



<http://bd.camara.leg.br>

“Dissemina os documentos digitais de interesse da atividade legislativa e da sociedade.”



UNIVERSIDADE FEDERAL DE MINAS GERAIS

ESCOLA DE CIÊNCIA DA INFORMAÇÃO

MARCUS VINÍCIUS CHEVITARESE ALVES

**Da Necessidade e Viabilidade da Adoção do Padrão *Dados Abertos Ligados* pelo Congresso Nacional: um estudo no contexto do Orçamento Público**

Brasília

2013

Marcus Vinícius Chevitarese Alves

**Da Necessidade e Viabilidade da Adoção do Padrão *Dados Abertos Ligados* pelo Congresso Nacional: um estudo no contexto do Orçamento Público**

Monografia apresentada como requisito parcial para aprovação no curso de Especialização em Arquitetura e Organização da Informação, do Programa de Pós-Graduação do Centro de Formação e Aperfeiçoamento da Câmara dos Deputados e da Universidade Federal de Minas Gerais.

Orientador: Marcello Peixoto Bax

**Brasília**

2013



UFMG

Universidade Federal de Minas Gerais  
Escola de Ciência da Informação  
Núcleo de Informação Tecnológica e Gerencial – NITEG  
Curso de Especialização em Arquitetura e Organização da Informação

ATA DA DEFESA DE MONOGRAFIA DE **MARCUS VINÍCIUS CHEVITARESE ALVES**, matrícula: 2011720944

Às 15:00 horas do dia 30 de abril de 2013, reuniu-se nas Instalações do CEFOR/Câmara dos Deputados em Brasília-DF a Comissão Examinadora, para julgar o trabalho intitulado ***Da necessidade e viabilidade da adoção do padrão dados abertos ligados pelo Congresso Nacional: um estudo no contexto do Orçamento Público***, requisito final para obtenção do Grau de ESPECIALISTA em ARQUITETURA E ORGANIZAÇÃO DA INFORMAÇÃO. Abrindo a sessão, o Presidente da Comissão, Prof. Dr. Marcello Peixoto Bax, após dar conhecimento aos presentes do teor das Normas Regulamentares do Trabalho Final, passou a palavra ao candidato para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do candidato. Logo após, a Comissão se reuniu sem a presença do candidato e do público, para julgamento e expedição do resultado final. Foram atribuídas as seguintes indicações:

Prof. Dr. Marcello Peixoto Bax – Orientador


Prof. Dr. Jorge Tadeu de Ramos Neves

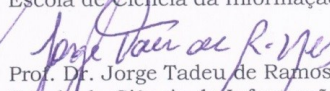
João Alberto de Oliveira Lima

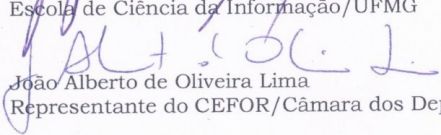
Pelas indicações, o candidato foi considerado aprovado com nota 100.

O resultado final foi comunicado publicamente ao candidato pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a sessão, da qual foi lavrada a presente ATA que será assinada por todos os membros participantes da Comissão Examinadora.

Brasília, 30 de abril de 2013

  
Prof. Dr. Marcello Peixoto Bax  
Escola de Ciência da Informação/UFMG (Orientador)

  
Prof. Dr. Jorge Tadeu de Ramos Neves  
Escola de Ciência da Informação/UFMG

  
João Alberto de Oliveira Lima  
Representante do CEFOR/Câmara dos Deputados

Obs: Este documento não terá validade sem a assinatura e carimbo do Coordenador do Curso.

Dedico este trabalho à minha querida esposa Aléssia, que  
nunca deixou de me apoiar, estimular e acalantar.

## **AGRADECIMENTOS**

Agradeço à Câmara dos Deputados por essa nova oportunidade de aprimorar meus conhecimentos e espero que este trabalho possa dar alguma contribuição para o Parlamento e para o Brasil.

Agradeço ao Centro de Formação e Aperfeiçoamento de Pessoal (Cefor) pela infraestrutura e apoio dados durante o curso.

Agradeço à direção do Centro de Informática (Cenin) pela confiança em mim depositada.

Agradeço aos colegas e amigos da Coordenação de Disseminação de Informações (Codis), especialmente Michael, Roberta, Carlos Jaques e Paulo Henrique.

Agradeço aos colegas e amigos da Coordenação de Engenharia de Sistemas e Análise de Negócios (Cesan), especialmente Sandro, Mário, Valmir, Douglas, Márcio e Fernando.

Agradeço aos ex-colegas do Cenin, mas ainda amigos: Dilson, Fernando Teixeira, Idelfonso, Leonardo, Ricardo Oliveira, Ricardo Vilarins e Rosinaldo.

Agradeço aos colegas e professores do curso, em especial ao meu orientador, Marcello Peixoto Bax.

Agradeço ao colega e grande amigo Maurílio e à sua família.

Agradeço à minha esposa e aos nossos familiares, pelo apoio e incentivo constantes.

*O aperfeiçoamento da inteligência coletiva (que supõe a liberdade) é o produto e o sentido da evolução cultural.*

---

Pierre Lévy, *Ciberdemocracia*.

## RESUMO

O objetivo do trabalho foi analisar a necessidade e a viabilidade da adoção do padrão Dados Abertos Ligados pelo Congresso Nacional no que concerne à informação sobre o Orçamento Federal. O trabalho inicia-se com uma revisão dos fundamentos teóricos da Web Semântica. A seguir, foi empreendida uma pesquisa acerca das principais tecnologias da Web Semântica, especialmente aquelas associadas ao padrão Dados Abertos e Ligados. Posteriormente, foi elaborado e aplicado um estudo de casos múltiplos, no qual se avaliou os sistemas de informação orçamentários do Congresso Nacional que possuem interface Web: o Siga Brasil (Senado) e o Fiscalize (Câmara dos Deputados). Dando prosseguimento, avaliou-se a viabilidade da adoção do padrão Dados Abertos e Ligados pelo Congresso Nacional, no que diz respeito aos dados e informações orçamentárias. Para isso, ampliou-se a revisão bibliográfica e telemática inicial, de modo a se ter mais clareza sobre os princípios do padrão em estudo, bem como das tecnologias e metodologias necessárias para a sua implantação. Além disso, fez-se outro estudo de casos múltiplos, no qual se avaliou duas aplicações computacionais baseadas nesse padrão. Os resultados do trabalho indicam a necessidade e a viabilidade da adoção do padrão Dados Abertos e Ligados pelo Congresso Nacional no contexto das informações associadas ao Orçamento Público.

**Palavras-chave:** 1. Direito à informação. 2. Transparência administrativa. 3. Orçamento Público. 4. Tecnologias da informação e da comunicação. 5. Web Semântica.



## LISTA DE ABREVIATURAS

Sigla	Significado
<b>BNDES</b>	Banco Nacional de Desenvolvimento Econômico e Social
<b>CSV</b>	<i>Comma Separated Values</i> – Valores Separados por Vírgula
<b>DL</b>	<i>Description Logics</i> – Lógica Descritiva; Lógica das Descrições
<b>DW</b>	<i>Data Warehouse</i> – Armazém de Dados
<b>EC</b>	<i>European Comission</i> – Comissão Europeia
<b>FNDE</b>	<i>Fundo Nacional de Desenvolvimento de Educação</i>
<b>FOAF</b>	<i>Friend of a Friend Ontology</i> – Ontologia Amigo de um Amigo
<b>FOL</b>	<i>First-Order Logics</i> – Lógica de Primeira Ordem
<b>FOSS</b>	<i>Free and Open Source Software</i> – Programa de Código Aberto e Gratuito
<b>G2B</b>	<i>Government to Business</i> – (do) Governo para as Empresas
<b>G2C</b>	<i>Government to Citizen</i> – (do) Governo para os Cidadãos
<b>G2G</b>	<i>Government to Government</i> – (do) Governo para o Governo
<b>HTML</b>	<i>HyperText Markup Language</i> – Linguagem para a Marcação de Hipertextos
<b>HTTP</b>	<i>HyperText Transfer Protocol</i> – Protocolo para a Transferência de Hipertextos
<b>LAI</b>	Lei de Acesso à Informação
<b>LDO</b>	Lei de Diretrizes Orçamentárias
<b>LOA</b>	Lei Orçamentária Anual
<b>LOD</b>	<i>Linked and Open Data</i> – Dados Abertos Ligados
<b>LOGD</b>	<i>Linked and Open Government Data</i> – Dados Abertos Ligados de Governo
<b>LPO</b>	Ver FOL.
<b>LRF</b>	Lei de Responsabilidade Fiscal
<b>MPOG</b>	Ministério do Orçamento, Planejamento e Gestão
<b>OWL</b>	<i>Web Ontology Language</i> – Linguagem de Ontologias para a Web
<b>PDF</b>	<i>Portable Document Format</i> – Formato Portável de Documento
<b>PPA</b>	Plano PluriAnual
<b>RDF</b>	<i>Resource Description Framework</i> – Arcabouço para a Descrição de Recursos
<b>RDFS</b>	<i>Resource Description Framework Schema</i> – Esquema para o RDF
<b>SBC</b>	Sociedade Brasileira de Computação
<b>SELOR</b>	Sistema de Elaboração Orçamentária
<b>SIAFI</b>	Sistema Integrado de Administração Financeira
<b>SICONV</b>	Sistema de Informações sobre Convênios e Transferências Voluntárias da União
<b>SIOP</b>	Sistema Integrado de Planejamento e Orçamento
<b>SPARQL</b>	<i>Protocol And RDF Query Language</i> – Protocolo e Linguagem de Consulta RDF
<b>XML</b>	<i>eXtensible Markup Language</i> – Linguagem Extensível de Marcação
<b>UFMG</b>	Universidade Federal de Minas Gerais
<b>UFSC</b>	Universidade Federal de Santa Catarina
<b>UFRS</b>	Universidade Federal do Rio Grande do Sul
<b>URI</b>	<i>Universal Resource Identifier</i> – Identificador Universal de Recursos
<b>W3C</b>	<i>WWW Consortium</i> – Consórcio para a WWW
<b>WS</b>	Web Semântica

## LISTA DE FIGURAS

Figura 1. Camadas da Web Semântica .....	20
Figura 2. Diferentes níveis de expressividade de modelos de uma molécula de água .....	22
Figura 3. Espectro semântico dos instrumentos da Web Semântica .....	23
Figura 4. Padrões de sentenças da lógica aristotélica .....	26
Figura 5. Silogismo aristotélico .....	26
Figura 6. Exemplo de raciocínio que não pode ser expresso corretamente em lógica proposicional .....	28
Figura 7. Raciocínio expresso corretamente em LPO .....	29
Figura 8. Categorias da ontologia de Aristóteles .....	32
Figura 9. Ontologia simples expressa em uma rede semântica .....	34
Figura 10. Exemplo de conhecimento intensional expresso em lógica descritiva TBox .....	36
Figura 11. Exemplo de conhecimento extensional expresso em lógica descritiva ABox .....	37
Figura 12. Fragmento de conteúdo em HTML .....	40
Figura 12. XML – exemplo 1 .....	41
Figura 13. XML – exemplo 2 .....	42
Figura 14. XML Schema – exemplo .....	43
Figura 15. Grafo orientado representando uma sentença RDF .....	44
Figura 16. Exemplos de assertivas em triplas (sujeito, propriedade, objeto) .....	45
Figura 17. Tradução de assertivas lógicas para RDF/XML .....	45
Figura 18. Fragmento de um documento RDFS/OWL com propriedade <code>rdfs:subClassOf</code> .....	47
Figura 19. Fragmento de um documento RDFS/OWL com propriedades <code>rdfs:domain</code> e <code>rdfs:range</code> .....	48
Figura 18. Tradução de assertivas lógicas para OWL (a) .....	51
Figura 19. Tradução de assertivas lógicas para OWL (b) .....	51
Figura 20. Trecho de um documento usando a ontologia FOAF .....	53
Figura 21. Consulta simples em SPARQL .....	53
Figura 22. Sistema Fiscalize – Transferências por Programação .....	59
Figura 23. Sistema Fiscalize – Transferências por Programação - relatório .....	59
Figura 24. Siga Brasil – “Universos” de informações disponíveis (detalhe) .....	62
Figura 25. Siga Brasil – “Universo” Convênios - SICONV .....	63
Figura 26. Siga Brasil – “Universo” Convênios - SICONV .....	63
Figura 27. Dados ligados de governo .....	67
Figura 28. Metadados de uma aplicativo de demonstração do portal Data-Gov Wiki .....	70
Figura 29. Consulta SPARQL sobre uma base descrita no portal Data-gov Wiki .....	70
Figura 30. Resultado da consulta sobre uma aplicação disponível no portal Data-gov Wiki .....	71
Figura 31. Visão geral da arquitetura do portal Ligados nos Políticos .....	72
Figura 32. Vocabulários e propriedades usados no portal Ligados nos Políticos .....	73
Figura 33. Detalhe da página inicial do portal Ligados nos Políticos .....	73
Figura 34. Método de Uschold para a construção de ontologias (skeletal methodology) .....	78
Figura 35. Mapeamento entre a Ontologia de Classificação da Despesa do Orçamento Federal e um recurso .....	83
Figura 36. Exercício de modelagem – Etapa Integrar .....	83
Figura 37. Tela de resultado de navegação do portal <orcamento.dados.gov.br> .....	85
Figura 38. Exemplo ilustrativo de tecnologias do tipo ‘mashup’ para visualização de dados abertos. ....	86

## LISTA DE TABELAS

Tabela 1 – Comparação entre a WWW e a Web Semântica .....	19
Tabela 2. Conectivos de uso comum na lógica proposicional .....	27
Tabela 3. Tabela-verdade para os conectivos da lógica proposicional .....	27
Tabela 2. Tradução de sentenças em linguagem natural para a lógica formal.....	30
Tabela 3. Construtores de uma lógica descritiva. ....	35
Tabela 4. Classes RDFS.....	46
Tabela 5. Equivalência entre algumas expressões OWL e DL.....	50
Tabela 6. Síntese da Avaliação do Sistema Fiscalize frente à Lei de Acesso à Informação .....	61
Tabela 7. Síntese da Avaliação do Sistema Siga Brasil frente à Lei de Acesso à Informação .....	64
Tabela 8. Tecnologias para conversão de dados para RDF, manipulação e gravação .....	77
Tabela 9. Exercício de modelagem – Etapa ‘Identificar o Propósito’ .....	78
Tabela 10. Exercício de modelagem – Etapa ‘Capturar’ .....	79
Tabela 11. Exercício de modelagem – Etapa ‘Codificar’ – Classes e Relacionamentos.....	79
Tabela 12. Exercício de modelagem – Etapa ‘Codificar’ – Axiomas em LPO e DL .....	80
Tabela 13. Ontologia da Classificação da Despesa do Orçamento Federal.....	82
Tabela 14. Ferramentas para visualização de dados do tipo ‘mashup’ .....	84

# SUMÁRIO

1.	Introdução .....	12
2.	Fundamentos teóricos .....	17
2.1	O contexto da Web Semântica .....	17
2.2	Representação do Conhecimento, Semântica e Linguagem .....	21
2.3	Linguagens Formais .....	25
2.4	As Lógicas de Descrições (DL) e as Ontologias .....	30
3.	As tecnologias da Web Semântica.....	39
3.1	Identificador Universal de Recursos: URI .....	39
3.2	Linguagens de Marcação: HTML e XML.....	40
3.3	Arcabouço para Descrição de Recursos: RDF .....	43
3.4	Esquema para Descrição de Recursos: RDFS.....	46
3.5	Linguagem para Descrição de Ontologias na Web: OWL .....	48
3.6	Linguagem para Consulta de Grafos RDF: SPARQL.....	52
3.7	Resumo do capítulo .....	54
4.	O Padrão ‘Dados Abertos Ligados’ na Representação e Disseminação da Informação Orçamentária....	55
4.1	Síntese do Processo Orçamentário .....	55
4.2	Análise dos Sistemas de Informação Orçamentária do Congresso Nacional frente à LAI .....	56
4.2.1	Sistema Fiscalize .....	57
4.2.2	Sistema Siga Brasil .....	61
4.3	O Padrão ‘Dados Abertos Ligados’ (LOD) .....	65
4.3.1	Data-Gov Wiki.....	68
4.3.2	Ligado nos políticos .....	71
4.4	Viabilidade dos Dados Abertos Ligados no Contexto da Informação Orçamentária .....	74
4.4.1	A viabilidade da etapa de conversão.....	76
4.4.2	A viabilidade da etapa de enriquecimento.....	77
4.4.3	A viabilidade da etapa de visualização .....	84
4.4.4	A viabilidade da etapa de proveniência do conhecimento sobre os dados .....	86
4.5	Resumo do Capítulo .....	87
5.	Considerações Finais .....	88
	Referências .....	90

## 1. INTRODUÇÃO

Conforme Platt Neto (2009), “as entidades que compõem a estrutura da administração pública brasileira são obrigadas pela Constituição Federal a prestar contas do uso de recursos públicos e a respeitar o princípio da publicidade” (p. 76).

Tal determinação constitucional ensejou a criação de órgãos de controle, como a Controladoria-Geral da União, bem como a edição de normas infraconstitucionais, tais como a chamada Lei de Responsabilidade Fiscal (BRASIL, 2000) e a Lei da Transparência (BRASIL, 2009).

Entretanto, há fortes indicativos de que as normas e as instituições públicas de controle, isoladamente, são incapazes de coibir a malversação de recursos públicos (cf. CARLOS et al, 2008; MELLO et al., 2002).

É fato que a Administração Pública está atenta à questão, uma vez que diversos produtos de informação para divulgação de dados foram criados, associados aos cidadãos, empresas e demais órgãos públicos: conhecidos como G2C – Government to Citizen, G2B – Government to Business e G2G – Government to Government. Dentre esses produtos, destacam-se os chamados portais de governo e alguns tipos especiais dentre eles: os portais da transparência e portais de orçamento, por exemplo. No plano jurídico, pode-se destacar a promulgação da Lei Complementar 131/2009, conhecida como Lei da Transparência (BRASIL, 2009). Essa Lei veio aperfeiçoar a Lei de Responsabilidade Fiscal (LRF) – Lei Complementar 101, de 2000 (BRASIL, 2000), tornando obrigatório o uso dos meios eletrônicos para a divulgação de dados públicos. o parágrafo único da art. 48 passou a vigorar dessa forma (BRASIL, 2009):

Art. 48. ....

Parágrafo único. A transparência será assegurada também mediante:

*I – incentivo à participação popular e realização de audiências públicas, durante os processos de elaboração e discussão dos planos, lei de diretrizes orçamentárias e orçamentos;*

*II – liberação ao pleno conhecimento e acompanhamento da sociedade, em tempo real, de informações pormenorizadas sobre a execução orçamentária e financeira, em meios eletrônicos de acesso público;*

*III – adoção de sistema integrado de administração financeira e controle, que atenda a padrão mínimo de qualidade estabelecido pelo Poder Executivo da União e ao disposto no art. 48-A.*

No entanto, alguns problemas ficaram evidenciados a partir desse fenômeno, dos quais serão destacados dois. O primeiro é que os dados da Administração Pública estão fragmentados, pouco

integrados, em formatos distintos e, muitas vezes, proprietários e, portanto, de difícil reutilização (ALANI et al, 2007). O segundo é que as aplicações desenvolvidas para a exploração e visualização dessas informações em geral possibilitam baixa flexibilidade na sua manipulação. A literatura denomina essa forma de disseminação como *broadcasting* (W3C Escritório Brasil, 2010), ou seja, um emissor (no caso, o governo) e vários receptores (os cidadãos e organizações civis). Portanto, nesse caso, se o usuário da informação quiser ver ou explorar a informação de forma diferente daquelas que foram previstas e implementadas pelos sistemas governamentais, terá grandes dificuldades para fazê-lo.

Os problemas citados estimularam, por parte da comunidade acadêmica, da Sociedade Civil e dos governos, a criação de um conjunto de recomendações denominado “Dados Abertos” (*Open Data*). “Dados governamentais abertos é a disponibilização de informações governamentais representadas em formato aberto e acessível de tal modo que possam ser reutilizadas, misturadas com informações de outras fontes, gerando novos significados” (W3C Escritório Brasil, 2010 , p. 8). Posteriormente, com o desenvolvimento da chamada Web Semântica (*Semantic Web*), emergiu um novo padrão, denominado Dados Ligados (*Linked Data*), o qual vem absorvendo o padrão anterior, dando origem aos Dados Abertos Ligados<sup>1</sup> (*Linked and Open Data – LOD*). Conforme Bizer et al (2009), os dados ligados referem-se a dados publicados na Web de modo que sejam legíveis por máquina, seus significados sejam explicitamente definidos, estejam ligados a outros conjuntos de dados e, por sua vez, possam ser ligados a partir de conjuntos de dados externos. A disseminação dos dados ligados tem mostrado que eles podem ser uma solução para os problemas mencionados, pois aumentam a reutilização e a integração dos dados e metadados, e tornam mais flexíveis sua exploração e visualização; ou seja, a estratégia de Dados Abertos Ligados, segundo a literatura, tem o potencial de agregar valor aos dados, tornando-os informações úteis aos seus usuários. No caso dos dados de governo, ou melhor, em posse do governo, esses usuários são todos os cidadãos, que são, por assim dizer, seus legítimos donos.

Nesse contexto, é importante destacar também a promulgação da Lei de Acesso à Informação – LAI (BRASIL, 2011), que vem ao encontro dos anseios da sociedade, ao mesmo tempo em que parece estar em sintonia com as novas tecnologias. Nesse particular, citam-se os parágrafos 2º e 3º, do artigo 8º:

*§ 2º Para cumprimento do disposto no caput, os órgãos e entidades públicas deverão utilizar todos os meios e instrumentos legítimos de que dispuserem, sendo*

---

<sup>1</sup> É comum, na literatura da área, o uso da expressão Linked Data como sinônimo de Linked and Open Data. Optou-se pela segunda forma, para enfatizar o caráter aberto dos dados.

*obrigatória a divulgação em sítios oficiais da rede mundial de computadores (internet).*

*§ 3º Os sítios de que trata o § 2º deverão, na forma de regulamento, atender, entre outros, aos seguintes requisitos:*  
*I - conter ferramenta de pesquisa de conteúdo que permita o acesso à informação de forma objetiva, transparente, clara e em linguagem de fácil compreensão;*  
*II - possibilitar a gravação de relatórios em diversos formatos eletrônicos, inclusive abertos e não proprietários, tais como planilhas e texto, de modo a facilitar a análise das informações;*  
*III - possibilitar o acesso automatizado por sistemas externos em formatos abertos, estruturados e legíveis por máquina;*  
*IV - divulgar em detalhes os formatos utilizados para estruturação da informação;*  
*V - garantir a autenticidade e a integridade das informações disponíveis para acesso;*  
*VI - manter atualizadas as informações disponíveis para acesso;*

.....

O inciso III do parágrafo 3º do artigo 8º é bastante claro no que tange à necessidade de se tornar disponíveis as informações no formato de “dados abertos”. As estruturas de dados não são definidas a priori – e nem é conveniente que uma lei discipline detalhes técnicos. Mas, como será visto ao longo desse texto, já existe uma gama de técnicas, tecnologias e ferramentas das quais governo e sociedade podem lançar mão para auxiliar a implementação efetiva dessa lei.

No entanto, como ressalta Klischewski *et al.* (2006), a área de Web Semântica para governo eletrônico enfrenta um desafio para migração de dados e informações para um formato 'semântico': afinal, trata-se de “páginas HTML feitas à mão e páginas web geradas automaticamente por sistemas gerenciadores de conteúdo, os quais foram projetados para exibir conteúdo para ser lido por humanos [não por máquinas]” (p. 1, tradução livre). Os autores explicam que, para a Web Semântica, é necessário, pelo menos, que o conteúdo seja enriquecido de modo a tornar-se legível por máquinas.

Isso pode ser feito de duas formas, reorganizando o código dos programas e ou conteúdo; ou criando uma “camada semântica” em cima da “camada não-semântica”. Este trabalho irá explorar alguns projetos que optaram pela segunda abordagem, por ser aquela que apresenta, em tese, os menores custos.

Em particular, o foco será na análise dos sistemas dos sistemas de informações relativas ao Orçamento Público, na esfera federal. O ciclo orçamentário é complexo, o que pode torná-lo obscuro e, assim, uma potencial fonte de corrupção. Isso leva à necessidade de sua decodificação para ampliar o entendimento pelas instituições governamentais, organizações da sociedade civil, enfim, por todos os atores envolvidos na fiscalização e controle das contas públicas. Nesse sentido

Mello et al. (2002) afirmam: “o conhecimento do processo orçamentário, seus prazos, sua forma de tramitação e execução dos gastos e a prestação de contas é de fundamental importância para o controle social e o combate à corrupção” (p. 172).

Dentre esses atores, o Congresso Nacional, com o auxílio do Tribunal de Contas da União, tem papel chave tanto na elaboração do Orçamento Federal, quanto na fiscalização do Executivo, inclusive no que tange às contas públicas. A Constituição Federal é clara a esse respeito, quando determina, no art. 48:

*Cabe ao Congresso Nacional, com a sanção do Presidente da República, [...] dispor sobre todas as matérias de competência da União, especialmente sobre:*

*I - sistema tributário, arrecadação e distribuição de rendas;*

*II - plano plurianual, diretrizes orçamentárias, orçamento anual, operações de crédito, dívida pública e emissões de curso forçado;*

*[...]*

Nesse contexto, o problema de pesquisa é: como o Congresso Nacional pode cumprir seu papel institucional e legal, especialmente os estabelecidos pela Constituição Federal e pela Lei de Acesso à Informação, no que concerne à abertura dos dados e informações sobre o Orçamento Público?

Na tentativa de se dar uma resposta ao problema, o objetivo do trabalho é analisar a necessidade e a viabilidade da adoção do padrão Dados Abertos Ligados pelo Congresso Nacional no que concerne à abertura dos dados e informações sobre o Orçamento Público, especialmente os estabelecidos pela Constituição Federal e pela Lei de Acesso à Informação. Importante ressaltar aqui uma restrição metodológica: o enfoque será dado na esfera federal, ou seja, nas informações orçamentárias da União; no entanto, acredita-se que esse padrão possa ser estendido aos orçamentos públicos dos demais entes da Federação, bem como a outras informações associadas às atribuições do Poder Legislativo.

A justificativa da pesquisa proposta se dá tanto na perspectiva teórica quanto aplicada. Na primeira, tem-se o fato que o padrão Dados Abertos Ligados é baseado na Web Semântica, a qual possui, como será visto ao longo do trabalho, um forte embasamento teórico, e contribui para análise de questões associadas à Representação e Recuperação da Informação e à Representação do Conhecimento, tratando questões relevantes como a expressividade e o formalismo dessas representações. Com relação à perspectiva aplicada, como foi descrito nos parágrafos anteriores, o padrão Dados Abertos Ligados tem o potencial – como a pesquisa visa demonstrar – de auxiliar na solução dos problemas decorrentes da necessidade de se promover a abertura dos dados públicos,



nos termos da legislação vigente, com ênfase na Lei de Acesso à Informação.

O trabalho possui dois pressupostos:

- Pressuposto 1 (P1): Os sistemas Web de informações orçamentárias do Congresso Nacional hoje disponíveis não atendem plenamente os requisitos estabelecidos pela Lei de Acesso à Informação .
- Pressuposto 2 (P2): A adoção do padrão Dados Abertos Ligados pelo Congresso Nacional, no tocante à informação sobre o Orçamento Federal, é viável e possibilita o atendimento aos requisitos da Lei de Acesso à Informação.

A fim de cumprir o objetivo proposto, foi adota a metodologia descrita a seguir. A fim de revisar os fundamentos teóricos da Web Semântica (WS), que, por sua vez, é a base do padrão Dados Abertos Ligados (LOD), fez-se uma revisão bibliográfica e telemática. O resultado encontra-se no Capítulo 1. A seguir, foi empreendida uma pesquisa acerca das principais tecnologias da WS e do LOD, cuja síntese está descrita no Capítulo 2. Posteriormente, para investigar o pressuposto (P1), foi elaborado e aplicado um estudo de casos múltiplos, no qual se avaliou os sistemas de informação orçamentários do Congresso Nacional que possuem interface Web: o Siga Brasil (Senado) e o Fiscalize (Câmara dos Deputados). Dando prosseguimento, no intuito de confirmar ou refutar o pressuposto (P2), avaliou-se a viabilidade da adoção do padrão LOD pelo Congresso Nacional, no que diz respeito aos dados e informações orçamentárias. Para isso, ampliou-se a revisão bibliográfica e telemática inicial, de modo a se ter mais clareza sobre os princípios do padrão LOD, bem como das tecnologias e metodologias necessárias para a sua implantação. Além disso, fez-se outro estudo de casos múltiplos, no qual se avaliou duas aplicações computacionais baseadas nesse padrão. Os resultados das avaliações foram relatados no Capítulo 3. Finalmente, algumas reflexões são propostas nas Considerações Finais.

## 2. FUNDAMENTOS TEÓRICOS

*Tudo era um caos até que surgiu a mente e pôs ordem nas coisas.*

---

Anaxágoras

O objetivo deste capítulo é apresentar uma síntese dos fundamentos teóricos do trabalho, em especial da Web Semântica (WS), base sobre a qual se assenta o padrão Dados Abertos Ligados. Ele está organizado como se segue. Na primeira seção, são apresentados o contexto, os objetivos e os componentes da WS. Na segunda seção, os principais aspectos da WS são associados a teorias relacionadas, advindas, principalmente, da Matemática, da Filosofia, da Linguística e da Ciência da Computação (especialmente da Inteligência Artificial). A terceira seção tem foco nas Linguagens Formais; ela apresenta um breve panorama histórico dessas linguagens e da relação delas com a Lógica. Na quarta e última seção, enfatiza-se a Lógica de Descrições, visto que ela é a base sob a qual foi desenvolvida a linguagem que permite descrever ontologias, no contexto da WS. Cabe ressaltar que, como se trata de um trabalho na área de Ciência da Informação, mas com forte interface com a Ciência da Computação, os conceitos básicos de Lógica Formal foram apresentados, uma vez que, eventualmente, o público oriundo da primeira área não tem familiaridade com o tema. No entanto, algumas questões relacionadas à formalização dos mecanismos de inferência e à análise dos algoritmos a eles relacionados foram propositalmente omitidas: para um maior aprofundamento, recomenda-se a leitura das referências indicadas, especialmente Mortari (2001), Russel e Norvig (2004).

### 2.1 O contexto da Web Semântica

Como ressalta Concepción (2010), a Web tornou-se um instrumento de uso cotidiano para o intercâmbio de informação em nossa sociedade. No entanto, seu crescimento foi tão vertiginoso que aquilo que, inicialmente, era uma vantagem – a disponibilidade da informação, logo se transformou em um problema potencial: a sobrecarga de informação (*information overload*).

De fato, de acordo com dados do Internet World Stats<sup>2</sup>, o número de usuários da Internet passou de cerca de 16 milhões, em 1995, para cerca de 2 bilhões e 400 milhões, em junho de 2012. O número de fontes de informação baseadas em websites também aumentou em proporções significativas. Um indicador disso é o incremento no número de domínios Web. Segundo o Núcleo

---

<sup>2</sup> Endereço: <<http://www.internetworldstats.com/>>. Trata-se de portal que agrega informações estatísticas sobre a Web, usando fontes confiáveis, como o ITU (*International Telecommunication Union*), agência da ONU especializada nas tecnologias da informação e da comunicação (TIC).

de Estudos sobre as Tecnologias da Informação e da Comunicação<sup>3</sup> (CETIC), no Brasil (domínio ‘.br’), esse número passou de 851, em janeiro de 1995, para mais de 360.000, em 2001; ou seja, multiplicou-se por mais de 400 ao longo dos seis primeiros anos. Atualmente são mais de três milhões de domínios no âmbito nacional – dados de dezembro de 2012.

