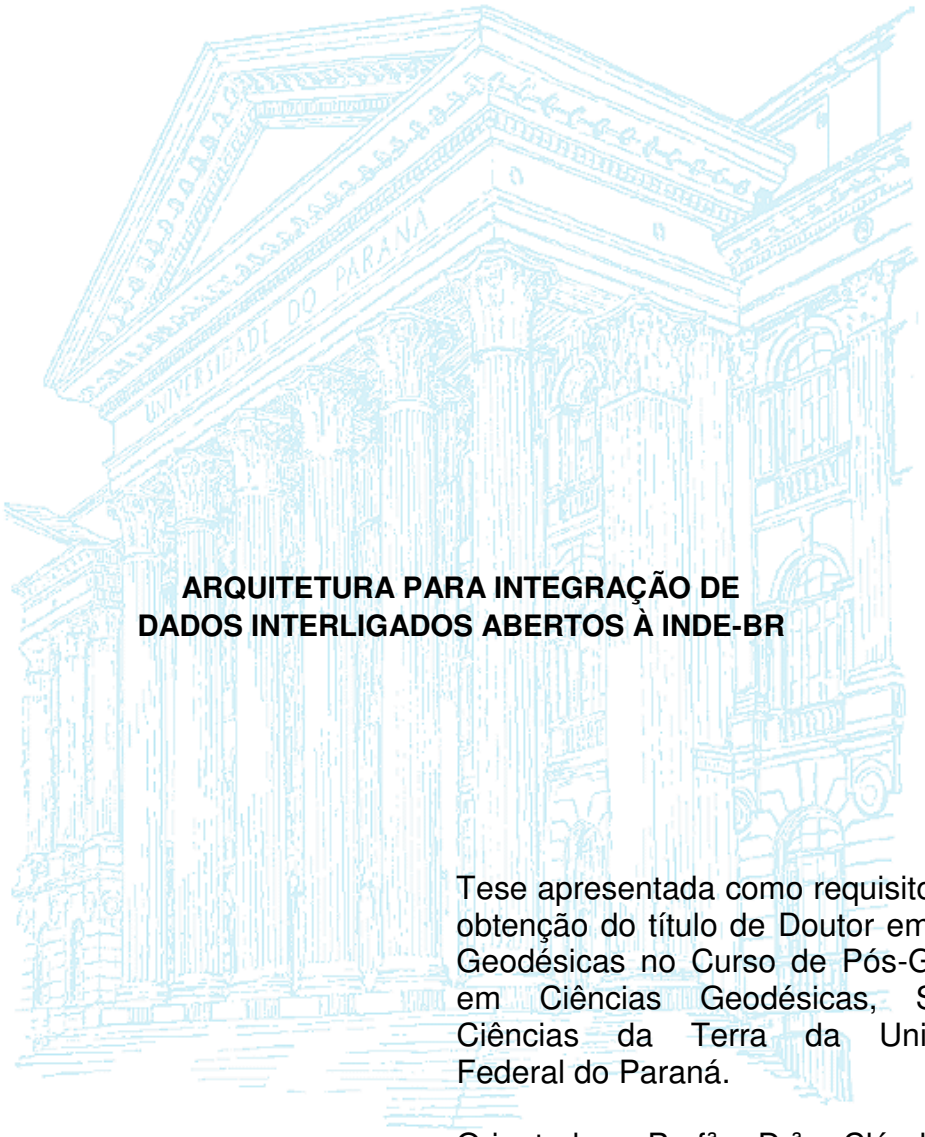
UNIVERSIDADE FEDERAL DO PARANÁ

SILVANA PHILIPPI CAMBOIM

ARQUITETURA PARA INTEGRAÇÃO DE DADOS INTERLIGADOS ABERTOS À INDE-BR

CURITIBA
2013

SILVANA PHILIPPI CAMBOIM



ARQUITETURA PARA INTEGRAÇÃO DE DADOS INTERLIGADOS ABERTOS À INDE-BR

Tese apresentada como requisito parcial à obtenção do título de Doutor em Ciências Geodésicas no Curso de Pós-Graduação em Ciências Geodésicas, Setor de Ciências da Terra da Universidade Federal do Paraná.

Orientador: Profª. Drª. Cláudia Robbi Sluter

CURITIBA
2013

C176a Camboim, Silvana Philippi
Arquitetura para integração de dados interligados abertos à INDE-BR. [manuscrito] / Silvana Philippi Camboim. – Curitiba, 2013.
140f. : il. [algumas color.] ; 30 cm.

Tese (doutorado) - Universidade Federal do Paraná, Setor de Ciências da Terra, Programa de Pós-graduação em Ciências Geodésicas, 2013.

Orientadora: Cláudia Robbi Sluter.

1. Pesquisa Geodésica. 2. Bando de Dados (INDE-BR). I. Universidade Federal do Paraná. II.Sluter, Cláudia Robbi. III. Título.

CDD: 526.3

TERMO DE APROVAÇÃO

SILVANA PHILIPPI CAMBOIM

"ARQUITETURA PARA INTEGRAÇÃO DE DADOS DA INDE-BR COM DADOS INTERLIGADOS ATRAVÉS DE SERVIÇO WEB"

Tese nº 86 aprovada como requisito parcial do grau de Doutor no Curso de Pós-Graduação em Ciências Geodésicas, Setor de Ciências da Terra da Universidade Federal do Paraná, pela seguinte banca examinadora:

Orientadora: Profª. Drª. Claudia Robbi Sluter
Departamento de Geomática, UFPR

Profª. Drª. Andrea Lopes Iescheck
Universidade Federal do Rio Grande do Sul, UFRGS

Profª. Drª. Julia Célia Mercedes Strauch
Escola Nacional de Ciências Estatísticas, ENCE

Prof. Dr. Henrique Firkowski
Departamento de Geomática, UFPR

Profª. Drª. Maria Cecília Bonato Brandalize
Departamento de Geomática, UFPR

Profª. Drª. Luciene Stamato Delazari
Departamento de Geomática, UFPR

Curitiba, 28 de fevereiro de 2013.

AGRADECIMENTOS

À minha orientadora, Profa. Dra. Claudia Robbi Sluter pela dedicação, paciência e sabedoria empenhados neste projeto.

Ao Curso de Pós-graduação em Ciências Geodésicas e seu Colegiado pela compreensão nos momentos de dificuldades.

Aos colegas discentes no curso, docentes na Universidade Federal do Paraná, funcionários do Instituto Brasileiro de Geografia e Estatística e integrantes da Comissão Nacional de Cartografia pelas contribuições valiosas e pelo companheirismo.

Ao analista de sistemas Anderson Carneiro pela colaboração no desenvolvimento do aplicativo.

E, finalmente, a todos que contribuíram direta ou indiretamente com a realização desta tese.

EPÍGRAFE

“Se eu vi mais longe, foi por estar de pé sobre ombros de gigantes”.

Isaac Newton

RESUMO

A crescente disponibilidade de dados da *web* gerou a demanda para que esta seja tratada como uma única base interligada de conhecimento a ser explorada. Adicionalmente, os governos, como grandes produtores de informação essencial, também estão sendo exigidos para que a publique de forma aberta, transparente e processável por máquina. Tais necessidades levaram a criação de um novo paradigma, dos dados interligados abertos, efetivada no governo brasileiro através de iniciativas como a Infraestrutura Nacional de Dados Abertos (INDA) e do portal dados.gov.br. O conhecimento do território surge neste contexto como forte componente na integração de informação de diversas naturezas. A comunidade de informação geoespacial tem uma história de busca por padrões e ações de interoperabilidade desde a década de 1990 já percebendo a necessidade de se superar obstáculos de conflitos sintáticos e semânticos. A arquitetura orientada a serviços vem embasando estes esforços, que culminaram com a criação das Infraestruturas de Dados Espaciais (IDEs), nas quais padrões, políticas públicas, tecnologia e formação de recursos são coordenados em busca da melhoria do compartilhamento e uso da informação geoespacial. No entanto, os dados publicados na Infraestrutura Nacional de Dados Espaciais (INDE-BR) através de serviços *web* padronizados não consegue ser acessada diretamente no formato orientado a recursos dos dados interligados. Para isso é necessário um conjunto de medidas que englobam a criação e manutenção de ontologias, a disponibilização de formas de publicação compatíveis e a ligação de bases de dados de diversas fontes Neste trabalho, foi proposta uma arquitetura para esta integração, que foi demonstrada através de um caso de uso, definindo ontologias para a divisão político-administrativa brasileira e com o desenvolvimento de um serviço *web* mediador que transforma dados diretamente de um serviço WFS compatível com a INDE-BR para o formato RDF, este por sua vez compatível com a INDA.

Palavras-chaves: Infraestrutura Nacional de Dados Espacias, Dados Interligados Abertos, Ontologias

ABSTRACT

The increasing availability of data on the web generated demand for it to be used as a unified interconnected knowledge database. Additionally, governments, as major producers of essential information, are also required to publish it in an open, transparent and machine processable way. These needs led to the creation of a new paradigm, the linked open data, carried out in the Brazilian government through initiatives such as the Infraestrutura Nacional de Dados Abertos (INDA) and the portal dados.gov.br. The knowledge about the territory arises in this context as a strong integration component of information from several sources. Since the 1990s, the geospatial community has been looking for interoperability measures and standards to overcome obstacles of syntactic and semantic conflicts. Service oriented architecture (SOA) is a key component of these efforts, which culminated with the creation of Spatial Data Infrastructures (SDIs), in which standards, policies, technology and capacity building are coordinated in pursuit of improving the sharing and use of geospatial information. However, data published in the National Spatial Data Infrastructure (NSDI-BR) through standardized web services cannot be accessed directly in the resource-oriented linked open data format.This requires a range of actions that include the creation and maintenance of ontologies, the availability of linked data delivery alternatives and the databases interconnection from external sources. In this work, the integration has been demonstrated through a use case, defining ontologies for the Brazilian political-administrative division and the development of a mediator web service that transforms data directly from a WFS service compatible with INDE-BR for the RDF format, compatible with INDA.

Palavras-chaves: Spatial Data Infrastructures, Linked Open Data

LISTA DE SIGLAS

ADGV - Especificação Técnica para Aquisição de Dados Geoespaciais Vetoriais
ANZLIC - *Australia and New Zealand Spatial Information Council*
APA - Área de Proteção Ambiental
CEMG - Comitê de Estruturação de Metadados Geoespaciais
CEMND - Comitê de Estruturação da Mapoteca Nacional Digital
CINDE - Comitê de Implantação da Infraestrutura Nacional de Dados Espaciais
CNGEO - Comitê de Nomes Geográficos
CNMC - Comitê de Normatização do Mapeamento Cadastral
COGIT - *Conception Objet et Généralisation de l'Information Topographique*
CONCAR - Comissão Nacional de Cartografia
CORINE - *Coordination of information on the environment* - Commission
CPRM - Serviço Geológico do Brasil
CSDGM - *Content Standard for Digital Geospatial Metadata*
DBDG - Diretório Brasileiro de Dados Geoespaciais
DHN - Diretoria de Hidrografia e Navegação da Marinha do Brasil
DNPM - Departamento Nacional de Produção Mineral
DNS - *Domain Name System*
DSG - Diretoria de Serviço Geográfico do Exército
EDGM – Estrutura Dados Geoespaciais Matriciais.
EDGV - Estruturação de Dados Geoespaciais Vetorias
e-PING – Padrões de Interoperabilidade do Governo Eletrônico
e-PMG - Padrão de Metadados do Governo Eletrônico
European
e-VoG - Vocabulários e Ontologias do Governo Eletrônico,
FES - *Filter Encoding*
FGDC - *Federal Geographic Data Comitee*
FIG - *International Federation of Surveyors*
FUNAI - Fundação Nacional do Índio
GML - *Geography Markup Language*
GPS - *Global Positioning System*
GRASS - *Geographical Resources Analysis Support System*

GSDI - *Global Spatial Data Infrastructure Association*

HTTP - *Hypertext Transfer Protocol*

HTML - *HyperText Markup Language*

IAG - *International Association of Geodesy*

IBGE - Instituto Brasileiro de Geografia e Estatística

ICA - Instituto de Cartografia Aeronáutica

ICA - *International Cartographic Association*

ICMBio - Instituto Chico Mendes de Conservação da Biodiversidade

IDE - Infraestruturas de Dados Espaciais

IEEE - *Geoscience and Remote Sensing Society*

IETF - *Internet Engineering Task Force*

IGN - *Institut national de l'information géographique et forestière*

INCRA - Instituto Nacional de Colonização e Reforma Agrária

INDA - Infraestrutura Nacional de Dados Abertos

INDE-BR - Infraestrutura Nacional de Dados Espaciais

INPE - Instituto Nacional de Pesquisas Espaciais

INSPIRE - *Infrastructure for Spatial Information in Europe*

IP - *Internet Protocol*

ISO - International Organization for Standardization

ISO/TC211 - *ISO Technical committee on Geographic information/Geomatics*

ISPRS - *International Society for Photogrammetry and Remote Sensing*

JAG - *Joint Advisory Group*

LGD - LinkedGeoData

MMA - Ministério do Meio Ambiente, dos Recursos Hídricos e da Amazônia Legal

OGC - Open Geospatial Consortium

OMT-G - Object Modelling Technique - Geographic

ONG - Organização não Governamental

OWL - *Web Ontology Language*

PERFIL MGB - Perfil de Metadados Geoespaciais do Brasil

RDF - *Resource Description Framework*

RDGV – Representação Dados Geoespaciais Vetoriais.

RIF - *Rule Interchange Format*

RS - *Reasoner Service*

SCN - Sistema Cartográfico Nacional

SDTS - *Spatial Data Transfer Standard*

SFA - *Simple feature Access*

SHP - ESRI Shapefile

SIG - Sistemas de Informação Geográfica

SISNAMA - Sistema Nacional do Meio Ambiente

SLD - *Styled Layer Descriptor*

SLTI - Secretaria de Logística e Tecnologia da Informação do Ministério do Planejamento, Orçamento e Gestão

SOA - *Service-oriented architecture*

SOS - *Sensor Observation Service*,

SPARQL - *SPARQL Protocol and RDF Query Language*

SQL - *Structured Query Language*

TBCD - Tabela de Bases Cartográficas Digitais

TCP - *Transmission Control Protocol*

TIC - Tecnologia de Informação e Comunicação

UML - *Unified Modeling Language*

URI - *Uniform Resource Identifier*

USGS - *United States Geological Survey*

VCGE - Vocabulário Controlado do Governo Eletrônico

VGI - *Volunteered geographic information*

VINDE - Visualizador da INDE-BR

W3C - *World Wide Web Consortium*

WCS - *Web Coverage Service*,

WFS - *Web Feature Service*,

WGS84 - *World Geodetic System 1984*

WKT - *Well-known text*

WMS - *Web Map Service*,

WOS - *Web Ontology Service*

XML - *eXtensible Markup Language*

LISTA DE FIGURAS

FIGURA 1 - CONJUNTOS DE DADOS INTERLIGADOS. 18
FIGURA 2 - APLICATIVO GEOLEIS 21
FIGURA 3 - APLICATIVO ACIDENTES DE TRABALHO POR CATEGORIA 22
FIGURA 4 - APLICATIVO SEGURANÇA PÚBLICA – RIO GRANDE DO SUL. 22
FIGURA 5 - A EVOLUÇÃO DAS IDES 25
FIGURA 6 - DISTRITO DE BREJO DE SÃO FÉLIX. 32
FIGURA 7 - SETORES CENSITÁRIOS DO DISTRITO DE BREJO DE SÃO FÉLIX. 32
FIGURA 8 - SEDE DO DISTRITO DE BREJO DE SÃO FÉLIX 33
FIGURA 9: BREJO DE SÃO FÉLIX – MA (715 HABITANTES – CENSO 2010), SOBRE SERVIÇO DE IMAGENS GOOGLE EARTH. 33
FIGURA 10 - DIAGRAMA DE INTERAÇÃO CLIENTE-PROVEDOR DE SERVIÇOS *WEB*. 37
FIGURA 11 - DIAGRAMA DE TECNOLOGIAS SEMÂNTICAS. 40
FIGURA 12 - RELAÇÃO ENTRE VOCABULÁRIOS EXISTENTES. 44
FIGURA 13 - TELA DO SITE DADOS.GOV.BR – PESQUISA POR DADOS DO IBGE 52
FIGURA 14 - PARADIGMA DOS QUATRO UNIVERSOS 53
FIGURA 15 - SERVIÇOS EM SIG. 55
FIGURA 16 - RELAÇÃO ENTRE AS INSTITUIÇÕES DE PADRONIZAÇÃO NA ÁREA GEOESPACIAL E NA INTERNET. 58
FIGURA 17 - RELACIONAMENTO ENTRE FEIÇÕES, OBJETOS ESPACIAIS E GEOMETRIA. FONTE: OGC, 2012a 70
FIGURA 18 - RELACIONAMENTOS DISPONÍVELS PARA FAMÍLIA *SIMPLE FEATURE* NA ESPECIFICAÇÃO GEOSPARQL. 71
FIGURA 19 - APLICATIVO *FLOOD ALERT* 74
FIGURA 20 - ETAPAS PARA A PUBLICAÇÃO DOS DADOS 75
FIGURA 21 - TELA DO SITE GEO.LINKEDDATA.ES (2013) 77
FIGURA 22 - TELA DO SITE GEONAMES.ORG 78
FIGURA 23 - EXTRATO DA ONTOLOGIA DO LINKEDGEODATA 79
FIGURA 24 - BUSCA NO *BROWSER* DO PROJETO LINKEDGEODATA (2013) ...79

FIGURA 25 - VISÃO DE IDE ORIENTADA A SERVIÇOS. 81
FIGURA 26 - ESTRUTURA DA INDE-BR 82
FIGURA 27 - DIAGRAMA CONCEITUAL DO DBDG 86
FIGURA 28 - DIAGRAMA DA CLASSE PONTO COTADO ALTIMÉTRICO 88
FIGURA 29 - CONTÍNUO ENTRE OS *MASHUPS* E AS IDES 92
FIGURA 30 – DIGRAMA DE CASOS DE USO DA ARQUITETURA PROPOSTA 98
FIGURA 31 - PASSOS PARA DISPONIBILIZAÇÃO DE DADOS INTERLIGADOS. 100
FIGURA 32 - PUBLICAÇÃO DE RDF 101
FIGURA 33 - PUBLICAÇÃO DE DADOS ESPACIAIS ATRAVÉS DE GEOSERVIÇOS 102
FIGURA 34 – ARQUITETURA PROPOSTA. 103
FIGURA 35 - CATEGORIA LOCALIDADES 106
FIGURA 36 - CATEGORIA LIMITES 107
FIGURA 37 - ONTOLOGIA OWL PARA DIVISÃO TERRITORIAL BRASILEIRA 108
FIGURA 38 - EXEMPLO DA HIERARQUIA DE GEODCÓDIGOS. 109
FIGURA 39 - ARQUITETURA DO SERVIÇO *WEB.* 111
FIGURA 40 - INTERFACE DO SERVIDOR GEOSERVER. 112
FIGURA 41: CAIXA DE SELEÇÃO NA INTERFACE. 112
FIGURA 42 - INTERFACE COM O USUÁRIO COM O RDF CRIADO DINAMICAMENTE. 114
FIGURA 43 - PONTOS DO PROJETO GEONAMES DENTRO DO POLÍGONO DO MUNICÍPIO DE CURITIBA 118
FIGURA 44 - RESULTADO DA SEGUNDA CONSULTA 120
FIGURA 45 - RESULTADO DA TERCEIRA CONSULTA 121
FIGURA 46 - APLICAÇÃO DE DADOS INTERLIGADOS ABERTOS DE DIVERSAS FONTES INTEGRADOS À INDE-BR 122

