volta para os provedores e compartilhamento entre os infomediários, os quais criam aplicações e adicionam valor aos conjuntos de dados. Pollock (2011) descreve três formas para desenvolver um ecossistema de dados abertos, como 1) infomediários deveriam publicar o que eles produzem, 2) melhores práticas para publicação e compartilhamento de dados e 3) notificação dos provedores de dados ao adicionar ou atualizar conjunto de dados.

Atualmente existem trabalhos que exploram o contexto dos ecossistemas de dados abertos governamentais, como os dados abertos estão sendo publicados e quais os resultados dessa experiência. Canares and Shekhar (2015) analisaram o contexto e as configurações de governança em 17 casos de estudo na África, Ásia e América Latina. Eles propuseram um sumário qualitativo da estrutura dos governos sub-nacionais e o que tem sido feito para acelerar o crescimento dos dados abertos governamentais. Dawes et al. (2016) desenvolve um modelo preliminar de ecossistema para planejar e projetar programas de dados abertos governamentais. Harrison et al. (2012) discute os conceitos para entender o que é um ecossistema e alguns princípios que caracterizam seu funcionamento; após isso eles aplicam esses conceitos ao funcionamento do governo e propõem que os governantes se engajem nas estratégias para construir ecossistemas de governos abertos. O cenário dos dados abertos governamentais foi também explorado por Gama and Lóscio

(2014) quando eles sugerem um ecossistema de dados abertos que é economicamente autossustentável e é baseado no uso de dados abertos como serviço. Zuiderwijk et al. (2014) foca nos elementos essenciais do ecossistema de dados abertos que possibilitam a publicação e uso dos dados abertos, foram analisados portais, aplicações e ferramentas que formam o ecossistema. Finalmente, Heimstädt et al. (2014) propõe uma conceitualização do ecossistema de dados abertos do Reino Unido.

2.3 Web Semântica

A Web Semântica tem como principal objetivo fornecer ferramentas para que os computadores possam interpretar os dados que estão na Web. Quando a máquina é capaz de apreender a semântica, consegue melhorar o resultado das buscas e automatizar trabalhos que eram feitos manualmente no passado. Ela é uma extensão da Web atual, onde as informações ganham um significado que pode ser processado por uma máquina e permite que humanos e computadores trabalhem em cooperação, segundo seu criador Tim Berners-Lee (Berners-Lee et al., 2001).

Mas, para que a Web Semântica venha estar presente no dia-a-dia das pessoas é necessário que exista um padrão para as tecnologias envolvidas nesse conceito, e a W3C (*World Wide Web Consortium*) tem padronizado os procedimentos para garantir a interoperabilidade dos dados.

Nessa seção serão abordados os conceitos de vocabulários e ontologias, as principais tecnologias no contexto da Web Semântica, como o RDF e RDFS (*Resource Description Framework e RDF Schema*), OWL (*Web Ontology Language*) que é a linguagem para criação de ontologias e o SPARQL (*SPARQL Protocol and RDF Query Language*) que é uma linguagem para consultar fontes de dados em RDF.

2.3.1 Vocabulários e Ontologias

Na Web Semântica, os vocabulários definem os conceitos e os relacionamentos usados para descrever e representar uma área específica. Vocabulários são utilizados para classificar os termos que podem ser usados em uma aplicação, caracterizar possíveis relacionamentos e definir possíveis restrições na utilização dos termos⁶. Atualmente, o uso de vocabulários tem facilitado a interoperabilidade dos dados na Web pois o seu uso facilita a

⁶<https://www.w3.org/standards/semanticweb/ontology>

comunicação entre os metadados. Existem repositórios online que armazenam conjuntos de vocabulários de diversos domínios, como Schema.org⁷, LOV⁸ e LinDA⁹. No âmbito dos dados abertos governamentais, o governo brasileiro criou um repositório de vocabulários e ontologias do governo eletrônico chamado e-VoG¹⁰, o qual reúne conjunto de padrões, ferramentas e metodologias para capacitação em ontologias, manual de boas práticas e processo de engenharia de ontologias e política de URIs para publicação de dados no governo. Para a criação de um novo vocabulário é preciso avaliar as necessidades de expressividade dos metadados, pois é essencial a reutilização do maior número possível de elementos de ontologias já existentes, evitando assim a duplicação de referências diferentes aos mesmos conceitos (Laufer, 2015).

Gruber (1993) definiu ontologia como sendo a especificação explícita de uma conceitualização, onde a conceitualização se refere ao significado de conceitos e relações, dado o contexto de domínio e a especificação significa uma representação formal, declarativa e explícita dos mesmos conceitos e relações. A criação de uma ontologia acontece em etapas e tem como objetivo gerar bases de conhecimento que possam ser reutilizadas em diferentes aplicações (Guarino, 1997). Com o aumento da publicação de dados na internet, surgiu o problema da dispersão de informação, o qual usuários procuram por informações específicas mas não encontram por estarem soltas na Web e de difícil acesso. Com isso, para que os dados estejam interligados e acessíveis surgiu a necessidade de aumentar o uso de ontologias, pois elas facilitam a interoperabilidade, realizam a fusão de informações, geram inferências e informações que podem ser reutilizadas e passam a gerar novos dados (Isotani and Bittencourt, 2015). As ontologias têm se mostrado essenciais na criação de aplicativos para lidar com grandes quantidades de dados (McGuinness et al., 2004).

De acordo com Mizoguchi (2004) uma ontologia é composta por um conjunto de conceitos base que representam o conhecimento do domínio e o corpo de conhecimento que descreve o domínio utilizando os conceitos. O corpo de conhecimento é composto por uma hierarquia entre os conceitos, por um conjunto de relações entre os conceitos e por uma axiomatização de restrições semânticas entre os conceitos e relações (Isotani and Bittencourt, 2015). Kiryakov (2006) representa uma ontologia como sendo uma quádrupla de elementos $O=C,R,I,A$ que são essenciais para a criação da estrutura que representa o

⁷<https://http://schema.org>

⁸<https://lov.okfn.org/dataset/lov>

⁹<http://linda.epu.ntua.gr/vocabularies>

¹⁰<https://vocab.e.gov.br/>

conhecimento de um domínio. Cada elemento é definido da seguinte forma:

- C – é o conjunto de classes que representam os conceitos em um dado domínio de interesse;
- R – é o conjunto de relações ou associações entre os conceitos do domínio;
- I – é o conjunto de instâncias derivadas das classes, ou ainda, os exemplos de conceitos representados em uma ontologia;
- A – é o conjunto de axiomas do domínio que servem para modelar restrições e regras inerentes às instâncias.

2.3.2 RDF, RDFS e OWL

O *Resource Description Framework* (RDF) é um modelo de dados que possibilita a definição de afirmações, chamadas sentenças, sobre um recurso. O RDF foi proposto como uma possível solução para a limitação do XML, pois o XML não possui os requisitos necessários para descrever adequadamente a semântica de uma informação (Wagner Filho and Lôscio, 2010). Recurso é qualquer coisa sobre a qual se quer expressar uma ideia, podendo estar relacionado com dados ou com outros recursos através das sentenças. Uma sentença é estruturada no formato sujeito + predicado + objeto onde (Wagner Filho and Lôscio, 2010):

- Sujeito: Tem como valor o recurso do qual se quer escrever uma sentença.
- Predicado: Especifica um relacionamento entre sujeito e objeto, é geralmente nomeado por um verbo e permite relacionar um recurso a dados ou a outros recursos.
- Objeto: Denomina o recurso ou dado que se relaciona ao sujeito.

A estrutura de qualquer expressão em RDF é uma coleção de triplas, cada tripla composta por um sujeito, um predicado (ou propriedade) e um objeto. O conjunto de triplas é chamado de grafo RDF e cada tripla representa a declaração de um relacionamento entre as coisas representadas pelos nós que ela liga. Os nós do grafo RDF são os sujeitos e os objetos, o arco é o predicado, o qual denota o relacionamento. O significado de um grafo RDF é o conjunto das declarações correspondentes a todas as triplas contidas (Klyne and Carroll, 2006).

RDF *Schema* (RDFS) surgiu para prover construtores que permitem especificar formalmente um esquema. A ideia principal é unir RDFS + RDF de tal forma que todas as sentenças descritas em RDF obedeam à semântica descrita no esquema especificado em RDFS. O esquema é a modelagem do domínio de interesse (Wagner Filho and Lóscio, 2010). Se um esquema é processável por máquina é possível que uma aplicação aprenda algumas das semânticas das propriedades de tipo nomeadas no esquema. Entender um esquema RDF em particular é entender as semânticas de cada propriedade da descrição (Miller, 1998).