Portanto, voltando ao problema apontado: como organizar, tratar e recuperar informação de forma eficiente em um repositório de informações tão grande e dinâmico? Os buscadores (Yahoo e Google, entre outros) foram projetados para solucionar um dos componentes do problema: a recuperação da informação. No entanto, mesmo nessa seara, eles ainda possuem limitações. Podem-se destacar duas: (i) eles são projetados para buscas genéricas, em domínios genéricos; (ii) eles são projetados para buscas baseadas na sintaxe, ou seja, nos termos, e não na semântica, isto é, no significado, ou melhor, na rede de significados evocados por um termo. Para ilustrar, um exemplo: se digitarmos a expressão “total despesa função saúde orçamento 2012” era de esperar, como resultado, um *número* que representasse o total de despesas orçadas para a função ‘saúde’ no Orçamento Federal de 2012. No entanto, são retornados 1.530.000 resultados, em forma de *documentos*. Os mais relevantes trazem cópias de documentos oficiais que, provavelmente, apresentam o número desejado. No entanto, é necessário que se obtenha o arquivo, abra-o e o vasculhe para, *talvez*, encontrar a informação desejada. Ademais, a informação estará em uma forma estática e sem metadados agregados, não possibilitando, diretamente, novos cruzamentos e inferências a partir dos dados preexistentes; ou seja, o modelo de representação e recuperação da informação baseado tão somente nas páginas Web ‘tradicionais’ (baseadas em HTML) e nos buscadores-padrão limitam o potencial de geração eficiente de nova informação a partir da informação existente.

Na mesma linha, Freitas (2003) afirma que, nesse contexto, “torna-se difícil *agregar valor* à informação disponível, ou seja, transformá-la em informação *útil* e facilmente *acessível*, convertendo informação *desestruturada* ou *semiestruturada* em *estruturada*” (p. 3, grifos nossos). Para o autor, “a falta de mecanismos capazes de captar a semântica do conteúdo das páginas da Web criou uma forte demanda de serviços que se ajusta adequadamente à classe de serviços estudada em Inteligência Artificial, que passou a ser vista como uma alternativa bastante factível para um melhor tratamento dos problemas relacionados à manipulação de informação na Internet”. Segundo ele, “dois tipos de solução foram propostos, que não são mutuamente exclusivas” (p. 4).

A primeira abordagem consiste em “dotar os sistemas de inteligência e autonomia para

---

<sup>3</sup> Endereço: <<http://www.cetic.br>>.

percorrer e selecionar informação relevante na imensidão da rede, deduzindo ou aprendendo quais as informações úteis” (FREITAS, 2003, p. 4). Essa abordagem é conhecida na literatura como *agentes inteligentes*. Já a segunda abordagem intenta “dotar a própria Internet de inteligência, fazendo com que as páginas possuam uma semântica clara e definida e que agentes possam raciocinar sobre esta semântica” (p. 4). Segundo o autor, foi essa a abordagem que deu origem à Web Semântica (WS). No entanto, todo o potencial da WS só poderá ser atingido com o uso dos agentes computacionais inteligentes.

Conforme Berners-Lee *et al.* (2001), a WS seria uma visão de futuro para a Web, a qual traria “estrutura ao conteúdo inteligível das páginas Web, criando um ambiente onde agentes de software percorreriam página por página e teriam a capacidade de efetuar tarefas sofisticadas para os usuários [humanos]” (p. 1, tradução livre). Os autores esclarecem, no mesmo trabalho, que a WS “não é uma Web separada, mas uma extensão da existente, na qual é acrescentado significado bem-definido [ou seja, semântica] à informação, habilitando computadores e pessoas a trabalhar em cooperação” (p. 1, tradução livre). Acrescentam ainda que a principal propriedade da Web é sua universalidade; ou seja, o link hipertextual (*hyperlink*) permite ligar ‘qualquer coisa a qualquer coisa’. Dessa forma, a tecnologia Web não pode discriminar aquilo que ela está referenciando, por exemplo, “entre informação comercial ou acadêmica”, para ficar no exemplo dos autores (p. 1). Na verdade, a informação “varia ao longo de vários eixos” (p. 1), ou seja, possui muitas dimensões. Uma delas, destacam os autores, é a diferença entre a informação produzida primariamente para consumo humano e aquela cujo foco principal são as máquinas. E o volume e taxa de crescimento das informações disponíveis na Web exigem que haja, cada vez mais, informações ‘legíveis’ por máquinas, para que elas nos auxiliem a estruturá-la, processá-la, recuperá-la; em suma, torná-la útil. A tabela 1 sintetiza as características principais da Web 'tradicional' e da WS:

*Tabela 1 – Comparação entre a WWW e a Web Semântica*

<i>Aspecto</i>	<i>Web 'tradicional'</i>	<i>Web Semântica</i>
<b>Componente fundamental</b>	Conteúdo não-estruturado	Declarações formais
<b>Audiência primária</b>	Humanos	Aplicações
<b>Links</b>	Indicam localização	Indicam localização e significado
<b>Vocabulário primário</b>	Instruções de formatação	Semântica e lógica
<b>Lógica</b>	Informal/não padronizada	Lógica descritiva

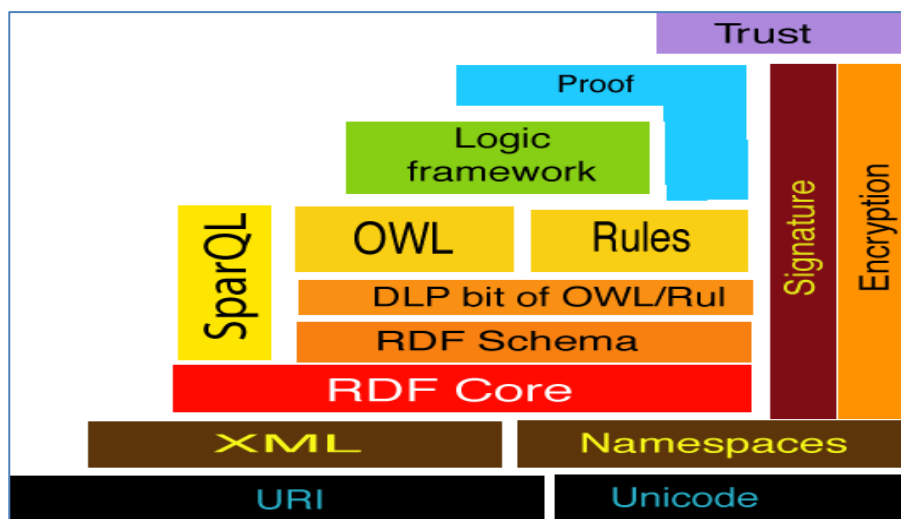
Fonte: HEBELER, John; FISHER, Matthew; BLACE, Ryan; PEREZ-LOPES, Andrew. **Semantic Web Programming**. Indianapolis: Wiley, 2009.

A partir do trabalho de Berners-Lee e seus colegas, houve uma grande evolução nas tecnologias e estudos envolvendo a WS. Atualmente, os padrões associados à WS são definidos no

âmbito do W3C<sup>4</sup>. E, na definição proposta por este organismo, a WS possibilita a transição da Web de *documentos* para a Web de *dados*. Ou, mais precisamente, para a Web de dados e metadados.

Berners-Lee *et al.* (2001) inicialmente propôs quatro aspectos principais que a WS deveria atender: expressar o significado; permitir o acesso às representações do conhecimento; as ontologias e os agentes de software. Para atender esses aspectos, a W3C propôs uma arquitetura, composta de diversas camadas (figura 1).

Figura 1. Camadas da Web Semântica



Fonte: <<http://www.w3.org/DesignIssues/diagrams/sw-stack-2005.png>>.

Essas camadas podem ser agrupadas em quatro níveis: o nível de infraestrutura, o nível de definição de esquemas, o nível de definição de ontologias e o nível lógico. Essa não é uma divisão estanque, mas serve como uma referência para um melhor entendimento da arquitetura. O *nível de infraestrutura* fornece os elementos estruturais básicos: URI (que identifica um recurso) e Unicode (que permite codificá-lo). O *nível de esquema* fornece meios para se definir as estruturas dos recursos e declarações sobre eles, por meio das linguagens XML, RDF e RDFS. O *nível de ontologia* possibilita construir declarações mais sofisticadas, que definem conceitos, relacionamentos e restrições (HEBELER et al., 2009); o principal componente dessa camada é a linguagem OWL. O *nível lógico* define os mecanismos pelos quais é possível se fazer inferências sobre os dados. Segundo Berners-Lee (1998), ao se inserir lógica nos documentos, abre-se a possibilidade de permitir regras de dedução de tipo de documento; verificação de um documento com base em um conjunto de regras de consistência e a resolução de uma consulta por meio da conversão de termos desconhecidos em termos conhecidos. É importante ressaltar que as camadas

<sup>4</sup> World Wide Web Consortium: entidade responsável pela definição de padrões para a Web. Fonte: <<http://www.w3c.org/sobre>>.

são interdependentes: por exemplo, as declarações sobre os recursos devem ser suficientemente explícitas e formalizadas para que um mecanismo automático possa fazer inferências acerca deles.

Os aspectos tecnológicos serão descritos no capítulo 2. Antes, porém, é importante descrever algumas das teorias subjacentes à Web Semântica, ainda que de forma sucinta.

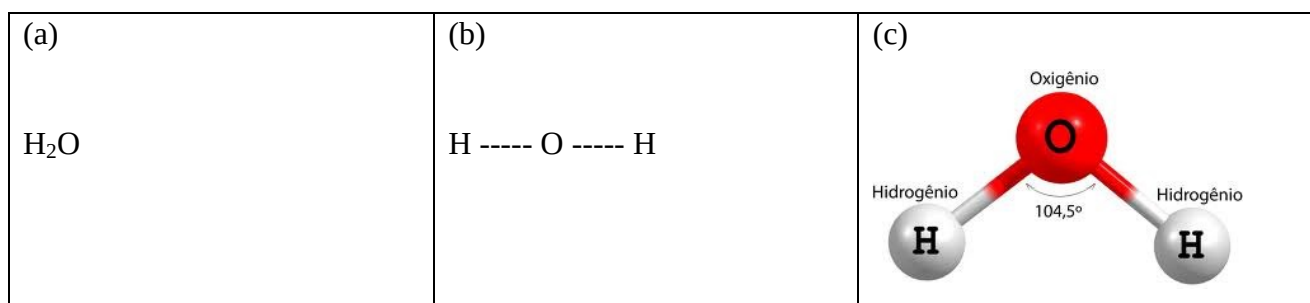
## 2.2 *Representação do Conhecimento, Semântica e Linguagem*

Cada uma das camadas descritas na seção 1.1 acrescenta um maior nível de semântica ao conteúdo. No entanto, o próprio termo *semântica* possui, paradoxalmente, diversos significados. Um dos motivos pelos quais isso ocorre é o fato de a semântica ser objeto de estudo de várias disciplinas, tais como: Filosofia, Linguística e Semiótica. Dessa forma, é importante especificar quais são os significados relevantes para o contexto deste trabalho. Além disso, é necessário também discutir como a semântica pode ser expressa, o que leva à reflexão sobre a Representação do Conhecimento (*Knowledge Representation*).

Representar o conhecimento implica em construir uma *representação*; uma forma de representação é um *modelo*, expresso em alguma *linguagem*, no sentido enfocado neste trabalho. Portanto, o grau de formalismo do modelo (ou seja, da representação), depende do grau de formalismo da linguagem. Allemang e Hendler (2011) explicam que quando um modelo “recai sobre as particularidades do contexto do seu leitor para que seu conteúdo seja interpretado, [...] este modelo é *informal*”; isto é, “o modelo carece de um formalismo no qual o significado dos termos [...] podem ser definidos de forma unívoca” (p. 15).

Além da questão do grau de *formalismo* de um modelo, existe outro aspecto importante, que é a *expressividade* de um modelo. Allemang e Hendler (2011) exemplificam este aspecto por meio de um exemplo, no qual são apresentados três modelos de uma molécula de água, com níveis crescentes de expressividade (figura 2).

Figura 2. Diferentes níveis de expressividade de modelos de uma molécula de água



Fonte: adaptado de ALLEMANG, Dean; HENDLER, Jim. Semantic Web for the Working Ontologist: effective modelling in RDFS and OWL. 2<sup>nd</sup> ed. Waltham: Morgan Kaufmann, 2011.

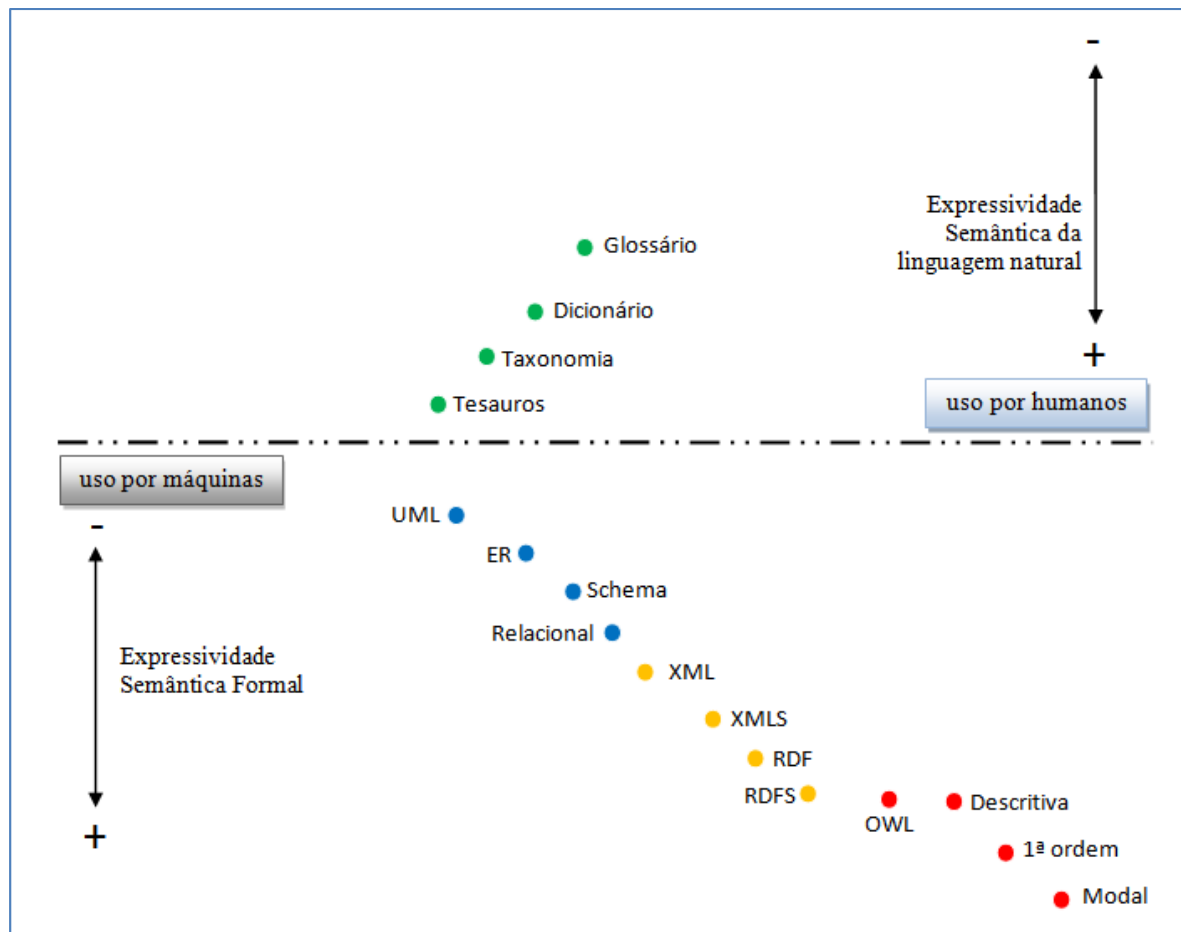
Na figura 2(a), o modelo expressa os elementos que compõe a molécula de água, e suas quantidades; na figura 2(b), além de o modelo expressar as informações do modelo anterior, ele mostra ainda como são as conexões (os dois átomos de hidrogênio ligam-se ao átomo de oxigênio); finalmente, na figura 2(c), o modelo acrescenta informação sobre a geometria da molécula, incluindo o ângulo entre seus átomos constituintes (ALLEMANG e HENDLER, 2011, p. 23).

A Web Semântica busca representar o conhecimento de forma que a semântica seja acessível também às máquinas. Mas que tipo de semântica pode ser acessível às máquinas? Existem semânticas mais ‘pobres’, ou seja, menos expressivas, mas que são mais facilmente computáveis; por outro lado, existem semânticas mais ‘ricas’, mas intratáveis do ponto de vista computacional.

Almeida e Souza (2011), debruçando-se sobre o tema, estudaram a expressividade e o formalismo de alguns instrumentos para organização da informação, com ênfase na Web Semântica. Dessa forma, elaboraram um “espectro semântico” desses instrumentos, exibido na figura 3. Nela é exibido um gráfico composto por dois hemisférios: o “superior” está associado às linguagens naturais; o “inferior”, às linguagens artificiais. Em ambos, há diferentes graus de expressividade semântica, crescentes em relação ao eixo vertical. O foco do presente trabalho está nas linguagens que expressam semântica formal, especialmente aquelas utilizadas na Web Semântica.

Com base na revisão empreendida, os autores buscaram identificar qual abordagem teórica reunia os aspectos mais importantes da representação do conhecimento subjacentes à Web Semântica. A partir do trabalho de Sowa (2004), listaram os modelos propostos na segunda metade do século XX para compreensão da linguagem e para sua implementação em computadores: o *modelo estatístico*, o *modelo sintático*, o *modelo lógico*, o *modelo léxico* e o *modelo neural*. Segundo o autor, cada uma dessas abordagens se fundamenta em uma teoria específica, como estatística, matemática, regras gramaticais, dentre outras, mas ignora aspectos da linguagem, aos quais a tecnologia não consegue se adaptar (*apud* ALMEIDA e SOUZA, 2011).

Figura 3. Espectro semântico dos instrumentos da Web Semântica



Fonte: ALMEIDA, Maurício Barcellos; SOUZA, Renato Rocha. Avaliação do espectro semântico de instrumentos para organização da informação. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação* [1518-2924], v. 16, iss. 31, 2011.

De todo modo, os modelos lógicos, “baseados na lógica filosófica, produziram teorias semânticas formais de qualidade superior em relação às abordagens concorrentes” (p. 35). Elas são particularmente importantes, pois “têm sido amplamente utilizadas nas pesquisas sobre ontologias” (p. 35), as quais possuem “papel relevante no contexto da WS” e que, por isso mesmo, serão vistas no decorrer desta seção. Por isso, neste trabalho, o enfoque de linguagem será o da Lógica, entendida como Semiótica Formal.

Para “representar conhecimento”, como foi dito, é necessário algum tipo de *linguagem*. Linguagem, nesse sentido, transcende o conceito de linguagem verbal. De fato, como ressalta Santaella (2012): “existe uma linguagem verbal, linguagem de sons que veiculam conceitos”; estes sons, “no Ocidente, receberam uma tradução visual alfabética (linguagem escrita)”. No entanto, simultaneamente, “existe uma enorme variedade de outras linguagens que também se constituem



em sistemas [...] de *representação do mundo*” (p. 16, grifos nossos). Essa perspectiva de representação coaduna-se com a Semiótica.

A Semiótica é considerada a “ciência de todas as linguagens” (SANTAELA, 2012). Um dos principais conceitos tratados por essa ciência é o *signo*. De forma simplificada, signo é algo que *representa* outro objeto, podendo ser, inclusive, outro signo. Peirce<sup>5</sup>, citado por Santaela (2012), afirma: “se um lógico for falar das operações da mente, ele deve significar por mente algo bem diferente do objeto de estudo do psicólogo. [...] Uma definição de signo será dada, sem se referir ao pensamento humano” (p. 87). Como bem destaca a autora, Peirce, ao generalizar a noção de signo “a ponto de não ter de referi-la à mente humana [...] soa como antecipação, visto que, com o advento da Cibernética, tal necessidade se patenteou histórica e concretamente” (p. 88). Ou seja, “para falarmos dos processos de comunicação entre máquinas, não temos necessariamente de nos referir às peculiaridades da consciência humana” (p. 88). Esse aspecto casa perfeitamente com o objetivo da Web Semântica, conforme já mencionado: prover “estrutura ao conteúdo inteligível das páginas Web, criando um ambiente onde agentes de software percorreriam página por página e teriam a capacidade de efetuar tarefas sofisticadas para os usuários [humanos]” (BERNERS-LEE *et al.*, 2001, p. 1).

O enfoque da Semiótica formal possui características comuns ao da Teoria dos Modelos o qual, por sua vez, é associado à Semântica Teórica dos Modelos (ALMEIDA e SOUZA, 2011). Esta teoria é utilizada para a elaboração da linguagem RDF – *Resource Description Framework* (W3C, 2004a), uma das linguagens fundacionais da Web Semântica. A Teoria dos Modelos é uma teoria semântica formal que relaciona expressões a interpretações. O termo *teoria dos modelos* advém do uso, tradicional na semântica lógica, na qual uma interpretação satisfatória é chamada *modelo* (W3C, 2004b). Esta teoria, ainda segundo o mesmo documento, “assume que a linguagem se refere ao ‘mundo’ e descreve condições mínimas que o mundo deve satisfazer, de forma a atribuir significado apropriado às expressões da linguagem” (s/p). O documento explica que o é considerado ‘significado’ em uma assertiva (sentença lógica) expressa em RDF ou RDFS pode depender de uma série de fatores, incluindo “convenções sociais, comentários em linguagem natural ou *links* para outros documentos” (s/p). No entanto, os autores do documento ressaltam que muito desses significados podem não ser acessíveis ao processamento por máquinas; com isso, querem enfatizar que o documento em questão não possui a intenção de prover uma análise completa do termo *significado*. Acrescentam que a semântica estabelecida no documento a restringe a uma

---

<sup>5</sup> Charles Sanders Peirce (1839-1914): cientista e filósofo norte-americano. Seus trabalhos apresentam importantes contribuições à lógica, matemática, filosofia e, principalmente à semiótica. Fonte: Wikipedia.

noção *formal* de significado, “o qual poderia ser caracterizado como a parte que é comum às outras acepções de significado, e pode ser capturado por regras mecânicas de inferências” (s/p).

Isso não quer dizer que as linguagens não formais, ou melhor, que outras abordagens das linguagens não são importantes no contexto da Representação do Conhecimento e da Web Semântica: certamente o são. Na verdade, a construção de uma linguagem formal e a expressão de conteúdo nessa linguagem exigem interação entre pessoas e, conseqüentemente, não podem prescindir das linguagens naturais (não formais), uma vez que os modelos (linguagens) estritamente formais “padecem da impossibilidade de lidar com um texto normal, escrito por pessoas para fins de comunicação” (SOWA, 2004 *apud* ALMEIDA e SOUZA, 2011, p. 35-36). Nesse aspecto, como ressalta o próprio Sowa (2011b), as linguagens naturais constituem “o mais sofisticado sistema de comunicação já desenvolvido” (p. 1). Feita a ressalva, é importante registrar que o tipo de Representação do Conhecimento tratado pelo trabalho é aquele *intrínseco* à Web Semântica, ou seja, assentado sobre linguagens formais.

### 2.3 *Linguagens Formais*

Como foi visto, as linguagens naturais possuem alto grau de expressividade. No entanto, não possuem ainda tratamento computacional eficiente; logo, no âmbito da Web Semântica, a qual exige esse tipo de tratamento, elas não são viáveis até o presente momento. Assim sendo, é necessário sacrificar a expressividade, em prol de um maior controle e formalismo. Dessa forma, são criadas *linguagens artificiais* e as *linguagens naturais controladas*.

Conforme explica Mortari (2011), “ao contrário de uma língua, que surge e evolui com um grupo de indivíduos, estando, portanto, em constante mudança, uma linguagem artificial tem uma gramática rigorosamente definida, que não se altera com o passar do tempo” (p. 33). A lógica, prossegue o autor, “faz uso dessas linguagens, também chamadas *linguagens formais*” (p. 33).

Uma espécie de meio-termo entre as linguagens artificiais e as naturais é a *linguagem natural controlada*, que pode ser conceituada como “um subconjunto de uma linguagem natural que possui um mapeamento bem definido de e para uma forma computável” (SOWA, 2011a, p. 3, tradução livre). Forma computável aqui não significa que a linguagem, necessariamente, devesse ser tratada por um computador, mas que possuísse um grau de formalismo suficiente para que houvesse um procedimento sistemático e objetivo capaz de processar conteúdo expresso nessa linguagem. É claro que, hodiernamente, esse procedimento seria transformado em um algoritmo e implementado em um computador.

Segundo Sowa (2011a), Aristóteles criou a primeira linguagem natural controlada (CNL), “a partir de um subconjunto de palavras do grego, para expressar a lógica e os padrões dos silogismos [usados] para raciocinar acerca dessa lógica” (p. 3, tradução livre). Esta CNL permitia expressar não só uma lógica, mas também uma ontologia. A figura 4 apresenta os *padrões de sentenças* da lógica aristotélica, também denominadas *proposições categóricas* (MORTARI, 2001).

Figura 4. Padrões de sentenças da lógica aristotélica

Prefixo	Nome	Padrão
A	Afirmativa universal:	Todo X é Y.
I	Afirmativa particular:	Algum X é Y.
E	Negativa universal:	Nenhum X é Y.
O	Negativa particular:	Algum X não é Y.

Fonte: adaptado de SOWA. John F. Controlled Natural Languages For Representing Ontology. In: 4º Seminário de Pesquisa em Ontologias no Brasil (Ontobras), 2011. Disponível em: <<http://www.inf.ufrgs.br/ontobras-most2011/arquivos/curso1/OntobrasMostSlidesMinicurso1Sowa.pdf>>. Acesso em: 10 fev. 2013.

A partir desses padrões, era possível construir silogismos. O *silogismo*, no sentido aristotélico, é um “raciocínio dedutivo estruturado formalmente a partir de duas proposições, ditas premissas, das quais, por inferência, se obtém necessariamente uma terceira, chamada conclusão” (HOUAISS, 2013). Um dos silogismos propostos, bastante conhecido, chamado *Barbara*<sup>6</sup>, possui a estrutura descrita na figura 5.

Figura 5. Silogismo aristotélico

Prefixo	Sentença
A	Todo animal é um ser vivente.
A	Todo homem é um animal.
A	Logo, todo homem é um ser vivente.

Fonte: adaptado de SOWA. John F. Controlled Natural Languages For Representing Ontology. In: 4º Seminário de Pesquisa em Ontologias no Brasil (Ontobras), 2011. Disponível em: <<http://www.inf.ufrgs.br/ontobras-most2011/arquivos/curso1/OntobrasMostSlidesMinicurso1Sowa.pdf>>. Acesso em: 10 fev. 2013.

O silogismo apresentado possui duas *premissas* e uma *conclusão*, a qual é uma *consequência lógica*<sup>7</sup> das premissas. Quaisquer outros pares de premissas cuja *estrutura* seja análoga à apresentada levam ao mesmo tipo de conclusão, utilizando esse silogismo como base (Todo X é Y. Todo Y é Z. Logo, todo X é Z). Portanto, Aristóteles formalizou, de maneira consistente, *alguns* tipos de raciocínio, por meio dos silogismos.

A lógica criada por Aristóteles foi aperfeiçoada ao longo dos séculos<sup>8</sup>. A ela foram acrescentadas:

- Padrões de linguagem natural que usam os termos “e”, “ou”, “não”, “se”.

<sup>6</sup> Os demais silogismos são descritos em Sowa (2011a).

<sup>7</sup> Consequência lógica é a relação entre uma sentença e outra sentença que decorre dela (RUSSEL e NORVIG, 2004).

<sup>8</sup> O trabalho citado (SOWA, 2011a) apresenta uma síntese da evolução da lógica aristotélica.

- Silogismo hipotético: padrões de raciocínio baseados em “se-então”.
- Silogismo disjuntivo: padrões de raciocínio baseados em “ou-ou” (ou exclusivo – exemplo: um polo magnético ou é positivo ou é negativo).

Os aperfeiçoamentos resultaram na chamada *lógica proposicional*. A *sintaxe* da lógica proposicional define as sentenças permitidas (ou seja, aquelas que são sintaticamente corretas). Essas sentenças podem ser atômicas ou complexas. As *sentenças atômicas* consistem em um único símbolo proposicional, o qual representa uma *proposição*, que pode ser *verdadeira* ou *falsa*. Existem ainda dois símbolos proposicionais com significados fixos: *Verdadeiro* é a proposição sempre verdadeira e *Falso* é a proposição sempre falsa. Já as *sentenças complexas* são construídas a partir de sentenças mais simples, por meio dos *conectivos lógicos* (RUSSEL e NORVIG, 2004). A tabela 2 apresenta os cinco conectivos de uso comum na lógica proposicional.

*Tabela 2. Conectivos de uso comum na lógica proposicional*

Símbolo	Descrição
$\neg$ (não)	Uma sentença como $\neg P$ é chamada <i>negação</i> de P.
$\wedge$ (e)	Uma sentença cujo principal conectivo é $\wedge$ , como $P \wedge Q$ (lê-se P e Q), é chamada <i>conjunção</i> .
$\vee$ (ou)	Uma sentença que utiliza $\vee$ , como $P \vee Q$ (lê-se P ou Q), é chamada <i>disjunção</i> .
$\Rightarrow$ (implica)	Uma sentença como $P \Rightarrow Q$ (lê-se P implica Q), é chamada <i>implicação</i> . Sua <i>premissa</i> ou <i>antecedente</i> é P; sua <i>conclusão</i> ou <i>consequente</i> é Q. As implicações são conhecidas como <i>regras</i> ou <i>declarações se-então</i> .
$\Leftrightarrow$ (se e somente se)	Uma sentença como $P \Leftrightarrow Q$ (lê-se P se e somente se Q) é uma bicondicional. Esse conectivo é logicamente equivalente à expressão $(P \Rightarrow Q) \wedge (Q \Rightarrow P)$ .

Fonte: adaptado de RUSSEL, Stuart; NORVIG, Peter. *Inteligência Artificial*. 2. ed. Rio de Janeiro: Campus, 2004.

A *semântica*, em lógica proposicional, apenas “define as regras para determinar a *verdade* de uma sentença com respeito a um *modelo* específico” (RUSSEL e NORVIG, 2004, p. 200, grifos nossos). Um modelo, nesse contexto, simplesmente fixa o valor-verdade (verdadeiro ou falso) para todo símbolo proposicional. O cálculo do valor-verdade de uma sentença é feito recursivamente com base nos valores dos componentes da sentença e dos conectivos. As regras para cada conectivo podem ser resumidas em uma tabela-verdade, a qual especifica o valor-verdade de uma sentença complexa para cada atribuição possível de valores-verdade a seus componentes (RUSSEL e NORVIG, 2004, p. 200-201). A tabela 3 apresenta os valores-verdade para os cinco conectivos da lógica proposicional.

*Tabela 3. Tabela-verdade para os conectivos da lógica proposicional*

P	Q	$\neg P$	$P \wedge Q$	$P \vee Q$	$P \Rightarrow Q$
‘falso’	‘falso’	‘verdadeiro’	‘falso’	‘falso’	‘verdadeiro’
‘falso’	‘verdadeiro’	‘verdadeiro’	‘falso’	‘verdadeiro’	‘verdadeiro’
‘verdadeiro’	‘falso’	‘falso’	‘falso’	‘verdadeiro’	‘falso’
‘verdadeiro’	‘verdadeiro’	‘falso’	‘verdadeiro’	‘verdadeiro’	‘verdadeiro’

Fonte: RUSSEL, Stuart; NORVIG, Peter. *Inteligência Artificial*. 2. ed. Rio de Janeiro: Campus, 2004.

O problema da lógica proposicional é que ela é limitada mesmo para expressar raciocínios bastante simples. A esse respeito, Krause (2011) fornece o seguinte exemplo: o raciocínio ‘se uma vaca é um animal, então um chifre de vaca é um chifre de um animal’ não pode ser expresso *corretamente* pela teoria do silogismo aristotélico – ver figura 6.