LISTA DE TABELAS

TABELA 1 - DISTRIBUIÇÃO DOS CONJUNTOS DE DADOS INTERLIGADOS. 20
TABELA 2 - FUNÇÕES ESPACIAS DO GEOSPARQL 71
TABELA 3 - QUANTIFICAÇÃO DE TRIPLAS PUBLICADAS.....75
TABELA 4 - EXEMPLOS DE PRODUTORES E DADOS DA INDE-BR. 84
TABELA 5 - DESCRIÇÃO DE DADOS DA FEIÇÃO PONTO COTADO ALTIMÉTRICO 89
TABELA 6 - COMPARAÇÃO IDES E DADOS INTERLIGADOS.....93
TABELA 7 - RESULTADO DA CONSULTA 116
TABELA 8 - RESULTADO DA TERCEIRA CONSULTA 121

SUMÁRIO

1. INTRODUÇÃO 17
1.1. HIPÓTESE **27**
1.2. OBJETIVOS **28**
1.2.1. OBJETIVO GERAL **28**
1.2.2. OBJETIVOS ESPECÍFICOS **28**
1.3. JUSTIFICATIVA **28**
1.3.1. Buscas por Dados Espaciais na INDE-BR **29**
2. *WEB* SEMÂNTICA E DADOS INTERLIGADOS ABERTOS 35
2.1. PROCESSO DE PADRONIZAÇÃO DA INTERNET **35**
2.2. PADRÃO XML – LINGUAGEM DE MARCAÇÃO EXTENSÍVEL **36**
2.3. ARQUITETURA ORIENTADA A SERVIÇOS (SOA) **37**
2.4. *WEB* SEMÂNTICA **38**
2.5. DIAGRAMA DE TECNOLOGIAS SEMÂNTICAS (*SEMANTIC WEB STACK)* **39**
2.6. DADOS INTERLIGADOS (*LINKED DATA*) **43**
2.7. DADOS INTERLIGADOS ABERTOS GOVERNAMENTAIS **45**
2.7.1. Governo Aberto e Transparência **45**
2.7.2. Dados Interligados Abertos no Brasil **47**
2.7.2.1. e-PING – Padrões de Interoperabilidade de Governo Eletrônico **47**
2.7.2.2. Lei de Acesso à Informação **48**
2.7.2.3. Infraestrutura Nacional de Dados Abertos **49**
2.7.2.4. Portal Brasileiro de Dados Abertos **51**
3. *WEB* SEMÂNTICA E DADOS GEOESPACIAIS 53
3.1. INFORMAÇÃO GEOESPACIAL, INTEROPERABILIDADE E SERVIÇOS **53**
3.2. INSTITUIÇÕES DE PADRONIZAÇÃO PARA A INFORMAÇÃO GEOESPACIAL **56**
3.2.1. *Open Geospatial Consortium* (OGC) **56**
3.2.2. ISO/TC211 ..... **56**
3.2.3. *Joint Advisory Group* (JAG) – OGC e ISO/TC211 ..... **57**
3.3. PADRÕES EXISTENTES **58**
3.3.1. Metadados - ISO 19115 e ISO 19139 ..... **59**
3.3.2. Bancos de Dados Geográficos, SQL, SFA e WKT (ISO 19125) **59**

3.3.3. *Geography Markup Language* - GML (ISO 19136) **60**
3.3.4. *Web Map Service Interface* - WMS (ISO 19128) **61**
3.3.5. *Web Feature Service Interface* - WFS(ISO 19142) **62**
3.3.6. *Filter Encoding* - FES (ISO 19143) **62**
3.4. HISTÓRICO DA PESQUISA EM GEOSSEMÂNTICA **63**
3.4.1. Estados Unidos **66**
3.4.2. Reino Unido **66**
3.4.3. França **67**
3.5. INFORMAÇÃO GEOESPACIAL E DADOS INTERLIGADOS **68**
3.5.1. Vocabulário Básico RDF GEO (W3C) **68**
3.5.2. *Ontology* – Parte 1 e 2 (ISO 19150) **68**
3.5.3. GeoSPARQL **69**
3.6. EXPERIÊNCIAS EM DADOS GEOESPACIAIS INTERLIGADOS ABERTOS **72**
3.6.1. Reino Unido – *UK LOCATION* **72**
3.6.2. Estados Unidos – geo.data.gov **74**
3.6.3. Espanha – *GeoLinked Data* **76**
3.6.4. GeoNames **77**
3.6.5. *LinkedGeoData (* baseado no *Open Street Map*) **78**
4. INFRAESTUTURA DE DADOS ESPACIAIS 80
4.1. IDEs – HISTÓRICO E CONCEITOS FUNDAMENTAIS **80**
4.2. INDE-BR **82**
4.2.1. Diretório Brasileiro de Dados Geoespaciais e Portal SIG Brasil **85**
4.2.2. NORMAS E PADRÕES VIGENTES DA INDE **87**
4.2.2.1. ESPECIFICAÇÃO TÉCNICA PARA ESTRUTURAÇÃO DE DADOS GEOESPACIAIS VETORIAIS (ET-EDGV) **87**
4.2.2.2. PERFIL DE METADADOS GEOESPACIAIS DO BRASIL **90**
5. INTEGRAÇÃO DADOS INTERLIGADOS e IDEs 91
5.1. DIFERENÇAS CONCEITUAIS E VISÕES SOBRE A INTEGRAÇÃO **91**
5.2. A ESCOLHA DO USO DO GML COMO INTEGRADOR **95**
5.3. ABORDAGENS DE INTEGRAÇÃO BASEADA EM SERVIÇOS **95**
6. METODOLOGIA 98
6.1. DIAGRAMA DE CASOS DE USO DA SOLUÇÃO **98**
6.2. EFEITOS DA ARQUITETURA NAS ETAPAS DE PUBLICAÇÃO DE DADOS INTERLIGADOS GEOESPACIAIS **99**

6.3. PREMISSAS PARA A CRIAÇÃO DO SERVIÇO *WEB* **100**
6.4. DESENVOLVIMENTO DO SERVIÇO MEDIADOR **111**
7. RESULTADOS **115**
7.1. CONSULTA GEOSPARQL EM DUAS BASES DE DADOS: GEONAMES (RDF) e LIMITES MUNICIPAIS (WFS) **115**
7.2. CONSULTA GEOSPARQL EM UMA MESMA BASE DE DADOS: LIMITES MUNICIPAIS (WFS) **119**
7.3. CONSULTA GEOSPARQL EM DUAS BASES DE DADOS: DBPEDIA (RDF) e LIMITES MUNICIPAIS (WFS) **120**
7.4. POTENCIAL PARA UTILIZAÇÃO DE CONSULTAS GEOSPARQL INTEGRANDO DADOS DA INDE-BR **121**
8. CONCLUSÕES **124**
8.1. AVALIAÇÂO DA SOLUÇÃO PROPOSTA **124**
8.2. RECOMENDAÇÕES **124**
8.3. PESQUISAS FUTURAS **126**
8.4. CONSIDERAÇÕES FINAIS **127**

1. INTRODUÇÃO

Todos os dias, são publicados na *web* cerca de 2,5 quintilhões de dados (IBM, 2012), vindos de sensores, postagens em redes sociais, fotos e vídeos digitais, coordenadas oriundas de GPS em dispositivos móveis, entre outros. Este universo de dados é chamado de "*Big Data*", tópico que tem atraído interesse e fomentado investimentos de instituições como o Banco Mundial (WORLD ECONOMIC FORUM, 2012) e o governo dos Estados Unidos (WHITE HOUSE, 2012). Esta disponibilidade, sem precedentes, apresenta um novo desafio relacionado à forma de interagir com esses dados.

A expansão na quantidade de bases de dados conectadas globalmente caracteriza a existência da "*web* de dados", definida como "uma rede de coisas do mundo físico descritas por dados na *web*" (BIZER, HEATH e BERNERS-LEE, 2009). Dentre as diversas informações a respeito das "coisas" do meio físico se destacam as suas relações no espaço geográfico, o que mostra a importância da informação geoespacial neste novo espaço de dados interligados.

A *Web* Semântica, termo proposto por Berners-Lee, Hendler e Lassila (2001) é a rede que estrutura o significado do conteúdo da *Web*, na qual os relacionamentos entre as diversas bases de dados pudessem ser executados pelos computadores. Sua principal característica é estender as capacidades atuais da Internet com o acréscimo dos significados dos conceitos, para que desta forma diversos aplicativos e usuários possam ampliar o seu potencial de trabalhar cooperativamente. Outro conceito importante é o de ontologias, definido por Guarino (1998), como um "vocabulário específico usado para descrever certa realidade, somado com um conjunto de premissas explícitas sobre o sentido pretendido pelas palavras no vocabulário".

O conjunto de boas práticas que permite a publicação de dados e seus relacionamentos dentro de uma estrutura semântica é chamado de Dados Interligados (*Linked Data*) (BIZER, HEATH e BERNERS-LEE, 2009). "Boas práticas" é o termo aplicado em diversos contextos para identificar estratégias avaliadas por especialistas como claramente superiores a outras alternativas (PETERS e HERON, 1993). Estas práticas, quando aplicadas no âmbito de Dados Abertos Governamentais, são chamadas de Dados Interligados Abertos (*Linked Open Data).*

Os dados interligados abertos baseiam-se na ligação entre URIs (*Uniform Resource Identifier*), um conjunto de caracteres que identifica um nome ou um recurso na *web* (BERNERS-LEE, 1994). Essa concepção é uma evolução da forma de navegação entre documentos através de *links*, mas agora com a possibilidade de relacionamento direto entre os dados. Com esta abordagem, é possível a criação de *mashups*, aplicativos de busca e outras funcionalidades que operam sobre um espaço de dados global. *Mashup* é uma técnica para construir aplicações que combinam dados de fontes múltiplas para que o usuário os utilize de forma integrada (CLARKIN e HOLMES, 2007).

Atualmente, um grande número de bases de dados já se encontra disponível na *web* segundo os princípios dos Dados Interligados. Na FIGURA 01 estão representados os 295 conjuntos de dados publicados levantados por Cyganiak e Jentzsch (2011).

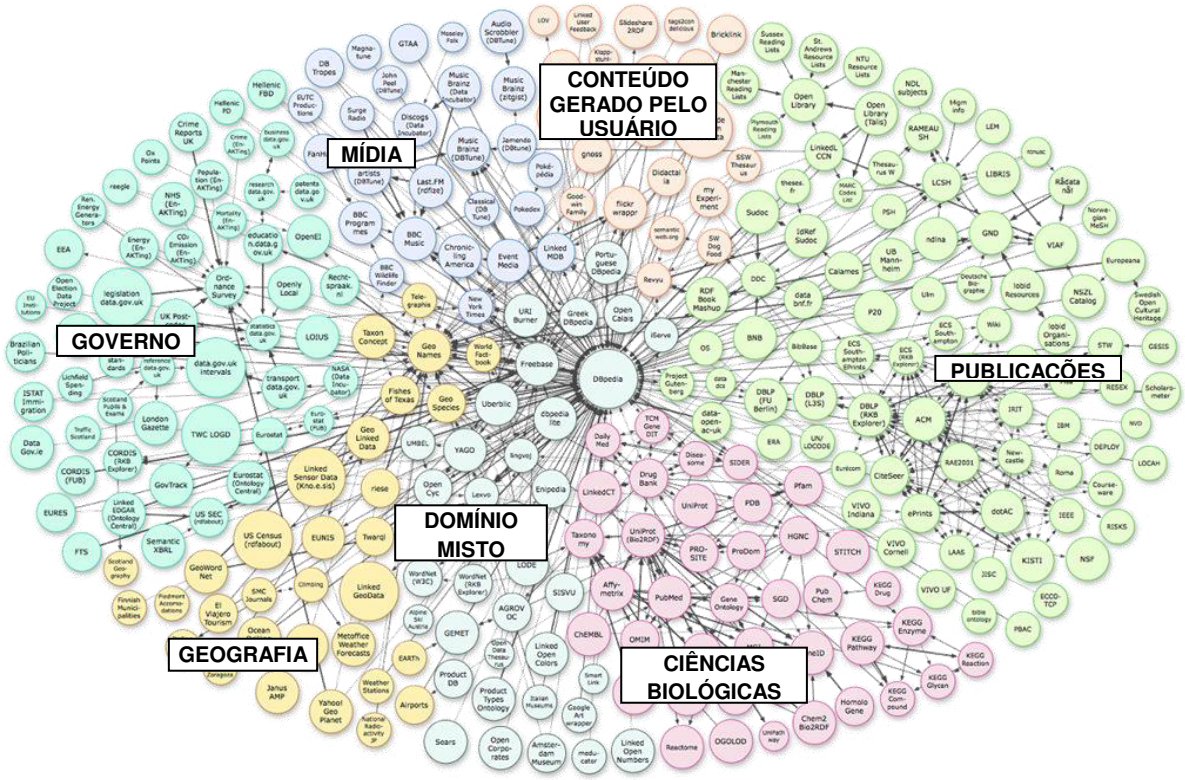


FIGURA 1 - CONJUNTOS DE DADOS INTERLIGADOS.
FONTE: CYGANIAK e JENTZSCH (2011)

Tais inovações na área da tecnologia da informação e comunicação geram um grande impacto também na participação dos indivíduos na sociedade. Os

governos produzem, em suas diversas esferas, uma quantidade bastante expressiva de dados, seja para executar suas atribuições ou por determinação legal. Adicionalmente, o aumento das interações entre pessoas e entidades governamentais cria uma demanda por transparência na implantação de políticas públicas, que consequentemente deflagra uma pressão pela publicação de dados de forma aberta.

Em setembro de 2011, o governo brasileiro adere à Parceria Governo Aberto, entidade internacional de afiliação voluntária na qual os governos se comprometem a seguir compromissos com a transparência e a participação popular (OPEN GOVERNMENT PARTNERSHIP, 2012). Dentre os compromissos estabelecidos no Plano de Ação do Governo Brasileiro Parceria para Governo Aberto (BRASIL, 2011), estão a criação do Portal Brasileiro de Dados Abertos e a criação da Infraestrutura Nacional de Dados Abertos (INDA), que se concretizou em abril de 2012. A INDA faz parte da política do governo brasileiro para dados abertos, um conjunto de padrões, tecnologias, procedimentos e mecanismos de controle necessários para atender às condições de disseminação e compartilhamento de dados e informações públicas dentro dos compromissos e legislação estabelecidos.

A Lei de Acesso à Informação (Lei nº 12.527), publicada em 18 de novembro de 2011, torna obrigatória a gestão transparente da informação, seu amplo acesso e divulgação, bem como a proteção de sua disponibilidade, autenticidade e integridade.

Tanto no âmbito dos Dados Interligados em geral, quanto dos dados governamentais, a informação geoespacial ocupa lugar de destaque. Segundo Cyganiak e Jentzsch *et al*. (2011), a categoria de base de dados geográficos era a segunda com o maior número de triplas RDF publicadas, 19,43% das mais de 31 bilhões disponíveis em setembro de 2011, conforme TABELA 1. O primeiro lugar era ocupado pelo grupo “Governo”. Em termos de ligações com outras bases de dados, o grupo “Geográfico” ficou em 5º lugar, com 7,11%. Considerando o potencial de espacialização dos dados dos demais grupos, através de nomes geográficos, endereços e outras informações presentes em notícias da mídia, ações governamentais, entre outros, pode-se inferir que a informação geoespacial carece de ligação com os demais tipos de dados.

TABELA 1 - DISTRIBUIÇÃO DOS CONJUNTOS DE DADOS INTERLIGADOS.

Domínio	Número de conjuntos de dados	Triplas	%	Ligações externas	%
Mídia	25	1.841.852.061	5,82 %	50.440.705	10,01 %
Geografia	31	6.145.532.484	19,43 %	35.812.328	7,11 %
Governo	49	13.315.009.400	42,09 %	19.343.519	3,84 %
Publicações	87	2.950.720.693	9,33 %	139.925.218	27,76 %
Domínio Misto	41	4.184.635.715	13,23 %	63.183.065	12,54 %
Ciências Biológicas	41	3.036.336.004	9,60 %	191.844.090	38,06 %
Conteúdo Gerado pelo Usuário	20	134.127.413	0,42 %	3.449.143	0,68 %
	295	31.634.213.770		503.998.829	

FONTE: CYGANIAK e JENTZSCH, 2011.

Das principais ontologias ou vocabulários do domínio geográfico destacados neste estudo, a principal é *geonames.org*, que disponibiliza dados em RDF para mais de 8.3 milhões de topônimos (GEONAMES, 2013), mas sem vínculos diretos com nenhum fornecedor oficial de informações cartográficas. Outro vocabulário geográfico bastante difundido é o *Basic Geo Vocabulary* (WGS84 lat/long), criado pela W3C em 2003, mas que apenas armazena pares de coordenadas latitude e longitude relacionadas ao sistema WGS84, portanto, bastante limitado em termos de aplicações geoespaciais. Apesar destas limitações, este vocabulário era usado por 8,47% das bases de dados na época da pesquisa.

Dentre as informações disponíveis no Portal Brasileiro de Dados Abertos (BRASIL, 2012b) em dezembro de 2012, a única que contém a geometria de feições geográficas é a chamada “Malha Geométrica dos Municípios Brasileiros”, a qual está disponível para *download* em formato *Shapefile (SHP), e* portanto não indexada automaticamente de um repositório de dados. Este conjunto de dados não segue completamente os princípios dos dados abertos, por não estar em formato RDF. Assim, é utilizado em forma de cópia de arquivo pelo usuário, o que pode causar inconformidades em caso de atualizações. Os dados também não estão estruturados semanticamente com o uso de ontologias para a documentação dos

significados. Embora não possuam geometria associada, dados diversos do portal têm informação de localização implícita, como nome de municípios ou endereços. Dentre os aplicativos disponíveis no portal, vários possuem interface de mapa para acesso às informações, como nos exemplos a seguir:

- Geoleis (geoleis.herokuapp.com): Sistema que mostra a distribuição espacial dos projetos de lei da Câmara Municipal de São Paulo promulgados por ano.

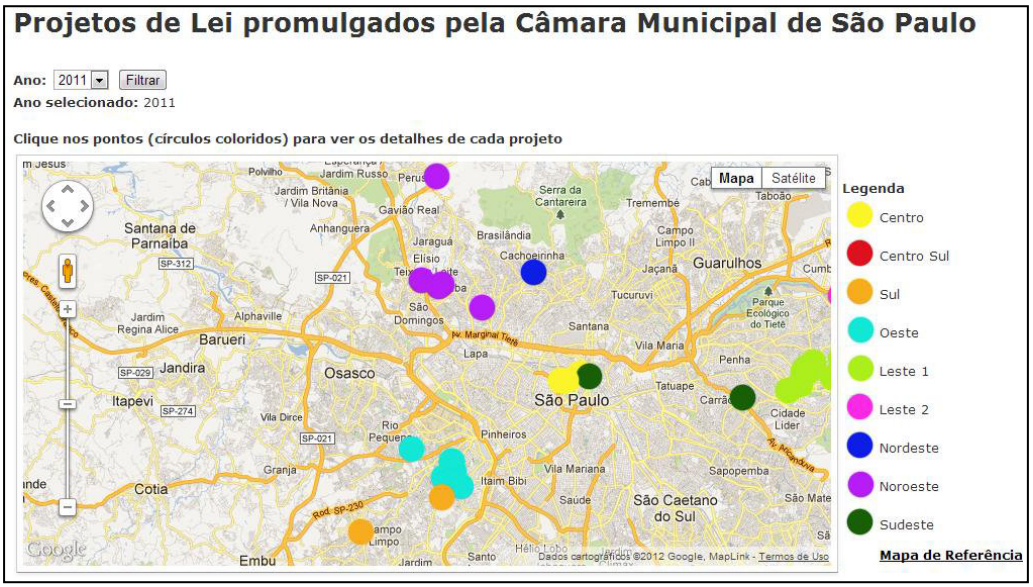


FIGURA 2 - APLICATIVO GEOLEIS
FONTE: GEOLEIS.HEROKUAPP.COM, 2013

- Mapa dos acidentes de trabalho (api.dataprev.gov.br/doc/visualizacao-mapa.html): Visualização do mapa com os acidentes de trabalho por município e por categoria.

