

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO
MESTRADO EM CIÊNCIA DA INFORMAÇÃO
UNIVERSIDADE FEDERAL FLUMINENSE
INSTITUTO DE ARTE E COMUNICAÇÃO SOCIAL

LUCAS DE LIMA ROCHA

**ONTOLOGIAS COMO UM AGENTE PARA A INTERLIGAÇÃO DE DADOS
ABERTOS EM CONTEXTO GOVERNAMENTAL: características e
aplicabilidade**

NITERÓI
2016



LUCAS DE LIMA ROCHA

**ONTOLOGIAS COMO UM AGENTE PARA A INTERLIGAÇÃO DE DADOS
ABERTOS EM CONTEXTO GOVERNAMENTAL: características e
aplicabilidade**

Dissertação apresentada ao curso de Mestrado em Ciência da Informação do Programa de Pós-Graduação da Universidade Federal Fluminense, como requisito parcial para obtenção do Grau de Mestre em Ciência da Informação.

Linha de pesquisa: Fluxos e mediações sócio técnicas da informação.

Orientadora: Prof^a. Dr^a. Maria Luiza de Almeida Campos
Co-orientador: Prof. Dr. Leonardo Cruz da Costa

Niterói
2016

R672o Rocha, Lucas de Lima

Ontologias como um agente para a interligação de dados abertos em contexto governamental: características e aplicabilidade / Lucas de Lima Rocha – 2016.

137 f.: il.

Orientadora: Maria Luiza de Almeida Campos.

Co-orientador: Leonardo Cruz da Costa

Dissertação (Mestrado) – Universidade Federal Fluminense, Departamento de Ciência da Informação.

1. Ciência da Informação. 2. Ontologia. 3. Dados Governamentais Interligados e Abertos. 4. Representação do Conhecimento. 5. Web Semântica. 6. Lei de Acesso à Informação. I. Campos, Maria Luiza de Almeida. II. Costa, Leonardo Cruz da. III. Universidade Federal Fluminense. Departamento de Ciência da Informação. IV. Título.

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO
MESTRADO EM CIÊNCIA DA INFORMAÇÃO
UNIVERSIDADE FEDERAL FLUMINENSE
INSTITUTO DE ARTE E COMUNICAÇÃO SOCIAL

FOLHA DE APROVAÇÃO

**ONTOLOGIAS COMO UM AGENTE PARA A INTERLIGAÇÃO DE DADOS
ABERTOS EM CONTEXTO GOVERNAMENTAL: características e
aplicabilidade**

LUCAS DE LIMA ROCHA

Dissertação apresentada ao curso de Mestrado em Ciência da Informação do Programa de Pós-Graduação da Universidade Federal Fluminense, como requisito parcial para obtenção do Grau de Mestre em Ciência da Informação.

Linha de pesquisa: Fluxos e mediações sócio técnicas da informação.

Orientadora: Prof^a. Dr^a. Maria Luiza de Almeida Campos

Co-orientador: Prof. Dr. Leonardo Cruz da Costa

Aprovado em: _____

Banca examinadora:

Prof^a. Dr^a. Maria Luiza de Almeida Campos (Orientadora)

Prof. Dr. Leonardo Cruz da Costa (Co-Orientador)

Prof. Dr. Carlos Henrique Marcondes (Membro da banca) UFF

Prof^a. Dr^a. Linair Maria Campos (Membro da banca) UFF

Prof. Dr. Mauricio Barcellos de Almeida (Membro da banca) UFMG

Prof^a. Dr^a. Regina de Barros Cianconi (Membro suplente interno) UFF

Hagar Espanha Gomes (Membro suplente externo) Livre docente

Niterói
2016

AGRADECIMENTOS

A minha família – Rodolfo, Suely, Diego –, pelo apoio incondicional nos momentos de incerteza, crises de ansiedade e alegria pelas conquistas alcançadas;

A Maria Luiza de Almeida Campos, por ser a melhor orientadora que qualquer aluno de mestrado poderia ter. Pelos sempre sábios conselhos, pela paciência, pelo incentivo nos momentos de dúvida e por me ajudar a modelar esse domínio de conhecimento chamado dissertação de mestrado;

Ao co-orientador Leonardo de Cruz da Costa, pela resolução das dúvidas acerca dos campos de conhecimento computacionais que não me são muito conhecidos;

Aos amigos de mestrado, em especial Thales e Lucas, por saberem a hora de ouvir e falar sobre seus próprios problemas, e ao apoio incondicional responsável pela solidificação da amizade que vêm desde a graduação;

Aos amigos fisicamente distantes ou próximos – Fábio, Jean, Bárbara, Gui, Babi, Pam, Taissa, Valéria, Vitor, Dayse, Fernanda, Mariana, Jéssyca, Ana Cristina, Julya –, por serem a segunda melhor família que qualquer pessoa poderia desejar. Por me ouvirem e insistirem pra eu largar a procrastinação e ir estudar;

Ao Programa de Pós-Graduação em Ciência da Informação da Universidade Fluminense, todo o seu corpo docente e seu excelente secretário Vitor Geraldo, por me ajudarem na resolução das dúvidas mais simples ou complexas, sempre com boa vontade e presteza;

À biblioteca Capital Humano da Fundação Getúlio Vargas e sua bibliotecária Virginia Spinelli Hottz, por ser o meu cantinho de refúgio nos dias de calor ou quando eu estava simplesmente muito cansado de encarar as paredes da minha própria casa.

Onde está a Vida que perdemos vivendo?
Onde está a sabedoria que perdemos no conhecimento?
Onde está o conhecimento que perdemos na informação?

T. S. Eliot

RESUMO

Apresentam-se diretrizes para aplicação de ontologias na interligação de dados governamentais na Web, evidenciando uma discussão sobre Dados Interligados Governamentais e Abertos no contexto brasileiro e a importância das ontologias como ferramentas semânticas para propiciar esta ligação. Através de pesquisa de caráter qualitativo e exploratório, elenca-se diretrizes a partir de análise realizada utilizando a metodologia ONTOMETRIC, e as quatro abordagens de interligação de dados governamentais. A metodologia ONTOMETRIC propõe cinco dimensões diferentes que dizem respeito à natureza das ontologias, sendo elas: conteúdo, linguagem, ferramentas, metodologia e custos. As quatro abordagens apresentam uma proposta de especificação de métodos diferentes para o processo de interligação de dados governamentais e abertos. A junção dos estudos possibilitou demonstrar quais são as maneiras possíveis de se interligar dados governamentais, e quais seriam as diretrizes apropriadas para que esta interligação ocorresse com o auxílio de ontologias, levando em consideração o contexto brasileiro e as especificações da Lei de Acesso à Informação. Conclui-se que o domínio governamental é plural e interdisciplinar, e deve-se levar em conta aspectos de multiplicidade de domínios, custos e treinamento de pessoal na escolha mais apropriada do tipo de ontologia a ser adotada. Desta forma, pretende-se que as diretrizes funcionem como um guia para que futuros idealizadores ou implementadores de ontologias possam ter um modelo inicial a partir do qual seguir no momento de escolha de um sistema informacional que tenha por objetivo organizar informações digitais de caráter governamental.

Palavras-chave: Ontologia. Dados Interligados Abertos. Dados Governamentais Interligados e Abertos. Representação do Conhecimento. Web Semântica. Lei de Acesso à Informação.

ABSTRACT

This study presents guidelines to the application of ontologies in the linking of governmental data on the Web, bringing forward a discussion about Governmental Linked Open Data in the Brazilian context, and the importance of ontologies as semantic tools to provide these connections. Through a qualitative and exploratory research, guidelines are listed from an analysis accomplished by the use of the ONTOMETRIC methodology, and the four approaches of linking governmental data. The ONTOMETRIC methodology proposes five different dimensions that concerns the ontology nature, being them: content, language, tools, methodology and costs. The four approaches presents a proposal of different methods specifications to the process of linking governmental open data. The union of the studies made possible to show different possibilities to linking governmental data, and which guidelines should be appropriate for this links to be made with the support of ontologies, taking into account the Brazilian context and the specifications of the Brazilian Information Access Law. The study concludes that the governmental domain is plural and interdisciplinary, and it should take in account aspects of domain multiplicity, costs and personnel training in the choosing of the most appropriate type of ontology to be adopted. Thus, it is intended that guidelines works as a guide for the future ontology creator or implementer can have an initial model from which to follow in the moment of choosing an information system that has for aim organizing digital information of governmental type.

Keywords: Ontology. Linked Open Data. Linked Open Governmental Data. Knowledge Representation. Semantic Web. Brazilian Information Access Law.

SUMÁRIO

1 INTRODUÇÃO	1
1.1 Objetivos	9
1.1.1 Objetivo geral	9
1.1.2 Objetivos específicos.....	9
2 METODOLOGIA	10
2.1 Etapas metodológicas	12
3 A WEB E O PAPEL DOS DADOS INTERLIGADOS: UMA NOVA CONFIGURAÇÃO	14
3.1 A Web de Documentos e a Web de Dados	15
3.2 Dados Interligados	19
3.2.1 Dados Interligados Abertos	24
4 DADOS GOVERNAMENTAIS INTERLIGADOS E ABERTOS	28
4.1 O cenário brasileiro de dados governamentais: especificações da lei versus realidade atual.....	38
4.1.1 Lei de Acesso à Informação brasileira.....	38
4.1.2 Análise do Portal Brasileiro de Dados Abertos à luz da Lei de Acesso à Informação	43
4.1.2.1 Critérios e análise de conteúdo da informação	47
4.1.2.2 Critérios e análise de estrutura da informação	49
5 ONTOLOGIA E OS DADOS INTERLIGADOS ABERTOS	51
5.1 Tipos e partes de ontologias	57
5.2 O papel das ontologias na padronização de dados governamentais interligados e abertos	62
5.2.1 Exemplos de interligação de dados através de ontologias.....	65
6 ANÁLISE DA METODOLOGIA ONTOMETRIC À LUZ DAS ABORDAGENS DA INTERLIGAÇÃO DE DADOS GOVERNAMENTAIS ABERTOS	70

6.1 Quatro abordagens para publicação de dados interligados abertos no cenário governamental.....	72
6.2 Metodologia ONTOMETRIC: uma análise das recomendações para escolha de uma ontologia apropriada.....	76
6.3 A integração das abordagens de Cordeiro e da metodologia Ontometric: especificação pelas dimensões.....	77
6.3.1 Dimensão: Ferramentas.....	78
6.3.2 Dimensão: Linguagem.....	82
6.3.3 Dimensão: Conteúdo.....	85
6.3.4 Dimensão: Metodologia.....	87
6.3.5 Dimensão: Custos.....	89
7 DIRETRIZES PARA A APLICAÇÃO DE ONTOLOGIAS NA INTERLIGAÇÃO DE DADOS GOVERNAMENTAIS ABERTOS.....	91
7.1 Diretrizes da dimensão ‘Ferramentas’.....	92
7.2 Diretrizes da dimensão ‘Linguagem’.....	93
7.3 Diretrizes da dimensão ‘Conteúdo’.....	95
7.4 Diretrizes da dimensão ‘Metodologia’.....	96
7.5 Diretrizes da dimensão ‘Custos’.....	97
8 CONSIDERAÇÕES FINAIS.....	99
REFERÊNCIAS.....	102
ANEXO A – EXPLICAÇÃO DAS DIMENSÕES ONTOMETRICS (LOZANO-TELLO, 2002, <i>tradução nossa</i>).....	110

1 INTRODUÇÃO

O cenário de produção informacional, atualmente, é desproporcional ao cenário de organização informacional, principalmente no espaço digital: se por um lado a Web se apresenta como um lugar democrático, no qual qualquer pessoa com as ferramentas e o conhecimento necessários para disponibilizar informações pode publicá-las, por outro é difícil recuperar com precisão essas informações, uma vez que elas são produzidas e indexadas, majoritariamente, em linguagem natural.

Ramalho, Vidotti e Fujita (2007) afirmam que

Nos últimos anos, o avanço exponencial na quantidade de recursos informacionais disponíveis no ambiente Web vem conduzindo a um estágio em que **os modelos clássicos de representação e recuperação de informações precisam ser (re)pensados sob diferentes perspectivas**, pois considerando a representação como elemento fundamental, para a garantia de qualidade na recuperação, apresenta-se como desafio a necessidade de singularização contextual na reconstrução do conhecimento, a partir da determinação de requisitos de qualidade e relevância das informações, que permitam categorizar e organizar, de maneira eficiente, o “oceano” de dados disponíveis, favorecendo a identificação de informações que realmente interessam ao usuário. (RAMALHO; VIDOTTI; FUJITA, 2007, n.p., grifo nosso)

Repensar esses modelos de representação e recuperação significa encontrar uma maneira de tornar a Web um lugar em que o processo de busca informacional não se torne apenas mais ágil, mas também mais inteligente e assertivo. A linguagem natural possui inúmeros problemas no que tange à recuperação informacional – o número de sinonímias e homonímias, por exemplo, é um aspecto que mascara ou torna excessiva a quantidade de informações recuperadas. De acordo com Breitman (2014)

Atualmente, a maior parte dos recursos primários presentes na Web está em linguagem natural, de modo que só podem ser interpretados por seres humanos. Deparamo-nos com essa situação diariamente, quando realizamos buscas na Web e somos forçados a “filtrar” informações que, mesmo dentro dos critérios de busca solicitados, pertencem a diferentes contextos. Um exemplo simples é fazer uma pesquisa sobre árvores – carvalhos. O resultado dessa busca, além de páginas referentes a esse tipo de árvore, também traz pessoas cujo sobrenome é Carvalho, empresas de guindastes e escritórios de advocacia. (BREITMAN, 2014, p. xi)

É pensando nesse sentido que esforços vêm sendo feitos para tornar a Web um espaço informacional não só recuperável através de palavras, mas também pelo conceito que as palavras representam dentro de um dado contexto. Com isso, também busca-se modelar essa Web de forma que ela não seja interpretável só por seres humanos, mas também por máquinas de processamento inteligente de dados, que, através de regras pré-estabelecidas, podem transformar esse espaço informacional em um lugar de sentido.

Essa perspectiva vem sendo estudada desde o início do século XXI através da proposta batizada de Web Semântica. Nessa Web, os recursos informacionais são descritos com significados bem definidos, proporcionando uma cooperação no trabalho entre pessoas e computadores (BERNERS-LEE; HENDLER; LASSILA, 2001).

A ideia da Web Semântica é a de que os dados disponíveis na Web sejam indexados de uma forma explícita, onde o relacionamento entre as páginas seja diferente do que é proposto por um *hiperlink*, por exemplo, no qual o ser humano deve interpretar o tipo de ligação existente entre duas coisas distintas. Essa perspectiva é chamada de Dados Interligados (BIZER, 2009; BIZER, HEATH, BERNERS-LEE, 2009; PARUNDEKAR, KNOBLOCK, AMBIT, 2010; BAUER, KALTENBÖCK, 2012), que traz a ideia de que todas as informações disponíveis na Web, para que se atinja o cenário ideal da Web Semântica, estejam interligadas.

Essa nova perspectiva de informação como um conjunto de dados estruturados é útil a diversos domínios, uma vez que, nesse novo contexto semântico, as informações podem ser reutilizadas e apropriadas às particularidades de cada pessoa ou grupo de pessoas interessadas em determinado aspecto daquela(s) informação(ões). Nesse sentido, destacamos a importância desse cenário semanticamente modelado para as informações de caráter governamental.

Atualmente, a maior parte das informações disponibilizadas pelo governo brasileiro é apropriada apenas para a leitura humana. Essas informações, que versam sobre diferentes assuntos governamentais, estão majoritariamente organizadas em formatos proprietários e em linguagem

natural, se utilizando de textos, planilhas e imagens que não podem ser interpretados por agentes inteligentes de software.

Recuperar dados nesse cenário é uma tarefa que, indo contra o estabelecimento das próprias diretrizes da Lei de Acesso à Informação – que manifesta que “é dever do Estado garantir o direito de acesso à informação, que será franqueada, **mediante procedimentos objetivos e ágeis, de forma transparente, clara e em linguagem de fácil compreensão**” (BRASIL, 2011, grifo nosso) –, não é fácil para o usuário comum: os próprios portais de transparência disponibilizam seus dados em formatos que não estabelecem conexões entre si e que não respondem bem a buscas.

É necessário que exista um mecanismo de recuperação da informação que lide com todas as particularidades de informações de caráter governamental, ao mesmo tempo em que sejam acessíveis aos diferentes tipos de usuário e respondam bem a buscas, das mais simples às mais sofisticadas.

Se por um lado o esforço pela transparência informacional vem se mostrando cada vez mais relevante, com investimentos que propiciam informações ao cidadão, entendemos que essa informação deva ser apresentada de forma inteligente, não-ambígua, e que possibilite uma recuperação mais precisa e ágil daquilo que se busca. Além disso, também entendemos que essas informações devem ser disponibilizadas de forma que cada cidadão possa não só utilizá-la isoladamente, mas que também tenha a possibilidade de interligá-la a outras informações, também modeladas dentro dessa perspectiva semântica, para que tanto seres humanos quanto os agentes inteligentes de software possam gerar novos significados às interligações estabelecidas.

O objetivo deste estudo é o de apresentar diretrizes para a aplicação de ontologias na interligação de dados governamentais, e para atingi-lo, pretende-se atingir duas metas, configuradas em objetivos específicos. A primeira delas é identificar de que maneira os Dados Interligados Abertos estão atualmente definidos e publicados na Web, na esfera governamental; e a segunda é a de evidenciar as características de aplicação de ontologias que visariam uma melhora na interligação dos dados na esfera governamental, objetivando sua publicação de forma consistente.

Para atingir a primeira meta, será necessário analisar a perspectiva de Dados Interligados Abertos como uma proposta, baseada na Web Semântica, de que as informações governamentais possam ser modeladas de forma altamente reutilizável e facilmente manipulável, com a explicitação de seus relacionamentos e a disponibilização tanto dos dados brutos quanto daqueles passados por tratamento semântico. Com isso, os dados semanticamente estruturados serão capazes de prover maneiras mais inteligentes de se trabalhar com essas informações, que precisam ser assertivas e corretas para atenção às questões dos cidadãos, seus donos originários.

Para que essas informações estejam interligadas, é necessário um agente no processo desses conteúdos de forma criteriosa, não-ambígua e colaborativa, de forma que qualquer pessoa possa ser capaz de colaborar para que essa visão se torne realidade. Nesse sentido, Campos (2006, p. 2) destaca que “na perspectiva da Web Semântica, este procedimento de controle de vocabulário vem sendo representado por um instrumento, que permite a representação e recuperação de dados e informações, denominado ontologia”. Breitman (2014, p. xi) também afirma a importância do controle do vocabulário através das ontologias, ao dizer que “a comunidade de Web Semântica aposta na utilização de ontologias, que fornecem uma *língua franca* na qual máquinas possam interagir de modo significativo”.

Para atingir a segunda meta, será necessário analisar as ontologias como especificações formais de uma conceitualização compartilhada, que fornecem uma compreensão comum e compartilhada de um domínio que pode ser comunicada a pessoas e sistemas (GRUBER, 1993; DING; FOO, 2002). Marcondes e Campos (2008) afirmam que esses artefatos tecnológicos, no contexto da Web Semântica, trabalham associados a sistemas de informação e utilizam linguagens padronizadas para que sua semântica seja compreensível por agentes de software, o que permite a sua interoperabilidade.

As ontologias, no entanto, não são simplesmente vocabulários controlados, como expõe Campos (2006):

Muita confusão se tem feito em torno do conceito de ontologia, que não pode ser considerado somente como um vocabulário controlado. Uma ontologia possui informações de natureza distinta, ou seja: terminológica – possui um conjunto básico de conceitos e relações –; e assertivas aplicadas aos conceitos e relações –, que constituem um conjunto de axiomas,

diferentemente de instrumentos de controle terminológico como os tesouros, por exemplo. Além disso, os softwares para elaboração de ontologias possuem, em sua grande maioria, funcionalidades que possibilitam o “entendimento” das necessidades dos usuários, pois se propõem à interpretação das necessidades de busca, agregando também a resposta a estas necessidades. Desta forma, estamos considerando sob um ponto de vista comparativo, que as ontologias com todo seu potencial de funcionamento, podem ser analogamente definidas como Sistemas de Recuperação de Informação. (CAMPOS, 2006, p. 2)

A literatura sobre a importância da modelização de domínios governamentais e da utilização de ontologias como ferramentas que permitem uma precisão mais apurada na busca e na representação dessas informações tem mostrado que o assunto vem sendo amplamente discutido tanto na esfera da Ciência da Informação quanto da Ciência da Computação, além de áreas como Administração (BAUER; KALTENBOCK, 2012; DING, 2010; ALBANO, 2014; ALVES; BAX, 2014; FONSECA; AZEVEDO; ALMEIDA, 2014; CAMPOS; CAMPOS, 2014). Isso só demonstra o caráter interdisciplinar desse assunto, que deve se preocupar com aspectos técnicos, tecnológicos, sociais e epistemológicos para compreender qual é a melhor maneira de utilizar ontologias para a disponibilização da informação mais apurada visando uma recuperação com qualidade.

Visualizando a ontologia como uma ferramenta sofisticada que permite a interligação de informações e a possibilidade de descoberta de novos conhecimentos – tanto pelos seres humanos quanto pelos agentes inteligentes de software –, este estudo busca elencar diretrizes úteis a uma ontologia para que funcione em um cenário de interligação de dados governamentais, além de entender de que forma essa ontologia pode ser aplicada nesse domínio específico. Assim, este estudo pretende apresentar uma lista de diretrizes úteis para que as ontologias funcionem como agentes no processo de interligação de informações governamentais, objetivando uma recuperação informacional precisa.

O foco da pesquisa verá as ontologias como linguagens computacionais responsáveis pelo controle de vocabulário, pelo estabelecimento de relações em um domínio e pela interoperabilidade semântica que proporcione a descoberta de novos conhecimentos, bem como às diretrizes necessárias para

que esses artefatos cumpram o objetivo de aumentar a precisão semântica na busca por informações governamentais interligadas e abertas.

A escolha pelo termo 'diretriz' mostrou-se mais adequada ao escopo deste trabalho. De acordo com Camargo e Vidotti (2011, p. 87), diretrizes "são consideradas recomendações ou atividades a serem realizadas ou checadas como oferecer menus de seleção, disponibilizar funções-chave, etc.". Outros termos, como 'critérios' e 'princípios' também foram levados em consideração, mas não se mostraram tão apropriados para a realização deste estudo.

Para que a lista de diretrizes possa ser estabelecida, este trabalho utilizará a metodologia ONTOMETRIC, apresentada por Lozano-Tello e Gómez-Pérez (2004), que possui a finalidade de apresentar uma série de recomendações para a escolha de uma ontologia para a interligação de dados. Além da metodologia, este trabalho também utilizará a análise de Cordeiro *et al.* (2011) sobre quatro métodos diferentes para o processo de interligação de dados governamentais abertos.

A junção dos trabalhos de Lozano-Tello e Gómez-Pérez (2004) e Cordeiro *et al.* (2011) possibilitará demonstrar quais são as maneiras possíveis de se interligar dados governamentais, e quais as diretrizes apropriadas para que esta interligação ocorra com o auxílio de ontologias e no contexto brasileiro, levando em consideração as especificações da Lei de Acesso à Informação (BRASIL, 2011).

O tema desse estudo surgiu a partir de algumas questões e constatações:

a. Não foram encontrados, durante o período de levantamento bibliográfico deste estudo, diretrizes para que uma ontologia funcione especificamente na interligação de informações em cenários governamentais. Acreditamos que, nesse sentido, poderemos contribuir com os estudos e as práticas daqueles que estejam interessados em transformar os dados brutos dos portais governamentais brasileiros em dados que possam promover interligações explícitas;

b. O trabalho pretende cumprir a função de entregar à sociedade um produto que possa ser de serventia aos mais diversos interessados em desenvolver ontologias que auxiliem na realização do projeto Web Semântica,

e, com isso, aumentar o potencial das informações disponibilizadas em formato bruto nos portais de dados abertos.

A organização desse trabalho se divide da seguinte maneira:

Capítulo 1 – Introdução: onde são introduzidos o tema da pesquisa e sua proposta, além da motivação e da justificativa sobre a importância do mesmo para a área de Ciência da Informação. Também é apresentada a organização geral do trabalho e os objetivos geral e específicos;

Capítulo 2 – Metodologia, onde é relatada a metodologia escolhida para que os objetivos da pesquisa pudessem ser alcançados

Capítulo 3 – A Web e o Papel dos Dados Interligados: Uma Nova Configuração: onde é apresentado o cenário de discussão desse estudo, comparando dois estágios diferentes da Web: a Web de Documentos – ou Web Sintática – e a Web de Dados – ou Web Semântica –, explicitando suas principais diferenças; também são discutidos os conceitos de Dados Interligados e Dados Interligados Abertos, com a finalidade de apresentar o contexto informativo no qual se insere o tema de pesquisa. Por fim, é apresentada uma seção sobre a Lei de Acesso à Informação brasileira, no qual é discutida a sua importância e suas especificidades quanto à publicação de dados governamentais.

Capítulo 4 – Dados Governamentais Interligados e Abertos: onde é discutido o conceito de Dados Governamentais Interligados e Abertos, como uma proposta de cenário no qual as informações de caráter público estejam interligadas de forma semântica. O capítulo demonstra os esforços coletivos para que os governos tornem suas informações mais inteligentes, realçando os estágios pelos quais a informação deve passar e quais são as problemáticas para a implementação desse projeto. Além disso, é efetuado um panorama histórico sobre o desenvolvimento da Lei de Acesso à Informação e, posteriormente, uma análise do Portal Brasileiro de Dados Abertos tendo como ferramenta norteadora a própria Lei de Acesso à Informação.

Capítulo 5 – Ontologia e os Dados Interligados Abertos: onde é exposto o conceito de ontologias como ferramentas tecnológicas indispensáveis para a realização do projeto Web Semântica e para a interligação de dados. É discutida a conceituação da ontologia, seus tipos, suas partes, suas funções e sua importância para a proposta de Dados Governamentais Interligados e

Abertos. Também são descritas as potenciais vantagens de utilização das ontologias para a interligação de dados governamentais abertos, além de exemplos onde são demonstrados esforços para que esse projeto se torne realidade, tanto a nível nacional quanto internacional.

Capítulo 6 – Análise da Metodologia ONTOMETRIC à luz das Abordagens da Interligação de Dados Governamentais Abertos: onde são apresentadas as abordagens de interligação de dados governamentais abertos (CORDEIRO *et al.*, 2011) e a metodologia ONTOMETRIC (LOZANO-TELLO; GÓMEZ-PÉREZ, 2004), estudos nos quais este trabalho se apoiou para que pudesse ser desenvolvido, além de uma proposta de análise das recomendações da metodologia ONTOMETRIC dentro das quatro abordagens de interligação de dados governamentais abertos.

Capítulo 7 – Diretrizes para a Aplicação de Ontologias na Interligação de Dados Governamentais Abertos: onde são apresadas as diretrizes consideradas mais adequadas à interligação de qualidade e consistência dos dados governamentais brasileiros e abertos, levando em consideração as recomendações estabelecidas por Lozano-Tello e Gómez-Pérez (2004), as quatro abordagens de interligação propostos por Cordeiro *et al.* (2011) e a realidade brasileira, configurada através da Lei de Acesso à Informação (BRASIL, 2011).

Capítulo 8 – Considerações Finais: onde são apresentadas as considerações finais e as perspectivas de continuação de estudos nesse domínio de conhecimento.

Por fim, são apresentadas as referências, seguidas pelo anexo.

1.1 Objetivos

1.1.1 Objetivo geral

Apresentar diretrizes para a aplicação de ontologias na interligação de dados governamentais.

1.1.2 Objetivos específicos

- Identificar de que maneira os Dados Interligados Abertos estão atualmente definidos e publicados na Web, na esfera governamental;
- Evidenciar as características de aplicação de ontologias que visariam uma melhora na interligação dos dados na esfera governamental, objetivando sua publicação de forma consistente.

2 METODOLOGIA

Neste capítulo são apresentados os aspectos metodológicos utilizados ao longo deste estudo. A princípio, são apresentados o método científico, os objetivos da pesquisa e a abordagem utilizada; depois, são elencados os procedimentos metodológicos utilizados para o alcance dos objetivos da pesquisa.

As etapas sobre as quais se debruçam a produção intelectual são de grande importância para o desenvolvimento do pensamento que será documentado na forma de um trabalho acadêmico. Lima (2004) diz que “toda metodologia é, ela própria, criação engenhosa do espírito humano, de processos cuidadosamente construídos, por gerações, para o sucesso prático de certas intenções que vieram, culturalmente, a nortear a produção do conhecimento”.

O caráter desta investigação se configura em uma pesquisa social. Para Gil (1989, p. 43), a pesquisa social pode ser definida como “o processo que, utilizando a metodologia científica, permite a observação de novos conhecimentos no campo da realidade social”. O trabalho está fundamentado na observação de quais são os benefícios da utilização das ontologias – linguagens responsáveis pela modelagem de um domínio, onde se pressupõe a descoberta automática de novos conhecimentos não previstos no momento de sua criação – em um contexto de Dados Governamentais Interligados e Abertos.

Gil (1989, p. 44-6) apresenta três níveis diferentes de pesquisa social, dependendo do escopo de seus objetivos específicos: as exploratórias, as descritivas e as explicativas.

As pesquisas exploratórias se caracterizam por desenvolver e modificar conceitos e ideias, com o objetivo de formular hipóteses que só serão completamente respondidas com estudos posteriores. Se caracterizam por envolverem principalmente o levantamento bibliográfico e documental, ou estudos de caso e entrevistas. A questão quantitativa – como coleta de dados ou amostragens – é pouco costumeira nesse nível de pesquisa.

As pesquisas descritivas se caracterizam por pesquisarem características de um determinado grupo a partir da coleta de amostragens, seja através de coleta de dados para avaliar estatisticamente características como sexo, escolaridade, estado de saúde física e mental, etc, ou através de pesquisas e questionários que buscam levantar as opiniões e crenças de determinada população.

Já as pesquisas explicativas são aquelas que buscam explicar o porquê das coisas, identificando fatores que determinam ou contribuem para a ocorrência dos fenômenos. É o tipo mais complexo de pesquisa, uma vez que as chances de errar são muito altas. Esse tipo de pesquisa é mais utilizado nas ciências naturais, através dos experimentos. Já para as ciências sociais, o processo se torna mais difícil por conta da irregularidade e a falta de controle e padrão para analisar um experimento; por isso, o método mais utilizado para pesquisas explicativas em ciências sociais é a observação.

Além de exploratório, este estudo também possui caráter qualitativo, a partir dos documentos levantados sobre o tema estudado. Lima (2004, p. 29) argumenta que

Os qualitativistas rejeitam o pressuposto que reconhece a existência de um único método de investigação, válido para todas as ciências, sejam estas físicas ou culturais. Para tanto, argumentam que a especificidade do objeto de investigação próprio das ciências humanas e sociais pressupõe a concepção e adoção de um método que seja capaz de respeitar tais singularidade. (LIMA, 2004, p. 29)

Dentro desta perspectiva, encaramos o trabalho como uma pesquisa exploratória, uma vez que o objetivo geral parte do pressuposto de que ontologias auxiliam no enriquecimento semântico de um domínio; e também a consideramos qualitativa, uma vez que o campo empírico possui particularidades que devem ser respeitadas ao longo de sua investigação.

Para que seja possível alcançar o objetivo geral desta pesquisa – a saber, *Apresentar diretrizes para a aplicação de ontologias na interligação de dados governamentais* –, é necessário estabelecer etapas metodológicas que irão evidenciar quais serão os caminhos tomados para respondê-lo. As etapas envolvem mecanismos que auxiliarão na fundamentação necessária para o

desenvolvimento dos objetivos específicos e, conseqüentemente, do objetivo geral da pesquisa.

A seguir, apresentaremos as etapas metodológicas.

2.1 Etapas metodológicas

As etapas metodológicas estão em conformidade com os objetivos específicos desta pesquisa e, para atingi-los, foram empregadas ações distintas.