A OWL é uma linguagem desenvolvida e aprovada pelo W3C. Ela tenta satisfazer o formalismo exigido pela comunidade de Web Semântica para que programas possam compreender e responder a consultas de agentes (pessoas ou outros programas) por meio do uso de descrições ontológicas (Horrocks et al., 2003). Atualmente, a OWL é a linguagem mais utilizada para representar ontologias formalmente, possui variantes da linguagem que lidam com a escalabilidade e expressividade das ontologias e permite que aplicações com diferentes propósitos sejam construídas (Mizoguchi, 2004). A Web Ontology Language (OWL) foi projetada para o uso de aplicações que necessitam processar o conteúdo de informação ao invés de apenas apresentar informações aos usuários. OWL agrega uma melhor interpretabilidade de conteúdo Web para as máquinas do que XML, RDF e RDF Schema, provendo vocabulário adicional com uma semântica formal McGuinness et al. (2004).

OWL possui três sublinguagens designadas para o uso específico das comunidades de desenvolvedores e usuários. As três linguagens são (Antoniou and Van Harmelen, 2004):

- OWL DL: A sublinguagem OWL DL (abreviação de *Description Logic*) surgiu para aumentar a eficiência computacional e restringir a forma como construtores de OWL e RDF são utilizados. Ela dá suporte a raciocínio eficiente pois não permite que sejam aplicados construtores OWL uns sobre os outros, garantindo assim que a linguagem corresponda a uma lógica de descrição bem estudada. Um documento RDF geralmente precisa sofrer restrições para ser um documento OWL DL válido e todo documento DL válido também é um documento RDF válido, sendo essa uma desvantagem da utilização da OWL DL.
- OWL Lite: Esta sublinguagem é uma aplicação de restrições a OWL DL, tomando um subconjunto de seus construtores. Ela serve para os usuários que desejam clas-

sificação hierárquica e características simples de limitação. Traz como vantagem o fato de ser uma linguagem tanto simples de compreender como de implementar. Possui expressividade restringida.

- **OWL Full:** É a linguagem que inclui todas as primitivas de OWL. A vantagem é que ela é totalmente compatível com RDF, ou seja, qualquer documento RDF válido é também um documento OWL *Full* válido. Ela é destinada para os usuários que desejam o máximo de expressividade possuindo uma liberdade sintática de RDF. A desvantagem é que é improvável que qualquer software de raciocínio seja capaz de suportar todos os recursos da OWL *Full*.

2.3.3 SPARQL

Como visto anteriormente, RDF é um modelo de dados que permite armazenar informações com conteúdo semântico agregado. Com o seu lançamento, no entanto, surgiu naturalmente o problema de como consultar dados RDF, de forma que aplicações pudessem utilizar esses dados de maneira eficiente. Depois de várias propostas de design e implementação, o W3C adotou a linguagem SPARQL como padrão para recuperação de informações em documentos RDF.

SPARQL é uma linguagem e protocolo para consulta em RDF, baseada em casamento de padrões. De forma análoga a SQL, possui uma estrutura *SELECT-FROM-WHERE* onde (Procópio et al., 2011):

- **SELECT:** Especifica uma projeção sobre os dados como a ordem e a quantidade de atributos e/ou instâncias que serão retornados.
- **FROM:** Declara as fontes que serão consultadas. Esta cláusula é opcional. Quando não especificada, assumimos que a busca será feita em um documento RDF particular.
- **WHERE:** Determina restrições na consulta. Os registros retornados pela consulta deverão satisfazer as restrições impostas por essa cláusula.

Uma consulta SPARQL consiste basicamente de três partes (Pérez et al., 2006):

- Casamento de padrões - Inclui características interessantes de casamento de padrões de grafos, como união de padrões, aninhamento, filtragem e restrição de valores, entre outros.
- Modificadores da solução - Permite modificar os valores da saída do casamento de padrões, aplicando operadores clássicos, como `distinct`, `order`, `limit` e `offset`.
- Saída - A saída de uma consulta SPARQL pode ser de diferentes tipos: consultas sim/não, seleção de valores das variáveis que casaram com os padrões, construção de novas triplas a partir desses valores, e descrições sobre consultas de recursos.

3 Criação da Ontologia de Ecossistemas de Dados Abertos Governamentais

3.1 Metodologia

O processo de criação da ontologia EcOGD dividiu-se essencialmente em 2 fases. A primeira fase foi dedicada ao estudo e especificação de uma estrutura básica para os ecossistemas de dados abertos governamentais, baseando-se nos portais de dados abertos do governo brasileiro, de outros países e na estrutura dos ecossistemas de dados abertos governamentais reais. Foram estudados exemplos de países como: Argentina (Canares and Shekhar, 2015), Brasil (Oliveira et al., 2016) e (Canares and Shekhar, 2015), Filipinas (Canares and Shekhar, 2015), Índia (Chattapadhyay, 2015), Irlanda (Lee, 2014), Estados Unidos (Dawes et al., 2016), Quênia (Canares and Shekhar, 2015), Reino Unido (Heimstädt et al., 2014) e Rússia (Dawes et al., 2016). Como resultado, foi possível extrair informações que ajudaram no mapeamento dos atores envolvidos na publicação e consumo de dados abertos, dos financiadores de iniciativas de dados abertos, dos relacionamentos entre eles, dos tipos de soluções desenvolvidas e quais os papéis exercidos por cada ator.

Na segunda fase, iniciou-se a criação da ontologia seguindo os passos da metodologia 101 (Noy et al., 2001). Durante o processo de criação da ontologia foram definidos os termos, vocabulários a serem reusados, quais termos seriam criados, as propriedades e a ferramenta a ser utilizada, a Protégè¹¹.

¹¹<http://protege.stanford.edu/>

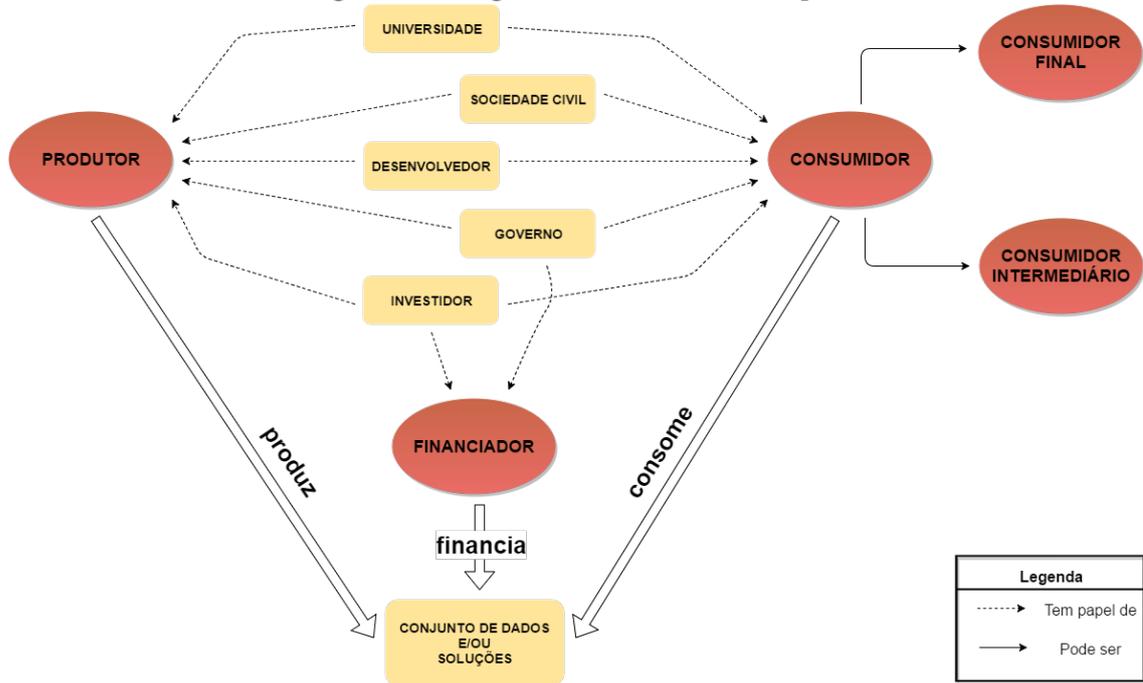
3.2 Mapeamento de atores, papéis e relacionamentos

Os dados abertos governamentais são publicados por órgãos públicos com o intuito de aumentar a transparência e a prestação de contas dos governos. A publicação é feita online por meio dos portais de dados abertos, os quais foram criados para ser um repositório de conjuntos de dados, auxiliar os governos a divulgar dados importantes para os usuários e aumentar a transparência. Desde o surgimento dos primeiros portais, os conjuntos de dados têm sido divulgados através de planilhas e arquivos com informações dispostas em formatos que dificultam a compreensão dos usuários comuns que não têm familiaridade com as tecnologias utilizadas.