*Figura 6. Exemplo de raciocínio que não pode ser expresso corretamente em lógica proposicional*

<b>Raciocínio</b>	<b>‘Toda vaca é um animal; logo, todo chifre de vaca é também um chifre de animal’.</b>
<b>A</b>	Denota ‘vaca’
<b>B</b>	Denota ‘animal’
<b>C</b>	Denota ‘chifre de vaca’
<b>D</b>	Denota ‘chifre de animal’
<b>Premissa</b>	Todo A é B.
<b>Conclusão</b>	Todo C é D.

Fonte: adaptado de KRAUSE, Décio. Tópicos em Ontologia Analítica. Florianópolis: UFSC, 2011. Disponível em: <<http://www.cfh.ufsc.br/~dkrause/pg/TopicosOntologia.pdf>>. Acesso em: 14 fev. 2013.

Krause explica que, “quando usamos variáveis (como *A*, *B*, *C* e *D*), explicitamos a forma do raciocínio, e ele deve permanecer válido sempre que substituirmos as variáveis por sujeitos e predicados específicos” (p. 68). No caso do exemplo da figura 6 isso não ocorre. O próprio autor fornece um contraexemplo, onde *A* = ‘morcego’, *B* = ‘mamífero’, *C* = ‘brasileiro’ e *D* = ‘jogador de futebol’; daí decorre: ‘Todo morcego é mamífero, logo, todo brasileiro é jogador de futebol’, o qual, claramente, é um raciocínio inválido.

Prosseguindo com a evolução da lógica, George Boole, no século XIX, criou a hoje chamada álgebra booleana. Ela de um caráter algébrico à lógica proposicional, mas ainda padecia das mesmas limitações. Apesar disso, a contribuição de Boole foi essencial para a criação dos circuitos lógicos, base dos microprocessadores atuais.

No entanto, segundo Mortari (2001), o maior avanço para a lógica contemporânea veio com a obra de Frege: “ao contrário de Aristóteles, e mesmo de Boole, que procuravam identificar as formas válidas de argumento, a preocupação de Frege era a sistematização do raciocínio matemático”; ainda segundo o autor, “Frege procurou formalizar as regras de demonstração [matemática], iniciando com regras bem simples, sobre cuja aplicação não houvesse dúvidas”. Krause (2009) acrescenta que, a partir dos trabalhos de Frege, “inicia-se uma visão linguística da lógica, em distinção à visão algébrica do período booleano” (p. 9).

A formalização da lógica teve ainda importantes contribuições dos trabalhos de Peirce

(mencionado na seção 1.2) e Peano<sup>9</sup>, os quais criaram notações baseadas em símbolos. Isso fez “com que a lógica contemporânea passasse a ser denominada ‘simbólica’” (MORTARI, 2001, p. 29). Ou seja, a lógica pôde ser expressa por *linguagens artificiais*. Uma das primeiras lógicas simbólicas é a chamada *lógica de primeira ordem* (LPO), também conhecida por *cálculo de predicados*.

Sowa (2011a) recorre a um exemplo simples para ilustrar as notações propostas por esses autores. Ele consiste em representar a sentença<sup>10</sup> “*The cat is on mat*” (O gato está no capacho).

Na notação de Peirce, a expressão seria:

$$\Sigma_x \Sigma_y \text{Cat}_x \bullet \text{Mat}_y \bullet \text{On}_{x,y}$$

Já na notação proposta por Peano:

$$\exists x \exists y \text{Cat}(x) \wedge \text{Mat}(y) \wedge \text{On}(x,y)$$

A LPO, além de prover maior formalismo, tem maior expressividade do que a lógica proposicional. Isso permite, por exemplo, expressar corretamente o raciocínio da figura 6, conforme pode ser visto na figura 7.

Figura 7. Raciocínio expresso corretamente em LPO

Raciocínio	‘Toda vaca é um animal; logo, todo chifre de vaca é também um chifre de animal’.
$V(x)$	x é uma vaca
$A(x)$	x é um animal
$C(x, y)$	x é chifre de y
<b>Premissa</b>	$\forall x(V(x) \rightarrow A(x)).$
<b>Conclusão</b>	$\forall z(\forall x(V(x) \wedge C(z, x)) \rightarrow \forall x(A(x) \rightarrow C(z, x)).$

Fonte: adaptado de KRAUSE, Décio. Tópicos em Ontologia Analítica. Florianópolis: UFSC, 2011. Disponível em: <<http://www.cfh.ufsc.br/~dkrause/pg/TopicosOntologia.pdf>>. Acesso em: 14 fev. 2013.

Além disso, do ponto de vista de várias áreas da ciência e da tecnologia, incluindo a Web Semântica, a lógica simbólica é crucial, pois se abriu a possibilidade de se representar um corpo de conhecimento, e de fazer inferências lógicas sobre ele, por meio de uma linguagem artificial:

*Embora o objetivo inicial da lógica fosse a análise de argumentos, o uso de linguagens artificiais ampliou seu âmbito de atuação: as linguagens de lógica passaram a ter outros usos. Por exemplo, passamos a poder representar informação em geral por meio delas. Hoje em dia, nota-se o grande papel da lógica em investigações científicas de ponta, como é o caso da Inteligência Artificial, particularmente nas áreas de representação do conhecimento e demonstração automática (MORTARI, 2001, p. 30, grifos nossos).*

<sup>9</sup> Giuseppe Peano (Spinetta, Piemonte, 27 de Agosto de 1858 — Turim, 20 de Abril de 1932) foi um matemático italiano que fez importantes contribuições teóricas nas áreas de análise matemática, lógica, teoria dos conjuntos, equações diferenciais e análise vetorial. Fonte: Wikipedia.

<sup>10</sup> As sentenças em lógica são sentenças *declarativas*.

Com base em Mortari (2001), foi elaborada uma tabela que mostra algumas etapas para traduzir sentenças em linguagem natural associadas às categorias aristotélicas para a lógica de primeira ordem, com notação Peirce-Peano.

*Tabela 2. Tradução de sentenças em linguagem natural para a lógica formal.*

<i>Categoria</i>	<i>Linguagem Natural (1)</i>	<i>Linguagem Natural (2)</i>	<i>Linguagem Formal</i>
<b>Afirmativa particular</b>	Alguns peixes são azuis.	Há ao menos um x tal que: x é um peixe e x é azul.	$\exists x(P_x \wedge A_x)$
<b>Negativa particular</b>	Algum pinguim não é azul.	Há ao menos um x tal que: x é um pinguim e x não é azul.	$\exists x(P_x \wedge \neg A_x)$
<b>Afirmativa universal</b>	Qualquer peixe é azul.	Para qualquer x, se x é um peixe então x é azul.	$\forall x(P_x \wedge A_x)$
<b>Negativa universal</b>	Nenhum peixe é azul.	Não existe x tal que: x é peixe e x é azul.	$\forall x(P_x \wedge \neg A_x)$

Fonte: adaptado de: MORTARI, Cezar A. **Introdução à lógica**. São Paulo: Unesp, 2001.

Pelos exemplos dados, é possível identificar que a utilização dos quantificadores e a possibilidade de se expressar relações foram as grandes contribuições da LPO. Na próxima seção, serão apresentadas as lógicas de descrições, as quais são subconjuntos da LPO e importantes na definição das ontologias, no contexto da Web Semântica.

## 2.4 As Lógicas de Descrições (DL) e as Ontologias

No contexto da Web, que é um sistema de informação *aberto*, dois conjuntos de dados “podem utilizar terminologias diferentes para referir-se à mesma informação”; por outro lado, existe também a possibilidade “de uma mesma terminologia estar sendo utilizada com significados diferentes, por aplicações distintas” (MORAIS e SOARES, 2004, p. 3). É necessário, portanto, um mecanismo de possibilite a *interoperabilidade semântica* entre os recursos. Segundo a Comissão Europeia (EC, 2010, p. 23):

*A interoperabilidade semântica diz respeito ao significado dos elementos de dados [recursos, na WS] e aos relacionamentos entre eles. Isso inclui o desenvolvimento de vocabulários que descrevam as trocas de dados, e assegure que os elementos sejam compreendidos da mesma forma por entre as partes que estão se comunicando.*

A interoperabilidade semântica depende da *interoperabilidade sintática*, que “diz respeito à descrição do formato exato da informação a ser trocada em termos de gramática, formato e esquemas” (EC, 2010, p. 23).

Na WS, o papel da interoperabilidade semântica cabe às ontologias. Ou seja, uma ontologia estabelece um vocabulário controlado (terminologia) e uma rede de relacionamentos para esse conjunto específico de conceitos ou classes, na forma de uma *rede semântica*. No entanto, para que essa rede seja processável por dispositivos computacionais, é necessário haver uma representação

formal.

Como asseveram Vieira *et al* (2008), tem havido muitas pesquisas no campo da representação do conhecimento aliadas ao raciocínio automático. O foco é a definição de linguagens formais para representar o conhecimento, combinadas a métodos que permitam fazer inferências a partir do conhecimento exposto nessas linguagens. Os *métodos de inferências* têm origem na lógica formal, a qual foi descrita, de forma resumida, na seção anterior.

Sowa (2011a) afirma que a lógica de primeira ordem (LPO) é um subconjunto ou superconjunto da maior parte das linguagens usadas para representação do conhecimento. Ademais, a LPO é também um subconjunto de todas as linguagens naturais. As principais linguagens artificiais utilizadas na Web Semântica – RDF, RDFS e a maior parte da OWL – são compatíveis com a LPO.

As chamadas Lógicas de Descrições, ou lógicas descritivas, são ‘famílias’ de linguagens derivadas da LPO. Elas constituem uma “evolução dos formalismos de representação do conhecimento baseados em objetos (como as redes semânticas...), ao qual corresponde um *subconjunto estruturado da lógica de primeira ordem*” (VIEIRA *et al.*, p. 22, grifos nossos).

Os autores esclarecem que as Lógicas de Descrição (DL – *Description Logics*) tem menor expressividade que a LPO. No entanto, possuem duas características fundamentais: “grande capacidade de representação para sistemas baseados em conhecimento” e “métodos de dedução [computacionalmente] eficientes para os serviços de raciocínio associados” (p. 22)<sup>11</sup>. A OWL (*Ontology Web Language*), que é a linguagem adotada pelo W3C para a definição de ontologias no contexto da Web Semântica, é baseada em DL. Dessa forma, essas linguagens são suficientes para os objetivos deste trabalho.

Antes de prosseguir com o tema no âmbito da Web Semântica, faz-se necessária uma breve digressão sobre a origem do termo *ontologia*. Conforme Segaran *et al* (2011), desde Aristóteles, filósofos tem-se preocupado sobre conhecer o *quê existe* e como *descrevê-lo*. Uma vez que se pensa saber que *algo* existe, o próximo passo é situá-lo *entre todas as outras coisas* (p. 127, tradução livre). No domínio da filosofia, prosseguem os autores, tal esforço é chamado ontologia.

Uma das formas de situar algo entre outras coisas é por meio da elaboração de categorias. Para ilustrar, a figura 8 apresenta as categorias da ontologia proposta por Aristóteles. Elas expressam *propriedades* das coisas e *relações* entre elas. Originalmente, eram baseadas na

---

<sup>11</sup> Um desses procedimentos é o algoritmo Tableau, descrito em Mortari (2001).



semântica da Língua Grega. Para cada categoria, há uma questão correspondente, que auxilia sua definição (SOWA, 2011a).

*Figura 8. Categorias da ontologia de Aristóteles*

Substância – O quê é isso?
Relação – Em direção ao quê?
Quantidade – O quanto?
Qualidade – Que tipo?
Agente – Fazendo o quê?
Paciente – Esperando o quê?
Posse – Tendo o quê?
Situação – Situado como?
Lugar – Onde?
Tempo – Quando?

Fonte: adaptado de SOWA, John F. Controlled Natural Languages For Representing Ontology. In: 4º Seminário de Pesquisa em Ontologias no Brasil (Ontobras), 2011. Disponível em: <<http://www.inf.ufrgs.br/ontobras-most2011/arquivos/curso1/OntobrasMostSlidesMinicurso1Sowa.pdf>>. Acesso em: 10 fev. 2013.

A Ciência da Computação, em especial a Inteligência Artificial, passou a usar o termo *ontologia* para descrever modelos da realidade utilizados para expressar semântica em sistemas computacionais. Nesse contexto, Ontologia, em Filosofia, é algo muito mais amplo, como explica Chauí (1995): “a ontologia investiga a essência ou sentido do ente físico ou natural, do ente psíquico, lógico, matemático, estético, ético, temporal, espacial, etc.” (p. 242).

No âmbito da Ciência da Computação, ontologia é definida como “uma especificação explícita e formal de uma conceitualização compartilhada” (STUDER *et al.*, 1998 *apud* GUARINO, 2009). Freitas (2003) propõe a seguinte explicação para a definição: (1) *especificação explícita* são as definições de conceitos, instâncias, relações, restrições e axiomas; (2) *formal* porque é definida por meio de alguma linguagem declarativa baseada na lógica formal – portanto, é compreensível para agentes e sistemas; (3) *conceitualização*: trata-se de um modelo abstrato de uma área do conhecimento (domínio) ou de um universo limitado do discurso; (4) *compartilhada*: trata-se de um conhecimento consensual, por exemplo, uma terminologia comum do domínio modelado.

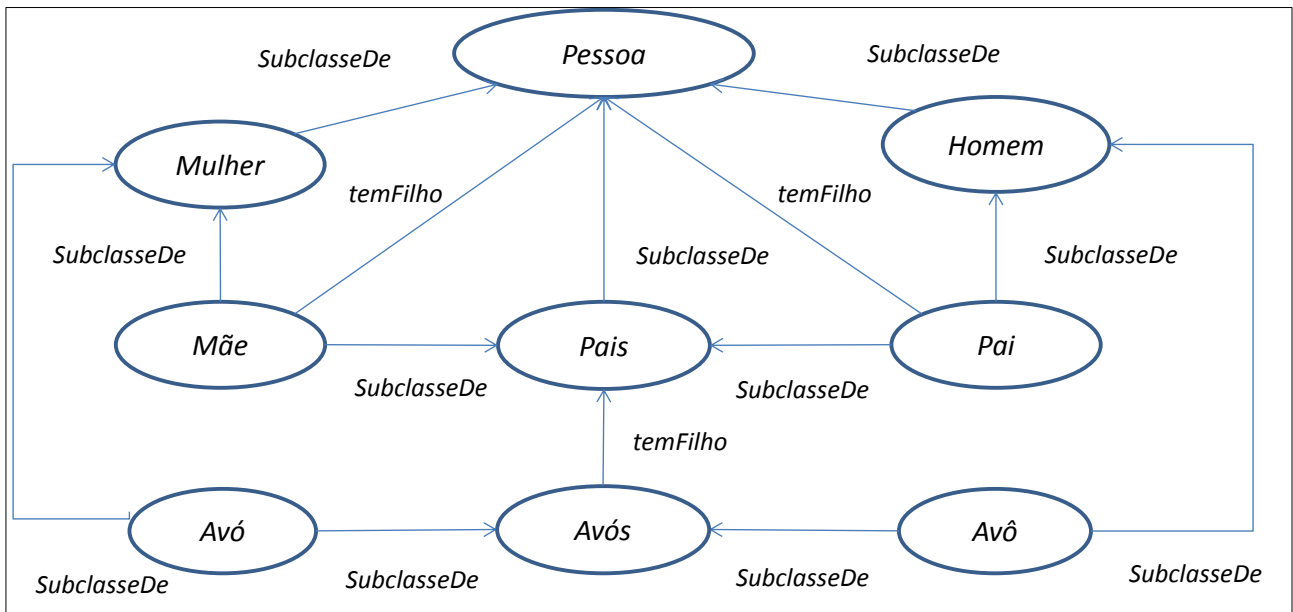
Analogamente às ontologias da filosofia clássica, os sistemas de representação do conhecimento modernos também utilizam as categorias. Conforme Russel e Norvig (2004), as *categorias* “são os principais blocos de construção de qualquer esquema de representação de conhecimento em grande escala” (p. 338). Os autores explicam ainda que existem duas famílias de sistemas relacionadas para representação por meio de categorias: *redes semânticas* e *lógicas descritivas*.

Um dos trabalhos fundamentais para o desenvolvimento das redes semânticas foi a técnica de *grafos*<sup>12</sup> *existenciais*, desenvolvida por Peirce (vide seção 1.2.1). Ela era baseada em uma notação gráfica de nós e arcos, que auxiliava a visualização de uma base de conhecimento. Além disso, ela provia “algoritmos eficientes para dedução de propriedades de um objeto, de acordo com sua pertinência a uma categoria” (RUSSEL e NORVIG, p. 338). Os autores ressaltam que as redes semânticas bem definidas são uma forma de lógica. A figura 9 apresenta um exemplo de rede semântica (a mesma ontologia é expressa, mais adiante, em lógica descritiva).

---

<sup>12</sup> Grafos são estruturas de representação de dados semelhantes a árvores, porém ainda mais poderosos. Na verdade, uma árvore é uma espécie de grafo, onde não são permitidos ciclos ou circuitos fechados (ou seja, a partir de um nó da árvore, percorrer um 'caminho' e chegar ao mesmo nó). Nos grafos, isso é possível.

Figura 9. Ontologia simples expressa em uma rede semântica



Fonte: elaborado pelo autor com base em Russel e Novig (2004) e Vieira *et al* (2011)

Já as lógicas descritivas, ou Lógicas de Descrições (DL – *Description Logics*), “fornecem uma linguagem formal para construção e combinação de definições de categorias e algoritmos eficientes para definir relacionamentos de subconjuntos e superconjuntos entre categorias” (RUSSEL e NORVIG, 2004, p. 338). A sintaxe da DL é formada por símbolos representando conceitos e papéis, construtores e quantificadores. Os conceitos representam classes; uma *classe* é um conjunto de indivíduos que possuem características comuns. Os conceitos dividem-se em conceitos primitivos e conceitos complexos. Um *conceito primitivo*, como o nome indica, não depende de outros conceitos para ser definido; já um *conceito complexo* é formado a partir de conceitos primitivos. Os *construtores* são “operadores que permitem a criação de conceitos complexos, dando um significado especial à interpretação do conceito”. Os *papéis* são propriedades dos conceitos e representam relacionamentos entre os elementos da base de conhecimento (conceitos e indivíduos). Finalmente, os *quantificadores* são operadores que quantificam os papéis (VIEIRA *et al*, 2008).

De maneira um pouco mais formal, uma *interpretação* de um conceito  $A$ , representada por  $A^I$ , pode ser definida como um conjunto de valores que torna este conceito verdadeiro. O conceito mais geral, do qual todos os outros conceitos são subconceitos, é representado por  $D^I$ , e é equivalente ao valor Verdadeiro (*True*). O Verdadeiro e o Falso são representados, respectivamente, pelos símbolos  $\top$  e  $\perp$  (VIEIRA *et al*, 2008). Supondo que  $A$  e  $B$  sejam nomes de conceitos e  $P$  é um nome de propriedade, podem-se ter os construtores descritos na Tabela 3.

Tabela 3. Construtores de uma lógica descritiva.

Construtor: Notação	Descrição	Exemplo
<b>Conjunção:</b> $A \sqcap B$	Representa a <i>interseção</i> entre as interpretações dos conceitos que fazem parte da conjunção.	$Humano \sqcap Masculino$ Um indivíduo pertence a $Humano \sqcap Masculino$ sss ele é Humano (pertence ao conceito Humano) e é Masculino (pertence ao conceito Masculino).
<b>Disjunção:</b> $A \sqcup B$	Representa a <i>união</i> entre as interpretações dos conceitos que fazem parte da disjunção.	$Médico \sqcup Advogado$ Um indivíduo pertence a $Médico \sqcup Advogado$ sss ele é Médico ou Advogado.
<b>Quantificação Universal:</b> $\forall P.A$	Determina que, por meio da propriedade $P$ , todos os indivíduos do conceito declarado devem se relacionar com a classe determinada pelo quantificador universal (conceito $A$ ). Os indivíduos do conceito declarado podem se relacionar com outras classes por meio da própria propriedade $P$ .	$\forall \text{temFilho. Professor}$ Um indivíduo, p.ex., João, pertence à $\forall \text{temFilho. Professors}$ todas as crianças (i.e., todos os indivíduos relacionados a um indivíduo por meio da propriedade $\text{temFilho}$ ) forem professores.
<b>Quantificação Existencial:</b> $\exists P.A$	Determina que, por meio da propriedade $P$ , o indivíduo do conceito declarado deve se relacionar com a classe determinada pelo quantificador existencial (conceito $A$ ). O indivíduo do conceito declarado pode se relacionar com outras classes por meio da própria propriedade $P$ .	$\exists \text{éCasadoCom. Médico}$ Um indivíduo, p.ex., João, pertence à $\exists \text{éCasadoCom. Médicos}$ existe um indivíduo que é casado com João (i.e., é relacionado com João pela propriedade $\text{éCasadoCom}$ ) e também é médico (i.e., pertence ao conceito Médico).
<b>Negação:</b> $\neg A$	Representa a representação de $D^I$ menos a representação de $A^I$ . Por isso, a Negação também é chamada de Complemento.	$\neg \text{Professor}$ Um indivíduo pertence a $\neg \text{Professor}$ sss ele pertence a qualquer um dos conceitos do domínio, <i>exceto</i> o conceito Professor.
<b>Restrição de Número:</b> $=<n.P$ e $>=n.P$	Determinam o número máximo ( $=<$ ) e mínimo ( $>=$ ) de relacionamentos que devem ser realizados por meio da propriedade $P$ .	$=<1.\text{temFilho}$ Um indivíduo pertence a $=<1.\text{temFilho}$ sss ele tem, no máximo, um filho (i.e., sss a propriedade $\text{temFilho}$ ocorre somente uma vez para aquele indivíduo).

Fonte: VIEIRA *et al*, 2008; BAADER *et al*, 2003.

Nota: sss: se e somente se.

Uma *base de conhecimento (knowledge base)* contém um conjunto de informações acerca de determinado domínio. Segundo Vieira *et al* (2008), para melhor representar este conhecimento, é necessário que a base seja dividida em conhecimento intensional e conhecimento extensional.

O *conhecimento intensional (intensional knowledge)* é o “conhecimento geral sobre o domínio do problema”, ou seja, “representa o conhecimento sobre os grupos (conjuntos) de indivíduos que apresentam as mesmas características” (VIEIRA *et al*, 2008). Esses conjuntos são chamados de categorias ou classes. As categorias “servem para organizar e simplificar a base de conhecimento por *herança*” (RUSSEL e NORVIG, 2004). Os autores fornecem um exemplo: se for dito que todos os indivíduos (instâncias) da categoria (classe) *Alimento* são comestíveis e que todos os indivíduos da categoria *Maçã* pertencem também à categoria *Alimento*, é válido concluir que uma maçã (instância de *Maçã*) é comestível. Prosseguem os autores: “as relações de subclasses organizam as categorias em uma *taxonomia*” (p. 312). As taxonomias são utilizadas em várias

outras áreas, como a Biologia, a Biblioteconomia e a Ciência da Informação. Existem outras relações importantes, além das relações de subclasses, como a de composição física (uma coisa é parte de outra coisa – ‘o motor é parte do carro’), por exemplo; tal relação, aliás, é bastante usada em diagramas de classes, na Engenharia de Software.

Já o *conhecimento extensional* (*extensional knowledge*) “especifica um problema em particular”, ou seja, “representa o conhecimento sobre cada indivíduo que faz parte do conjunto” (p. 24). Dito de outra forma, o conhecimento extensional define as instâncias, associando-as às classes e suas relações. Uma assertiva comum é aquela que define um indivíduo como uma instância de uma classe (GIACOMO e LENZERINI, 1996).

Em Lógica de Descrições, o conhecimento intensional é tradicionalmente denominado inTensional Box – TBox; o extensional, Asserted Box – ABox (GIACOMO e LENZERINI, 1996). As figuras 10 e 11 fornecem exemplos de descrições nessas linguagens.

*Figura 10. Exemplo de conhecimento intensional expresso em lógica descritiva TBox*

<p>Mulher <math>\equiv</math> Pessoa <math>\sqcap</math> Fêmea  Homem <math>\equiv</math> Pessoa <math>\sqcap</math> <math>\neg</math>Mulher  Pais <math>\equiv</math> Pessoa <math>\sqcap</math> <math>\exists</math>temFilho.Pessoa  Mãe <math>\equiv</math> Pessoa <math>\sqcap</math> Pais  Pai <math>\equiv</math> Homem <math>\sqcap</math> Pais  Avós <math>\equiv</math> Pessoa <math>\sqcap</math> <math>\exists</math>temFilho.Pais  Avó <math>\equiv</math> Mulher <math>\sqcap</math> Avós  Avô <math>\equiv</math> Pai <math>\sqcap</math> Avós</p>
--

Fonte: adaptado de VIEIRA et al (2008).

A figura 10 expressa conhecimento acerca de *classes* (conhecimento *intensional*); para efeito de comparação, ela representa a mesma semântica da rede mostrada, anteriormente, na figura 9. Como Vieira et al. (2008) explicam (p. 24):

*os conceitos Pessoa e Fêmea [...], definidos pelo próprio nome, são chamados de conceitos-base. O conceito Mulher é declarado como o conceito resultante da interseção entre as interpretações dos conceitos Pessoa e Fêmea, ou seja, o conceito Mulher é formado pelos indivíduos que pertencem aos conceitos Pessoa e Fêmea simultaneamente.*

A figura 11 descreve *indivíduos* dessas classes (conhecimento *extensional*). Outra observação importante é que, dada a forma como o conhecimento sobre o domínio em questão (família) foi expresso, é possível que um mecanismo de raciocínio automático faça inferências e ‘deduza’ fatos (sentenças) que não estão explícitas. Por exemplo, o mecanismo pode inferir que Paulo é avô, ou seja, que o indivíduo *Paulo* pertence à classe *Avô*, uma vez que é *Homem* e tem um

filho – Mauro, que é pai de alguém – no caso, Pedro.

Figura 11. Exemplo de conhecimento extensional expresso em lógica descritiva ABox

Mulher(Ana)
Mulher (Joana)
Mulher(Maria)
Homem(Mauro)
Homem(Paulo)
Homem(Pedro)
temFilho(Mauro, Pedro)
temFilho(Mauro, Ana)
temFilho(Paulo, Mauro)
temFilho(Joana, Maria)
temFilho(Maria, Pedro)
temFilho(Maria, Ana)

Fonte: VIEIRA et al (2008).

Embora a explicação dos mecanismos de inferência não seja foco do presente trabalho, é importante, ao menos, destacar quais são os possíveis tipos de inferência automática no contexto da Lógica das Descrições. A forma mais simples é chamada *subsunção* e consiste em verificar se uma expressão sempre denota um subconjunto de objetos denotados por outra expressão; ou seja, verificar se uma categoria é um subconjunto de outra pela comparação de suas definições. Uma segunda forma é a *classificação*: verificar se um objeto pertence a uma categoria. Existem ainda outras, mais complexas, como verificar se uma determinada assertiva é logicamente implicada por uma base de conhecimento (RUSSEL e NORVIG, 2004; GIACOMO e LENZERINI, 1996).

Portanto, bases de conhecimento expressas em Lógica Descritiva possibilitam a definição de ontologias, entendidas aqui como descrições explícitas e compartilhadas de conhecimentos acerca de determinado domínio. Essas ontologias, ao seu turno, possibilitam a interoperabilidade entre as camadas da Web Semântica e a reutilização da informação *estruturada*, a qual expressa a *representação do conhecimento* (ou seja, um *modelo*) de um domínio, conforme explica Freitas (2003):

*O engajamento ontológico ou de conhecimento, ou seja, a semântica das sentenças codificadas em lógica guarda uma relação mais direta com o domínio modelado e permite o reuso de um grande volume de informação que desempenha o papel de conhecimento estruturado, disponível para reuso em larga escala por sistemas e programas.*

No contexto da Web Semântica, existem diversas linguagens para a representação do conhecimento; cada uma delas acrescenta um maior grau de expressividade. As linguagens são:

RDF, RDFS, SPARQL e OWL (ALLEMANG e HENDLER, 2011).

A linguagem recomendada pelo W3C e mais utilizada para a descrição de ontologias é a OWL (*Web Ontology Language*); ou seja, ela é o principal componente da camada *ontologia* da Web Semântica. Sua definição oficial é dada pelo próprio W3C (2012): “OWL é uma *linguagem* [...] projetada para *representar conhecimento* rico e complexo sobre coisas, grupo de coisas e relações entre coisas” (s/p, tradução livre, grifos nossos). O documento complementa a definição, asseverando que a OWL é uma linguagem computacional baseada em lógica, na qual o conhecimento expresso pode ser processado por programas de computador, ou para verificar a consistência do conhecimento expresso, ou para tornar explícitos conhecimentos implícitos, ou seja, derivar conclusões a partir de premissas.

Explicações mais detalhadas acerca das linguagens mencionadas, bem como questões tecnológicas serão tratadas no próximo capítulo. Antes, porém, é importante mencionar que todas as linguagens citadas foram definidas pelo W3C (*World Wide Web Consortium*), o que significa que elas constituem um padrão para a Web. Ademais, são abertas e gratuitas, ou seja, sua especificação está disponível e não é necessário se pagar *royalties* para usá-las. Portanto, pessoas ou organizações podem utilizá-las livremente para a criação de ontologias e bases de conhecimento. Isso implica que, nesse contexto, “passa-se de um domínio de conceitos isolados na Web, próprios de uma aplicação específica, para conceitos universalmente conhecidos entre as aplicações” (MORAIS e SOARES, 2004).

Assim, pelo que foi exposto ao longo desta seção, a Web Semântica, ao contrário do que se poderia supor, não é meramente um conjunto de tecnologias; na verdade, ela possui um forte embasamento teórico, que remonta à Filosofia Clássica, passando pela Matemática, Lógica, Linguística, Semiótica e Ciência da Computação, entre outras. No entanto, como em outras áreas da atuação humana, não se pode prescindir da tecnologia para a materialização de ideias, projetos e processos. O próximo capítulo explora as principais tecnologias de suporte à Web Semântica.

### 3. AS TECNOLOGIAS DA WEB SEMÂNTICA

*...quando se une a prática com a teoria tem-se a práxis, a ação criadora e modificadora da realidade.*

Paulo Freire

#### 3.1 *Identificador Universal de Recursos: URI*

Como foi mencionado, o projeto da Web Semântica consiste em diversas camadas interdependentes, onde cada uma serve de suporte às demais. Assim, a camada de infraestrutura, fornece os componentes básicos, denominados recursos, e os meios pelos quais os recursos serão referenciados e/ou descritos pelos componentes das demais camadas.

Um *recurso*, no contexto da Web Semântica, é qualquer *conteúdo* que tenha *identidade*. Assim, um recurso pode ser uma simples sequência de caracteres, uma imagem, um documento eletrônico completo ou mesmo uma coleção de outros recursos. A *referência* a um recurso na Rede Mundial de Computador é feita por um identificador único, denominado URI (do inglês, *Universal Resource Identifier* – Identificador Universal de Recursos). Interessante notar que, em termos semióticos, um URI é um signo que *refere* a um *objeto* (ou *conteúdo*). Um exemplo comum é um endereço de um sítio eletrônico (*web site*), chamada URL (do inglês, *Universal Resource Locator* – Localizador Universal de Recursos): trata-se, na verdade, de um URI.

Do ponto de vista técnico, o URI é um identificador único, em formato padronizado, para um recurso disponível na Rede Mundial de Computadores (Internet). Através do protocolo HTTP (*HyperText Transfer Protocol* – Protocolo de Transferência de Hipertexto) podemos acessar um URI na Web a fim de obter as informações sobre o recurso referenciado. A sintaxe de uma HTTP URI permite que todas as informações nela contidas (nome do servidor, protocolo, número da porta, nome do arquivo) sejam usadas para localizar um arquivo de documento (ou a localização de um arquivo) na Web (ALLEMANG e HENDLER, 2011).