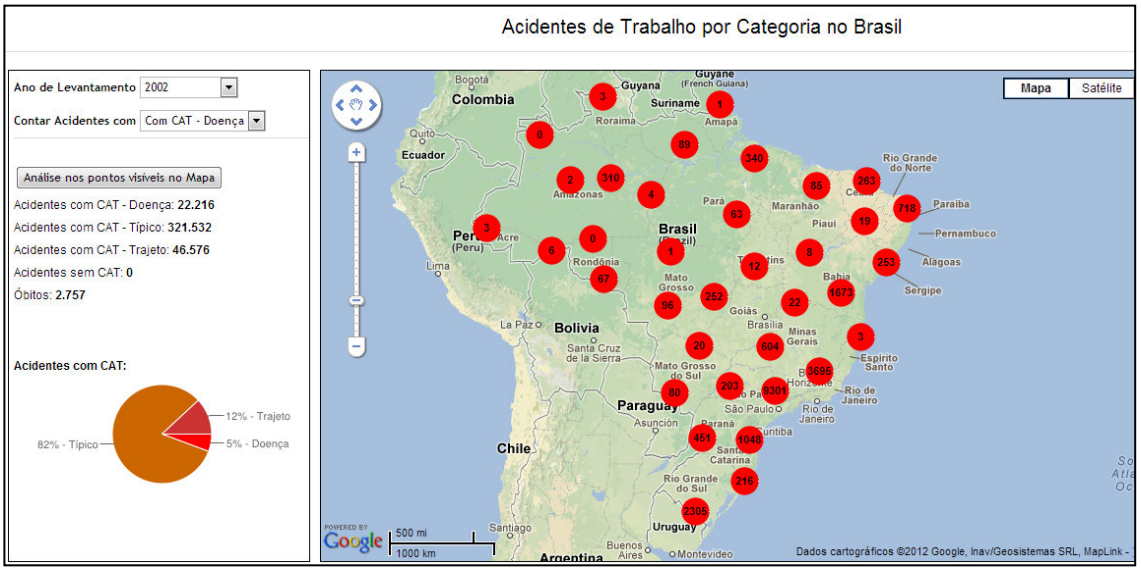


FIGURA 3 - APLICATIVO ACIDENTES DE TRABALHO POR CATEGORIA
FONTE: API.DATAPREV.GOV.BR/DOC/VISUALIZACAO-MAPA.HTML, 2013

- Onde Acontece (ondeacontece.com.br): Mostra os dados de ocorrências de crimes, agregados por município, no Rio Grande do Sul

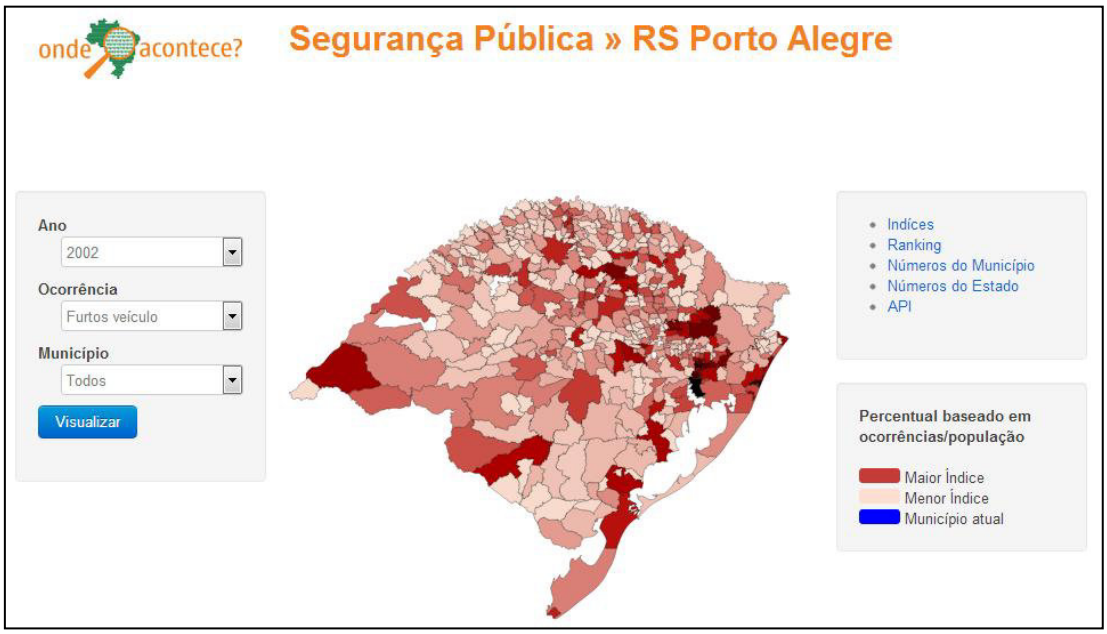


FIGURA 4 - APLICATIVO SEGURANÇA PÚBLICA – RIO GRANDE DO SUL.
FONTE: ONDEACONTECE.COM.BR, 2013

Estes aplicativos mostram a existência de demanda para *mashups* que usam informação geoespacial para a visualização dos dados. No entanto, pela falta de disponibilidade de bases cartográficas oficiais em formato adequado, e talvez também de conhecimento em cartografia por parte dos desenvolvedores, não é explorada a potencialidade de recursos disponíveis para este tipo de aplicação.

As informações geoespaciais deveriam ser obtidas diretamente da fonte oficial dos dados e não compiladas separadamente pelos criadores dos aplicativos (ou mesmo utilizando dados de terceiros como Google ou Geonames). Este é um indicativo da demanda por dados espaciais oficiais publicados em formato de Dados Interligados Abertos.

Em termos de informação geoespacial, o desenvolvimento de pesquisas e aplicações sobre a questão semântica teve uma abordagem específica, desde trabalhos pioneiros que destacam a questão com relação à informação geográfica como o de Worboys e Deen (1991) que previa as vantagens de dados distribuídos em "nós" locais (antevendo a utilização da *web*) e definia heterogeneidade semântica como o *"desacordo a respeito do significado, interpretação ou uso pretendido para o dado".*

Tornou-se vital, a partir da década de 1990, que a comunidade de informação geoespacial se reunisse para resolver conflitos de interoperabilidade, principalmente uma integração melhor e mais rápida entre *softwares* da área (OGC, 2012). O Consórcio Geoespacial Aberto, OGC, formado por 483 empresas, entidades governamentais e universidades foi então constituído em 1994, e tem como objetivo desenvolver padrões baseados em consenso para este campo. Outra entidade criada com objetivos complementares é o ISO/TC211 (Comitê Técnico especializado em informação geográfica da Organização Internacional para Padronização – ISO), criado também em 1994. Ambas as entidades trabalham em cooperação, sendo os padrões discutidos pelo ISO/TC211 de caráter mais amplo, envolvendo vocabulário, conteúdo e modelagem abstrata de objetos e os tratados pelo OGC, de competência mais tecnológica, com a definição de interface de software e outros padrões de implementação.

Vários destes padrões se referem a serviços *web* (*Web Feature Service* – WFS , *Web Map Service* – WMS, *Web Coverage Service* – WCS, *Sensor Observation Service* - SOS, entre outros), dentro de um conceito de arquitetura orientada a serviços, uma arquitetura de sistemas cuja finalidade é a junção entre agentes de software que interagem entre si (HE, 2003).

Com o estabelecimento de padrões de interoperabilidade, fez-se necessária a articulação na comunidade de informações geoespacial para a coordenação de outros aspectos do compartilhamento, como políticas públicas, recursos humanos e

arranjos interinstitucionais, em nível local, nacional e global. As Infraestruturas de Dados Espaciais (IDEs) foram então formadas, da união de usuários, produtores e reguladores do uso de geoinformação. Já em 2001 foram identificados mais de 120 países com Infraestruturas Nacionais de Dados Espaciais (INDEs) em processo de estabelecimento (CROMPVOETS *et al.*, 2004).

É significativa a mudança de paradigma da primeira geração de IDEs criada na década de 1990, essencialmente voltada a dados e processos, para as atuais, mais orientadas a serviços e voltadas às demandas dos usuários (BERNARD e CRAGLIA, 2005; FERNÁNDEZ e CASTELLANOS, 2006; LEITE JR. *et al.*, 2007 e MASSER, 2009). Da análise de casos internacionais e da INDE-BR, Borba *et al. (*2012), sumariza alguns dos sintomas que impedem o sucesso das IDEs, tais como: alta complexidade, falta de participação ativa e carência de integração.

Portanto, se estabelece como uma questão importante para as INDEs o esforço em facilitar o acesso aos dados pelos usuários. A FIGURA 05 ilustra o desenvolvimento das IDEs, em paralelo com o desenvolvimento da Internet em geral. As diversas fases da evolução da Internet podem ser comparadas, analogamente, às permissões de sistemas de arquivo (TIE-JUN *et al.*, 2009) tal que a primeira fase, ou *web* 1.0, na qual seriam possíveis ações de “somente leitura” de conteúdo, a seguinte ou *web* 2.0, com a possibilidade de “leitura e escrita”, com o advento das redes sociais e participativas no estilo wiki e, finalmente, a *web* 3.0, que permite a “leitura, escrita e execução”, numa alusão à inclusão de capacidades semânticas e à orientação a serviços. Igualmente, as IDEs em geral inicialmente estavam focadas em divulgar dados governamentais através de portais de dados geoespaciais centralizados. Com a expansão do número e quantidade de atores, incluindo produtores privados de dados, teve início uma fase centrada em processos e padrões criados em fóruns especializados como o OGC. Entretanto, segundo Sadeghi-Niaraki *et al.*, (2010) no desenho desses processos as preferências e interesses dos usuários finais não foram considerados. Desta forma, surge atualmente uma nova fase, baseada na participação mais ativa do usuário e que promove o papel do indivíduo como produtor e consumidor da informação geoespacial através de dispositivos móveis, cada vez mais populares. É a fase da interação social e do compartilhamento de informações em tempo real que tem se mostrado especialmente útil em situações que demandam respostas ágeis como

desastres naturais e crises de emergência social (GOODCHILD e GLENNON, 2010; LAITURI e KODRICH, 2008; e KWAN e LEE, 2005).

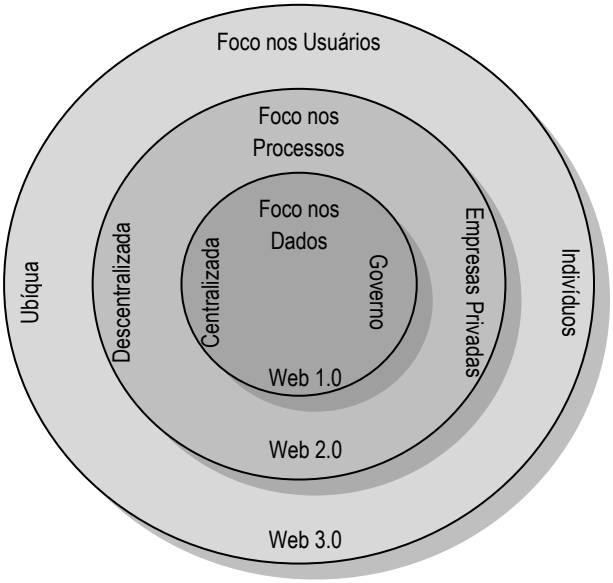


FIGURA 5 - A EVOLUÇÃO DAS IDES
FONTE: SADEGHI-NIARAKI *et al.*, 2010

No Brasil, a INDE-BR foi criada através do Decreto n° 6.666 de novembro de 2008, tendo a CONCAR (Comissão Nacional de Cartografia) como a sua entidade regulamentadora. O Plano de Ação da INDE-BR (CONCAR, 2010) foi lançado em abril de 2010 e estabelece as principais ações para implementação da infraestrutura e seu respectivo cronograma. No capítulo de Normas e Padrões e de Tecnologia é adotada a concepção orientada a serviços de acordo com os padrões do OGC. No capítulo de Dados Geoespaciais, dentre outras definições, estão caracterizados os dados geoespaciais de referência e os atores que são os responsáveis legais pela sua publicação. Foram considerados como de referência os dados agrupados da seguinte forma: Controle Geodésico, Cartografia Terrestre Básica (escalas geográfica, topográfica e cadastral), Cartografia Especial (Náutica e Aeronáutica) e Dados subsidiários e acessórios (Divisão Política Administrativa, Imagens, Terras Indígenas, Bacias Hidrográficas, Unidades de Conservação, Nomes Geográficos, Dados Fundiários e Modelos Numérico de Terreno). Estes são de responsabilidade de uma série de atores do governo federal, estadual e municipal bem como produzidas em convênios e contratos com outros órgãos ou com a iniciativa privada.

Estas informações de referência são a base de todas as análises espaciais e são as mais demandadas para atuar como base em uma série de aplicações de base geoespacial. Algumas destas informações (como, por exemplo, o mapeamento sistemático na escala 1:1.000.000), está disponível ao público atualmente através de serviços WMS. A capacitação da INDE-BR conta com um módulo específico de tecnologia com instruções para a publicação de geoserviços, funcionalidade disponível em uma série de soluções encontradas tanto em software livre como proprietários. Portanto, a arquitetura orientada a serviços está estabelecida como fundamental na definição da infraestrutura brasileira.

Atualmente, a INDE-BR ainda está no seu ciclo inicial de implementação, fortemente identificado com as características de IDEs com foco nos dados. No entanto, com o crescimento da adoção de normas internacionais e produção de normas locais, como o Perfil Brasileiro de Metadados e a Estruturação de Dados Geoespaciais Vetorias (ET-EDGV), a infraestrutura nacional está avançando para a segunda etapa do modelo. Os nós do Diretório Brasileiro de Dados Geoespaciais (DBDG) espelham essa configuração de servidores distribuídos da IDE focada em processos. Nesse esquema de evolução das IDEs, uma fase não suplanta a anterior, ao contrário, ambas convivem com a extensão das capacidades inicialmente instaladas para novos horizontes de aplicação, o que aponta para o crescimento da preocupação com o usuário também na INDE-BR.

Dentre as propostas para esta terceira geração de IDEs, é ressaltada a importância do *Volunteered Geographic Information* (VGI), informação geográfica criada e disseminada voluntariamente pelos usuários, um caso específico de conteúdo gerado pelo usuário (GOODCHILD, 2007). O VGI demonstra a capacidade e o interesse do usuário em criar informação, aplicativos, *mashups* e gerar suas análises personalizadas.

Outras pesquisas e aplicações para as IDEs buscam a solução para aproximar-se dos usuários utilizando funcionalidades semânticas ou ontologias (ou *Ontology-driven SDI*) (SADEGHI-NIARAKI *et al.*, 2010), ou mesmo mais recentemente tentam propor integração com os Dados Interligados Abertos.

Algumas das propostas para integração de conceitos semânticos às IDEs se concentram na criação e adaptação de ontologias para dados geoespaciais e também de sistemas de conversão e referência entre tais ontologias, tais como:

Miguel, 2009; Kuhn, 2003 e Janowicz e Wilkes, 2009. Outros, como Janowicz *et al*., 2010, propõem a criação de uma camada ontológica para os serviços OGC. Esta camada possibilitaria a criação e manutenção de ontologias, ao mesmo tempo em que proveria um serviço de *reasoner* associado, caracterizando então um WOS (*Web Ontology Service*) e um RS (*Reasoner Service*). Propostas semelhantes foram apresentadas por Lacasta *et al.*, 2007 e Lutz *et al.*, 2009.

As abordagens orientada a serviços das IDEs e a abordagem de dados interligados podem conviver de forma complementar, e não excludentes (ARCTUR, 2011). Os especialistas produtores de dados geoespaciais e usuários especializados ainda precisarão de informação estruturada detalhada, contendo informações técnicas sobre qualidade, abrangência e outros aspectos específicos da informação geoespacial, enquanto a comunidade de TI em geral demanda pelo uso da informação de forma integrada com a *web* de dados, por fazer inferências, por descobrir tendências, por criar *mashups* e por outras aplicações.

A questão chave é que as IDEs não foram estruturadas para publicação de dados RDF e sim serviços *web* como WFS. Não é um problema de pura conversão de formatos, mas sim de possibilitar o acesso direto ao recurso, e também às inter-relações semânticas que são inerentes a cada conjunto de dados, através de ontologias que descrevam os conceitos envolvidos na criação e manipulação do dado. Neste trabalho é abordada uma arquitetura para que os dados da INDE-BR possam ser publicados em formato compatível com a INDA e assim possibilitar a integração de informação geoespacial de fontes oficias à *web* de dados no Brasil.

1.1. HIPÓTESE

Para que o usuário consiga explorar o aspecto geográfico das informações publicadas na forma de Dados Interligados Abertos no Brasil, é importante utilizar os serviços da INDE-BR como fonte de informações oficiais, e, para tanto se deve descrever de forma semântica a informação geoespacial existente e torná-la acessível em formato compatível.

Desta forma, é mantida a publicação das informações espaciais dentro dos serviços padronizados pelo OGC, que, além de já considerarem as peculiaridades

inerentes a este tipo de informação, também possibilitam a utilização da infraestrutura de tecnologia da informação e de recursos humanos já estabelecida nas instituições participantes da INDE-BR.

1.2. OBJETIVOS

1.2.1. OBJETIVO GERAL

O objetivo deste trabalho é propor uma arquitetura que possibilite buscas semânticas utilizando de forma integrada os dados geoespaciais publicados na INDE-BR e os dados interligados abertos.

1.2.2. OBJETIVOS ESPECÍFICOS

a) Propor e descrever a arquitetura de integração.
b) Gerar a ontologia que permita a estruturação semântica de uma camada da INDE-BR a ser selecionada.
c) Desenvolver um serviço web que disponibilize dados geoespaciais da INDE-BR de forma automática.
d) Detalhar a sintaxe de buscas no padrão GeoSparql com critérios espaciais, usando dados da INDE-BR e outros dados interligados abertos publicados.

1.3. JUSTIFICATIVA

Uma integração entre as duas visões, a orientação a serviços em que se baseiam as IDEs, e a orientação a recursos que permite os dados interligados, consolidaria a posição da INDE-BR como a fornecedora de dados espaciais oficiais para a INDA. Adicionalmente, seria possível a integração da INDE-BR com informação geoespacial publicada em formato interligado aberto de diversas origens, inclusive as informações geográficas voluntárias que podem, por exemplo, contribuir para a dinamização do processo de atualização cartográfica.

Outra vantagem seria a criação de aplicativos que, além de se servirem da informação geoespacial oficial, atualizada e fidedigna, como publicada pelos provedores de produtos cartográficos de referência, também teriam a flexibilidade das inúmeras aplicações semânticas que podem ser criadas utilizando a abundância de dados interligados abertos disponíveis.

Da forma que está sendo proposta, esta arquitetura permite disponibilização de informação geoespacial em formato RDF a partir dados gerados por um serviço WFS, e portanto toda a dinâmica de atualização cartográfica pode ser retratada através de conexões automáticas. A Divisão Político-Administrativa foi escolhida para este caso de uso primeiramente porque muitos dados, principalmente governamentais, têm como informação de localização o nome do estado ou do município na qual as ações ocorreram. Adicionalmente, é um tema de atualização constante para o qual não é indicada a replicação da base de dados. De forma similar, a mesma integração pode ser extrapolada para os demais dados geoespaciais de referência definidos no Plano de Ação da INDE.