Com o estudo das iniciativas de abertura dos dados governamentais e as relações presentes entre os produtores e consumidores dos dados abertos governamentais, aplicou-se o conceito de ecossistema à esfera dos dados abertos governamentais. Os ciclos existentes no ecossistema dependem dos atores, dos papéis exercidos por cada um, como eles se relacionam e quais os produtos oriundos dessas conexões. A depreciação das soluções produzidas utilizando os dados abertos e a falta de um canal de comunicação entre os atores, dificulta a troca de informações entre os atores e o estreitamento de relações entre eles. Além disso, a ausência de informações sobre o funcionamento desse tipo de ecossistema, quais as principais partes interessadas, seus papéis e os relacionamentos existentes, também influencia no grau de satisfação dos produtores e consumidores de soluções que utilizam tais dados. Por isso, a criação da ontologia EcOGD, além de explorar e revelar novos fluxos de comunicação entre os atores, torna possível a troca de informações que sejam relevantes entre eles e ajuda no surgimento de novas parcerias, investimentos, clientes e soluções.

De acordo com o estudo do ecossistema de dados abertos governamentais, foi feito um diagrama que exemplifica de forma ilustrativa os papéis que cada ator exerce e a relação desses papéis com os conjuntos de dados e soluções produzidas:

Figura 1: Diagrama dos Atores e Papéis



Fonte: Próprio autor

Na Figura 1 os papéis são representados pelos círculos e os atores pelos cinco quadrados da parte superior. O quadrado inferior, que tem escrito conjuntos de dados e/ou soluções, é o produto das relações de produção, financiamento e consumo.

Cada papel da Figura 1 tem o seguinte significado:

- Produtor: O papel de produtor produz e disponibiliza conjuntos de dados;
- Financiador: O papel de financiador financia projetos e iniciativas que utilizam os conjuntos de dados abertos para gerar soluções e/ou novos conjuntos de dados;
- Consumidor: O papel de consumidor consome os conjuntos de dados e/ou as soluções por duas perspectivas:
 - Consumidor Final: O papel de consumidor final consome aplicações e soluções que utilizam os dados abertos governamentais;
 - Consumidor Intermediário: O papel de consumidor intermediário consome os dados de forma bruta extraindo-os de portais na Web, por exemplo.

A Tabela 1 resume os papéis que cada ator assume nos ecossistemas de dados abertos governamentais, os quais estão ilustrados na Figura 1.

Tabela 1: Atores e Papéis do Ecossistema

Atores	Produtor	Financiador	Consumidor
Governo	X	X	X
Desenvolvedor	X		X
Universidade	X		X
Sociedade Civil	X		X
Investidor		X	X

O ator governo exerce os papéis de produtor, financiador e consumidor. O governo é o principal produtor dos ecossistemas de dados abertos governamentais pois, composto pelos órgãos e secretarias federais, é a fonte principal de dados abertos governamentais. Ele é considerado financiador pois possui órgãos de incentivo à pesquisa e desenvolvimento que possuem relações com a publicação de dados abertos.

O ator desenvolvedor representa empresas, *startups* e profissionais que desenvolvem soluções e produtos utilizando os conjuntos de dados abertos governamentais.

O ator universidade representa a instituição de ensino superior, assim como centros de pesquisa e desenvolvimento, que desenvolvem projetos, pesquisas e soluções com dados abertos governamentais, além de agregar dados produzindo novos conjuntos de dados.

A sociedade civil representa os cidadãos e as organizações não-governamentais que consomem os dados tanto de forma bruta através de planilhas quanto através de aplicações. A sociedade civil pode ser considerada uma consumidora final e intermediária, respectivamente. Além de consumidora, a sociedade civil é considerada produtora indireta pois faz demandas de conjuntos de dados ao governo.

O ator investidor representa os investidores privados que fomentam as iniciativas que fazem uso de dados abertos governamentais.

Todos os cinco atores são considerados consumidores, pois consomem conjuntos de dados e soluções que utilizam dados abertos governamentais. Esses atores são os primeiros cinco conceitos definidos da ontologia. Abaixo segue uma lista com outros conceitos definidos para a ontologia:

- **Cidadão** - O conceito cidadão representa o cidadão comum que consome soluções de dados abertos, aplicações móveis ou portais, conjuntos de dados em formato bruto e cobram do governo a abertura de novos dados e de novas áreas;

- **Organização Não-governamental** - O conceito de organização não-governamental representa as ONGs existentes nos ecossistemas de dados abertos governamentais que se interessam por dados abertos, utilizam soluções com esses dados e cobram do governo a abertura de novos dados e de novas áreas;
- **Conjuntos de dados** - Conjuntos de dados são publicados nos portais de dados abertos ou por alguns atores, e consumidos por todos os atores do ecossistema através de soluções de dados abertos;
- **Soluções de Dados Abertos Governamentais** - Soluções de dados abertos representam os produtos criados a partir do uso dos dados abertos. Essas soluções podem ser aplicações móveis e portais de dados abertos que utilizam os dados abertos governamentais;
- **Departamento** - Departamento representa os setores da universidade que reúnem profissionais da mesma área.
- **Grupo de Pesquisa** - Grupo de pesquisa representa um grupo composto por estudantes, professores e pesquisadores com o intuito de promover os dados abertos governamentais através de pesquisas e geração de soluções;
- **Pesquisador** - Pesquisador representa o membro do grupo de pesquisa, o qual pode ser estudante ou professor.
- **Projeto** - Projeto representa o trabalho que é feito pelos pesquisadores e que recebe investimento de investidores.

Além dos atores e papéis, a Figura 1 mostra os relacionamentos de produção, de financiamento e de consumo, os quais estão representados pelas setas grossas com os nomes produz, financia e consome, respectivamente. Esses são os relacionamentos-base dos ecossistemas de dados abertos governamentais, pois são eles que ligam os atores aos conjuntos de dados e às soluções.

Analisando os relacionamentos existentes entre os atores e possíveis novos relacionamentos que podem ser inferidos com a criação da ontologia, detectou-se alguns problemas:

- Falta de comunicação entre o governo e desenvolvedores que têm interesse em desenvolver soluções utilizando os dados abertos do governo;

- Falta de informação sobre quais produtos e quais empresas estão trabalhando e desenvolvendo aplicações com esses dados;
- Governo não tem um canal de comunicação direto com a sociedade civil para saber quais domínios ela tem interesse em ter acesso;
- Falta de comunicação entre os desenvolvedores e o governo para fazer demandas de conjuntos de dados;
- Falta de comunicação entre os desenvolvedores e os investidores;
- Falta de informação sobre investidores que desejam financiar projetos de pesquisa e desenvolvimento;
- Falta de comunicação entre a universidade e empresas sobre parcerias para colaboração;
- Falta de divulgação sobre as soluções e aplicações que estão sendo produzidas e os domínios;
- Falta de informação sobre quais grupos de pesquisa ou empresas estão procurando por investimento;

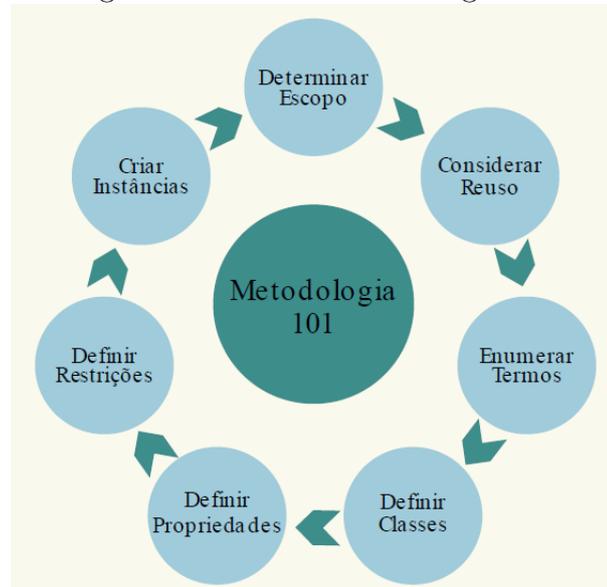
Como resultado do mapeamento, pode-se concluir que a criação da ontologia do ecossistema de dados abertos governamentais trará novas informações para os atores, aprimorando assim a comunicação entre as partes interessadas para a produção de novas soluções. Com o estreitamento das relações entre as partes interessadas, o conhecimento sobre as áreas que estão sendo mais requisitadas, novas parcerias sendo construídas, as soluções que serão produzidas serão mais sustentáveis, duradouras e os problemas supracitados serão sanados.

3.3 Criação da ontologia EcOGD

Após a especificação inicial dos ecossistemas de dados abertos governamentais, optou-se pelo desenvolvimento de uma especificação mais formal para o ecossistema. Desta forma, decidiu-se pela criação de uma ontologia, denominada EcOGD, para a representação dos principais conceitos dos ecossistemas de dados abertos governamentais e seus

relacionamentos. A metodologia 101, proposta por Noy et al. (2001), foi escolhida para auxiliar na construção da ontologia por ser composta por passos objetivos e simples. Ela é composta por 7 passos, os quais estão representados na Figura 2.