Visando atender ao primeiro objetivo específico desta pesquisa – *Identificar de que maneira os Dados Interligados Abertos estão atualmente definidos e publicados na Web, na esfera governamental* –, foram empregadas as seguintes ações metodológicas:

1) Levantamento bibliográfico contemplando os campos da Ciência da Computação (CC) e Ciência da Informação (CI). O corpus foi feito através dos periódicos nacionais e internacionais, restringindo seu escopo para os últimos 10 anos nas seguintes bases de dados: Portal de Periódicos da CAPES, com a seleção dos artigos de Ciências Sociais Aplicadas de escopo nacional e internacional; Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI); Biblioteca Brasileira de Teses e Dissertações (BDTD); Annual Review of Information Science and Technology; Bulletin of the American Society for Information Science and Technology; IEEE Intelligent Systems; IEEE Latin America Transactions; International Journal of Human-Computer Studies. Além das bases de dados, foram feitas pesquisas individualizadas em diversas revistas acadêmicas de Ciência da Informação e Ciência da Computação de âmbito nacional e internacional, de forma a verificar se a pesquisa feita nas bases poderia ter omitido algum resultado. Também foi feita uma busca nos últimos 10 anos dos anais dos Encontros Nacionais de Pesquisa em Ciência da Informação (ENANCIB), através do repositório BENANCIB.

Os termos para a pesquisa por palavras-chave foram adotados nos idiomas inglês e português. Os termos em inglês foram: “*Linked data*”, “*Linked open data*”, “*Governmental data*”, “*Linked open governmental data*”, “*Ontology*”,

“*Data linkage*” e “*Semantic Web*”. Já em português os termos foram: “*Dados interligados*”, “*Dados interligados abertos*”, “*Dados abertos*”, “*Dados governamentais interligados e abertos*”, “*Dados governamentais abertos*”, “*ontologia*” e “*Web Semântica*”. As pesquisas também incluíram a combinação de dois ou mais destes termos, de forma a tornar os resultados mais refinados.

2) Após o levantamento bibliográfico, partiu-se para a etapa de análise e síntese do conteúdo levantado, que resultou nos capítulos que fundamentam a pesquisa. No processo de análise, foi necessário aprofundar as questões referentes ao entendimento do cenário sobre o qual esta pesquisa se encontra – ou seja, a Web Semântica –, e conceitos como o de dados interligados, dados interligados abertos, dados governamentais interligados e abertos e ontologias, através da literatura pertinente no âmbito da CI e da CC, tanto no escopo nacional quanto no internacional.

Os movimentos de levantamento bibliográfico, síntese e análise possibilitaram atender ao segundo objetivo específico desta pesquisa – *Evidenciar as características de aplicação de ontologias que visariam uma melhora na interligação dos dados na esfera governamental, objetivando sua publicação de forma consistente* –, que ocorreu através das seguintes etapas:

1) Identificação, na literatura, de procedimentos que visem a interligação de dados governamentais;

2) Identificação, na literatura, de propriedades que uma dada ontologia deve ter para atender as necessidades de interligação de dados;

3) Determinação de um procedimento de análise para verificação das propriedades de uma ontologia em um ambiente de interligação de dados governamentais;

4) Identificação de diretrizes para aplicação de ontologias na interligação de dados governamentais.

Acreditamos que, através deste processo metodológico, seja possível demonstrar quais são as diretrizes que as ontologias devem possuir para que, através dos dados interligados, possam recuperar as informações de caráter governamental de forma mais apurada e otimizada, estando em conformidade com a realidade brasileira e suas especificidades legislativas.

3 A WEB E O PAPEL DOS DADOS INTERLIGADOS: UMA NOVA CONFIGURAÇÃO

A cada minuto que passa, mais dados são inseridos na web. Seja através da criação de novas páginas, da publicação de novos conteúdos informacionais, da disponibilização de tabelas e dados brutos, textos, vídeos, imagens, sons e outros formatos de mídia, fato é que o crescimento informacional é constante e descontrolado. Produtores e consumidores de informação se confundem e se entrelaçam em um meio colaborativo onde todos que dispõem da tecnologia e do conhecimento necessário podem publicar algo.

Allemang e Hendler (2008) dizem que esse fenômeno pode ser resumido pelo slogan AAA (*Anyone can say Anything about Any topic* – ou seja, qualquer um pode dizer qualquer coisa sobre qualquer assunto). A Web atual é estruturada de maneira que publicar novos conteúdos é fácil, rápido e barato, criando um ambiente aparentemente democrático onde as informações, em teoria, estão disponíveis ao alcance de apenas alguns cliques.

Mas nem tudo nessa Web onde todos podem dizer o que quiserem é simples. Se publicar novos conteúdos é relativamente fácil, recuperar as informações que se busca é um processo que se torna cada vez mais complicado. Allemang e Hendler (2008) afirmam que a abundante quantidade de publicações na Web cria uma sensação de que há muita informação disponível sobre qualquer tipo de assunto, mas que tudo é extremamente raso e repetitivo; que as informações vêm de fontes não confiáveis, e que os esforços para conseguir encontrar o que é relevante em fontes distintas precisam ser cada vez maiores. A Web é o exemplo máximo de que quantidade nem sempre é sinônimo de qualidade.

O desenvolvimento da Web se deu de forma gradual, através do que Allemang e Hendler (2008) chamam de '*efeito de rede*': quando buscadores Web eram uma novidade e poucas pessoas possuíam a tecnologia e o conhecimento necessários para publicar informações online, essas poucas pessoas se esforçavam para entender como aquela nova ideia funcionava e, depois disso, contribuíam com sua parcela para o crescimento da Web. Com o

passar do tempo e o desenvolvimento de ferramentas que facilitavam o acesso e a publicação de documentos, esses esforços foram se tornando cada vez mais simples, para que agora não apenas especialistas, mas também usuários com menos conhecimento técnico pudessem participar do crescimento desse cenário. A partir do momento em que pessoas começaram a publicar documentos online, outras passaram a se interessar e desenvolver meios para que essa publicação se tornasse menos complexa.

No entanto, esse desenvolvimento não foi orgânico ou controlado, e a quantidade de informações pouco confiáveis, repetidas ou mal estruturadas também aumentou consideravelmente. E mesmo as informações relevantes estão majoritariamente organizadas em uma lógica que, de acordo com Breitman (2014), é sintática e não semântica.

Atualmente, a maior parte das informações disponíveis na Web está organizada em formatos como o HTML, que são próprios para a leitura humana, mas não para a busca inteligente de um software. Segundo Heath e Bizer (2011), formatos como o HTML sacrificam a estrutura e a semântica da informação, uma vez que são orientados de forma a expor documentos textuais e não dados estruturados. Pode-se encarar documentos textuais, aqui, como aqueles feitos para a leitura e o consumo dos seres humanos, e os dados estruturados como aqueles que podem ser interpretados tanto por seres humanos quanto por máquinas.

Essa diferenciação traça dois estágios distintos da Web: a primeira fase, conhecida como Web de Documentos, e a segunda, conhecida como Web de Dados. (ALLEMANG; HENDLER, 2008; BIZER; HEATH; BERNERS-LEE, 2009; BAUER; KALTENBÖCK, 2012; AUER, 2014). Breitman (2014) diferencia os dois momentos utilizando as nomenclaturas Web Sintática e Web Semântica.

3.1 A Web de Documentos e a Web de Dados

A Web de Documentos, também conhecida como Web Sintática, nasceu com os primeiros esforços para a publicação de conteúdos informacionais na

Web, quando essas publicações eram pouco populares e difíceis de serem feitas. Allemang e Hendler (2008) afirmam que, nessa Web, as informações são disponibilizadas através de formatos orientados para o consumo e a leitura de humanos, com uma grande variedade de fontes, estilos e organização da informação. Os mecanismos de recuperação informacional são, em sua maioria, organizados em linguagem natural, sendo necessários esforço e criatividade por parte do usuário para encontrar termos adequados que consigam expressar para o sistema aquilo que está sendo procurado. E, apesar de existirem alguns mecanismos de recuperação e catalogação de documentos, tais como marcações sociais (*social bookmarking*), sistemas estatísticos de informações mais procuradas nos serviços de busca e os próprios esforços dispendidos pelas empresas, essas informações podem não ser utilizadas em toda a sua potencialidade ou mesmo encontradas por aqueles que necessitam delas.

Breitman (2014) compara a organização da Web de Documentos a de uma biblioteca cujos livros estão espalhados sem distinção de assunto, e onde, para recuperá-los, seria necessário utilizar termos que remetessem ao livro desejado. Ela exemplifica a busca por um livro sobre TCP/IP, dizendo que usaria o termo 'rede' para a pesquisa. Com isso, os livros recuperados nessa biblioteca hipotética e sem organização seriam das mais diferentes naturezas: redes de computadores, redes telefônicas e até mesmo redes de artesanato. Caberia a ela, usuária, identificar dentre os livros recuperados quais seriam os mais interessantes para responder aquilo que procurava.

Esse cenário de desorganização que depende da interpretação humana gerou um movimento que pode ser encarado como o segundo estágio da Web: a Web de Dados (ALLEMANG; HENDLER, 2008; BIZER; HEATH; BERNERS-LEE, 2009; BAUER; KALTENBÖCK, 2012). A diferença entre a Web de Documentos e a Web de Dados não está na quantidade de informações que são disponibilizadas, mas sim na qualidade dessas informações, que, a partir de agora, são encaradas como dados. Enquanto na Web de Documentos as informações eram disponibilizadas para consumo humano, a Web de Dados se preocupa em dispor informações de forma que tanto pessoas quanto agentes

inteligentes de software possam ser capazes de consumi-las e gerar novos conhecimentos a partir delas.

A Web de Dados também é conhecida como Web Semântica (BREITMAN, 2014). O termo 'Web Semântica' foi melhor conhecido em 2001 através da publicação do artigo "The Semantic Web" por Tim Berners-Lee, James Hendler e Ora Lassila, na revista *Scientific American*. No texto, os autores preconizavam um novo tipo de Web, "que trará estrutura aos conteúdos significativos das páginas web, criando um ambiente no qual os agentes de software, transitando pelas páginas, poderão executar prontamente tarefas sofisticadas propostas pelos usuários" (BERNERS-LEE; HENDLER; LASSILA, 2001). Os autores afirmam que a Web Semântica não é uma Web separada da de Documentos, mas sim uma extensão da que já existe, na qual são atribuídos significados bem definidos às informações, permitindo que pessoas e computadores trabalhem de forma cooperativa.

A grande diferença entre esses dois cenários está na estrutura informacional. Allemang e Hendler (2008) dizem que a Web de Dados não torna os dados disponíveis mais inteligentes, mas dá a capacidade para que aplicações inteligentes possam ser feitas de forma mais simples e dinâmica. Os autores utilizam o seguinte exemplo: ao acessar um site para participar de um congresso, um usuário procura por um hotel que fique próximo do local da conferência. Na Web de Dados, ele poderia utilizar um mapa que traça a rota entre os hotéis da região e encontrar aquele que está à uma menor distância de seu destino final. Esse mapa também poderia disponibilizar os números de táxis que estão na região, bem como os restaurantes mais próximos e as opiniões das pessoas que já os frequentaram. Caso um restaurante não funcionasse em determinado dia, essa informação poderia ser atualizada apenas uma vez e estaria disponível nas informações presentes no próprio mapa. Já na Web de Documentos, todas essas informações também existem, mas estão em lugares diferentes, de forma mínima ou nulamente integradas, fazendo com que o tempo de busca do usuário e a possibilidade de encontrar informações desatualizadas seja maior.

É necessário ressaltar, no entanto, que a Web de Dados não significa uma quebra de paradigma revolucionária no que diz respeito à disponibilização

de dados na Web. As duas formas de tratamento informacional acontecem ao mesmo tempo, e esforços para que a Web de Dados se torne cada vez mais conhecida e de fácil acesso estão lado a lado de práticas comuns à Web de Documentos. A tabela abaixo sintetiza as principais diferenças entre esses dois tipos distintos de Web.

Quadro 1. Diferença entre Web de Documentos e Web de Dados

Web de Documentos	Web de Dados
Orientado para consumo informacional humano	Orientado para consumo informacional humano e computacional
Recuperação da informação baseada em linguagem natural	Recuperação informacional baseada em linguagem controlada
Pouca ou nenhuma integração entre diferentes páginas web	Páginas web integradas
Permite modelos estruturados como XML, mas não se comunicam entre si	Utiliza modelo de dados RDF, que se comunica com outras bases de dados
Ligações implícitas feitas através de hiperlinks	Ligações explícitas feitas através de ontologias

Fonte: Elaboração própria

Heath e Bizer (2011) levantam três questões-chave sobre a utilização de dados na Web: (i) como prover acesso aos dados de forma que possam ser facilmente reutilizados?; (ii) como permitir a descoberta de dados relevantes na grande quantidade de conjunto de dados?; (iii) como permitir que aplicações integrem dados de um grande número de fontes de dados desconhecidas?

Essas questões são importantes para entender que o projeto Web Semântica visa possibilitar esse novo momento da Web. Hoje, grande parte das informações que farão parte da Web Semântica já estão disponíveis online, mas em formatos que não são adequados, como textos em formatos proprietários, tabelas e planilhas. É necessário que esses conteúdos passem por um tratamento informacional que os deixem prontos para serem integrados em uma perspectiva semântica. Auer (2014) comenta que, da mesma forma

que hoje é possível publicar conteúdos desestruturados na Web – como páginas HTML – e fazer pesquisas através de motores de busca utilizando um esquema de palavras-chave, também é possível publicar informações estruturadas e ligá-las a outras informações também estruturadas e publicadas por terceiros, além de utilizar métodos mais expressivos de busca do que simples palavras-chave.

Esses dados, publicados de forma estruturada para que permitam uma recuperação mais precisa e sofisticada, são conhecidos como Dados Interligados (BAUER; KALTENBÖCK, 2012; BERNERS-LEE, 2006; BIZER; CYGANIAK; HEATH, 2006; AUER, 2014; BIZER; HEATH; BERNERS-LEE, 2009; HEATH; BIZER, 2011), e sua disponibilização é essencial para que a Web de Dados se consolide de forma que seus conteúdos informacionais possam ser melhor reaproveitados, manipulados e integrados a aplicações inteligentes.

3.2 Dados Interligados

Os Dados Interligados são a base da Web Semântica. De acordo com Marcondes (2012), esta é uma Web diferente da convencional, estruturada de forma semanticamente inteligível para a realização de inferências por agentes de software. É uma Web muito mais rica nas possibilidades de interação e recuperação semântica da informação.

O World Wide Web Consortium (W3C) “representa o consórcio de empresas, profissionais, cientistas e instituições acadêmicas responsável pela criação de padrões tecnológicos que regulam a World Wide Web” (VILLALOBOS; SILVA, 2010) e define Dados Interligados como “aqueles que devem ser disponibilizados em um formato padrão, acessível e manipulável, além de levar em conta os relacionamentos entre esses dados, que também devem ser explícitos”.

Os Dados Interligados têm por objetivo prover conexões compreensíveis por máquina entre os dados da Web (PARUNDEKAR; KNOBLOCK; AMBIT, 2010). Essas conexões são conhecidas como links semânticos, que funcionam como explicitações do tipo de relação existente entre dois dados que se

conectam. Marcondes (2012) utiliza o seguinte exemplo: em uma aplicação da Web de Documentos, tudo o que haveria de uma relação entre uma obra de Machado de Assis e a cidade do Rio de Janeiro seria um *hiperlink*, que guiaria o leitor de uma página para a outra, sem, no entanto, explicitar qual é o tipo de relação estabelecida entre os dois recursos; caberia ao ser humano inferir que essas informações se relacionam por conta da utilização da cidade como cenário para os romances do autor. No entanto, em uma aplicação semântica da Web de Dados, essas ligações seriam explícitas, de forma que tanto os seres humanos veriam que tipo de relação é essa – o que facilitaria a compreensão de alguém que, por exemplo, não tem conhecimento sobre a obra do autor –, quanto a máquina seria capaz de interpretá-la. Essa capacidade da máquina poderia fazer com que ela ligasse outras obras literárias que têm o Rio de Janeiro por cenário de forma automática, provendo aos programas navegadores a explicitação de significado da ligação entre os recursos, visando uma exploração mais rica da informação.

Bizer, Heath e Berners-Lee (2009) conceituam Dados Interligados como um conjunto de boas práticas que levam a Web a uma extensão de interligação de dados nos mais diversos domínios, tais como: pessoas, companhias, livros, publicações científicas, filmes, músicas, programas de TV e rádio, genes, proteínas, entre outros. Para os autores, esse tipo de boas práticas – que, segundo eles, giram em torno de utilizar a Web para criar novos links entre dados de diferentes fontes e formatos – permite novos tipos de aplicações.

Enquanto na Web de Documentos os *hiperlinks* são conectados através de relações sintáticas em HTML, os documentos conectados através dos Dados Interligados estruturam-se semanticamente a partir do formato RDF (*Resource Description Framework*). O RDF, desenvolvido pela W3C, é uma infraestrutura que permite a codificação, o intercâmbio e o reuso de metadados estruturados. Esta infraestrutura permite a interoperabilidade de metadados por meio da concepção de mecanismos que suportam convenções comuns de semântica, de sintaxe e de estrutura (MILLER, 1998).

Ainda segundo Miller (1998), diferente do HTML – que simplesmente conecta documentos sem explicitação de como essa conexão é efetuada –, os Dados Interligados utilizam o RDF para efetuar relacionamentos das coisas

conectadas de forma aparentemente arbitrária no mundo. A maior parte dos Dados Interligados são gerados automaticamente através de fontes de dados estruturados já existentes (como, por exemplo, uma base de dados relacional) na linguagem RDF, ou são interligados através da linguagem OWL, que instancia que dois *links* se referem à mesma coisa.

Berners-Lee (2006) lista uma série de passos, aos quais chama de 'regras', para que links semânticos sejam construídos dentro da lógica dos Dados Interligados. Segundo o autor, a Web Semântica não se trata apenas de acrescentar dados à Web, mas sim de criar links onde pessoas ou máquinas possam explorar os dados disponíveis. Quando corretamente estruturados, os Dados Interligados proporcionam a descoberta de outros dados relacionados.

Os passos listados por Berners-Lee (2006) são os seguintes:

1. Utilizar URIs para nomear as coisas;
2. Utilizar URIs em HTTP para que as pessoas possam visualizar esses nomes;
3. Quando alguém procurar por uma URI, proporcionar informação relevante, utilizando os padrões tais como RDF e SPARQL;
4. Incluir links às URIs para que possam ser descobertas novas coisas.

Essas quatro regras ou passos ficaram conhecidas como "Princípios dos Dados Interligados" (BIZER; HEATH; BERNERS-LEE, 2009) e versam sobre as práticas básicas para publicar e conectar dados utilizando a estrutura da Web.

O primeiro passo é essencial para a identificação dos dados. As URIs (*Uniform Resource Identifiers*) oferecem formas mais genéricas de identificar qualquer entidade existente no mundo, seja ela um documento, uma pessoa ou uma abstração, e são identificadores únicos no espaço de um servidor, o que torna seu desaparecimento muito mais difícil do que o de uma URL (*Uniform Resource Locator*). No entanto, as URLs não são indispensáveis, uma vez que funcionam como um endereço para a localização das URIs na Web, podendo ser utilizadas como identificadores em documentos que descrevam recursos disponíveis na Web (FERREIRA; SANTOS, 2013). Nisso entra o passo (2): um

endereço HTTP é a URL de um tipo de URI, ou seja, é um localizador para a representação de uma entidade existente dentro da base de dados.

Os passos (3) e (4) deixam mais claro o propósito mais interessante de se interligar dados: o de descobrir novas informações relevantes, seja através do esforço humano ou da capacidade das máquinas de produzir inferências. Proporcionar informações relevantes na URI de uma entidade (3) é importante para deixar claro o que ela pretende representar, e incluir links às URIs (4) demonstra que o esforço da Web Semântica é coletivo e pode ser feito por qualquer pessoa com conhecimentos técnicos e que possa contribuir com algo relevante.

Bizer, Cyganiak e Heath (2006) dizem que existem dois passos simples para interligar dados. São eles: utilizar um modelo de RDF para publicar dados estruturados na Web; e utilizar links RDF para interligar dados de diferentes locais.

Os autores citados acima afirmam que a linguagem RDF pressupõe um modelo de dados estruturados dentro de uma tripla que contempla sujeito, predicado e objeto de uma determinada relação de dados, sendo esses os elementos de seu metamodelo que representam entidades concretas ou abstratas do mundo real. O sujeito e o objeto desta tripla são links URIs, ou um URI e um *string literal*¹, respectivamente. O predicado desta tripla especifica como o sujeito e o objeto estão relacionados, e também é representado por uma URI.

Marcondes (2012) afirma também que os metamodelos reúnem os pressupostos semânticos de cada elemento do vocabulário da linguagem, e, através de um modelo ontológico expresso em RDF, sempre se poderá identificar, dentro de uma assertiva, a entidade (sujeito), que tipo de assertiva está sendo dita sobre essa entidade (predicado) e o que, especificamente, está sendo dito sobre a entidade (objeto).

Bizer, Heath e Berners-Lee (2009) apresentam o exemplo dessas relações triplas com o filme 'Pulp Fiction – Tempos de Violência', de Quentin

¹ Sequência de caracteres que, quando reunidos, formam um caractere nulo (MICROSOFT, 2013. Disponível em: <http://msdn.microsoft.com/en-us/library/vstudio/69ze775t.aspx>)

Tarantino. A linguagem RDF conecta os links da descrição do filme na Linked Movie Database (*subject*) com a descrição do filme disponível na DBpedia (*object*), sentenciando que ambos os links se referem à mesma entidade no mundo real (*predicate*) através da declaração '*owl:sameAs*'.

Figura 1. Exemplo de link em RDF

```
Subject: http://data.linkedmdb.org/resource/film/77
Predicate: http://www.w3.org/2002/07/owl#sameAs
Object: http://dbpedia.org/resource/Pulp_Fiction_%28film%29
```

Fonte: Bizer, Heath e Berners-Lee (2009)

Da mesma forma que a informação precisa ser estruturada ao entrar na Web, a pergunta a ser feita para que possa ser recuperada também deve ser estruturada. A linguagem que possibilita a busca desses documentos interligados dentro dos padrões RDF é a SPARQL (*SPARQL Protocol and RDF Query Language*), uma linguagem de formulação de questões para busca estabelecida dentro da ideia de triplas, que possui particularidades diferentes de uma busca por palavras-chave efetuada em linguagem natural (W3C, 2008).

Auer (2014) lista uma série de benefícios significativos para o uso de Dados Interligados. São eles:

- **Uniformidade:** Todas as bases de dados publicadas como Dados Interligados compartilham um modelo de dados uniforme, baseado em RDF. Com esse modelo, todas as informações são representadas através de fatos expressos em triplas que consistem em sujeito, predicado e objeto; e os elementos utilizados nessas representações são, principalmente, identificadores universais (URIs). Outros identificadores, sobretudo para informações de valor, também podem ser utilizados na posição de objeto.
- **Referenciabilidade:** URIs não são utilizadas apenas para identificar entidades, mas também podem ser utilizadas da mesma forma que URLs, possibilitando localizar e recuperar recursos que descrevem e representam essas entidades na Web.
- **Coerência:** Quando uma tripla RDF contém URIs de diferentes fontes de dados nas posições de sujeito e objeto, essa tripla basicamente estabelece uma ligação entre a entidade identificada pelo sujeito (e descreve a origem daquela

fonte de dados através do endereço A) com a entidade identificada pelo objeto (descrita em outra fonte de dados utilizando o endereço B). Através do link RDF que estabelece a relação entre as duas entidades, os dados estão efetivamente interligados.

- **Integrabilidade:** Uma vez que todas as fontes de Dados Interligados compartilham o modelo RDF – que é baseado em um único mecanismo para representação informacional – é muito fácil alcançar uma integração sintática e, de forma simples, uma integração semântica de diferentes fontes de Dados Interligados. Uma integração semântica de nível mais alto pode ser alcançada com a aplicação de técnicas de compatibilização a nível de esquema e de instância, para nova compatibilização posterior expressa como alinhamentos de vocabulários RDF e ontologias através de triplas adicionais de fatos.
- **Menor consumo de tempo:** Publicar e atualizar Dados Interligados é relativamente simples e faz com que o tempo seja poupado. Adicionalmente, uma vez que uma base de Dados Interligados é atualizada, é possível acessá-la e utilizá-la imediatamente, uma vez que o consumo de tempo advindo da análise de erros, transformações e carregamento de atualizações não é requerido (AUER, 2014, p. 5-6).

A proposta da Web Semântica é apostar em linguagens que estruturam os conteúdos de forma crescente, promovendo relações semânticas cada vez mais complexas que permitam a programas seguir essas relações através de inferências computacionais. Os conteúdos estruturados devem estar o mais público possível, permitindo que programas genéricos os processem.

3.2.1 Dados Interligados Abertos

Dentro dessa perspectiva, pode-se falar na proposta de Dados Interligados Abertos, que pressupõem a interligação, de acordo com os padrões RDF, das informações disponíveis em domínio público. Entender o conceito de Dados Interligados Abertos é fundamental para aplicá-los às informações governamentais, uma vez que grande parte do que está disponível é de natureza aberta e livre de direitos autorais.

O maior esforço por parte da comunidade acadêmica para interligar Dados Abertos é o *Linking Open Data W3C SWEO Community Project*, formado por uma comunidade que objetiva identificar os dados disponíveis com

licenças abertas, convertê-los dentro dos padrões RDF de acordo com os princípios dos Dados Abertos e publicá-los na Web (BIZER; HEATH; BERNERS-LEE, 2009). O Open Data Project tem por objetivo fazer com que os dados sejam disponíveis livremente para todos, através da ampliação da Web de Dados. Essa ampliação é feita através da publicação de várias fontes de Dados Abertos na Web e da criação de links RDF entre os diferentes itens de diferentes bases de dados.

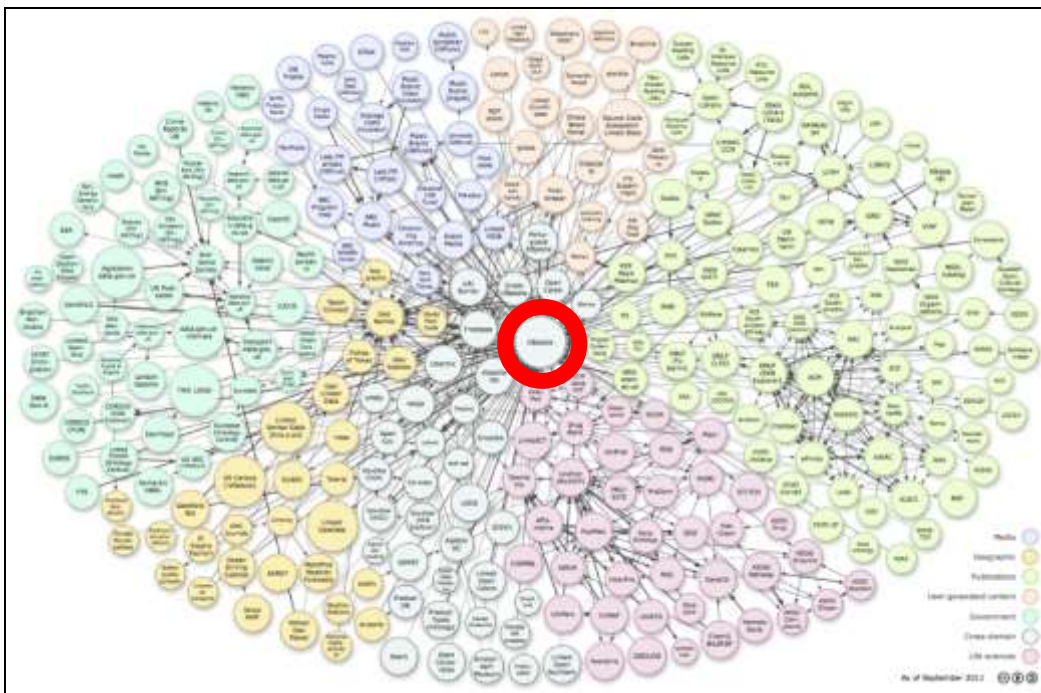
Bizer, Heath e Berners-Lee (2009) expõem que o Open Data Project conta com organizações significativas, tais como a BBC e a Biblioteca do Congresso Americano, e sua filosofia de cooperação e abertura a todos é um dos motivos de seu sucesso: qualquer um pode participar da interligação de dados, bastando apenas publicar novos dados de acordo com os parâmetros estabelecidos e relacioná-los a outros dados previamente publicados.

Outro exemplo de iniciativa de Dados Interligados Abertos é o projeto VIVO (2012), liderado por Michael Conlon, que possibilita a descoberta de pesquisadores entre diferentes instituições. Os participantes desta rede incluem instituições que podem agregar valor semântico para as informações compartilhadas, e a iniciativa fornece ferramentas para utilizar ontologias e gerenciar os Dados Interligados Abertos a ela associados.

O elemento central para interligar Dados Abertos atualmente existente é a DBpedia, uma *wiki* formada com o propósito de estruturar informações presentes na Wikipédia e disponibilizá-la de forma semântica. Segundo a apresentação do site,

(...) DBpedia permite ao usuário fazer perguntas sofisticadas aos verbetes da Wikipédia e linkar diferentes informações disponíveis na Web para as informações da Wikipédia. Esperamos que este trabalho faça com que o acesso à grande quantidade de informação disponível na Wikipédia possa ser efetuado de maneiras mais interessantes. Além disso, esperamos inspirar novos mecanismos para navegação, interligação e aperfeiçoamento da própria enciclopédia. (DBPEDIA, 2014)

Figura 2. O universo do conhecimento disponível na DBpedia, onde a DBpedia é o ponto central



Fonte: DBPEDIA (2013)

Para que essa interligação entre dados seja efetivamente possível, é necessário pensar na melhor maneira de modelar um domínio. Horrocks (2008) afirma que uma alternativa é concordar com um tipo de linguagem para definir novos termos, através da combinação e/ou restrição de termos já existentes. Também afirma que a linguagem precisa ser relativamente simples e precisamente especificada para que seja mais simples de ser processada por ferramentas de software. Isso aumenta a flexibilidade da ideia de Dados Interligados Abertos, de forma que novos termos possam ser incluídos sempre que necessário.

A ideia de Dados Interligados Abertos está voltada, principalmente, para informações de caráter e interesse públicos, como, por exemplo, as informações governamentais. O site Open Definition², um projeto do grupo Open Knowledge Foundation, diz que os dados são abertos quando “todos podem acessá-los, utilizá-los, modificá-los e compartilhá-los livremente –

² Open Definition. Disponível em: <<http://opendefinition.org/od/>>. Acesso em: 23 mar. 2016.

limitado, na maior parte das vezes, a medidas que preservem a proveniência e a abertura desses dados”. No capítulo seguinte, apresentaremos a ideia de Dados Governamentais Interligados e Abertos, como uma iniciativa que busca fazer com que as informações públicas possam ser disponibilizadas de forma que os cidadãos possam recuperá-las de forma mais rápida e menos burocrática.

4 DADOS GOVERNAMENTAIS INTERLIGADOS E ABERTOS

Um importante aspecto dos Dados Abertos é a utilização governamental das regras e diretrizes para a maior transparência sobre dados de caráter público. Os Dados Governamentais Abertos são aqueles que devem ser representadas em formato aberto e acessível, de tal modo que possam ser reutilizados, misturados com informações de outras fontes e gerar novos significados (W3C ESCRITÓRIO BRASIL, 2010). O principal movimento nesse sentido foi o de “Governo Aberto”, que tem como principal ideia “estabelecer uma cooperação moderna entre políticos, administração pública, indústria e cidadãos privados, permitindo mais transparência, democracia, participação e colaboração” (BAUER e KALTENBÖCK, 2012).

Segundo Bauer e Kaltenböck (2012) a parceria do Governo Aberto foi estabelecida em setembro de 2011, quando governos de oito países, entre eles o Brasil, aprovaram uma declaração de Governo Aberto, anunciando planos de ação e convidando outros trinta e oito países a se comprometerem com a participação. Os fatores que foram importantes para a aceitação do Governo Aberto diziam respeito, principalmente, ao acesso livre às informações e à possibilidade de utilização e reutilização livre desses dados e/ou conteúdos.