A formação de um URI demanda o uso de *namespaces* (espaços de nomes). Eles nomeiam e especificam contextos para os recursos, o que permite, dentre outras coisas, evitar ambiguidade entre eles. Por exemplo, o termo “laranja” pode estar em dois domínios (cores e frutas). Nesse caso, deveria ser criado um *namespace* para cada contexto. Logo, os *namespaces* são especialmente úteis quando se deseja combinar em um só documento metadados oriundos de diferentes origens. Também são úteis quando diferentes organizações necessitam compartilhar recursos, mas possuem metadados cujas marcações coincidem (PEDRAZA-JIMENEZ *et al*, 2011).



Os *namespaces* padrão são determinados pela W3C. Assim, tem-se `xsd:` para definição de esquema XML; `xmlns:` para definição de *namespaces* XML; `rdf:` para indicar os identificadores usados no RDF (que será explicado a seguir) e assim por diante (ALLEMANG e HENDLER, 2011). Importante notar que tanto um URL quanto um URI podem se referenciar a um mesmo termo; é o caso de `<http://www.w3.org/2000/01/rdf-schema#subClassOf>` e `<rdfs:subClassOf>`: ambos referem-se ao termo em particular (`subClassOf`).

Enquanto o URI *identifica* um recurso, a intenção do RDF (do inglês *Resource Description Framework* – Arcabouço para Descrição de Recursos) é *descrever* o recurso, ou seja, prover uma forma simples de declarar sentenças acerca de um recurso (W3C, 2004a). Por meio dessas sentenças, o RDF descreve e estrutura este recurso em termos de suas propriedades e das relações com outros recursos. Usando uma expressão comum à Biblioteconomia e à Computação, o RDF adiciona *metadados* ao recurso. Para isso, ele utiliza a metalinguagem XML (do inglês, *eXtensible Markup Language* – Linguagem de Marcação Extensível). A XML tornou-se a *língua franca* da Web, uma vez que a própria linguagem-base HTML é, agora, descrita em XML, e denominada XHTML. Essas linguagens serão descritas um pouco mais detalhadamente nas próximas seções.

### 3.2 Linguagens de Marcação: HTML e XML

No início da Web, a principal linguagem utilizada – HTML<sup>13</sup>, não distinguia a *representação* da informação da sua *apresentação*. Em uma mesma página, havia marcadores (*tags*) que dividiam o texto em seções, ou seja, proviam-na de estrutura, e marcadores que diziam como a informação seria apresentada. Um exemplo simples pode ser visto na figura 12.

Figura 12. Fragmento de conteúdo em HTML

```
...
<head>
...
</head>
<body>
    <b>O gato está no capacho.</b>
</body>
```

Fonte: do autor.

Os marcadores `<head>`, `</head>`, `<body>` e `</body>` estruturam o texto, separando qual conteúdo estava no cabeçalho (*head*) e qual no corpo (*body*). Já os marcadores `<b>` e `</b>`

---

<sup>13</sup> Hypertext Markup Language – linguagem de marcação de hipertextos.

apenas instruem o navegador a apresentar a informação ‘O gato está no capacho’ em negrito.

Além disso, em HTML, mesmo os marcadores ‘estruturais’ não dizem nada a respeito da semântica do conteúdo: este é expresso em linguagem natural. A sentença ‘O gato está no capacho’, por exemplo, contém uma relação binária entre as entidades ‘gato’ e ‘capacho’. Tal relação, no entanto, está *implícita*: é necessário que haja uma decodificação, antes que ela possa ser reutilizada.

O XML (do inglês *eXtensible Markup Language* – Linguagem de Marcação Extensível), ao contrário do HTML, não descreve como o conteúdo será visualizado nem ações que incidirão sobre elementos desse conteúdo; ele descreve o *conteúdo* do documento em si. O XML é, por isso, a linguagem base da camada esquema da Web Semântica. É importante ressaltar que o XML já é largamente utilizado também na Web ‘tradicional’.

Conforme Dias e Santos (2003), “o fator que realmente diferencia XML de outras linguagens [...] é sua capacidade extensiva, pois provém um formato de dados para estruturação de documentos sem utilização de um vocabulário específico” (p. 9). A figura 12 mostra um exemplo de documento com conteúdo expresso em XML.

Figura 12. XML – exemplo 1

```
<?xml version="1.0" encoding="UTF-8"?>
<enderecos xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">

  <endereço>
    <logradouro>Av. Antônio Carlos 6627 - Pampulha</logradouro>
    <cidade>Belo Horizonte</cidade>
    <UF>MG</UF>
  </endereço>

  <endereço>
    <logradouro>Palácio do Congresso Nacional - Zona Cívico-Administrativa
    </logradouro>
    <cidade>Brasília</cidade>
    <UF>DF</UF>
  </endereço>

</enderecos>
```

Fonte: adaptado pelo autor a partir de <http://www.xmlvalidation.com/example.html?&L=0>.

O conteúdo acima é facilmente legível por falantes da Língua Portuguesa. Contudo, em termos de *formalismo*, tanto do ponto de vista sintático quanto semântico, surgem alguns problemas. Por exemplo: não há nenhum mecanismo definindo que <UF> é um marcador que representa uma sigla composta de uma sequência de dois caracteres alfabéticos. Logo, é possível

que se introduza, nesse marcador, números, caracteres especiais etc. E o documento continuaria válido do ponto de vista da *estrutura*.

Ademais, surge um problema de interoperabilidade. Imaginemos que uma organização possua várias filiais, cada uma das quais coletando endereços de forma autônoma e gerando, a partir deles, documentos XML. Alguns podem ter a forma do exemplo da figura 12; outros, contudo, poderiam assumir outras formas e, ainda assim, expressar o mesmo conteúdo, como no exemplo da figura 13. Na figura em questão, foi acrescido o marcador `<bairro>`. Para um ser humano, a semântica dos documentos é a mesma, porque a informação de bairro estava implícita no marcador `<logradouro>`. No entanto, os dois documentos não seriam automaticamente compatíveis para um sistema de informação computacional.

Figura 13. XML – exemplo 2

```
<?xml version="1.0" encoding="UTF-8"?>
<enderecos xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">

  <endereco>
    <logradouro> Av. Antônio Carlos, 6627</logradouro>
    <bairro>Pampulha</bairro>
    <cidade>Belo Horizonte</cidade>
    <UF>MG</UF>
  </endereco>

  <endereco>
    <logradouro>Palácio do Congresso Nacional</logradouro>
    <bairro>Zona Cívico-Administrativa</bairro>
    <cidade>Brasília</cidade>
    <UF>DF</UF>
  </endereco>

</enderecos>
```

Fonte: adaptado pelo autor a partir de <http://www.xmlvalidation.com/example.html?&L=0>.

O exemplo deixa claro que é necessário agregar outras linguagens ao XML, de modo a enriquecer a informação sintática e semanticamente. O DTD e o XML Schema (esquema XML) são “mecanismos utilizados para especificar a estrutura de documentos escritos em linguagem XML” (DIAS e SANTOS, 2001). Ou seja, eles agregam mais uma camada de metadados, que define um conjunto de regras estruturais e sintáticas. A grande vantagem desses mecanismos é a possibilidade de validar se um documento XML está de acordo com regras de domínio, sintaxe, formato etc.

Dentre os dois mecanismos, o XML Schema é o mais sofisticado e possui vantagens sobre o

DTD, por isso foi o enfatizado neste trabalho. Dias e Santos (2001) destacam duas delas: (i) “a utilização de uma gramática mais rica e elaborada na prescrição da estrutura dos elementos, como por exemplo, especificação da quantidade exata de ocorrências possíveis de elementos filhos”; (ii) “a formatação para digitação de informação, ou seja, é possível estabelecer máscaras para determinado valor de atributo”. A figura 14 ilustra o uso do XML Schema; nesse exemplo, são definidas regras para validação do XML da figura 13. O XML Schema representa um avanço em relação ao XML, mas a utilização das outras linguagens da WS permite uma representação ainda mais precisa.

*Figura 14. XML Schema – exemplo*

```
<xs:schema xmlns:xs='http://www.w3.org/2001/XMLSchema'>
  <xs:element name="enderecos">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="endereco" minOccurs='1' maxOccurs='unbounded' />
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="endereco">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="logradouro" minOccurs='0' maxOccurs='1' />
        <xs:element ref="cidade" minOccurs='0' maxOccurs='1' />
        <xs:element ref="bairro" minOccurs='0' maxOccurs='1' />
        <xs:element ref="UF" minOccurs='0' maxOccurs='1' />
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="logradouro" type='xs:string' />
  <xs:element name="cidade" type='xs:string' />
  <xs:element name="bairro" type='xs:string' />
  <xs:element name="UF" type='xs:string' />
</xs:schema>
```

Fonte: adaptado pelo autor a partir de <http://www.xmlvalidation.com/example.html?&L=0>.

### 3.3 *Arcabouço para Descrição de Recursos: RDF*

O RDF é o acrônimo de *Resource Description Framework* ou, Arcabouço para Descrição de Recursos. Do ponto de vista semântico, o RDF é “uma linguagem baseada em assertivas, cujo objetivo é ser usada para expressar proposições usando vocabulários formais, em particular aqueles especificados pelo RDFS, para acesso e uso por toda a Web”; além disso, “pretende-se ainda que

ela seja usada como fundação básica para linguagens baseadas em assertivas mais avançadas e com o mesmo propósito” (W3C, 2004b, s/p). Portanto, o RDF é uma linguagem formal, na acepção dada no capítulo 1. De fato, RDF é definida como uma “lógica baseada em assertivas, na qual cada tripla expressa uma proposição simples” (s/p).

Ainda segundo o documento citado, a técnica usada para construir a semântica do RDF é a *teoria dos modelos* – também mencionada. A teoria dos modelos “assume que a linguagem refere-se a um ‘mundo’, e descreve as mínimas condições que um mundo precisa satisfazer para atribuir significado para cada expressão na linguagem”. Acrescenta-se ainda que um “mundo em particular é chamado *interpretação*”, logo a teoria dos modelos “poderia ser mais bem denominada ‘teoria da interpretação’” (W3C, 2004b, s/p).

Quando se trata de linguagem, faz-se necessário ‘anexar’ a semântica a uma sintaxe. A terminologia usada na especificação consiste em: referência a um URI, literal, literal ‘plano’, literal ‘tipado’, literal XML, valor XML, nó, nó vazio, tripla e grafo RDF (W3C, 2004b).

Um *grafo RDF* é um conjunto de triplas. Cada *tripla* representa uma assertiva e é constituída de sujeito, predicado e objeto. Os sujeitos e objetos são os *nós* do grafo. O *sujeito* é uma *referência a um URI*, que identifica o recurso descrito. O *objeto* pode ser um *literal*, como uma sequência de caracteres, um número ou uma data, ou ainda o URI de outro recurso que está relacionado ao sujeito. O *predicado* (ou *propriedade*) “permite descrever as características de um recurso, assim como as relações que guardam com outros recursos” (PEDRAZA-JIMENEZ *et al*, 2011, p. 23); ele também é associado a um URI. Os conceitos são ilustrados na figura 15.

O RDF “tem um modelo de dados fácil de ser processado e manipulado por aplicações” e “é independente de qualquer sintaxe de serialização específica” (W3C, 2004a, s/p). Sintaxes de serialização são as formas pelos quais o modelo de dados torna-se, efetivamente, um documento. As formas mais comuns são a RDF/XML e a Turtle; a primeira é a única com recomendação formal do W3C, mas a última também está em vias de ser recomendada.

Figura 15. Grafo orientado representando uma sentença RDF

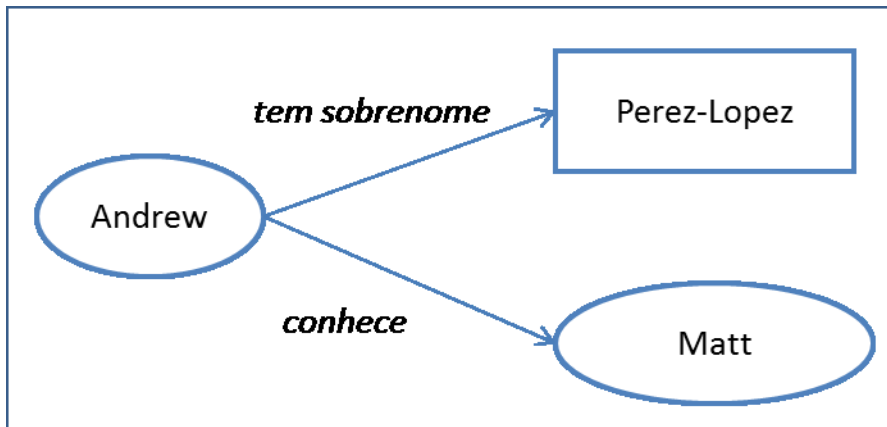


Fonte: adaptado de Hebelet et al (2009).

Para ilustrar, a figura 16 exhibe a representação simplificada de um RDF em forma de grafo, no qual há duas assertivas (sentenças lógicas). Optou-se por utilizar um tipo de notação frequente na

literatura da área, no qual um *recurso associado a um URI* é representado por uma *elipse* e um *literal*, por um *retângulo*.

Figura 16. Exemplos de assertivas em triplas (sujeito, propriedade, objeto)



Fonte: adaptado de Hebel et al (2009).

Sua tradução para RDF/XML é exibida na figura 17.

Figura 17. Tradução de assertivas lógicas para RDF/XML

```
<rdf:Description rdf:about="http://semwebprogramming.net/people#Andrew">
  <foaf:surname>Perez-Lopez</foaf:surname>
  <foaf:knows rdf:resource="http://semwebprogramming.net/people#Matt"/>
</rdf:Description>
```

Fonte: Hebel et al (2009).

Foge dos objetivos deste texto fornecer explicações detalhadas acerca do RDF. No entanto, é importante destacar certos aspectos que serão relevantes para o entendimento dos Dados Ligados. Um desses aspectos são os espaços de nomes ou *namespaces*, já mencionados anteriormente. No exemplo da figura 17, há dois *namespaces*, indicados pelos prefixos `rdf` e `foaf`. O primeiro é associado aos elementos que compõem o vocabulário da linguagem RDF<sup>14</sup>. O segundo se refere a uma ontologia para descrever pessoas e suas redes sociais de forma semântica, denominada FOAF (*Friend of a Friend*)<sup>15</sup>. Com base nisso, depreende-se que, por exemplo, o rótulo `foaf:surname` indica que o elemento seguinte refere-se a um *sobrenome* definido pela ontologia FOAF. Analogamente, `foaf:knows` indica que o elemento seguinte refere-se a uma outra pessoa (Matt), conhecida pela primeira (Andrew).

Outro aspecto relevante é que o RDF representa um grafo *orientado*: isso porque o *sentido* do predicado importa. Usando o exemplo do capítulo 1 (figura 11), a sentença

<sup>14</sup> A lista completa de elementos é descrita no documento RDF/XML Syntax Specification (W3C, 2004a).

<sup>15</sup> Descrição completa disponível em: <<http://semanticweb.org/wiki/FOAF>>.

`temFilho(Mauro, Pedro)`, a qual se pode ler “Mauro tem filho(a): Pedro”, contém o predicado `temFilho` (‘tem filho(a)’), cujo sentido é de *Mauro* para *Pedro*: Mauro é pai, e não Pedro.

No entanto, analisando essa expressão isoladamente, um aplicativo computacional não poderia decidir que Pedro *é filho(a) de* Mauro. Um leitor humano tem a capacidade de fazer tal inferência, uma vez que ele *sabe* que a relação ‘tem filho(a)’ possui uma relação inversa ‘é filho(a) de’. No entanto, uma máquina não possui tal capacidade *a priori*. Porém, é possível representar a segunda relação, por meio de outra sentença; ou, com a ajuda da linguagem OWL, é possível definir a relação inversa: no caso, “Y é filho(a) de X” é a inversa de “X tem filho(a): Y” – vide seção 2.5.

Outro tipo de relacionamento importante para a representação do conhecimento é a de hierarquia, onde uma categoria é uma *restrição* ou *especialização* de outra, como Pai e Homem, ou Mãe e Mulher. A Web Semântica também provê esse tipo de construção, de forma análoga a um tesouro – o qual utiliza os papéis *broader term* (BT) e *narrower term* (NT). Porém, para isso, é necessária a utilização da linguagem RDFS – *Resource Description Framework Schema*. Tanto a OWL quanto a RDFS serão vistas nas próximas seções.

### 3.4 Esquema para Descrição de Recursos: RDFS

Como foi dito, o RDFS estende o RDF, conferindo-lhe maior expressividade. O RDFS especifica como usar o RDF para descrever vocabulários, a partir de um conjunto predefinido de vocabulários (W3C, 2004c). Esse conjunto é formado por classes e propriedades. As classes são apresentadas sinteticamente na tabela 4.

Tabela 4. Classes RDFS

Classe	Descrição
<b>rdfs:Resource</b>	Todas as coisas descritas pelo RDF são chamadas recursos, e todas são instâncias da classe <code>rdfs:Resource</code> . Todas as classes são subclasses desta classe. <code>rdfs:Resource</code> é uma instância de <code>rdfs:Class</code> .
<b>rdfs:Class</b>	Esta é a classe de recursos que são classes RDF. <code>rdfs:Class</code> é uma instância de <code>rdfs:Class</code> .
<b>rdfs:Literal</b>	É a classe de valores literais, tais como <i>strings</i> e inteiros. Literais podem ser planos ou tipados. Um literal tipado é uma instância de uma classe de tipo de dados ( <i>datatype class</i> ). A especificação não define a classe dos literais planos.
<b>rdfs:Datatype</b>	É a classe dos tipos de dados. Todas as instâncias de <code>rdfs:Datatype</code> correspondem ao modelo RDF de um tipo de dados descrito na especificação dos conceitos RDF (W3C, 2004a).
<b>rdf:XMLLiteral</b>	É a classe dos valores literais XML. <code>rdf:XMLLiteral</code> é uma instância de <code>rdfs:Datatype</code> e uma subclasse de <code>rdfs:Literal</code> .
<b>rdf:Property</b>	É a classe das propriedades RDF. <code>rdf:Property</code> é uma instância de <code>rdfs:Class</code> .

Fonte: W3C (2004c).

Com relação às propriedades, nesta seção serão destacadas três: `rdfs:subClassOf`,

`rdfs:domain` e `rdfs:range`, as quais permitem um entendimento básico dos documentos da Web Semântica. Para as demais propriedades e explicações detalhadas, recomenda-se a leitura de Hebler *et al.* (2009), Allemang e Hendler (2011) e do próprio documento de especificação do W3C (2004c). Além disso, algumas propriedades da linguagem OWL serão utilizadas nos exemplos desta seção, com o intuito de torná-los mais significativos.

Conforme explicam Hebler *et al.* (2009), “uma das formas mais simples de restringir o pertencimento a uma classe é criar um relacionamento taxonômico entre ela e outras classes por meio da propriedade `rdfs:subClassOf`” (p. 111). Ou seja, a assertiva `Classe1 rdfs:subClassOf Classe2` implica:

- `Classe1` é uma especialização de `Classe2`.
- Cada membro de `Classe1` necessita ser um membro válido de `Classe2` (ou seja, precisa satisfazer todas as restrições de pertencimento especificadas por `Classe2`).
- As propriedades de `Classe2`, incluindo as restrições de pertencimento, são subsumidas (herdadas) por `Classe1`.

Utilizando novamente a ontologia ‘Família’, as assertivas (a) “Todo aquele que é avô também é pai” e (b) “Toda aquela que é avó também é mãe” serão expressas em RDFS/OWL, conforme a figura 18. Ressalta-se que as categorias ‘Avô’ e ‘Avó’ são representadas pelas classes OWL ‘Avo’ e ‘Avoh’ (`owl:class`); o prefixo ‘f:’ representa a ontologia ‘Família’.

*Figura 18. Fragmento de um documento RDFS/OWL com propriedade `rdfs:subClassOf`*

f:Avo	<code>rdf:type owl:Class;</code> <code>rdfs:subClassOf Pai .</code>
f:Avoh	<code>rdf:type owl:Class;</code> <code>rdfs:subClassOf Mae .</code>

Fonte: do autor.

Nota: alguns termos estão sem sinais de acentuação propositalmente, para simplificar a codificação do documento.

É possível também descrever o domínio e o escopo das relações entre as propriedades e classes ou tipos de dados por meio das propriedades `rdfs:domain` e `rdfs:range` (HEBELER *et al.*, 2009, p. 114-115):

- `rdfs:domain` – especifica o tipo de todos os indivíduos que são *sujeitos* das assertivas que usam a propriedade descrita;
- `rdfs:range` – especifica o tipo de todos os indivíduos ou tipos de dados de todos os literais que são *objetos* das assertivas que usam a propriedade descrita.



Lançando mão mais uma vez da ontologia ‘Família’, a figura 19 exibe uma especificação da propriedade “tem filho(a)” (temFilho). No caso, tanto o sujeito (pai ou mãe) quanto o objeto (filho(a)) são, necessariamente, pessoas – ou seja, pertencem à classe Pessoa.

*Figura 19. Fragmento de um documento RDFS/OWL com propriedades rdfs:domain e rdfs:range*

```
f:temFilho rdf:type owl:ObjectProperty ;
    rdfs:domain f:Pessoa ;
    rdfs:range f:Pessoa .
```

Fonte: do autor.

O RDFS, juntamente com a OWL, permitem ainda expressar taxonomias entre propriedades (rdfs:subPropertyOf), propriedades inversas (owl:inverseOf), propriedades disjuntas (owl:propertyDisjointWith), entre outras, ampliando a expressividade e a precisão das representações.

### 3.5 *Linguagem para Descrição de Ontologias na Web: OWL*

Conforme visto no capítulo 1, a OWL é a linguagem utilizada pela Web Semântica para expressar ‘conhecimento rico’ sobre coisas, conjuntos de coisas ou relações entre essas coisas. Em outras palavras, é a linguagem para descrever ontologias, na acepção computacional do termo.

A OWL é derivada da linguagem DAML+OIL (W3C, 2002). Esta linguagem, por sua vez, foi projetada para ter um mapeamento com a Lógica Descritiva (BAADER *et al.*, 2003). A versão atual da OWL, conhecida como OWL 2, foi publicada em 2009, e teve um segundo lançamento em 2012 (W3C, 2012).

A OWL possui três variações: OWL Lite, OWL DL e OWL Full. A OWL Full, como o nome indica, é a especificação completa e irrestrita da OWL. A OWL DL introduziu algumas restrições ao uso da OWL Full, incluindo a separação de classes e indivíduos (ou seja, de conhecimento intensional e extensional). Finalmente, a OWL Lite aumentou ainda mais as restrições, no intuito de torna-la mais prática para os programadores. No entanto, como destacam Allemang e Hendler (2011), os benefícios em termos de praticidade foram inferiores à supressão dos recursos que lhe conferiam maior expressividade.

Com isso, a OWL DL tornou-se o mais bem sucedido ‘dialeto’ da OWL. A expressão DL vem, justamente, da Lógica Descritiva, a qual, como se viu, é um importante subconjunto da Lógica de Primeira Ordem. Daí a importância do estudo dessas lógicas para uma compreensão mais profunda da Web Semântica.

Como toda linguagem, a OWL possui sintaxe e semântica. Existem duas formas alternativas

de prover significado (ou seja, semântica) para as ontologias OWL2: a Semântica Direta (OWL 2 Direct Semantics) e Semântica Baseada em RDF (OWL 2 RDF-Based Semantics). Há um teorema de correspondência que provê uma associação entre as duas formas (W3C, 2012b).

Conforme o W3C (2012a), “na prática, uma sintaxe concreta é necessária para armazenar as ontologias OWL2 e fazer seu intercâmbio entre ferramentas e aplicações” (s/p). A opção primária é o padrão RDF/XML. No entanto, há alternativas, como Turtle, OWL2 XML e OWL2 Manchester Syntax. Existe ainda uma sintaxe funcional, que permite ver com mais clareza os aspectos formais da linguagem (W3C, 2012a).

A OWL, assim como a DL, possui construtores e axiomas. Um subconjunto dos construtores e axiomas mais usados para a descrição de ontologias em OWL é exibido na tabela 5, utilizando-se notações diferentes; na última coluna, são mostrados alguns exemplos informais.

Tabela 5. Equivalência entre algumas expressões OWL e DL

Sintaxe Funcional	Sintaxe RDF	Sintaxe DL	Exemplo
ObjectIntersectionOf( $CE_1 \dots CE_n$ )	<code>_:x rdf:type owl:Class .</code> <code>_:x owl:intersectionOf T(SEQ CE<sub>1</sub> ... CE<sub>n</sub>) .</code>	$CE_1 \wedge \dots \wedge CE_n$	Humano $\wedge$ Masculino
ObjectUnionOf( $CE_1 \dots CE_n$ )	<code>_:x rdf:type owl:Class .</code> <code>_:x owl:unionOf T(SEQ CE<sub>1</sub> ... CE<sub>n</sub>) .</code>	$CE_1 \vee \dots \vee CE_n$	Advogado $\vee$ Médico
ObjectComplementOf( $CE$ )	<code>_:x rdf:type owl:Class .</code> <code>_:x owl:complementOf T(CE) .</code>	$\neg CE$	$\neg$ Masculino
ObjectMinCardinality( $n$ OPE $CE$ )	<code>_:x rdf:type owl:Restriction .</code> <code>_:x owl:onProperty T(OPE) .</code> <code>_:x owl:minQualifiedCardinality "n"^^xsd:nonNegativeInteger .</code> <code>_:x owl:onClass T(CE) .</code>	$(\geq n \text{ OPE}.CE)$	$(\geq 2 \text{ temFilhos}.Advogado)$
ObjectInverseOf( $OP$ )	<code>_:x owl:inverseOf T(OP) .</code>	$OP^{-}$	$\text{temFilhos}^{-}$
SubClassOf( $CE_1$ $CE_2$ )	<code>T(CE<sub>1</sub>) rdfs:subClassOf T(CE<sub>2</sub>) .</code>	$CE_1 \subseteq CE_2$	Humano $\subseteq$ Animal $\wedge$ Bípede
EquivalentClasses( $CE_1 \dots CE_n$ )	<code>T(CE<sub>1</sub>) owl:equivalentClass T(CE<sub>2</sub>) .</code> ... <code>T(CE<sub>n-1</sub>) owl:equivalentClass T(CE<sub>n</sub>) .</code>	$CE_1 \equiv CE_2$ ... $CE_{n-1} \equiv CE_n$	Homem $\equiv$ Humano $\wedge$ Masculino
SubObjectPropertyOf( $OPE_1$ $OPE_2$ )	<code>T(OPE<sub>1</sub>) rdfs:subPropertyOf T(OPE<sub>2</sub>) .</code>	$OPE_1 \subseteq OPE_2$	$\text{temFilha} \subseteq \text{temFilhos}$
EquivalentObjectProperties( $OPE_1 \dots OPE_n$ )	<code>T(OPE<sub>1</sub>) owl:equivalentProperty T(OPE<sub>2</sub>) .</code> ... <code>T(OPE<sub>n-1</sub>) owl:equivalentProperty T(OPE<sub>n</sub>) .</code>	$OPE_1 \equiv OPE_2$ ... $OPE_{n-1} \equiv OPE_n$	valor $\equiv$ preço
DisjointObjectProperties( $OPE_1$ $OPE_2$ )	<code>T(OPE<sub>1</sub>) owl:propertyDisjointWith T(OPE<sub>2</sub>) .</code>	$OPE_1 \vee \neg OPE_2$	Masculino $\vee \neg$ Feminino
SameIndividual( $a_1 \dots a_n$ )	<code>T(a<sub>1</sub>) owl:sameAs T(a<sub>2</sub>) .</code> ... <code>T(a<sub>n-1</sub>) owl:sameAs T(a<sub>n</sub>) .</code>	$\{a_1\} \equiv \{a_2\}$ ... $\{a_{n-1}\} \equiv \{a_n\}$	$\{JK\} \equiv \{\text{Juscelino Kubistchek}\}$

Fonte: adaptado de W3C (2012c) e BAADER *et al* (2003).

Para facilitar o entendimento, será dado um exemplo, utilizando novamente a *ontologia* ‘família’, por meio de uma versão em OWL/RDF, conforme figura 18.

Figura 18. Tradução de assertivas lógicas para OWL (a)

```
<owl:ObjectProperty rdf:about="#temFilhoH">
  <rdfs:range rdf:resource="#Homem"/>
  <rdfs:domain rdf:resource="#Pessoa"/>
  <rdfs:subPropertyOf rdf:resource="#temFilho"/>
  <owl:inverseOf rdf:resource="#ehFilhoDe"/>
</owl:ObjectProperty>
```

Fonte: adaptado de <http://www.cs.man.ac.uk/~stevensr/ontology/family.rdf.owl>.

O poder expressivo da OWL permitiu, nesse caso, adicionar um *axioma* relacionando uma propriedade à sua inversa, por meio do construtor `owl:inverseOf`, tornando a ontologia mais completa e precisa. No exemplo, é definida a propriedade `temFilhoH` (no sentido de filho homem) como uma ‘subpropriedade’ de `temFilho` (no sentido geral). Lembrando que a OWL é baseada na lógica formal, um mecanismo de inferência agora poderia deduzir a assertiva `ehFilhoDe(Pedro, Mauro)`, ou seja, “Pedro é filho de Mauro”, a partir da assertiva `temFilhoH(Mauro, Pedro)`, isto é: “Mauro tem filho: Pedro”. A figura 19 apresenta um exemplo de uma descrição RDF da última assertiva. Nesse exemplo, a expressão “`#mauro_1946`” é um URI que identifica Mauro, “`#pedro_1972`” é uma URI que identifica Pedro e `temFilhoH` é um URI que faz referência à propriedade (predicado) `temFilho`.

Figura 19. Tradução de assertivas lógicas para OWL (b)

```
<owl:NamedIndividual rdf:about="#mauro_1946">
  <temAnoNascimento rdf:datatype="xml:Integer">1946</ temAnoNascimento>
  <temFilhoH rdf:resource="#pedro_1972"/>
</owl:NamedIndividual>
```

Fonte: adaptado de <http://www.cs.man.ac.uk/~stevensr/ontology/family.rdf.owl>.

Portanto, as descrições RDF associadas aos esquemas RDFS e aos axiomas expressos em OWL possibilitam a ação de mecanismos de inferência, os quais podem deduzir novas assertivas a partir das existentes: ou seja, tornam explícitas associações outrora implícitas na base de conhecimento. A linguagem SPARQL, descrita a seguir, é a forma mais usual de interação com os mecanismos de inferência.

Importante ainda ressaltar que a combinação dos mecanismos do RDFS e da OWL confere às ontologias da Web Semântica um grau de expressividade maior que alguns recursos tradicionais de representação da informação e do conhecimento, como os tesauros e as taxonomias. Conforme Bräscher (2008, p. 44), a ontologia “agrega valor aos demais esquemas de representação, por meio

de uma semântica mais profunda e também sob os aspectos conceitual, relacional e tecnológico”.

### 3.6 *Linguagem para Consulta de Grafos RDF: SPARQL*

O RDF provê uma forma simples de representar dados distribuídos: no entanto, uma representação de dados é inútil sem os meios adequados para o acesso aos dados (ALLEMANG e HENDLER, 2011). A linguagem SPARQL foi desenvolvida com o intuito de preencher essa lacuna. SPARQL é a sigla para *SPARQL Protocol And RDF Query Language* (Protocolo e Linguagem de Consulta RDF).

Como o próprio nome indica, a linguagem SPARQL é feita para dados em formato RDF, e é baseada em padrões de grafos e correspondência (*matching*) entre subgrafos. O componente básico dos padrões de consulta mais complexos em SPARQL é baseado em uma estrutura denominada padrão de grafo básico – BGP (*basic graph pattern*). Um BGP é um conjunto de triplas que são, na verdade, triplas RDF, as quais podem conter variáveis (parâmetros) de consulta na posição do sujeito, do predicado ou do objeto. Durante a execução da consulta, soluções que associam valores às variáveis são determinadas (HARTIG, BIZER e FREITAG, 2009).