Figura 2: Ciclo da Metodologia 101



Fonte: Isotani, S., & Bittencourt, I. I. (2015). Dados Abertos Conectados. Novatec Editora.

O primeiro passo é a definição do domínio e do escopo da ontologia.

As seguintes perguntas ajudam a definir o domínio (Noy et al., 2001):

- Qual o domínio que a ontologia deve cobrir?
- Para que nós usaremos a ontologia?
- Para que tipos de perguntas as informações na ontologia gerarão respostas?
- Quem usará e manterá a ontologia?

O domínio da ontologia se concentra no ecossistema de dados abertos governamentais, mais especificamente como os atores do ecossistema se relacionam para consumir e gerar soluções. Essa ontologia tem como objetivo otimizar a criação e uso de aplicações a partir do estreitamento da comunicação e troca de informações entre os atores identificados anteriormente. Por meio da ontologia serão geradas informações sobre potenciais relacionamentos entre os atores, conexões que facilitarão a criação de soluções mais eficazes

e, como consequência, aumentará o consumo e publicação dos conjuntos de dados abertos governamentais.

Um esboço de questões que a base de conhecimento da ontologia deve ser capaz de responder é composto por questões de competência, do inglês *competency questions*, que segundo Grüninger and Fox (1995) servem para determinar o escopo. As questões de competência se baseiam em projeção temporal do futuro, planejamento, monitoramento e eventos externos, por exemplo. A partir da leitura de exemplos de questões de competência dos artigos de Gruber (1993) e Noy et al. (2001), questões de competência referente a esse domínio foram formuladas e respondidas com o intuito de definir o escopo:

- Quais domínios de conjuntos de dados a sociedade civil está interessada em ter acesso?
- Quais projetos de pesquisa que trabalham com dados abertos governamentais estão sendo financiados por órgãos privados e quais os órgãos?
- Quais aplicações foram desenvolvidas por *startups* e em quais áreas?
- Quais os portais de dados abertos governamentais do Brasil?
- Quais universidades tem projetos com dados abertos governamentais na área da saúde?

Para a definição do domínio do ecossistema avaliou-se também quais conceitos e dados, referentes ao ecossistema de dados abertos governamentais, foram levantados durante o mapeamento e quais elementos podem fornecer conteúdo para que aplicações possam extrair informações relevantes da base de conhecimento.

O primeiro passo para a criação do esquema foi avaliar quais conceitos, dados e metadados referentes aos ecossistemas de dados abertos governamentais são importantes para uma modelagem coerente do domínio. Os conceitos utilizados para a criação da ontologia EcOGD foram definidos no capítulo 3.2 e são: governo, desenvolvedor, universidade, sociedade civil, investidor, cidadão, organização não-governamental, conjunto de dados, soluções de dados abertos governamentais, portais de dados abertos governamentais, aplicações móveis, departamento, grupo de pesquisa, pesquisador e projeto.

Os passos seguintes da metodologia 101 foram executados nas próximas seções.

3.3.1 Reuso de vocabulários

O reuso de termos de vocabulários é muito recomendado para uma aplicação que segue os princípios de dados ligados (Isotani and Bittencourt, 2015), pois facilita a integração de dados de diferentes fontes, facilitando o processamento desses dados. Caso nenhum vocabulário consiga descrever os dados, é necessário a criação de novos termos que correspondam ao que os dados expressam (Noy et al., 2001).

Durante o processo de pesquisa por termos que expressassem o sentido dos dados, foram encontrados quatro vocabulários que tiveram seus termos reusados, o FOAF (*Friend of a Friend*)¹², o VoID (*Vocabulary of Interlinked Datasets*)¹³, o Schema.org (*Schema.org Vocabulary*)¹⁴ e o DBPedia (*DBPedia Ontology*)¹⁵.

O projeto *Friend of a Friend* (FOAF) é voltado para unir pessoas e informações utilizando a Web. Ele integra três tipos de rede: redes sociais de colaboração humana, amizades e associações, redes representativas que descrevem numa visão simplificada do universo de desenho animado em termos factuais, e redes de informações que usam links baseados na Web para compartilhar descrições publicadas independentemente do mundo interconectado (Brickley and Miller, 2012). Ele é largamente usado para representar redes sociais entre pessoas e muitos websites de redes sociais o utilizam para produzir perfis de seus usuários na Web Semântica (Golbeck and Rothstein, 2008).

O VoID é um vocabulário de conjuntos de dados interligados, do inglês *Vocabulary of Interlinked Datasets*, e um conjunto de instruções que facilita a descoberta e uso de conjuntos de dados ligados (Alexander et al., 2009). Ele é um vocabulário em RDF *Schema* para expressar metadados sobre conjuntos de dados RDF. Tem o objetivo de ser uma ponte entre os publicadores e consumidores de dados em RDF por meio de aplicações que vão desde a descoberta de dados para catalogação até o arquivamento de conjuntos de dados (Alexander et al., 2011).

O vocabulário Schema.org pode ser usado com muitas codificações diferentes, incluindo RDFa, Microdata e JSON-LD. Estes vocabulários abrangem entidades, relações entre entidades e ações, e podem ser facilmente estendidos através de um modelo de extensão bem documentado. Mais de 10 milhões de sites usam Schema.org para marcar suas páginas da

¹²<http://xmlns.com/foaf/0.1/>

¹³<http://rdfs.org/ns/void>

¹⁴<http://schema.org/>

¹⁵<http://dbpedia.org/ontology/>

Web e mensagens de e-mail. Muitos aplicativos do Google, Microsoft, Pinterest, Yandex e outros já usam esses vocabulários para poder experiências ricas e extensíveis.

A ontologia DBpedia é uma ontologia de vários domínios, que foi criada com base nas *infoboxes* mais usadas na *Wikipedia*. A ontologia abrange atualmente 685 classes as quais são descritas por 2.795 propriedades diferentes.

Na Tabela 2 pode-se encontrar os termos reutilizados para as classes da ontologia Ececgd (*Ecosystem of Open Government Data*).

Tabela 2: Termos Reusados para Classes

Conceito	Termo Reusado
Órgão Governamental	schema:GovernmentOrganization
Universidade	schema:CollegeOrUniversity
Organização Não-governamental	dbo:Non-ProfitOrganisation
Projeto	dbo:project
Departamento	dbo:department
Desenvolvedor	dbo:developer
Pessoa	foaf:Person
Conjunto de Dados	void:Dataset

Além das classes, várias propriedades foram reaproveitadas. Na Tabela 3 pode-se encontrar os termos reusados para *Datatype Properties*. É importante notar que não foram reusados termos para *Object Properties*.

Tabela 3: Termos Reusados para *Datatype Properties*

Termo Reusado	Domain	Range
schema:description	void:Dataset	String
dbo:numberOfMembers	ecogd:ResearchGroup	Integer

Termo Reusado	Domain	Range
dbo:zipCode	schema:CollegeOrUniversity ecogd:Investor schema:GovernmentOrganization foaf:Person dbo:Non-ProfitOrganisation dbo:developer	String
schema:addressCountry	schema:CollegeOrUniversity ecogd:Investor schema:GovernmentOrganization foaf:Person dbo:Non-ProfitOrganisation dbo:developer	String
schema:streetAddress	schema:CollegeOrUniversity ecogd:Investor schema:GovernmentOrganization foaf:Person dbo:Non-ProfitOrganisation dbo:developer	String
foaf:name	foaf:Person ecogd:ResearchGroup dbo:developer dbo:Non-ProfitOrganization dbo:department dbo:project ecogd:Application void:Dataset schema:GovernmentOrganization ecogd:OpenDataPortal ecogd:Researcher ecogd:Investor schema:CollegeOrUniversity	String

Termo Reusado	Domain	Range
schema:telephone	foaf:Person ecogd:Investor dbo:department dbo:developer dbo:Non-ProfitOrganization schema:GovernmentOrganization	String
schema:email	foaf:Person ecogd:Investor dbo:department dbo:developer dbo:Non-ProfitOrganization schema:GovernmentOrganization ecogd:Researcher	String

3.3.2 Criação de novos termos

Apesar dos vocabulários existentes cobrirem uma ampla parte dos termos necessários para a criação de uma ontologia, ainda foi necessário definir alguns novos termos, já que em alguns casos não foi encontrado nenhum termo existente refletindo a semântica do conceito que se desejava representar.

Desta forma, um conjunto de termos foi criado com o objetivo de complementar o esquema da ontologia. O *namespace* deste novo vocabulário é o ecogd, que possui a IRI <http://www.semanticweb.org/ecogd>. Ao invés de criar termos para todo o domínio em questão, nesta fase o foco foi na criação apenas dos termos essenciais ao esquema projetado.