Os dados dentro da ideia de Governo Aberto tiveram diversas especificações que visavam padronizar as melhores práticas para sua publicação. Eaves (2009), especialista em políticas públicas e ativista de Dados Abertos, estabeleceu três leis para os Dados Governamentais Abertos. São eles:

- 1) Se o dado não pode ser encontrado e indexado na Web, ele não existe;
- 2) Se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser reaproveitado; e
- 3) Se algum dispositivo legal não permitir sua replicação, ele não é útil.

Essas três leis são extremamente relevantes quando se discute informações abertas de caráter governamental. Atualmente, duas das três leis

são, parcialmente, uma realidade para a política de acesso aos dados governamentais brasileiros. Muitos dos dados existem (1), uma vez que podem ser encontrados na Web – de forma desestruturada, o que dificulta sua indexação –, e a Lei de Acesso à Informação é o mecanismo legal que permite que tais informações sejam replicadas (3). No entanto, a lei mais relevante para a prática de boa utilização desses dados (2) ainda não é uma realidade. Alani *et al.* (2007 *apud* ALVES; BAX, 2014) destacam problemáticas no que diz respeito à prestação de contas governamentais via dados disponíveis nos portais Web, afirmando que os dados da Administração Pública são fragmentados, pouco integrados, em formatos distintos e, muitas vezes, proprietários, o que dificulta sua reutilização.

Outros princípios foram elaborados para definir Dados Governamentais Abertos, e os adotados pelo Brasil foram escritos ainda em 2007, antes da parceria de Governo Aberto. Os oito princípios, disponíveis no Portal Brasileiro de Dados Abertos³, são os seguintes:

- 1) **Completos.** Todos os dados públicos são disponibilizados. Dados são informações eletronicamente gravadas, incluindo, mas não se limitando a, documentos, bancos de dados, transcrições e gravações audiovisuais. Dados públicos são dados que não estão sujeitos a limitações válidas de privacidade, segurança ou controle de acesso, reguladas por estatutos.
- 2) **Primários.** Os dados são publicados na forma coletada na fonte, com a mais fina granularidade possível, e não de forma agregada ou transformada.
- 3) **Atuais.** Os dados são disponibilizados o quão rapidamente seja necessário para preservar o seu valor.
- 4) **Acessíveis.** Os dados são disponibilizados para o público mais amplo possível e para os propósitos mais variados possíveis.
- 5) **Processáveis por máquina.** Os dados são razoavelmente estruturados para possibilitar o seu processamento automatizado.
- 6) **Acesso não discriminatório.** Os dados estão disponíveis a todos, sem que seja necessária identificação ou registro.

³ Dados abertos. Disponível em: <<http://dados.gov.br/dados-abertos/>>. Acesso em: 5 fev. 2015.

- 7) **Formatos não proprietários.** Os dados estão disponíveis em um formato sobre o qual nenhum ente tenha controle exclusivo.
- 8) **Livres de licenças.** Os dados não estão sujeitos a regulações de direitos autorais, marcas, patentes ou segredo industrial. Restrições razoáveis de privacidade, segurança e controle de acesso podem ser permitidas na forma regulada por estatutos.

Além desses, outros dois princípios foram adicionados pela *Sunlight Foundation*⁴. São eles:

- 9) **Permanência.** Refere-se à capacidade de encontrar informações através do tempo.
- 10) **Custos de uso.** Uma das grandes barreiras para o uso e para a disponibilização ostensiva de informações públicas é o custo imposto para o público para acessá-las – mesmo que esse custo seja mínimo.

Todas essas leis e princípios versam sobre boas práticas que devem ser feitas para que a ideia de Web Semântica para Dados Governamentais Abertos seja uma realidade. E, para se beneficiar dessa estruturação em exemplos concretos, fica claro que os princípios de interoperabilidade e padrões de dados são questões centrais. Nesse contexto, a ideia de Dados Interligados Abertos é fundamental. Bauer e Kaltenböck (2012) afirmam que

Para se beneficiar completamente dos Dados Abertos, é crucial que se coloque informações e dados em um contexto que crie novos conhecimentos e que possibilite serviços e aplicações mais poderosos. Os Dados Interligados Abertos facilitam inovação e criação de conhecimento através de dados interligados, e é um importante mecanismo para gestão e integração de informação. (BAUER; KATENBÖCK, 2012, p. 17)

Ainda segundo Bauer e Kaltenböck (2012), existem dois importantes pontos de vista para a concretização dos Dados Interligados Abertos: publicação e consumo. Para que a proposta de Dados Governamentais Abertos (DGA) possa se concretizar de forma amplamente acessível e

⁴ Ten Open Data Principles. Disponível em: <<http://sunlightfoundation.com/policy/documents/ten-open-data-principles/>>. Acesso em: 2 maio 2015.

reutilizável, é preciso pensar na transição entre DGA para Dados Governamentais Interligados e Abertos (DGIA). Segundo os autores, esse caminho foi melhor discutido por Berners-Lee durante a apresentação do Modelo de Cinco Estrelas, onde são explicados os custos e os benefícios para a interligação de Dados Abertos através de cinco fases.

Quadro 2. O Modelo de Cinco Estrelas

Quantidade de estrelas	Benefícios para consumidores	Benefícios para publicadores
* A informação está disponível na web (em qualquer formato) sob licença aberta	Você pode ver, imprimir, armazenar localmente (no disco rígido ou em um armazenador USB) e pode armazenar dados manualmente em outro sistema.	É fácil de publicar.
* * A informação está disponível de forma estruturada (por exemplo, uma tabela Excel ao invés da imagem de uma tabela)	Você pode acessar diretamente com softwares proprietários para agregar, fazer cálculos, visualizar, etc; e pode exportar para outro formato (estruturado).	É fácil de publicar.
* * * Formatos não-proprietários são utilizados (por exemplo, CSV ao invés de Excel)	Você não precisa pagar por um formato do qual uma única entidade possui controle exclusivo.	É fácil de publicar.
* * * * Identificadores URI são utilizados para que as pessoas possam buscar por dados individuais	Você pode ligar os dados de qualquer outro lugar, seja na web ou em um servidor local, além de poder marcá-los e reutilizá-los, em parte ou totalmente.	Você deverá investir algum tempo dividindo e organizando seus dados, precisará atribuir URIs para os dados e pensar em formas de representar esses dados. Você também tem certo controle local sobre os dados e pode otimizar o seu acesso.
* * * * * Dados estão ligados a outros dados	Você pode descobrir novos dados interessantes enquanto consome outras informações,	Você precisará investir recursos para ligar seus dados para outros dados na

para prover contexto	além de possuir acesso ao esquema dos dados.	web. Você faz com que seus dados se tornem detectáveis e aumenta o valor deles.
----------------------	--	---

Fonte: Bauer e Kaltenböck (2012) (Adaptado)

Os Dados Interligados Abertos são extremamente importantes para os campos da informação e gestão de dados, e já estão sendo utilizados por diversas organizações, produtos e serviços na criação de portais, plataformas e serviços baseados em internet e aplicações. Uma vez que são independentes de domínio e atravessam vários domínios diferentes, os Dados Interligados Abertos são muito mais vantajosos do que uma gestão tradicional de dados. (BAUER; KALTENBÖCK, 2012)

Janssen, Charalabidis e Zuiderwijk (2012) listam uma série de benefícios dos DGIA nas esferas políticas e sociais, econômicas, e técnico-operacionais.

Quadro 3. Benefícios dos Dados Governamentais Interligados Abertos

Categoria	Benefícios
Políticos e econômicos	<p>Maior transparência</p> <p>Responsabilidade democrática</p> <p>Maior participação e responsabilidade por parte do cidadão</p> <p>Maior confiança no governo</p> <p>Maior participação pública</p> <p>Maior análise criteriosa dos dados</p> <p>Igualdade de acesso aos dados</p> <p>Novos serviços governamentais para os cidadãos</p> <p>Aperfeiçoamento de serviços prestados aos cidadãos</p> <p>Aperfeiçoamento da satisfação dos cidadãos</p> <p>Aperfeiçoamento dos processos de criação de políticas</p> <p>Maior visibilidade para os provedores de informação</p> <p>Maior estímulo ao desenvolvimento de conhecimento</p> <p>Criação de novas ideias para o setor público</p> <p>Serviços sociais novos e inovadores</p>
Econômicos	<p>Crescimento econômico e estímulo à competitividade</p> <p>Estímulo à inovação</p> <p>Contribuições no que diz respeito ao aprimoramento de</p>

	<p>processos, produtos e/ou serviços</p> <p>Desenvolvimento de novos produtos e serviços</p> <p>Uso da sabedoria da população: indo de encontro à inteligência coletiva</p> <p>Criação de um novo setor, aumentando valor à economia</p> <p>Disponibilidade da informação para investidores e empresas</p>
Técnico-operacionais	<p>A possibilidade de reutilizar dados sem a necessidade de coletá-los novamente ou enfrentar problemas de duplicações desnecessárias e custos associados (também por outras instituições públicas)</p> <p>Otimização de processos administrativos</p> <p>Aperfeiçoamento de políticas públicas</p> <p>Acesso a pessoas de outros lugares para a resolução de problemas</p> <p>Tomada de decisão justa, por possibilitar comparação</p> <p>Fácil acesso aos dados e à descoberta de dados</p> <p>Criação de novos dados baseados em dados combinados</p> <p>Avaliações externas sobre a qualidade dos dados (validação)</p> <p>Sustentabilidade dos dados (não há perda de dados)</p> <p>Capacidade de integrar, misturar e comparar dados públicos e privados</p>

Fonte: Janssen, Charalabidis e Zuiderwijk (2012)

Da mesma forma, Janssen, Charalabidis e Zuiderwijk (2012) também listam uma série de barreiras para a implementação dos DGIA. Os autores dividem as barreiras em seis grandes categorias: institucionais, de complexidade de tarefas, de uso e participação, de legislação, de qualidade da informação, e barreiras técnicas.

Quadro 4. Barreiras dos Dados Governamentais Interligados Abertos

Categoria	Barreiras
Institucionais	<p>Ênfase nas barreiras e negligência de oportunidades</p> <p>Troca pouco transparente entre valores públicos (transparência versus valores privados)</p> <p>Cultura de aversão a riscos (falta de empreendedorismo)</p> <p>Nenhuma política uniforme para publicar dados</p> <p>Tornar públicos apenas os dados que não são relevantes</p> <p>Falta de recursos para publicar dados (especialmente em setores pequenos)</p>

	<p>Sistema de receita é baseado em criar renda a partir dos dados</p> <p>Priorizar os interesses das organizações locais às custas dos interesses dos cidadãos</p> <p>Nenhum processo para lidar com o usuário que insere dados</p> <p>Qualidade discutível da inserção dos dados</p>
Complexidade de tarefa	<p>Falta de habilidade para descobrir os dados apropriados</p> <p>Falta de acesso aos dados originais (acesso apenas aos dados processados)</p> <p>Não há explicações sobre o significado dos dados</p> <p>Não há informações sobre a qualidade dos dados abertos</p> <p>Aplicativos que mascaram a complexidade, mas que podem ser potenciais novas utilizações dos dados</p> <p>Duplicação dos dados, disponibilidade em vários formatos ou disponíveis antes e depois do processamento resultam em discussões sobre qual é a fonte de dados</p> <p>Dificuldade para pesquisar uma vez que não há índice ou outras formas de assegurar busca facilitada para os dados que se pretende encontrar</p> <p>Mesmo que os dados possam ser encontrados, usuários podem não estar a par de suas potenciais utilizações</p> <p>Formatos dos dados e bases de dados são muito complexas para serem facilmente utilizadas</p> <p>Falta de suporte para auxílio de busca aos dados</p> <p>Foco na utilização de bases de dados únicas, quando o real valor está na utilização combinada de várias bases de dados</p> <p>Resultados contraditórios baseados na utilização dos mesmos dados</p> <p>Conclusões inválidas</p>
Uso e participação	<p>Não há incentivo para os usuários</p> <p>Organizações públicas não aceitam inserções feitas por usuários</p> <p>Frustração por haver diversas iniciativas de aprimoramento de dados</p> <p>Falta de tempo para entender os detalhes ou completa falta de tempo</p> <p>Obrigações de pagar uma assinatura para os dados</p> <p>Registro necessário antes de ser possível acessar os dados</p> <p>Crescimento inesperado de custos</p> <p>Falta de tempo para utilizar os dados abertos</p> <p>Falta de conhecimento para utilizar ou fazer com que os dados tenham sentido</p> <p>Falta da capacidade necessária para utilizar as</p>

	<p>informações</p> <p>Falta de conhecimento estatístico ou conhecimento dos potenciais e limitações estatísticos</p> <p>Ameaça de processos judiciais ou outras penalidades</p>
Legislação	<p>Violação de privacidade</p> <p>Segurança</p> <p>Falta de licença para utilizar os dados</p> <p>Condições limitadas para utilizar os dados</p> <p>Disputas e litígios</p> <p>Permissão por escrito antecipadamente necessária para ter acesso e para reproduzir os dados</p> <p>Reutilização de contratos/acordos</p>
Qualidade da informação	<p>Falta de informação</p> <p>Falta de informação confiável</p> <p>Informações incompletas, apenas parte do cenário exposto ou uma certa parcela</p> <p>Dados não-válidos ou obsoletos</p> <p>Valores pouco claros: a informação pode parecer irrelevante ou benigna quando vista isoladamente, mas quando ligada e analisada coletivamente, pode resultar em novos conhecimentos</p> <p>Muita informação para ser processada e falta de certeza para o que procurar primeiro</p> <p>Informações (essenciais) faltando</p> <p>Dados similares alocados em sistemas diferentes significa resultados diferentes</p>
Técnicas	<p>Dados devem estar em formatos bem definidos para que sejam facilmente acessíveis: enquanto o formato dos dados é arbitrário, o formato das definições dos dados deve ser rigorosamente definido</p> <p>Falta de padrões</p> <p>Falta de um portal ou uma arquitetura centrais</p> <p>Falta de suporte para tornar os dados disponíveis</p> <p>Falta de meta padrões</p> <p>Falta de um software padrão para o processamento de dados abertos</p> <p>Fragmentação de softwares e aplicações</p> <p>Sistemas legais que dificultam a publicação de dados</p>

Fonte: Janssen, Charalabidis e Zuiderwijk (2012)

Nesta dissertação, nossa preocupação se concentrará nos aspectos relacionados à utilização das ontologias em cenários governamentais, levando

em conta suas características e aplicabilidade para uma melhor disponibilização dos dados já existentes. A utilização de ontologias nesse contexto pode auxiliar na transposição de diversas barreiras comentadas por Janseen, Charalabidis e Zuiderwijk (2012), tais como:

- Na categoria Complexidade da Tarefa: Falta de habilidade para descobrir os dados apropriados; Não há explicações sobre o significado dos dados; Dificuldade para pesquisar uma vez que não há índice ou outras formas de assegurar busca facilitada para os dados que se pretende encontrar; Foco na utilização de bases de dados únicas, quando o real valor está na utilização combinada de várias bases de dados.

Nesta categoria, a aplicação de ontologias se mostra importante como facilitador sobre a explicação dos significados dos dados – uma vez que as instâncias apresentadas nas ontologias possuem explicações bem fundamentadas –; e na facilidade na pesquisa, que vasculha uma série de bases de dados diferentes e integradas, ao invés de utilizar uma única base de dados para uma pesquisa.

- Na categoria Uso e Participação: Obrigação de pagar uma assinatura para os dados; Falta da capacidade necessária para utilizar as informações.

Nesta categoria, a utilização de ontologias, na perspectiva de dados governamentais interligados e abertos, dispensa a obrigação de pagamento de assinatura para utilização dos dados, uma vez que o livre acesso é um dos pilares desse tipo de informação; além disso, a integração dos dados facilita sua utilização pelo usuário final, desde que aplicações que beneficiem a busca de informações interligadas por usuários não-especializados sejam disponibilizadas.

- Na categoria Qualidade da informação: Falta de informação; Falta de informação confiável; Informações incompletas, apenas parte do cenário exposto ou uma certa parcela; Dados não-válidos ou obsoletos; Valores pouco claros: a informação pode parecer irrelevante ou benigna quando vista isoladamente, mas quando ligada e analisada coletivamente, pode resultar em novos conhecimentos; Muita informação para ser processada e falta de certeza para o que procurar primeiro; Informações (essenciais) faltando; Dados similares alocados em sistemas diferentes significa resultados diferentes.

Nesta categoria, a ontologia poderia auxiliar na confiabilidade das informações apresentadas; na validação de dados, que passariam por tratamento semântico antes de sua interligação; no aumento do valor da informação, uma vez que estaria integrada a outras bases de dados e possibilitaria a descoberta de novos conhecimentos; possibilidade de diminuir a problemática de dados similares em lugares diferentes com a instanciação de entidades através de URIs.

- Na categoria Técnicas: Dados devem estar em formatos bem definidos para que sejam facilmente acessíveis: enquanto o formato dos dados é arbitrário, o formato das definições dos dados deve ser rigorosamente definido; Falta de padrões; Falta de um portal ou uma arquitetura centrais; Falta de suporte para tornar os dados disponíveis; Falta de meta padrões; Falta de um software padrão para o processamento de dados abertos; Fragmentação de softwares e aplicações.

Nesta categoria, a utilização de ontologias pode auxiliar na padronização do formato dos dados, que possuem rigor em seus formatos; a ontologia também funcionaria como um suporte central para tornar os dados governamentais disponíveis, provendo meta padrões pelos quais seria possível se basear no momento da modelagem do conhecimento.

4.1 O cenário brasileiro de dados governamentais: especificações da lei versus realidade atual

O objetivo da utilização das ontologias é o de diminuir as barreiras de utilização dos dados governamentais, e, no contexto brasileiro, diminuí-las se configura em um exercício que deve ser posto em prática, uma vez que os dados devem ser disponibilizados indiscriminadamente dentro dos parâmetros da Lei de Acesso à Informação para todos os cidadãos, seus donos por direito.

Na seção seguinte, apresentamos a Lei de Acesso à Informação como instrumento norteador para a política brasileira de disponibilização de dados governamentais abertos. Apresentamos seu histórico e as motivações para a sua criação; e, em um momento posterior, analisamos alguns de seus aspectos em confrontação com a realidade de disponibilização de dados governamentais no Brasil, ilustrado através do Portal Brasileiro de Dados Abertos.

4.1.1 Lei de Acesso à Informação brasileira

O Brasil não está livre das barreiras e dificuldades que dizem respeito à implementação de acesso aos Dados Interligados Abertos, mas vem realizando, desde 2000, ações para uma política de promoção à transparência da informação. De acordo com o Tribunal de Contas da União – TCU (BRASIL, 2015), existe todo um arcabouço normativo em vigência que diz respeito à promoção da transparência e da participação social na gestão pública. Duas leis complementares e um decreto foram importantes para dar o primeiro passo em direção a uma lei que dissesse respeito especificamente ao acesso à informação. São elas:

- Lei complementar 101/2000 (Lei de Responsabilidade Fiscal – LRF), que, em seu capítulo IX, trata da transparência, do controle e da fiscalização da gestão fiscal;
- Lei complementar 131/2009 (Lei de Transparência, também conhecida como Lei Capibaribe), que alterou a LRF a fim de determinar a disponibilização, em tempo real, de informações

pormenorizadas sobre a execução orçamentária e financeira da União, dos Estados, do Distrito Federal e dos Municípios;

- Decreto s/n de 15 de setembro de 2011, que instituiu o plano de ação nacional por meio do qual o Brasil celebrou a Parceria para Governo Aberto (OGP).

Todas essas ações culminaram na Lei 12.527/2011, também conhecida como Lei de Acesso à Informação, que dispõe sobre os procedimentos a serem observados pela União, Estados, Municípios e Distrito Federal, com o fim de garantir o acesso a informações conforme estabelecido pela Constituição Federal no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216, com o intuito de promover a efetivação deste direito humano fundamental (MEDEIROS; MAGALHÃES; PEREIRA, 2014; AROUCK; AMARAL, 2013; BRASIL, 2015). A lei conceitua informação como “dados, processados ou não, que podem ser utilizados para produção e transmissão de conhecimento, contidos em qualquer meio, suporte ou formato” (BRASIL, 2011).

Além da Lei de Acesso à Informação, outras duas ações – na forma de uma instrução normativa e um decreto – instituíram políticas que auxiliam na disponibilização transparente da informação. São elas:

- Instrução Normativa SLTI – 4/2012, que instituiu a Infraestrutura Nacional de Dados Abertos (Inda);
- Decreto 8.243/2014, que instituiu a Política Nacional de Participação Social – PNPS, com o objetivo de fortalecer e articular os mecanismos e as instâncias democráticas de diálogo e a atuação conjunta entre a Administração Pública Federal e a sociedade civil.

O TCU (BRASIL, 2015) elabora uma cartilha na qual lista uma série de motivos para a abertura de dados da Administração Pública. Segundo a publicação, “uma política que incentive a disponibilização de dados abertos possui potenciais diversos, como a melhoria na gestão pública, o provimento

da transparência, o estímulo ao controle e participação social, a geração de emprego e renda e o fomento à inovação tecnológica” (BRASIL, 2015, p. 5). Os cinco motivos para a abertura de dados na Administração Pública, de acordo com o TCU, são:

1) Porque a sociedade exige mais transparência na gestão

pública: a publicação dos dados governamentais em formato aberto permite que a sociedade possa avaliar as ações e decisões do governo. Dessa forma, os cidadãos assumem o papel de agente de transformação social por meio do acompanhamento e da fiscalização do desempenho do governo;

2) Porque a própria sociedade pode contribuir com os

serviços inovadores dos cidadãos: organizações, cidadãos, acadêmicos e até mesmo instituições públicas têm a possibilidade de utilizar bases de dados públicos para a produção e o compartilhamento de novos conhecimentos e de novos serviços, numa concepção de coparticipação entre entes privados e governo na oferta de serviços públicos à sociedade.

3) Porque ajuda a aprimorar a qualidade dos dados

governamentais: a abertura dos dados governamentais deve ser enxergada como uma oportunidade de menor custo para a melhoria desses dados, pois a própria sociedade pode identificar erros e apontar as devidas correções, o que reduz o esforço da administração na realização dessa tarefa.

4) Para viabilizar novos negócios:

percebe-se que o setor privado pode fazer uso de dados abertos governamentais para gerar produtos e serviços que são comercializados à população e que até então eram inexistentes. Dessa forma, pode-se dizer que a disponibilização de dados públicos em formato aberto pelo governo potencializa um retorno positivo pois, ao serem criados novos negócios, tem-se a geração de novos empregos e, por consequência, o aumento da receita pública mediante o recolhimento de tributos.

5) Porque é obrigatório por Lei:

a abertura de dados governamentais não se apresenta como uma mera alternativa de viabilização da transparência pública, mas como um dever a ser cumprido pelo administrador público.

O Estado é responsável pela garantia de direitos de acesso à informação, que deve ser veiculada de forma transparente, ágil, clara e em linguagem de fácil compreensão, uma vez que o usuário interessado nessas informações é diverso e plural, e poderia se sentir afastado pelos tecnicismos de uma linguagem pouco adequada. Com isso, as informações precisam ser claras e compreensíveis para que possam ser interpretadas da forma mais direta possível (MEDEIROS; MAGALHÃES; PEREIRA, 2014). Além disso, a Lei

de Acesso à Informação também prevê a disponibilização de dados em formatos abertos (ALVES; BAX, 2014; ALBANO, 2014).

A Lei de Acesso à Informação diz que toda a informação produzida ou gerenciada pelo governo é pública, exceto as informações de cunho pessoal. Outras informações que entram na categoria de exceções são aquelas que dizem respeito à segurança do Estado, aos projetos de pesquisa e desenvolvimento científico e tecnológico, além de informações utilizadas para tomada de decisões (ALBANO, 2014). A Lei de Acesso à Informação distingue as informações de dados considerados sigilosos em três grupos diferenciados, a saber: ultrassecretos, secretos e reservados, e estipula prazos máximos para que essas exceções informacionais mantenham sua natureza sigilosa. No caso de informações ultrassecretas, o prazo para divulgação é de 25 anos; para as informações secretas, 15 anos; e para as informações reservadas, que dizem respeito à segurança do Presidente e Vice-Presidente da República, bem como a seus eventuais filhos(as) ou cônjuges, as informações ficam sob sigilo até o término do mandato em exercício (BRASIL, 2011).

As informações dispostas para o cidadão não devem apenas constar nas bases de dados; elas devem ter, antes de mais nada, qualidade. Arouck e Amaral (2013) chegaram à definição de cinco atributos de qualidade essenciais para o cumprimento da Lei 12.527/2011. São elas:

- **Disponibilidade:** diz respeito à possibilidade de utilização por determinado indivíduo, equipamento ou sistema autorizado. A disponibilidade qualifica a informação em disponível ou indisponível.
- **Autenticidade:** diz respeito à produção, expedição, recebimento ou modificação da informação por determinado indivíduo, equipamento ou sistema automatizado. A autenticidade qualifica a informação como autêntica ou inautêntica.
- **Integridade:** diz respeito à informação não modificada, inclusive quanto à origem, trânsito e destino. A integridade qualifica a informação como íntegra ou adulterada, modificada.
- **Primariedade:** diz respeito à coleta da informação na fonte, com o máximo de detalhamento possível, sem modificações. Nesse quesito, existem outros atributos que o complementam: completude, que diz respeito à informação a qual não faltam partes ou elementos que a constitui ou deve ter, e que é qualificada como completa ou incompleta; e integridade, já definida acima.

- **Atualidade:** diz respeito à reunião mais recente dos dados mais recentes sobre o tema, de acordo com a natureza, os prazos previstos em normas específicas ou conforme a periodicidade estabelecida nos sistemas informatizados que a organizam. (AROUCK; AMARAL, 2013)

Um caminho importante para o pleno cumprimento das diretrizes da Lei é o da utilização de Tecnologias da Informação e Comunicação (TICs) como meio facilitador para que as informações cheguem ao cidadão de forma mais rápida e segura. Utilizar essas tecnologias é importante para manter um diálogo acessível com a população, uma vez que, através dessas ferramentas, é possível que o cidadão fiscalize os poderes públicos e participe da tomada de decisão sem que sua presença física em um lugar específico seja necessária. O governo brasileiro propôs algumas importantes iniciativas para o aumento da transparência de informações públicas baseada na utilização de TICs, tais como a criação do Portal Brasileiro de Dados Abertos, que pretende disponibilizar qualquer tipo de dado que tenha sua origem na esfera pública.

No entanto, esse tipo de portal ainda carece de medidas que tornem os dados disponíveis realmente livres para utilização. Pereira e Marcondes (2014) constatam que “as informações disponibilizadas pelo governo brasileiro, hoje, estão majoritariamente publicadas em formatos proprietários ou com tecnologias que restringem o seu acesso”. Entendemos por formatos proprietários os tipos de informações que só podem ser acessadas mediante presença de um software específico que carece de pagamento para que sua utilização seja possível.

A legislação é clara ao prever que os órgãos e entidades públicas divulguem as informações em locais de fácil acesso, sem a necessidade de requerimentos. Assim, tendo em vista todo esse contexto, fica claro que ainda existem diversos problemas que dizem respeito ao cumprimento da Lei de Acesso à Informação. De acordo com Sant’Ana e Rodrigues (2013, p. 1), “mesmo com os recursos proporcionados pelas Tecnologias da Informação e Comunicação, são grandes as dificuldades de ampliação da participação popular nas administrações públicas”.

A Lei de Acesso à Informação, por mais que seja um instrumento que consolida as obrigações da Administração Pública para com o cidadão, ainda carece de cumprimento efetivo, uma vez que suas diretrizes, na prática, são apenas parcialmente cumpridas. Em um contexto de informações semânticas essa problemática se torna ainda mais nítida, já que, mesmo que a Lei de Acesso à Informação esteja em sintonia com as novas tecnologias (ALVES; BAX, 2014), o cumprimento das obrigações ainda não é plenamente atendido.

Um dos objetivos específicos desta pesquisa é o de identificar de que maneira os Dados Interligados Abertos estão atualmente definidos e publicados na Web, na esfera governamental e, no contexto brasileiro, a utilização da Lei de Acesso à Informação se apresenta como instrumento norteador para elencar critérios de análise.

4.1.2 Análise do Portal Brasileiro de Dados Abertos à luz da Lei de Acesso à Informação

Uma das questões levantadas durante a investigação foi a de tentar entender em que medida a Lei de Acesso à Informação se configura em realidade para uma política de dados interligados abertos no Brasil. Para tal, levando em consideração os próprios aspectos da Lei no que diz respeito à disponibilidade e qualidade da informação apresentada, estabelecemos alguns critérios de análise da informação disponibilizada pelo governo brasileiro, restringindo-nos ao acesso através de Tecnologias da Informação e Comunicação. Dividimos os aspectos em duas esferas diferentes, sendo elas (1) Critérios de conteúdo da informação; e (2) Critérios de estrutura da informação. Entendemos por critérios de conteúdo da informação os aspectos que dizem respeito à natureza informacional, sua proveniência e confiabilidade; e por critérios de estrutura da informação os aspectos que dizem respeito à forma como essas informações estão disponibilizadas. Os critérios podem ser distribuídos da seguinte forma:

Quadro 5. Critério de conteúdo da informação

A informação é autêntica? (Art 4º, VII)
A informação é íntegra? (Art 4º, VIII)
A informação é primária? (Art 4º, IX)
A informação é transparente, clara e em linguagem de fácil compreensão? (Art. 5º)

Fonte: Elaboração própria

Quadro 6. Critérios de estrutura da informação

A informação está disponível para utilização de equipamentos e sistemas automatizados? (Art 4º, VI)
É possível gravar a informação em diversos formatos, inclusive abertos e não proprietários? (Art 8º, §3º, II)
É possível o acesso automatizado por sistemas externos em formatos abertos, estruturados e legíveis por máquina? (Art 8º, §3º, III)
São divulgados em detalhes os formatos utilizados para a estruturação da informação? (Art 8º, §3º, IV)

Fonte: Elaboração própria

Como campo de análise, optamos pelo Portal Brasileiro de Dados Abertos (PBDA), que é “a ferramenta disponibilizada pelo governo para que todos possam encontrar e utilizar os dados e as informações públicas⁵”. Segundo o PBDA, as informações e os dados são disponibilizados de forma que encontrá-los seja fácil, além de promover a interlocução entre atores da sociedade e com o governo para pensar a melhor utilização dos dados em prol de uma sociedade melhor.

⁵ Disponível em: <<http://dados.gov.br/sobre/>>. Acesso em: 23 mar. 2016.

Figura 2. PBDA, página inicial. Pesquisa, dados em destaque e publicações mais recentes



Fonte: <http://dados.gov.br>

A página inicial do PBDA é simples e organizada da seguinte forma: no canto superior esquerdo há quatro opções: (1) Ir para o conteúdo; (2) Ir para o menu; (3) Ir para a busca; e (4) Ir para o rodapé. Há um menu simples ao lado do título, no lado direito, com as seguintes opções: “dados”, onde são listados todos os dados disponíveis no sítio; “aplicativos”, onde estão listados os aplicativos desenvolvidos com licença aberta a partir dos dados disponibilizados no PBDA; “perguntas frequentes”, onde são listadas informações relevantes sobre o desenvolvimento do portal; “contato”, onde há um formulário onde podem ser colocadas dúvidas que serão sanadas pelos desenvolvedores do PBDA; e “sobre o portal”, onde existem informações sobre as motivações que levaram à construção do PBDA.