1.3.1. Buscas por Dados Espaciais na INDE-BR

Uma característica básica de uma IDE é a facilitação do acesso e troca de informações geoespaciais. Para tanto, é fundamental a existência de ferramentas que possibilitem ao usuário interagir com as diversas bases de dados envolvidas, procurando por informações que atendam às suas necessidades e, para isso, usando critérios espaciais, temporais ou outras características associadas às informações. Estas interações com as IDEs em busca de informação se dividem em duas frentes (KUHN, 2003):

• Descoberta de Informação Geoespacial: numa estrutura de produtores e usuários distribuídos através da rede, a estrutura da IDE prevê a existência de catálogos de metadados onde os produtores possam publicar as características das informações disponíveis. Geralmente, estes catálogos possuem ferramentas de buscas nos metadados para selecionar conjuntos de dados que atendam determinados critérios que estejam descritos pelos metadados (área de abrangência, forma de execução, data de atualização, nível de detalhamento, entre

outros). Para descrição do conteúdo da base de dados, são usadas palavras-chave, o que pode acabar gerando problemas na busca caso seja usada terminologia na descrição do conteúdo diferente da empregada pelo usuário.

• Recuperação de Informação Geoespacial: uma vez descoberta uma fonte de dados geoespaciais na etapa anterior, existe a interação com o conjunto de dados em si, através da criação de consultas ou filtros das informações armazenadas. Este processo pode ser feito interagindo diretamente com Banco de Dados, ou através de serviços *online*. Novamente, problemas semânticos podem ocorrer na interpretação do esquema conceitual usado quando da modelagem dos dados utilizados. Uma represa, por exemplo, pode ser entendida diferentemente em contextos distintos, como por exemplo, no contexto de produção energética, de lazer ou como limite entre dois municípios.

A seguir são ilustrados os passos que um usuário deve seguir para realizar uma tarefa de busca de informações usando os dados disponíveis atualmente na INDE-BR e no site do IBGE (www.ibge.gov.br). O exemplo foi escolhido por ilustrar tanto a Divisão Territorial Brasileira, informação de referência para diversas aplicações, e dados do Censo Demográfico, a única pesquisa a visitar todos os domicílios do Brasil e que fornece informações bastante ricas e detalhadas.

No exemplo, um usuário gostaria de saber a população de uma localidade chamada “Brejo de São Félix” a partir de um mapa de localização (perímetro). Ele não sabe se é uma sede municipal, distrital ou apenas um lugarejo isolado. Usando mecanismos de busca tradicional, ele precisaria seguir os seguintes passos:

1 – Acessar o serviço WMS “Localidades” disponibilizado pelo IBGE na INDE, de geometria “pontos”, com dados provenientes da Base Cartográfica ao Milionésimo.

2 – Com a ferramenta de identificação no visualizador WMS utilizado, clicar sobre um ponto para descobrir que os nomes estão armazenados em um campo chamado “[nomeloc]”.

3 - Gerar a consulta SQL:

```
SELECT * FROM localidades WHERE nomeloc = ‘Brejo de São Félix’
```

4 – Como resultado, obtém-se a seguinte informação:

```
Brejo de São Félix-MA - localidades
 gid: '1336'
 nomemun: 'Parnarama'
 tipo: 'Vila'
 nomeloc: 'Brejo de São Félix'
 anocria: '1948.00000'
 estado: 'MA'
```

5 – Portanto, Brejo de São Félix é uma vila do município de Parnarama no estado do Maranhão.

6 – Para o próximo passo, o usuário precisaria obter a informação de que o termo “Vila” se refere à denominação oficial para uma Sede Distrital.

7 – Como a malha distrital não está publicada em forma de Geoserviço, o usuário precisa obter o arquivo vetorial no formato *shapefile* da malha distrital do Estado do Maranhão, acessando o seguinte endereço:

```
ftp://geoftp.ibge.gov.br/malhas_digitais/censo_2010/setores_censitarios/shape/ma_v1.3.zip
```

8 - Adicionalmente, é preciso obter a informação que os arquivos correspondem às seguintes camadas:

```
21SEE250GC_SIR.shp = Setores Censitários
21SDE250GC_SIR.shp = Subdistritos
21MUE250GC_SIR.shp = Municípios
21DSE250GC_SIR.shp = Distritos
```

9 – Portanto, o usuário deve fazer uma nova consulta SQL no shapefile 21SDE250GC_SIR:

```
SELECT * FROM distritos WHERE NM_DISTRIT  =  'BREJO DE SÃO FÉLIX'
```

Com o resultado, tem-se que este distrito possui o código CD_GEOCODI = ‘210780310’.

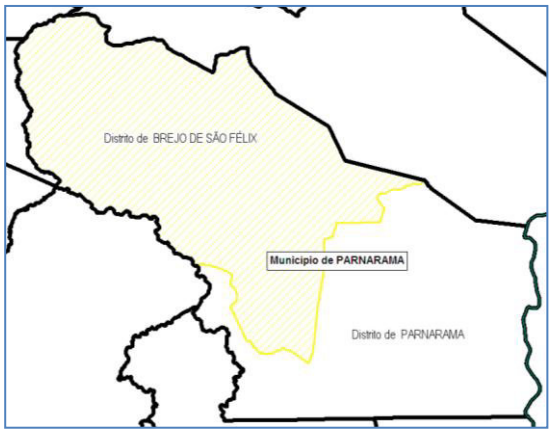


FIGURA 6 - DISTRITO DE BREJO DE SÃO FÉLIX.
FONTE: IBGE, 2010

10 – É possível então selecionar os setores censitários que formam o distrito de Brejo de São Félix, com a consulta SQL no arquivo vetorial em formato *shapefile* 21SEE250GC_SIR:

```
SELECT * FROM setores WHERE CD_GEOCODI like '210780310%'
```

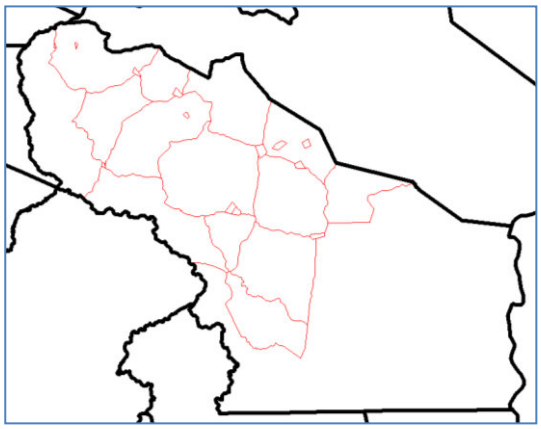


FIGURA 7 - SETORES CENSITÁRIOS DO DISTRITO DE BREJO DE SÃO FÉLIX.
FONTE: IBGE, 2010

11 – Destes setores censitários, deve-se realizar uma nova consulta para obter a sede do distrito (é necessário mais uma informação adicional: que a sede do distrito é representada pelo setor 01):

```
SELECT * FROM setores WHERE CD_GEOCODI = 210780310000001)
```

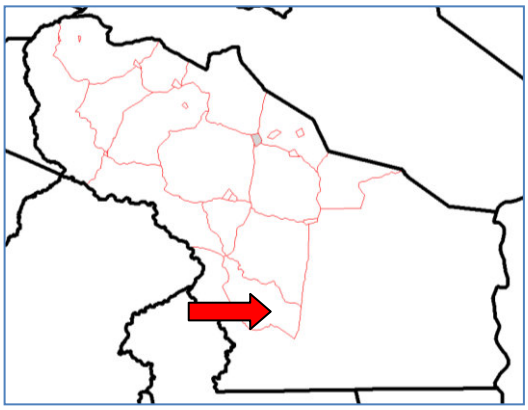


FIGURA 8 - SEDE DO DISTRITO DE BREJO DE SÃO FÉLIX
FONTE: IBGE, CENSO 2010

12 – Agora, com o código do Setor Censitário, é possível baixar as tabelas do resultado do Censo 2010: (ftp://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Sinopse/Agregados_por_Setores_Censitarios/Base_informacoes_setores2010_sinopse_MA.zip) e chega-se à informação desejada, ou seja, que a vila de Brejo de São Félix possuía, no Censo 2010, 715 residentes.

FIGURA 9: BREJO DE SÃO FÉLIX – MA (715 HABITANTES – CENSO 2010), SOBRE SERVIÇO DE IMAGENS GOOGLE EARTH.
FONTE: A AUTORA, 2013

São, portanto, doze passos para a obtenção da informação, exigindo conhecimento da estrutura e das nomenclaturas de arquivos e tabelas, dos

conceitos e da codificação empregados, além da localização da informação distribuída em diferentes páginas da *web*.

2. *WEB* SEMÂNTICA E DADOS INTERLIGADOS ABERTOS

A visão da *web* semântica trata da evolução da *Internet*, de uma interligação entre páginas e documentos para um verdadeiro repositório de recursos interligados. A *Internet*, ao contrário de ter uma governança única e centralizada, mantém seu caráter livre e segue padrões construídos em grupo pelos membros da comunidade e também adotados de forma voluntária. Neste capítulo são descritos os padrões mais importantes para o tópico, que formam o arcabouço de conceitos ligados à Tecnologia da Informação envolvidos no entendimento das questões semânticas relacionadas ao problema.

Os Dados Abertos Interligados proporcionam as melhores práticas para a concretização da *web* de dados. Os governos, com suas atribuições legais de produção maciça de dados e com a pressão popular por mais inclusão e transparência, estão investindo em políticas públicas que garantam o acesso à informação, enquanto trabalham em padrões tecnológicos que possam prover os dados de forma mais direta e interligada. Algumas experiências internacionais na área são estudadas para compreensão de como os governos estão trabalhando estas questões, bem como as ações brasileiras em termos de legislação, políticas públicas e disponibilização de dados abertos.

2.1. PROCESSO DE PADRONIZAÇÃO DA INTERNET

A *Internet* é uma junção internacional de redes interconectadas autônomas que colaboram entre si (BRADNER, 1996). Ela suporta a comunicação de servidor para servidor através da adesão voluntária a procedimentos e protocolos abertos que são definidos por uma série de padrões. Há também redes isoladas, que não fazem parte da Internet global, mas também utilizam os mesmos padrões. Dentre os principais atores envolvidos com a criação dos padrões que possibilitam a governança na Internet estão o W3C e o IETF.

O *World Wide Web Consortium* (W3C) foi criado em 1994 e constitui uma comunidade internacional para a criação de padrões para a *web*. Seu papel é trabalhar com a evolução técnica da Internet e garantir sua interoperabilidade.

Dentre os padrões relevantes criados pela W3C estão o *HyperText Markup Language* - HTML (1995), o *eXtensible Markup Language* - XML (1996). A partir de 2001 deu início a atividades relacionadas à *web* semântica (W3C, 2012).

O *Internet Engineering Task Force* (IETF) é provavelmente a instituição de padronização com mais influência no desenvolvimento da tecnologia usada para a construção da Internet (ALVESTRAND e LIE, 2009). Foi criado em 1986 e é responsável por padrões tecnológicos relevantes como *Transmission Control Protocol* (TCP), *Internet Protocol* (IP), *Domain Name System* (DNS) e *Hypertext Transfer Protocol* (HTTP), dentre milhares de outros publicados. As organizações, tais como o IETF e o W3C, não trabalham sozinhas, mas em conjunto e com outros parceiros internacionais.

A Internet tem possibilitado uma troca de informações sem precedentes. Uma grande parte deste êxito se dá através da adoção de padrões comuns, pois esta simplifica o processo de desenvolvimento de aplicações, possibilita a visualização compatível de conteúdo em diversos sistemas e plataformas e minimiza a possibilidade de erros e garante uma integração mais segura e confiável. A seguir, alguns dos padrões e conceitos relevantes que formam a base tanto dos conceitos de IDEs como de dados interligados serão detalhados.

2.2. PADRÃO XML – LINGUAGEM DE MARCAÇÃO EXTENSÍVEL

Dentre os padrões mais importantes para a realização deste trabalho está o XML (*eXtensible Markup Language),* uma especificação do W3C usada para conter e administrar informação (RAY, 2009). É uma metalinguagem, que proporciona uma estrutura formal padronizada para qualquer documento sem especificar o vocabulário utilizado.

O XML não inclui uma interpretação específica dos dados. Essa característica, segundo Klein (2001) é tanto a força quanto a fraqueza da linguagem, uma vez que se pode estruturar todos os tipos de dados com uma sintaxe não ambígua, mas também não especifica a semântica destes dados. Os atores envolvidos na troca de dados usando XML precisam, de alguma forma, concordar previamente com o vocabulário, seu uso e significado.

Uma recomendação adicional da W3C se refere a esquemas XML. Este padrão ainda está em fase de "*candidate recommendation*", portanto, ainda pode passar por algumas modificações. É uma forma de definir o vocabulário utilizado nos documentos XML, de modo que fique documentada a sua estrutura. É possível, por exemplo, definir valores *default* para atributos, máscaras e várias regras para as instâncias que forem criadas com a utilização do esquema.

2.3. ARQUITETURA ORIENTADA A SERVIÇOS (SOA)

Uma das formas de se obter a interoperabilidade na Internet se dá através da arquitetura orientada a serviços *Service Oriented Architecture* (SOA), conforme ilustrado na FIGURA 10. Um serviço é então definido como uma unidade de trabalho realizado por um provedor de serviços para atingir os objetivos desejados pelo consumidor do serviço.

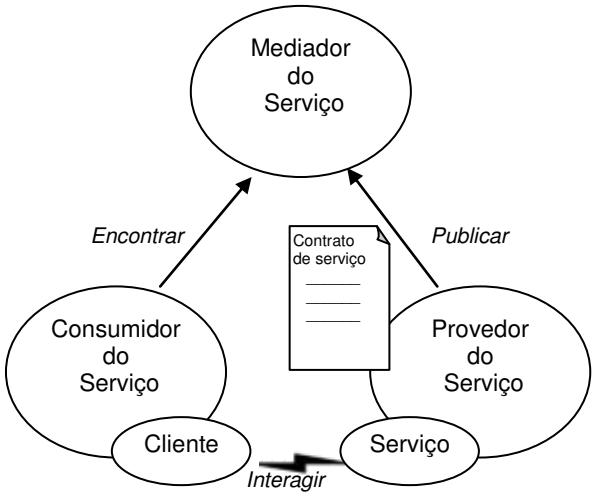


FIGURA 10 - DIAGRAMA DE INTERAÇÃO CLIENTE-PROVEDOR DE SERVIÇOS *WEB*.
FONTE: ADAPTADO DE W3.ORG

Na FIGURA 10 está esquematizada a interação de um provedor de serviços com o cliente. Para que o cliente possa encontrar o serviço que execute determinada tarefa, ele procura um intermediador de serviços, o qual funciona como um catálogo de publicações feitas pelos servidores disponíveis. Para que esta arquitetura atinja seus objetivos de interoperabilidade, é necessário observar duas restrições (HE, 2003):

- A existência de um conjunto pequeno de interfaces para todos os agentes de software participantes. Estas interfaces devem ser disponibilizadas universalmente para todos os provedores e consumidores através da utilização de padrões.
- Os serviços devem disponibilizar mensagens descritivas através de um esquema extensível de suas interfaces. O esquema limita o vocabulário e a estrutura das mensagens e permite que, por exemplo, novas versões do serviço possam ser introduzidas sem prejuízo às versões já existentes.

As interfaces devem usar protocolos da Internet como o *Hypertext Transfer Protocol* - HTTP (FIELDING *et al.*, 1999) ou *File Transfer Protocol* – FTP (POSTEL e REYNOLDS, 1985) e usar as mensagens no formato XML.

2.4. *WEB* SEMÂNTICA

O termo "*Web* Semântica" foi proposto por Berners-Lee, Hendler e Lassila, (2001) e significa a estruturação dos significados de conteúdo na *web*, fazendo uma extensão da *web* existente na época. A proposta é que a informação tenha um significado bem definido, para que assim, máquinas e usuários possam trabalhar de forma cooperativa. Esta visão utilizou-se da pesquisa existente na área de Engenharia do Conhecimento (*Knowledge Engineering*), definida por Feigenbaum e Mccorduck, (1983) como: "a disciplina que envolve a integração de conhecimento em sistemas computacionais para resolver problemas complexos que normalmente requerem um alto nível de perícia humana".

Uma das maneiras de se representar o conhecimento de uma forma estruturada é com o uso de ontologias. As ontologias permitem a descrição de conceitos e possibilitam um entendimento a respeito dos significados destes e seus relacionamentos por parte dos agentes envolvidos. Como o modelo cognitivo do indivíduo é produto da sua cultura, linguagem, educação e experiência de vida (STOCK, 2008), então, mesmo numa mesma comunidade de informação, as visões do mundo diferem.

2.5. DIAGRAMA DE TECNOLOGIAS SEMÂNTICAS (*SEMANTIC WEB STACK)*

As tecnologias envolvidas na concretização da visão da *web* semântica são representadas num diagrama proposto por Berners-Lee (2000) e que tem evoluído com a tecnologia disponível desde então. Na FIGURA 11 encontra-se a versão mais atualizada do diagrama denominado de *Semantic Web Stack*.

Dentre os diversos componentes do diagrama, algumas tecnologias estão mais maduras, com padrões criados e em operação, enquanto outras ainda estão sendo discutidas pela comunidade e têm necessidade de aperfeiçoamento. A cada nova versão de uma das tecnologias envolvidas, as demais também são afetadas e rediscutidas, em um processo contínuo e dinâmico. Na base do diagrama estão os componentes iniciais da interoperabilidade: a codificação Unicode (ISO/IEC 10646:2012) para conjuntos de caracteres em linguagens diversas e a organização de recursos na *web* via URI (*Unified Resource Identifier).* Em seguida, no papel de harmonização sintática está a linguagem XML, descrita anteriormente. O intercâmbio de dados se dá através do padrão RDF (*Resource Description Framework),* e as taxonomias correspondentes se utilizam de esquemas RDFS (*Resource Description Framework Schema)*. Para as ontologias, a linguagem padrão corrente é OWL (*Web Ontology Language).* Há ainda a linguagem para o intercâmbio de regras RIF *(Rule Interchange Format)* e o padrão SPARQL (*SPARQL Protocol and RDF Query Language*). A camada seguinte corresponde à Lógica Unificadora, na qual se processam as inferências baseadas nas regras já definidas, através de motores de raciocínio ou *reasoners.* A etapa de aprovação depende da explicação lógica para os passos usados nas inferências e, finalmente, a autenticação diz respeito ao reconhecimento das fontes e da confiabilidade das informações derivadas. Acompanhando todas as fases, está a possibilidade de criptografia das informações.

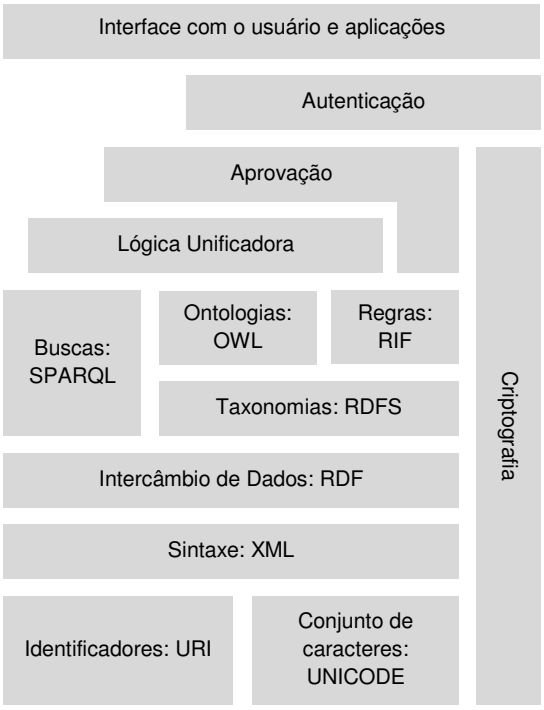


FIGURA 11 - DIAGRAMA DE TECNOLOGIAS SEMÂNTICAS
FONTE: ADAPTADO DE SIGNER, 2012 E BRATT, 2007

Na sequência, estão descritas as codificações, formatos e padrões que formam o diagrama:

- **UNICODE -** O Unicode (ISO/IEC 10646:2012) é um padrão internacional mantido por um consórcio específico, que trata da representação, codificação e manipulação de caracteres na maioria dos sistemas de escrita existentes no mundo (UNICODE, 2012). O padrão, que começou a ser desenvolvido em 1987, contém hoje 110.181 caracteres e é pré-requisito para aplicações semânticas em múltiplas linguagens.