Os padrões de linguagem SPARQL são representados numa variante da linguagem Turtle (ALLEMANG e HENDLER, 2009). As duas principais palavras-chave, ou cláusulas, são *SELECT* e *WHERE*. A cláusula *SELECT* indica *quais* elementos serão selecionados. Esses elementos podem ser *sujeitos*, *propriedades* ou *objetos*. Já a cláusula *WHERE* indica as *condições* ou *padrões* de seleção dos elementos. O mecanismo de consulta fará a recuperação das triplas RDF que contenham os elementos selecionados e que satisfaçam as condições estabelecidas.

Para facilitar a compreensão, alguns exemplos de consultas expressas em SPARQL serão exibidos a seguir, extraídos do tutorial produzido por Feigenbaum e Prud'hommeaux (2011), cujos autores são associados ao grupo Cambridge Semantics e ao W3C. Primeiramente, é necessário definir um conjunto de dados (*dataset*), sobre o qual será realizada a consulta. Um extrato desse *dataset* é exibido na figura 20.

Figura 20. Trecho de um documento usando a ontologia FOAF

```
@prefix card: <http://www.w3.org/People/Berners-Lee/card#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

card:i foaf:name "Timothy Berners-Lee" .
<http://bblfish.net/people/henry/card#me>foaf:name "Henry Story" .
<http://www.cambridgesemantics.com/people/about/lee>foaf:name "Lee Feigenbaum" .
card:amy foaf:name "Amyvan der Hiel" .

...
```

Fonte: FEIGENBAUM, Lee; PRUD'HOMMEAUX, Eric. SPARQL By Example: a tutorial. Disponível em: <<http://www.cambridgesemantics.com/pt/semantic-university/sparql-by-example>>. Acesso em 12 mar. 2013.

As expressões `card` e `foaf` são prefixos. A função de um *prefixo* é tornar o documento mais enxuto e legível, uma vez que não é necessário repetir a URI: o prefixo a substitui. No caso do exemplo, `card` substitui o URI da ‘agenda de nomes’ da pessoa “Timothy Berners-Lee”. Existem outras pessoas representadas no trecho destacado, como “Henry Story”, “Lee Feigenbaum” e “Amyvan der Hiel”. Interessante observa que cada um é originário de um documento diferente, mas compartilham a mesma *estrutura semântica* de metadados, definida pela ontologia FOAF.

Supondo que o trecho acima estivesse em um documento armazenado no URI <http://dig.csail.mit.edu/2008/webdav/timbl/foaf.rdf>, poder-se-ia formular a seguinte consulta (FEIGENBAUM e PRUD'HOMMEAUX, 2011):

No *grafo* *localizado* em <http://dig.csail.mit.edu/2008/webdav/timbl/foaf.rdf>, *encontre-me todos os sujeitos* (*?person*) *e objetos* (*?name*) *associados ao predicado* *foaf:name*. *Retorne então todos os valores de ?name*. *Sinteticamente, encontre todos os nomes mencionados no arquivo FOAF de Tim Berners-Lee's*.

A consulta equivalente em SPARQL é exibida na figura 21.

Figura 21. Consulta simples em SPARQL

```
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
SELECT ?name
WHERE {
    ?person foaf:name ?name .
}
```

Fonte: FEIGENBAUM, Lee; PRUD'HOMMEAUX, Eric. SPARQL By Example: a tutorial. Disponível em: <<http://www.cambridgesemantics.com/pt/semantic-university/sparql-by-example>>. Acesso em 12 mar. 2013.

Conforme o próprio tutorial explica (p. 8):

- As *variáveis* SPARQL iniciam com um ponto de interrogação (?) e podem corresponder a qualquer nó (recurso ou literal) no conjunto de dados expresso em

RDF – ou seja, no grafo RDF.

- Os *padrões de triplas* (*triple patterns*) são semelhantes às triplas, exceto pelo fato de seus componentes poderem ser substituídos por uma variável.
- A cláusula SELECT retorna uma tabela de variáveis e valores que satisfazem os critérios estabelecidos pela cláusula WHERE; ou seja, que correspondem aos padrões de triplas.
- O conjunto de dados utilizado foi:  
<http://dig.csail.mit.edu/2008/webdav/timbl/foaf.rdf>.

Cumpramos ressaltar que o fato de a semântica do modelo de dados do RDF ser especificada por um grafo, permite que as sentenças (triplas) sejam escritas em qualquer ordem dentro do documento. E isso implica, em grande parte, que um padrão de grafo (padrão de tripla) escrito em uma ordem diferente, retornará o mesmo resultado (ALLEMANG e HENDLER (2011)).

A linguagem SPARQL possui muitos outros recursos, os quais podem ser estudados com maior profundidade no tutorial citado, bem como em Allemang e Hendler (2011).

### 3.7 *Resumo do capítulo*

Conforme foi discutido ao longo dos capítulos 1 e 2, uma das características distintivas da Web Semântica em relação à Web tradicional é a representação do conhecimento apoiada em modelos e linguagens formais. Quando se utiliza apenas linguagens naturais, as relações e inferências não estão totalmente explícitas e padronizadas, o que as impede de serem processados computacionalmente de forma eficiente. Ou seja, boa parte do conteúdo hoje existente na Web ‘tradicional’ depende da interpretação humana, mesmo quando seria interessante que fossem processáveis automaticamente.

E esse é justamente o caso dos dados governamentais que, salvo raras exceções previstas legalmente, são públicos. No próximo capítulo será visto como a utilização dos princípios e tecnologias da Web Semântica possibilitam que esses dados tornem-se efetivamente ‘abertos’, ou seja, amplamente disponíveis e inteligíveis, inclusive por sistemas computacionais, os quais podem auxiliar na transformação desses dados em informação útil à sociedade e ao próprio governo.

## 4. O PADRÃO ‘DADOS ABERTOS LIGADOS’ NA REPRESENTAÇÃO E DISSEMINAÇÃO DA INFORMAÇÃO ORÇAMENTÁRIA

O objetivo deste capítulo é apresentar uma análise da necessidade e da viabilidade da aplicação dos conceitos e tecnologias da Web Semântica e, em especial, do padrão *dados abertos ligados*, na representação e disseminação das informações sobre o Orçamento Público pelo Congresso Nacional.

Na primeira seção, o funcionamento do Orçamento será explicado de forma sucinta. Posteriormente, na segunda seção, os sistemas de informação orçamentária baseados na Web do Congresso Nacional serão analisados frente à Lei de Acesso à Informação (LAI). Na terceira seção serão apresentadas algumas definições e tecnologias associadas ao padrão Dados Abertos Ligados. Finalmente, na quarta seção, serão discutidos e ilustrados alguns aspectos relativos à aplicabilidade do padrão dados abertos ligados pelo Congresso Nacional no contexto da informação orçamentária.

### 4.1 *Síntese do Processo Orçamentário*

O processo orçamentário compreende as fases de elaboração e execução das leis orçamentárias – PPA (Plano Plurianual), LDO (Lei de Diretrizes Orçamentárias) e LOA (Lei Orçamentária Anual) (BRASIL, 2013). O PPA “estabelece as diretrizes, os objetivos e as metas da Administração Pública, para as despesas de capital e outras delas decorrentes, e para as relativas aos programas de duração continuada” (BEZERRA FILHO, 2006, p. 22). Já a LDO “estabelece metas e prioridades da administração pública para o exercício seguinte” (p. 22). Finalmente a LOA estabelece a peça orçamentária ou orçamento propriamente dito, composto de Orçamento Fiscal, Orçamento de Investimentos das Estatais e Orçamento da Seguridade Social. Como o foco do estudo de caso será o acompanhamento da execução orçamentária da despesa, ela será explicada com mais detalhes a seguir. Algumas expressões-chave foram destacadas em *itálico* e serão retomadas nas próximas seções.

A execução da *despesa* possui duas etapas: execução orçamentária e execução financeira. A execução orçamentária consiste em utilizar os créditos fixados no Orçamento mediante o *empenho* da despesa para que as *ações* sejam efetivadas.

Já a execução financeira corresponde à etapa final da realização da despesa, e se subdivide em liquidação e pagamento. *Liquidação* é a “verificação da certeza de liquidez do credor, ou seja, se este cumpriu integralmente com o implemento da condição ou do objeto do contrato”. A liquidação da despesa “gera uma *ordem de pagamento*, que determina que a despesa será paga ao



*credor*”. Esta ordem de pagamento é associada a um documento chamado *Nota de Liquidação*. O *pagamento* é a “extinção da obrigação financeira da entidade governamental, mediante a entrega de dinheiro correspondente ao valor devido ao credor”, o qual “só será efetuado após a liquidação da mesma” (NASCIMENTO, 2001, p. 39-40).

## 4.2 *Análise dos Sistemas de Informação Orçamentária do Congresso Nacional frente à LAI*

A Câmara dos Deputados e o Senado, a partir do acesso às informações provenientes da base de dados do Sistema Integrado de Administração Financeira (SIAFI) e de suas próprias bases de dados, construíram *datawarehouses* (armazéns de dados), os quais vêm permitindo a extração de informações 'gerenciais' ou 'analíticas'. Ou seja, as Casas Legislativas passaram a ter ferramentas que possibilitam o cruzamento dos dados orçamentários e sua análise sob diversas dimensões e níveis de agregação. Isso permitiu, por exemplo, verificar qual o montante que determinado programa orçamentário recebeu em determinado período de tempo (CHEVITARESE ALVES, 2009).

Ainda segundo o autor, com base nesses armazéns de dados, foram construídos portais eletrônicos de orçamento por ambas as Casas do Poder Legislativo Federal: Portal do Orçamento – Senado (implantado em 2005); Portal Orçamento Brasil – Câmara dos Deputados (implantado em 2007). No contexto desses portais, foram criadas aplicações computacionais para a divulgação de informações orçamentárias a partir desses dados: o Sistema Fiscalize<sup>16</sup> e o Sistema Siga Brasil<sup>17</sup>, na Câmara e no Senado, respectivamente.

Ambos os sistemas serão avaliados em termos da conformidade com o art. 8º da Lei de Acesso à Informação (LAI), a fim de verificar o pressuposto (P1):

*Os sistemas Web de informações orçamentárias do Congresso Nacional disponíveis atualmente não atendem plenamente os requisitos estabelecidos pela Lei de Acesso à Informação (LAI).*

Como foco de análise, foram escolhidas as informações sobre repasses de recursos financeiros da União para os municípios. Esses repasses podem ser efetivados de duas formas distintas: por meio das transferências obrigatórias ou voluntárias. As transferências obrigatórias podem ser constitucionais ou legais. Já as transferências voluntárias compreendem os repasses de

---

<sup>16</sup> <<http://www2.camara.leg.br/atividade-legislativa/orcamentobrasil/fiscalize/transferencia>>.

<sup>17</sup> <[http://www9.senado.gov.br/portal/page/portal/orcamento\\_senado/SigaBrasil](http://www9.senado.gov.br/portal/page/portal/orcamento_senado/SigaBrasil)>. Neste trabalho, foi utilizada a opção 'Acesso Livre'.

recursos aos municípios para a execução de programas e ações governamentais não decorrentes de determinação constitucional, legal ou os destinados ao Sistema Único de Saúde. Os instrumentos utilizados nas transferências voluntárias para os municípios são o convênio e o contrato de repasse. Um convênio é uma parceria formalizada entre a União e o município para a execução de programa de governo. Já o contrato de repasse é o instrumento por meio do qual a transferência voluntária dos recursos financeiros é realizada por instituição financeira pública federal<sup>18</sup> (BRASIL, 2013).

Optou-se por esse recorte devido: (a) ao interesse dos parlamentares no acompanhamento de emendas a programas de governo associados às transferências voluntárias; (b) ao interesse dos próprios cidadãos das localidades convenientes, uma vez que, em tese, eles seriam beneficiários diretos; (c) ao fato de as informações sobre transferências estarem, hoje, dispersas em várias bases de dados, situação na qual a aplicação de dados abertos ligados é recomendável.

Para ambos será adotada a mesma sistemática: os sistemas serão brevemente descritos e, para cada detalhe descrito, eventualmente será feita uma observação crítica, destacada com sublinhado. Essas observações serão posteriormente classificadas como limitações ou não-conformidades em relação a alguns dos dispositivos do art. 8º, §3º, da Lei de Acesso da Informação (LAI). No último caso, haverá uma referência ao dispositivo.

É importante mencionar que tais sistemas foram concebidos antes da promulgação da LAI. Portanto, o objetivo principal da avaliação foi mostrar a necessidade de se desenvolver estratégias complementares para o cumprimento pleno da LAI, uma vez que os sistemas avaliados, isoladamente, são insuficientes, conforme será visto. E, dentre as estratégias complementares possíveis, defende-se aqui a utilização de uma plataforma baseada no padrão LOD (*Dados Abertos Ligados*).

#### **4.2.1 Sistema Fiscalize**

O Sistema Fiscalize<sup>19</sup> extrai informação do armazém de dados orçamentários da Câmara dos Deputados (DW Orçamento), e, por meio da tecnologia Oracle Reports, gera relatórios em formato estático (PDF). Sua atualização é diária, o que implica que ele atende ao art. 8º, §3º, inc. VI, da LAI; suas fontes são o Sistema Integrado de Administração Financeira (SIAFI operacional) e sua variante SIAFI Gerencial, o que, em tese, garante a autenticidade e a integridade das informações

---

<sup>18</sup> Atualmente são: Banco do Brasil, Banco do Nordeste, Banco da Amazônia e Caixa Econômica Federal.

<sup>19</sup> Informações obtidas verbalmente junto à Seção de Soluções de Inteligência de Negócios (SESIN) do Centro de Informática da Câmara dos Deputados.

disponíveis para acesso – LAI, art. 8º, §3º, inc. V.

O sistema possui quatro módulos: Transferências a Estados, Transferências a Municípios, Transferências a Entidades Privadas e Transferências por Programação. Todos utilizam mecanismos de funcionamento análogos, baseado nas seguintes etapas: o usuário seleciona o tipo de relatório; o usuário seleciona período de tempo coberto pelo relatório: o mês e o ano; o usuário seleciona critérios de busca em lista de opções; o sistema devolve um relatório em formato PDF.

Para a avaliação será descrito o módulo Transferências a Municípios, com foco nas informações sobre execução de convênios, conforme mencionado. Como os demais módulos têm estrutura e funcionamento análogos, a avaliação de um aplica-se aos demais. O funcionamento do módulo se dá pelos seguintes passos:

1. O usuário seleciona o tipo de relatório: valores pagos ou valores empenhados/conveniados. Não é possível visualizar ambos os valores em um mesmo relatório.
2. O usuário seleciona período de tempo coberto pelo relatório: o mês e o ano. Não é possível selecionar um intervalo de tempo envolvendo dois ou mais meses.
3. O usuário seleciona o município destinatário do recurso. Não é possível selecionar mais de um município simultaneamente.
4. O usuário solicita a geração do relatório. O relatório é gerado apenas no formato PDF.

Um exemplo de uso pode ser visto na figura 22. No caso, foi escolhido o relatório de valores empenhados/conveniados, com os seguintes critérios: (a) mês: ‘junho/2010’; (b) UF: ‘Minas Gerais’; (c) município: ‘Bonfim’. A figura 23 exhibe um detalhe do relatório gerado.

Figura 22. Sistema Fiscalize – Transferências por Programação

Fonte: Portal Orçamento Brasil – Sistema Fiscalize

Figura 23. Sistema Fiscalize – Transferências por Programação - relatório

CÂMARA DOS DEPUTADOS

Orçamento da União

Execução Orçamentária

Orçamento Fiscal e Seguridade Social

Orçamento Brasil

Fiscalize

Recursos do Orçamento da União empenhados aos Municípios

Data Posição SIAFI: 28/03/2013

Data Emissão: 02/04/2013

Página:1 / 1

UF: MG Município: BONFIM

Favorecido: CNPJ 183639450/0013-3 - PREFEITURA MUNICIPAL DE BONFIM

Unidade Orçamentária / Projeto-Atividade / Convênio(*2)-Empenho	Valores Empenhados (*1) - Em R\$1,00	
	Jun/2010	Jan a Jun/2010
54.101 - MINISTERIO DO TURISMO		
4620.0031 PROMOCAO DE EVENTOS PARA DIVULGACAO DO TURISMO INTERNO	100.000,00	100.000,00
CONVÊNIO: 738757 - EMPENHO: 901103 - FESTA DO PEÃO BOIADEIRO - 738757/2010, CV 00913/2010, PROPOSTA 027065/2010, PROCESSO 72031.010308/2010-97.		
56.101 - MINISTERIO DAS CIDADES		
1D73.0204 APOIO A POLITICA NACIONAL DE DESENVOLVIMENTO URBANO	146.950,00	146.950,00
CONVÊNIO: 738942 - EMPENHO: 901111 - PAVIMENTAÇÃO ASFÁLTICA EM RUAS DO PERÍMETRO URBANO DO MUNICÍPIO, CITA - SE AS REFERIDAS RUAS: JOÃO DA COSTA NEVES, VICENTE S. MOREIRA, GERALDO PEREIRA BORGES, JOÃO BATISTA PARREIRAS, VICENTE C. BICALHO DO BAIRRO SAGRADO CORAÇÃO DE JES		
Total Favorecido:	246.950,00	246.950,00
Total do Município:	246.950,00	246.950,00

Fonte: Portal Orçamento Brasil – Sistema Fiscalize

Prosseguindo a análise, tem-se que o relatório é gerado apenas no formato PDF: trata-se de uma não-conformidade e, ao mesmo tempo, de uma limitação. O fato de o Sistema Fiscalize não possibilitar a geração dos dados em formatos abertos simples, como o CSV, o qual pode ser lido pelos principais editores de planilha eletrônica disponíveis, é uma não-conformidade relativa a dois dispositivos da LAI: art. 8º, §3º, inc. II: “possibilitar a gravação de relatórios em diversos formatos

*eletrônicos, inclusive abertos e não proprietários, tais como planilhas e texto, de modo a facilitar a análise das informações”; art. 8º, §3º, inc. III: “possibilitar o acesso automatizado por sistemas externos em formatos abertos, estruturados e legíveis por máquina” (BRASIL, 2012).* Ao mesmo tempo, trata-se de uma limitação, já que não está disponível em formatos abertos e ligados, como o RDF. Em tese, caso os metadados necessários estivessem disponíveis, seria possível a transformação dos dados para o RDF. Ocorre que, embora o PDF seja um formato aberto, o processo de extração dos dados e eventuais metadados existentes exigiria um esforço considerável de ‘raspagem de dados’ – conforme será visto no caso “Ligado nos Políticos” (Seção 3.3.2).

As demais observações são no sentido das limitações do sistema em termos de flexibilidade na combinação e visualização dos dados:

- não é possível visualizar ambos os valores em um mesmo relatório;
- não é possível selecionar um intervalo de tempo envolvendo dois ou mais meses;
- não é possível selecionar mais de um estado simultaneamente;
- não é possível selecionar mais de um estado município simultaneamente.

Apenas para ilustrar essas limitações, não seria possível fazer análises comparativas do tipo ‘comparar os valores empenhados e conveniados transferidos da União para municípios com a mesma população, dentro de um mesmo estado’. Naturalmente, seria possível obter esta informação a partir da junção das informações obtidas de vários relatórios separados e de consultas a bases oficiais, como o IBGE; no entanto, além de ser trabalhoso, a princípio, exigiria um agente humano. Caso a informação estivesse disponível em RDF, no formato de dados abertos e ligados, ela poderia ser recuperada por um agente computacional, ou seja, por um programa de computador. Nesse caso, poderia haver um vocabulário compartilhado sobre municípios, por meio de uma ontologia, entre dois conjuntos de dados: por exemplo, um oriundo do armazém de dados da Câmara dos Deputados e outro, do IBGE.

Além disso, a linguagem utilizada no Fiscalize não pode ser considerada de fácil compreensão para um usuário da informação leigo, pois contém siglas sem tradução (como FNDE e LOA, por exemplo) e muitos termos técnicos (vide figuras 22 e 23), como ‘Unidade Orçamentária’, ‘Projeto-Atividade’ e ‘Empenho’. É fato que no Portal Orçamento Brasil existe uma página denominada ‘Entenda o Orçamento’<sup>20</sup>. No entanto, a página em si não apresenta informações explicativas sobre o orçamento, mas contém *hyperlinks* para documentos, os quais inclui uma

---

<sup>20</sup> Vide <<http://www2.camara.leg.br/atividade-legislativa/orcamentobrasil/entenda>> .

cartilha. No entanto, mesmo essa cartilha está em formato PDF, ou seja, não está em um formato aberto e legível por máquina. O problema da falta de clareza da linguagem é uma não-conformidade com o disposto na LAI, art. 8º, §3º, inc. I: “Conter ferramenta de pesquisa de conteúdo que permita o acesso à informação de forma objetiva, transparente, clara e em linguagem de fácil compreensão”.

Por último, em relação à LAI, o Sistema Fiscalize não divulga como a informação é estruturada; ou seja, não são explicitados os metadados que mostram como as categorias e os itens de informação se relacionam. Portanto, a falta de divulgação dos metadados é uma não-conformidade com o disposto na LAI, art. 8º, §3º, inc. IV: “Divulgar em detalhes os formatos utilizados para estruturação da informação”.

A síntese do resultado da avaliação do Sistema Fiscalize em relação à LAI pode ser visto na tabela 6.

*Tabela 6. Síntese da Avaliação do Sistema Fiscalize frente à Lei de Acesso à Informação*

<b>Requisito</b>	<b>Dispositivo Legal</b>	<b>Conformidade</b>
<b>Conter ferramenta de pesquisa de conteúdo que permita o acesso à informação de forma objetiva, transparente, clara e em linguagem de fácil compreensão.</b>	LAI, art. 8º, §3º, inc. I	Atende parcialmente
<b>Possibilitar a gravação de relatórios em diversos formatos eletrônicos, inclusive abertos e não proprietários, tais como planilhas e texto, de modo a facilitar a análise das informações.</b>	LAI, art. 8º, §3º, inc. II	Não atende
<b>Possibilitar o acesso automatizado por sistemas externos em formatos abertos, estruturados e legíveis por máquina.</b>	LAI, art. 8º, §3º, inc. III	Não atende
<b>Divulgar em detalhes os formatos utilizados para estruturação da informação.</b>	LAI, art. 8º, §3º, inc. IV	Não atende
<b>Garantir a autenticidade e a integridade das informações disponíveis para acesso.</b>	LAI, art. 8º, §3º, inc. V	Atende
<b>Manter atualizadas as informações disponíveis para acesso</b>	LAI, art. 8º, §3º, inc. VI	Atende

Fonte: dados da pesquisa

#### 4.2.2 Sistema Siga Brasil

O Sistema Siga Brasil extrai dados do armazém de dados orçamentários do Senado e, por meio de técnicas de modelagem dimensional de dados e da tecnologia Business Objects, gera informações orçamentárias por meio de *relatórios dinâmicos*. Sua atualização é diária, o que implica que ele atende ao art. 8º, §3º, da LAI.

Suas fontes são as bases de dados de três sistemas de informação do Governo e do Legislativo Federal (BRASIL, 2013): Sistema Integrado de Administração Financeira (SIAFI), Sistema de Elaboração Orçamentária (SELOR) e Sistema Integrado de Orçamento Público (SIOP). Aliás, tal fato, em tese, também garante a autenticidade e a integridade das informações disponíveis para acesso – LAI, art. 8º, §3º, inc. V.

O Siga Brasil possui diversos módulos, tais como: Orçamento Fiscal e Seguridade – Execução; Emendas – Orçamento; Regiões, Estados e Municípios. Existe ainda uma funcionalidade na qual o próprio usuário pode construir seus relatórios, com bases nos ‘universos’ existentes na aplicação<sup>21</sup>. Optou-se por esta funcionalidade, porque ela permite acesso aos dados sobre convênios; além disso, ela demonstra a flexibilidade do sistema em relação ao Sistema Fiscalize.

*Figura 24. Siga Brasil – “Universos” de informações disponíveis (detalhe)*

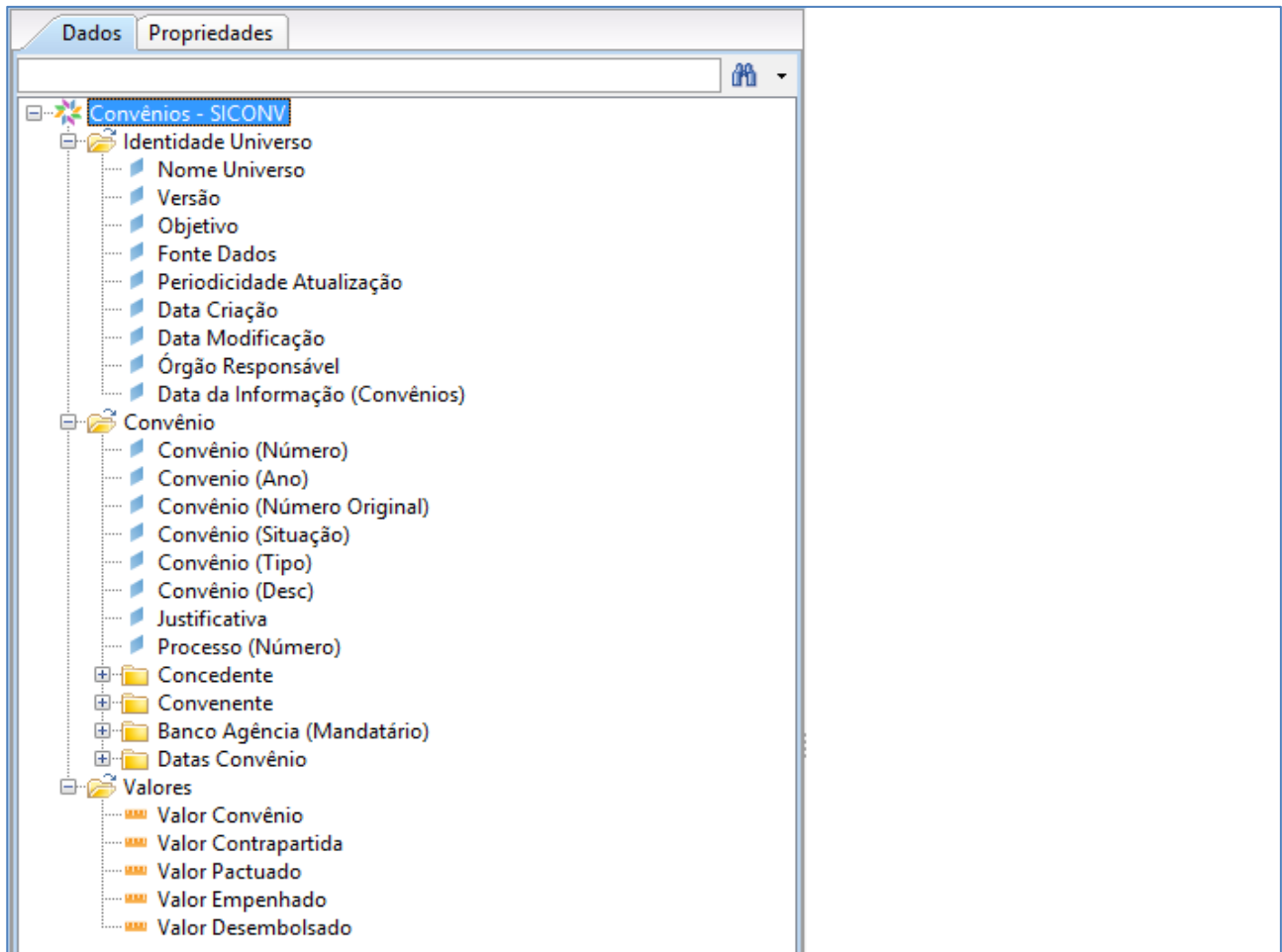
Universo	Descrição	Pasta
Censo 2010 - TIControl	Universo de EXEMPLO utilizado no MINICURSO do TIControl em 2011.	/Universos_Gabinets
Convênios	Contêm informações dos convênios firmados pelos entes da federação que até 2008 eram operacionalizados por meio do SIAFI.	/Universos_Gabinets
Convênios - SIAFI	Universo com informações das transferências voluntárias realizadas pelo Governo Federal com Estados, Municípios e entidades privadas em fins lucrativos utilizando o SIAFI.	/Universos_Gabinets
Convênios - SICONV	A partir de 2009 o Ministério do Planejamento lançou o sistema SINCOV, cujo objetivo é permitir a operacionalização dos convênios do governo federal com estados, municípios e entidades privadas. O DW de convênios apresenta esses convênios para assim promover a transparência e permitir o acompanhamento pela sociedade.	/Universos_Gabinets
LDO2007 - Emendas Apresentadas	Universo que disponibiliza as informações sobre as emendas propostas ao Congresso Nacional referentes ao ano de 2007, que posteriormente poderão ou não ser anexadas às leis. Contêm informações como as autorias das emendas, os textos emendados assim como os valores propostos às metas.	/Universos_Gabinets
LDO2011 - Emendas Apresentadas	Este universo disponibiliza as informações das Emendas Apresentadas ao PL da LDO2011.	/Universos_Gabinets

Fonte: Siga Brasil: <<http://www.sigabrasil.gov.br>>.

Para se chegar a essa funcionalidade, deve-se escolher a opção “Novo – Documento do Web Intelligence”; após a escolha será exibida uma tela com a lista dos universos disponíveis. A figura 24 exibe um detalhe desta tela. Ao se selecionar um dos conjuntos, o sistema abre uma janela com várias informações que podem ser combinadas para a geração dos relatórios – figura 25. A figura 26 exibe um relatório construído a partir do universo ‘Convênios – SICONV’, o qual contém dados semelhantes aos exibidos na avaliação do Sistema Fiscalize.

<sup>21</sup> O Sistema Siga Brasil possui boa documentação, incluindo materiais didáticos, que podem ser encontrado no endereço: <[http://www9.senado.gov.br/portal/page/portal/orcamento\\_senado/SIGA\\_EAD2](http://www9.senado.gov.br/portal/page/portal/orcamento_senado/SIGA_EAD2)>.

Figura 25. Siga Brasil – “Universo” Convênios - SICONV



Fonte: Siga Brasil: <<http://www.sigabrasil.gov.br>>.

Figura 26. Siga Brasil – “Universo” Convênios - SICONV

Título do relatório						
Convênio (Número)	Convenio (Ano)	Convênio (Situação)	Convênio (Desc)	Conveniente	Valor Convênio	Valor Empenhado
738942	2010	Em execução	Pavimentação asfáltica em ruas do perímetro URBANO do município, cita - se as referidas ruas: João da Costa Neves, Vicente S. Moreira, Geraldo Pereira Borges, João Batista Parreiras, Vicente C. Bicalho do bairro Sagrado Coração de Jesus.	PREFEITURA MUNICIPAL DE BONFIM	146.950	146.950

Fonte: relatório elaborado pelo autor a partir do sistema Siga Brasil, do Senado <[http://www9.senado.gov.br/portal/page/portal/orcamento\\_senado/SigaBrasil](http://www9.senado.gov.br/portal/page/portal/orcamento_senado/SigaBrasil)>.

É importante, nesse ponto, mostrar algumas diferenças relativas ao Sistema Fiscalize, no que tange à conformidade com a LAI. No Siga Brasil, por exemplo, é possível, a partir de então,



exportá-lo para PDF, XLS (Excel) e CSV, em conformidade com o que dispõe o art. 8º, §3º, inc. II da LAI, da referida lei. Outra diferença é a possibilidade se fazer modificações no relatório, tais como: reordená-lo com base em algum outro critério; acrescentar cálculos estatísticos simples e acrescentar filtros. É possível ainda observar alguns relacionamentos entre as categorias; no entanto, esses metadados só podem ser visualizados. Portanto, o Siga Brasil atende parcialmente o disposto no art. 8º, §3º, inc. IV da LAI. O sistema apresenta ainda metadados descritivos, que explicam alguns termos utilizados. Mesmo assim a linguagem não é de fácil compreensão para o público leigo. Dessa forma, o Siga Brasil atende apenas parcialmente o disposto no art. 8º, §3º, inc. I da LAI.