A busca no portal é feita através de uma caixa de diálogo em linguagem natural, e no canto direito da tela há uma janela de destaque para os últimos conjuntos de dados publicados na base. Abaixo da caixa de pesquisa há uma janela com os dados em destaque e, mais abaixo, há uma caixa na esquerda com os dados categorizados por etiquetas (*tags*) e na direita uma caixa com informações relacionadas ao PBDA. O rodapé da página concentra links com informações sobre o próprio portal e a comunidade que o alimenta, além de

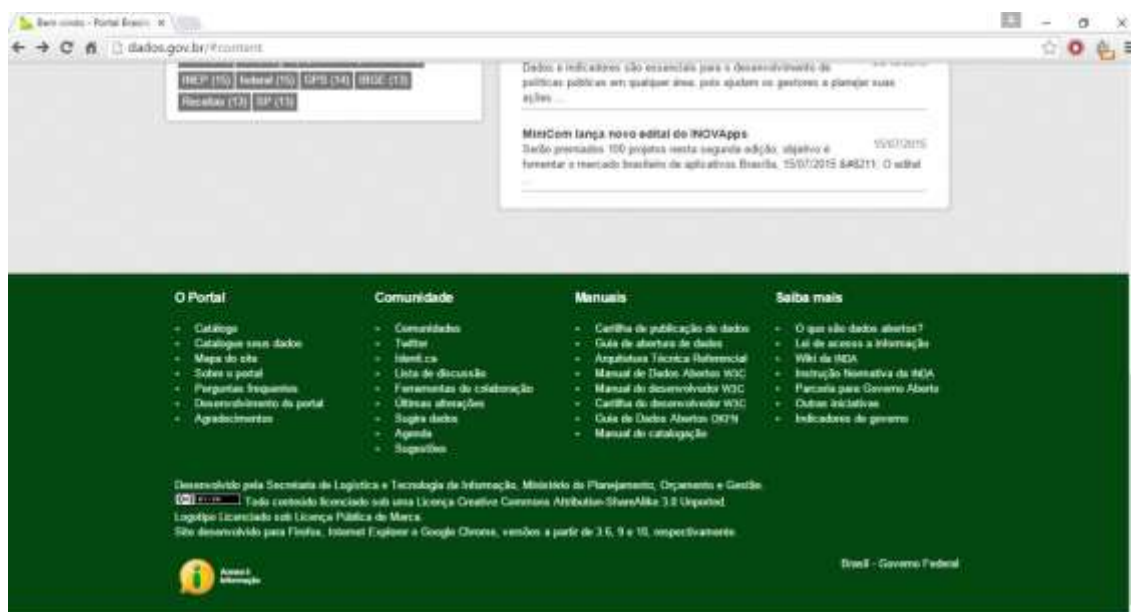
links para manuais que versam sobre a publicação de dados abertos e a política de dados abertos, e outros links intitulados 'Saiba mais', onde informações sobre o que são dados abertos e quais as iniciativas que estão sendo feitas são disponibilizadas.

Figura 3. PBDA, página inicial. Dados por etiqueta e notícias



Fonte: <http://dados.gov.br>

Figura 4. PBDA, rodapé



Fonte: <http://dados.gov.br>

Ao acessar um conjunto de dados, somos direcionados a uma página que contém o título do conjunto de dados, uma breve descrição e os dados em si, disponibilizados em diversos formatos. Abaixo, há uma seção de marcadores (*tags*) descritos em linguagem natural pelo próprio produtor da informação, sem nenhum tipo de controle de vocabulário. No campo de Informações adicionais, são disponibilizados a fonte do conjunto de dados, o(s) autor(res), o(s) mantenedor(es), a que esfera governamental aquele conjunto de dados pertence (Federal, Estadual, Distrital ou Municipal) e a qual poder governamental pertence (Executivo, Legislativo ou Judiciário).

4.1.2.1 Critérios e análise de conteúdo da informação

1) A informação é autêntica? (Art 4º, VII)

A proveniência dos dados disponibilizados no PBDA é informada pelo próprio produtor da informação, mas o PBDA não obriga que a autoria do conjunto de dados seja especificada. Na página de dados, há a possibilidade de filtrar os resultados de busca pela organização produtora da informação, no

entanto, ao se clicar em uma organização produtora, os resultados encontrados sempre são zero. Em alguns casos, não há informação sobre o autor do conjunto de dados – no campo de autoria a informação aparece como ‘não informado’.

2) A informação é íntegra? (Art 4º, VIII)

Da mesma forma que é possível avaliar a autenticidade da informação através da proveniência, a integridade pode ser mensurada pelo ente produtor daquele conjunto de dados. No caso de informações cuja autoria não é especificada, a confiança na integridade é gravemente afetada.

3) A informação é primária? (Art 4º, IX)

A primariedade da informação deve ser avaliada de caso a caso, mas a recomendação do PBDA é que a informação esteja disponibilizada da forma mais bruta possível; mesmo que o produtor acredite que informações cruzadas também sejam valiosas, as informações sem cruzamento também possuem seu valor.

4) A informação é transparente, clara e em linguagem de fácil compreensão? (Art. 5º)

A Art. 5º da Lei de Acesso à Informação diz que “É dever do Estado garantir o direito de acesso à informação, que será franqueada, mediante procedimentos objetivos e ágeis, de forma transparente, clara e em linguagem de fácil compreensão.” (BRASIL, 2011); no entanto, é difícil mensurar em que medida uma informação possui características que podem mudar de acordo com quem a interpreta. No geral, os manuais de utilização da informação são claros, mas as informações brutas não são facilmente compreendidas, ainda mais por utilizarem formatos que não são costumeiros à população de uma maneira geral.

4.1.2.2 Critérios e análise de estrutura da informação

1) A informação está disponível para utilização de equipamentos e sistemas automatizados? (Art 4º, VI)

Existem informações em formatos proprietários onde não é possível utilizá-las em outros equipamentos e sistemas automatizados sem um pré-processamento informacional, como as disponibilizadas em formato PDF. No entanto, a maior parte das informações se encontram em formatos onde é possível a utilização de sistemas automatizados, mesmo que as informações estejam estruturadas em formatos proprietários.

2) É possível gravar a informação em diversos formatos, inclusive abertos e não proprietários? (Art 8º, §3º, II)

A Cartilha Técnica para Publicação de Dados Abertos⁶ recomenda a publicação em diferentes formatos, preferencialmente abertos, estáveis e não proprietários. No entanto, não há fiscalização no que diz respeito à publicação de dados proprietários e em formatos não-proprietários. Com isso, a obrigação para com a lei passa a ser facultativa.

3) É possível o acesso automatizado por sistemas externos em formatos abertos, estruturados e legíveis por máquina? (Art 8º, §3º, III)

Como as informações são estruturadas em formatos que possibilitam aplicações, é possível que estas sejam acessadas por sistemas automatizados; no entanto, esse acesso não é automatizado, uma vez que as informações só podem ser pré-visualizadas ou baixadas, para só então serem acessadas.

⁶ Cartilha de Publicação de Dados Abertos: Disponível em: <<http://dados.gov.br/cartilha-publicacao-dados-abertos/>>. Acesso em: 23 mar. 2016.

4) São divulgados em detalhes os formatos utilizados para a estruturação da informação? (Art 8º, §3º, IV)

A Cartilha Técnica para Publicação de Dados Abertos tem uma seção na qual especifica e explica brevemente quais são os melhores formatos para a publicação de informações e como cada um desses formatos funciona.

Todos esses pontos convergem para a ideia de que os dados devem ser estruturados de maneira que possam ser rápida e facilmente recuperados; que sua recuperação seja precisa e assertiva; e que o usuário se sinta confortável durante o processo de busca e sinta confiança no sistema para utilizá-lo novamente.

Por conta disso, destacamos a importância das ontologias como mecanismos de controle de vocabulário e explicitação de relações entre diferentes recursos. Eles funcionam como agentes no processo de modelagem de um domínio, e, através delas, é possível que as informações governamentais possam ser disponibilizadas de forma que exista uma recuperação informacional com uma semântica mais apurada, que leve em conta dados de fontes distintas que podem ser misturados, utilizados e reutilizados em aplicações inteligentes que servem melhor aos propósitos de busca dos mais diferentes tipos de usuário.

5 ONTOLOGIA E OS DADOS INTERLIGADOS ABERTOS

Uma das maiores problemáticas acerca da estruturação da Web atual é a de que os computadores não possuem a capacidade de abstração humana; com isso, as informações são representadas de forma matemática para seu posterior processamento. E uma das preocupações da Ciência da Informação é, provavelmente, a forma mais eficaz de representar o conhecimento; por isso, uma de suas linhas de pesquisa gira em torno da criação de mecanismos que auxiliem na busca e posterior recuperação da informação representada.

No contexto dos dados interligados abertos, a utilização de tecnologias que permitam a organização informacional e a explicitação do relacionamento entre as informações é indispensável. Com isso, a utilização de ontologias se mostra relevante para a interligação de dados, uma vez que essa ferramenta de controle terminológico possibilita a modelagem de um domínio e das relações entre os seus conteúdos.

O termo 'Ontologia' foi originalmente criado na área de Filosofia, e sua origem remonta ao início do século XVII, entre 1606 e 1613. Sua etimologia o define como estudo da existência e, no âmbito filosófico, o termo pode ser definido como “um braço da metafísica que se preocupa com a natureza e a relação do ser” e a “teoria que diz respeito aos tipos de entidades e, especificamente, aos tipos de entidades abstratas que podem ser admitidas em sistemas de linguagem” (MERRIAM-WEBSTER, 2004). Ela surge neste contexto por conta da “necessidade da organização do conhecimento, no contexto da investigação filosófica dos componentes da realidade”, e “independe de uma linguagem específica” (CAMPOS, 2011, p. 34). Smith (2003, p. 1), ao descrever a Ontologia no campo da Filosofia, afirma que ela “busca trazer a classificação definitiva e exaustiva de entidades em todas as esferas do ser”. A Ontologia busca responder a questões como: “Quais classes de entidades são necessárias para uma completa descrição e explicação de todas as coisas que acontecem no universo?” ou “Quais classes de entidades são necessárias para dar conta do que torna verdadeiro todas as verdades?”.

Partindo do discurso no campo da Filosofia, Chateaubriand (1998) afirma que

Em princípio, pode-se dizer que Ontologia é uma teoria do real, uma teoria do ser. Essa é uma concepção tradicional que afirma, em geral, que a Ontologia envolve alguma categorização muito ampla da realidade. Isto é, para se fazer uma teoria ontológica, uma das primeiras coisas a fazer, e é o que foi feito na tradição filosófica, é categorizar de alguma maneira a realidade em certas grandes distribuições do ser. (CHATEAUBRIAND, 1998).

No contexto da Ciência da Computação, Guarino (1998) distingue *Ontologia* como um sistema de categorias e *ontologia* como vocabulário – diferenciando-as através da grafia da palavra e de sua inicial maiúscula ou minúscula –, sendo a primeira uma teoria que distingue as entidades no mundo e caracteriza suas naturezas através de categorias de alto nível, como propriedade, qualidade, estado, papel, e outras; e a segunda como um artefato de engenharia da computação.

Para a Ciência da Computação, o termo ontologia relaciona-se à Inteligência Artificial, Engenharia do Conhecimento, Processamento de Linguagem Natural e Representação do Conhecimento, e é visto como a descrição de um domínio desde os seus termos – como um vocabulário – até as regras e descrições lógicas que governam o domínio. Diferente de um vocabulário convencional, a ontologia é caracterizada por uma série de conceituações dos termos de um determinado vocabulário (CHANDRASEKARAN; JOSEPHSON; BENJAMINS, 1999).

A abordagem da ontologia na Inteligência Artificial (CAMPOS, 2011) tem como foco “a resolução de problemas por agentes de software⁷ que fazem inferências sobre bases de conhecimento específicas para um determinado propósito, e geralmente utilizam linguagens lógicas”.

Na Ciência da Informação, o termo ontologia passou a ser utilizado no final dos anos 1990, com o desenvolvimento da Web e dos ambientes digitais (SOERGEL, 1999; VICKERY, 1997). Smith (2003, p. 6) salienta a problemática da ‘Torre de Babel’, afirmando que diferentes grupos de organizadores de conhecimento possuem vocabulários diferentes para expressar suas representações informacionais. Ele afirma que “diferentes bases de dados

⁷ Agente de software: “um agente é uma entidade que executa um conjunto de operações que lhes foram incumbidas por um usuário ou outro programa, com algum grau de independência ou autonomia e, executando estas operações, emprega algum conhecimento dos objetivos ou desejos do usuário” (IBM RESEARCH, 1998).

podem utilizar nomes iguais com diferentes significados; ou o mesmo significado pode ser expresso através de nomes diferentes”.

Gruber (1993) diz que uma “ontologia é uma especificação de uma conceituação”. Para o autor, o termo – emprestado da Filosofia – é uma proposta sistemática de existência, onde o que existe para a máquina é o que pode ser representado. O formalismo declarativo é o responsável por modelar o conhecimento daquele domínio específico, onde o universo do discurso é composto através dos objetos constituintes daquele domínio. O conjunto de objetos e os relacionamentos entre esses objetos são representados através de um vocabulário, que é apresentado em um programa baseado em conhecimento. As definições associam os nomes das entidades ao universo do discurso através de textos legíveis por humanos, e os axiomas formais garantem a restrição interpretativa da máquina.

Chandrasekaran, Josephson e Benjamins (1999) concordam com a definição clássica e entendem ontologia em Inteligência Artificial como

um vocabulário de representação, frequentemente especializado em um domínio ou um assunto. Mais precisamente, não é o vocabulário, como tal, que se qualifica como uma ontologia, mas as conceituações que os termos do vocabulário destinam-se a capturar. (CHANDRASEKARAN; JOSEPHSON; BENJAMINS, 1999).

Gruber (1993, 1995) diz que a conceituação é “um conjunto de relações extensionais que descrevem um estado particular das coisas” e que “é uma visão abstrata e simplificada do mundo que desejamos representar com algum propósito”. Guarino e Giaretta (1995) discordam do autor e apontam a conceituação como uma “estrutura semântica intensional que codifica as regras implícitas que restringem a estrutura de uma porção da realidade”.

Para Campos (2011) conceituação é entendida como as regras implícitas ao entendimento humano que devem ser traduzidas para que possam ser utilizadas pelas máquinas. Almeida e Bax (2003, p. 8) a definem como “uma visão abstrata e simplificada do mundo que se deseja representar”.

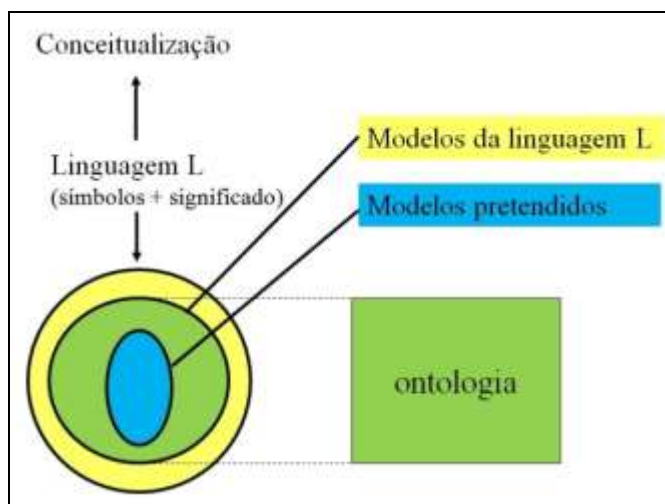
Almeida (2010) salienta a dificuldade em definir o termo ‘conceituação’, se apoiando em Smith (2003) ao afirmar que existem diversas definições conflitantes sobre o termo. Smith (2003, p. 11) afirma que o comportamento de um usuário, ao comprar um seguro, investir em derivativos ou até mesmo

suplicar aos deuses dos nossos ancestrais envolve certo tipo de conceituação, o que quer dizer que “envolve um sistema de conceitos através dos quais o universo de discurso correspondente é dividido em objetos, processos e relações através de diferentes maneiras”.

Ao tratar a conceituação como extensional, afirma-se que ela liga elementos do mundo em uma linguagem lógica, como uma definição por enumeração de elementos; já a conceituação intensional busca restringir os significados dos elementos através da abstração das suas características.

Levando essa diferença em consideração, Guarino (1998) define ontologia como “um artefato de engenharia, constituído por um vocabulário específico usado para descrever uma certa realidade, mais um conjunto de afirmativas que se referem ao significado pretendido das palavras do vocabulário”. Para o autor, ontologias devem levar em consideração o compromisso ontológico, que restringe um modelo pretendido através de uma linguagem lógica. Quanto mais significativas forem as restrições para as possibilidades de interpretação, mais próximo estaremos do modelo pretendido.

Figura 5. Aproximação do modelo pretendido



Fonte: Almeida (2014), adaptado de Guarino (1998)

Guarino e Giaretta (1995) utilizam como exemplo um universo de blocos sobre uma mesa para diferenciar as relações extensionais – aquelas que se referem a um estado particular das coisas – e as relações intensionais –

aquelas que se referem ao significado dessas relações. Para os autores, um arranjo de blocos ‘a’ e um arranjo de blocos ‘b’ não possuem conceituações diferentes – uma vez que continuam sendo arranjos de blocos –, mas o estado das coisas em cada cenário é diferente. Portanto, mais do que entender o estado particular das coisas, é necessário entender o significado das relações e como elas acontecem.

Por isso, o conceito de compromisso ontológico é importante para entender o cenário em que a ontologia está inserida. De acordo com Campos, Campos e Medeiros (2011)

O compromisso ontológico (...) pode ser definido como um acordo firmado por uma comunidade sobre o significado que estabelece e é expresso em uma ontologia, tanto do ponto de vista da compreensão pelo homem quanto do tratamento pela máquina, através dos agentes de software. Isso implica definir o vocabulário de uma forma que venha a minimizar ambiguidades, de modo que seu uso possa ser compartilhado para representar e recuperar conhecimento entre comunidades afins. (CAMPOS; CAMPOS; MEDEIROS, 2011, p. 142).

Para Souza e Alvarenga (2004, p. 137) o termo ontologia foi utilizado no âmbito da Web e Inteligência Artificial como “um documento ou um arquivo que define formalmente as relações entre termos e conceitos”. Essas questões também são discutidas por estudiosos em âmbito internacional (VICKERY, 1997; SOERGEL, 1999; SMITH, 2003; GARCÍA-MARCO, 2007). Nesse sentido, mantiveram semelhanças com as taxonomias e os tesouros no que diz respeito à possibilidade de restrição da linguagem natural.

Apesar dessa semelhança com tesouros, as ontologias permitem um maior número de relações entre diferentes entidades – que, em tesouros, se restringem às relações genérico-específicas, partitivas e associativas –, além de permitir a explicitação dessas relações. Essa explicitação é essencial para que os agentes de software sejam capazes de interpretar os dados e interligá-los a outras bases de dados através da possibilidade de inferências. Desta forma, diferente dos tesouros – que permitem recuperação de documentos e informações –, as ontologias permitem a descoberta de conhecimento.

Ontologias são capazes de potencializar a representação dos relacionamentos existentes entre os conceitos de um determinado domínio. São elas as responsáveis por explicitar formalmente qual é a ligação entre as

entidades descritas, demonstrando de que forma se relacionam; com isso, possibilitam representações semânticas que não poderiam ser obtidas a partir de linguagem natural, e possibilitam novas formas de acesso através da realização de inferências, que aumentam a expressividade na recuperação da informação (VICKERY, 1997; SMITH, 2003; ZAFALON; RAMALHO, 2014).

Uma das grandes vantagens das inferências e do raciocínio automático é a possibilidade de permitir aos agentes de software que entendam e tirem conclusões lógicas sobre os dados existentes. Através das inferências é possível descobrir relações entre dados que se conectam, permitindo a busca de conexões que não são tão simples de ser observadas. Isso também possibilita a elaboração de perguntas mais sofisticadas que o sistema será capaz de responder, desde que também sejam elaboradas dentro de princípios pré-estabelecidos. (USCHOLD; GRUNINGER, 1996; VIEIRA *et al.*, 2005; FONSECA; AZEVEDO; ALMEIDA, 2014).

Para os Dados Interligados Abertos, a definição de conceituação que melhor se enquadra para o trabalho de interligação é a proposta por Guarino e Giaretta (1995). Ao buscar restringir uma porção da realidade e entender os significados das relações, e não apenas o estado particular das coisas – como Gruber (1993) propõe em sua definição extensional de conceituação –, é possível estabelecer de que forma os relacionamentos acontecem em um determinado domínio de conhecimento, e qual seria a melhor forma de interligar as instâncias através dessas relações.

Aliar uma ontologia à perspectiva de Dados Interligados Abertos é uma alternativa que visa facilitar a busca por assuntos que se mostrem correlacionados, uma vez que mecanismos menos complexos, como a atribuição de palavras-chave, traz consigo problemas concernentes à linguagem natural, tais como sinonímias, termos homônimos e ambiguidades. A utilização destas ferramentas de representação do conhecimento possibilita uma melhor estruturação daquilo que se quer representar, assim como restringe os significados que determinado termo pode ter no interior de um dado contexto.

A seguir, o capítulo estrutura-se da seguinte forma: em (5.1) apresentam-se os diferentes tipos de ontologia como vistas pela Ciência da

Informação, suas particularidades e funções; em (5.2) são apresentadas as partes da ontologia e a importância de cada uma delas para a modelagem conceitual e representação do conhecimento; em (5.3) são explicadas as vantagens de se utilizar uma ontologia em um cenário de interligação de informações governamentais, explicitando suas potenciais vantagens; e em (5.4) são apresentados exemplos nos quais as ontologias foram utilizadas para a interligação de dados governamentais abertos.

5.1 Tipos e partes de ontologias

Entender a classificação das ontologias é um exercício importante para que possamos estabelecer a importância que cada uma delas possui dentro de um determinado contexto. Cada tipo de ontologia possui suas limitações e especificidades e, para o entendimento deste trabalho, é fundamental compreender como são classificadas para que, posteriormente, possa ser estabelecido o tipo mais adequado à interligação de dados governamentais no contexto brasileiro.

Dito isto, as ontologias possuem diversas abordagens e formas de classificação. Seu estudo em diferentes áreas do conhecimento faz com que suas definições atendam a especificidades de domínios diferentes. Almeida e Bax (2003) propõem diversas classificações sobre tipologias de ontologia e, no escopo deste trabalho, vamos nos apropriar dos conceitos correspondentes às funções, grau de formalismo e estrutura das ontologias.

No que diz respeito à função, Mizoguchi, Vanwelkenhuysen e Ikeda (1995) dizem que as ontologias podem ser divididas em:

- Ontologias de domínio: reutilizáveis no domínio, fornecem vocabulários sobre conceitos, seus relacionamentos, sobre atividades e as regras que os governam.
- Ontologia de tarefa: fornecem um vocabulário sistemático de termos, especificando tarefas que podem ou não estar no mesmo domínio.

- Ontologias gerais: incluem um vocabulário relacionado a coisas, eventos, tempo, espaço, casualidade, comportamento, funções, etc.

Quanto ao grau de formalismo, Uschold e Gruninger (1996) dizem que as ontologias podem ser:

- Ontologias altamente informais: expressas livremente em linguagem natural.
- Ontologias semi-informais: expressa em linguagem de forma restrita e estruturada.
- Ontologias semiformais: expressa em uma linguagem artificial definida formalmente.
- Ontologia rigorosamente formal: os termos são definidos com semântica formal, teoremas e provas.

Quanto à estrutura, Haav e Lubi (2001) dizem que as ontologias podem ser:

- Ontologias de alto nível: descrevem os conceitos gerais relacionados a todos os elementos da ontologia (espaço, tempo, matéria, objeto, evento, ação, etc) os quais são independentes do problema ou domínio.
- Ontologias de domínio: descrevem o vocabulário relacionado a um domínio, como, por exemplo, medicina ou automóveis.
- Ontologias de tarefa: descrevem uma tarefa ou atividade, como, por exemplo, diagnósticos ou compras, mediante inserção de termos especializados na ontologia.

Já Guarino (1998) diz que ontologias devem ser construídas de acordo com seu nível de generalidade, dividindo-as em ontologias genéricas, de domínio, de tarefa e de aplicação.

As ontologias de alto nível, também conhecidas como ontologias genéricas, de topo ou de fundamentação, caracterizam-se por representar uma visão geral de mundo adaptável para uma grande variedade de tarefas,

domínios e áreas de aplicação. Smith (2003, p. 7) as descreve como “a realização do antigo sonho filosófico de uma Grande Enciclopédia, compreendendo todo o conhecimento através de um único sistema”. Apesar dos obstáculos que dizem respeito ao desenvolvimento desse tipo de ontologia – como uma visão neutra de todas as descrições históricas, o que, por sua vez, exigiria que toda a visão política, religiosa, de crenças e poderes fosse compreendida através de uma lista de categorias –, Smith (2003, p. 7) acredita que esse tipo de ontologia serviria como um tipo de esqueleto para a criação de outras ontologias, com a descrição de relações independentes de domínio, tais como: “tempo, espaço, herança, instanciação, identidade, medidas, quantidade, dependência funcional, processo, evento, atributo, entre outros”.

Como exemplos de ontologias de alto nível, podem ser citadas a Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE); a Suggested Upper Merged Ontology (SUMO); a Unified Foundation Ontology (UFO); e a Basic Formal Ontology (BFO).

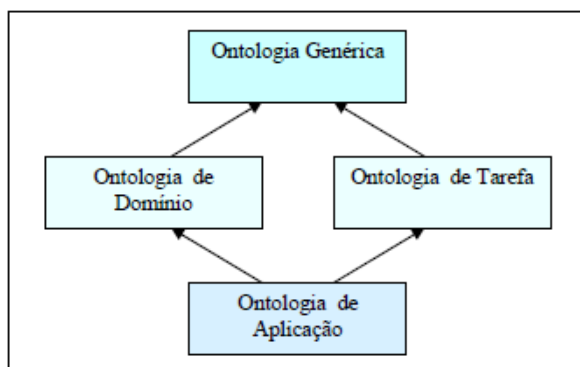
As ontologias de aplicação descrevem conceitos dependendo de um domínio e tarefa específicos, que são especializações das ontologias relacionadas. Essas ontologias têm como objetivo principal a sua implementação prática (GUARINO, 1998).

As ontologias de domínio caracterizam-se por conceituarem um domínio específico do conhecimento, como, por exemplo, ‘construção de rodovias’ ou ‘medicina’.

As ontologias de tarefa descrevem uma atividade ou tarefa genéricas, como ‘diagnóstico’ ou ‘venda’, expressando conceitos para a resolução de problemas.

O esquema de ontologias de acordo com seu nível de generalidade por Guarino (1998) pode ser melhor explicitada no esquema abaixo:

Figura 6. Ontologias de acordo com o nível de generalidade



Fonte: Guarino (1998)

Entender os diferentes tipos de ontologias se mostra importante, neste estudo, para identificar de que forma esses artefatos de organização do conhecimento podem ser utilizados na interligação de informações governamentais. Ao nosso entendimento, a utilização de ontologias de domínio, dentro de cada especificidade informacional daquilo que se deseja representar – se, por exemplo, informações de caráter social, ou econômico, ou financeiro – se mostra como a alternativa mais interessante, uma vez que ela permite configurar especificidades dentro de cada domínio, que, posteriormente, podem ser interligados através de suas relações, explicitando os conceitos que poderão gerar novos conhecimentos de forma automatizada, levando em conta as diferentes maneiras que os domínios tratam os significados dos termos existentes no seu universo de conhecimento.

Da mesma forma que é importante conhecer o tipo de ontologia mais apropriado para o contexto do nosso trabalho, também é importante apresentar a formação da ontologia e a descrição de suas partes, uma vez que, através da exploração de suas classes, propriedades e, principalmente, relações, é possível explicitar informações importantes para a modelagem do domínio através da lógica de dados interligados abertos. Entender as partes de uma ontologia faz com que fique claro em que momento cada uma delas será importante para a construção ou utilização de um sistema de representação do conhecimento que possa ser eficaz no contexto governamental.

Gruber (1993) afirma que

em uma ontologia, definições associam os nomes das entidades no universo do discurso (classes, relações, funções ou outros objetos) com textos descritivos sobre os significados

desses nomes e axiomas formais que contém a interpretação e melhor forma de utilização desses termos. (GRUBER, 1993).

Com isso, divide as partes componentes da ontologia em (a) classes, (b) relações e (c) funções, além de (d) axiomas que interpretam de que forma os termos devem ser utilizados. Martins (2002) *apud* Pickler (2007) diz que uma ontologia requer um vocabulário específico para descrever uma realidade e um conjunto de axiomas lógicos necessários para dar semântica ao significado pretendido por esse vocabulário estruturado. Com isso, é possível dividir as partes da ontologia em:

- Entidades ou classes: descrevem conceitos (elementos de um domínio estudado) e providenciam uma representação lógica;
- Atributos ou funções: descrevem as propriedades das entidades;
- Relações: descrevem as ligações entre os objetos no modelo (entidade e atributos);
- Restrições ou axiomas: condições que o modelador do domínio impõe sobre as entidades, atributos e relações.

Ramalho (2010) também esquematiza e sintetiza as partes de uma ontologia. Além das entidades – que o autor divide entre classes e subclasses, caracterizadas como as coisas do mundo real que são representadas e categorizadas de acordo com suas similaridades –, dos atributos, das relações e dos axiomas, o autor também caracteriza as instâncias, que, para ele, são “os valores das classes e subclasses, constituindo uma representação de objetos ou indivíduos pertencentes ao domínio modelado, de acordo com as características das classes, relacionamentos e restrições definidas”. No âmbito da interligação de dados, as instâncias são fundamentais, uma vez que particularizam uma entidade e possibilitam que ela possa se interligar a várias outras entidades. Essa interligação só é possível através das relações, que formam essas pontes entre as instâncias e as conectam.

5.2 O papel das ontologias na padronização de dados governamentais interligados e abertos

Após apresentar a classificação e as partes de uma ontologia como vista pela Ciência da Informação e Ciência da Computação, é importante, para o escopo deste estudo, entender de que forma esse instrumento terminológico pode auxiliar na representação do conhecimento de informações governamentais, e como ele é capaz de aumentar a precisão semântica na recuperação informacional.

Sales (2007) diz que as ontologias, diferente de tesouros e vocabulários controlados, se destacam pela utilização da linguagem formal para a compreensão da máquina, e essa linguagem é que capacita a interoperabilidade entre sistemas. E complementa, dizendo que

O modelo de relações que aparece em Vocabulários Controlados e Tesouros é um modelo diádico que revela apenas as categorias as quais os conceitos pertencem, por ex: Coisa-propriedade; Material-produto; Processo-resultado, etc.

Em Instrumentos como Ontologias, o modelo diádico se torna inadequado se torna inadequado, pois requer que suas definições/relações sejam explicitadas da maneira mais clara possível, com a finalidade de garantir consistência na estrutura terminológica, e a adoção de algoritmos consistentes para a inferência pela máquina (SALES, 2007, n.p.)

O grande diferencial das ontologias é a explicitação das relações entre diferentes instâncias com uma linguagem específica, que é capaz de ser compreendida tanto semanticamente por humanos quanto matematicamente por máquinas; e, a partir disso, essas relações são capazes de gerar conhecimento através de ligações entre diferentes instâncias que não foram previstas no momento da modelagem informacional.

Enquanto em modelos clássicos, como Tesouros e Vocabulários Controlados, a regra é a utilização de um modelo diádico – ou seja, dois elementos –, nas ontologias os elementos de uma declaração devem ser enunciados, assim como a relação existente entre elas. Enquanto em Tesouros

temos, por exemplo, Coisa-propriedade, em ontologias temos, por exemplo, Coisa-<relação>-propriedade.

Outra função do uso dessas linguagens formais para definir relações em ontologias é a capacidade que elas possuem para se integrar com outras ontologias. A questão da integração de ontologias e suas dificuldades vem sendo discutida na Ciência da Informação com alguma frequência nos últimos anos (CAMPOS, 2005; CAMPOS, 2006; FARINELLI; MELO; ALMEIDA, 2013; SILVA; FARINELLI, 2014; MENDONÇA; ALMEIDA, 2014), e é importante salientar que os dados interligados abertos estão diretamente relacionados a esse assunto, uma vez que uma interligação consistente entre diferentes domínios é facilitada através dessas linguagens.

Como já explicitado no Capítulo 4, o Portal Brasileiro de Dados Abertos expõe oito princípios para publicação de dados abertos – sendo eles: dados completos, primários, atuais, acessíveis, processáveis por máquina, de acesso não discriminatório, em formatos não-proprietários e livres de licença –, e as ontologias são agentes facilitadores que atendem a esses princípios. Através dessas ferramentas representacionais, é possível atender a todos os princípios estabelecidos no Portal:

- Uma ontologia é capaz de representar dados completos e primários, uma vez que pode utilizar URI's que representam uma determinada instância na Web;
- Uma ontologia é capaz de representar dados atuais, contanto que seja constantemente alimentada;
- Uma ontologia é ferramenta importante para tornar a informação acessível de forma mais rápida e criteriosa;
- Uma ontologia preconiza dados processáveis por máquina, já que uma de suas funções é a produção de conhecimento através de inferências;
- Uma ontologia possibilita o acesso a informações de forma não-discriminatória, em formatos não-proprietários e de livre acesso, uma vez que, pela lógica de dados interligados abertos, trabalha com informações disponíveis a todos os cidadãos.