- **URI e REFERÊNCIAS URI -** Os identificadores unificados de recursos (URIs) são centrais para aplicações semânticas (Berners-Lee, Fielding e Masinter, 2005). O fato de ser unificado garante que sejam seguidos padrões conhecidos para sua criação. Como recursos, entende-se todo o tipo de conceito que possa ser identificado, desde um documento, um arquivo, um site, um serviço, ou um conjunto de outros arquivos. O recurso não precisa necessariamente ser acessível via *web*, pode ser um ser humano, objeto, instituição, conceito abstrato ou mesmo números e operadores matemáticos. Como identificador, entende-se o conjunto de informações necessário para individualizar aquele recurso no universo de recursos possíveis.

Referências URI (URIRefs) são usadas para representar através de URIs caminhos relativos de determinadas porções de um mesmo esquema lógico, ou um fragmento de um recurso. O fragmento fica separado do restante da URI pelo sinal #.

- ***RESOURCE DESCRIPTION FRAMEWORK* (RDF) -** O RDF é um padrão para intercâmbio de dados na *Web* mantido pelo W3C. Ele se estrutura em declarações no formato sujeito-predicado-objeto ou recurso-propriedade-valor (conhecidos como triplas) que permitem o compartilhamento de dados em diversas aplicações através da *Web*. Os nós, como são chamados cada elemento das triplas, podem ser representados por um URIRef, estar em branco, ou ser um texto.

- ***RESOURCE DESCRIPTION FRAMEWORK SCHEMA* (RDFS) -** O RDF possibilita declarar entidades e seus relacionamentos, mas o esquema RDFS é o que permite registrar semanticamente a definição das classes e propriedades utilizadas. O RDFS permite, por exemplo, a definição do tipo de dados que podem ser instanciados em uma classe, o domínio dos valores possíveis, comentários, definição de subclasse, entre outras possibilidades.

Usando o RDF em conjunto com o RDFS, é possível criar classes e subclasses, propriedades e subpropriedades, e associá-los entre si. As classes se referem a conceitos gerais e os indivíduos ou instâncias são os membros destes conjuntos.

- ***WEB ONTOLOGY LANGUAGE* (OWL) -** Os padrões RDF e RDFS possibilitam as funcionalidades básicas detalhadas anteriormente para a descrição de recursos. No entanto, definições mais complexas como cardinalidade, ou número máximo de ocorrências para determinada classe, combinação de classes para criação de outras classes, entre outros, exigem recursos mais avançados. O padrão OWL - *Web Ontology Language* estende as capacidades dos vocabulários dos padrões anteriores. O OWL é uma linguagem para definir e instanciar ontologias na *Web* (SMITH, WELTY, e MCGUINNESS, 2004). No final de 2009, esta especificação foi aprimorada, e como consequência foi lançado o OWL 2, uma expansão e revisão dos conceitos da especificação anterior, de 2004.

- ***REASONERS* ou MECANISMOS DE INFERÊNCIAS -** *Reasoners* são mecanismos computacionais criados para se realizar inferências lógicas a partir de

um conjunto de fatos ou axiomas (DILLI, 2009). Eles funcionam com base nas regras estabelecidas previamente na ontologia. O *reasoner* também ajuda a buscar inconsistências, redundâncias e inferir relacionamentos, de forma a ajudar na manutenção da ontologia e gerar novos conhecimentos quando as ontologias são utilizadas. A maioria dos *reasoners* utiliza o OWL para realizar o processamento das inferências (SANTOS, SOARES e MATOS, 2008).

- **SPARQL -** SPARQL é uma linguagem de consulta para triplas RDF (PRUD'HOMMEAUX e SEABORNE, 2008). Ela funciona tanto para dados armazenados em RDF quanto para os que são vistos como RDF através de algum *middleware*, ou software intermediário.

A busca consiste de algumas partes:

- Declaração de prefixo, para abreviar as referências URI.
- Definição de qual conjunto de dados será objeto da busca. Existem *endpoints* (sites que executam buscas SPARQL) genéricos, que aceitam qualquer conjunto de dados; e específicos, que apenas aceitam determinados conjuntos. Pode ser usada a cláusula FROM.
- Cláusula de resultado, que especifica quais informações aparecerão no resultado.
- Para execução de buscas, usa-se a cláusula SELECT e o filtro WHERE. Para que essas buscas resultem em novas triplas de forma permanente, usa-se a cláusula CONSTRUCT.
- Modificadores, que reagrupam ou ordenam os resultados (ORDER BY, LIMIT, etc.)

As variáveis na linguagem SPARQL começam com símbolo ? e podem ser associadas a qualquer nó (texto ou recurso) do RDF pesquisado. Padrões de triplas são como triplas comuns, mas com a diferença de que cada uma das partes pode ser substituída por uma variável. Se for usada a clausula SELECT *, todas as variáveis mencionadas na busca serão selecionadas.

2.6. DADOS INTERLIGADOS (*LINKED DATA*)

Apesar de todo o arcabouço tecnológico descrito nos itens anteriores, a realização da *Web* de Dados requer um esforço ainda maior por parte da comunidade para garantir que a informação seja compartilhada e utilizada de forma adequada. Neste contexto surgiu o conceito de Dados Interligados (*Linked Data).*

A *Web* de Dados pode ser considerada uma camada adicional da *web* tradicional, genérica, que pode conter qualquer tipo de dados e é aberta à publicação. Com a conexão de diversas bases através de *links* RDF, forma-se uma interconexão global de dados que permite a descoberta de novas bases de dados. Este dados são auto-descritivos, usam os padrões existentes para a *web* semântica e são separados de sua forma de apresentação e formatação.

Uma recomendação a ser seguida é a padronização das URIs e reutilização de termos já definidos em vocabulários disponíveis, dos quais alguns estão descritos na lista abaixo. Na FIGURA 12 observa-se de forma gráfica a relação entre estes vocabulários, apontando quais fazem referência a termos de outros.

- FOAF - *Friend-of-a-Friend*: termos para descrição de pessoas e redes sociais
- SIOC - *Semantically-Interlinked Online Communities*: possui uma ontologia para comunidades *online* como grupos de *e-mail*, *blogs*, *wikis*, etc.
- DOAP - *Description of a Project*: vocabulário para descrição de projetos de software, em especial de software livre.
- *Dublin Core ®*: Metadados de uso geral, para descrição de objetos tais como: sons, imagens, textos e *sites*.
- *Review Vocabulary* : termos para representação de resenhas.
- *GoodRelations*: vocabulário profissional para *e-commerce*.
- *Music Ontology*: termos para descrição de artistas, álbuns faixas, shows, etc.
- *Organization Ontology:* descreve a estrutura das organizações.

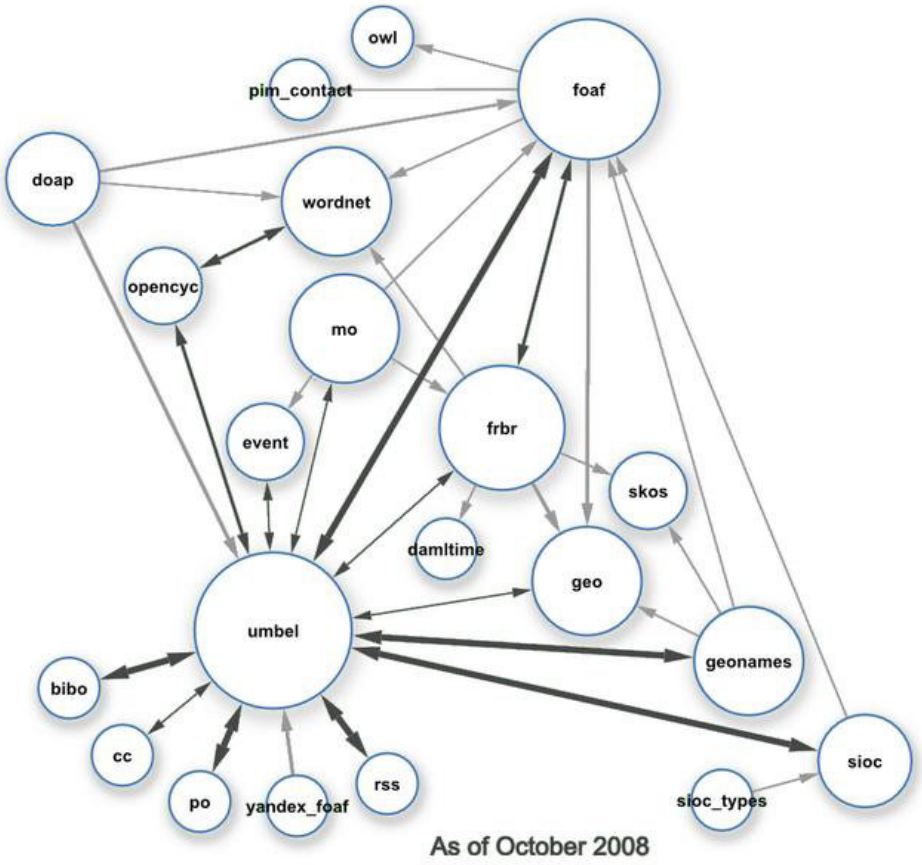


FIGURA 12 - RELAÇÃO ENTRE VOCABULÁRIOS EXISTENTES.
FONTE: CYGANIAK e JENTZSCH, 2011

Para fazer a conexão entre bases de dados diversas, é uma prática comum usar a propriedade owl:sameAs para se estabelecer que duas URIs em bases de dados distintas na realidade, apontam para o mesmo recurso, fazendo então uma ponte entre as duas publicações.

Um exemplo de iniciativa, no tópico de dado interligados, é o DBpedia, um projeto comunitário que visa extrair conteúdo estruturado da Wikipedia (BIZER, LEHMANN *et al.*, 2009). Existe um subgrupo trabalhando com a versão em português, a qual se chama DBpedia-PT, com pesquisadores de diversas universidades brasileiras. O conjunto de dados da Dbpedia (BECKER, 2009) possui identificadores para 2,6 milhões de objetos, em 80 linguagens. Dentre eles, tem-se pelo menos 213 mil pessoas, 328 mil lugares e 274 milhões de triplas no total. Estão incluídas também 609 mil figuras e 4,9 milhões de ligações para outros conjuntos de dados interligados.

Outros exemplos são o Projeto Gutemberg, de criação e distribuição de ebooks gratuitos, que conta hoje com cerca de 40 mil títulos e possui seu acervo

para consulta em RDF; além do jornal NY Times, que a partir de 2009 começou a publicar, em formato RDF, seu acervo que cobre mais de 150 anos de dados.

2.7. DADOS INTERLIGADOS ABERTOS GOVERNAMENTAIS

2.7.1. Governo Aberto e Transparência

Os governos sempre foram grandes produtores e consumidores de dados, seja por conta de seus procedimentos de rotina ou por determinação legal. “Dados Governamentais Abertos” é o termo que determina a “disponibilização, através da Internet, de informações e dados governamentais de domínio público para a livre utilização pela sociedade” (AGUNE, GREGORIO FILHO e BOLLIGER, 2010). Esta prática se baseia no princípio de que as pessoas têm o direito à informação sobre os documentos e procedimentos do governo. Esta percepção enquanto direito do cidadão remonta ao Iluminismo, e no Brasil, o direito de acesso à informação pública foi previsto na Constituição Federal de 1988, no inciso XXXIII do Capítulo I - dos Direitos e Deveres Individuais e Coletivos - que dispõe que:

> “todos têm direito a receber dos órgãos públicos informações de seu interesse particular, ou de interesse coletivo ou geral, que serão prestadas no prazo da lei, sob pena de responsabilidade, ressalvadas aquelas cujo sigilo seja imprescindível à segurança da sociedade e do Estado”.

O conceito fundamental é que dados públicos pertencem ao cidadão que, por consequência, deve ter acesso irrestrito às informações governamentais (DINIZ, 2009). Esta interpretação foi sendo difundida ao longo do tempo, inclusive pela influência da comunidade de Software Livre, criando na sociedade uma expectativa maior de responsabilidade social e transparência por parte de órgãos do governo.

Atualmente, novas tecnologias permitem não apenas uma comunicação mais ampla, mas também que os dados sejam gerenciados e reutilizados de forma efetiva. Como colocado no manifesto assinado por Berners-Lee e Shadbolt (2010) no jornal britânico *The Guardian*: “um número maior de pessoas tem ideias

inovadoras e habilidade para usar dados. E organizações, do governo à iniciativa privada, de grupos voluntários à mídia, estão famintos por dados."

A Parceria de Governo Aberto, uma organização internacional de adesão voluntária, foi criada em 2011 com oito países fundadores: Brasil, Indonésia, México, Noruega, Filipinas, África do Sul, Reino Unido e Estados Unidos. Seu principal objetivo é assegurar o compromisso dos governos em promover a transparência e o "empoderamento" dos cidadãos, lutar contra a corrupção e fortalecer novas tecnologias para reforçar a governança (OPEN GOVERNMENT PARTNERSHIP, 2012).

No entanto, não basta que as informações sejam compartilhadas, mas que esta ação aconteça de forma a atingir de fato os objetivos de gestão transparente e reuso das informações. Para isso, em 2007, um grupo de especialistas denominado OpenGovData criou os oito princípios dos dados governamentais abertos (OPEN GOVERNMENT WORKING GROUP, 2007), que estabelecem que eles devem ser:

1. Completos: Todos os dados públicos estão disponíveis. Um dado público é o dado que não está sujeito a limitações válidas de privacidade, segurança ou privilégios de acesso.

2. Primários: os dados são disponibilizados tal como os coletados na fonte, com o maior nível possível de granularidade e sem agregação ou modificação.

3. Atuais: os dados são colocados à disposição tão rapidamente quanto necessário para preservar o seu valor.

4. Acessíveis: os dados estão disponíveis para o maior escopo possível de usuários e para o maior escopo possível de finalidades.

5. Processáveis por máquinas: os dados são estruturados para permitir processamento automatizado.

6. Não-discriminatórios: os dados estão disponíveis para todos, sem necessidade de cadastro.

7. Não-proprietários: os dados estão disponíveis em um formato sobre o qual nenhuma entidade tem controle exclusivo.

8. Livres de licenças: os dados não estão sujeitos a nenhuma regulação de direitos autorais, patentes, propriedade intelectual ou segredo industrial. Restrições

sensatas relacionadas à privacidade, segurança e privilégios de acesso podem ser permitidas.

Os Dados Interligados Abertos, descritos nas sessões anteriores, formam a plataforma tecnológica e de boas práticas para a publicação dos Dados Governamentais Abertos, que uma vez acessados em seu estado bruto, servem como base para diversas aplicações que podem ser construídas, tanto pelo próprio governo, quanto pela sociedade.

2.7.2. Dados Interligados Abertos no Brasil

No Brasil, os esforços para garantir o acesso à informação remontam à Constituição de 1988. No âmbito da Administração Pública Federal, a Secretaria de Logística e Tecnologia da Informação (SLTI) do Ministério do Planejamento, Orçamento e Gestão é responsável pelas normas relacionadas ao uso da Tecnologia da Informação. As iniciativas têm início no ano de 2003 com a e-PING (Padrões de Interoperabilidade de Governo Eletrônico) e foram reforçadas com a adoção da Lei de Acesso à Informação em 2011 e criação na Infraestrutura Nacional de Dados Abertos (INDA) em 2012.

2.7.2.1. e-PING – Padrões de Interoperabilidade de Governo Eletrônico

Uma das principais ações da área está no estabelecimento e contínua atualização da e-PING (Padrões de Interoperabilidade de Governo Eletrônico), cujas primeiras ações iniciaram no ano de 2003 tendo sido estabelecido oficialmente pela Portaria SLTI/MP nº 5, de 14 de julho de 2005.

A e-PING define um conjunto mínimo de premissas, políticas e especificações técnicas que regulamentam a utilização da Tecnologia de Informação e Comunicação (TIC) na interoperabilidade de serviços de Governo Eletrônico e estabelece as condições de interação com os demais poderes e esferas de governo e com a sociedade em geral (BRASIL, 2012)

Para os órgãos do poder executivo do governo federal, a adoção dos padrões e políticas contidos na e-PING é obrigatória. As normas, além de preconizarem o alinhamento com a Internet e a adoção de padrões internacionais, preferencialmente abertos, além da priorização de uso de software livre, também tem uma dimensão semântica com três principais ações:

- Desenvolvimento e manutenção de ontologias e outros recursos de organização da informação;
- Desenvolvimento e adoção de um padrão de modelagem de dados para Governo;
- Desenvolvimento e adoção de uma política de disseminação de dados e informações.

A e-PING define o RDF e o OWL como padrões recomendados e o SPARQL como padrão em estudo.

Uma parte importante da e-PING é o e-VoG - Vocabulários e Ontologias do Governo Eletrônico, um conjunto de padrões, ferramentas e metodologias cuja finalidade é possibilitar: o intercâmbio de informações com acordo semântico; o uso de metodologias de modelagem conceitual nas áreas de negócio de governo; o uso de ontologias como ferramenta para explicitar conhecimentos de maneira formal e coerente; e o alinhamento conceitual das diversas áreas do conhecimento do governo (BRASIL, 2012). O e-VoG é constituído de duas partes, o Vocabulário Controlado do Governo Eletrônico (VCGE) e do Padrão de Metadados do Governo Eletrônico (e-PMG).

2.7.2.2. Lei de Acesso à Informação

A legislação sobre acesso à informação e transparência foi sendo fortalecida ao longo do tempo. O Decreto 6.932, de 11 de agosto de 2009, por exemplo, trouxe diversas diretrizes importantes para melhorar a prestação de serviços ao cidadão, tais como: o compartilhamento de informações entre órgãos públicos; a inversão do ônus da prova para informações comprobatórias do cidadão; e a carta de serviços ao cidadão. Destaca-se que o referido Decreto tem diversas interfaces com o e-

PING, pois suas diretrizes são fortemente ancoradas em sistemas de informação e na interoperabilidade entre eles.

Posteriormente tem-se a Lei 12.527, de 18 de novembro de 2011, ou Lei de Acesso à Informação, que inova em termos internacionais por determinar não somente a chamada transparência passiva, ou seja, quando a administração disponibiliza ao cidadão a informação solicitada, mas também a transparência ativa, com a publicação padronizada de dados interligados abertos prontos para o processamento por máquinas e reuso. Ficam subordinados a essa lei:

> "I – os órgãos públicos integrantes da administração direta dos Poderes Executivo, Legislativo, incluindo as Cortes de Contas, e Judiciário e do Ministério Público;
>
> II – as autarquias, as fundações públicas, as empresas públicas, as sociedades de economia mista e demais entidades controladas direta ou indiretamente pela União, Estados, Distrito Federal e Municípios."

Nela são assegurados princípios importantes, como a observância da publicidade como preceito geral e do sigilo como exceção, e o desenvolvimento do controle social da administração pública.