A ferramenta utilizada para a criação do vocabulário foi o Protégé. O Protégé é uma ferramenta que permite construir ontologias, personalizar formulários de entrada de dados, inserir e editar dados, possibilitando, então, a criação de bases de conhecimento guiadas por uma ontologia. Sua interface gráfica provê acesso à barra de menus e à barra de ferramentas, além de apresentar cinco áreas de visualização (views) que funcionam como módulos de navegação e edição de classes, atributos, formulários, instâncias e pesquisas na base de conhecimento, propiciando a entrada de dados e a recuperação das informações

(Semprebom et al., 2007).

A Tabela 4 exibe as classes criadas. As Tabelas 5 e 6 contêm as propriedades criadas, tendo a primeira as *Datatype Properties* e a segunda as *Object Properties*.

Tabela 4: Termos Criados para Classes

Classe	Descrição	Mapeamento com outros vocabulários
ecogd:Investor	Investidor	Não houve
ecogd:OpenDataSolutions	Soluções de dados abertos	Não houve
ecogd:Application	Aplicação	Sub-classe de ecogd:OpenDataSolutions
ecogd:OpenDataPortal	Portal de dados abertos	Sub-classe de ecogd:OpenDataSolutions
ecogd:Researcher	Pesquisador	Não houve
ecogd:CivilSociety	Sociedade Civil	Não houve
ecogd:ResearchGroup	Grupo de Pesquisa	Não houve

Tabela 5: Termos Criados para *Object Properties*

Propriedade	Domain	Range
ecogd:publishes	schema:Government-Organisation dbo:developer schema:CollegeOrUniversity ecogd:CivilSociety	void:Dataset
ecogd:consumes	ecogd:CivilSociety dbo:developer schema:Government-Organization schema:CollegeOrUniversity ecogd:ResearchGroup ecogd:Investor	ecogd:OpenDataSolutions void:Dataset

Propriedade	Domain	Range
ecogd:develops	dbo:developer ecogd:CivilSociety ecogd:ResearchGroup	ogb:OpenDataSolutions
ecogd:requests	ecogd:CivilSociety dbo:developer ecogd:ResearchGroup ecogd:Investor	void:Dataset
ecogd:worksFor	foaf:Person	schema:Government-Organization dbo:developer
ecogd:isMemberOf	ecogd:Researcher	ecogd:ResearchGroup
ecogd:has	ecogd:ResearchGroup schema:CollegerOrUniversity dbo:department dbo:developer	dbo:project dbo:department ecogd:ResearchGroup ecogd:Researcher
ecogd:sponsors	ecogd:Investor schema:Government-Organization	dbo:project
ecogd:isHeldBy	dbo:project	ecogd:ResearchGroup dbo:developer

Tabela 6: Termos Criados para *Datatype Properties*

Propriedade	Domain	Range
ecogd:companyType	dbo:developer	String
ecogd:sourceOfInvestment	ecogd:Investor	String
ecogd:position	ecogd:Researcher	String

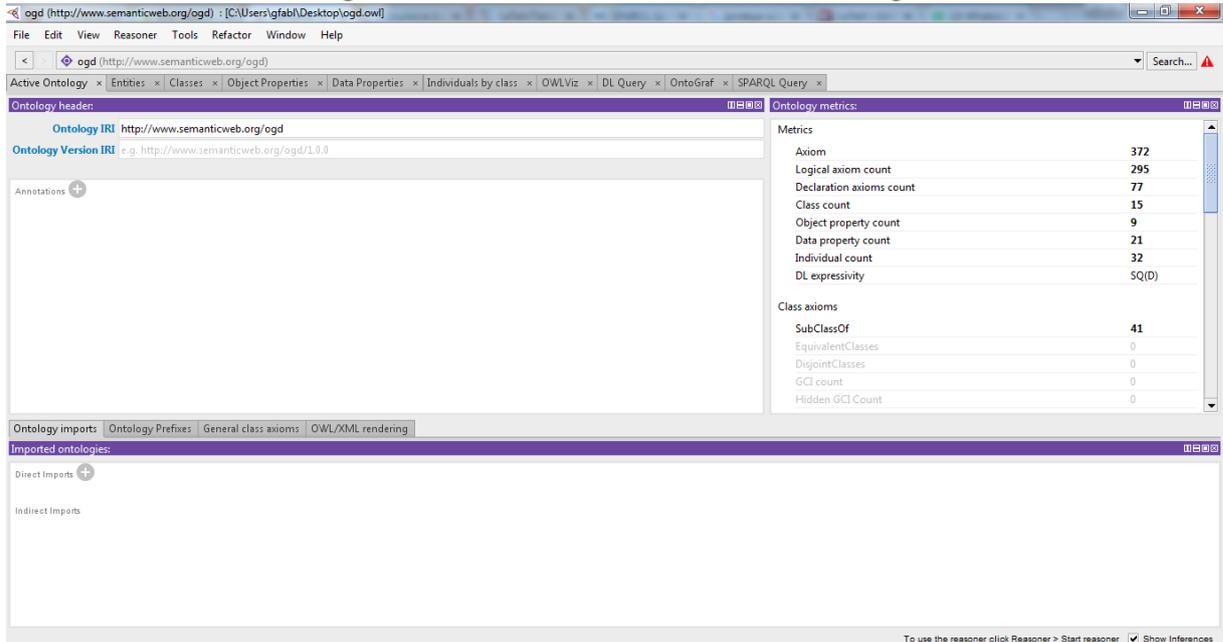
Propriedade	Domain	Range
ecogd:state	schema:CollegeOrUniversity ecogd:Investor schema:Government-Organization foaf:Person dbo:Non-ProfitOrganisation dbo:developer	String
ecogd:city	schema:CollegeOrUniversity ecogd:Investor schema:Government-Organization foaf:Person dbo:Non-ProfitOrganisation dbo:developer	String
ecogd:webAddress	schema:CollegeOrUniversity ecogd:Investor schema:Government-Organization dbo:Non-ProfitOrganisation dbo:developer ecogd:OpenDataPortal ecogd:ResearchGroup dbo:project ecogd:Application dbo:department	String

Propriedade	Domain	Range
ecogd:CNPJ	dbo:Non-ProfitOrganisation ecogd:Investor dbo:developer	String
ecogd:personalID	foaf:Person	String
ecogd:acronym	ecogd:Investor schema:CollegeOrUniversity dbo:developer	String
ecogd:areaOfExpertise	schema:Government-Organization	String
ecogd:areaOfInterest	ecogd:Investor ecogd:ResearchGroup dbo:developer foaf:Person dbo:Non-ProfitOrganisation	String
ecogd:companyTypeToInvest	ecogd:Investor	String

3.3.3 Ferramenta Protègè

Para criar a ontologia do ecossistema de dados abertos governamentais fez-se uso da ferramenta Protègè. Na Figura 3 é mostrada a tela inicial dela. Essa ferramenta possibilitou de forma simplificada o reuso e a criação das classes, das *datatype properties* e das *object properties*. Além disso, possibilitou a instanciação manual de cada indivíduo para que consultas pudessem ser feitas.

Figura 3: Tela inicial da ferramenta Protégè



Abaixo estão as interfaces do Protégè para a criação das classes, das *datatype properties* e das *object properties*, nas figuras 4, 5 e 6, respectivamente.