A utilização de ontologias não é encarada como uma alternativa em um universo de diferentes métodos dentro da lógica de dados interligados abertos. Esses artefatos tecnológicos já eram apontados no texto seminal sobre Web Semântica (BERNERS-LEE; HENDLER; LASSILA, 2001) como instrumentos que poderiam diminuir os conflitos terminológicos de diferentes bases de dados e aumentar a eficiência de buscas na Web, através de seu formalismo e de suas regras de inferência.

No Brasil, as discussões sobre as ontologias como mecanismos de representação de informações governamentais têm aumentado nos últimos cinco anos dentro da Ciência da Informação (ALVES; BAX, 2014; FARINELLI; MELO; ALMEIDA, 2013; RODRIGUES; SANT'ANA, 2012; MENEZES; DUARTE, 2011; RIBEIRO; ALMEIDA, 2011), sendo um dos principais aspectos analisados as capacidades de representação desses mecanismos. Para Sales, Campos e Gomes (2008), a representação do conhecimento em ontologias envolve, principalmente, a análise semântica, uma vez que esse artefato visa otimizar a recuperação da informação. Os conceitos da ontologia estão ligados por relacionamentos, e estudar de que maneira essas relações entre as informações acontecem é importante para que se chegue à melhor forma de disponibilizar a informação.

Além disso, é necessário ter em mente a adequação da ontologia para o cenário com o qual se deseja trabalhar. Schiessl e Bräscher (2012) afirmam que escolher uma ontologia tem por base o uso pretendido por uma comunidade de usuários e a aceitação desta por eles, através do compartilhamento de uma determinada visão de realidade, que Guarino (1998) chama de compromisso ontológico – ou seja, um acordo, implícito ou explícito, entre uma comunidade de usuários que está em consonância com suas necessidades.

Esse compromisso é importante porque ontologias representam a visão específica de um domínio através de um vocabulário, e essa representação pode ser contraditória quando entra em contato com outros domínios, tendo em vista questões ligadas a interoperabilidade.

Apesar de ainda incipiente, alguns avanços no sentido de interligação de dados abertos já estão sendo feitos, como é possível observar na seção a

seguir, que trata de esforços nacionais e internacionais no estudo e aplicação de ontologias dentro da lógica governamental de dados interligados abertos.

5.2.1 Exemplos de interligação de dados através de ontologias

Ao longo deste estudo, achamos pertinente apresentar exemplos de aplicações de dados governamentais interligados e abertos que utilizem ontologias como ferramentas de controle de vocabulário para melhor explicitar as relações existentes dentro de um domínio modelado. Acreditamos que esse tipo de exemplo demonstre como as ontologias vêm ajudado no avanço de pesquisas nesse cenário.

Bauer e Kaltenböck (2012) citam o Portal Legislativo de Dados do Reino Unido, que utiliza os padrões preconizados pelo W3C para que os cidadãos possam consumir e reutilizar Dados Interligados Abertos de forma que as buscas possam ser feitas através de um mecanismo de interface intuitiva e simples para um usuário não-especialista. O portal é modelado de forma a incentivar e possibilitar a interligação de novos dados, com os conceitos e ideias utilizadas naquele domínio bem definidas por um estatuto.

Essas definições, criteriosamente articuladas, podem ser utilizadas para conectar termos em um tesouro ou uma ontologia, que melhora drasticamente a eficiência do que está sendo pesquisado. Bauer e Kaltenböck (2012) citam um Kit de Ferramentas Online de Educação para o Desenvolvimento Sustentável⁸, dizendo que esse kit criou um vocabulário controlado de serviços fornecidos por diferentes autoridades e os ligou aos identificadores URI encontrados no domínio modelado do Portal Legislativo de Dados, o que possibilitou que diferentes formas de pesquisar por um mesmo conceito – incluindo sinônimos e até mesmo erros de digitação – fossem reunidas em um único termo, o que potencializou o poder de busca dos dados e, conseqüentemente, toda a experiência de navegação do usuário.

Outro exemplo é o da conversão do portal de dados norte-americano para uma perspectiva de interligação, feita por Ding *et al.* (2010). A ideia da

⁸ Disponível em: <<http://www.esdtoolkit.org>>. Acesso em: 17 mar. 2016.

pesquisa era a de criar uma página onde os dados brutos do governo fossem convertidos de forma a possibilitar a interligação automática entre diferentes fontes de dados. A motivação do projeto partiu da evidência de que as informações disponíveis se encontravam em diversos formatos e, para que fossem convertidos, seria necessário tornar os dados governamentais interligáveis a partir dos processos de conversão e aprimoramento e, posteriormente, interligá-los.

O processo de conversão foi feito através da 'limpeza' de dados brutos, posteriormente preservados em uma representação baseada em RDF. Depois, as bases de dados convertidas foram referenciadas através de URIs, fazendo com que tanto as bases de dados quanto as ontologias por trás delas pudessem ser referenciados por terceiros. Durante o processo de aprimoramento, o foco foi o de extração semântica dos dados governamentais e da representação dessas em URIs relevantes, a interligação de fontes de dados diferentes que possuíam as mesmas URIs mencionadas. Esses passos, feitos de forma tanto automática quanto manual, envolveram adição, derivação, ligação e integração dos dados governamentais, e a pesquisa certificou-se de que os usuários poderiam acessar tanto os dados originais e brutos quanto os dados enriquecidos semanticamente.

Após os processos de enriquecimento semântico, a pesquisa preocupou-se também com a forma de utilização destes dados, através da utilização de ferramentas tanto semanticamente relevantes quanto de tecnologias baseadas na web tradicional. O mecanismo de busca utilizado é baseado em SPARQL para fazer a ponte entre os dados governamentais interligados e as interfaces convencionais de busca e visualização.

O projeto *Data.gov Wiki* foi capaz de publicar, até 2010, cinco bilhões de triplas derivadas dos dados brutos do Data.gov, e uma das conclusões do projeto é que, por mais que a automação seja possível, a curadoria humana ainda é importante para lidar com as exceções que a inteligência artificial não consegue identificar.

Pereira e Marcondes (2014) também apresentam um exemplo no qual informações governamentais abertas foram enriquecidas através da utilização de ontologias. A pesquisa se debruçou em modelar os dados governamentais

abertos brasileiros que dissessem respeito ao domínio do acidente de trabalho, e a utilização da lógica de dados interligados abertos e das ontologias como ferramentas de controle terminológico justificou-se pela pluralidade de formatos nos quais os dados brutos publicados encontravam-se: dos 112 conjunto de dados publicados, foram encontrados formatos como xls, pdf, csv, html, zip, entre outros, o que vai contra as boas práticas de publicação de dados interligados abertos, uma vez que grande parte desses formatos são proprietários e inadequados à interligação.

A pesquisa utilizou categorias de ontologias já existentes, evidenciando o alto potencial de reutilização dessas ferramentas, e elaborou um vocabulário de triplas em RDF, que posteriormente receberam URIs e tiveram seus relacionamentos explicitados com o auxílio do Linked Open Vocabularies (LOV), um portal no qual são compilados vocabulários que tornem explícitas as representações entre as entidades de um determinado domínio.

Como resultado, a pesquisa constatou que a elaboração de um modelo conceitual descrito em RDF fornece um nível semântico mínimo aos dados abertos, o que facilita sua organização, recuperação e reuso por parte dos usuários.

Todas essas pesquisas e modelos propostos para a interligação de dados governamentais abertos são esforços relevantes para aprimorar a organização de um cenário no qual as boas práticas, muitas vezes, não estão sendo postas em prática. Tendo em vista que as ontologias podem, entre outras funções, servirem como ferramentas de controle terminológico, sua utilização diminui não apenas os problemas de duplicação de informação, sinonímias ou erros de digitação, mas também são extremamente relevantes por permitir que as interligações explicitadas possam ser conectadas com outros dados através de inteligência artificial, possibilitando a descobertas de novos conhecimentos que não foram previamente identificados pelo esforço intelectual humano.

Outros exemplos podem ser complementados pelo relato de Cordeiro *et al.* (2011), que citam esforços, ressaltando o tratamento semântico na interligação de dados governamentais abertos.

Os autores relatam as iniciativas do governo espanhol (através da Asociación Española de Linked Data – AELID), que objetiva estimular pesquisas em dados interligados não só na Espanha, mas em toda a Europa, materializando-se através de projetos como o Data-Gov Wiki, desenvolvido pelo Instituto Politécnico de Rensslear, e o GeoLinkedData, desenvolvido pela Universidade de Madri. Esses projetos utilizam informações existentes no Instituto Nacional de Estatísticas e no Instituto Nacional Geográfico da Espanha e as transformam em estruturas padrões de dados interligados abertos, o que possibilita sua interligação consistente.

Cordeiro *et al.* (2011) também relatam discussões da União Europeia sobre um projeto em larga escala de integração de dados de diferentes fontes – que envolvem pesquisadores, empresas e produtores de informação de sete países europeus, que se uniram para a criação do projeto LOD2⁹ – e de discussões a nível global, como, por exemplo, a criação da Associação Internacional para Ontologias e Aplicações (IAOA, em sua sigla em inglês), uma organização internacional liderada pelos EUA que objetiva promover pesquisa e desenvolvimento em ontologias ao redor do mundo, com foco em aspectos de modelagem conceitual e enriquecimento semântico.

O trabalho de Cordeiro *et al.* (2011), como apresentado acima ao discutir o cenário internacional, também problematiza a publicação de dados governamentais em formato aberto e interligado no contexto brasileiro, afirmando que a publicação de dados não acontece de forma padronizada e integrada, o que dificulta a utilização destas informações por parte tanto do governo quanto dos cidadãos.

Os autores dizem que o processo de publicação de informações interligadas na esfera governamental implica a definição e criação de vocabulários, e confirmam a ideia de que ontologias são mecanismos interessantes que possibilitam uma visão mais integrada dos dados e maximizam a capacidade de interoperabilidade semântica entre bases de dados diferentes, bem como entre produtores e consumidores de informação.

Eles também apontam para a tendência do cenário governamental brasileiro de evoluir do atual status de dados abertos para um em que os dados

⁹ Linked Open Data 2 Project. Disponível em: <<http://lod2.eu>>. Acesso em: 23 mar. 2016.

estejam não só disponíveis, mas interligados, e apresentam projetos que têm como objetivo atingir esse fim, como o e-Gov, e-Ping e Consegi (Congresso Internacional Software Livre e Governo Eletrônico) como iniciativas que visam entender de que forma esse cenário está sendo modelado e como ele pode ser aprimorado para que a informação seja bem estruturada.

O trabalho de Cordeiro *et al.* (2011) se mostra interessante para o presente estudo, uma vez que apresenta diferentes abordagens para a publicação de dados governamentais em formato interligado e aberto, bem como suas vantagens e desvantagens. Ao considerá-las, é possível ter uma visão ampla das práticas de estruturação das informações governamentais, suas diferentes formas de execução e suas facilidades e dificuldades.

6 ANÁLISE DA METODOLOGIA ONTOMETRIC À LUZ DAS ABORDAGENS DA INTERLIGAÇÃO DE DADOS GOVERNAMENTAIS ABERTOS

Uma das maiores problemáticas acerca da disponibilização da informação governamental não reside no fato desse tipo de informação existir ou não, nem se ela está ou não acessível ao cidadão comum; mais do que discutir a sua disponibilidade – que é um ponto importante, mas não é objeto direto de estudo deste trabalho –, é necessário discutir *de que forma* essa informação se encontra presente para o acesso de alguém interessado em obtê-la.

A informação governamental, de acordo com a Lei de Acesso à Informação, “deve ser transparente, clara e em linguagem de fácil compreensão” (BRASIL, 2011, Art. 5º), e é necessário entender em que medida esses dados podem ser melhor disponibilizados para que sua recuperação seja mais assertiva e seu processo de busca seja menos dispendioso.

Como já demonstrado em análise ao Portal de Dados do governo federal (seção 4.1.2), as informações brasileiras disponibilizadas em dados abertos encontram-se minimamente estruturadas. Encontrar as informações necessárias é um processo que depende de estratégias de busca que não são simples, o que faz com que o tempo gasto com a exploração do Portal seja longo e, muitas vezes, infrutífero, já que a informação não está descrita de forma que sua recuperação seja ágil e fácil. Com isso, é evidente a necessidade de algum modelo que possa auxiliar na representação dessa informação.

Modelos, de acordo com Almeida, Oliveira e Coelho (2010) representam a realidade de modo simplificado, uma vez que o mundo é complexo e suas relações não conseguem ser abrangidas em sua totalidade por um mecanismo de representação. Ao representar, lida-se com perdas e ambiguidades, com restrições e com escolhas que devem ser levadas em conta para que o domínio consiga ser – mesmo com suas limitações – representado de forma adequada.

A escolha de um modelo está diretamente ligada à natureza da informação que deve se tornar disponível. No caso da informação

governamental, esta deve estar de acordo com a lei a que está submetida, levando em conta aspectos como autenticidade, integridade, primariedade, transparência e facilidade de acesso e utilização, entre outros.

Com a produção cada vez maior de informação na sociedade, se tornou evidente a necessidade de encontrar mecanismos que pudessem representar de forma mais adequada um universo de conhecimento. E, em paralelo, o desenvolvimento das Tecnologias de Informação e Comunicação propiciou avanços nesse sentido, fazendo com que a informação deixasse de ser organizada de uma maneira analógica e passasse a ser esquematizada e organizada em formato digital. E a internet, como meio de troca dessas informações, fez com que elas pudessem ser compartilhadas de uma maneira mais ágil.

Todo esse cenário culminou na ideia de que os mecanismos de representação informacional conhecidos como ontologias fossem interessantes à representação de um domínio de conhecimento. De acordo com Almeida e Bax (2003)

Uma ontologia é criada por especialistas e define as regras que regulam a combinação entre termos e relações de um domínio do conhecimento. Os usuários formulam consultas usando conceitos definidos pela ontologia. O que se busca, em última instância, são melhorias nos processos de recuperação da informação. (ALMEIDA; BAX, 2003, p. 1)

Este capítulo estrutura-se da seguinte maneira: em (6.1) apresentamos uma análise dos estudos de Cordeiro *et al.* (2011), onde são propostas quatro diferentes abordagens para a publicação de dados interligados governamentais e abertos; em (6.2), apresentamos uma análise da metodologia ONTOMETRIC, proposta por Lozano-Tello e Gómez-Pérez (2004), que consiste em uma série de recomendações para a escolha de uma ontologia para ambientes de interligação de dados; em (6.3), são apresentados quadros comparativos nos quais são feitas análises das recomendações da metodologia ONTOMETRIC, aplicando-as às quatro abordagens expostas por Cordeiro *et al.* (2011), com o objetivo de selecionar as recomendações que melhor se adequam a cada uma das abordagens de interligação de dados governamentais.

6.1 Quatro abordagens para publicação de dados interligados abertos no cenário governamental

Dentro da proposta de trabalho de Cordeiro *et al.* (2011), que está ligada ao tratamento semântico dado às informações governamentais e à sua disponibilização, são encontradas quatro abordagens que dizem respeito à estruturação e publicação de dados interligados abertos no cenário governamental. Traçando uma escala que é inversamente proporcional entre facilidade de publicação e relevância semântica, os autores estabelecem as abordagens da seguinte maneira:

A abordagem 1 é caracterizada pela total falta de preocupação com semântica. O processo de publicação dos dados abertos consiste em:

- a) Extrair os dados relevantes das fontes de dados brutos;
- b) Fazer triplas dos dados;
- c) Ligar os dados com outras bases que estejam estruturadas na perspectiva de dados interligados abertos, utilizando relações simples como, por exemplo, *owl:sameAs*¹⁰.

As vantagens da abordagem 1 são a facilidade e a velocidade no processo de publicação dos dados abertos; as desvantagens são a baixa confiabilidade nos dados e nas ligações resultantes desse processo.

A abordagem 2 é caracterizada pela pouca preocupação com semântica. O processo de publicação dos dados abertos consiste em:

- a) Extrair os dados relevantes das fontes de dados brutos;
- b) Identificar o vocabulário com o qual cada dado será anotado;
- c) Fazer triplas dos dados e ligá-los com outras bases de dados.

As vantagens da abordagem 2 são a facilidade e a velocidade no processo de publicação e a necessidade de identificar os instrumentos terminológicos apropriados para sua utilização, que são as ontologias e os vocabulários utilizados por elas no cenário de dados governamentais interligados e abertos.

¹⁰ The built-in OWL property *owl:sameAs* links an individual to an individual. Such an *owl:sameAs* statement indicates that two URI references actually refer to the same thing: the individuals have the same "identity". Disponível em: <<https://www.w3.org/TR/owl-ref/#sameAs-def>>. Acesso em: 5 mar. 2016.

A abordagem 3 é caracterizada pela média preocupação com semântica. Ela se mostra mais complexa ao objetivar resolver o processo de integração ao nível conceitual, independente da tecnologia empregada em sua utilização. O processo de publicação dos dados consiste em:

- a) Pré-processamento semântico, que é desmembrado em:
 - a. Extração do modelo conceitual, ou seja, modelar as fontes de dados brutos e também das bases de dados já existentes com quais conexões serão efetuadas (se o modelo conceitual não estiver disponível);
 - b. Integração do modelo conceitual, ou seja, os conceitos nos quais cada modelo está interligado.
- b) Pré-processamento dos dados extraídos das fontes de dados brutos;
- c) Fazer triplas e criar um *rdf-schema*¹¹ a partir do modelo conceitual estabelecido;
- d) Interligar a base de dados com outras bases através da integração do modelo conceitual.

A vantagem da abordagem 3 é o aumento na confiabilidade potencial das ligações entre os dados e o esquema proposto, e a desvantagem reside no fato de que ainda podem haver problemas semânticos no processo de integração dos modelos conceituais, principalmente por conta da qualidade desses modelos, que podem conter ambiguidades de termos linguísticos que, mesmo que sejam lexicalmente idênticos, possuam significados diferentes por conta do compromisso ontológico diferenciado, o que pode acarretar problemas no momento da interligação entre bases de dados interligados abertos diferentes.

A quarta e última abordagem é muito semelhante a terceira, no entanto sua preocupação com a semântica é altamente levada em consideração, uma vez que seu objetivo principal é deixar explícita a semântica dos dados. Para isso, Cordeiro *et al.* (2011) propõem a utilização de uma ontologia de domínio como base para a integração dos modelos conceituais.

¹¹ RDF Schema is a semantic extension of RDF. It provides mechanisms for describing groups of related resources and the relationships between these resources. Disponível em: <<https://www.w3.org/TR/rdf-schema/>>. Acesso em: 5 mar. 2016.

O tratamento semântico entre as abordagens 3 e 4 são iguais, no entanto a diferença reside no fato de que, no cenário da abordagem 4, os modelos conceituais das bases de dados são mapeados para uma ontologia de domínio bem fundamentada na etapa de pré-processamento semântico. Com o mapeamento, é possível designar maior precisão semântica aos modelos conceituais das bases de dados; e, na etapa de interligação, os esquemas locais das bases de dados são ligados à ontologia de domínio, guiadas através do seu mapeamento, o que ameniza as problemáticas de ambiguidades e fortalece o compromisso ontológico daquele domínio. Com isso, a base de dados é ligada a outras bases em uma grande nuvem de dados interligados abertos, guiada pelo mapeamento entre as bases de dados e a implementação da ontologia de domínio.

De acordo com Cordeiro *et al.* (2011)

A última abordagem procura resolver o problema da expressividade dos dados em LOD. É principalmente baseada na utilização de um nível conceitual e ontologias de domínio bem fundamentadas. A utilização de um nível conceitual é importante porque abstrai os aspectos tecnológicos, provê uma descrição conceitual das bases de dados e aumenta a compreensão humana e as designações semânticas. A utilização de ontologias bem fundamentadas, por sua vez, melhora principalmente a qualidade da representação dos dados em LOD. (CORDEIRO *et al.*, 2011, p. 92, *tradução nossa*)

De forma a melhor expor as diferenças e semelhanças entre as quatro abordagens propostas por Cordeiro *et al.* (2011), o quadro abaixo sintetiza suas características:

Quadro 7. Abordagens de interligação de dados governamentais, síntese

Abordagem 1: Nenhuma preocupação semântica	<ul style="list-style-type: none"> - Total falta de preocupação com semântica; - O processo de publicação dos dados abertos consiste em: <ul style="list-style-type: none"> a) Extrair os dados relevantes das fontes de dados brutos; b) Fazer triplas dos dados; c) Ligar os dados com outras bases que estejam estruturadas na perspectiva de dados interligados abertos, utilizando relações simples. <p>Vantagens: a facilidade e a velocidade no processo de publicação dos dados abertos; Desvantagens: a baixa confiabilidade nos dados e nas ligações resultantes desse processo.</p>
Abordagem 2: pouca preocupação	<ul style="list-style-type: none"> - Pouca preocupação com semântica; - O processo de publicação dos dados abertos consiste em: <ul style="list-style-type: none"> a) Extrair os dados relevantes das fontes de dados brutos;

semântica	<p>b) Identificar o vocabulário com o qual cada dado será anotado; c) Fazer triplas dos dados e ligá-los com outras bases de dados. Vantagens: a facilidade e a velocidade no processo de publicação e a necessidade de identificar os instrumentos terminológicos apropriados para sua utilização, que são as ontologias e os vocabulários utilizados por elas no cenário de dados governamentais interligados e abertos.</p>
Abordagem 3: Média preocupação semântica.	<p>- Média preocupação com semântica; O processo de publicação dos dados consiste em:</p> <ol style="list-style-type: none"> a) Pré-processamento semântico, que é desmembrado em: <ol style="list-style-type: none"> a. Extração do modelo conceitual, ou seja, modelar as fontes de dados brutos e também das bases de dados já existentes com quais conexões serão efetuadas (se o modelo conceitual não estiver disponível); b. Integração do modelo conceitual, ou seja, os conceitos nos quais cada modelo está interligado. b) Pré-processamento dos dados extraídos das fontes de dados brutos; c) Fazer triplas e criar um <i>rdf-schema</i> a partir do modelo conceitual estabelecido; d) Interligar a base de dados com outras bases através da integração do modelo conceitual. <p>Vantagens: aumento na confiabilidade potencial das ligações entre os dados e o esquema proposto; Desvantagens: ainda podem haver problemas semânticos no processo de integração dos modelos conceituais, principalmente por conta da qualidade desses modelos.</p>
Abordagem 4: Alta preocupação semântica	<p>Muito semelhante a terceira, no entanto sua preocupação com a semântica é altamente levada em consideração, uma vez que seu objetivo principal é deixar explícita a semântica dos dados. Propõe-se a utilização de uma ontologia de domínio como base para a integração dos modelos conceituais.</p> <p>Nesta abordagem, os modelos conceituais das bases de dados são mapeados para uma ontologia de domínio bem fundamentada na etapa de pré-processamento semântico. Com o mapeamento, é possível designar maior precisão semântica aos modelos conceituais das bases de dados; e, na etapa de interligação, os esquemas locais das bases de dados são ligados à ontologia de domínio, guiadas através do seu mapeamento, o que ameniza as problemáticas de ambiguidades e fortalece o compromisso ontológico daquele domínio.</p>

Fonte: Elaboração própria, baseado em Cordeiro *et al.* (2011)

As diferentes abordagens sistematizadas no *quadro 6* demonstram que existem diferentes possibilidades de interligar dados governamentais abertos. E essas capacidades possuem vantagens e desvantagens para o processo de enriquecimento semântico do domínio governamental.

Nessas quatro abordagens, por exemplo, podemos citar a vantagem que as abordagens 1 e 2 possuem de serem mais velozes no processo de publicação das informações. No entanto, as duas perdem em possibilidade de enriquecimento semântico do domínio, uma vez que seus processos dão mais valor à velocidade de publicação do que às restrições de vocabulário e

relações, o que faz com que conhecimento gerado automaticamente seja mais passível a erros do que as abordagens 3 e 4.

6.2 Metodologia ONTOMETRIC: uma análise das recomendações para escolha de uma ontologia apropriada

A metodologia ONTOMETRIC, desenvolvida por Lozano-Tello e Gómez-Pérez (2004), consiste em um conjunto de características onde é possível analisar se uma ontologia é adequada ou não àquilo que se deseja modelar. Os autores expõem uma lista de recomendações para a modelagem de uma ontologia qualquer, sem levar em conta seu domínio de conhecimento e suas particularidades, e, a partir deste conjunto de recomendações, aqueles que desejam representar conhecimento através de uma ontologia podem selecionar as que são interessantes ao seu domínio e à sua comunidade, e buscar, a partir daí, ontologias que sejam interessantes para uma posterior utilização.

A metodologia visa apresentar uma sequência de recomendações em uma estrutura denominada pelos autores como multi-nível, objetivando a escolha de elementos que sejam apropriados para uma ontologia dentro do cenário informacional com o qual se deseja trabalhar. Nos níveis superiores das taxonomias em que as recomendações estão estruturadas, há cinco aspectos básicos para o qual qualquer interessado em implementar uma ontologia deve estar atento. São eles

o **conteúdo** da ontologia e a organização de seu conteúdo, a **linguagem** na qual é implementada, a **metodologia** que foi seguida para construí-la, as **ferramentas** de software utilizadas para construir e editar a ontologia e os **custos** que a ontologia necessitará em determinado projeto. (LOZANO-TELLO; GOMÉZ-PÉREZ, 2004, p. 6, *tradução nossa*)

A taxonomia da metodologia ONTOMETRIC é disposta da seguinte maneira: as dimensões (ferramentas, linguagem, conteúdo, metodologia e custos) são divididas em características; as características são subdivididas em fatores; e os fatores são elencados dentro de suas categorias como as recomendações de utilização. Os fatores das características seguem uma hierarquia de intensidade que começam com 'muito baixa' e seguem por

‘baixa’, ‘média’, ‘alta’ e ‘muito alta’ em uma ontologia, enquanto as recomendações dos fatores por vezes seguem o mesmo modelo de intensidade e, em outras ocasiões, se restringem à presença ou não de determinada recomendação na ontologia (‘suporta’ ou ‘não suporta’). Os quadros com as cinco dimensões da metodologia ONTOMETRIC, bem como as explicações de todas as características citadas, se encontram disponíveis no Anexo A.

A motivação do uso desta metodologia em específico mostra-se interessante uma vez que foi a que apresentou uma estrutura mais completa para a análise da ontologia em um cenário de interligação governamental. Através de seus diferentes fatores de recomendação e sua subdivisão em cinco dimensões diferentes, é possível visualizar as diferentes partes de uma ontologia e entender como cada uma delas pode ou não ser interessante para interligar informações no domínio governamental, foco deste estudo.

Para o escopo deste trabalho, optamos por alterar as intensidades (‘muito baixa’, ‘baixa’ [...] ‘muito alta’) dos fatores e das recomendações de Lozano-Tello e Gómez-Pérez (2004) pela presença ou não das mesmas (‘suporta’, ‘não suporta’) dentro das quatro abordagens propostas por Cordeiro *et al.* (2011), uma vez que, para atingir o objetivo geral deste estudo, mostra-se interessante a apresentação da presença das diretrizes, e não a intensidade das mesmas dentro de uma ontologia. Também optamos, como critério de análise, excluirmos as recomendações que não possuíssem explicações, obtidas em (LOZANO-TELLO, 2002). As recomendações existentes na lista cujas explicações não puderam ser encontradas foram retiradas da lista final, uma vez que, ao procurá-las em outras fontes, poder-se-ia incorrer no erro de interpretação de algumas das recomendações propostas pelos autores.

6.3 A integração das abordagens de Cordeiro e da metodologia Ontometric: especificação pelas dimensões

A seguir, as cinco dimensões apresentadas pela metodologia ONTOMETRIC são detalhadas em quadros, onde também estão apresentadas as quatro abordagens para interligação de dados governamentais abertos

propostos por Cordeiro *et al.* (2011). Para a leitura dos quadros, optou-se marcar com um X as recomendações que estão presentes dentro de cada uma das abordagens apresentadas; e, quando não há a presença dessa recomendação, o espaço mantém-se em branco.

Toda a análise das dimensões apresentadas ao longo desta seção é importante para estabelecer que diretrizes são ou não necessárias ao se interligar informações governamentais no cenário brasileiro. A discussão deste trabalho se foca na realidade nacional e, através da leitura do Portal de Dados Abertos, do Programa de governo aberto brasileiro e do Programa de governo eletrônico brasileiro será possível identificar quais das diretrizes apontadas por Lozano-Tello e Gómez-Pérez (2004) melhor se enquadram dentro da perspectiva de interligação de dados interligados governamentais abertos no Brasil, apoiadas por ontologias como agentes no processo de interligação.

6.3.1 Dimensão: Ferramentas

A primeira dimensão da metodologia ONTOMETRIC (LOZANO-TELLO; GOMÉZ-PÉREZ, 2004) consiste nas ferramentas “utilizadas para construir e editar a ontologia” (p. 6), e se divide nos fatores “Capacidades”, “Visualização”, “Edição”, “Interação”, “Aspectos metodológicos”, “Aspectos cooperativos”, “Tradução” e “Integração”. A dimensão ferramenta é a mais relevante dentro do contexto de interação máquina-usuário, uma vez que diz respeito aos aspectos diretamente envolvidos no manuseio humano e na capacidade que a ontologia possui de ser modificada de acordo com as necessidades do domínio.

Quadro 8. Análise da dimensão ‘Ferramentas’ à luz das abordagens propostas por Cordeiro *et al.* (2011)

Dimensão: Ferramentas	Abordagem 1*	Abordagem 2**	Abordagem 3***	Abordagem 4****
CARACTERÍSTICA				
CAPACIDADES (FATOR)				
Uso local	X	X	X	X
Uso em rede	X	X	X	X
Uso baseado em internet	X	X	X	X
Interface clara para usuário	X	X	X	X
Tempo de resposta	X	X	X	X
Confiabilidade			X	X

VISUALIZAÇÃO (FATOR)				
Navegadores mostram informação completa dos termos	X	X	X	X
Navegadores permitem seleção de detalhes		X	X	X
Navegadores mostram taxonomia	X	X	X	X
Navegadores mostram relações implícitas (<i>ad-hoc</i>)	X	X	X	X
EDIÇÃO (FATOR)				
Ferramenta utiliza todas as capacidades de representação da linguagem		X	X	X
Ferramenta permite edição a qualquer momento	X	X	X	X
Ferramenta explicita graficamente a taxonomia		X	X	X
Ferramenta permite a definição de novas relações	X	X	X	X
INTERAÇÃO (FATOR)				
Ferramenta permite uso independente	X	X	X	X
Ferramenta proporciona interfaces de acesso	X	X	X	X
Documentação da utilização de interfaces de acesso			X	X
Interfaces de acesso são de formatos abertos	X	X	X	X
Documentação da programação de interfaces de acesso			X	X
ASPECTOS METODOLÓGICOS (FATOR)				
Ferramenta dá suporte todo o ciclo de vida	X	X	X	X
Ferramenta dá suporte a desenvolvimentos importantes de atividades	X	X	X	X
Ferramenta proporciona documentação sobre processos de construção			X	X
Ferramenta checa a consistência			X	X
ASPECTOS COOPERATIVOS (FATOR)				
Ferramenta cria grupos de trabalho		X	X	X
Ferramenta permite trabalho simultâneo		X	X	X
Ferramenta visualiza ontologias editadas			X	X
Ferramenta visualiza termos editados			X	X
Ferramenta notifica mudanças ao grupo		X	X	X
Ferramenta identifica as mudanças do usuário	X	X	X	X
TRADUÇÃO (FATOR)				
Ferramenta importa de outras linguagens			X	X
Ferramenta importa de linguagens de marcação			X	X
Ferramenta exporta para linguagens			X	X
Ferramenta exporta para			X	X

linguagens de marcação				
Tradução perde o mínimo de semântica				X
Tradução é supervisionada		X	X	X
INTEGRAÇÃO (FATOR)				
Dificuldade de referenciar novos termos				
Ferramenta permite seleção de termos para integração			X	X
Ferramenta checa consistência na integração ou fusão			X	X
Assistência para fusão manual			X	X
Fusão semi-automática			X	X

Legenda: * Abordagem 1: Nenhuma preocupação semântica

** Abordagem 2: Pouca preocupação semântica

*** Abordagem 3: Média preocupação semântica

**** Abordagem 4: Alta preocupação semântica

O fator **capacidade** diz respeito à utilização da ontologia por uma pessoa ou uma comunidade dentro de um contexto de rede, com a utilização baseada em internet. Além disso, leva em conta a interface da ontologia com o usuário e a confiabilidade da mesma no processo de representação informacional. Todas as abordagens se destacam pela possibilidade de utilização seja local, de rede, ou baseada em internet, mas a abordagem 1 e 2, como comentadas por Cordeiro *et al.* (2011), mostram-se com pouco grau de confiabilidade, uma vez que as informações não passam por um tratamento semântico consistente antes de sua publicação.