2.7.2.3. Infraestrutura Nacional de Dados Abertos

Como consequência da adesão do Brasil aos compromissos da Pareceria Governo Aberto, em setembro de 2011, um decreto S/N institui o Plano de Ação Nacional sobre Governo Aberto. Nele consta o compromisso do governo de implantar a Infraestrutura Nacional de Dados Abertos (INDA), entre outros objetivos, tais como:

- Estruturar e coordenar a política de dados abertos;
- Definir e disciplinar os padrões e os aspectos técnicos referentes à disponibilização e disseminação de dados;
- Fornecer suporte e capacitação aos órgãos e entidades que aderirem à INDA, além de disponibilizar tecnologia e apoio às ações;

- Promover a colaboração entre governos dos diferentes níveis da federação e entre o Poder Executivo federal e a sociedade.

Em abril de 2012, a Instrução Normativa nº4, da SLTI, institui a INDA, e estabelece seus objetivos, Comitê Gestor e um prazo de cento e cinquenta dias para a criação do seu Plano de Ação (BRASIL, 2012a), que foi lançado em setembro do mesmo ano. O Comitê Gestor, definido na Instrução Normativa, é composto por membros de instituições diversas do governo federal, além de membros da sociedade civil e setor acadêmico. A instância coube, além da gestão da INDA, a criação do Plano de Ação e a definição e articulação de Grupos de Trabalho, aos quais estão atribuídas tarefas específicas.

Dentre outras interações com políticas da área, o Plano de Ação cita a Infraestrutura Nacional de Dados Espaciais (INDE), a qual será descrita no capítulo 4 deste trabalho.

O plano também estabelece os princípios da disponibilização de dados na administração pública, que foram baseados nos adotados pelo Reino Unido. São eles:

> "I. A política de dados abertos será impulsionada pela sociedade civil (inclusive pelas empresas) e pelos interessados em utilizar os dados.
>
> II. Os dados públicos serão publicados de forma reutilizável e legível por máquina.
>
> III. Os dados públicos devem ser publicados sob licença aberta que permita a livre reutilização, incluindo a reutilização comercial.
>
> IV. Os dados públicos estarão disponíveis e fáceis de encontrar através de um único e fácil ponto de acesso *on-line* (dados.gov.br).
>
> V. Os dados públicos devem ser publicados com rapidez e precisão.
>
> VI. Publicar dados rapidamente, e então republicá-los como *Linked Data.*
>
> VII. Os dados públicos estarão disponíveis gratuitamente para uso em qualquer forma legal.
>
> VIII. Os Órgãos públicos devem incentivar ativamente o reuso de seus dados públicos."

Nos princípios destacam-se, além de disposições essenciais dos Dados Interligados Abertos, importantes definições como a gratuidade dos dados, a reutilização comercial e a criação do portal dados.gov.br. O item VI, no entanto, abre a possibilidade de que os dados sejam publicados no formato disponível e adia a publicação como *Linked Data*. É importante que as melhores práticas dos dados interligados abertos, que exigem adaptação tecnológica mais trabalhosa, sejam incentivadas e priorizadas, e não postergadas.

Dentre as ações previstas no plano estão: ações de disseminação; criação do Plano de Capacitação da INDA; avaliação das licenças existentes para Dados Abertos sob o ordenamento jurídico brasileiro; e oferta de tecnologias de suporte à publicação de dados abertos.

2.7.2.4. Portal Brasileiro de Dados Abertos

No Plano de Ação, a entrada na INDA, tanto pelos usuários quanto pelos fornecedores de dados, será feita pelo Portal Brasileiro de Dados Abertos (http://dados.gov.br), cuja interface pode ser visualizada na FIGURA 13. Esse funciona como um catálogo de metadados das informações disponíveis, pelo endereço, autor, formato e responsável pelo dado.

Neste momento, o fornecedor dos dados cadastra os metadados no portal e publica seus dados em ambiente próprio. Ainda não há informações semânticas ou ferramentas específicas para procura e interligação de dados no portal. O portal dispõe também de uma lista de aplicativos que usam dados publicados e material educativo sobre como executar a publicação.

FIGURA 13 - TELA DO SITE DADOS.GOV.BR – PESQUISA POR DADOS DO IBGE.
FONTE: WWW.DADOS.GOV.BR, 2013

Ding *et al. (*2011) criaram um sistema de classificação dos estágios de publicação de dados governamentais abertos. O primeiro é o de catálogo, no qual é criado um inventário de dados disponíveis *online*. Em seguida, acontece a fase de extração de dados, na qual uma cópia do conjunto de dados, em determinado momento no tempo, é usada como entrada para conversão para dados interligados abertos. Neste caso, os metadados do conjunto de dados apontam para a fonte original. O terceiro estágio é o da conversão, com a criação de configurações e a criação de uma versão interligada do dado original, automaticamente. No momento, o portal dados.gov.br se encontra no primeiro estágio, funcionando primordialmente como catálogo de dados disponíveis, sem informações ainda convertidas em formato de dados interligados abertos.

3. *WEB* SEMÂNTICA E DADOS GEOESPACIAIS

Dificuldades com interoperabilidade de aplicações geoespaciais são relatadas desde a década de 1990 (BISHR, 1998; DEVOGELE, PARENT E SPACCAPIETRA,1998 e YUAN, 1997). Com a necessidade da integração cada vez maior de aplicações e bases de dados, a comunidade se reuniu em instituições que se debruçaram para a criação de diversos padrões que permitiram as soluções que conhecemos atualmente. Grande parte dessas normas apresentam soluções que utilizam a arquitetura baseada em serviços (SOA), que proporciona uma forma flexível de montagem de cadeias de serviços com interfaces padronizadas, sem que haja conhecimento exato das características por trás de cada *interface*.

Neste capítulo também é tratada a questão da pesquisa na área de semântica e informação geográfica e são relatadas experiências de países na criação de ontologias para o mapeamento topográfico de referência. Também são destacados os conceitos e iniciativas relacionados à informação geoespacial na abordagem de dados interligados abertos, que tem sido objeto de interesse particular, por ter na sua própria constituição um potencial agregador de conhecimento sobre fenômenos.

3.1. INFORMAÇÃO GEOESPACIAL, INTEROPERABILIDADE E SERVIÇOS

Como definido por Câmara (2005), o problema fundamental da Geoinformação é a "produção de representações computacionais do espaço geográfico". O processo para essas representações é ilustrado pelo Paradigma dos Quatro Universos, conforme FIGURA 14:

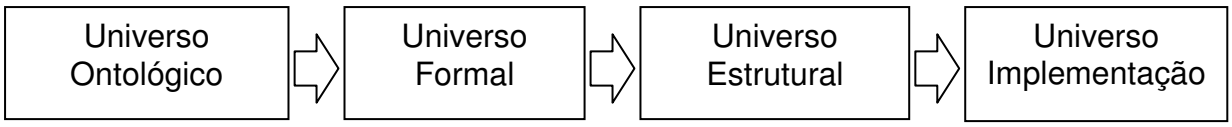


FIGURA 14 - PARADIGMA DOS QUATRO UNIVERSOS
FONTE: CÂMARA, 2005

O Universo Ontológico compreende os conceitos da realidade a representar computacionalmente e, portanto, é o primeiro nível de abstração. Em muitos

Sistemas de Informação, esta etapa não é formalizada. O Universo Formal corresponde às abstrações formais necessárias, como modelos de dados e modelos conceituais. O Universo Estrutural é composto dos algoritmos necessários à implementação computacional, à conversão e ao armazenamento dos dados e seus relacionamentos topológicos. Por fim, no Universo de Implementação executa-se a conversão dos modelos anteriores em implementações computacionais.

Com o crescimento do uso de Sistemas de Informação Geográfica (SIG) na década de 1990, foi identificada na comunidade a demanda para a troca de dados entre sistemas. Bishr (1998) identificou o potencial dos avanços na área de tecnologia da informação e as demandas crescentes dos usuários de SIG como o início do compartilhamento e transferência de dados entre as ilhas que formavam os primeiros sistemas existentes. No entanto, estudos como os de Devogele, Parent e Spaccapietra (1998) apontaram o reuso de dados SIG como um “pesadelo”, devido à existência de pouca documentação, heterogeneidade de modelos conceituais, diferentes formatos de codificação dos dados, estruturas de armazenamento incompatíveis, entre outros problemas.

A partir de então se inicia a ênfase ao conceito de interoperabilidade, definida como a propriedade de “possibilitar o compartilhamento e a troca de informação e processos em ambientes computacionais heterogêneos, autônomos e distribuídos (YUAN, 1997). O mesmo autor coloca a interoperabilidade em SIG como muito mais desafiadora do que em outras áreas, por causa da complexidade nos processos de aquisição, representação e operação de dados geoespaciais. Essa complexidade é que leva a questões fundamentais relacionadas, na época, à incompatibilidade de representações, estruturas e semântica que devem ser tratadas para se obter a interoperabilidade da informação geográfica. Nota-se que, muitos anos antes do artigo que propõe a *web* semântica (BERNERS-LEE et al., 2001), a comunidade de informação geoespacial já tinha preocupações sobre este aspecto.

Foram identificados por Bishr (1998) três tipos de heterogeneidades:

- Heterogeneidade semântica: deve-se ao fato de que um mesmo objeto no mundo real pode ter mais de uma descrição no banco de dados, de acordo com a aplicação em diferentes disciplinas. É, portanto, referente à representação do Universo Ontológico em

diferentes universos do discurso. A primeira descrição deste tipo de heterogeneidade aparece em Worboys e Deen, 1991.

- Heterogeneidade esquemática: diferenças nos modelos de dados usados, considerando agregações e hierarquias diversas, na fase de representação do Universo Formal.
- Heterogeneidade sintática: cada base de dados pode ser implementada com o uso de diferentes paradigmas de bancos de dados, ou através de representações geométricas diversas, como estrutura matricial ou vetorial. Esta é uma questão relacionada aos Universos Estrutural ou de Implementação.

A utilização de serviços abertos como solução das heterogeneidades aparece como tema para pesquisa em Goodchild *et al.* (1996). Esta visão foi sendo consolidada através dos anos como a estrutura apresentada na FIGURA 15, na qual os serviços intermediam os processos entre as diversas bases de dados e as aplicações clientes.

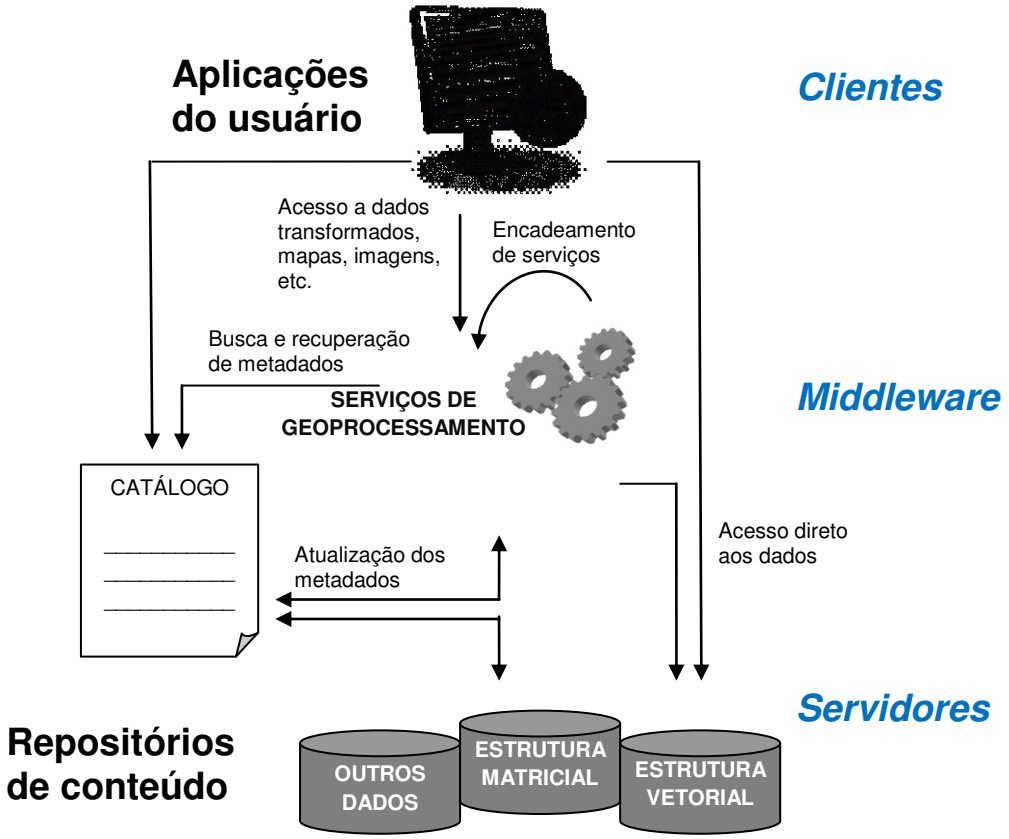


FIGURA 15 - SERVIÇOS EM SIG.
FONTE: ADAPTADO DE EVANS, 2003

3.2. INSTITUIÇÕES DE PADRONIZAÇÃO PARA A INFORMAÇÃO GEOESPACIAL

3.2.1. *Open Geospatial Consortium* (OGC)

O OGC tem origem na Fundação Open GRASS. O GRASS (*Geographical Resources Analysis Support System*) é o primeiro SIG de código aberto a chegar ao *status* de produção. Este *software* foi originalmente desenvolvido (entre 1982 e 1995) pelo exército americano e, desde então, pela comunidade internacional de desenvolvedores. A Fundação *Open GRASS* começou a operar em 1992. Em 1994 ela se transformou em Fundação *Open GIS* e, no mesmo ano, houve a alteração para o formato atual, de *Open Geospatial Consortium*, um consórcio internacional formado atualmente por 480 membros de empresas, instituições governamentais e universidades.

A visão do OGC é *"*criar um mundo no qual todos se beneficiem do uso da informação geoespacial e tecnologias associadas" (OGC, 2012). A instituição desenvolve padrões abertos *de facto*, isto é, criado pelo consenso de mercado e não por atribuição legal. Atualmente, existem grupos de trabalho do OGC para diversas questões, entre elas: Qualidade de Dados, Geossemântica, Gestão de Direitos Geoespaciais Digitais e Metadados.

3.2.2. ISO/TC211

A ISO (*International Standards Association)* é uma associação cujos membros são as instituições nacionais de padronização de cerca de 130 países, com o apoio de um escritório central localizado na Suíça (ISO, 2012). Os padrões são criados com base no consenso mútuo seguindo as seguintes fases: 1 – Proposta: a confirmação por meio dos votos dos membros, que um novo padrão é necessário; 2 – Preparação: através de um grupo de especialistas, é criado um anteprojeto de trabalho que é enviado ao comitê para apreciação; 3 – Comitê: é a fase de apresentação para comentários até que se obtenha consenso sobre o anteprojeto; 4 – Inquérito: a versão inicial da proposta de padrão é posta em

circulação pelos membros para votação e comentários durante cinco meses; 5 – Aprovação: a versão final da proposta do padrão é enviada para um voto positivo ou negativo por um período de dois meses; e 6 – Publicação: o padrão é publicado.

Dentre os vários comitês técnicos que discutem áreas de trabalho específicas, foi fundado, em 1994, o ISO/TC211 (ISO/TC211, 2012). O objetivo deste comitê é estabelecer um conjunto de padrões para a informação geográfica que especificam a infraestrutura e serviços necessários para a manipulação de dados geográficos incluindo sua gestão, aquisição, processamento, análise, acesso, apresentação e transferência. Atualmente o comitê possui 35 países membros, com 30 outros membros observadores, dos quais o Brasil não faz parte em nenhuma categoria.

Além dos membros, o ISO/TC211 possui várias conexões com outras organizações internacionais, entre elas: OGC, FIG (*International Federation of Surveyors)*, GSDI (*Global Spatial Data Infrastructure)*, IAG *(International Association of Geodesy)*, ICA (*International Cartographic Association*), IEEE *Geoscience and Remote Sensing Society* e ISPRS (*International Society for Photogrammetry and Remote Sensing*).

Os padrões produzidos pelo comitê formam a série ISO 19000.

3.2.3. *Joint Advisory Group* (JAG) – OGC e ISO/TC211

Os padrões ISO/TC têm uma visão de prazo mais longo e são mais abstratos, enquanto os padrões do OGC são mais voltados para questões tecnológicas e de implementação. Como têm uma certa sobreposição do escopo de suas atividades, as duas organizações trabalham em conjunto, possuindo um grupo especial para esta coordenação, denominiado JAG (*Joint Advisory Group).*

Um acordo de cooperação mútua existe desde 1998 (ISO/TC211, 2009a). Em função deste acordo, o OGC adotou vários padrões ISO/TC211 como especificações gerais nas quais são baseados os padrões de implementação. Na FIGURA 16, apresentam-se as inter-relações entre o OGC, o ISO/TC211 e as demais organizações de padronização da Internet. Enquanto a ISO trata de questões de domínio, como conteúdo e vocabulário, e tem um mandato legal mais

claro outorgado pelos países que formam a organização, e portanto classificados como "*De Jure*", no outro lado do espectro estão as organizações como W3C e IETF, cujas normas são discutidas com a comunidade, formando padrões de adoção voluntária por que acabam se fortalecendo por exigência do mercado, sendo assim padrões "*De Facto*". Estes também são voltados para questões de infraestrutura como formatos e linguagens para construção e troca de informações. Situando-se em uma posição intermediária, o OGC instancia os padrões de domínio de forma compatível com a infraestrutura previamente definida, formando uma interligação entre os padrões existentes e tornando possível sua implementação.

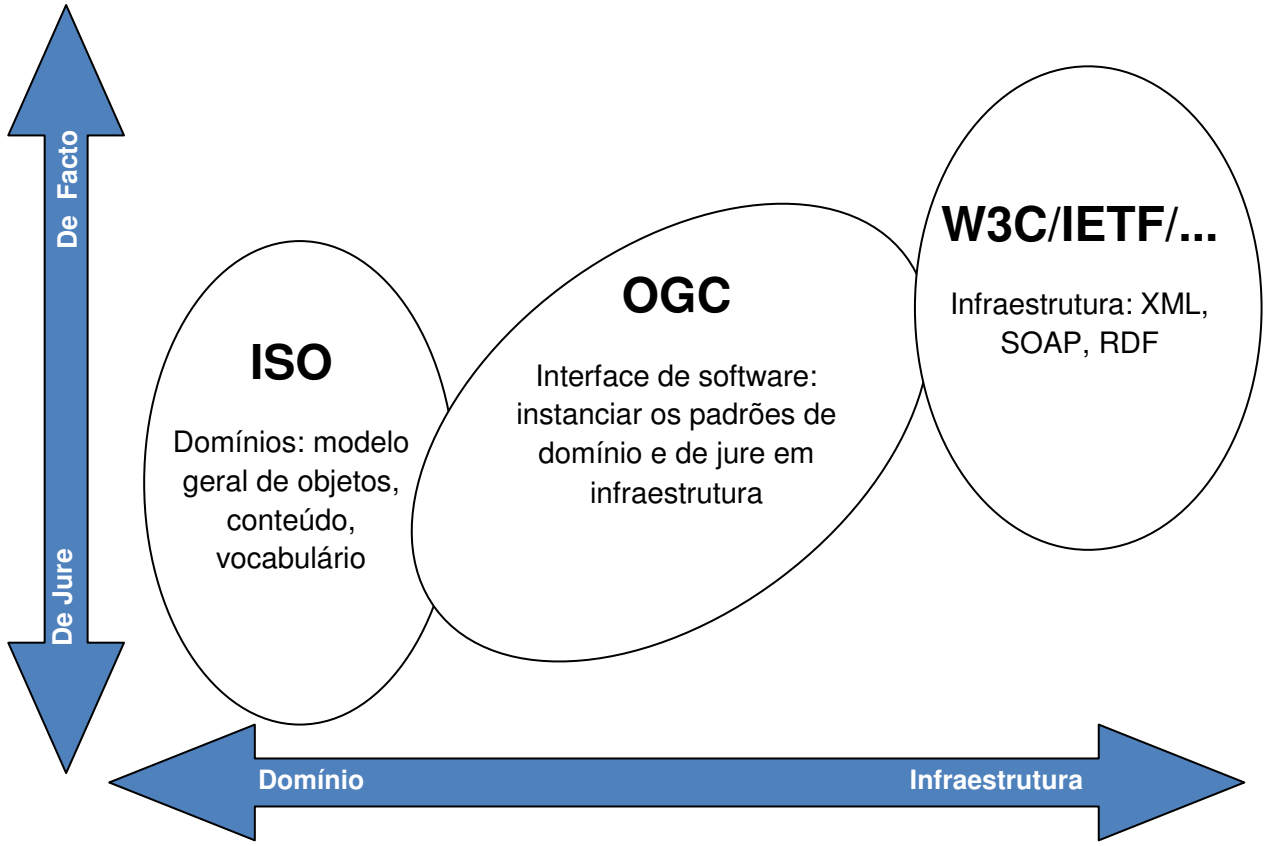


FIGURA 16 - RELAÇÃO ENTRE AS INSTITUIÇÕES DE PADRONIZAÇÃO NA ÁREA GEOESPACIAL E NA INTERNET.
FONTE: ADAPTADO DE RAMAGE E REED, 2012

3.3. PADRÕES EXISTENTES

Nesta sessão se encontram os principais padrões internacionais existentes para informação geográfica relacionados a este trabalho. Os padrões semânticos

internacionais existentes para a informação geográfica são tratados adiante neste mesmo capítulo, e os padrões brasileiros, no capítulo 04.