A despeito dos avanços do Siga Brasil em relação ao Fiscalize, *em termos de dados abertos* persistem alguns problemas. Apesar de ser bastante flexível e possibilitar a geração em formatos não proprietários, como o CSV, o acesso aos dados é indireto, via uma aplicação, e não diretamente, via um URI. Com isso, não seria possível automatizar o processo de extração dos dados. O formato CSV, por sua vez, é mais fácil de ser transformado em RDF, se comparado ao PDF. No entanto, o Siga Brasil não fornece os metadados semânticos em formatos recomendáveis pelo W3C, como o RDFS e a OWL. Na verdade, os metadados podem ser visualizados, mas não estão disponíveis em um local físico acessível por um URI baseado em HTTP. Logo, o Siga Brasil não atende o disposto no art. 8º, §3º, inc. III da LAI. A síntese do resultado da avaliação do Siga Brasil em relação à LAI é apresentada na tabela 7.

*Tabela 7. Síntese da Avaliação do Sistema Siga Brasil frente à Lei de Acesso à Informação*

<b>Requisito</b>	<b>Dispositivo Legal</b>	<b>Nível de Atendimento</b>
<b>Conter ferramenta de pesquisa de conteúdo que permita o acesso à informação de forma objetiva, transparente, clara e em linguagem de fácil compreensão.</b>	LAI, art. 8º, §3º, inc. I	Atende parcialmente
<b>Possibilitar a gravação de relatórios em diversos formatos eletrônicos, inclusive abertos e não proprietários, tais como planilhas e texto, de modo a facilitar a análise das informações.</b>	LAI, art. 8º, §3º, inc. II	Atende
<b>Possibilitar o acesso automatizado por sistemas externos em formatos abertos, estruturados e legíveis por máquina.</b>	LAI, art. 8º, §3º, inc. III	Não atende
<b>Divulgar em detalhes os formatos utilizados para estruturação da informação.</b>	LAI, art. 8º, §3º, inc. IV	Atende parcialmente
<b>Garantir a autenticidade e a integridade das informações disponíveis para acesso.</b>	LAI, art. 8º, §3º, inc. V	Atende
<b>Manter atualizadas as informações disponíveis para acesso</b>	LAI, art. 8º, §3º, inc. VI	Atende

Fonte: dados da pesquisa

Logo, a análise efetuada ao longo da seção indica que os Sistemas Fiscalize e Siga Brasil, isoladamente, não atendem os requisitos estabelecidos pelo art. 8º da Lei de Acesso à Informação, no tocante às informações orçamentárias. Ou seja, o pressuposto P1 foi confirmado.

Nas próximas seções serão discutidos e ilustrados aspectos referentes à aplicabilidade do padrão *dados abertos ligados* para a construção de aplicações que atendam plenamente o disposto na legislação e permitam, no futuro, a interligação com outras informações de interesse governamental e público, as quais hoje, embora disponíveis, encontram-se dispersas e em formatos distintos, prejudicando sua interoperabilidade. A discussão dar-se-á em três aspectos: jurídico – principalmente no que tange à Lei de Acesso à Informação, administrativo e técnico. A seção 3.3 apresentará os conceitos básicos dos dados abertos ligados e apresentará dois casos que implementam esses conceitos. A seção 3.4 apresenta uma síntese dessa análise.

### 4.3 O Padrão ‘Dados Abertos Ligados’ (LOD)

O conceito de Dados Abertos Ligados, conhecido pela sigla LOD (do inglês *Linked and Open Data*) deriva de dois outros conceitos: dados abertos (*open data*) e dados ligados (*linked data*).

O W3C (W3C Escritório Brasil, 2010) adota as três leis formuladas por Eaves (2009) para estabelecer se um dado é ou não um *dado governamental aberto* (*government open data*):

1. Se ele não pode ser encontrado na web e indexado, ele não existe.
2. Se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser utilizado.
3. Se qualquer dispositivo legal não permitir que ele seja reutilizado, ele não é útil.

É possível, no entanto, ir além dos dados governamentais abertos: ao invés de se pensar apenas em abrir os dados acerca de determinado tema ou órgão, pode-se pensar em interligar os dados de temas relacionados entre si, ampliando o poder de análise sobre esses dados. Por exemplo, interligar dados de transferências orçamentárias da União para determinadas localidades a índices socioeconômicos dessas localidades, ampliam o poder de análise sobre efetividade das políticas públicas.

Essa evolução dos dados abertos de governo tem sido referenciada pela literatura como *dados ligados de governo* (*government linked data*). Conforme Bizer *et al.* (2009), os dados ligados “referem-se a dados publicados na Web de modo que sejam legíveis por máquina, seus significados sejam *explicitamente* definidos, estejam ligados a outros conjuntos de dados e, por sua vez, possam ser ligados a partir de conjuntos de dados externos”. Outro conceito, extraído da organização Linkeddata.org é: um termo usado para descrever uma boa prática recomendada para expor, compartilhar e conectar pedaços de dados, informações e conhecimentos na Web Semântica usando URIs e RDF” (LINKEDDATA, 2013). Portanto, a despeito do nome ‘dados ligados’, o conceito

transcende a exposição somente dos dados: por meio da formalização das relações entre as categorias associadas aos dados e suas instâncias, ou seja, os dados propriamente ditos, também informações e conhecimento a respeito dos dados são compartilhados. E o uso das tecnologias URI e RDF, ao invés de simplesmente URL e HTML, é fundamental. De fato, como asseveram Heath and Bizer (2011), os *links* HTML normalmente indicam que dois documentos são de alguma forma relacionados, mas deixa ao usuário a tarefa de inferir a *natureza* do relacionamento. O RDF, por outro lado, possibilita a quem publica os dados definir, *de forma explícita*, a natureza da conexão.

A ideia básica de dados ligados foi elaborada por Berners-Lee (2006), que também criou o conceito de Web Semântica, como já foi visto. Posteriormente, essa ideia foi sendo refinada por outros colaboradores da Web Semântica, como os já citados Heat e Bizer. Esses autores sintetizam os princípios dos dados ligados da seguinte forma (HEAT e BIZER, 2011):

- Princípio 1. Usar referências URI para identificar não apenas conteúdos digitais e documentos Web, mas também objetos do mundo real e conceitos abstratos: “essas [referências] podem incluir coisas tangíveis, tais como pessoas, lugares e carros, ou aquelas que são mais abstratas, como o relacionamento do tipo *conhecer alguém*, ou o conjunto de todos os carros verdes no mundo, a própria cor verde” (s/p).
- Princípio 2. Usar URIs baseadas em HTTP para identificar objetos e conceitos abstratos, habilitando esses URIs a ser ‘derreferenciáveis’ (*dereferenceable*). Isso significa que os clientes HTTP podem procurar o URI usando o protocolo HTTP e recuperar a descrição do recurso referenciado pelo URI.
- Princípio 3. Usar um único modelo de dados para publicar dados estruturados na Web: o RDF (brevemente descrito no capítulo 3).
- Princípio 4. Os *hyperlinks* devem ser usados para conectar não apenas documentos Web, mas qualquer tipo de coisa: “um *hyperlink* pode ser estabelecido entre uma pessoa e um lugar, ou entre uma pessoa e uma companhia” (s/p). No contexto dos Dados Ligados, os *hyperlinks* que conectam coisas possuem tipos que *descrevem o relacionamento* entre essas coisas.

Os princípios dos dados abertos têm sido também incorporados na esfera governamental, no contexto do chamado Governo Aberto (*Open Government*). Esse fenômeno é denominado Dados Abertos de Governo (OGD – *Open Government Data*). Segundo Bauer e Kaltenböck (2012, p. 10, tradução livre):

*O OGD é um movimento em nível global para promover a abertura de dados, informações e conteúdos públicos e governamentais – tanto para humanos quanto para formatos legíveis por máquina e não-proprietários – para reuso pela sociedade civil, economia, imprensa e academia assim como por políticos e administradores públicos. Ele deve se aplicar somente aos dados e informações produzidos ou patrocinados pelo governo ou entidade controladas pelo governo e não é relacionado aos dados sobre indivíduos.*

Em particular, no âmbito do governo, essa estratégia é denominada ‘dados abertos ligados de governo’ – LOGD (*Linked Open Government Data*). Ela vem emergindo da Web de dados ligados como uma forma de facilitar a abertura, a interligação e o reuso dos dados abertos de governo (DING *et al.*, 2012). A figura 27 apresenta uma síntese desse conceito.

*Figura 27. Dados ligados de governo*

<p><i>dados abertos de governo + estratégia baseada em dados ligados</i></p> <p><i>= dados abertos ligados de governo</i></p>
---

Fonte: adaptado de Berners-Lee (2009) e Ding *et al.* (2012)

No contexto deste trabalho, é relevante notar que os princípios dos dados ligados, aplicados ao governo, são compatíveis com os requisitos estabelecidos pela Lei de Acesso à Informação (LAI). O atendimento ao Princípio 1 (“Usar URIs baseadas em HTTP para identificar objetos e conceitos abstratos”, de modo que “os clientes HTTP podem procurar a URI usando o protocolo HTTP e recuperar a descrição do recurso referenciado pelo URI”) e ao Princípio 3 (“Usar um único modelo de dados para publicar dados estruturados na Web: o RDF”) implica no atendimento a três dispositivos da LAI: (1) art. 8º, §3º, inc II: “possibilitar a gravação de relatórios em diversos formatos eletrônicos, inclusive abertos e não proprietários, tais como planilhas e texto, de modo a facilitar a análise das informações”; (2) art. 8º, §3º, inc. III: “possibilitar o acesso automatizado por sistemas externos em formatos abertos, estruturados e legíveis por máquina”; (3) art. 8º, §3º, inc. IV: “divulgar em detalhes os formatos utilizados para estruturação da informação”.

Os quatro princípios dos dados ligados só podem ser adotados plenamente com a utilização da Web Semântica, uma vez que: (i) ela fornece tecnologias e padrões abertos, bem definidos e com forte embasamento teórico, que estabelecem especificações claras e consistentes para representar, interligar, consolidar e recuperar a informação a partir dos dados ligados; (ii) atualmente há um ecossistema de organizações que desenvolvem ferramentas computacionais que implementam essas especificações.

Do ponto de vista técnico, os dados de governo ligados podem ser produzidos em duas etapas: conversão e aprimoramento (*enhancement*). A etapa de conversão é constituída, por sua vez, de duas fases principais. Na primeira, os dados de governo “brutos” (*raw data*) são “limpos” e gravados mediante representação baseada na estrutura RDF, estudada no capítulo 2. Na segunda, estes conjuntos de dados convertidos podem ser referenciados por meio de URIs, de forma que tanto os conjuntos quanto suas ontologias podem ser estendidas por terceiros: cidadãos, ONGs, outros órgãos de governo (BERNERS-LEE, 2009).

Uma vez convertidos os dados para a estrutura RDF, pode-se iniciar a etapa de aprimoramento (*enhancement*) destes, a qual pode ser feita em quatro fases principais: adicionar dados aos dados de governo preexistentes; derivar dados de governo; ligar dados de governo; integrar dados de governo.

Portanto, a implementação de *linked data* depende das tecnologias URI, RDF e SPARQL, as quais foram brevemente descritas no capítulo 2. No escopo deste texto, o importante é notar que, no percurso que vai do dado “bruto” até chegar ao RDF e à OWL, significados vão sendo adicionados aos dados, por meio da sofisticação da sintaxe das estruturas que descrevem estes dados. Dessa maneira, os usuários e as aplicações conseguem transformar esses dados em informações mais ricas e contextualizadas, as quais, por sua vez, podem ser associadas a outras informações, possibilitando uma recuperação da informação mais precisa e completa. A seguir serão descritas duas aplicações construídas a partir dos pressupostos dos dados abertos ligados e que utilizam as tecnologias mencionadas.

#### 4.3.1 Data-Gov Wiki

Segundo Ding et al. (2010) o Data-gov Wiki é um projeto que está sendo conduzido pelo Tetherless World Constellation, no Instituto Politécnico Rensselaer. Eles estão pesquisando o uso de tecnologias da Web Semântica na exploração e visualização de conjuntos de dados abertos de governo. A maior parte dos dados é oriunda do portal de governo norte-americano [data.gov](http://www.data.gov)<sup>22</sup>; o restante é oriundo de bases de dados de outros países, bem como fontes não-governamentais.

Utilizou-se a seguinte estratégia para permitir a navegação semântica nos dados (DING et al., 2010):

1. Converter os dados do portal [data.gov](http://www.data.gov) em formato legível por máquina – no caso, RDF;
2. *Enriquecer e ligar* os dados por meio de vários tipos de processamento (por exemplo: extração, normalização e mapeamento);
3. Mostrar o *valor dos dados enriquecidos* usando aplicações interessantes de exploração e visualização.
4. Mostrar a *proveniência do conhecimento sobre os dados* por meio de metadados e processos de rastreabilidade.

---

22      Endereço: <<http://www.data.gov>>.

O que se percebe nesse processo é uma transformação paulatina de dados brutos em informações úteis e acessíveis a diversos públicos, de forma flexível e variada. E isso só é possível pelo uso das tecnologias da Web Semântica, que agregam significado aos dados e, em paralelo, possibilitam que eles sejam estruturados de forma a facilitar seu processamento por máquinas e sua interligação com outros dados.

Conforme mencionado, uma das etapas na estratégia da Data-Gov Wiki consiste em mostrar o valor dos dados (previamente enriquecidos), usando aplicações de exploração e visualização. Para isso, foram construídos vários aplicativos demonstrativos. Para cada um deles, há uma página na Data-Gov Wiki, descrevendo seus objetivos e citando as tecnologias utilizadas na sua construção.

Para ilustrar, são apresentadas, nas figuras 28 e 30, descrições do aplicativo “Broadband Internet Use versus Total Internet Use – United States Households, 2009”<sup>23</sup>, que permite visualizar, de forma geograficamente referenciada, a informação da taxa de adesão à banda larga em comparação com o acesso total à Internet.

Pela figura, vê-se que as tecnologias utilizadas nesta aplicação demonstrativa são: RDF, SPARQL, SparqlProxy e Google Visualization API<sup>24</sup>. Importante notar que todas são gratuitas; ademais, as duas primeiras são ainda abertas e padronizadas pelo W3C, conforme mencionado anteriormente (em especial no capítulo 2). Além disso, são *aplicações genéricas*, ou seja, basta que os dados sejam passados em um formato predeterminado, a saída será produzida e a interação, possibilitada, não importando o conteúdo desses dados. Com isso, não há necessidade de se produzir novos aplicativos para cada contexto de uso, economizando recursos preciosos dos governos e das organizações civis. Além disso, são apresentadas também informações da origem dos dados, como a identificação do conjunto de dados (*Dataset 10040*) e suas características mais relevantes.

Do ponto de vista técnico, o aplicativo fornece outras informações importantes, como a consulta SPARQL utilizada, conforme figura 29.

---

23 <<http://data-gov.tw.rpi.edu/demo/stable/demo-10040-broadband-home.html>> .

24 A lista completa das tecnologias utilizadas em cada aplicativo encontra-se em <[http://data-gov.tw.rpi.edu/wiki/Property:Technology\\_used](http://data-gov.tw.rpi.edu/wiki/Property:Technology_used)>.

Figura 28. Metadados de uma aplicativo de demonstração do portal Data-Gov Wiki

Live Demo(s)	<ul style="list-style-type: none"> <li>• <a href="http://data-gov.tw.rpi.edu/demo/stable/demo-10040-broadband-home.html">http://data-gov.tw.rpi.edu/demo/stable/demo-10040-broadband-home.html</a></li> </ul>
Video Demo(s)	
Data.gov Data source(s)	
Other Data source(s)	<ul style="list-style-type: none"> <li>• <a href="#">Dataset 10040</a> (Households using the Internet in and outside the home, by selected characteristics: Total, Urban, Rural, Principal City, 2009, <a href="#">National Telecommunications and Information Administration</a>)</li> </ul>
Technology Used	<ul style="list-style-type: none"> <li>• <a href="#">RDF</a></li> <li>• <a href="#">SPARQL</a></li> <li>• <a href="#">SparqlProxy</a></li> <li>• <a href="#">Google Visualization API</a></li> </ul>
Related SPARQL	<ul style="list-style-type: none"> <li>• <a href="http://data-gov.tw.rpi.edu/demo/stable/demo-10040-broadband-home.sparql">http://data-gov.tw.rpi.edu/demo/stable/demo-10040-broadband-home.sparql</a></li> </ul>
Related Demo(s)	

Fonte: <[http://data-gov.tw.rpi.edu/wiki/Demo:\\_Broadband\\_Internet\\_Use\\_versus\\_Total\\_Internet\\_Use\\_-\\_United\\_States\\_Households,\\_2009](http://data-gov.tw.rpi.edu/wiki/Demo:_Broadband_Internet_Use_versus_Total_Internet_Use_-_United_States_Households,_2009)>

Figura 29. Consulta SPARQL sobre uma base descrita no portal Data-gov Wiki

```

PREFIX d: <http://data-gov.tw.rpi.edu/vocab/p/10040/>
SELECT ?state ?urban_home_broadband ?urban_home_total
WHERE {
  graph <http://data-gov.tw.rpi.edu/vocab/Dataset_10040>{
    ?s d:state ?state .
    ?s d:urban_internet_use_in_the_home_broadband_no ?urban_home_broadband .
    ?s d:urban_internet_use_in_the_home_urban_total_no ?urban_home_total .
    FILTER (?state != "TOTAL HOUSEHOLDS")
  }
}
ORDER BY ?state

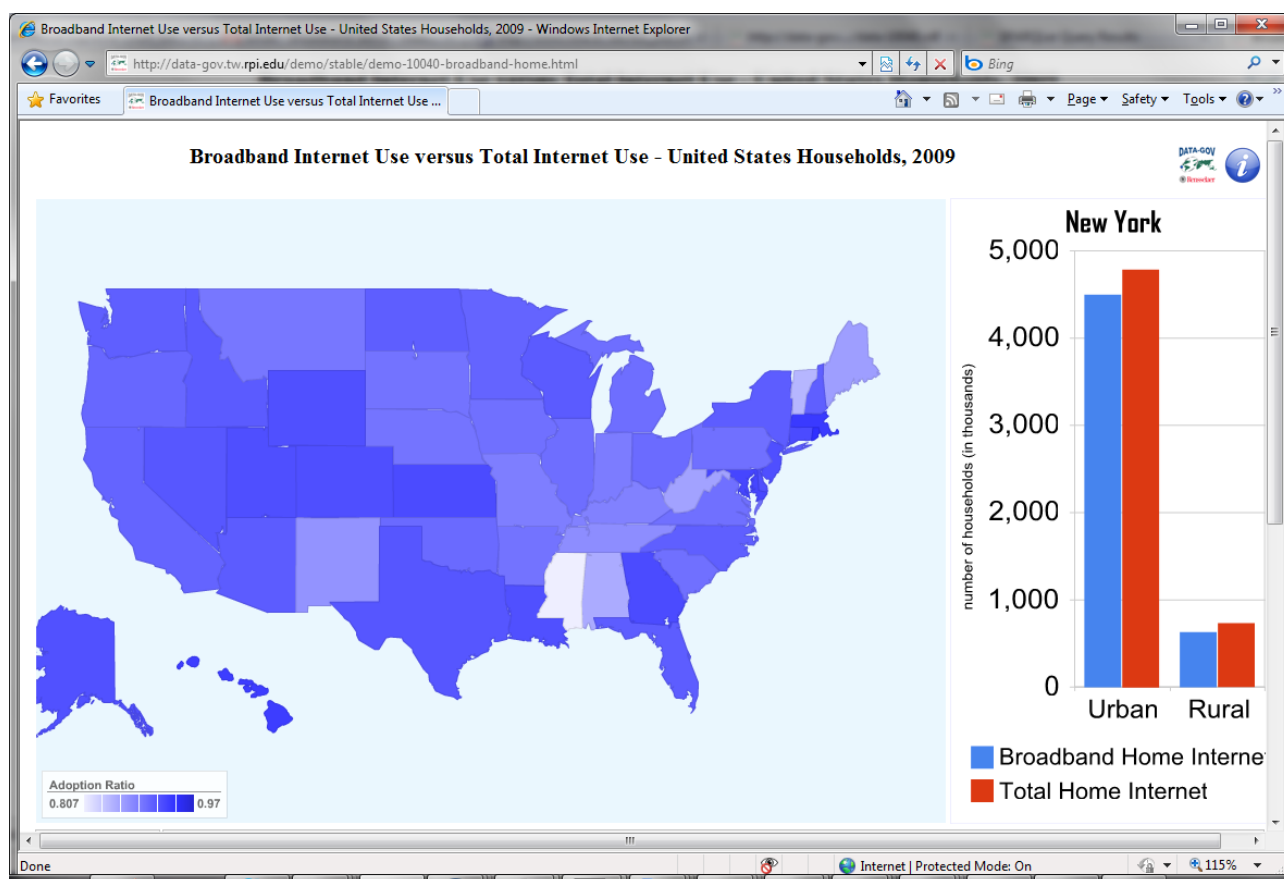
```

A consulta acima pode ser explicada da seguinte forma: a cláusula `PREFIX` indica a origem dos dados: no caso, a variável `d` passa a representar esse valor, o que elimina a necessidade de repeti-lo no restante da consulta; a cláusula `SELECT` indica quais atributos (campos) serão selecionados do conjunto de dados indicado; a cláusula `WHERE` indica as condições (critérios) que deverão ser satisfeitas: ou seja, somente o subconjunto de dados que atender aos critérios será recuperado; a cláusula indica um filtro, no caso, que só serão retornados dados cujo atributo `?state` seja diferente da expressão `"TOTAL HOUSEHOLDS"` (que corresponde à totalização das residências); finalmente, a cláusula `ORDER BY` indica que os dados serão classificados por ordem de nome de estado.

Na figura 30, vê-se uma tela do aplicativo propriamente dito exibindo o resultado da

consulta. Conforme mencionado, ele é referenciado geograficamente: quando se posiciona o ponteiro do mouse em um estado, ele exibe as informações sobre taxa de uso de banda larga de Internet, conforme mencionado. Trata-se, claramente, de aplicativo com interface intuitiva, o que facilita o acesso à informação. Além disso, a descrição logo abaixo do mapa – oriunda dos metadados da fonte original (data.gov) – possibilitam a compreensão da informação. Tal constatação indica que é possível, a partir do uso das tecnologias disponíveis e utilizando o padrão dados abertos ligados, atender ao que dispõe o inc. I do art. 3 da LAI : “Conter ferramenta de pesquisa de conteúdo que permita o acesso à informação de forma objetiva, transparente, clara e em linguagem de fácil compreensão”.

*Figura 30. Resultado da consulta sobre uma aplicação disponível no portal Data-gov Wiki*



Fonte: <[http://data-gov.tw.rpi.edu/wiki/Demo:\\_Broadband\\_Internet\\_Use\\_versus\\_Total\\_Internet\\_Use\\_-\\_United\\_States\\_Households,\\_2009](http://data-gov.tw.rpi.edu/wiki/Demo:_Broadband_Internet_Use_versus_Total_Internet_Use_-_United_States_Households,_2009)>

#### 4.3.2 Ligado nos políticos

Este projeto<sup>25</sup> foi proposto e desenvolvido por Araújo e Souza (2011). Segundo os autores, foram utilizados como fontes de dados: os portais da Administração Pública do Tribunal Superior

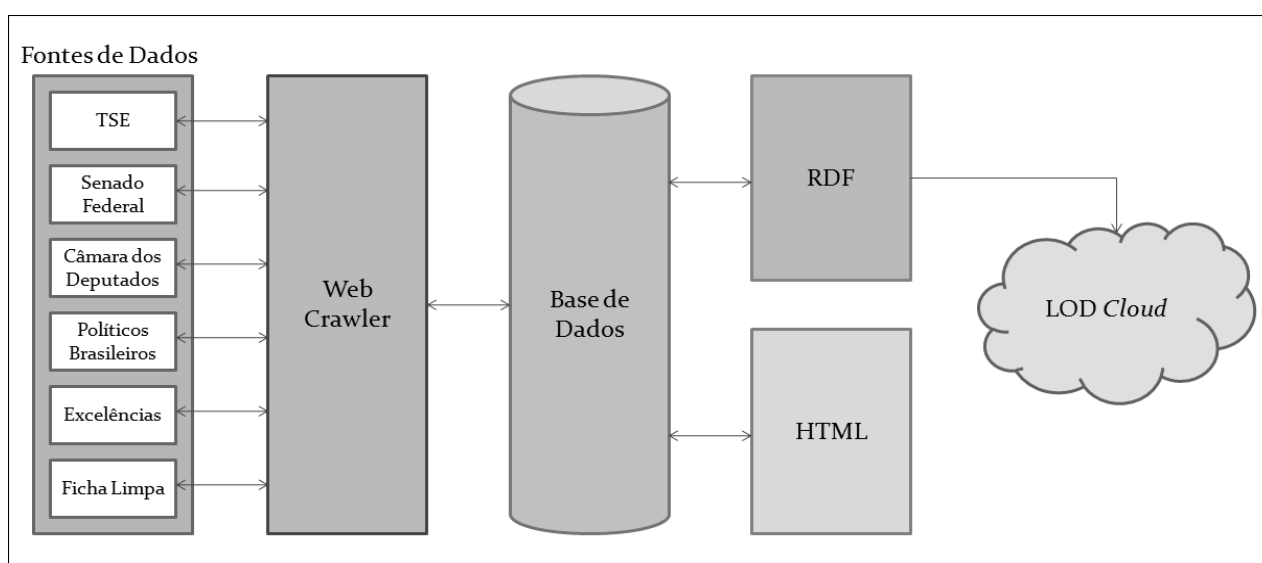
25 Endereço: <<http://ligadonospoliticos.com.br>>.



Eleitoral (TSE), do Senado e da Câmara dos Deputados; os portais das ONGs (Organização Não Governamentais) Políticos Brasileiros, Ficha Limpa e do projeto Excelências da Transparência Brasil (ARAÚJO e SOUZA, 2011). Os seguintes dados foram coletados: “dados pessoais, dados da eleição, divulgação de bens, dados parlamentares, lideranças, missões, mandatos, afastamentos, pronunciamentos, comissões, proposições e ocorrências” (p. 8).

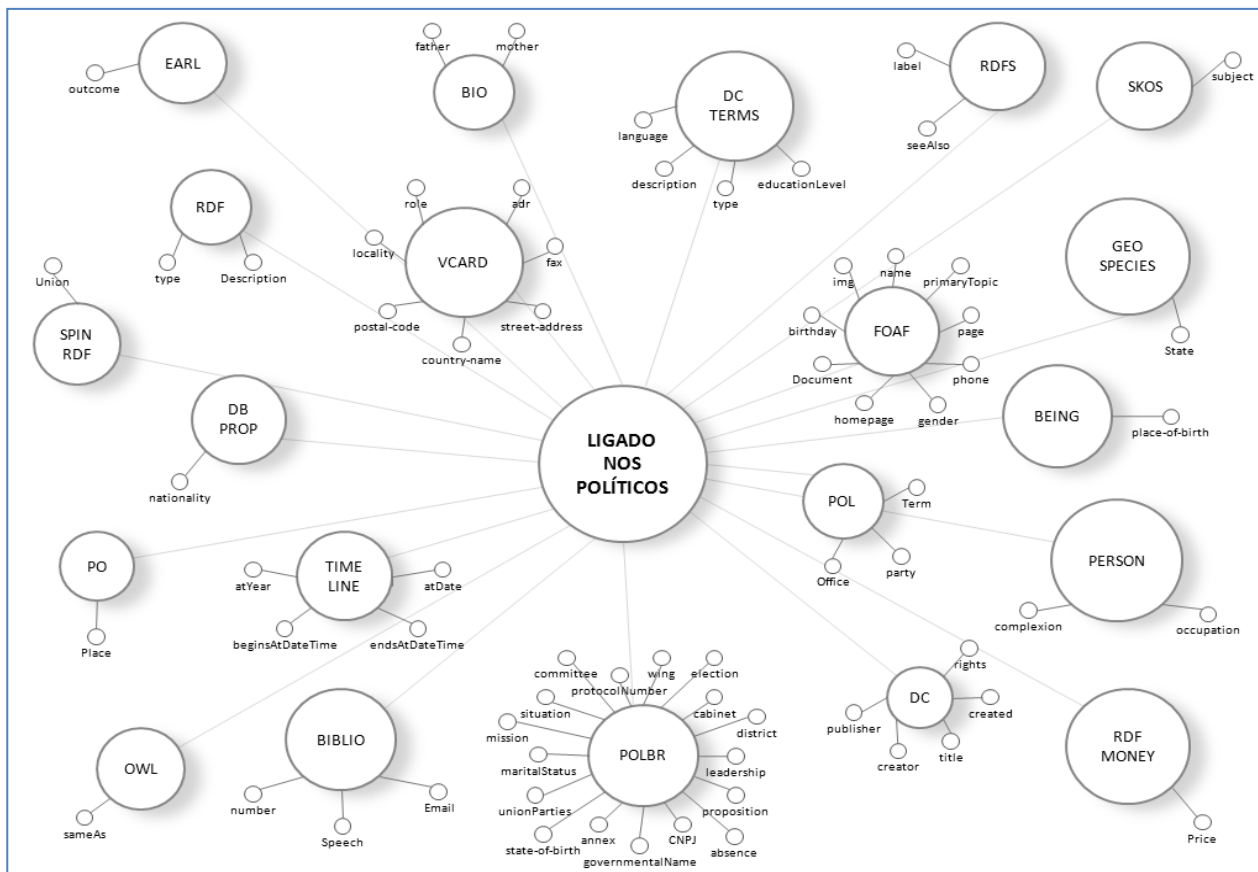
Os autores relatam que poucos dados estão disponíveis em formatos mais abertos e de mais fácil tratamento, como CSV; a maior parte estava disponível somente em formato HTML. Por isso, “foi necessária a criação e utilização de Web Crawlers para extrair os dados de uma forma metódica e automatizada” (p. 8). Esta técnica para extrair dados disponíveis na Internet e que não estão em formatos abertos é denominada *screen scraping* ou “raspagem de dados” (ARAÚJO e SOUZA, 2011). Uma visão geral da arquitetura utilizada pode ser visto na figura 31.

Figura 31. Visão geral da arquitetura do portal Ligados nos Políticos



Fonte: ARAÚJO, Lucas de Ramos; SOUZA, Jair Francisco de. Aumentando a Transparência do Governo por Meio da Transformação de Dados Governamentais Abertos em Dados Ligados . *Revista Eletrônica de Sistemas de Informação*, v. 10, n. 1, artigo 7, 2011.

Figura 32. Vocabulários e propriedades usados no portal Ligados nos Políticos



Fonte: ARAÚJO, Lucas de Ramos; SOUZA, Jair Francisco de. Aumentando a Transparência do Governo por Meio da Transformação de Dados Governamentais Abertos em Dados Ligados . *Revista Eletrônica de Sistemas de Informação*, v. 10, n. 1, artigo 7, 2011.

Figura 33. Detalhe da página inicial do portal Ligados nos Políticos

Home	Visualizações	Downloads	Dados Abertos	Dados Ligados	Sobre o Projeto	Links	Contato
------	---------------	-----------	---------------	---------------	-----------------	-------	---------

O site **Ligado nos Políticos** tem como objetivo fornecer dados de políticos brasileiros usando os conceitos de [Dados Ligados](#) e [Dados Governamentais Abertos](#).  
Utilize os mecanismos de busca abaixo para encontrar dados sobre os políticos desejados.

Políticos Cadastrados:  
**22475**

### Busca

Nome:

Situação:

Cargo:


Estado:

Partido:

Sexo:

Naturalidade:

### Estados



Fonte: <<http://ligadonospoliticos.com.br>>.

A figura 32 mostra os vocabulários e propriedades utilizados na elaboração do conjunto de dados ligados; além disso, possibilita ainda a identificação da origem desses vocabulários (ontologias) e das relações entre eles.