O fator **visualização** diz respeito à interação ontologia-usuário, sobretudo no que tange à sua navegação. As abordagens são equilibradas no que diz respeito à navegação, no entanto, a primeira abordagem peca ao não permitir a seleção de detalhes, uma vez que as triplas são feitas apenas entre termos e relações entre termos, não sendo possível a visualização de seus significados.

O fator **edição** é autoexplicativo e diz respeito à capacidade de edição que determinada ontologia pode ter. As quatro abordagens são equilibradas nesse fator, com exceção da abordagem 1, que não tem capacidade de representar graficamente a taxonomia, uma vez que consiste apenas em uma lista de termos e suas relações. Ela também não possui a possibilidade de aproveitar todas as capacidades de representação da linguagem, uma vez que suas triplas são simples e suas relações, pouco expressivas.

O fator **interação** diz respeito à interface e à utilização da ontologia por parte do usuário, bem como à documentação das mesmas nesse sentido. Quanto à documentação, apenas as abordagens 3 e 4 se destacam por possuírem esse tipo de característica; as duas primeiras abordagens dão pouca (abordagem 2) ou nenhuma (abordagem 1) importância ao registro documentário de suas ontologias.

O fator **aspectos metodológicos** diz respeito à construção e inserção de novos termos na ontologia, à consistência da inserção destes termos e à documentação referente a eles. As duas primeiras abordagens pecam ao não se preocuparem com documentação e com a checagem da consistência das ontologias – que não passam por nenhum tipo de ferramenta que indica se as relações fazem ou não sentido em um contexto geral.

O fator **aspectos cooperativos** diz respeito à construção da ontologia a partir de múltiplos produtores de ontologias. Nesse aspecto, a abordagem 1, por se tratar de uma lista de termos ligados através de relacionamentos, e por sua velocidade de publicação, tem pouca preocupação com os aspectos cooperativos, sobretudo de trabalho simultâneo. Quanto mais expressiva for a modelagem do conhecimento, maiores serão as preocupações com a consistência do que está sendo inserido na ontologia.

O fator **tradução** diz respeito à manutenção da consistência da ontologia com a inserção de termos em outras línguas, importando-as ou exportando-as de e para outras ontologias. As duas primeiras abordagens não possuem preocupações com tradução de termos, apenas com os relacionamentos entre eles, enquanto as duas últimas abordagens, por serem melhor fundamentadas e passarem por etapa de pré-processamento, possuem maior preocupação com esse aspecto.

O último fator da dimensão 'Ferramentas' é **integração**: nele, estão ligados aspectos referentes à inserção de novos termos e fusão de ontologias. Nas duas primeiras abordagens, não há dificuldade de referenciar novos termos, uma vez que eles possuem pouco ou nenhum tratamento semântico; já nas abordagens 3 e 4, essa inserção é mais trabalhosa, uma vez que deve ser bem fundamentada. No entanto, as abordagens 3 e 4, apesar de se mostrarem mais trabalhosas na questão de inserção de termos, possuem a vantagem de

uma fácil integração com outras ontologias, enquanto as abordagens 1 e 2 (principalmente a 1) possuem deficiência nesse aspecto.

6.3.2 Dimensão: Linguagem

A segunda dimensão da metodologia ONTOMETRIC (LOZANO-TELLO; GOMÉZ-PÉREZ, 2004) consiste na linguagem “na qual [a ontologia] é implementada” (p. 6), e se divide em apenas dois fatores: “Domínio de Conhecimento” e “Mecanismo de Inferência”. A dimensão da linguagem é importante para avaliar as informações inerentes à própria ontologia, como seus termos, o relacionamento entre eles, as facetas, os atributos, as taxonomias, a possibilidade de axiomas, as regras de produção da ontologia e a capacidade de raciocínio e de produção de inferências por parte desse mecanismo informacional.

Quadro 9. Análise da dimensão ‘Linguagem’ à luz das abordagens propostas por Cordeiro *et al.* (2011)

Dimensão: Linguagem	Abordagem 1*	Abordagem 2**	Abordagem 3***	Abordagem 4****
DOMÍNIO DE CONHECIMENTO (FATOR)				
CONCEITOS, INSTÂNCIAS, FATOS, AFIRMAÇÕES				
Permite instâncias nas classes	X	X	X	X
Tem metaclasses		X	X	X
Pode definir classes sem metaclasses	X	X		
Permite fatos	X	X	X	X
Permite afirmações	X	X	X	X
ATRIBUTOS				
Pode definir atributos de classes		X	X	X
Pode definir atributos de instâncias	X	X	X	X
Pode definir atributos locais				X
Pode definir atributos globais				X
Pode definir atributos polimorfos				X
FACETAS				
Possui valores padrões de atributos			X	X
Possui tipos de atributos			X	X
Pode definir cardinalidade em atributos			X	X
Permite conhecimento regulamentado definido			X	X

Permite novas facetas			X	X
RELAÇÕES				
N relações arbitrárias	X			
Pode formar o tipo em relações			X	X
Pode formar o valor em relações			X	X
Possui definição operacional			X	X
TAXONOMIAS				
Contém relação subclasseDe	X	X	X	X
Contém relação nãoEhUmaSubclasseDe	X	X	X	X
Pode definir decomposição exaustiva	X	X	X	X
Pode definir decomposição disjuntiva	X	X	X	X
AXIOMAS				
Permite axiomas independentes			X	X
REGRAS DE PRODUÇÃO				
Cada regra possui um mecanismo em cadeia definido			X	X
Cada regra possui uma prioridade definida			X	X
Procedimentos nas consequências das RP			X	X
Valores seguros nas RP			X	X
MECANISMO DE INFERÊNCIA (FATOR)				
POTENCIAL DE RACIOCÍNIO				
Axiomas mantêm a consistência			X	X
MECANISMO DE INFERÊNCIA				
Mecanismo é completo			X	X
Mecanismo faz classificações automáticas	X	X	X	X
Mecanismo lida com exceções			X	X

Legenda: * Abordagem 1: Nenhuma preocupação semântica

** Abordagem 2: Pouca preocupação semântica

*** Abordagem 3: Média preocupação semântica

**** Abordagem 4: Alta preocupação semântica

O fator **domínio de conhecimento** diz respeito às partes da ontologia e de que forma elas se relacionam com a interligação de dados. A primeira divisão do fator é o de **conceitos, instâncias, fatos, afirmações**, e é possível perceber que está relacionado à capacidade que a abordagem possui de representar instâncias, classes, metaclasses e afirmações. Nesse sentido, a abordagem 1 se destaca pela falta de metaclasses, uma vez que suas representações são feitas através de triplas; as abordagens 3 e 4 se destacam por não permitirem a definição de classes sem metaclasses – como suas

ontologias são bem fundamentadas, é necessário que todo o caminho hierárquico seja percorrido.

A divisão de **atributos** trata das propriedades que a ontologia deve ter. Nesse aspecto, as abordagens 2, 3 e 4 se destacam por permitir tanto atributos de classe quanto de instâncias, enquanto a abordagem 1 não tem a possibilidade de definir atributos de classe, já que suas ligações são feitas diretamente pelas instâncias através de relacionamentos simples.

As divisões **facetadas** e **relações** estão diretamente ligados aos atributos. Nessas divisões, destacam-se a presença das abordagens 3 e 4, que possibilitam a capacidade de uma interligação com facetadas, definição de funções, tipos e valores de relações. A única ressalva fica por conta das relações arbitrárias, que não são possíveis de serem criadas nas abordagens 2, 3 e 4, uma vez que a utilização de ontologias restringe o escopo das relações e só possibilita que sejam feitas quando bem fundamentadas. A criação de relações implícitas não se mostra em nenhuma das abordagens, uma vez que todas preconizam o modelo entidade-relacionamento.

A subdivisão **taxonomia** se adequa quase que totalmente às quatro abordagens, uma vez que as relações *subclasseDe* e *nãoEhSubclasseDe* podem ser estabelecidas em qualquer modelo simples de entidade-relacionamento, possibilitando a existência de hierarquia genérico-específica. No entanto, nas duas primeiras abordagens, uma instância só pode se ligar à outra instância, impossibilitando a existência de heranças múltiplas, uma vez que as abordagens 1 e 2 só possibilitam heranças simples, onde a classe subordinada só pode possuir uma classe ordenada.

Os **axiomas** e as **regras de produção** são as últimas subdivisões do fator 'Domínio de conhecimento'. O primeiro diz respeito às restrições que determinado relacionamento de instâncias pode ter, e o segundo diz respeito à forma com a qual as informações são construídas dentro da lógica ontológica, definindo regras e valores. Nessas duas subdivisões, apenas as abordagens 3 e 4 possibilitam restrições e regras de produção, já que são as únicas que se preocupam com a consistência das informações interligadas.

O segundo fator da dimensão linguagem é o de **mecanismo de inferência**, que se divide em **potencial de raciocínio** e **mecanismo de inferência**. Na primeira divisão, fica claro que a abordagem 1 não possui

nenhum potencial de raciocínio, uma vez que suas ligações são arbitrárias e só interligam duas instâncias com relações simples; da mesma forma, a abordagem 1 também é pobre no que diz respeito às inferências, já que sua capacidade de interligar conhecimentos não previstos por seres humanos é extremamente limitado. A abordagem 2 também se destaca por possuir pouca possibilidade de inferência, mas todas as abordagens se destacam pela possibilidade de classificações automáticas – nas primeiras abordagens, menos consistentes e com maiores possibilidades de ambiguidades, e nas duas últimas abordagens, com um risco menor de erros e um maior comprometimento ontológico.

6.3.3 Dimensão: Conteúdo

A terceira dimensão da metodologia ONTOMETRIC (LOZANO-TELLO; GOMÉZ-PÉREZ, 2004) consiste no “conteúdo da ontologia e na organização do seu conteúdo” (p. 6), e se divide em quatro fatores: “Conceitos”, “Relações”, “Taxonomias” e “Axiomas”. Apesar de a nomenclatura ser idêntica a algumas subdivisões da dimensão da linguagem, aqui o importante é entender como as informações da ontologia se relacionam, e se a presença ou não desses fatores torna a ontologia menos ou mais eficaz.

Quadro 10. Análise da dimensão ‘Conteúdo’ à luz das abordagens propostas por Cordeiro *et al.* (2011)

Dimensão: Conteúdo	Abordagem 1*	Abordagem 2**	Abordagem 3***	Abordagem 4****
CARACTERÍSTICA				
CONCEITOS (FATOR)				
Conceitos essenciais	X	X	X	X
Conceitos essenciais em níveis superiores			X	X
Conceitos devidamente descritos em linguagem natural			X	X
Especificações formais de conceitos coincidem com linguagem natural			X	X
Atributos descrevem conceitos	X	X	X	X
Número de conceitos	X	X	X	X
RELAÇÕES (FATOR)				
Relações essenciais	X	X	X	X
Relações relacionam conceitos		X	X	X

apropriados				
Especificações formais de relações coincidem com linguagem natural	X	X	X	X
Especificação de aridade			X	X
Propriedades formais de relações	X	X	X	X
Número de relações	X	X	X	X
TAXONOMIA (FATOR)				
Várias perspectivas			X	X
NãoEhUmaSubclasseDe apropriadas	X	X	X	X
PartiçõesDisjuntivas apropriadas			X	X
PartiçõesExaustivas apropriadas			X	X
Profundidade máxima			X	X
Média de subclasses		X	X	X
AXIOMAS (FATOR)				
Axiomas resolvem questões			X	X
Axiomas inferem conhecimento	X	X	X	X
Axiomas verificam consistência			X	X
Axiomas não ligados a conceitos			X	X
Número de axiomas			X	X

Legenda: * Abordagem 1: Nenhuma preocupação semântica

** Abordagem 2: Pouca preocupação semântica

*** Abordagem 3: Média preocupação semântica

**** Abordagem 4: Alta preocupação semântica

O primeiro fator da dimensão conteúdo, **conceitos**, diz respeito à capacidade de a ontologia descrever e armazenar conceitos em seu sistema. Apesar de todas as abordagens possibilitarem a descrição de conceitos essenciais – já que as instâncias são descritas através de URIs, é relativamente fácil descrevê-los – e nenhuma ter algum tipo de limitação de número de conceitos, alguns aspectos como subdivisão de conceitos em classes são limitadas às abordagens mais sofisticadas (3 e 4). As especificações formais dos conceitos, que correspondem as descrições em linguagem natural com as restrições da ontologia, também só são possíveis em abordagens que levam em conta o pré-processamento semântico das informações.

O segundo fator, **relações**, fala sobre a quantidade e adequação das relações explícitas entre as instâncias. Assim como o fator anterior, não há limitação de relações.

O terceiro fator, **taxonomias**, fala sobre a quantidade de subclasses e as várias perspectivas que a ontologia pode ter, dependendo da quantidade de classes e subclasses. Nesse sentido, a abordagem 4 é a mais completa, já que

traz uma ontologia de domínio como base de sua modelagem conceitual, determinando previamente as classes e subclasses permitidas nas taxonomias. As outras abordagens pecam pela arbitrariedade de suas interligações, que não necessariamente precisam de taxonomias para serem efetuadas.

O quarto e último fator da dimensão conteúdo, **axiomas**, versa sobre as restrições e a inferência de conhecimento que a ontologia pode ter na recuperação informacional. Todas as abordagens possibilitam a presença de inferências, mas elas são mais ricas a partir do momento em que há um elemento conceitualmente consistente por trás delas. Nesse caso, as abordagens 3 e 4 se destacam mais por possuírem fase de pré-processamento no qual a informação ganha mais riqueza semântica.

6.3.4 Dimensão: Metodologia

A quarta dimensão da metodologia ONTOMETRIC (LOZANO-TELLO; GOMÉZ-PÉREZ, 2004) consiste na “metodologia que foi seguida para construí-la [a ontologia]” (p. 6), e se divide em três fatores: “Precisão”, “Usabilidade” e “Maturidade”. A metodologia diz respeito à construção da ontologia através da inserção dos conceitos, termos e restrições, delimitando as fases de produção, os manuais para a melhor execução das atividades e a relevância e quantidade de ontologias que necessitam ser desenvolvidas para melhor descrever o conhecimento de determinado domínio.

Quadro 11. Análise da dimensão ‘Metodologia’ à luz das abordagens propostas por Cordeiro *et al.* (2011)

Dimensão: Metodologia	Abordagem 1*	Abordagem 2**	Abordagem 3***	Abordagem 4****
PRECISÃO (FATOS)				
Delimitação de fases	X	X	X	X
Especificação de atividades por fase			X	X
Especificação de pessoal por fase			X	X
Especificação de técnicas por fase			X	X
Especificação de produtos finais por fase			X	X
USABILIDADE (FATOR)				
Clareza na descrição de atividades e técnicas	X	X	X	X
Qualidade dos manuais			X	X

Manuais com exemplos completos			X	X
MATURIDADE (FATOR)				
Número de ontologias desenvolvidas	X	X	X	
Número de domínios diferentes	X	X	X	X
Importância de ontologias desenvolvidas	X	X	X	X

Legenda: * Abordagem 1: Nenhuma preocupação semântica

** Abordagem 2: Pouca preocupação semântica

*** Abordagem 3: Média preocupação semântica

**** Abordagem 4: Alta preocupação semântica

No fator **precisão**, todas as abordagens possuem delimitações de fases, mesmo que nas abordagens 1 e 2 essas fases sejam mais simples do que nas abordagens 3 e 4. No entanto, não existe especificação de quais atividades, pessoas ou técnicas são empregadas nas abordagens 1 e 2, que não se preocupam muito com a documentação de suas atividades. Nesse quesito, as abordagens 3 e 4 se destacam.

No fator **usabilidade**, as abordagens 1 e 2 ganham vantagem por serem mais claras em suas técnicas de implementação – como as etapas são mais simples, sua usabilidade é mais fácil –; no entanto, como as abordagens 3 e 4 necessitam de manuais para a sua implementação, a dificuldade pode ser compensada com a utilização de manuais de qualidade e que possuam exemplos completos, que auxiliam na modelagem do conhecimento.

O fator **maturidade** diz respeito ao desenvolvimento e qualidade da ontologia como um todo. A abordagem 4 se destaca por não possuir um grande número de ontologias desenvolvidas – como as ontologias são bem fundamentadas e requerem uma grande especialização, seu desenvolvimento é mais lento em relação a uma ontologia de abordagem mais simples. Na quantidade de domínios, as abordagens se equilibram, uma vez que suas metodologias podem ser utilizadas abordagem 4 também se destaca por não possuir inconsistências, já que possui a ontologia de domínio que leva em conta o compromisso ontológico e diminui as ambiguidades. Em todas as abordagens, as ontologias desenvolvidas são importantes para a interligação de dados, já que são a base das ligações.

6.3.5 Dimensão: Custos

A quinta e última dimensão da metodologia ONTOMETRIC (LOZANO-TELLO; GOMÉZ-PÉREZ, 2004) consiste nos “custos que a ontologia necessitará em determinado projeto” (p. 6), e se divide em quatro fatores: “Licenças de uso da ontologia”, “Custos estimados de hardware e software”, “Custos de interface de acesso” e “Licenças de uso das ferramentas da ontologia”. A dimensão de custos, apesar de ser a que possui menor número de itens, se configura como um dos aspectos mais relevantes para a implementação da ontologia em um sistema de representação do conhecimento, já que os custos estão diretamente ligados à manutenção de pessoal e tecnologia responsável para dar vida ao sistema.

Quadro 12. Análise da dimensão ‘Custos’ à luz das abordagens propostas por Cordeiro *et al.* (2011)

Dimensão: Custos	Abordagem 1*	Abordagem 2**	Abordagem 3***	Abordagem 4****
CARACTERÍSTICA				
Licenças de uso da ontologia (FATOR)			X	X
Custos estimados de <i>hardware</i> e <i>software</i> (FATOR)			X	X
Custos de interface de acesso (FATOR)			X	X
Licenças de uso das ferramentas da ontologia (FATOR)			X	X

Legenda: * Abordagem 1: Nenhuma preocupação semântica

** Abordagem 2: Pouca preocupação semântica

*** Abordagem 3: Média preocupação semântica

**** Abordagem 4: Alta preocupação semântica

Na dimensão de custos, as abordagens 3 e 4 – que, até o momento, se destacaram por possuírem maiores critérios de representação do conhecimento, o que auxilia no compromisso ontológico de uma comunidade e na amenização das ambiguidades do domínio – se destacam negativamente, uma vez que, para sua implementação, são necessários maiores gastos com software, hardware, interfaces, licenças de uso e de ferramentas. No que diz respeito aos custos, as abordagens 1 e 2 se destacam por sua simplicidade e por sua facilidade de implementação.

A partir da análise apresentada, foi possível estabelecer quais seriam as diretrizes que melhor se adequariam a uma interligação de dados governamentais abertos, utilizando as ontologias como agentes neste processo de interligação.

7 DIRETRIZES PARA A APLICAÇÃO DE ONTOLOGIAS NA INTERLIGAÇÃO DE DADOS GOVERNAMENTAIS ABERTOS

A natureza das informações eletrônicas do governo brasileiro necessita de mecanismos que auxiliem na sua melhor recuperação. A Lei de Acesso à Informação (BRASIL, 2011) preconiza que as informações sejam claras e de fácil acesso, no entanto é possível perceber, ao analisar um portal como o de Dados Abertos¹² como a realidade e a lei entram em desacordo.

A ideia da apresentação das diretrizes – como “recomendações ou atividades a serem realizadas ou checadas” (CAMARGO; VIDOTTI, 2011) – para a utilização de ontologias como um mecanismo auxiliar à recuperação da informação se mostra relevante em um contexto no qual a produção informacional é alta e vêm de diversos locais diferentes. A pluralidade que diz respeito às informações governamentais faz com que os produtores de informação, quando não possuem nenhuma recomendação a qual seguir, as produzam de forma desorganizada, levando em conta suas próprias experiências ao formato no qual esse compartilhamento deve ser feito. Ao se estabelecer diretrizes, objetiva-se obter uma lista de recomendações a partir das quais todos possam se basear, para que, com isso, novas aplicações sejam feitas através das informações, culminando em um compartilhamento mais eficaz.

Este capítulo busca analisar as recomendações e as abordagens de interligação de dados governamentais abertos, propondo um conjunto de diretrizes que sejam mais adequadas à interligação de dados governamentais abertos de qualidade e consistência, levando em consideração as recomendações estabelecidas por Lozano-Tello e Gómez-Pérez (2004), as quatro abordagens de interligação de dados governamentais abertos propostos por Cordeiro *et al.* (2011) e a realidade brasileira, configurada através da Lei de Acesso à Informação (BRASIL, 2011). Ao fim do estudo, o Anexo A apresenta todas as explicações para as recomendações da metodologia ONTOMETRIC, apresentadas em Lozano-Tello (2002), deixando claro o escopo que utilizamos ao definir as diretrizes.

¹² Portal Brasileiro de Dados. Disponível em: <<http://dados.gov.br/>>. Acesso em: 23 mar. 2016.

7.1 Diretrizes da dimensão 'Ferramentas'

As diretrizes da dimensão 'Ferramentas' são compostas pelos seguintes fatores: Capacidades, Visualização, Edição, Interação, Aspectos Metodológicos, Aspectos Cooperativos e Integração.

Na dimensão **Capacidades**, julgamos que os itens 'uso local' e 'uso em rede' fossem excluídos, uma vez que o item 'uso baseado em internet' já possibilita a execução das duas características anteriores.

Nas dimensões **Visualização, Edição e Interação**, julgamos que todos os itens são importantes e devem estar presentes em uma ontologia.

Uma das dimensões propostas por Lozano-Tello e Gómez-Pérez (2004) que não consta na lista de diretrizes da dimensão ferramentas é a de Tradução. Julgamos que esta dimensão não é imprescindível para a publicação de dados governamentais no contexto brasileiro, e, levando em conta o fator de custos, a ferramenta de tradução poderia significar um grande aumento nos preços de manutenção da ontologia, além de poder trazer consigo um sem-número de questões que influem em diversos aspectos do sistema no que diz respeito à integração. Apesar de esse ser um fator interessante para manter as informações a nível global, em um primeiro momento pode ser encarado como não tão importante para atender às demandas específicas da população brasileira. Por isso, foi decidido que, para o contexto de interligação de informações nacionais, sua presença não seja necessária.

Também julgamos que os itens das dimensões **Aspectos metodológicos, Aspectos cooperativos e Integração** são importantes e devem estar presentes em uma ontologia, com exceção da recomendação 'Ferramenta visualiza ontologias editadas', que pressupõe o bloqueio completo de uma ontologia quando uma pessoa estiver editando-a. No contexto de cooperação dos dados governamentais, essa ação dificultaria a edição de uma ontologia.

Quadro 13. Apresentação das diretrizes para utilização de ontologias em ambiente de dados interligados governamentais abertos (Dimensão: Ferramentas)

Dimensão: Ferramentas
CARACTERÍSTICA
CAPACIDADES (FATOR)
Uso baseado em internet
Interface clara para usuário
Tempo de resposta
Confiabilidade
VISUALIZAÇÃO (FATOR)
Navegadores mostram informação completa dos termos
Navegadores permitem seleção de detalhes
Navegadores mostram taxonomia
Navegadores mostram relações implícitas (<i>ad-hoc</i>)
EDIÇÃO (FATOR)
Ferramenta utiliza todas as capacidades de representação da linguagem
Ferramenta permite edição a qualquer momento
Ferramenta explicita graficamente a taxonomia
Ferramenta permite a definição de novas relações
INTERAÇÃO (FATOR)
Ferramenta permite uso independente
Ferramenta proporciona interfaces de acesso
Documentação da utilização de interfaces de acesso
Interfaces de acesso são de formatos abertos
Documentação da programação de interfaces de acesso
ASPECTOS METODOLÓGICOS (FATOR)
Ferramenta dá suporte todo o ciclo de vida
Ferramenta dá suporte a desenvolvimentos importantes de atividades
Ferramenta proporciona documentação sobre processos de construção
Ferramenta checa a consistência
ASPECTOS COOPERATIVOS (FATOR)
Ferramenta cria grupos de trabalho
Ferramenta permite trabalho simultâneo
Ferramenta visualiza termos editados
Ferramenta notifica mudanças ao grupo
Ferramenta identifica as mudanças do usuário
INTEGRAÇÃO (FATOR)
Dificuldade de referenciar novos termos
Ferramenta permite seleção de termos para integração
Ferramenta checa consistência na integração ou fusão
Assistência para fusão manual
Fusão semi-automática

Fonte: Lozano-Tello e Gómez-Pérez (2004)

7.2 Diretrizes da dimensão 'Linguagem'

As diretrizes da dimensão 'Linguagem' são compostas pelos seguintes fatores: Domínio de Conhecimento e Mecanismo de Inferência. O fator Domínio de Conhecimento se desdobra em: Conceitos, Instâncias, Fatos, Afirmações;

Atributos; Facetas; Relações; Taxonomias; Axiomas; e Regras de Produção. O fator Mecanismo de Inferência se desdobra em: Potencial de Raciocínio; e Mecanismo de Inferência.

Na dimensão **Domínio de Conhecimento**, todos os subitens foram mantidos, uma vez que os julgamos importantes para que a ontologia funcione bem dentro do cenário de dados governamentais interligados e abertos. **Conceitos, Atributos e Relações** são a base para a interligação dentro da perspectiva do modelo entidade-relacionamento; as **Taxonomias** são importantes para hierarquizar os termos e a relação que um possui com o outro dentro do domínio, facilitando a visualização; as **Facetas** são importantes para estabelecer as diferentes perspectivas de um mesmo termo e suas diferentes possibilidades de utilização; e os **Axiomas** são o que permitem que as ligações possam produzir conhecimento, diminuindo as ambiguidades através das restrições dos seus significados, e estão intimamente ligados ao mecanismo de inferência da ontologia; e as **Regras de Produção** permitem que valores de verdade ou mecanismos de encadeamento de produção de dados sejam efetuados, trazendo uma rotina que visa a consistência da informação gerada e interligada através das ontologias.

A dimensão **Mecanismo de Inferência** é dividida em **Potencial de Raciocínio e Mecanismo de Inferência** – uma subdivisão que repete o nome da dimensão acima. Nela, julgamos importante que todas as recomendações fossem mantidas, uma vez que suas funções permitem a descoberta de conhecimento não previsto pelo humano no momento da organização das informações disponibilizadas na ontologia.

Quadro 14. Apresentação das diretrizes para utilização de ontologias em ambiente de dados interligados governamentais abertos (Dimensão: Linguagem)

Dimensão: Linguagem
DOMÍNIO DE CONHECIMENTO (FATOR)
CONCEITOS, INSTÂNCIAS, FATOS, AFIRMAÇÕES
Permite instâncias nas classes
Tem metaclasses
Pode definir classes sem metaclasses
Permite fatos
Permite afirmações
ATRIBUTOS
Pode definir atributos de classes

Pode definir atributos de instâncias
Pode definir atributos locais
Pode definir atributos globais
Pode definir atributos polimorfos
FACETAS
Possui valores padrões de atributos
Possui tipos de atributos
Pode definir cardinalidade em atributos
Permite conhecimento regulamentado definido
Permite novas facetas
RELAÇÕES
N relações arbitrárias
Pode formar o tipo em relações
Pode formar o valor em relações
Possui definição operacional
TAXONOMIAS
Contém relação subclasseDe
Contém relação nãoEhUmaSubclasseDe
Pode definir decomposição exaustiva
Pode definir decomposição disjuntiva
AXIOMAS
Permite axiomas independentes
REGRAS DE PRODUÇÃO
Cada regra possui um mecanismo em cadeia definido
Cada regra possui uma prioridade definida
Procedimentos nas consequências das RP
Valores seguros nas RP
MECANISMO DE INFERÊNCIA (FATOR)
POTENCIAL DE RACIOCÍNIO
Axiomas mantêm a consistência
MECANISMO DE INFERÊNCIA
Mecanismo é completo
Mecanismo faz classificações automáticas
Mecanismo lida com exceções

Fonte: Lozano-Tello e Gómez-Pérez (2004)

7.3 Diretrizes da dimensão 'Conteúdo'

As diretrizes da dimensão 'Conteúdo' são compostas pelos seguintes fatores: Conceitos; Relações; Taxonomias; e Axiomas.

Os conceitos, relações, taxonomias e axiomas são importantes nas informações presentes na ontologia, para que assim possam produzir conhecimento de forma criteriosa. Os **conceitos** devem ser essenciais, assim como as **relações**; ambos também devem poder existir em um grande número, para que seja possível traduzir a pluralidade de instâncias no mundo dos dados governamentais interligados e abertos. Os **axiomas** são responsáveis por restringir os significados destes conceitos e relações, para que não existam

problemas referentes a ambiguidades; e as **taxonomias** são essenciais para criar hierarquias de organização, nas quais será possível identificar como os conteúdos estão organizados.

Quadro 15. Apresentação das diretrizes para utilização de ontologias em ambiente de dados interligados governamentais abertos (Dimensão: Conteúdo)

Dimensão: Conteúdo
CARACTERÍSTICA
CONCEITOS (FATOR)
Conceitos essenciais
Conceitos essenciais em níveis superiores
Conceitos devidamente descritos em linguagem natural
Especificações formais de conceitos coincidem com linguagem natural
Atributos descrevem conceitos
Número de conceitos
RELAÇÕES (FATOS)
Relações essenciais
Relações relacionam conceitos apropriados
Especificações formais de relações coincidem com linguagem natural
Especificação de aridade
Propriedades formais de relações
Número de relações
TAXONOMIA (FATOR)
Várias perspectivas
NãoEhUmaSubclasseDe apropriadas
PartiçõesExaustivas apropriadas
PartiçõesDisjuntivas apropriadas
Profundidade máxima
Média de subclasses
AXIOMAS (FATOR)
Axiomas resolvem questões
Axiomas inferem conhecimento
Axiomas verificam consistência
Axiomas não ligados a conceitos
Número de axiomas

Fonte: Lozano-Tello e Gómez-Pérez (2004)

7.4 Diretrizes da dimensão 'Metodologia'

As diretrizes da dimensão 'Metodologia' são compostas pelos seguintes fatores: Precisão; Usabilidade; e Maturidade.

A dimensão de metodologia, apesar de importante, é menos visualizável dentro do cenário brasileiro de utilização de ontologias, sobretudo

governamental. São escassas as ontologias que possuem o caminho de seu desenvolvimento, a utilização de exemplos e de manuais que auxiliem na construção ou escolha de uma ontologia apropriada para determinado domínio. Mantemos as diretrizes de metodologia que dizem respeito à **precisão**, **usabilidade** e **maturidade** porque julgamos importante uma descrição de tudo o que possa auxiliar na seleção de um sistema para o contexto governamental, mas salientamos que os fatores ficam comprometidos, por ser muito difícil visualizá-las na prática dentro do cenário brasileiro, mesmo que sejam fundamentais para garantir a consistência e atualização do sistema.

Quadro 16. Apresentação das diretrizes para utilização de ontologias em ambiente de dados interligados governamentais abertos (Dimensão: Metodologia)

Dimensão: Metodologia
PRECISÃO (FATOS)
Delimitação de fases
Especificação de atividades por fase
Especificação de pessoal por fase
Especificação de técnicas por fase
Especificação de produtos finais por fase
USABILIDADE (FATOR)
Clareza na descrição de atividades e técnicas
Qualidade dos manuais
Manuais com exemplos completos
MATURIDADE (FATOR)
Número de ontologias desenvolvidas
Número de domínios diferentes
Importância de ontologias desenvolvidas

Fonte: Lozano-Tello e Gómez-Pérez (2004)

7.5 Diretrizes da dimensão 'Custos'

As diretrizes da dimensão 'Custos' são compostas pelos seguintes fatores: Licenças de uso de ontologia; Custos estimados de hardware e software; e Custos de interfaces de acesso.