Figura 4: Tela de criação das classes da ferramenta Protégè

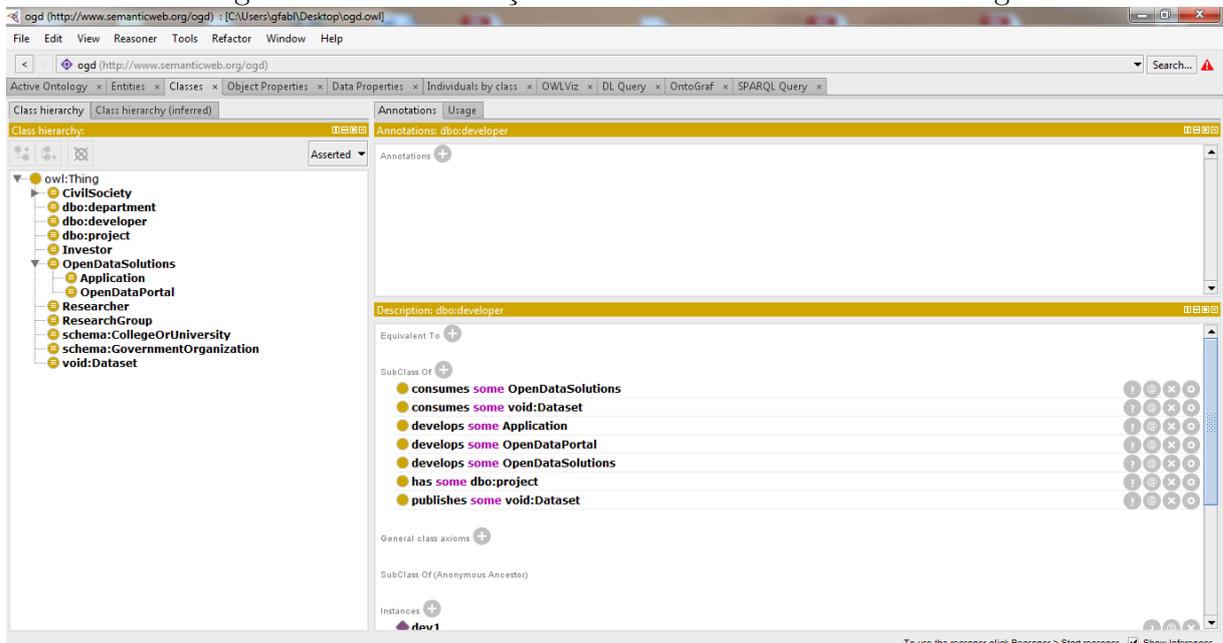


Figura 5: Tela de criação das *datatype properties* da ferramenta Protégè

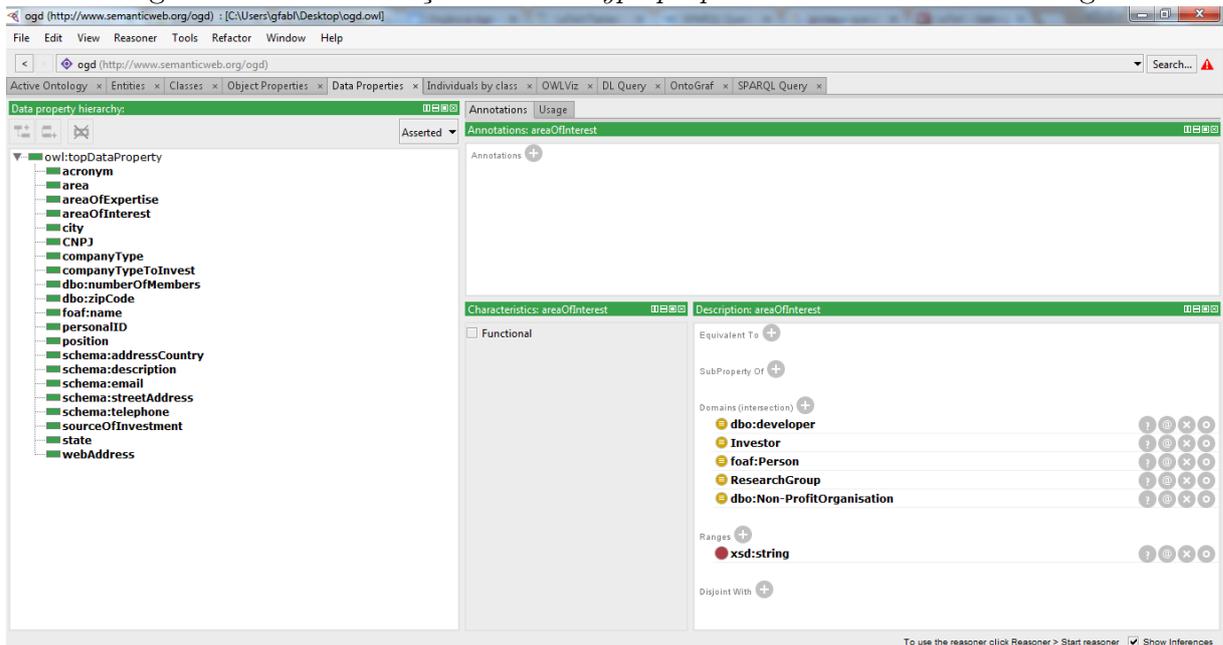
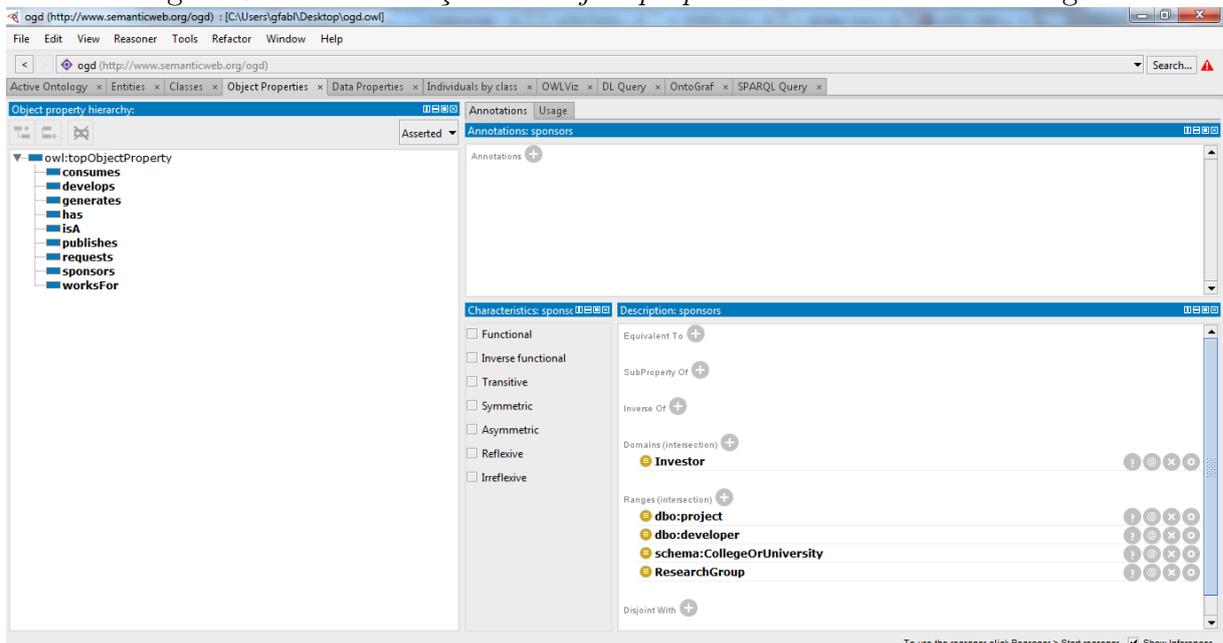


Figura 6: Tela de criação das *object properties* da ferramenta Protégè



A ferramenta Protégè permite a visualização da hierarquia de classes através do plugin OWLViz¹⁶. A hierarquia de classes pode ser visualizada na Figura 7.

¹⁶<http://protegewiki.stanford.edu/wiki/OWLViz>

Figura 7: Hierarquia de classes



Com a criação destes termos, o esquema da ontologia foi finalizado.

4 Criação da base de conhecimentos

Após a criação da ontologia, fez-se possível a criação de uma base de conhecimentos que possui instâncias referentes a cada conceito relacionado ao ecossistema de dados abertos governamentais da cidade do Recife. Na base de conhecimentos, pode-se realizar consultas para a extração de relatórios e de novas informações que otimizem as relações entre os atores e os processos de publicação e consumo nesse ecossistema. A base explora e revela novos fluxos de comunicação entre os atores e ajuda no surgimento de novas parcerias, investimentos, clientes e soluções.

A ferramenta Protégè dá suporte à instanciação dos indivíduos da ontologia. O Protégè possui uma interface intuitiva e permite a criação de indivíduos de forma simples. Primeiramente, seleciona-se a classe que se deseja instanciar, cria-se o indivíduo, nomeando-o, e depois adiciona-se as propriedades do tipo *datatype* e as *object*, as quais já devem ter sido criadas e especificadas quanto ao *domain* e ao *range*.

Pode-se ver na Figura 8 a interface para criação das intâncias e na Figura 9 a visualização de todas as instâncias da ontologia.