A figura 33 exhibe a página principal do portal. Trata-se de uma interface de busca simples, baseada em 'filtros' nos campos “Nome”, “Situação”, “Cargo”, “Estado”, “Partido”, “Sexo” e “Naturalidade”. Apesar da simplicidade da interface, o tratamento da informação torna a busca eficiente e a informação recuperada com um bom nível de semântica agregado, pela interligação de dados oriundos de bases distintas, porém complementares.

Também esta aplicação, assim como a anterior (Data.gov Wiki), demonstra o potencial das soluções baseadas em dados abertos ligados na representação e disseminação da informação governamental, atendendo aos requisitos da Lei de Acesso à Informação. As aplicações indicam ainda a viabilidade técnica desse tipo de solução.

Na próxima seção a aplicação dos dados abertos ligados será discutida no âmbito da representação e disseminação da informação orçamentária.

#### **4.4 Viabilidade dos Dados Abertos Ligados no Contexto da Informação Orçamentária**

Pelo exposto, a conjunção dos dados abertos de governo e as técnicas, tecnologias e ferramentas da Web Semântica já é uma realidade. Em particular, os casos apresentados buscam, justamente, a construção dos dados ligados (*linked data*) como etapa fundamental para agregar valor ao conjunto de dados original, enriquecendo-o com a associação a outros conjuntos de dados conexos. Ou seja, a utilização de conjuntos de dados (*datasets*) expressos em formatos abertos e referenciáveis, no caso, o RDF, e associados a vocabulários comuns, também abertos e referenciáveis, expressos em RDFS e OWL.

No caso dos sistemas Siga Brasil e Fiscalize, ao contrário, os conjuntos de dados estão em bases de dados proprietárias e a semântica das relações entre as categorias (classes) e indivíduos (instâncias) está implícita nos sistemas e nas próprias bases. Isso acarreta algumas desvantagens: (1) os dados não podem ser acessados diretamente por uma aplicação via URI; (2) qualquer interligação entre os conjuntos de dados demanda grande esforço, uma vez que a semântica deverá ser inferida (por um processo de ‘engenharia reversa’), uma vez que não está explícita e formalizada, o que leva a (3) os usuários finais da informação, mesmo que tenham acesso aos dados, não conhecem as relações entre eles: e isso, certamente, reduz o valor da informação.

O caso das informações sobre transferências a municípios, explorado na seção anterior, ilustra bem essa situação, uma vez que estão dispersas em diversos sistemas: Selor, Siafi, Siconv e Sistema de Acompanhamento de Obras da Caixa Econômica Federal. No Selor encontram-se informações sobre a elaboração orçamentária, nos quais os programas de trabalho são definidos. O Siafi contém informações sobre a execução orçamentária desses programas de trabalho. O Siconv possui o cadastro das informações das transferências voluntárias (incluindo convênios), os quais, como se viu, são associados a programas de trabalho. Finalmente, o Sistema de Acompanhamento de Obras possui informações detalhadas do andamento de obras, as quais podem ser objetos de contratos de repasse firmados entre a União e os municípios. Portanto, os dados são complementares, mas não estão disponíveis a partir de um só ponto de acesso, nem são totalmente processáveis por máquina.

Um dos pressupostos deste trabalho (P2) é que a construção de aplicações e conjuntos de dados (*datasets*) baseadas no padrão *dados abertos ligados* é uma solução viável para a representação e disseminação da informação orçamentária, não só atendendo ao que determina a Lei de Acesso à Informação, mas permitindo a interligação entre dados e metadados, aumentando, assim, o valor dessa informação.

Baseado no que foi visto, seriam três as etapas básicas para a construção de uma aplicação baseada em dados ligados para representação e disseminação das informações orçamentárias (adaptado de DING *et al.*, 2010):

1. *Converter os dados orçamentários* em formato legível por máquina – no caso, RDF;
2. *Enriquecer e ligar os dados orçamentários* por meio de vários tipos de processamento (por exemplo: extração, normalização e mapeamento);
3. *Mostrar o valor dos dados orçamentários enriquecidos* usando aplicações interessantes de exploração e visualização.
4. *Mostrar a proveniência do conhecimento sobre os dados orçamentários* por meio de metadados e processos de rastreabilidade.

Portanto, uma forma de avaliar a viabilidade da adoção do padrão dados abertos ligados é analisar a viabilidade da implementação das etapas citadas. Com esse intuito, será explicado no que consiste cada etapa. Também serão destacados exemplos de tecnologias usadas em cada uma delas, bem como aplicações que as utilizam. Além disso, a título de ilustração das etapas 1 e 2, será apresentado um exercício de modelagem de uma ontologia.

#### 4.4.1 A viabilidade da etapa de conversão

Segundo DING *et al.* (2010), o processo de conversão é “razoavelmente direto e altamente extensível”. Inicialmente os dados ‘crus’ de governo (*raw data*) “passam por um processo de ‘limpeza’ e preservação por meio de representações baseadas em RDF”. Posteriormente, “esses conjuntos de dados convertidos passam a ser navegáveis via URIs, de forma que *tanto os conjuntos de dados quanto suas ontologias possam ser estendidos por terceiros*” (p. 39, grifos nossos).

Para se iniciar a etapa de conversão, é necessário se definir uma ou mais fontes iniciais de dados orçamentários. Conforme foi visto na seção 3.2, são fontes as bases de dados dos sistemas a seguir: Sistema de Elaboração de Leis Orçamentárias (Selor), Sistema Integrado de Administração Financeira (Siafi), Siafi Gerencial, Sistema de Gestão de Convênios e Contratos de Repasse (Siconv) e Sistema de Acompanhamento de Obras da Caixa Econômica Federal (Siurb)<sup>26</sup>. Naturalmente, não há necessidade de se converter todas as bases para se passar à próxima etapa. O trabalho pode ser feito de forma iterativa e incremental.

Importante ressaltar que várias dessas bases de dados estão disponíveis, sendo que algumas delas em formato de dados abertos, ainda que não de dados ligados. A base de dados do Selor é hospedada e gerida pelo próprio Senado e contém dados sobre elaboração orçamentária: PPA, LDO e LOA (BRASIL, 2007); ela ainda não está disponível diretamente em formatos abertos. A base de dados do Siconv está disponível nos formatos CSV e XML e com acesso via HTTP e já é aderente ao padrão dados abertos<sup>27</sup>, mas não ao padrão dados ligados. Os dados do Siafi Gerencial são transmitidos diariamente tanto à Câmara dos Deputados, quanto ao Senado – são esses dados que alimentam os armazéns de dados, que, por sua vez, são utilizados pelos Sistemas Fiscalize e Siga Brasil, conforme seção 3.2.

Portanto, entre as principais fontes de dados, restaria apenas o acesso às bases de dados do Siafi e do Siurb. No entanto, é importante ressaltar que parte dos dados da base do Siafi encontra-se no Siafi Gerencial – apenas de forma mais agregada. Acredita-se, portanto, que os dados do Siafi Gerencial disponíveis hoje no Congresso Nacional sejam suficientes para se iniciar o trabalho de conversão. Quanto aos dados do Siurb, há a possibilidade de se firmar um convênio com seu gestor, a Caixa Econômica Federal, para a geração dos dados em algum formato aberto, que facilitasse a conversão para RDF. De qualquer forma, as bases já disponíveis fornecem dados suficientes para, após a realização das três etapas de construção dos dados abertos ligados, prover o

<sup>26</sup> Vide <[http://www1.caixa.gov.br/gov/gov\\_social/municipal/assistencia\\_tecnica/siurb\\_index.asp](http://www1.caixa.gov.br/gov/gov_social/municipal/assistencia_tecnica/siurb_index.asp)>.

<sup>27</sup> Vide <<http://api.convenios.gov.br/siconv/doc/>>.

acompanhamento integrado de grande parte da elaboração e da execução orçamentária.

A etapa de conversão exige também o uso de tecnologias, as quais poderiam ser desenvolvidas *interna corporis*; no entanto, felizmente, várias delas já estão disponíveis, em licenças *free and open source* (FOS). Algumas delas são destacadas na tabela 8. Para conhecer as tecnologias citadas e como utilizá-las, recomenda-se a leitura de Hebeler et al. (2009) e Allemang e Hendler (2009), além dos *websites* citados no rodapé da própria tabela.

Portanto, a etapa de conversão é viável, desde que ela seja feita de forma iterativa e incremental, uma vez que já existem bases de dados e tecnologias disponíveis. A conversão poderia se iniciar com as bases já existentes em formatos abertos (Siconv, LOA 2012). Posteriormente, poderiam ser convertidas aquelas as quais o Congresso Nacional tem acesso (Selor, Siafi Gerencial). Futuramente, na medida em que outras bases sejam abertas (Siurb, Siafi), elas também poderiam ser incorporadas. Para isso, seria necessário que as ontologias fossem estendidas, o que corresponde à etapa de enriquecimento.

*Tabela 8. Tecnologias para conversão de dados para RDF, manipulação e gravação*

<i>Tecnologia</i>	<i>Descrição</i>
csv2rdf4lod <sup>(1)</sup>	Permite a conversão de dados em formato ‘valores separados por vírgulas’ ( <i>Comma-Separated Values – CSV</i> ) para RDF.
Virtuoso Triple Store <sup>(2)</sup>	Permite armazenar triplas RDF.
Apache Jena <sup>TM</sup> <sup>(3)</sup>	Framework que provê uma coleção de ferramentas e bibliotecas escritas em Java que auxiliam o desenvolvimento de aplicações baseadas em Web Semântica e dados ligados.
RDFLib <sup>(4)</sup>	Biblioteca baseada na linguagem de programação Python, permite gerar, manipular e recuperar informação a partir de triplas RDF.
pyrdfa3 <sup>(5)</sup>	Consegue extrair dados expressos em RDFa 1.1 a partir de (X)HTML, SVG e XML. Baseada em Python.
R2RML <sup>(6)</sup>	Uma linguagem para expressar mapeamentos customizados de bases de dados relacionais para conjuntos de dados RDF.

Fontes: (1) <[http://logd.tw.rpi.edu/tutorial/getting\\_started\\_csv2rdf4lod\\_create\\_verbatim\\_rdf\\_conversions\\_tabular\\_data](http://logd.tw.rpi.edu/tutorial/getting_started_csv2rdf4lod_create_verbatim_rdf_conversions_tabular_data)>; (2) <<http://www.openlinksw.com/wiki/main/>>; (3) <<http://jena.apache.org/>>; (4) <<https://github.com/RDFLib>>; (5) <<https://github.com/RDFLib>>; (6) <<http://www.w3.org/TR/r2rml/>>.

#### **4.4.2 A viabilidade da etapa de enriquecimento**

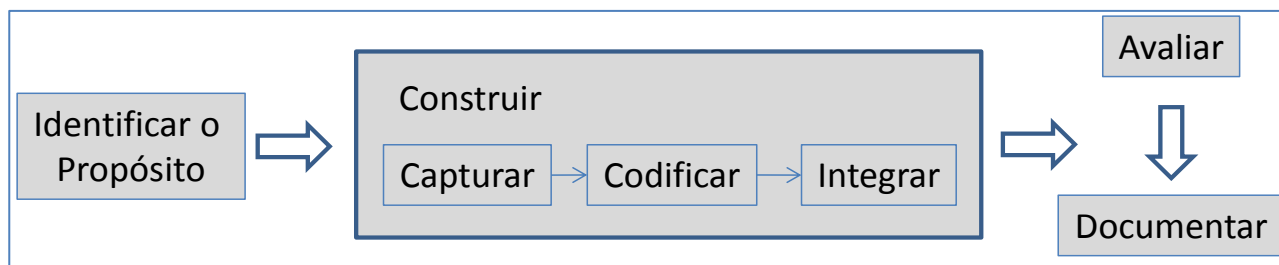
A etapa de enriquecimento visa transformar os dados que já estão em RDF, de forma que eles possam ser ligados com outros conjuntos de dados, por meio do uso de ontologias comuns. Uma das formas, segundo Ding *et al.* (2010) é fazer com que as propriedades usadas nos dados RDF sejam associadas a termos oriundos de ontologias bem conhecidas, tais como FOAF (citada no capítulo 2) e Dublin Core. Essa abordagem, ainda segundo os autores, permite que os conjuntos de dados possam ser usados por ferramentas da Web Semântica, como o Tabulator, que é um navegador construído para dados ligados. Esse tipo de ferramenta funciona do seguinte modo:

quando se navega em um conjunto de dados, ela irá procurar por valores associados a marcadores ‘rdfs:label’ (ou subpropriedades correspondentes), os quais ela exibirá no lugar das URIs, tornando assim os dados mais legíveis para os usuários humanos. Outra estratégia é ligar os conjuntos de dados à nuvem *Linking Open Data* (Ligando Dados Abertos).

Tanto para a etapa de conversão, quanto para a etapa de *enriquecimento e ligação dos dados orçamentários* (2), há a necessidade da definição de vocabulários comuns, por meio de ontologias. A literatura aponta várias metodologias para a construção de ontologias. Em Breitman (2006), há uma revisão de algumas delas: KACTUS, Methontology, Método 101, Cyc, Uschold, Gruninger & Fox.

Conforme mencionado, será apresentado um exercício de modelagem, o qual ilustrará alguns passos da metodologia proposta por Mike Uschold, da Universidade de Edimburgo. Essa metodologia foi escolhida por ser baseada em *cenários de motivação*, o qual é justamente o caso da informação orçamentária: há um cenário que motiva a adoção de ontologias que permitam a descrição dos dados orçamentários em RDF e a ligação destes com outras bases de interesse público que hoje estão dispersas e, às vezes, pouco acessíveis. A figura 34 dá uma visão geral do método.

Figura 34. Método de Uschold para a construção de ontologias (*skeletal methodology*)



Fonte: BREITMAN (2006).

Prosseguindo a ilustração baseada na metodologia, na primeira etapa, o *propósito* seria o descrito na tabela 9:

Tabela 9. Exercício de modelagem – Etapa ‘Identificar o Propósito’

“Descrever dados referentes à execução orçamentária da União em RDF, de modo a facilitar a construção de aplicações que permitam uma visão integrada entre a elaboração e a execução orçamentária pela sociedade e pelo próprio governo”.

Fonte: dados da pesquisa.

A segunda etapa, *capturar*, consiste em “definir textualmente conceitos e relacionamentos” (BREITMAN, 2006, p. 70). No caso do Orçamento Público, há um ponto de partida bastante útil: as categorias e os conceitos são comuns às diversas aplicações, pois são definidos por normas federais, que devem ser seguidas por todas elas. Assim, por exemplo, o termo ‘empenho’ é utilizado nos diversos sistemas para designar o mesmo conceito, o qual é associado à categoria na qual se

enquadra todos os empenhos específicos (indivíduos ou instâncias dessa categoria). O mesmo se dá em relação aos termos ‘convênio’, ‘programa’, ‘conveniente’, ‘concedente’, ‘unidade orçamentária’ etc. Isso significa que já existe um vocabulário implicitamente compartilhado entre as aplicações.

Continuando o exercício de modelagem, a partir dos conceitos descritos em 3.1, são definidas as seguintes relações (Tabela 10).

*Tabela 10. Exercício de modelagem – Etapa ‘Capturar’*

<b>Tipo de Relação</b>	<b>Descrição em Linguagem Natural (LN)</b>
<b>Relação entre ‘despesa’ e ‘empenho’</b>	‘Todo item de despesa é empenhado por uma nota de empenho’.
<b>Relação entre ‘despesa’, ‘empenho’ e ‘liquidação’</b>	‘Todo item de despesa é liquidado por uma nota de liquidação, desde que esse item já tenha sido empenhado’ (ou seja, desde que haja uma reserva orçamentária associada).
<b>Relação entre ‘despesa’, ‘pagamento’ e ‘liquidação’</b>	‘Todo item de despesa é realizado (pago) por uma ordem de pagamento de uma despesa, desde que esse item já tenha sido liquidado’ (ou seja, o pagamento da despesa já tenha sido autorizado).

Fonte: dados da pesquisa.

A próxima etapa (terceira), *codificar*, consiste na representação dos conceitos e relacionamentos definidos na etapa de captura, por meio de uma linguagem formal (BREITMAN, 2006). É necessário “decidir se cada termo será modelado como classe, entidade ou relacionamento” (p. 71).

Continuando o exercício proposto e utilizando os exemplos construídos anteriormente, ter-se-iam as classes, relacionamentos expressos na tabela 11 e os axiomas – em lógica de primeira ordem e em lógica descritiva – expressos na tabela 12.

*Tabela 11. Exercício de modelagem – Etapa ‘Codificar’ – Classes e Relacionamentos*

<b>Expressão</b>	<b>Nome OWL</b>	<b>Símbolo</b>	<b>Tipo</b>
<b>‘item de despesa’</b>	ItemDeDespesa	ID(x)	Classe
<b>‘nota de empenho’</b>	NotaDeEmpenho	NE(x)	Classe
<b>‘nota de liquidação’</b>	NotaDeLiquidacao	NL(x)	Classe
<b>‘ordem de pagamento’</b>	OrdemDePagamento	OP(x)	Classe
<b>‘é liquidado por’</b>	ehLiquidadoPor	lp(x,y)	Relacionamento/Propriedade
<b>‘é empenhado por’</b>	ehEmpenhadoPor	ep(x,y)	Relacionamento/Propriedade
<b>‘é realizado por’</b>	ehRealizadoPor	rp(x,y)	Relacionamento/Propriedade

Fonte: dados da pesquisa.



Tabela 12. Exercício de modelagem – Etapa ‘Codificar’ – Axiomas em LPO e DL

Axioma em Linguagem Natural	Axioma em Lógica de Primeira Ordem	Axioma em Lógica Descritiva
Todo item de despesa é empenhado por uma nota de empenho.	$\exists y(\forall x(ep(x, y) \rightarrow NE(y)))$ $\exists y(\forall x(ep(x, y) \rightarrow ID(x)))$	(1) $T \sqsubseteq \forall ep. NE$ (2) $T \sqsubseteq \forall ep. ID$
Todo item de despesa é liquidado por uma nota de liquidação, desde que esse item já tenha sido empenhado.	$\exists y(\forall x(lp(x, y) \rightarrow NL(y)))$ $\exists y(\forall x(lp(x, y) \rightarrow ID(x) \wedge \exists z(ep(x, z) \wedge NE(z))))$	(3) $T \sqsubseteq \forall lp. NL$ (4) $T \sqsubseteq \forall lp. ID \sqcap \exists ep. NE$
Todo item de despesa é realizado (pago) por uma ordem de pagamento de uma despesa, desde que esse item já tenha sido liquidado.	$\exists y(\forall x(rp(x, y) \rightarrow OP(y)))$ $\exists y(\forall x(rp(x, y) \rightarrow ID(x) \wedge \exists z(lp(x, z) \wedge NL(z))))$	(5) $T \sqsubseteq \forall rp. OP$ (6) $T \sqsubseteq \forall rp. ID \sqcap \exists lp. NL$

Fonte: dados da pesquisa.

Os axiomas da última coluna da tabela 12 indicam a imagem (*range*) e o domínio (*domain*) das propriedades – exemplos: axiomas (1) e (2), respectivamente. O símbolo  $T$  representa uma tautologia, e equivale à classe Thing da OWL. Para inferir os axiomas, foi necessário o uso de equivalências lógicas, omitidas por simplicidade. Para referências sobre lógica de primeira ordem, recomenda-se Mortari (2001). Para equivalências entre Lógica de Primeira Ordem (FOL) e Lógica Descritiva (DL), recomenda-se Grosz *et al.* (2003) e Baader *et al.* (2003). Os axiomas da última coluna podem ser convertidos para OWL, uma vez que esta linguagem é um tipo de lógica descritiva, conforme foi visto.

Também é importante ressaltar que existem ferramentas que podem auxiliar as etapas de captura, codificação, integração, validação e codificação. A principal delas é a plataforma Protégé<sup>28</sup>. Trata-se de uma ferramenta gratuita e de código aberto (*free and open source*) que, além de permitir a edição de ontologias, é considerada um arcabouço (*framework*) para representação do conhecimento. Além disso, ela tem o suporte de uma comunidade ativa de acadêmicos e profissionais (STANFORD, 2013).

Finda a etapa de codificação, pode-se iniciar a etapa de integração. No caso dos dados orçamentários, já existe uma ontologia, denominada Ontologia da Classificação da Despesa do Orçamento Federal, desenvolvida pela Secretaria de Orçamento e Finanças (SOF), e expressa em OWL. Com base nessa ontologia, foi gerado um conjunto de dados (*dataset*), referentes à base de dados da Lei Orçamentária Anual (LOA) de 2012. O *dataset*, descrito em RDF, contém a descrição de todos os itens de despesa que estavam programados para serem executados em 2012, com seus

<sup>28</sup><http://protege.stanford.edu/>.

valores de dotação inicial. Ele já está disponível no portal <<http://orcamento.dados.gov.br>>. O arquivo OWL contém a descrição dos metadados associados aos itens de despesa e como eles se relacionam. Existem ainda manuais de utilização disponíveis na Web (BRASIL, 2012).

A tabela 13 mostra as classes definidas nessa ontologia e as relaciona ao Manual Técnico do Orçamento de 2012. A figura 35 exibe um mapeamento entre a ontologia e alguns recursos expressos em RDF.

Hipoteticamente, poderia ser criada, a partir do exercício de modelagem proposto, uma ontologia OWL de Execução da Despesa compatível com a Ontologia de Classificação da Despesa do Orçamento Federal. A classe ItemDeDespesa (Item de Despesa) seria o elo entre ambas. A figura 36 ilustra essa ideia.

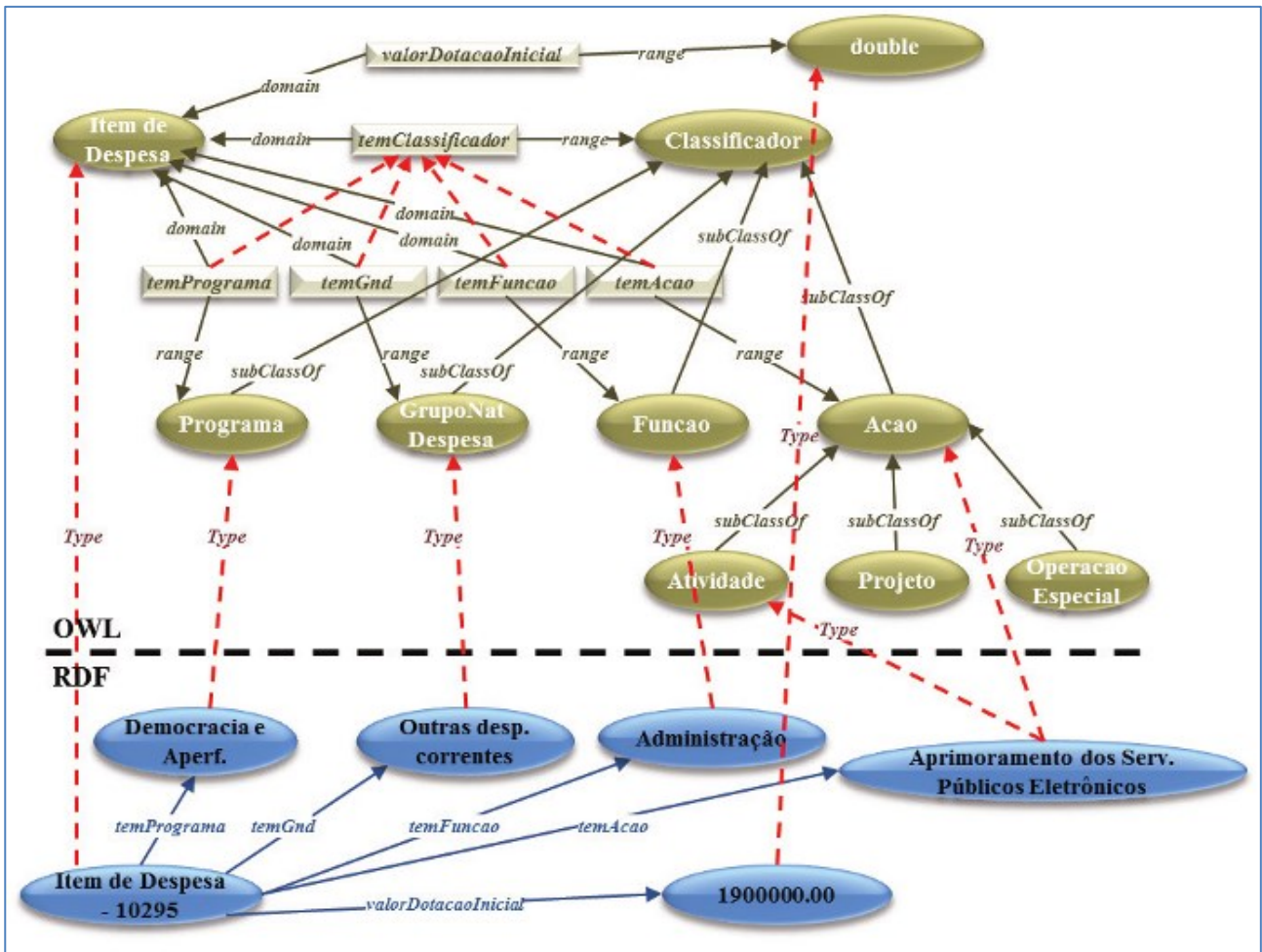
Como já foi dito, trata-se de uma ontologia simplificada. Mas ela indica que a etapa de enriquecimento no contexto dos dados do Orçamento Federal é viável, utilizando os métodos citados na literatura, como o Uschold, bem como as ontologias e tecnologias disponíveis. Uma ontologia mais sofisticada demandaria a execução da etapa de avaliação, além de novas iterações das etapas já mencionadas, as quais ampliariam o vocabulário, por meio de novas classes, propriedades e axiomas. Após sucessivas iterações, caso a ontologia fosse julgada suficientemente madura, ela poderia ser, finalmente, documentada.

Tabela 13. Ontologia da Classificação da Despesa do Orçamento Federal

Classes	MTO 2012
<b>CategoriaEconômica</b>	Categoria Econômica
<b>Classificador</b>	Sem correspondência. Classe de suporte ao conceito de classificação, das quais, todas as outras classes que se referem ao conceito de classificação da despesa se especializam. Tem como propriedades a descrição ( <i>label</i> ) e o código ( <i>codigo</i> ).
<b>ElementoDespesa</b>	Elemento de Despesa
<b>Esfera</b>	Esfera
<b>FonteRecursos</b>	Fonte de Recursos
<b>GrupoNatDespesa</b>	Grupo de Natureza da Despesa (GND)
<b>CategoriIndicadorUsoaEconômica</b>	Indicador de Uso
<b>ItemDeDespesa</b>	Sem correspondência. Cada elemento da classe ItemDeDespesa é caracterizado por ter propriedades ( <i>object properties</i> ) específicas para cada uma das subclasses da classe Classificadores, além de possuir um <i>data type property</i> denominado ValorDotaçãoInicial que é do tipo double.
<b>Localizador</b>	Subtítulo (Localizador do gasto)
<b>ModalidadeAplicacao</b>	Modalidade de Aplicação
<b>Programa</b>	Programa
<b>Acao</b>	Ação
<b>Projeto</b>	Projeto
<b>Atividade</b>	Atividade
<b>OperacaoEspecial</b>	Operação Especial
<b>Funcao</b>	Função
<b>Subfuncao</b>	Subfunção
<b>UnidadeOrcamentaria</b>	Unidade Orçamentária (UO)
<b>Orgao</b>	Órgão
<b>ResultadoPrimario</b>	Identificador de Resultado Primário

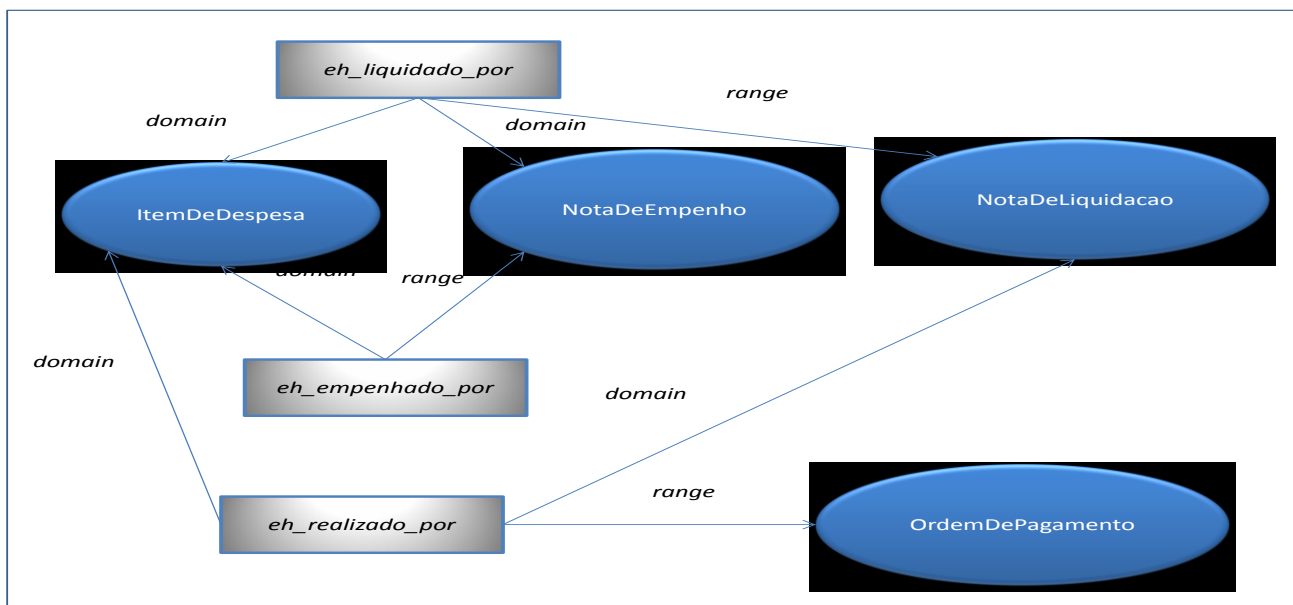
Fonte: <<http://vocab.e.gov.br/2012/08/loa2012#modelo>>.

Figura 35. Mapeamento entre a Ontologia de Classificação da Despesa do Orçamento Federal e um recurso



Fonte: <<http://vocab.e.gov.br/2012/08/loa2012#modelo>>.

Figura 36. Exercício de modelagem – Etapa Integrar



Fonte: elaborada pelo autor.

#### 4.4.3 A viabilidade da etapa de visualização

Após os dados estarem abertos, ligados e enriquecidos, é necessário mostrar o *valor dos dados orçamentários* usando aplicações interessantes de exploração e visualização.

Conforme visto nas aplicações Data.gov Wiki e Ligado nos Políticos, uma vez que os dados estão no padrão Dados Abertos Ligados, é possível lançar mão de várias tecnologias de busca, exploração e visualização de dados. O portal do Executivo Federal de Dados Abertos de Orçamento possui também formas de visualização dos dados. Uma delas, baseada em navegação, é exibida na figura 37.

Além disso, existem ainda outras ferramentas de visualização, como as baseadas nos chamados *mashups*: aplicações que combinam dados de múltiplas fontes e os apresentam por meio de novas interfaces (BIZER *et al.*, 2006). Um exemplo já mencionado neste trabalho são aplicações que possuem uma interface para navegação por coordenadas geográficas (ver figuras 30 e 33). A tabela 14 apresenta algumas das tecnologias que podem ser usadas para visualização de dados. A figura 38 exibe uma aplicação que ilustra a utilização dessas tecnologias.

Em relação à busca, a principal tecnologia é o SPARQL, descrita no capítulo 2. O portal citado também possui uma interface de busca SPARQL, que funciona sobre uma plataforma Open Link Virtuoso. Ela pode ser acessada pelo endereço: <<http://orcamento.dados.gov.br/sparql/>>.