Apesar de ser a dimensão que apresenta o menor número de fatores, os custos são muito importantes ao se implementar uma ontologia, sobretudo no contexto governamental, em que são necessárias aprovações de orçamento e a garantia de que o sistema seja eficaz. Não é possível ignorar os custos em

uma ontologia, no entanto essa mensuração deve ser feita em relação ao custo-benefício.

Nesta dimensão, optamos por excluir o fator 'Licenças de uso das ferramentas da ontologia', uma vez que, de acordo com a lógica de dados interligados abertos, não devem haver custos relacionados a ferramentas de utilização por parte do usuário, que deve ter tudo o que precisa sem que seja necessário pagar por isso.

Quadro 17. Apresentação das diretrizes para utilização de ontologias em ambiente de dados interligados governamentais abertos (Dimensão: Custos)

Dimensão: Custos
CARACTERÍSTICA
Licenças de uso da ontologia (FATOR)
Custos estimados de hw e sw (FATOR)
Custos de interface de acesso (FATOR)

Fonte: Lozano-Tello e Gómez-Pérez (2004)

8 CONSIDERAÇÕES FINAIS

Ao propor um estudo que objetiva apresentar uma lista de diretrizes úteis para que uma ferramenta tecnológica como a ontologia possa funcionar em um ambiente como o de informações governamentais, esbarramos em uma série de desafios que, a princípio, parecem pequenos, mas que, ao serem discutidos, tomam grandes proporções e se tornam centrais durante questionamentos sobre a melhor aplicação das ontologias para o contexto governamental.

Uma das primeiras considerações a serem feitas sobre o processo de produção desta pesquisa é o de que ela, mesmo que tenha tentado estabelecer um resultado sólido e fundamentado, é apenas o início do que pode se transformar em algo diferente com o passar dos anos. Uma lista de diretrizes tais como as que foram apresentadas estão à mercê das mudanças que o tempo e o impacto dos avanços tecnológicos causarão à sociedade futura. Estudar um tema cujo centro está na discussão de um artefato computacional é, na verdade, um exercício em constante mutação, uma vez que as tecnologias da computação se desenvolvem de forma rápida e novas necessidades estão constantemente sendo colocadas em pauta de acordo com sua evolução.

O universo de Dados Interligados Abertos e do Projeto Web Semântica proposto por Berners-Lee, Hendler e Lassila (2001) é ambicioso e se desenvolve paulatinamente, com esforços de pequenos coletivos que tentam fazer com que a ideia se transforme em realidade. Atualmente, o projeto avança através de pesquisas e desenvolvimento nos mais diferentes domínios que vão além do governamental. Áreas como engenharia, medicina e educação desenvolvem suas aplicações com um olhar que objetiva interligar suas informações, em um movimento interdisciplinar.

No entanto, essa interligação esbarra em algumas problemáticas. A primeira e a principal delas está no campo da linguagem: cada domínio possui suas próprias nomenclaturas, relações e definições para termos. E, de um campo para outro, a mesma palavra pode ter significados diferentes. As ontologias, nesse contexto, funcionam como elementos responsáveis pela

restrição do significado de um dado conjunto de termos, mas interligá-los se torna um exercício dispendioso, porque as relações responsáveis por essas interligações também precisam concordar entre os domínios, o que nem sempre é possível.

Órgãos como o World Wide Web Consortium (WC3) buscam funcionar como mediadores na definição de meta relações que possam ser utilizadas em diferentes domínios e desenvolvendo linguagens que objetivam a utilização unificada para a interligação de informações. A presença desse tipo de mediador é importante para que o projeto Web Semântica possa se estender e persistir a longo prazo, renovando-se constantemente e estabelecendo relações que possam ser aproveitadas por diferentes domínios.

Este trabalho propôs um movimento de observação de um domínio e uma alternativa para que suas informações pudessem ser melhor disponibilizadas. Ao estabelecer um paralelo entre dois diferentes estudos – a apresentação de quatro diferentes abordagens de publicação de dados interligados governamentais, de acordo com os estudos de Cordeiro *et al.* (2011), e a utilização da metodologia ONTOMETRIC, elaborada por Lozano-Tello (2002) e exposta resumidamente em (LOZANO-TELLO; GOMÉZ-PÉREZ, 2004) –, tivemos por objetivo central alcançar uma lista de diretrizes de utilização de ontologias em cenários governamentais – tendo como elemento norteador a legislação brasileira – e chegamos a algumas considerações.

Com o auxílio das designações da Lei de Acesso à Informação (BRASIL, 2011), foi possível analisar quais são as necessidades específicas do contexto brasileiro e, a partir delas, selecionar quais características podem ser encaradas como diretrizes para uma melhor configuração do cenário de informações governamentais brasileiras.

A metodologia ONTOMETRIC propõe cinco dimensões diferentes que dizem respeito à natureza das ontologias, sendo elas: conteúdo, linguagem, ferramentas, metodologia e custos. E, analisando essas dimensões e fazendo um paralelo com as quatro abordagens de interligação de dados governamentais apresentadas por Cordeiro *et al.* (2011), chegamos à primeira conclusão: a abordagem 4, apesar de mostrar a maior fundamentação conceitual e, em um primeiro momento, partir do pressuposto de que seria a mais adequada para a interligação de dados governamentais no Brasil, esbarra

em problemáticas que dizem respeito à realidade brasileira, sobretudo orçamentária e de especialização de pessoal.

O domínio governamental é plural e interdisciplinar, e um dos fundamentos da abordagem 4 é a utilização de uma ontologia de domínio – ou uma série delas – para dar maior precisão à recuperação informacional e restringir os significados, o que seria extremamente ambicioso, dada a realidade brasileira. Chegamos à conclusão de que a abordagem 4 envolveria custos que iriam contra a ideia de liberdade financeira proposta pelos dados governamentais interligados e abertos.

Com isso, consideramos que a melhor abordagem identificada para a utilização de ontologias seria a abordagem 3, que se assemelha à abordagem 4, mas não necessita de uma ontologia de domínio ou uma série delas. Apesar de algumas perdas e da menor capacidade conceitual apresentada por essa abordagem, ela se configura como mais adequada, levando em conta aspectos de multiplicidade de domínios, custos e treinamento de pessoal.

A ideia das diretrizes funciona como um guia para que futuros idealizadores ou implementadores de ontologias possam ter um modelo inicial a partir do qual seguir no momento de escolha de um sistema informacional que tenha por objetivo organizar informações digitais de caráter governamental. Por se tratar de um primeiro movimento, esta lista é uma sugestão que deverá sempre estar em modificação, levando em consideração as opiniões dos que trabalham modelando determinado domínio de conhecimento ou na construção informática das ontologias.

Esse modelo pode se estender para outros domínios: o que se iniciou com uma pesquisa em um cenário governamental pode se tornar algo maior, envolvendo aspectos culturais, econômicos, políticos e sociais. Para perspectivas futuras, trabalhar em domínios cujas informações necessitam de cada vez mais especialização e pluralidade de vozes pode se mostrar uma tarefa desafiadora e interessante, uma vez que deveria levar em conta não só especificações do sistema computacional e do compromisso ontológico do domínio, mas também as diversas falas – muitas vezes completamente antagonistas – em um mesmo domínio de conhecimento.

REFERÊNCIAS

ALBANO, C. S. **Dados governamentais abertos**: proposta de um modelo de produção e utilização de informações sob a ótica conceitual da cadeia de valor. Tese (Doutorado). Universidade de São Paulo, 2014.

ALLEMANG, D.; HENDLER, J. **Semantic web for the working ontologist**: effective modeling in RDFS and OWL. 2nd. ed. Amsterdam: Morgan Kaufmann, 2011.

ALMEIDA, M. B. Uma abordagem integrada sobre ontologias: Ciência da Informação, Ciência da Computação e Filosofia. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 19, n. 3, p. 242-258, jan./set. 2014.

ALMEIDA, M. B.; BAX, M. P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. **Ci. Inf.**, Brasília, v. 32, n. 3, p. 7-20, set./dez. 2003.

ALMEIDA, M. B.; OLIVEIRA, V. N. P.; COELHO, K. C. Estudo exploratório sobre ontologias aplicadas a modelos de sistemas de informação: perspectivas de pesquisa em Ciência da Informação. **Enc. Bibli.**, Florianópolis, v. 15, n. 30, p. 32-56, 2010.

ALVES, M. V. C.; BAX, M. P. Da Necessidade e Viabilidade da Adoção do Padrão LOD pelo Congresso Nacional. **Inf. & Soc.:Est.**, João Pessoa, v.24, n.1, p. 73-94, jan./abr. 2014.

AROUCK, O. **Atributos de qualidade da informação**. (2011). 117 f. Dissertação (Mestrado) - Faculdade de Ciência da Informação, Universidade de Brasília, Brasília, 2011. Disponível em:<<http://hdl.handle.net/10482/9501>>. Amaral, Sueli Angélica do (orient.). Acesso em: 23. Mar. 2015.

AROUCK, O.; AMARAL, S. A. do. Atributos de qualidade da informação e a lei de acesso à informação. In: Congresso Brasileiro de Biblioteconomia, Documento e Ciência da Informação. 25., 2013. **Anais...** Florianópolis, SC, 2013. Disponível em: <<http://portal.febab.org.br/anais/article/viewFile/1610/1611>>. Acesso em 23. mar. 2015.

AUER, S. Introduction to LOD2. In: AUER, S. et. al. **Linked Open Data – Creating Knowledge Out of Interlinked Data. Lecture Notes in Computer Science**, Springer, 2014.

BAUER, F.; KALTENBÖCK, M. **Linked Open Data**: The Essentials. A quick start guide for decision makers. Viena: edition mono, 2012. Disponível em: <<http://www.semantic-web.at/LOD-TheEssentials.pdf>>. Acesso em: 13. jan. 2015.

BERNERS-LEE, T. **Linked data** – design issues, 2006. Disponível em: <<http://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: nov. 2014.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. **The semantic web**. Scientific America, Maio 2001. Disponível em: <<http://www.cs.umd.edu/~golbeck/LBSC690/SemanticWeb.html>>. Acesso em: 13. jan. 2015.

BIZER, C.; CYGANIAK, R.; HEATH, T. **How to publish linked data on the web**. 2006. Disponível em: <<http://wifo503.informatik.unimannheim.de/bizer/pub/LinkedDataTutorial/#RDFlinks>>. Acesso em: 13. jan. 2015.

BIZER, C.; HEATH, T.; BERNES-LEE, T. Linked Data. The story so far. Preprint to the special issue on linked data. **International Journal on Semantic Web and Information Systems (IJSWIS)**, 2009.

BRASIL. Lei nº 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei no 8.112, de 11 de dezembro de 1990; revoga a Lei no 11.111, de 5 de maio de 2005, e dispositivos da Lei no 8.159, de 8 de janeiro de 1991; e dá outras providências. **Diário Oficial [da] União**, Brasília, DF, 18 dez. 2011. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2011/Lei/L12527.htm>. Acesso em: 15 mar. 2015.

_____. Tribunal de Contas da União (TCU). **5 motivos para a abertura de dados na Administração pública**. Brasília, 2015.

BREITMAN, K. **Web Semântica: a internet do futuro**. Rio de Janeiro: LTC, 2014.

CAMARGO, L. S. A.; VIDOTTI, S. A. B. **Arquitetura da informação: uma abordagem prática para o tratamento de conteúdo e interface em ambientes informacionais digitais**. São Paulo: GEN, 2011.

CAMPOS, L. M. **Diretrizes para definição de recorte de domínio no reuso de ontologias biomédicas: uma abordagem interdisciplinar baseada na análise do compromisso ontológico**. Orientadoras: Maria Luiza de Almeida Campos e Maria Luiza Machado Campos. Niterói: UFF/IBICT, 2011. Tese (Doutorado em Ciência da Informação).

CAMPOS, L. M.; CAMPOS, M. L. A. Aplicação de dados interligados abertos apoiada por ontologia. In: Encontro Nacional de Pesquisa em Pós-Graduação em Ciência da Informação. 15., 2014, Belo Horizonte. **Anais...** Belo Horizonte: UFMG, 2014.

CAMPOS, M. L. A. A problemática da compatibilização terminológica e a integração de ontologias: o papel das definições conceituais. In: Encontro Nacional de Pesquisa em Pós-Graduação em Ciência da Informação. 6., 2005, Florianópolis, SC. **Anais...** Florianópolis, SC: UFSC, 2005.

_____. Integração de ontologias: o domínio da bioinformática e a problemática da compatibilização terminológica. In: Encontro Nacional de Pesquisa em Pós-

Graduação em Ciência da Informação. 7., 2006, Marília, SP. **Anais...** Marília, SP: UNESP, 2006.

CAMPOS, M. L. de A.; CAMPOS, L. M.; MEDEIROS, J. da S. A representação de domínios do conhecimento e uma teoria de representação: a ontologia de fundamentação. **Inf. Inf.**, Londrina, v. 16, n. 3, p. 140 – 164, jan./jun. 2011.

CHANDRASEKARAN, B.; JOSEPHSON, J. R.; BENJAMINS, V. R. What are ontologies, and why do we need them? **IEEE Intelligent Systems**, v. 14, n. 1, p. 20-26, jan. 1999.

CHATEAUBRIAND, O. A filosofia, a linguagem e o mundo. In: BRITO, A. N.; VALE, O. A. **Filosofia, linguística, informática**: aspectos da linguagem. Goiânia: Ed. UFG, 1998.

CORDEIRO, K. F. *et al.* An approach for managing and semantically enriching the publication of Linked Open Governmental Data. Workshop de Computação Aplicada em Governo Eletrônico (WCGE). **Simpósio Brasileiro de Banco de Dados (SBDD)**, Florianópolis, SC, 2011.

DAVIES, J; STUDER, R; WARREN, P. **Semantic Web Technologies trends and research in ontology-based systems**. West Sussex: John Wiley & Sons, Ltd, 2006.

DBPEDIA. **Página inicial**. Disponível em: <<http://dbpedia.org/About>>. Acesso em: 13. jan. 2015.

MERRIAM-WEBSTER. **Dicionário**. [Internet], 2004. Disponível em: <<http://www.m-w.com>>. Acesso em: 11 abr. 2016.

DING, L. *et al.* **Data-gov Wiki: Towards Linking Government Data**. [S.I.]: Association for the Advancement of Artificial Intelligence, 2010. Disponível em: <<http://data-gov.tw.rpi.edu/2010/linkedai-2010-datagov.pdf>>. Acesso em: 13 mar. 2015.

DING, Y.; FOO, S.. Ontology research and development. Part 2 – a review of ontology generation. **Journal of Information Science**, v.28, n.4, p. 375-388, 2002.

EAVES, D. **The three laws of open government data**, 2009. Disponível em:<<http://eaves.ca/2009/09/30/three-law-of-open-government-data/>>. Acesso em: 13. jan. 2015.

FARINELLI; MELO; ALMEIDA. O papel das ontologias na interoperabilidade de sistemas de informação: reflexões na esfera governamental In: Encontro Nacional de Pesquisa em Pós-Graduação em Ciência da Informação. 14., 2013, Florianópolis, SC. **Anais...** Florianópolis, SC: UFSC, 2013.

FERREIRA, J. A.; SANTOS, P. L. V. A. da C. O modelo de dados Resource Description Framework (RDF) e o seu papel na descrição de recursos. **Inf. & Soc.: Est.**, João Pessoa, v. 23, n. 2, p. 13-23, maio/ago. 2013.

FONSECA, L. B. R.; AZEVEDO, C. L. B.; ALMEIDA, J. P. A. Mapeando dados governamentais com uma ontologia de organizações. In: LOD Brasil Linked Open Data. 1., 2014, Florianópolis. **Anais...** Florianópolis: UFSC, 2014.

GARCÍA-MARCO, F. J. Ontologías y organización del conocimiento: retos y oportunidades para el profesional de la información. **El profesional de la información**, v. 16, n. 6, p. 541- 550, Nov.-Dez. 2007. Disponível em: <<http://www.elprofesionaldeinformacion.com/contenidos/2007/noviembre/01.pdf>>. Acesso em: 5 mar. 2016.

GIL, A. C. **Métodos e técnicas de pesquisa social**. São Paulo: Atlas, 1989.

GRUBER, T. R. Toward principles for the design of ontologies used for knowledge sharing. **International Journal Human-Computer Studies**, v. 43, 1993, p. 907-928. Disponível em: <<http://tomgruber.org/writing/onto-design.pdf>>. Acesso em: 15. Jan. 2015.

GUARINO, N. Formal ontology and information systems, **Formal Ontology in Information Systems**, edited by N. Guarino. Amsterdam, IOS Press. p. 3-15, 1998. Disponível em:<<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.29.1776>>. Acesso em 15 jan. 2015.

_____. Some Ontological Principles for Designing Upper Level Lexical Resources, In: First International Conference on Language Resources and Evaluation. **Proceedings...** Granada, Espanha, maio. 1998.

GUARINO, N.; GIARETTA, P. **Ontologies and Knowledge Bases: Towards a Terminological Clarification**. In N. Mars (ed.) Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing 1995. IOS Press, Amsterdam: 25-32, 1995.

GUIZZARDI, G. **Ontological foundations for structural conceptual models**. 416f. Tese (PhD in Computer Science) – Twente University of Technology, Twente, Holanda, 2005.

HAAV, H. M.; LUBI, T. L. A survey of concept-based information retrieval tools on the web. In: **Proceedings of east-european conference ADBIS**. 5. 2001.

HEATH, T.; BIZER, C. Linked data: evolving the web into a global data space. **Synthesis lectures on the semantic web: theory and technology**, Lecture #1. United Kingdom: Morgan & Claypool publishers, 2011.

HORROCKS, I. Ontologies and the semantic web. **Communications of the ACM**, v. 51, n. 12, dez. 2008. p. 58-67

IBM RESEARCH. **IBM Corporation**. Intelligent Agents Project at IBM T. J. Watson Research, 1998.

- JANSSEN, M.; CHARALABIDIS, Y.; ZUIDERWIJK, A. **Benefits, adoption barriers and myths of open data and open government**. Information Systems Management, v. 29, p. 258-268, 2012.
- LIMA, M. C. **Monografia**: a engenharia da produção acadêmica. São Paulo: Saraiva, 2004.
- LOZANO-TELLO, A. **Métrica de idoneidad de ontologias**. Tese (Doutorado em Informática). 236 f. Madrid, 2002 – Universidad de Extremadura, Madrid, 2002.
- LOZANO-TELLO, A.; GÓMEZ-PÉREZ, A. Ontometric: A method to choose the appropriate ontology. **J. Datab. Mgmt.**, v. 15, n. 2, p.1-18, 2004.
- MARCONDES, C. H. Linked data – dados interligados abertos – e interoperabilidade entre arquivos, bibliotecas e museus na Web. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, v. 17, n. 34, p.171-192, maio./ago., 2012.
- MARCONDES, C. H.; CAMPOS, M. L. A. Ontologia e web semântica: o espaço da pesquisa em Ciência da Informação. **PontodeAcesso**, Salvador, v. 2, n. 1, p. 107-136, jun./jul. 2008.
- MARTIN, P. A. Organizing Linked Data Quality Related Methods. **Proceedings of the International Conference on Information and Knowledge Engineering IKE'12**, p. 376-382, 2012.
- MEDEIROS, S. A.; MAGALHÃES, R.; PEREIRA, J. R. Lei de Acesso à Informação: em busca da transparência e do combate à corrupção. **Inf. Inf.**, Londrina, v. 19, n. 1, p. 55 – 75, jan./abr. 2014.
- MENDONÇA, F. M.; ALMEIDA, M. B. Princípios metodológicos para desenvolvimento de ontologias: análise das práticas correntes e proposição de melhorias. In: Encontro Nacional de Pesquisa em Pós-Graduação em Ciência da Informação. 15., 2014, Belo Horizonte. **Anais...** Belo Horizonte: UFMG, 2014.
- MENEZES; DUARTE. Web semântica e o governo eletrônico brasileiro: uma nova roupagem para a representação da informação. In: Encontro Nacional de Pesquisa em Pós-Graduação em Ciência da Informação. 12., 2011, Brasília, DF. **Anais...** Brasília, DF: UnB, 2011.
- MILLER, E. An Introduction to the Resource Description Framework. **D-Lib Magazine**, v. 4, n. 5, May, 1998. Disponível em: <<http://www.dlib.org/dlib/may98/miller/05miller.html>>. Acesso em: 13. jan. 2015.
- MIZOGUCHI, R.; VANWELKENHUYSEN, J.; IKEDA, M. Task ontology for reuse of problem solving knowledge. In: **Proceedings of ECAI'94 towards very large knowledge bases**, 1994, Amsterdam. [S. l.]: IOS Press, 1995, p. 46-59.

- NASCIMENTO, R. B.; TROMPIERI FILHO, N.; BARROS, F. G. F. Avaliação da qualidade dos serviços prestados nas unidades de informação universitárias. **Transinformação**, Campinas, v. 17, n. 3, p. 235-251, set./dez. 2005. Disponível em: <<http://periodicos.puc-campinas.edu.br/seer/index.php/transinfo/article/view/688/668>>. Acesso em: 25 mar. 2015.
- PARUNDEKAR, R.; KNOBLOCK, C. A.; AMBITE, J. L. Linking and Building Ontologies of Linked Data. **9th international semantic web conference**. University of Southern California, Information Sciences Institute and Department of Computer Science, 2010.
- PEREIRA, D. V.; MARCONDES, C. H. Modelagem e representação semântica de dados governamentais abertos da previdência social brasileira. In: Encontro Nacional de Pesquisa em Pós-Graduação em Ciência da Informação. 15., 2014, Belo Horizonte. **Anais...** Belo Horizonte: UFMG, 2014.
- PICKLER, M. E. V. Web semântica: ontologias como ferramentas de representação do conhecimento. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 12, n. 1, p. 65-83, jan./abr. 2007.
- RAMALHO, R. A. S. **Desenvolvimento e utilização de ontologias em Bibliotecas Digitais**: uma proposta de aplicação. Tese (Doutorado em Ciência da Informação). 145 f. Marília, 2010 – Universidade Estadual Paulista, 2010.
- RAMALHO, R. A. S.; VIDOTTI, S. A. B. G.; FUJITA, M. S. L. Web semântica uma investigação sob o olhar da Ciência da Informação. **DataGramZero – Revista de Ciência da Informação**. Rio de Janeiro, v. 8, n. 6, dez. 2007.
- RIBEIRO; ALMEIDA. Dados abertos governamentais (open government data): instrumento para exercício de cidadania pela sociedade. In: Encontro Nacional de Pesquisa em Pós-Graduação em Ciência da Informação. 12., 2011, Brasília, DF. **Anais...** Brasília, DF: UnB, 2011.
- RODRIGUES; SANT'ANA. Restrições tecnológicas e de acesso a dados disponíveis sobre destinos de repasses financeiros federais para a saúde pública em ambientes informacionais digitais. In: Encontro Nacional de Pesquisa em Pós-Graduação em Ciência da Informação. 13, 2006, Rio de Janeiro, RJ. **Anais...** Rio de Janeiro, RJ: Fiocruz, 2012.
- SALES, L. F. Relações conceituais para instrumentos de padronização terminológica: um novo modelo para o uso em Ontologias. In: Encontro Nacional de Pesquisa em Pós-Graduação em Ciência da Informação. 8., 2007, Salvador. **Anais...** Salvador: UFBA, 2007.
- SALES, L. F.; CAMPOS, M. L. A.; GOMES, H. E. Ontologias de domínio: um estudo das relações conceituais. **Perspec. Cienc. Inf.**, Belo Horizonte, v. 13, n. 2, p. 62-76, maio/ago. 2008.
- SANT'ANA, R. C. G.; RODRIGUES, F. de A. Visualização de afinidades entre parlamentares mediante dados de votação do Senado brasileiro. **Inf. & Soc.:Est.**, João Pessoa, v.23, n.1, p. 49-59, jan./abr. 2013.

SCHIESSL, M.; BRÄSCHER, M. Ontologia: ambiguidade e precisão. **Enc. Bibli.**, v. 17, n. esp. 1, p. 125-141, 2012.

SILVA, A. C. S.; AMARAL, R. M. do. Qualidade da informação: elaboração de uma sistemática para diagnóstico. In: Congresso Brasileiro de Biblioteconomia, Documento e Ciência da Informação. 25., 2013. **Anais...** Florianópolis, SC, 2013. Disponível em: <<http://portal.febab.org.br/anais/article/view/1553/1554>>. Acesso em: 23. Mar. 2015.

SILVA, S. M.; FARINELLI, F. um roteiro para modelagem conceitual de sistemas de informação baseada em princípios ontológicos. In: Encontro Nacional de Pesquisa em Pós-Graduação em Ciência da Informação. 15., 2014, Belo Horizonte. **Anais...** Belo Horizonte: UFMG, 2014.

SMITH, B. Ontology. **The Blackwell Guide to the Philosophy of Computing and Information**, p. 153-166, 2003. Disponível em: <<http://philpapers.org/archive/SMIO-2.pdf>>. Acesso em: 5 mar. 2016.

SOERGEL, D. The rise of ontologies or the reinvention of classification. **Journal of the American Society for Information Science**, v. 50, n. 12, p. 1119-1120, Oct. 1999. Disponível em: <<http://www.dsoergel.com/cv/B70.pdf>>. Acesso em: 5 mar. 2016.

SOUZA, R. R.; ALVARENGA, L. A Web Semântica e suas contribuições para a ciência da informação. **Ci. Inf.**, Brasília, v. 33, n. 1, p. 132-141, jan./abr. 2004.

USCHOLD, M.; GRUNINGER, M. Ontologies: principles, methods an applications. **Knowledge Engineering Review**, v. 11, n. 2, 1996.

VICKERY, B. C. Ontologies. **Journal of Information Science**, v. 23, n. 4, p. 277-286, 1997.

VIEIRA, R. *et al.* WS: ontologias, lógica de descrição e inferências. In: TEIXEIRA, C.; BARRERE, E.; ABRAÃO, I. (Org.). **Web e Multimídia: Desafios e Soluções (WebMedia 2005 - Minicursos)**. Porto Alegre: SBC, 2005, v. 1, p. 127-167. Disponível em: <<http://www.inf.pucrs.br/~rvieira/cursos/webmidia.pdf>>. Acesso em: 17 mar. 2016.

VILLALOBOS, A. P. de O.; SILVA, D. C. As potencialidades da web semântica para a ciência da informação. **PontodeAcesso**, Salvador, v. 4, n. 2, p. 58-75, set. 2010.

VIVO. **About**, 2012. Disponível em: <<http://vivoWeb.org/about>>. Acesso em: 13. jan. 2015.

W3C ESCRITÓRIO BRASIL. O governo de inovação na Copa 2014: uso de redes sociais e dados abertos. In: Seminário de Inovação em Governo Eletrônico. 2., 2010, Porto Alegre. **Anais...** Porto Alegre: [s.n.], 2010. Disponível em: <http://www.procergs.rs.gov.br/uploads/1285856001W3C_Seminario_Inovacao_eGov_POA_17092010.pdf>. Acesso em: 19 nov. 2014.

W3C. **SPARQL Query Language for RDF**. 2008. Disponível em: <
<http://www.w3.org/TR/rdf-sparql-query/>>. Acesso em: 20. abr. 2015.

ZAFALON, Z. R.; RAMALHO, R. A. S. Ontologia baseada nos FRBR: proposta de aplicação em catálogos online. In: Encontro Nacional de Pesquisa em Pós-Graduação em Ciência da Informação. 15., 2014, Belo Horizonte. **Anais...** Belo Horizonte: UFMG, 2014.

ANEXO A – EXPLICAÇÃO DAS DIMENSÕES ONTOMETRICS (LOZANO-
TELLO, 2002, *tradução nossa*)

DIMENSÃO: FERRAMENTAS

CARACTERÍSTICA	TIPO
CAPACIDADES (FATOR)	Muito baixa, baixa, média, alta, muito alta
Uso local	Suporta, não suporta
Uso em rede	Suporta, não suporta
Uso baseado em internet	Suporta, não suporta
Interface clara para usuário	Muito baixa, baixa, média, alta, muito alta
Tempo de resposta	Muito baixa, baixa, média, alta, muito alta
Confiabilidade	Muito baixa, baixa, média, alta, muito alta
VISUALIZAÇÃO (FATOR)	Muito baixa, baixa, média, alta, muito alta
Navegadores mostram informação completa dos termos	Suporta, não suporta
Navegadores permitem seleção de detalhes	Suporta, não suporta
Navegadores mostram taxonomia	Suporta, não suporta
Navegadores mostram relações implícitas (ad-hoc)	Suporta, não suporta
EDIÇÃO (FATOR)	Muito baixa, baixa, média, alta, muito alta
Ferramenta utiliza todas as capacidades de representação da linguagem	Suporta, não suporta
Ferramenta permite edição a qualquer momento	Suporta, não suporta
Ferramenta explicita graficamente a taxonomia	Suporta, não suporta
Ferramenta permite a definição de novas relações	Suporta, não suporta
INTERAÇÃO (FATOR)	Muito baixa, baixa, média, alta, muito alta
Ferramenta permite uso independente	Suporta, não suporta
Ferramenta proporciona interfaces de acesso	Suporta, não suporta
Documentação da utilização de interfaces de acesso	Muito baixa, baixa, média, alta, muito alta
Interfaces de acesso são de formatos abertos	Suporta, não suporta
Documentação da programação de interfaces de acesso	Muito baixa, baixa, média, alta, muito alta
ASPECTOS METODOLÓGICOS (FATOR)	Muito baixa, baixa, média, alta, muito alta
Ferramenta dá suporte todo o ciclo	Suporta, não suporta

de vida	
Ferramenta dá suporte a desenvolvimentos importantes de atividades	Suporta, não suporta
Ferramenta proporciona documentação sobre processos de construção	Suporta, não suporta
Ferramenta checa a consistência	Suporta, não suporta
ASPECTOS COOPERATIVOS (FATOR)	Muito baixa, baixa, média, alta, muito alta
Ferramenta cria grupos de trabalho	Suporta, não suporta
Ferramenta permite trabalho simultâneo	Suporta, não suporta
Ferramenta visualiza ontologias editadas	Suporta, não suporta
Ferramenta visualiza termos editados	Suporta, não suporta
Ferramenta notifica mudanças ao grupo	Suporta, não suporta
Ferramenta identifica as mudanças do usuário	Suporta, não suporta
TRADUÇÃO (FATOR)	Muito baixa, baixa, média, alta, muito alta
Ferramenta importa de outras linguagens	Suporta, não suporta
Ferramenta importa de linguagens de marcação	Suporta, não suporta
Ferramenta exporta para linguagens	Suporta, não suporta
Ferramenta exporta para linguagens de marcação	Suporta, não suporta
Tradução perda mínima de semântica	Suporta, não suporta
Tradução é supervisionada	Suporta, não suporta
INTEGRAÇÃO (FATOR)	Muito baixa, baixa, média, alta, muito alta
Facilidade de integração	Muito baixa, baixa, média, alta, muito alta
Dificuldade de referenciar novos termos	Muito baixa, baixa, média, alta, muito alta
Ferramenta permite seleção de termos para integração	Suporta, não suporta
Ferramenta checa consistência na integração ou fusão	Suporta, não suporta
Assistência para fusão manual	Suporta, não suporta
Fusão semi-automática	Suporta, não suporta

CAPACIDADES (FATOR)

Uso local: a instalação exclusivamente em modo local supõe um aumento na segurança dos dados armazenados e, na maioria dos casos, são aplicações

mais cômodas de manusear, além de permitir um melhor desempenho do modo gráfico. No entanto, possuem o inconveniente de, uma vez instalados exclusivamente em modo local, não permitirem a construção cooperativa de ontologias.

Uso em rede: a instalação em uma rede de área local permite a uma equipe de trabalho desenvolver uma ontologia de forma cooperativa, mesmo que não permita que essa equipe trabalhe em diferentes localizações geográficas.