Figura 8: Interface do Protégè para instanciação

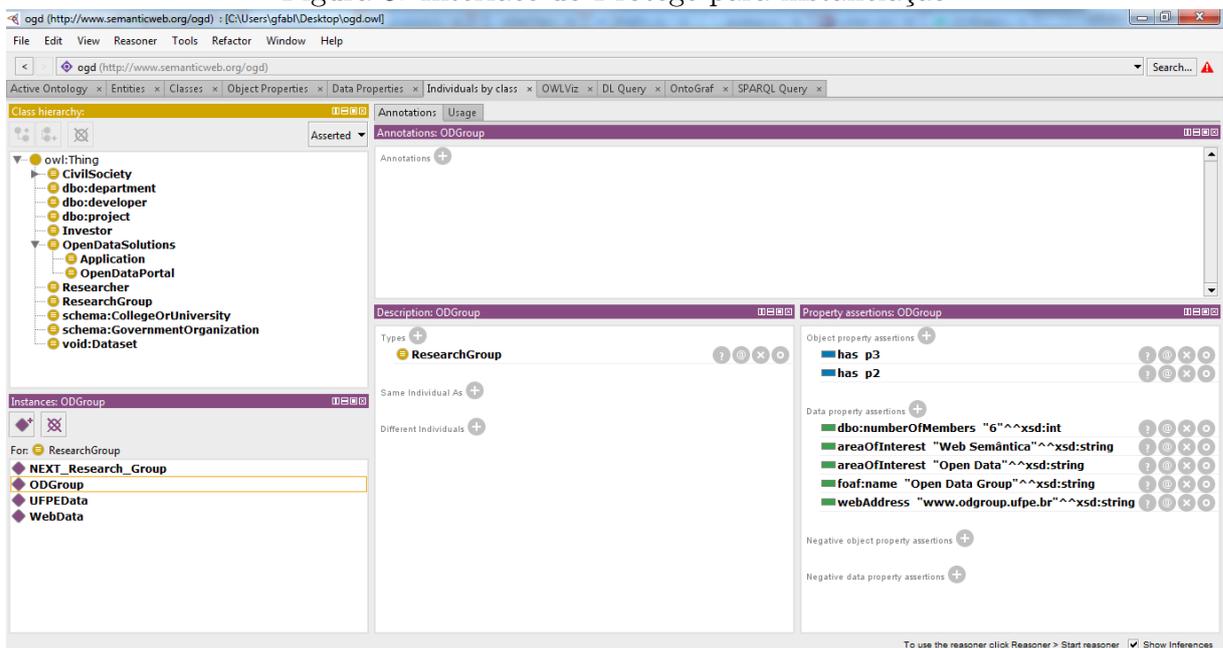
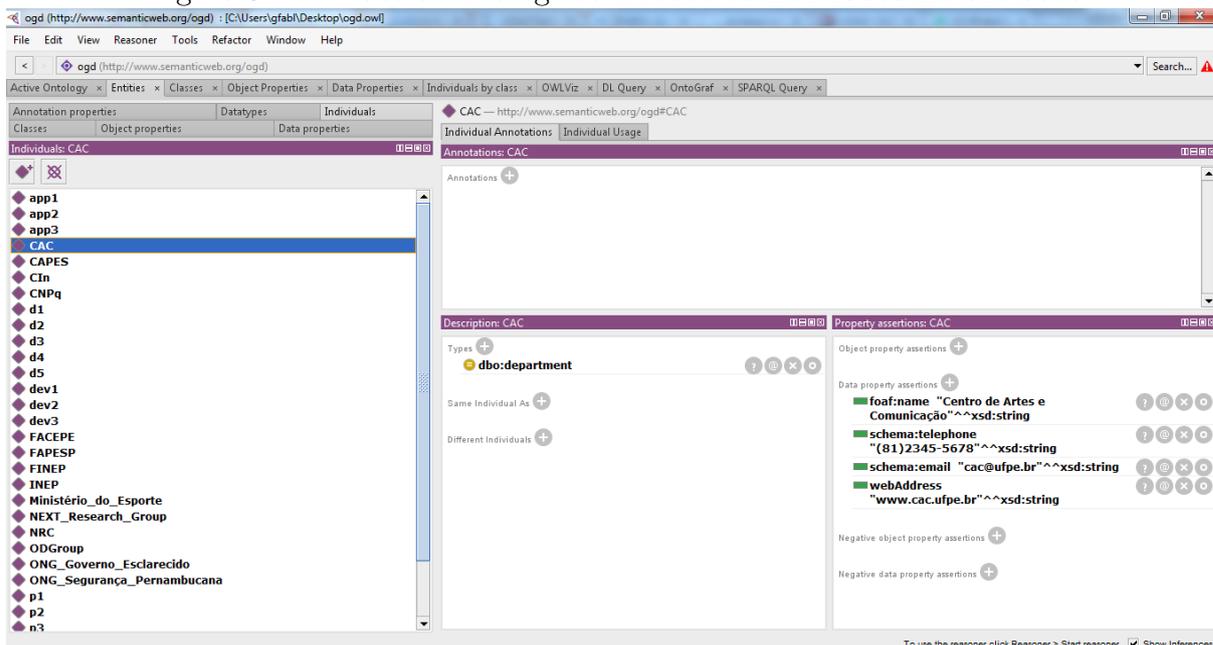
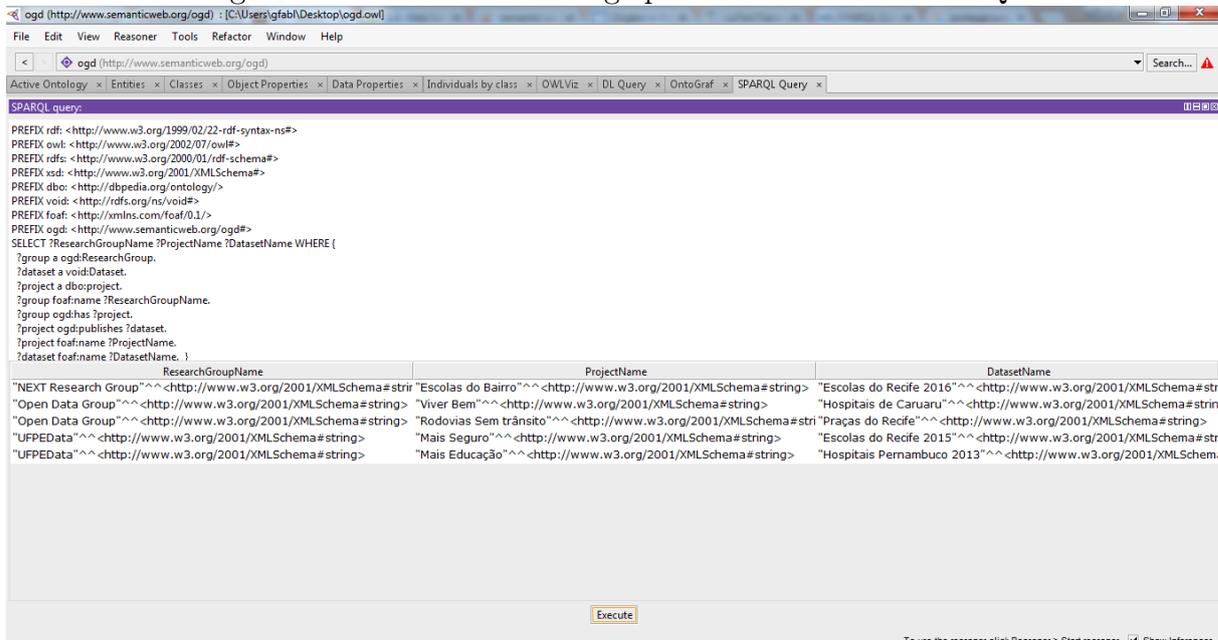


Figura 9: Interface do Protégè com todos os indivíduos instanciados



A ferramenta Protégè possui uma interface para consultas SPARQL. Nessa interface é preciso inserir os prefixos e as URIs dos vocabulários que estão sendo reutilizados e o prefixo e a URI do vocabulário que foi criado na ferramenta.

Figura 10: Interface do Protégè para consultas em SPARQL



A partir dessa interface foram realizadas consultas explorando as classes e propriedades definidas no mapeamento. A Tabela 7 mostra os prefixos utilizados na composição das consultas.

Tabela 7: Prefixos dos vocabulários utilizados nas consultas SPARQL

Prefixo	Vocabulário
foaf	http://xmlns.com/foaf/0.1/
dbo	http://dbpedia.org/ontology/
void	http://rdfs.org/ns/void#
schema	http://schema.org/
ecogd	http://www.semanticweb.org/ecogd#

Com a base de conhecimentos com informações sobre o ecossistema de dados abertos governamentais da cidade do Recife, é possível extrair informações sobre: quais áreas de conjuntos de dados a sociedade tem interesse em ter acesso, quais projetos desenvolvidos em universidades produzem aplicativos móveis que utilizam os dados abertos governamentais e em quais áreas, lista de portais de dados abertos governamentais de um país, estado ou cidade, quais áreas os investidores têm interesse em investir, quais investidores estão investindo atualmente em projetos de áreas específicas, quais os conjuntos de dados que foram publicados por um grupo de pesquisa em um ano específico, quais empresas estão desenvolvendo soluções de dados abertos governamentais em uma área específica, entre outros exemplos. Com a base de conhecimentos é possível realizar diversas consultas e explorar todas os conceitos e propriedades da ontologia.

Seguem abaixo algumas consultas realizadas e os respectivos resultados:

Consulta 1: Quais órgãos de investimentos estão interessados em investir em educação e quais projetos estão trabalhando nessa área?

```

SELECT ?Investor ?Project_Name WHERE {
  ?investor a ecogd:Investor.
  ?investor ecogd:acronym ?Investor.
  ?investor ecogd:areaOfInterest "Educação"^^xsd:string.
  ?project a dbo:project.
  ?project foaf:name ?Project_Name.
  ?project ecogd:areaOfInterest "Saúde"^^xsd:string.
}

```

A tabela 8 ilustra o resultado deste consulta.