*Tabela 14. Ferramentas para visualização de dados do tipo ‘mashup’*

<b>Tecnologia</b>	<b>Descrição</b>	<b>Endereço</b>
<b>Google Visualization API</b>	Permite acessar múltiplas fontes de dados estruturados, e exibi-los por diversas formas de visualizações.	<a href="http://code.google.com/apis/visualization/">http://code.google.com/apis/visualization/</a>
<b>Google Visualization GeoMap</b>	Extensão da Google Visualization API que permite exibir dados referenciados geograficamente	<a href="http://code.google.com/apis/visualization/documentation/gallery/geomap.html">http://code.google.com/apis/visualization/documentation/gallery/geomap.html</a>
<b>Google Visualization Pie Chart</b>	Extensão da Google Visualization API que permite construir gráficos estatísticos.	<a href="http://code.google.com/apis/visualization/documentation/gallery/piechart.html">http://code.google.com/apis/visualization/documentation/gallery/piechart.html</a>
<b>Simile Exhibit</b>	Permite criar páginas web com busca textual, filtragem, mapas interativos e linhas de tempo.	<a href="http://simile-widgets.org/exhibit/">http://simile-widgets.org/exhibit/</a>

Fonte: elaborado pelo autor a partir do portal Data-gov Wiki: <[http://data-gov.tw.rpi.edu/wiki/The\\_Data-gov\\_Wiki](http://data-gov.tw.rpi.edu/wiki/The_Data-gov_Wiki)>.

Figura 37. Tela de resultado de navegação do portal <orcamento.dados.gov.br>.

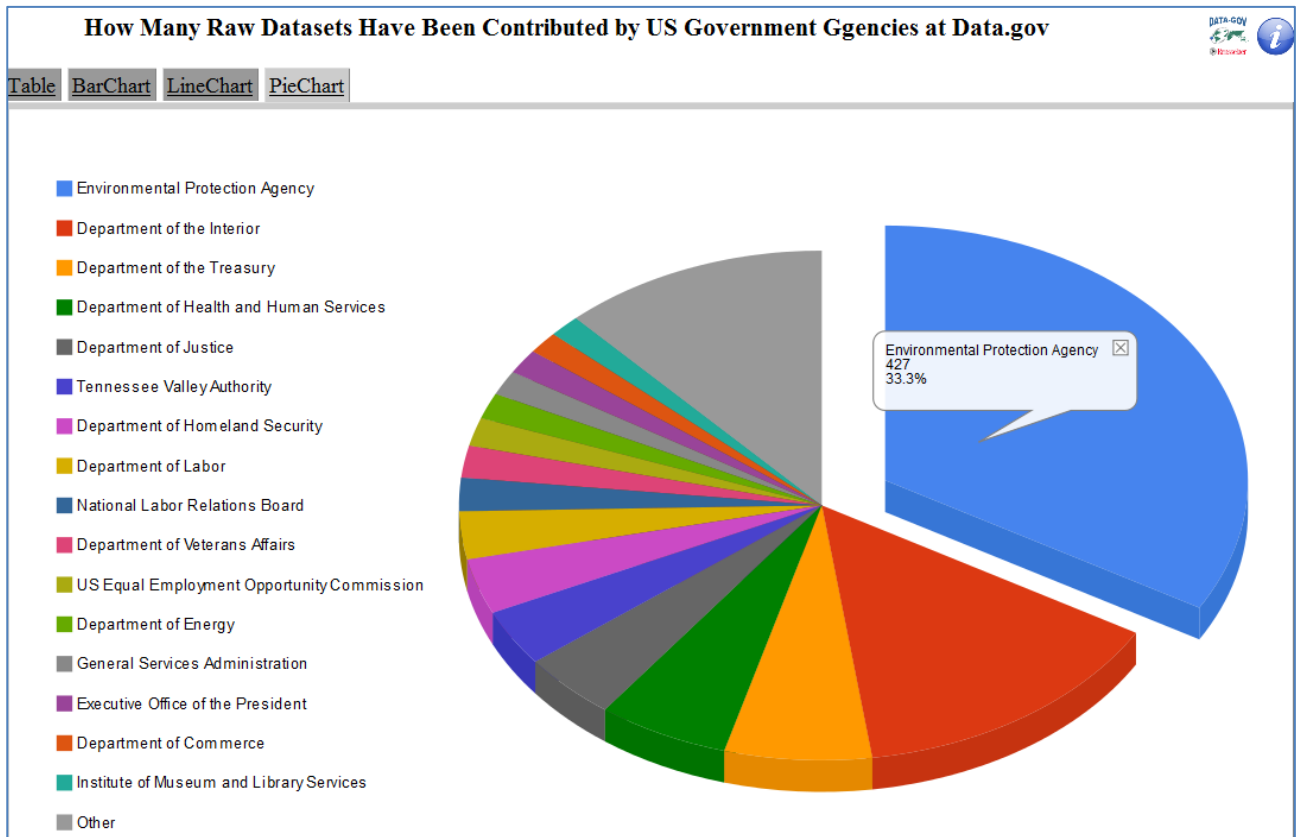
The screenshot displays the 'orcamento.dados.gov.br' website interface. At the top, there's a header with the site logo and navigation links for 'html', 'json', 'rdf', 'text', 'ttl', and 'xml'. Below the header, the main content area is divided into several sections:

- Resultados da Busca:** This section shows three search results. Each result includes a table with fields like 'loa\_valor dotacao inicial', 'loa\_tem acao', 'loa\_tem elemento despesa', 'loa\_tem esfera', 'loa\_tem funcao', 'loa\_tem localizador', and 'loa\_tem unidade orcamentaria'. The first result is for '475817400', the second for '753251222', and the third for '8197174'.
- Nesta Página:** A sidebar section with a list of links, including '> 1', '> 10', '> 100', '> 1000', '> 10000', '> 10001', '> 10002', '> 10003', '> 10004', '> 10005', and '> próxima >'. It also has a 'Ordenar por' section with a list of fields to sort by, such as 'loa\_valor dotacao inicial', 'loa\_tem acao', 'loa\_tem elemento despesa', 'loa\_tem esfera', 'loa\_tem funcao', 'loa\_tem localizador', and 'loa\_tem unidade orcamentaria'.
- Visualizar:** A section with a 'detalhado' link and a list of fields to visualize, including 'loa\_valor dotacao inicial', 'loa\_tem acao', 'loa\_tem elemento despesa', 'loa\_tem esfera', 'loa\_tem funcao', 'loa\_tem localizador', and 'loa\_tem unidade orcamentaria'.
- Itens por página:** A section with a dropdown menu showing '10', '25', and '50' items per page.

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

Há, portanto, uma gama considerável de tecnologias, que possibilitam a implementação de diversas formas de visualização e que já foram testadas em aplicações variadas, conforme as referências citadas e os exemplos descritos e ilustrados. Isso indica a viabilidade da implementação da etapa de visualização de dados abertos ligados pelo Congresso Nacional, provendo múltiplas formas de acesso aos dados orçamentários pelos cidadãos, empresas, organizações não-governamentais e pelo próprio governo. Na realidade, com o uso dessas tecnologias associadas às demais descritas ao longo desta seção, os dados orçamentários poderiam ser interligados com dados socioeconômicos, urbanísticos, demográficos e outros, possibilitando não apenas transparência, mas uma visão integrada do país aos gestores públicos, valiosa para o planejamento e acompanhamento das ações do Estado. Naturalmente chegar nesse ponto requer um esforço considerável, mas, a estratégia baseada em dados ligados é, até o presente momento, a que provê maior flexibilidade e interoperabilidade.

Figura 38. Exemplo ilustrativo de tecnologias do tipo ‘mashup’ para visualização de dados abertos.



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

#### 4.4.4 A viabilidade da etapa de proveniência do conhecimento sobre os dados

Assegurar a proveniência dos dados é importante, pois aumenta a credibilidade da informação gerada. É importante saber, por exemplo, qual é a base de origem e a data em que foi feita a extração dos dados. Ding *et al.* (2010) explicam que uma das estratégias é a utilização de metadados de proveniência. Explicam que no projeto Data-gov Wiki (descrito neste capítulo) eles lançaram mão de vários metadados, tais como: *título*, para o título do conjunto de dados; *agência* para a agência (órgão) governamental responsável pela publicação do conjunto de dados; *cobertura geográfica*, para indicar quais localidades geográficas foram mencionadas no conjunto de dados, *categoria*, para a classificação do conjunto de dados, e *data de alteração*, para quando o conjunto de dados foi modificado; lista de variáveis do dicionário de dados, para mostrar como obter a descrição detalhada dos metadados do conteúdo do conjunto de dados, e ponto de acesso do CSV/TXT, para definir as URLs a partir das quais os usuários podem recuperar as versões do conjunto de dados baseadas em CSV ou texto.

A implementação desses metadados pode ser trabalhosa, mas é plenamente viável. Para os conjuntos de dados que ainda não estão disponíveis diretamente em CSV, como o Siafi Gerencial,

duas estratégias podem ser adotadas: solicitar ao órgão gestor a geração e a disponibilização do conjunto de dados convertido, ou solicitar apenas a licença para a publicação, uma vez que, como foi mencionado, tanto a Câmara quanto o Senado já possuem cópias desses conjuntos de dados, porém em formato relacional e proprietário.

Na verdade, mesmo que, em uma etapa inicial, os dados ainda não estivessem convertidos para RDF, o registro da proveniência dos dados e o acesso aos dados originais seriam essenciais, uma vez que correspondem a um requisito da Lei de Acesso à Informação: mais precisamente, ao inc. IV do art. 3º: “garantir a autenticidade e a integridade das informações disponíveis para acesso”. Trata-se, portanto, de uma etapa necessária e viável.

#### 4.5 *Resumo do Capítulo*

Como conclusão deste capítulo, a primeira questão a ser destacada é o não-atendimento pleno aos requisitos da Lei de Acesso à Informação (LAI) pelos sistemas Siga Brasil e Fiscalize, no que tange às informações orçamentárias, conforme seção 4.2. Tal fato confirma o pressuposto (P1) e leva a uma tentativa de solução, que passa, necessariamente, pela adoção do padrão *dados abertos* de forma mais ampla pelo Congresso Nacional. A segunda questão envolve a adoção do padrão ‘dados abertos ligados’ (LOD), o qual, como se viu, é uma evolução do padrão anteriormente citado. É fato que a LAI não determina, explicitamente, a adoção do padrão LOD. No entanto, ela também não o restringe; ao contrário, estimula o uso de todos os instrumentos disponíveis, como se infere a partir de uma leitura mais atenta do parágrafo 2º do artigo 8º: “Para cumprimento do disposto no *caput*, os órgãos e entidades públicas deverão utilizar todos os meios e instrumentos legítimos de que dispuserem”. Além disso, os argumentos mostrados neste trabalho, especialmente nas seções 4.3 e 4.4, indicam a viabilidade da sua adoção, pelos motivos expostos, os quais podem ser sintetizados em: já existem conjuntos de dados em formatos abertos, que podem ser convertidos com relativa facilidade; o Congresso Nacional possui acesso ou gerencia importantes bases de dados orçamentários em formato relacional, as quais também podem ser convertidas; além disso, há conjuntos de dados em formatos abertos e ligados, que podem se somar às demais; o Congresso Nacional, como ator imprescindível em várias fases do processo orçamentário, pode e deve contribuir, tanto na questão da modelagem quanto no fornecimento de dados; finalmente, há tecnologias disponíveis, já testadas com sucesso em outros projetos. Dessa forma, o pressuposto (P2), relativo à adoção do padrão LOD pelo Congresso Nacional no que tange ao Orçamento também é confirmado, com a ressalva de que a adoção deve ser paulatina, trabalhando cada base de dados de acordo com a sua disponibilidade e com a capacidade dos setores envolvidos.



## 5. CONSIDERAÇÕES FINAIS

Nas democracias contemporâneas, a transparência é palavra de ordem. No entanto, a transparência não pode ser meramente formal: ela deve ser efetiva; ou seja, deve ser materializada por meio de ações concretas, alcançando o maior público possível.

As Novas Tecnologias da Informação e Comunicação não podem ser ignoradas nesse processo. A Web, em particular, é a principal plataforma tecnológica para a efetivação da transparência, uma vez que, além do seu longo alcance, possibilita a interatividade governo-sociedade.

A Lei de Acesso à Informação – LAI, ao mesmo tempo em que vem ao encontro dos anseios da sociedade por transparência, está em sintonia com as novas tecnologias, pois indica, claramente, a necessidade da exposição dos dados de forma aberta e acessível por máquinas.

No entanto, a própria Web vem evoluindo drasticamente: a Web Semântica, que poderia ser considerada apenas uma visão há alguns anos, possui, atualmente, um real potencial de implementação. Conforme visto ao longo deste texto, ela é um resultado de anos de estudos, com forte embasamento teórico e possui, como aliada, uma vasta gama de tecnologias que lhe dão suporte. Os órgãos públicos não deveriam ignorar tal fato. Na verdade alguns deles, ainda que de forma incipiente se comparados a outros países, vem tomando iniciativas no sentido de adotar padrões e tecnologias da Web Semântica para a concretização da transparência, por meio dos chamados Dados Abertos Ligados. Uma dessas iniciativas é a criação do portal de Dados Abertos do Orçamento Federal: <<http://orcamento.dados.gov.br>>. Trata-se de iniciativa recente, conduzida pelo Ministério do Planejamento. Por isso mesmo, contém apenas os dados da elaboração orçamentária de 2012.

O Congresso Nacional, como um dos atores institucionais responsáveis pela elaboração e fiscalização da execução orçamentária, não poderia ficar alijado desse processo. Contudo, apesar do avanço na abertura dos dados, tanto a Câmara quanto o Senado não possuem ainda iniciativa de peso na adoção dos Dados Abertos Ligados no que tange à área orçamentária. Na verdade, até o presente momento, nenhuma das Casas Legislativas tornou disponível nenhum conjunto de dados no padrão Dados Abertos Ligados. Ademais, conforme demonstrado, os dois principais sistemas de informações orçamentárias baseados na Web no âmbito do Congresso Nacional – Fiscalize e Siga Brasil – não satisfazem plenamente os requisitos da Lei de Acesso à Informação.

A revisão de literatura empreendida e os estudos de caso descritos indicam a viabilidade da adoção ampla do padrão Dados Abertos Ligados no contexto do Orçamento Público. Naturalmente

existe uma questão operacional, que envolve aspectos gerenciais e tecnológicos. Os aspectos gerenciais poderiam contar com o suporte da Comunidade TI Controle<sup>29</sup>, que já apoia importantes projetos, como o LexML<sup>30</sup>. Essa Comunidade coordenaria as etapas necessárias para a ampliação e aprimoramento dos dados abertos ligados de orçamento, desde a identificação das fontes de dados até a construção e integração de ontologias e esquemas. A execução das etapas poderia ser dividida entre os diversos órgãos envolvidos, de acordo com sua capacidade técnica e operacional.

Reconhece-se que algumas etapas, como a construção de ontologias, não são triviais, pois necessitam de conhecimento altamente especializado. Porém, tanto a Câmara dos Deputados quanto o Senado possuem um corpo técnico qualificado na área orçamentária, por meio das suas respectivas consultorias, bem como nas áreas de Ciência da Computação e Biblioteconomia. Além disso, poderiam ser firmados convênios com universidades e institutos de pesquisa, numa interação benéfica a todos os envolvidos. Até porque a elaboração e o controle do Orçamento Público, pela sua natureza, envolvem múltiplos atores. Portanto, a construção de sistemas de informação orçamentários deveria seguir a mesma lógica.

O estudo apresentado neste texto, embora preliminar, indica que a aplicação do padrão Dados Abertos Ligados nas informações relativas ao Orçamento Público seria proveitosa tanto para o Congresso Nacional e demais atores envolvidos – por reutilizar uma gama de técnicas, tecnologias e ferramentas maduras, flexíveis e poderosas, quanto para a sociedade em geral, que ganharia poder de (re)elaborar a informação pública de acordo com suas necessidades. E isso seria um importante passo no cumprimento dos requisitos legais e institucionais de transparência e controle social.

---

<sup>29</sup> Comunidade de Tecnologia da Informação Aplicada ao Controle (TIControl), reúne representantes dos Poderes Legislativo, Executivo, Judiciário, do Ministério Público e da Advocacia-Geral da União. Foi criada em 2006 e tem por finalidade contribuir para o incremento da eficiência, eficácia e efetividade na gestão pública. Fonte: <<http://www.ticontrol.gov.br>>.

<sup>30</sup> O LexML é uma rede de informação legislativa e jurídica que pretende organizar, integrar e dar acesso às informações disponibilizadas nos diversos portais de órgãos do governo na Internet. Fonte: <<http://lexml.gov.br>>.

## REFERÊNCIAS

ALANI, H.; DUPPLAW, D.; SHERIDAN, J.; O'HARA, K., DARLINGTON, J.; SHADBOLT, N. e TULLO, C. Unlocking the Potential of Public Sector Information with Semantic Web Technology. In: **ISWC – International Semantic Web Conference, 2007**.

ALLEMANG, Dean; HENDLER, Jim. **Semantic Web for the Working Ontologist: effective modeling in RDFS and OWL**. 2nd. ed. Waltham: Morgan Kaufmann, 2011.

ALMEIDA, Maurício Barcellos; SOUZA, Renato Rocha. Avaliação do espectro semântico de instrumentos para organização da informação. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação** [1518-2924], v. 16, iss. 31, 2011.

ARAÚJO, Lucas de Ramos; SOUZA, Jair Francisco de. Aumentando a Transparência do Governo por Meio da Transformação de Dados Governamentais Abertos em Dados Ligados. **Revista Eletrônica de Sistemas de Informação**, v. 10, n. 1, artigo 7, 2011. Disponível em: <<http://revistas.facecla.com.br/index.php/reinfo/article/download/880/pdf>>. Acesso em: maio de 2012.

BAADER, Franz; HORROCKS, Ian; SATTler, Ulrike. Description Logics as Ontology Languages for the Semantic Web. In: **Festschrift in honor of Jörg Siekmann, Lecture Notes in Artificial Intelligence**. Berlin: Springer Verlag, 2003, p. 228-248.

BERNERS-LEE, Tim. **Semantic web road map**. [S.l.]: W3C, 1998. Disponível em: <<http://w3.org/DesignIssues/Semantic.html>>. Acesso em: maio de 2012.

BERNERS-LEE, Tim. **Linked data**. [S.l.]: W3C, 2009. Disponível em: <<http://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: maio de 2012.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**, New York, maio 2001. Disponível em: <<http://trac.assembla.com/soray/export/230/user/Marcell/readings/Semantics/The%20Semantic%20Web.pdf>>. Acesso em: maio de 2012.

BEZERRA FILHO, João Eudes. **Contabilidade Pública: teoria, técnica de elaboração de balanços e 500 questões**. 2. ed. Rio de Janeiro: Elsevier, 2006.

BIZER, C.; CYGANIAK, R.; HEATH, T. **How to publish linked data on the web**. [S.l.]:[s:n], 2007. Disponível em: <<http://www4.wiwiw.fu-berlin.de/bizer/pub/LinkedDataTutorial>>. Acesso em: maio de 2012.

BRÄSCHER, Marisa. **Tesauros, Taxonomias e Ontologias**. IV Encontro de Arquivos e Bibliotecas da Administração Pública Federal. Brasília: 2008.

BRASIL. **Lei Complementar no. 101, de 4 de maio de 2000**. *Estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências*. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/Leis/LCP/Lcp101.htm](http://www.planalto.gov.br/ccivil_03/Leis/LCP/Lcp101.htm)>. Acesso em: 08 maio 2013.

BRASIL. **Lei Complementar nº 131, de 27 de Maio de 2009**. *Acrescenta dispositivos à Lei Complementar nº 101, de 4 de maio de 2000, que estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências, 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*. Texto atualizado disponível em: <[http://www.planalto.gov.br/ccivil\\_03/Leis/LCP/Lcp131.htm](http://www.planalto.gov.br/ccivil_03/Leis/LCP/Lcp131.htm)>. Acesso em: 08 maio 2013.

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.*

BRASIL. MPOG. **Orçamento Federal em Formato Aberto: manual de referência.** 2012. Disponível em: <[https://www.siop.planejamento.gov.br/rdf/manual\\_de\\_referencia\\_2012.pdf](https://www.siop.planejamento.gov.br/rdf/manual_de_referencia_2012.pdf)>. Acesso em: janeiro 2013.

BRASIL. MPOG. **O Sistema De Gestão Das Transferências Voluntárias Da União – Siconv Como Ferramenta Para Captação De Recursos Federais Pelos Municípios.** 2013. Disponível em: <[https://www.convenios.gov.br/portal/avisos/CARTILHA\\_SICONV\\_PARA\\_MUNICIPIOS\\_-\\_Jan\\_13-1.pdf](https://www.convenios.gov.br/portal/avisos/CARTILHA_SICONV_PARA_MUNICIPIOS_-_Jan_13-1.pdf)>. Acesso em: 02 abril 2013.

BRASIL. SENADO. **Portal de Trabalho Siga Brasil.** Brasília: Senado, 2007. Disponível em: <[http://www9.senado.gov.br/dav\\_portal/portal/orcamento\\_senado/arquivo/Outros/SigaBrasil/390-Introducao\\_ao\\_SIGA\\_BRASIL.pdf](http://www9.senado.gov.br/dav_portal/portal/orcamento_senado/arquivo/Outros/SigaBrasil/390-Introducao_ao_SIGA_BRASIL.pdf)>. Acesso em: 27 mar. 2013.

BRASIL. SENADO. Sistema Siga Brasil. Disponível em: <[http://www9.senado.gov.br/portal/page/portal/orcamento\\_senado](http://www9.senado.gov.br/portal/page/portal/orcamento_senado)>. Acesso em: 27 mar. 2013.

CARLOS, Fábio Alves; LOPES, Jorge Expedito de Gusmão; PEDERNEIRAS, Marcleide Maria Macedo; MACÊDO, João Marcelo Alves; AMARO, Rodrigo Gayger. Uma discussão sobre a Criação de Indicadores de Transparência na Gestão Pública como Suporte ao Ciclo da Política Pública. **Revista de Contabilidade do Mestrado em Ciências Contábeis da UERJ**, Rio de Janeiro, v.13, n.2, maio/ago, 2008.

CHAUÍ, Marilena. **Convite à Filosofia.** 3. ed. São Paulo: Ática, 1995.

CHEVITARESE ALVES, Marcus Vinícius. **Transparência orçamentária efetiva e Internet : um estudo a partir dos portais eletrônicos de orçamento da Câmara dos Deputados e do Senado Federal.** Monografia (especialização) - Curso em Instituições e Processos Políticos do Legislativo, Câmara dos Deputados, Centro de Formação, Treinamento e Aperfeiçoamento (Cefor), 2009. Disponível em: <<http://bd.camara.gov.br/bd/handle/bdcamara/5644>>. Acesso em: 27 mar. 2013.

CONCEPCIÓN, Luzmila Pró. Fundamentos de Ingeniería de la Web: Ontologías, Web Semántica y Agentes de Software. **Revista de Investigación de Sistemas e Informática** 7(1), 77 - 89 (2010). ISSN 1816-3823 (versión electrónica).

DIAS, Tatiane Domingos; SANTOS, Neide. **Web Semântica: conceitos básicos e tecnologias associadas.** Rio de Janeiro: UERJ, 2001.

DING, Li; DIFRANZO, Dominic; GRAVES, Alvaro; MICHAELIS, James R.; LI, Xian; MCGUINNESS, Deborah L.; HENDLER, Jim. **Data-gov Wiki: Towards Linking Government Data.** [S.l.]: Association for the Advancement of Artificial Intelligence, 2010. Disponível em: <<http://data-gov.tw.rpi.edu/2010/linkedai-2010-datagov.pdf>>. Acesso em: maio 2012.

EAVES, David. **The three laws of open government data.** [S.l.]:[s.n], 2009. Disponível em: <<http://eaves.ca/2009/09/30/three-law-of-open-government-data/>>. Acesso em: maio 2012.

EC (European Commission). Interoperability Solutions for European Public Administrations (ISA). **European Interoperability Framework (EIF) for European public services.** Bruxelles: EC, 2010.

FEIGENBAUM, Lee; PRUD'HOMMEAUX, Eric. **SPARQL By Example: a tutorial.** [S.l.]: Cambridge Semantics, 2011. Disponível em: <<http://www.cambridgesemantics.com/pt/semantic-university/sparql-by-example>>. Acesso em 12 mar. 2013.

FREITAS, Frederico Luiz Gonçalves de. Ontologias e a web semântica. **Jornada de Minicursos em Inteligência Artificial**, SBC, v. 8, 2003.

GROSOFF, Ian Horrocks; VOLZ, Raphael; DECKER, Stefan. **Description Logic Programs: Combining Logic Programs with Description Logic**. Working Paper 4437-03, February 2003. [S.l.]: MIT Sloan School of Management, 2003. Disponível em: <[http://papers.ssrn.com/sol3/Delivery.cfm/SSRN\\_ID460986\\_code285952.pdf?abstractid=460986&mirid=1](http://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID460986_code285952.pdf?abstractid=460986&mirid=1)>. Acesso em: 12 mar 2013.

GUARINO, Nicola; OBERLE, Daniel; STAAB, Steffen. What Is an Ontology? In: STAAB, Steffen; Studer, R. (eds.). **Handbook on Ontologies**. Berlin: Springer-Verlag, 2009.

HARTIG, Olaf; BIZER, Christian; FREYTAG, Johann-Christoph. Executing SPARQL Queries over the Web of Linked Data. In: 8th International Semantic Web Conference (ISWC), Washington, DC, USA, Oct. 2009. **Proceedings...** [Versão preimpressa]. Disponível em: <[http://www2.informatik.hu-berlin.de/~hartig/files/HartigEtAl\\_QueryTheWeb\\_ISWC09\\_Preprint.pdf](http://www2.informatik.hu-berlin.de/~hartig/files/HartigEtAl_QueryTheWeb_ISWC09_Preprint.pdf)>. Acesso em: 12 mar. 2013.

HEATH, Tom; BIZER, Christian. Linked Data: Evolving the Web into a Global Data Space. In: **Synthesis Lectures on the Semantic Web: Theory and Technology**, 1:1, 1-136. [S.l.]: Morgan & Claypool: 2011. [Versão online]. Disponível em: <<http://linkdatabook.com/editions/1.0/#htoc6>>. Acesso em: 27 mar. 2013.

HEBELER, John; FISHER, Matthew; BLACE, Ryan; PEREZ-LOPES, Andrew. **Semantic Web Programming**. Indianapolis: Wiley, 2009.

KLISCHEWSKI, Ralf; TAGAMOA, Al; KHAMES, Al. Migrating small governments' websites to the semantic web. In: The Semantic Web meets eGovernment, AAAI 2006, Stanford University, California. **Proceedings...** Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=2D23FBC733BA62F5639DB7332D9AC1CF?doi=10.1.1.106.8669&rep=rep1&type=pdf>>. Acesso em: maio 2012.

KRAUSE, Décio. **Tópicos em Ontologia Analítica**. Florianópolis: UFSC, 2011. Disponível em: <<http://www.cfh.ufsc.br/~dkrause/pg/TopicosOntologia.pdf>>. Acesso em: 14 fev. 2013.

MORAIS, Erikson Freitas de; SOARES, Marcelo Borghetti. **Web Semântica para Máquinas de Busca**. [S.l.]: UFMG, 2004. Disponível em: <<http://www.dcc.ufmg.br/~nivio/cursos/pa03/seminarios/seminario7/seminario7.pdf>>. Acesso em: fev. 2013.

MORTARI, Cezar A. **Introdução à lógica**. São Paulo: Unesp, 2001.

NASCIMENTO, Cláudio. **Acompanhamento da execução orçamentária**. Rio de Janeiro: IBAM, 2001. (Lei de Responsabilidade Fiscal, 5). Patrocínio: BNDES.

PEDRAZA-JIMÉNEZ, Rafael; CODINA, Lluís; ROVIRA; Cristòfol. Sistemas de Información y Metadatos em la Web Semántica. In: CODINA, Lluís; MARCOS, Mari-Carmen; PEDRAZA, Rafael (coords.). **Web Semántica y Sistemas de Información Documental**. Gijón. España: Trea, 2011.

RUSSEL, Stuart; NORVIG, Peter. **Inteligência Artificial**. 2. ed. Rio de Janeiro: Campus, 2004.

SANTAELLA, Lúcia. **O que é Semiótica?** São Paulo: Brasiliense, 2012.

SOWA. John F. (2011a). Controlled Natural Languages For Representing Ontology. In: 4º Seminário de Pesquisa em Ontologias no Brasil (Ontobras), Porto Alegre, UFRGS, 2011. **Anais...** Porto Alegre: UFRGS, 2011. Disponível em: <<http://www.inf.ufrgs.br/ontobras-most2011/arquivos/curso1/OntobrasMostSlidesMinicurso1Sowa.pdf>>. Acesso em: 10 fev. 2013.

SOWA, John F. (2011b). The Role of Logic and Ontology In Language and Reasoning. In: POLI, R.; SEIBT, J. **Theory and Applications of Ontology: Philosophical Perspectives**. Berlin: Springer, 2011. [Versão revisada]. Disponível em: <<http://www.jfsowa.com/pubs/rolelog.pdf>>. Acesso em: 10 fev. 2013.

VIEIRA, Renata ; SANTOS, Débora Abdalla dos ; SILVA, Douglas Michaelson da ; SANTANA, Menandro Ribeiro . Web semântica: ontologias, lógica de descrição e inferências. In: Cesar Teixeira; Eduardo Barrere; Iran Abraão. (Org.). **Web e Multimídia: Desafios e Soluções (WebMedia 2005 - Minicursos)**. 1ed.Porto Alegre: SBC, 2005, v. 1, cap. 5, p. 127-167. Capítulo disponível em: <<http://www.inf.pucrs.br/~rvieira/cursos/webmidia.pdf>>. Acesso em: 10 fev. 2013.

W3C. **RDF Primer**. (2004a). W3C Recommendation 10 February 2004. Disponível em: <<http://www.w3.org/TR/rdf-primer/>>. Acesso 09 fev. 2013.

\_\_\_\_\_. **RDF Semantics**. (2004b). W3C Recommendation 10 February 2004. Disponível em: <<http://www.w3.org/TR/rdf-mt/>>. Acesso 09 fev. 2013.

\_\_\_\_\_. **RDF Vocabulary Description Language 1.0: RDF Schema**. (2004c). *W3C Recommendation 10 February 2004*. Disponível em: <<http://www.w3.org/TR/rdf-schema/>>. Acesso em: 10 fev. 2013.

\_\_\_\_\_. **OWL Web Ontology Language Overview**. (2004d). Disponível em: <<http://www.w3.org/TR/2004/REC-owl-features-20040210/>>. Acesso em: 09 fev. 2013.

\_\_\_\_\_. **OWL 2 Web Ontology Language Document Overview (Second Edition)**. (2012a). Disponível em: <<http://www.w3.org/TR/2012/REC-owl2-overview-20121211/>>. Acesso em: 10 fev. 2013.

\_\_\_\_\_. **OWL 2 Web Ontology Language Direct Semantics (Second Edition)**. (2012b). Disponível em: <[http://www.w3.org/TR/2012/REC-owl2-direct-semantics-20121211/#Class\\_Expression\\_Axioms](http://www.w3.org/TR/2012/REC-owl2-direct-semantics-20121211/#Class_Expression_Axioms)>. Acesso em: 10 fev. 2013.

\_\_\_\_\_. **OWL 2 Web Ontology Language Mapping to RDF Graphs (Second Edition)**. (2012c). Disponível em: <[http://www.w3.org/TR/2012/REC-owl2-direct-semantics-20121211/#Class\\_Expression\\_Axioms](http://www.w3.org/TR/2012/REC-owl2-direct-semantics-20121211/#Class_Expression_Axioms)>. Acesso em: 10 fev. 2013.

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., Porto Alegre, Rio Grande do Sul. **Anais...** Setembro, 2010. Disponível em: <[http://www.procergs.rs.gov.br/uploads/1285856001W3C\\_Seminario\\_Inovacao\\_eGov\\_POA\\_17092010.pdf](http://www.procergs.rs.gov.br/uploads/1285856001W3C_Seminario_Inovacao_eGov_POA_17092010.pdf)>. Acesso em: maio de 2012.