3.3.1. Metadados - ISO 19115 e ISO 19139

Este conjunto de normas tem como objetivo estabelecer a estrutura para descrever a informação geográfica. Nos metadados, ou dados sobre os dados, encontram-se as informações necessárias para descobrir, acessar, utilizar adequadamente e transferir dados espaciais.

Historicamente, os padrões para metadados têm origem nos anos 1980s, nos Estados Unidos, com o desenvolvimento do SDTS (*Spatial Data Transfer Standard*) e, mais tarde, em 1994, com a aprovação pelo FGDC (*Federal Geographic Data Comitee*) do CSDGM – *Content Standard for Digital Geospatial Metadata* (TSOU, 2002).

Para adaptar o padrão de metadados com a realidade de determinado domínio ou país são criados perfis, respeitando as informações obrigatórias, mas com a possibilidade de criar traduções, designar como obrigatórios itens opcionais e criar novos domínios para o preenchimento de atributos.

A norma ISO 19139:2007 *Geographic information - Metadata XML schema implementation* (ISO/TC211, 2007a) estabeleceu o esquema XML para implementação computacional da norma ISO 19115:2003 (ISO/TC211, 2003), especificando o formato de registro que pode ser usado para descrever, validar e compartilhar metadados.

Em 2009, a norma ISO 19115 recebeu uma segunda parte (ISO 19115:2- *Geographic information – Metadata – Part 2: Extensions for imagery and gridded data* - ISO/TC211, 2009), na qual são definidos os elementos de metadados específicos para imagem e dados em estrutura matricial, com suas particularidades. A norma ISO 19139 foi também acrescida de uma segunda parte em 2012 para o esquema XML que possibilite a implementação da extensão para dados matriciais.

3.3.2. Bancos de Dados Geográficos, SQL, SFA e WKT (ISO 19125)

A SQL (*Structured Query Language* ou Linguagem de Consulta Estruturada), norma ISO/IEC 9075-1:2008, hoje usada pela maioria dos Sistemas Gerenciadores de Bancos de Dados existentes (FERREIRA *et al.*, 2005), não comporta originalmente dados e funções espaciais. Para isso o OGC teve a iniciativa de criar uma norma específica para a área. A norma OpenGIS "*Geographic information — Simple feature Access*" (SFA), também homologada como ISO 19125 (OGC, 2006) é dividida em duas partes: a primeira define a arquitetura e a segunda detalha a opção SQL. A norma se baseia em feições simples, formadas por entidades geométricas de ponto, curva e superfície em duas ou três dimensões. As feições são armazenadas em tabelas nas quais a parte não espacial é tratada pelas normas já definidas para a SQL, enquanto a parte geométrica é armazenada em colunas específicas cujas regras são definidas pela norma do OGC. A coluna espacial é associada a um sistema de referência espacial, também descrita na primeira parte da norma, na qual são especificados atributos, métodos e operações com geometrias. Também definida por esta norma está a linguagem WKT (*Well-known text*), uma linguagem de marcação de texto para representar a geometria de objetos vetoriais que é usada atualmente nos Bancos de Dados Geográficos como o PostGIS.

A segunda parte da norma ISO 19125 (OGC, 2006a), define um esquema SQL que suporta operações de armazenamento, recuperação, busca e atualização de coleções de feições através da interface de nível de chamada SQL definida pela norma ISO/IEC 9075-3:2003. A partir dessa norma, critérios espaciais podem ser definidos nas cláusulas de busca SQL, podendo retornar geometrias no lugar de apenas resultados tabulares. Alguns exemplos de métodos espaciais são: ST_Disjoint, ST_Intersects e ST_Touches, para as operações topológicas Disjuntas, Intersecção e Toca, respectivamente.

3.3.3. *Geography Markup Language* - GML (ISO 19136)

Publicada em 2007, a norma ISO 19136 (ISO/TC211, 2007) padroniza uma codificação XML em conformidade com a norma ISO 19118, que determina a codificação de dados geográficos com o uso de diagramas em UML (*Unified*

Modeling Language). A GML inclui tanto propriedades espaciais como não espaciais das informações geográficas.

Como é baseado em XML, o padrão proporciona um arcabouço aberto e independente da indústria de software para o armazenamento de informação geográfica. Assim, este permite perfis que tratem de subconjuntos do padrão GML, a descrição de esquemas de aplicações geoespaciais para comunidades e domínios específicos e a possibilidade de troca de esquemas entre organizações.

3.3.4. *Web Map Service Interface* - WMS (ISO 19128)

Este padrão (ISO/TC211, 2005) especifica um serviço que produz dinamicamente mapas georreferenciados. O serviço disponibiliza uma renderização do mapa solicitado em forma de figura, e não feições ou dados matriciais em si. O padrão define um “mapa”, neste contexto, como uma representação da informação geográfica na forma de uma imagem digital adequada à visualização em uma tela de computador. Os mapas produzidos por serviços WMS são imagens no formato PNG, GIF ou JPEG, ocasionalmente em formato de figuras vetoriais como *Scalable Vector Graphics* (SVG).

São as três operações principais realizadas por um WMS:

- *GetCapabilities*, que retorna os metadados do serviço;
- *GetMap*, que retorna o mapa como base em parâmetros geográficos e de representação definidos; e
- *GetFeatureInfo*, que retorna os atributos sobre uma feição em particular mostrada no mapa.

Se mapas com o parâmetro de extensão geográfica iguais forem criados com a utilização de formato de imagem com suporte a transparência (como GIF), estes podem ser tratados como camadas sobrepostas. Estas camadas podem ser provenientes de serviços fornecidos por servidores distintos.

A questão da simbologia das feições é tratada por uma especificação a parte, a OGC *Styled Layer Descriptor* (SLD). Esta norma permite que simbologias pré-definidas possam ser criada e compartilhadas em formato XML.

3.3.5. *Web Feature Service Interface* - WFS(ISO 19142)

A norma *Web Feature Service* (WFS) representa uma mudança na forma como a informação geográfica é criada, modificada e compartilhada na Internet (ISO/TC211, 2010). No lugar da simples troca de arquivos, o serviço oferece acesso à informação ao nível de feições e suas propriedades. Os dados podem ser recuperados ou modificados usando um protocolo HTTP, no formato GML (*Geography Markup Language*).

As principais operações disponíveis para o serviço WFS são:

- *GetCapabilities*, que retorna os metadados do serviço;
- *DescribeFeatureType*, que recupera o esquema XML da feição;
- *GetFeature*, que executa a obtenção dos dados em si. Parâmetros como o retângulo envolvente e outros devem ser informados e o serviço retornará o arquivo dos dados em formato GML com as geometrias e atributos pertinentes.

3.3.6. *Filter Encoding* - FES (ISO 19143)

O *Filter Encoding* (ISO/TC211, 2010b) é também definido em XML e foi projetado para ser originalmente parte da norma WFS, mas posteriormente foi definido em um documento à parte para que pudesse ser usado como filtro em outros serviços, como o *Web Coverage Service*, *Gazetteer* e *Web Registries*.

Um filtro é usado para identificar um subconjunto de recursos de uma coleção cujos valores das propriedades satisfaçam uma série de predicados conectados de forma lógica (ISO/TC211, 2010b). Se o valor da propriedade satisfizer todos os predicados do filtro, então esse recurso é considerado parte do subconjunto resultado. As buscas feitas usando o padrão FES podem ser transformadas para consultas SQL e vice-versa (GEOCONNECTIONS, 2012), bastando uma reestruturação entre os termos.

3.4. HISTÓRICO DA PESQUISA EM GEOSSEMÂNTICA

Apesar do rápido desenvolvimento dos padrões e das tecnologias de serviços *web* ter, sem dúvida, ajudado o compartilhamento de informação geoespacial entre diversas fontes, a interoperabilidade resultante não pôde resolver os problemas de heterogeneidade semântica (ZHAO *et al.*, 2009).

Como indicado por Cai (2007), o conhecimento espacial humano é incompleto, tendencioso, vago e, algumas vezes, inconsistente. Graças a esta condição, tanto a abstração das classes, que gera a modelagem, quanto os critérios da busca, são muito dependentes do contexto e da escala de representação. O conceito de proximidade, por exemplo, pode ser dependente do meio de transporte. Em uma busca: "Quais são os restaurantes próximos ao teatro?", os resultados próximos são dependentes da forma de execução do percurso, se a pé ou de carro, e em que horário, por exemplo. Outro artigo fundamental é o publicado por Bennett, (2001): "*What is a Forest*", que examinou em detalhe o quão vago é um conceito geográfico aparentemente simples como "Floresta", e chama a atenção para a necessidade de utilização de uma visão filosófica do problema, ainda que não seja possível definir claramente a forma de tratar estas questões.

Dada a complexidade de se retratar o espaço geográfico computacionalmente, foi natural a expansão das pesquisas no campo da semântica para os dados espaciais. Dentre os trabalhos pioneiros que destacam problemas semânticos na informação geográfica está o de Worboys e Deen (1991) que já previa as vantagens de dados distribuídos em "nós" locais (antevendo a utilização da *web*) e definia heterogeneidade semântica como o "desacordo a respeito do significado, interpretação ou uso pretendido para o dado", dividida em dois grupos:

- Genérica: dados em modelo vetorial ou matricial, por exemplo – que vêm se mostrando menos importante depois dos diversos esforços do setor em aspectos de interoperabilidade de dados e serviços geográficos;
- Contextual: causada por conceituações diferentes utilizadas em cada nó – que é mais similar ao significado atual do termo.

Wache *et al.* (2001) percebem três divisões principais nos conflitos semânticos:

- Conflitos de confusão: ocorrem quando dois itens parecem ter o mesmo significado, mas na realidade são diferentes;
- Conflitos de escala: ocorrem quando diferentes sistemas de referência são usados para a medição de um valor, como por exemplo, diferentes escalas geográficas;
- Conflitos de nome: ocorrem quando nomes em esquemas de informação diferem significativamente, como por exemplo, a presença de sinônimos.

Desde o início da utilização do termo no âmbito da Inteligência Artificial, a semântica se tornou também um tópico importante de pesquisa na comunidade da Ciência da Informação Geográfica, em aplicações como a do processamento da informação (recuperação ou integração), feito com base em significado no lugar de "meras palavras-chave" (FONSECA e SHETH, 2002). Câmara (2005) também destaca a dificuldade do compartilhamento de informações semânticas nos Sistemas de Informação Geográfica atuais, com a falta da inclusão de um conjunto de conceitos com um conjunto de relações semânticas, o que ele chama de Geo-Ontologias.

Em Winter (2001) encontra-se uma divisão das pesquisas na área em duas linhas: uma estuda o projeto de ontologias no desenvolvimento de sistemas, e representa o conhecimento de diversas áreas; enquanto a outra se concentra na construção de aplicações interoperáveis e aborda a integração de serviços, buscas, bancos de dados, padrões, entre outros.

No primeiro grupo encontram-se estudos como o de Ahlqvist (2005) que escolhe o problema de interoperabilidade semântica entre classificações de uso e cobertura do solo, e utiliza a avaliação de diversos aspectos da similaridade semântica entre classes de dois sistemas de classificação distintos (*US Natural Vegetation Classification Standard* e *European CORINE Land Cover*). O resultado é uma correspondência entre as duas classificações.

A pesquisa em aplicações interoperáveis semanticamente possui diversas frentes: Lindroos *et al.* (2007) propuseram a utilização de um servidor ontológico para buscar por nomes de lugares na Finlândia; Fallahi, Mesgari e Tehran (2006) estudaram a aplicação semântica de ontologias para geoserviços e Hess, Iochpe e Castano (2006) trabalharam na pesquisa de possíveis formas de integração entre

geo-ontologias diversas. Estudando uma forma semântica de desenvolvimento de portais geográficos, Athanasis *et al.* (2009) sugerem uma organização ontológica de metadados, pois pela própria natureza distribuída dos portais, ainda é um desafio a busca por informação, atualmente baseada em palavras-chave temáticas, espaciais ou temporais, isto cria dificuldades para o usuário. No estudo, eles explicitam os metadados usando RDF e também propõem formas de melhorar a interatividade da interface com o usuário.

No Brasil, Frozza e Mello (2006) propõem formas de determinação de equivalência semântica a partir do estabelecimento de esquemas no formato GML. Viegas e Soares (2006) estudaram formas de acesso a banco de dados geográficos sem que o usuário tenha conhecimento de sua estrutura interna, usando ontologias. Na pesquisa foi criado um protótipo para a Área de Proteção Ambiental dos Corais, no Rio Grande do Norte, no qual as buscas dos usuários em uma interface *web* utilizavam um tradutor Jena para serem repassadas, em formato SQL, para um banco de dados integrando ontologias feitas sob o ponto de vista dos biólogos, geólogos e turistas. Apesar de ser um exemplo bastante restrito, exemplifica bem a arquitetura proposta de criar interfaces semânticas para buscas. De Sousa e Leite (2005) também criaram ontologias para formalizar conhecimento de quatro áreas de um SIG específico: aspectos fisiográficos, geologia, mineração e socioeconomia.

Há também quem questione o uso de ontologias para dados espaciais, como Fernández e Fernández (2009) que destacam que esta ainda não é a solução ideal para se abordar a questão semântica e, portanto, não se deve abandonar a busca por melhores alternativas para refletir mais fielmente as variadas formas que os seres humanos pensam o mundo. Cai (2007) também questiona o uso da abordagem ontológica, que segundo ele, falha em gerar os resultados esperados no domínio geográfico. Segundo o mesmo autor, a abordagem, que é tida como panaceia para todos os problemas semânticos geoespaciais, é questionável ao se verificar a validade das suas premissas quando aplicadas para o domínio espacial, pois os conceitos geográficos são vagos e muito dependentes do contexto de utilização. Para superar esses problemas, ele cria ferramentas que incluem o contexto na manipulação de informação espacial semântica.

O mapeamento topográfico, fundamental como ferramenta para o desenvolvimento das demais aplicações da cartografia, pode ter sua utilização

facilitada e expandida com a adoção de conceitos semânticos. No entanto, Varanka (2008) ressalta a dificuldade de se estabelecer o escopo para definições em programas de mapeamento sistemático nacional, dada a natureza ampla e flexível dos mapas topográficos como mapas básicos para múltiplos e diversos usos e propósitos. Como também reforça Dolbear, Hart e Goodwin (2006) o produtor da ontologia fica na posição difícil de estabelecer definições sem conhecimento sobre o usuário final, pela própria natureza do Mapeamento Topográfico. Outra característica importante deste tipo de mapeamento, a integração de diversas escalas, também pode ser beneficiada com a utilização de Ontologias, como mostra Uitermark *et al.* (2005).

A seguir estão descritas as experiências de três países que já possuem ações específicas de integração de propriedades semânticas aos seus mapas topográficos: os Estados Unidos, o Reino Unido e a França.

3.4.1. Estados Unidos

Nos Estados Unidos, o USGS (*United States Geological Survey*), órgão responsável pelo mapeamento topográfico nacional, instituiu um projeto chamado "*Ontology for The National Map*" para estabelecer semântica de feições com o objetivo de melhorar os modelos de dados atuais. O projeto inclui *workshops* com especialistas, formalização de conceitos existentes nos padrões já estabelecidos e principalmente um grande esforço no levantamento e descrição de termos e nomes geográficos presentes no mapeamento (VARANKA, 2008; VARANKA, 2009; e VARANKA, 2009a). Foram criadas taxonomias para seis categorias de feições: terreno, águas superficiais, regimes ecológicos, áreas construídas, limites e eventos.

3.4.2. Reino Unido

O *Ordnance Survey*, instituição responsável pelo mapeamento topográfico no Reino Unido, também possui uma linha de pesquisa exclusiva para o campo de Semântica, denominado *GeoSemantics*. Um dos produtos das pesquisas é a criação do *Rabbit* (HART, JOHNSON e DOLBEAR, 2008), uma forma de inglês estruturado

para que os especialistas de cada área possam escrever ontologias sem conhecimentos específicos em OWL.

Como nos Estados Unidos, no caso inglês também foram realizados testes na execução de buscas semânticas no Banco de Dados (DOLBEAR e HART, 2006) A ferramenta Jena também foi escolhida como elemento integrador das partes, com uma preocupação especial na velocidade das buscas, uma vez que o Banco de Dados do *Ordnance Survey* (em formato *Oracle*) é o maior banco de dados espaciais vetoriais do mundo, segundo o artigo.

Existe, portanto, um esforço consolidado na formalização de Ontologias, de forma que, já em 2006, estava concluída a etapa referente ao tema Hidrografia, com 301 classes e 162 propriedades, representando aproximadamente 10% do tamanho estimado da Ontologia Topográfica Completa (DOLBEAR, HART e GOODWIN, 2006).

3.4.3. França

Na França, a pesquisa sobre Ontologias aplicadas ao Mapeamento Topográfico é realizada no IGN (*Institut national de l'information géographique et forestière*) pelo laboratório COGIT (*Conception Objet et Généralisation de l'Information Topographique)*, dentro do tópico Infraestrutura de Dados Geográficos.

Pressionado pelas demandas do projeto Europeu INSPIRE e por iniciativas governamentais de acesso múltiplo a diversas plataformas de dados espaciais, o IGN reconheceu um grande número de demandas de alta complexidade, principalmente com integração de bancos de dados diversos. Foi criado então o Projeto GéOnto (MUSTIERE *et al.*, 2009), que entre outros objetos de pesquisa procura criar ontologias para os dois bancos de dados principais de informações topográficas na França. O BD CARTO é um banco de dados geográfico de referência usado de escalas departamentais (1:50.000) a escalas regionais (1:250.000). O BD TOPO é o banco de dados geográfico tridimensional de referência com vetores de precisão métrica. O IGN (ABADIE e MUSTIERE, 2010) também usou as especificações existentes para os bancos de dados, em formato de

texto, para então extrair semi-automaticamente ontologias como um experimento de extrair informações semânticas do material disponível.