Tabela 8: Resultado da consulta 1

Investidores	Projetos
FACEPE	Data on the Web: Vantagens e Desafios
CNPq	Open Data Ecosystem
FAPESP	Escolas Brasileiras e o uso de aplicativos
FINEP	Portais de dados abertos e suas limitações
NRC	Datasets e os cidadãos

Consulta 2: Lista dos grupos de pesquisas e os nomes dos datasets publicados pelos projetos desenvolvidos por eles.

```

SELECT ?ResearchGroupName ?ProjectName ?DatasetName WHERE {
  ?group a ogd:ResearchGroup.
  ?dataset a void:Dataset.
  ?project a dbo:project.
  ?group foaf:name ?ResearchGroupName.
  ?group ecogd:has ?project.
  ?project ecogd:publishes ?dataset.
  ?project foaf:name ?ProjectName.
  ?dataset foaf:name ?DatasetName. }
GROUP BY ?ResearchGroupName ?ProjectName ?DatasetName
ORDER BY ASC(?ResearchGroupName)

```

A tabela 9 exhibe o resultado da consulta 2.

Tabela 9: Resultado da consulta 2

Grupo de Pesquisa	Conjunto de dado
NEXT	Escolas do Recife 2016
ODGroup	Hospitais de Caruaru
UFPEData	Praças do Recife 2013
WebData	Escolas do Recife 2015
WebData	Hospitais Pernambuco 2015

5 Conclusão

A Web Semântica tem o poder para evoluir a Web atual, oferecendo mecanismos e ferramentas para armazenar e recuperar informações que são extremamente eficientes, além de oferecer maior capacidade de inferência. Para que tudo isso ocorra, é necessário que os dados estejam organizados através de formatos bem definidos, como RDF, e OWL, permitindo que os dados sejam manipulados de forma eficaz.

Como resultado deste trabalho, a criação da ontologia de dados abertos governamentais (EcoGD) se tornou possível após o estudo e mapeamento de todas as partes envolvidas, assim como a identificação dos relacionamentos existentes e quais deficiências o uso dessa ontologia pode sanar. O maior desafio deste trabalho se deu ao tentar recolher dados de todos os participantes do ecossistema, pois não há muitos dados disponíveis online detalhando características dos projetos que estão sendo desenvolvidos, assim como empresas que fazem soluções de dados abertos, órgãos de fomento, áreas de interesse, por exemplo.

Além da criação da ontologia, a criação da base de conhecimentos do ecossistema de dados abertos governamentais da cidade do Recife possibilitou a realização de consultas para a extração de relatórios e de novas informações que otimizem as relações entre os atores e os processos de publicação e consumo no ecossistema de Recife. A base explora e revela novos fluxos de comunicação entre os atores e ajuda no surgimento de novas parcerias, investimentos, clientes e soluções.

É perceptível o poder que a Web terá se todos adotarem as técnicas e padrões propostos pela Web Semântica, como foi usado neste trabalho. Vários outros trabalhos podem ser

originados a partir deste, tanto para aperfeiçoar como para expandir. Aperfeiçoar com a extração de dados de bases de dados abertos que contém informações relacionadas com a publicação de soluções abertas, governo, ONGs, órgãos de investimento e desenvolvedores interessados no aumento do uso dos dados abertos. E para a expansão deste trabalho, criar visualizações mais elaboradas do *dataset* e inserir novos conceitos que possam vir a integrar o contexto do ecossistema de dados abertos governamentais.

Referências

- Agune, R. M. and Bolliger, S. P. (2010). Filho. *Álvaro Santos Gregório*.
- Alexander, K., Cyganiak, R., Hausenblas, M., and Zhao, J. (2009). Describing linked datasets. In *LDOW*.
- Alexander, K., Cyganiak, R., Hausenblas, M., and Zhao, J. (2011). Describing linked datasets with the void vocabulary.
- Antoniou, G. and Van Harmelen, F. (2004). Web ontology language: Owl. In *Handbook on ontologies*, pages 67–92. Springer.
- Berners-Lee, T., Hendler, J., Lassila, O., et al. (2001). The semantic web. *Scientific american*, 284(5):28–37.
- Brickley, D. and Miller, L. (2012). Foaf vocabulary specification 0.98. *Namespace document*, 9.
- Canares, M. and Shekhar, S. (2015). Creating an ecosystem for open sub-national governments: Lessons from a developing country perspective. In *Open Data Research Symposium*.
- Chattapadhyay, S. (2015). Whose open data community is it? reflections on the open data ecosystem in india. In *Open Data Research Symposium*.
- Dawes, S. S., Vidiasova, L., and Parkhimovich, O. (2016). Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly*, 33(1):15–27.
- Dietrich, D., Gray, J., McNamara, T., Poikola, A., Pollock, P., Tait, J., and Zijlstra, T. (2009). Open data handbook.
- Gama, K. and Lóscio, B. F. (2014). Towards ecosystems based on open data as a service. In *ICEIS (2)*, pages 659–664.

- Golbeck, J. and Rothstein, M. (2008). Linking social networks on the web with foaf: A semantic web case study. In *AAAI*, volume 8, pages 1138–1143.
- Gruber, T. (1993). What is an ontology. *WWW Site* <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html> (accessed on 07-09-2004).
- Grüninger, M. and Fox, M. S. (1995). Methodology for the design and evaluation of ontologies.
- Guarino, N. (1997). Understanding, building and using ontologies. *International Journal of Human-Computer Studies*, 46(2):293–310.
- Harrison, T. M., Pardo, T. A., and Cook, M. (2012). Creating open government ecosystems: A research and development agenda. *Future Internet*, 4(4):900–928.
- Heimstädt, M., Saunderson, F., and Heath, T. (2014). Conceptualizing open data ecosystems: A timeline analysis of open data development in the uk. In *Conference for E-Democracy and Open Governement*, page 245.
- Horrocks, I., Patel-Schneider, P. F., and Van Harmelen, F. (2003). From shiq and rdf to owl: The making of a web ontology language. *Web semantics: science, services and agents on the World Wide Web*, 1(1):7–26.
- Isotani, S. and Bittencourt, I. I. (2015). *Dados Abertos Conectados*. Novatec Editora.
- Kiryakov, A. (2006). Ontologies for knowledge management. *Semantic Web Technologies: trends and research in ontology-based systems*, pages 115–138.
- Klyne, G. and Carroll, J. J. (2006). Resource description framework (rdf): Concepts and abstract syntax.
- Laufer, C. (2015). Guia da web semântica. <http://ceweb.br/guias/web-semantica/>. Acesso em: 15-11-2016.
- Lee, D. (2014). Building an open data ecosystem: an irish experience. In *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance*, pages 351–360. ACM.
- Lóscio, B. F., O. M. I. S. B. I. I. (2015). Publicação e consumo de dados na web: Conceitos e desafios. *Dados na Web, Mini Curso SBBDD 2015*.

- Manyika, J., Chui, M., Groves, P., Farrell, D., Van Kuiken, S., and Doshi, E. A. (2013). Open data: Unlocking innovation and performance with liquid information. *McKinsey Global Institute*, page 21.
- McGuinness, D. L., Van Harmelen, F., et al. (2004). Owl web ontology language overview. *W3C recommendation*, 10(10):2004.
- Miller, E. (1998). An introduction to the resource description framework. *Bulletin of the American Society for Information Science and Technology*, 25(1):15–19.
- Mizoguchi, R. (2004). Part 3: Advanced course of ontological engineering. *New Generation Computing*, 22(2):193–220.
- Noy, N. F., McGuinness, D. L., et al. (2001). Ontology development 101: A guide to creating your first ontology.
- Oliveira, M. I. S., de Oliveira, H. R., Oliveira, L. A., and Lóscio, B. F. (2016). Open government data portals analysis: The brazilian case. In *Proceedings of the 17th International Digital Government Research Conference on Digital Government Research*, pages 415–424. ACM.
- Pérez, J., Arenas, M., and Gutierrez, C. (2006). Semantics and complexity of sparql. In *International semantic web conference*, pages 30–43. Springer.
- Pollock, R. (2011). Building the (open) data ecosystem. *Open Knowledge Foundation Blog*, 31.
- Procópio, P., Laender, A. H., and Moro, M. M. (2011). Análise da rede de coautoria do simpósio brasileiro de bancos de dados. *SIMPÓSIO BRASILEIRO DE BANCO DE DADOS, Florianópolis, 2011. Proceedings... Florianópolis*.
- Semprebom, T., Camada, M. Y., and Mendonça, I. (2007). Ontologias e protégé. *Documento extraído em*, 21(07):2007.
- Vaz, J. C., Ribeiro, M. M., and Matheus, R. (2010). Dados governamentais abertos e seus impactos sobre os conceitos e práticas de transparência no brasil. *Cadernos ppg-au/ufba*, 9(1).
- Wagner Filho, F. and Lóscio, B. F. (2010). Web semântica: Conceitos e tecnologias.
- Zuiderwijk, A., Janssen, M., and Davis, C. (2014). Innovation with open data: Essential elements of open data ecosystems. *Information Polity*, 19(1, 2):17–33.