Uso baseado em internet: diversos projetos de software geralmente são desenvolvidos por pessoas localizadas em diferentes espaços geográficos. Em alguns casos, a existência de locais preparados para funcionar com Internet ou Intranet, e que permitem trabalhar a ontologia de modo cooperativo, podem ser imprescindíveis.

Interface clara: o usuário deve ter em conta se a equipe de desenvolvimento do projeto vai ter ou não dificuldades para aprender a utilizar o ambiente. E também se as características do ambiente aceleram ou atrasam o processo de adaptação da ontologia. O usuário pode examinar, por exemplo, a clareza dos comandos das aplicações, a uniformidade das opções dos menus, a qualidade dos manuais de ajuda, as ajudas contextuais ou os exemplos de como utilizar as opções.

Tempo de resposta: o usuário deve analisar o tempo que demora para que o ambiente mostre as conclusões ou atualize as modificações. Isso geralmente tem mais relevância em ambientes usados através de Internet, já que, por exemplo, ambientes com grande quantidade de dados gráficos, ou que permitam trabalhar com um grande número de usuários de uma só vez, podem atrasar consideravelmente o tempo de resposta.

Confiabilidade: um mau funcionamento do ambiente do software pode ocasionar um considerável atraso no processo de adaptação da ontologia às necessidades do projeto. Um ambiente recentemente desenvolvido, ou que não tenha sido suficientemente testado, terá menos garantia do que outro mais utilizado. Para determinar essa característica, o usuário pode examinar: a data da última versão do ambiente, o número de ontologias desenvolvidas utilizando o ambiente ou os projetos importantes em que foram utilizados o ambiente para construir ou adaptar as ontologias.

VISUALIZAÇÃO (FATOR)

Navegadores mostram informação completa dos termos: indica se o ambiente mostra ao usuário toda a informação de todos os termos que se incluem na ontologia.

Navegadores permitem seleção de detalhes: indica se o usuário pode selecionar a informação que quiser visualizar sobre os termos.

Navegadores mostram taxonomia: indica se o ambiente permite visualizar a hierarquia de conceitos inclusos na ontologia.

Navegadores mostram relações implícitas (ad-hoc): a visualização gráfica de relações ad-hoc entre conceitos e taxonomias de conceitos oferece uma forma muito intuitiva de mostrar interconexões entre conceitos. Normalmente, nesse tipo de representação gráfica se mostram os conceitos como nós em um gráfico, e as relações entre conceitos como arcos no gráfico.

EDIÇÃO (FATOR)

Ferramenta utiliza todas as capacidades de representação da linguagem: nem sempre os ambientes conseguem aproveitar todas as possibilidades expressivas da linguagem. Por exemplo, na linguagem RDF(S) é possível dar um nome alternativo às classes com a etiqueta *label*; mas a linguagem OilEd 2.2, apesar de poder exportar ontologias para a linguagem RDF(S), não utiliza essa etiqueta.

Ferramenta permite edição a qualquer momento: pode ser que alguns ambientes não permitam alterar os termos uma vez que a ontologia tenha sido traduzida a alguma determinada linguagem de implementação. Por exemplo, a versão 1.0.3 do OntoEdit permite exportar ontologias a FLogic; mas, uma vez realizada essa ação, se um usuário – manualmente ou através de outro ambiente – modifica a ontologia em FLogic, o OntoEdit 1.0.3 não pode acessar a ontologia modificada porque não dispõe da opção de importação a partir dessa linguagem.

Ferramenta explicita graficamente a taxonomia: alguns ambientes permitem definir a hierarquia entre conceitos mediante algum tipo de representação gráfica, normalmente através de arcos representando relações taxonômicas (subclasseDe, partições exaustivas e disjuntivas). Isso oferece ao usuário uma forma rápida e intuitiva de modificar a taxonomia de conceitos.

Ferramenta permite a definição de novas relações: assim como no caso anterior, alguns ambientes permitem fazer modificações de forma gráfica das relações ad-hoc entre conceitos.

INTERAÇÃO (FATOR)

Ferramenta permite uso independente: se refere ao fato de a ontologia poder ser exportada e seu conteúdo poder ser acessado sem necessidade de utilizar nenhum outro ambiente de software. A maioria dos ambientes que possuem esta característica geralmente exportam a ontologia para um arquivo em formato ASCII ou XML.

Ferramenta proporciona interfaces de acesso: se refere a existência de modos de software que sirvam para acessar o conteúdo da ontologia, de forma local ou remota.

Documentação da utilização de interfaces de acesso: os módulos de acesso das ontologias podem ter manuais de referência, exemplos de usos das

interfaces, etc., que ajudem a equipe de desenvolvimento a compreender seu funcionamento, para utilizá-la em seu projeto.

Interfaces de acesso são de formatos abertos: estabelece se os módulos de acesso podem ser conseguidos sem dificuldades, e se seus códigos estão em formato aberto (*opensource*) para, se for o caso, adicionar ou modificar informações primárias de acesso.

Documentação da programação de interfaces de acesso: se refere à existência de manuais adequados, direcionados aos programadores, que permitam modificar os módulos de acesso dos termos da ontologia com maior facilidade.

ASPECTOS METODOLÓGICOS (FATOR)

Ferramenta dá suporte todo o ciclo de vida: será mais fácil para o usuário fazer modificações na ontologia (adicionar, modificar ou eliminar termos) se o ambiente colaborar em todas as atividades que indica a metodologia de desenvolvimento.

Ferramenta dá suporte a desenvolvimentos de atividades importantes: algumas atividades, como a conceitualização, formalização e implementação, são recomendadas de serem realizadas utilizando um ambiente de desenvolvimento de ontologias apropriado, já que podem mostrar os termos existentes em cada momento, indicar ao usuário as opções disponíveis e comprovar a consistência, uma vez que os termos tenham sido modificados.

Ferramenta proporciona documentação sobre processos de construção: no momento de realizar modificações em uma ontologia, é aconselhável dispor de informações sobre as modificações realizadas. Se os ambientes criam estes documentos de forma automática, isso facilitará essa atividade.

Ferramenta checa consistência: ao realizar modificações do modelo conceitual, é muito importante verificar a consistência das modificações realizadas.

ASPECTOS COOPERATIVOS (FATOR)

Ferramenta cria grupos de trabalho: para modificar a ontologia alguns ambientes permitem definir grupos de trabalho, com os devidos direitos de acesso e modificação, informando que integrantes do grupo estão conectados, etc.

Ferramenta permite trabalho simultâneo: o projeto pode requerer que o ambiente permita incluir, modificar ou excluir termos simultaneamente por várias pessoas da equipe. Por exemplo, um dos ambientes que mais favorece o trabalho colaborativo é o *Tadzebao-WebOnto*, que permite o uso de diagramas de discussão.

Ferramenta visualiza ontologias editadas: se refere ao fato de o ambiente de software bloquear totalmente a possibilidade de edição quando uma pessoa do grupo de trabalho já está editando.

Ferramenta visualiza termos editados: alguns ambientes bloqueiam a edição de termos que estão sendo editados por uma pessoa do grupo de trabalho, permitindo a outros membros do grupo modificar outros termos.

Ferramenta notifica mudanças ao grupo: em projetos colaborativos, é importante avisar ao resto dos integrantes do grupo sobre as modificações realizadas utilizando, por exemplo, mensagens de correio eletrônico, uma agenda comum de modificações, etc.

Ferramenta identifica as mudanças do usuário: da mesma forma, é interessante que as ferramentas registrem quem realizou as modificações em cada um dos termos, quando tais modificações foram efetuadas, assim como a evolução das modificações. Desta forma, o trabalho colaborativo é favorecido.

TRADUÇÃO (FATOR)

Ferramenta importa de outras linguagens: pode acontecer de, apesar de haver encontrado uma ontologia cujo conteúdo coincida em grande parte com o conteúdo buscado, o ambiente analisado não possa importar da linguagem em que está implementada.

Ferramenta importa de linguagens de marcação: da mesma forma que no caso anterior, se refere à capacidade do ambiente importar uma ontologia para uma linguagem de marcação *web*.

Ferramenta exporta para linguagens: uma vez que se foi construída a ontologia, pode ser que o projeto requeira a utilização da ontologia em linguagem clássica, diferente daquela presente no ambiente de software.

Ferramenta exporta para linguagens de marcação: da mesma forma que no caso anterior, pode ser que o projeto requeira a utilização de uma ontologia em linguagem de marcação *web*.

Tradução perde mínimo de semântica: se o tradutor estiver bem construído, as únicas perdas de conhecimento se darão devido a uma menor expressividade da linguagem traduzida.

Tradução é supervisionada: normalmente, as traduções se realizam de forma automática; nesta característica, se examina se o ambiente pergunta ao usuário quando encontra diferentes possibilidades de tradução, ou avisa quando detecta se haverá algum tipo de perda de conhecimento.

INTEGRAÇÃO (FATOR)

Facilidade de integração: se o processo de integração utilizando o ambiente é fácil de ser realizado.

Dificuldade de referenciar novos termos: se, ao realizar a integração por meio do software do ambiente, os termos incorporados da outra ontologia têm uma

notação uniforme. Se não for assim, o usuário deverá considerar as notações distintas a serem adicionadas desde sua aplicação até a ontologia totalmente integrada.

Ferramenta permite seleção de termos para integração: é possível que o usuário não deseje integrar todos os termos da ontologia, mas apenas uma parte deles.

Ferramenta checa consistência na integração ou fusão: ao se realizar um processo de integração, pode acontecer de que os termos integrados, juntamente com as novas relações criadas na integração, sejam inconsistentes. No processo de fusão, pode acontecer de, ao considerar ao mesmo tempo as definições dos termos das ontologias que serão fundidas, ocorrerem inconsistências. Será de grande ajuda se alguma ferramenta detectar estas inconsistências, entre outras.

Assistência para fusão manual: para favorecer o processo de fusão, seria conveniente que o ambiente permitisse visualizar, preferivelmente de forma gráfica, a taxonomia das ontologias que se desejam fundir e a ontologia resultado dessa fusão, o que permitiria, facilmente, modificar a nova taxonomia formada.

Fusão semi-automática: para favorecer o processo de fusão, seria conveniente que o ambiente fizesse a fusão semi-automática dos conceitos que tenham certo grau de semelhança. A única ferramenta de software encontrada até o final deste estudo que realiza parte desse processo é a *Chimaera*.

DIMENSÃO: LINGUAGEM

DOMÍNIO DE CONHECIMENTO (FATOR)	Muito baixa, baixa, média, alta, muito alta
CONCEITOS, INSTÂNCIAS, FATOS, AFIRMAÇÕES	Muito baixa, baixa, média, alta, muito alta
Permite instâncias nas classes	Suporta, não suporta
Tem metaclasses	Suporta, não suporta
Pode definir classes sem metaclasses	Suporta, não suporta
Permite fatos	Suporta, não suporta
Permite afirmações	Suporta, não suporta
ATRIBUTOS	Muito baixa, baixa, média, alta, muito alta
Pode definir atributos de classes	Suporta, não suporta
Pode definir atributos de instâncias	Suporta, não suporta
Pode definir atributos locais	Suporta, não suporta
Pode definir atributos globais	Suporta, não suporta
Pode definir atributos polimorfos	Suporta, não suporta
FACETAS	Muito baixa, baixa, média, alta, muito alta
Possui valores padrões de atributos	Suporta, não suporta
Possui tipos de atributos	Suporta, não suporta
Pode definir cardinalidade em	Suporta, não suporta

atributos	
Permite definições operacionais de conhecimento	Suporta, não suporta
Permite novas facetas	Suporta, não suporta
RELAÇÕES	Muito baixa, baixa, média, alta, muito alta
N relações arbitrárias	Suporta, não suporta
Permite definição de relações implícitas (ad-hoc)	Suporta, não suporta
Pode formar o tipo em relações	Suporta, não suporta
Pode formar o valor em relações	Suporta, não suporta
Possui definição operacional	Suporta, não suporta
TAXONOMIAS	Muito baixa, baixa, média, alta, muito alta
Contém relação subclasseDe	Suporta, não suporta
Contém relação nãoEhUmaSubclasseDe	Suporta, não suporta
Pode definir decomposição exaustiva	Suporta, não suporta
Pode definir decomposição disjuntiva	Suporta, não suporta
AXIOMAS	Muito baixa, baixa, média, alta, muito alta
Permite axiomas independentes	Suporta, não suporta
REGRAS DE PRODUÇÃO	Muito baixa, baixa, média, alta, muito alta
Cada regra possui um mecanismo em cadeia definido	Suporta, não suporta
Cada regra possui uma prioridade definida	Suporta, não suporta
Procedimentos nas consequências das RP	Suporta, não suporta
Valores seguros nas RP	Suporta, não suporta
MECANISMO DE INFERÊNCIA (FATOR)	Muito baixa, baixa, média, alta, muito alta
POTENCIAL DE RACIOCÍNIO	Muito baixa, baixa, média, alta, muito alta
Axiomas mantêm a consistência	Suporta, não suporta
MECANISMO DE INFERÊNCIA	Muito baixa, baixa, média, alta, muito alta
Mecanismo é completo	Suporta, não suporta
Mecanismo faz classificações automáticas	Suporta, não suporta
Mecanismo lida com exceções	Suporta, não suporta

CONCEITOS, INSTÂNCIAS, FATOS, AFIRMAÇÕES: conceitos são coleções de objetos do domínio; instâncias, fatos e afirmações são termos utilizados para representar elementos concretos do domínio.

Permite instâncias nas classes: representam elementos de um determinado conceito.

Tem metaclasses: possibilidade de criar classes como instâncias de outras classes.

Pode definir classes sem metaclasses: classes irmãs que não podem ter nenhuma instância em comum.

Permite fatos: os fatos são geralmente utilizados para representar uma relação que se estabelece entre elementos.

Permite afirmações: as afirmações se referem à declaração de um fato por parte de uma instância.

ATRIBUTOS

Pode definir atributos de classes: são atributos próprios da classe.

Pode definir atributos de instâncias: atributos definidos em uma classe que podem ter valores diferentes para cada instância.

Pode definir atributos locais: atributos que afetam a um único conceito.

Pode definir atributos globais: atributos que se declaram uma vez na ontologia e que, diretamente, não se vinculam a nenhum conceito.

Pode definir atributos polimorfos: atributos globais que podem ter comportamento diferente de acordo com o conceito ao qual se referem.

FACETAS

Possui valores padrões de atributos: é o valor atribuído se, ao criar uma instância, não for atribuído nenhum valor a ela.

Possui tipos de atributos: obriga a atribuição de um determinado dado.

Pode definir cardinalidade em atributos: se refere ao número mínimo e máximo de valores que o atributo pode ter.

Permite definições operacionais de conhecimento: será possível associar fórmulas ou regras, que serão ativadas ao ser realizada alguma ação (adicionar, modificar ou excluir) no valor do atributo.

Permite novas facetar: possibilidade de adicionar novas facetar às predefinidas na linguagem.

RELAÇÕES: representam interações entre conceitos do domínio. Geralmente aparecem relações do tipo *subclasseDe*, *parteDe*, *ligadoA*, etc. Podem ser binárias ou n-árias, dependendo dos argumentos que utilizem. As características identificadas são:

N relações arbitrárias: indica se é possível utilizar qualquer número de argumentos nas relações/funções.

Pode formar o tipo em relações: restringe os tipos de argumentos em que aparecem as relações e funções.

Pode formar o valor em relações: possui restrições de integridade para comprovar a correção de valores que os argumentos possuem.

Possui definição operacional: existem definições operacionais associadas às relações e funções para inferir valores dos argumentos por meio de regras, fórmulas ou procedimentos.

TAXONOMIAS: as taxonomias são utilizadas para organizar os conceitos do domínio, usando relações de generalização e especialização. As características identificadas são:

Contém relação subclasseDe: se está definida na linguagem a relação subclasseDe, para especializar conceitos gerais com outros mais específicos.

Contém relação nãoEhUmaSubclasseDe: está definida na linguagem a relação nãoEhUmaSubclasseDe para declarar que uma classe determinada não é especialização de outra classe.

Pode definir decomposição exaustiva: um conjunto X de subclasses de um classe Y é uma partição exaustiva se é uma partição disjunta e não pode existir nenhuma instância de Y que não seja uma classe de X.

Pode definir decomposição disjuntiva: um conjunto X de subclasses da classe Y é disjunta se uma instância só pode ser de uma das classes de X; pode existir uma instância que seja da classe Y e que não seja de nenhuma das classes de X.

AXIOMAS: são fórmulas, ocasionalmente expressas em lógica de primeira ordem, que se utilizam para modelar sentenças que são sempre verdadeiras e para verificar a correção da informação contida na ontologia. Os axiomas também são usados para realizar inferências sobre os conhecimentos armazenados.

Permite axiomas independentes: axiomas identificados com um nome que não se associam a nenhum conceito, mas como um elemento independente na ontologia.

REGRAS DE PRODUÇÃO: são regras que seguem a estrutura Se <<condição>>, então <<ação>>, e se utilizam para expressar heurísticas de condições e ações.

Cada regra possui um mecanismo em cadeia definido: permite definir claramente o mecanismo de encadeamento que se utilizará com as regras, encadeamento para trás, para frente ou ambos.

Cada regra possui uma prioridade definida: permite definir valores de verdade ou certeza nas regras.

Procedimentos nas consequências das RP: permite incluir procedimentos nas consequências, por exemplo, para mudar valores nos atributos dos conceitos.

Valores seguros nas RP: as regras, quando ativadas, atualizam a base de conhecimento adicionando ou eliminando atos, instâncias ou afirmações.

POTENCIAL DE RACIOCÍNIO

Axiomas mantêm a consistência: os axiomas são utilizados para realizar comprovações ou restrições, com o objetivo de manter a consistência da ontologia.

MECANISMO DE INFERÊNCIA: descreve como as estruturas estáticas podem ser utilizadas para levar a frente um processo de raciocínio. A forma com a qual é representado o conhecimento irá condicionar, em grande parte, os métodos de raciocínio que podem ser implementados. As características identificadas nessa dimensão são:

Mecanismo é completo: se a linguagem possui um motor de inferências implementado correto e completo. Correto indica que as soluções alcançadas pelo motor de inferência são sempre certas. Completo se refere a sempre proporcionar uma solução, se ela existir.

Mecanismo faz classificações automáticas: fazer classificações automáticas permite também detectar inconsistências nas taxonomias existentes.

Mecanismo lida com exceções: se o motor de inferências lida com exceções na herança dos atributos. Essa qualidade permite redefinir atributos nas subclasses para assinalar características particulares das subclasses.

DIMENSÃO: CONTEÚDO

CARACTERÍSTICA	TIPO
CONCEITOS (FATOR)	Muito baixa, baixa, média, alta, muito alta
Conceitos essenciais	Muito baixa, baixa, média, alta, muito alta
Conceitos essenciais em níveis superiores	Muito baixa, baixa, média, alta, muito alta
Conceitos devidamente descritos em linguagem natural	Muito baixa, baixa, média, alta, muito alta
Especificações formais de conceitos coincidem com linguagem natural	Muito baixa, baixa, média, alta, muito alta
Atributos descrevem conceitos	Muito baixa, baixa, média, alta, muito alta
Número de conceitos	Muito baixa, baixa, média, alta, muito alta

RELAÇÕES (FATOS)	Muito baixa, baixa, média, alta, muito alta
Relações essenciais	Muito baixa, baixa, média, alta, muito alta
Relações relacionam conceitos apropriados	Muito baixa, baixa, média, alta, muito alta
Especificações formais de relações coincidem com linguagem natural	Muito baixa, baixa, média, alta, muito alta
Especificação de aridade	Muito baixa, baixa, média, alta, muito alta
Propriedades formais de relações	Muito baixa, baixa, média, alta, muito alta
Número de relações	Muito baixa, baixa, média, alta, muito alta
TAXONOMIA (FATOR)	Muito baixa, baixa, média, alta, muito alta
Várias perspectivas	Muito baixa, baixa, média, alta, muito alta
NãoÉUmaSubclasseDe apropriadas	Muito baixa, baixa, média, alta, muito alta
PartiçõesExaustivas apropriadas	Muito baixa, baixa, média, alta, muito alta
PartiçõesDisjuntivas apropriadas	Muito baixa, baixa, média, alta, muito alta
Profundidade máxima	Muito baixa, baixa, média, alta, muito alta
Média de subclasses	Muito baixa, baixa, média, alta, muito alta
AXIOMAS (FATOR)	Muito baixa, baixa, média, alta, muito alta
Axiomas resolvem questões	Muito baixa, baixa, média, alta, muito alta
Axiomas inferem conhecimento	Muito baixa, baixa, média, alta, muito alta
Axiomas verificam consistência	Muito baixa, baixa, média, alta, muito alta
Axiomas não ligados a conceitos	Muito baixa, baixa, média, alta, muito alta
Número de axiomas	Muito baixa, baixa, média, alta, muito alta

CONCEITOS

Conceitos essenciais: o usuário deve buscar na ontologia a ser implementada os conceitos que considera fundamentais para o seu projeto, ou seja, aqueles que está seguro de que vai utilizar em seu sistema. Deve-se ter em conta que, quando se faz uma busca por nome para determinado conceito, podem aparecer sinônimos na ontologia. Por isso, para validar essa característica, não deve-se examinar apenas o nome, mas também os atributos que o conceito

possui e a descrição que se faz em linguagem natural sobre o conceito. O valor que essa característica terá indicará o grau de coincidência dos conceitos da ontologia com as necessidades do projeto.

Conceitos essenciais em níveis superiores: se os conceitos necessários para o projeto estão nos níveis superiores da ontologia, pode-se dizer que poderão ser reutilizados mais termos que aparecem em níveis inferiores de outra taxonomia. Neste último caso, deverá ser realizado um esforço adicional para especializar os conceitos da ontologia às necessidades do projeto.

Conceitos devidamente descritos em linguagem natural: se a definição informal dos conceitos encontrados na ontologia for breve e ambígua, será necessário um esforço adicional para entender sua definição formal. Definições pouco precisas ou ambíguas em linguagem natural dificultarão o uso da ontologia e sua aplicação.

Especificações formais de conceitos coincidem com linguagem natural: se a definição formal dos conceitos dada pela exposição de seus atributos, relações, axiomas, etc. não corresponde com a descrição em linguagem natural do conceito, isso indica que a ontologia não foi construída corretamente.

Atributos descrevem conceitos: deve-se comprovar se os atributos dos conceitos que existem na ontologia são aqueles de que o projeto a ser desenvolvido necessita, já que, caso não sejam, será necessário adicionar ou modificar esses atributos, com a consequente adição de esforço. Além da existência desses atributos, também deve-se examinar se estão bem definidos e se suas facetas correspondem com aquelas que se buscam para o projeto.

Número de conceitos: o usuário deve estimar se o tamanho da ontologia, relativo ao número de conceitos que contém, é adequado para o projeto. Um número pequeno de conceitos a respeito das necessidades do projeto implicará no aumento do esforço para adicionar novos conceitos; um número grande implicará uma ontologia menos maleável, e, talvez, um processo de seleção de conceitos.

RELAÇÕES

Relações essenciais: o usuário deve buscar na ontologia candidata as relações que considere fundamentais. Da mesma forma que com os conceitos, para identificar a relação deve-se examinar não somente o nome que lhe é designada no decorrer da ontologia, mas também a descrição que se faz em linguagem natural da relação, bem como sua definição formal.

Relações relacionam conceitos apropriados: mesmo que uma ontologia tenha uma relação definida, pode ser que ela não proporcione a relação de que se necessita no projeto, ou seja, que a relação está definida de forma diferente da qual se necessita para o projeto.

Especificações formais de relações coincidem com linguagem natural: se as definições informais das relações encontradas na ontologia forem muito curtas

ou ambíguas, será necessário realizar um esforço adicional para entender sua definição formal.

Especificação de aridade: deverá ser comprovado se existe coincidência com o número de conceitos que associa a relação.

Propriedades formais de relações: da mesma forma que ocorre com a aridade, deverão ser examinadas as propriedades formais das relações, e avaliar sua adequação com as necessidades do projeto. Sendo R uma relação, e x_1, \dots, x_n , conceitos, as propriedades matemáticas clássicas que podem ser definidas nas relações [Sta00] são:

- Reflexividade: $\forall x_1, R(x_1, x_1)$
- Irreflexividade: $\forall x_1 : \neg R(x_1, x_1)$
- Simetria: $\forall x_1, x_2, R(x_1, x_2) \rightarrow R(x_2, x_1)$
- Assimetria: $\forall x_1, x_2, R(x_1, x_2) \rightarrow \neg R(x_2, x_1)$
- Antissimetria: $\forall x_1, x_2, R(x_1, x_2) \wedge R(x_2, x_1) \rightarrow x_1 = x_2$
- Transitividade: $\forall x_1, x_2, x_3, R(x_1, x_2) \wedge R(x_2, x_3) \rightarrow R(x_1, x_3)$
- Intransitividade: $\forall x_1, x_2, x_3, R(x_1, x_2) \wedge R(x_2, x_3) \rightarrow \neg R(x_1, x_3)$

Este poderia ser um exemplo típico que o usuário decida detalhar essa característica com o estudo de cada uma das propriedades formais, que seriam consideradas como subcaracterísticas da característica “propriedades formais de relações”.

Número de relações: o usuário deve estimar se o número de relações que contém a ontologia é adequado para o projeto. Caso existam poucas relações para as necessidades do projeto, será necessário um maior esforço para adicionar as relações; caso existam muitas, a ontologia se tornará menos maleável e, talvez, possua uma série de relações inúteis.

TAXONOMIA

Várias perspectivas: pode ser que, para o projeto, necessite-se que a classificação dos conceitos não seja única. Deve ser examinado se a ontologia contém várias relações do tipo *subclasseDe* em um mesmo conceito, e se esta é a classificação de que necessita o sistema.

NãoEhUmaSubclasseDe apropriada: da mesma forma, pode ser interessante ao projeto especificar essa relação, normalmente utilizada para romper uma hierarquia entre conceitos.

PartiçõesDisjuntivas apropriada: uma partição disjuntiva entre classes indica que as ditas classes não podem ter instâncias nem subclasses em comum. As partições disjuntivas geralmente são importantes em processos de classificação automática.

PartiçõesExaustivas apropriada: uma partição exaustiva entre classes é uma partição disjunta, a qual se adiciona o requisito de plenitude. Quer dizer, o conjunto de classes que formam a partição define de forma completa a classe

superior. Nota-se que nas partições disjuntivas esse requisito de plenitude não é adicionado. Deve-se examinar se tanto as partições disjuntivas como as partições exaustivas estão definidas de forma que se tornem necessárias ao projeto.

Profundidade máxima: define-se como o maior caminho existente seguindo as relações de herança que a taxonomia pode alcançar. Para se obter uma validação do grau de detalhe da ontologia, deve-se analisar a taxonomia conjuntamente com o número de conceitos que a ontologia contém.

Média de subclasses: o número médio de subclasses por cada classe dará uma ideia do nível de detalhe em que está definida a taxonomia da ontologia.

AXIOMAS

Axiomas resolvem questões: alguns axiomas podem estar definidos de forma que possam ser utilizados no projeto para inferir conhecimentos não explicitados na base de conhecimento da aplicação (esses axiomas são denominados de 'regras'). Desta forma, o usuário deverá analisar o potencial desses axiomas com relação às necessidades de seu projeto para deduzir novos conhecimentos e para realizar buscas de conhecimentos não explícitos.

Axiomas inferem conhecimento: ao criar uma nova instância no sistema, alguns valores dos atributos de instância podem ser deduzidos utilizando os axiomas definidos na ontologia. Também podem ser utilizados para deduzir as instâncias de relações.

Axiomas verificam consistência: os axiomas podem ser utilizados para verificar os valores que os atributos dos conceitos podem ter, ou as relações permitidas entre conceitos que sejam definidas no novo sistema.

Axiomas não ligados a conceitos: os axiomas definidos fora das declarações das condições da ontologia dão maior facilidade de compreensão e de modificação, já que sua definição não dependerá das trocas que se façam com as outras condições da ontologia.

Número de axiomas: a quantidade de axiomas definidos pode dar uma ideia do potencial de dedução especificado na ontologia, e da capacidade de a ontologia manter a consistência.

DIMENSÃO: METODOLOGIA

PRECISÃO (FATOR)	Muito baixa, baixa, média, alta, muito alta
Delimitação de fases	Muito baixa, baixa, média, alta, muito alta
Especificação de atividades por fase	Muito baixa, baixa, média, alta, muito alta
Especificação de pessoal por fase	Muito baixa, baixa, média, alta, muito alta

Especificação de técnicas por fase	Muito baixa, baixa, média, alta, muito alta
Especificação de produtos finais por fase	Muito baixa, baixa, média, alta, muito alta
USABILIDADE (FATOR)	Muito baixa, baixa, média, alta, muito alta
Clareza na descrição de atividades e técnicas	Muito baixa, baixa, média, alta, muito alta
Qualidade dos manuais	Muito baixa, baixa, média, alta, muito alta
Manuais com exemplos completos	Muito baixa, baixa, média, alta, muito alta
MATURIDADE (FATOR)	Muito baixa, baixa, média, alta, muito alta
Número de ontologias desenvolvidas	Muito baixa, baixa, média, alta, muito alta
Número de domínios diferentes	Muito baixa, baixa, média, alta, muito alta
Importância de ontologias desenvolvidas	Muito baixa, baixa, média, alta, muito alta

PRECISÃO (FATOR)

Delimitação de fases: indica se estão bem estabelecidos e definidos os processos que devem ser realizados em cada uma das fases de desenvolvimento.

Especificação de atividades por fase: indica se as atividades que devem acontecer em cada uma das fases de desenvolvimento estão bem detalhadas.

Especificação de pessoal por fase: indica se as pessoas que devem intervir em cada uma das fases de desenvolvimento estão bem detalhados.

Especificação de técnicas por fase: indica se as técnicas que devem ser implementadas estão bem identificadas.

Especificação de produtos finais por fase: indica se os formatos e o conteúdo dos produtos que se obtiveram depois da execução de cada uma das fases estão descritos.

USABILIDADE (FATOR)

Clareza na descrição de atividades e técnicas: indica se é facilmente compreendida a forma em que são detalhadas as atividades e técnicas que devem ser realizadas em cada fase.

Qualidade dos manuais: se refere à qualidade dos materiais de apoio disponíveis para aprender a utilizar a metodologia.

Manuais com exemplos completos: indica se existem exemplos adequados que ajudem a compreender as técnicas e atividades da metodologia.

MATURIDADE (FATOR)

Número de ontologias desenvolvidas: indica o número de ontologias desenvolvidas seguindo a metodologia.

Número de domínios diferentes: indica o número de domínios de ontologias diferentes nos quais foi implementada a metodologia.

Importância das ontologias desenvolvidas: indica a importância das ontologias que foram desenvolvidas implementando a metodologia.

DIMENSÃO: CUSTOS

CARACTERÍSTICA	TIPO
Licenças de uso da ontologia (FATOR)	Muito baixa, baixa, média, alta, muito alta
Custos estimados de hardware e software (FATOR)	Muito baixa, baixa, média, alta, muito alta
Custos de interface de acesso (FATOR)	Muito baixa, baixa, média, alta, muito alta
Licenças de uso das ferramentas da ontologia (FATOR)	Muito baixa, baixa, média, alta, muito alta

Licenças de uso da ontologia: se refere aos custos de licença de aquisição e exploração da ontologia.

Custos estimados de hardware e software: se refere aos investimentos que se devem fazer para adquirir hardwares (equipamentos especiais com diferentes capacidades de processamento, armazenamento, comunicação, etc) e softwares (diferentes sistemas operacionais, programas de comunicação, etc) específicos. Também deverão ser incluídos nessa característica os gastos derivados de contratação de pessoal especializado, cursos de formação, instalação, manutenção, etc.

Custos de interface de acesso: se refere aos custos de licenças de aquisição, utilização e modificação das interfaces de acesso aos termos da ontologia.

Licenças de uso das ferramentas da ontologia: irá depender dos processos que sejam necessários realizar para adaptar a ontologia ao projeto; por exemplo, modificação do modelo conceitual, tradução a outras linguagens de implementação, processos de fusão de ontologias, etc. Nestes casos, deve-se calcular os custos de adquirir ferramentas que ajudem a realizar essas tarefas.