3.5. INFORMAÇÃO GEOESPACIAL E DADOS INTERLIGADOS

Analisando dados interligados abertos, Kulk e Van Loenen (2012) identificaram que a grande maioria desses dados se refere a dados geográficos. Não apenas este tipo de informação está presente em uma grande quantidade de bases de dados interligados, como tratado no capítulo 1 deste trabalho, mas estes dados têm a importante função de atuar como um fator de conexão entre diversas fontes, e podem contribuir com uma parte significativa da sustentação da *web* semântica (GOODWIN, DOLBEAR e HART, 2008).

A questão vem sendo abordada de forma diversa pela comunidade de dados interligados abertos e pela comunidade da informação geoespacial. A seguir, abordam-se alguns padrões existentes para o assunto e também algumas experiências desenvolvidas.

3.5.1. Vocabulário Básico RDF GEO (W3C)

Este foi o primeiro padrão desenvolvido para informação geoespacial na *web* semântica, proposto em 2003 pelo Grupo de Interesse em *Web* Semântica do W3C (W3C, 2003). É um vocabulário muito simples, que permite a descrição somente de pontos com latitude, longitude e, opcionalmente, altitude, sempre usando o Datum WGS 84.

3.5.2. *Ontology* – Parte 1 e 2 (ISO 19150)

O ISO/TC211 criou um grupo, em 2006, para pesquisar como a visão da *Web* Semântica poderia beneficiar o desenvolvimento de informação geoespacial interoperável (LOPEZ-PELLICER *et al.*, 2012). Este grupo recomendou, em 2009, que fosse revisto o modelo de referência do TC/211, além de ter participado da

proposição das normas ISO 19150 – Parte 1 - *Ontology – Framework* (ISO/TC211, 2012b) e Parte 2 - *Rules for developing ontologies in the Web Ontology Language*. A parte 1 foi publicada em novembro de 2012 e a parte 2 ainda está em revisão. Na primeira parte está definida a estrutura para interoperabilidade semântica da informação geográfica. Nesta estrutura, está um modelo de alto nível dos componentes necessários para manipulação semântica nos padrões ISO para informação geográfica com o uso de ontologias. A segunda parte definirá regras e orientações para o desenvolvimento de ontologias. Também estabelecerá a conversão dos elementos de diagramas de classes em UML, usados nos padrões ISO para informação geográfica, para OWL. Em maio de 2012, um grupo de especialistas em semântica foi convidado a apresentar um relatório sobre dados interligados que foi apresentado na 32ª plenária do ISO/TC211 em Delft, Holanda (LOPEZ-PELLICER *et al.*, 2012). No relatório, foram identificados alguns desafios, entre eles:

- Identificação: os padrões ISO/TC211 devem ser revisados para considerar o uso de URIs para identificar conceitos, feições e metadados. Além disso, como essas URIs requerem estabilidade e durabilidade, a governança das URIs deve também ser considerada, com regras bem definidas.
- Transformação de modelos: a ISO/TC211 deve definir vocabulários para suportar a descrição em RDF de dados e metadados geográficos, além de uma metodologia para transformação de modelos existentes para RDF e vice-versa, usando ontologias OWL e regras RIF.

3.5.3. GeoSPARQL

Em setembro de 2012, o OGC lançou a especificação GeoSPARQL, uma linguagem de busca geográfica para dados RDF. Ela contém três partes principais (BATTLE e KOLAS, 2012):

- A definição de um vocabulário que representa feições, geometrias e seus relacionamentos.

- Um conjunto de funções espaciais para ser usada em buscas SPARQL.
- Um conjunto de regras para transformação de buscas.

Pela norma, cada objeto espacial é formado por uma feição (*Feature)* e opcionalmente por uma ou mais geometrias, como por exemplo, um objeto pode ser representado por um polígono em uma escala e um ponto em outra. Esta relação está representada na FIGURA 17.

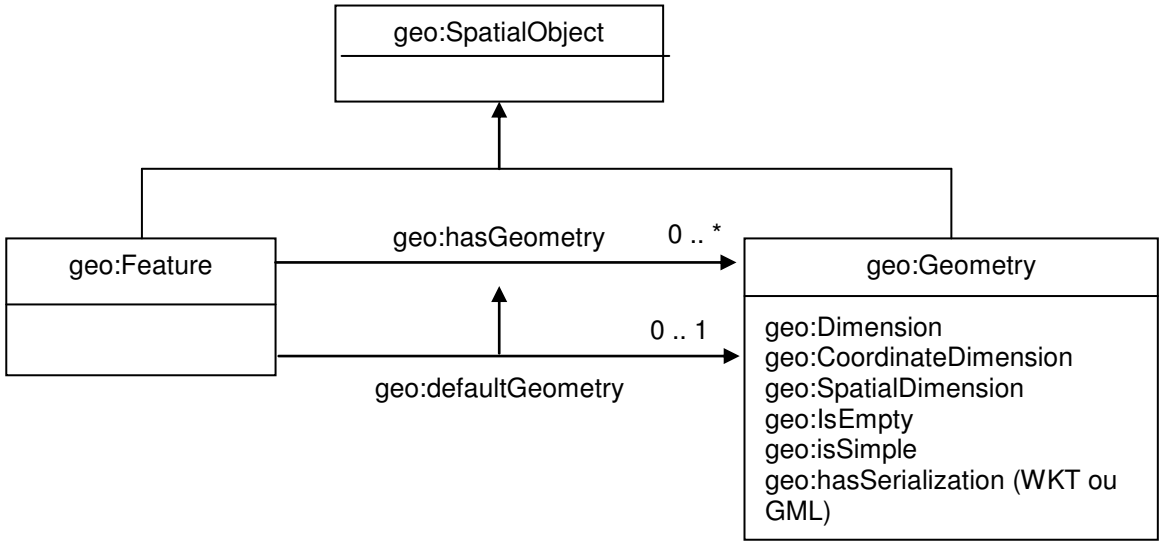


FIGURA 17 - RELACIONAMENTO ENTRE FEIÇÕES, OBJETOS ESPACIAIS E GEOMETRIA.
FONTE: OGC, 2012a

Em termos de relacionamentos topológicos, estão disponíveis três famílias ou grupos de relacionamentos consagrados: Egenhofer, RCC8 e *Simple Feature*. Para este último, por exemplo, as opções estão demonstradas na FIGURA 18. Sendo dois objetos, A e B, a função *geo:overlaps*, por exemplo, verifica se os objetos se sobrepõem.

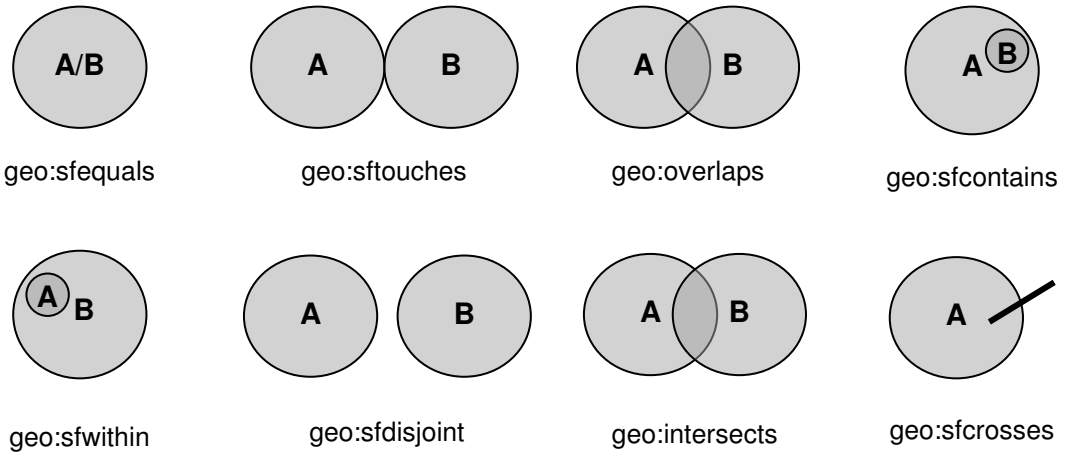


FIGURA 18 - RELACIONAMENTOS DISPONÍVELS PARA FAMÍLIA *SIMPLE FEATURE* NA ESPECIFICAÇÃO GEOSPARQL.
FONTE: OGC, a

Existe também a possibilidade de usar uma série de funções espaciais, que ficam definidas pelo prefixo “geof:”, descritas na TABELA 2. Por exemplo, a função *geof:distance* tem como parâmetros duas geometrias e uma unidade de medida, e retorna um valor numérico que é a distância entre elas. As demais funções na tabela incluem *buffer* de uma geometria, intersecção entre duas geométricas, entre outras. Na segunda coluna temos os parâmetros a serem fornecidos e a terceira nos dá o objeto que é resultado da função.

TABELA 2 - FUNÇÕES ESPACIAS DO GEOSPARQL

FUNÇÃO	PARÂMETROS	RESULTADO
geof:distance	Geom1, Geom2, unitsURI	xsd:double
geof:buffer	Geom1, radius, unitsURI	Geometry literal
geof:convexHull	Geom1	Geometry literal
geof:intersection	Geom1, Geom2	Geometry literal
geof:union	Geom1, Geom2	Geometry literal
geof:difference	Geom1, Geom2	Geometry literal
geof:symDifference	Geom1, Geom2	Geometry literal
geof:envelope	Geom1	Geometry literal
geof:boundary	Geom1	Geometry literal
geof:getsrid	Geom1	SRID of literal

FONTE :OGC, 2012a

3.6. EXPERIÊNCIAS EM DADOS GEOESPACIAIS INTERLIGADOS ABERTOS

3.6.1. Reino Unido – *UK LOCATION*

O Reino Unido decidiu abrir a informação geográfica, particularmente a fornecida pelo *Ordnance Survey,* para, entre outros objetivos, cumprir com diretivas da União Europeia no quesito de intercâmbio de informações, optando pelos Dados Interligados para cumprir os requisitos estabelecidos (SHERIDAN e TENNISON, 2010). O trabalho envolve, em especial, envolve as seguintes questões:

- A correspondência entre coisas do mundo real e os dados espaciais que as representam.
- A representação de fenômenos como fronteiras que podem tanto variar ao longo do tempo quanto serem disponibilizadas em diferentes resoluções.
- A representação de geometrias em RDF, tanto como recursos como conjunto de caracteres.

São três as bases de dados atualmente em conversão para o formato de dados interligados (GOODWIN, 2012):

- *Boundary-Line*: dados vetoriais de unidades administrativas lançados sobre a base 1:50.000 do país.
- *Code-Point Open*: pontos, originalmente em formato CSV, que retratam os códigos postais.
- *1:50.000 Scale Gazetteer*: pontos de mais de 250 mil nomes de lugares, provenientes do mapeamento topográfico na escala 1:50000;

Os passos usados para publicação dos dados acima (GOODWIN, DOLBEAR e HART, 2008) são:

- Criação de ontologias em formato OWL, conforme descrito no item 5.1.2.;
- Criação dos RDFs e URIs – foram criadas para o *Boundary-Line* 161 mil triplas;

- Criação de consultas, uma vez que a publicação inicial se deu antes da publicação da norma GEOSPARQL.

Além do *Ordnance Survey*, as seguintes instituições também publicam dados georreferenciados no site data.gov.uk, através do *UK Location Programme,* sendo que, em parênteses, está descrito o número de conjuntos de dados publicados:

- *British Geological Survey (369)*
- *Centre for Ecology & Hydrology (284)*
- *Northern Ireland Spatial Data Infrastructure (147)*
- *Scottish Government Spatial Data Infrastructure (112)*
- *Barrow Borough Council (89)*
- *Environment Agency (39)*
- *Welsh Government Spatial Data Infrastructure (18)*
- *Welsh Government (13)*
- *British Waterways (12)*
- *English Heritage (10)*
- *Department for Environment, Food and Rural Affairs (10)*
- *Forestry Commission (9)*
- *Land & Property Services (8)*
- *Land Registry (2)*
- *Geoplace (2)*
- *Worcestershire County Council (1)*
- *Marine Environmental Data & Information Network (1)*

Na FIGURA 19, estão telas de um aplicativo para *smartphone* que sinaliza alerta de enchente com base nos dados da agência de proteção ambiental publicados como dados interligados:

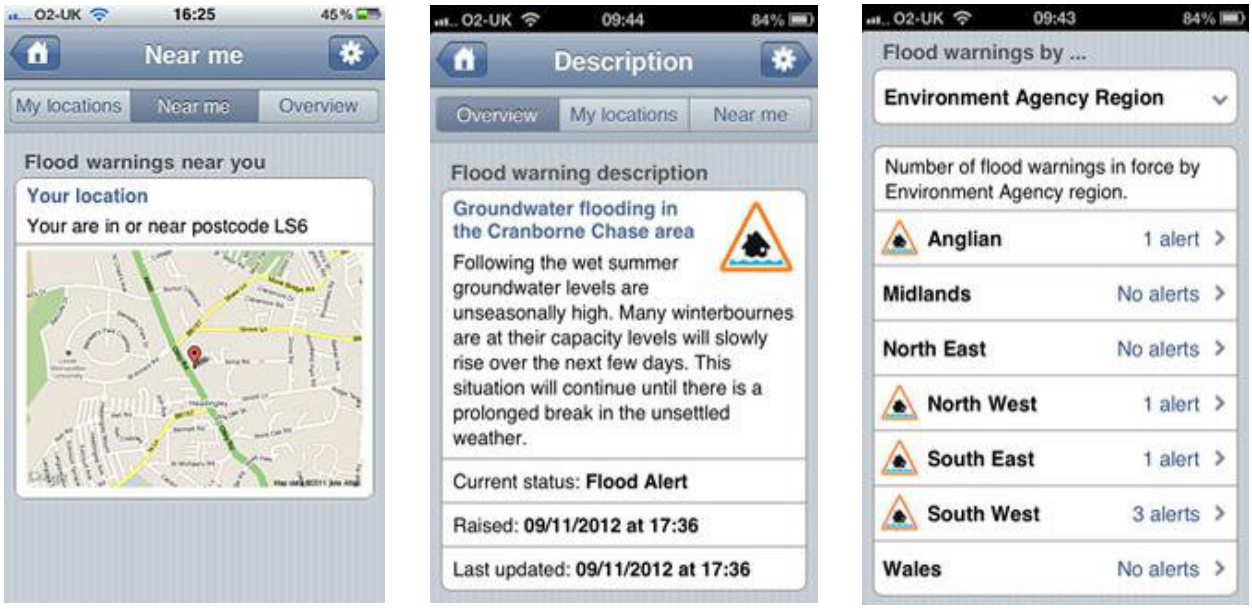


FIGURA 19 - APLICATIVO *FLOOD ALERT*
FONTE: HTTP://WWW.HALCROW.COM/FLOODALERT, 2013

3.6.2. Estados Unidos – geodata.gov

Desde 2005 o governo americano mantinha o site geodata.gov, um portal para mapas *web*, dados governamentais e também para geoserviços. Em 2011, o site geodata.gov foi migrado para o geo.data.gov, dentro do site data.gov. Esta alteração considerou as seguintes vantagens (GEOSPATIAL PLATFORM, 2011):

- Promover o reuso de padrões de arquitetura e tecnologia;
- Aumentar o acesso aos dados geoespaciais;
- Promover a comunicação governo-cidadão e a transparência;
- Facilitar a coordenação com outras atividades de TI da administração;
- Lançar um inventário de todos os dados, serviços e aplicações geográficas do governo;
- Funcionar como um braço operacional da Plataforma Geoespacial.

A base de dados do *The National Map* foi considerada a base primária para o processo de integração de informação geográfica com os dados interligados abertos nos Estados Unidos (USERY e VARANKA, 2012).

As principais etapas usadas na conversão dos arquivos existentes estão dispostas do diagrama da FIGURA 20:

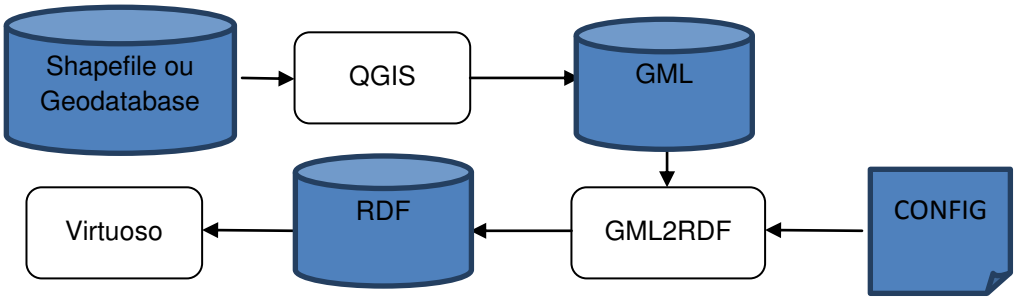


FIGURA 20 - ETAPAS PARA A PUBLICAÇÃO DOS DADOS
FONTE: Adaptado de BULEN, CARTER e VARANKA, 2011

Na FIGURA 20, inicialmente os dados em formato *shapefile* ou em banco de dados são transformados em GML pelo software livre QuantumGIS. Depois, um programa desenvolvido em Java pelo USGS, o GML2RDF, usa ferramentas da biblioteca *GeoTools* para converter os dados em RDF, e assim poderem ficar disponíveis para aplicações semânticas.

A TABELA 3 demonstra a quantidade de conjunto de dados já publicados:

TABELA 3 - QUANTIFICAÇÃO DE TRIPLAS PUBLICADAS

CONJUNTO DE DADOS	NÚMERO DE TRIPLAS	TAMANHO DO ARQUIVO
Hidrografia	20.000.000	2.7 Gb
Transportes	25.000.000	2.4 Gb
Limites	52.000	189 Mb
Estruturas	388.000	37 Mb

FONTE: USERY e VARANKA, 2012

Alguns dos desafios relatados no Manual do *Workshop* de Semântica e Tecnologia Espacial (USGS, 2012) foram:

- Recursos legados são ambíguos;
- Repositório de triplas acabam ficando muito volumosos;
- Padronização versus informação *bottom-up*;
- Formatação dos dados de entrada;

- Compromisso com os conceitos ou perspectivas múltiplas.

Dentre as vantagens da abordagem semântica destacadas no mesmo relatório estão a integração de dados de diversas fontes usando padrões já existentes como o USGS *Digital Line Graph* (DLG*), DLG-Enhanced*, *Spatial Data Transfer Standard* (SDTS), mas integrados no vocabulário semântico, e, portanto, dispensando novas formatações. E adicionalmente, tal abordagem possibilita ainda a construção de aplicações e usos dos dados não suportados pelos SIG e modelos existentes.

3.6.3. Espanha – *GeoLinked Data*

As informações abertas que foram escolhidas para ser disponibilizadas provém do INE (*Instituto Nacional de Estadistica*) e IGN (*Instituto Geográfico Nacional*), respectivamente fornecedoras das informações oficiais estatísticas e geográficas da Espanha (VILCHES-BLÁZQUEZ *et al.*, 2010). As bases de dados escolhidas foram o mapeamento sistemático 1:25.000, 1:200.000 e 1:1.000.000, bem como camadas de nomes geográficos em diversas escalas e Atlas.

O passo seguinte da escolha da base de dados foi a conversão dos dados GML em RDF. Para esse propósito, foi desenvolvido um software chamado GEOMETRYtoRDF, que cria as triplas usando a serialização WKT ou GML.

O arquivo gerado é manipulado então pelo GeoTools, a fim de recuperar a geometria e também para a conversão de coordenadas se necessário. Por fim, foi usada a ferramenta Jena para gerar o RDF final. Os resultados estão disponíveis no site geo.linkeddata.es, do qual um exemplo de busca na interface é mostrado na FIGURA 21.

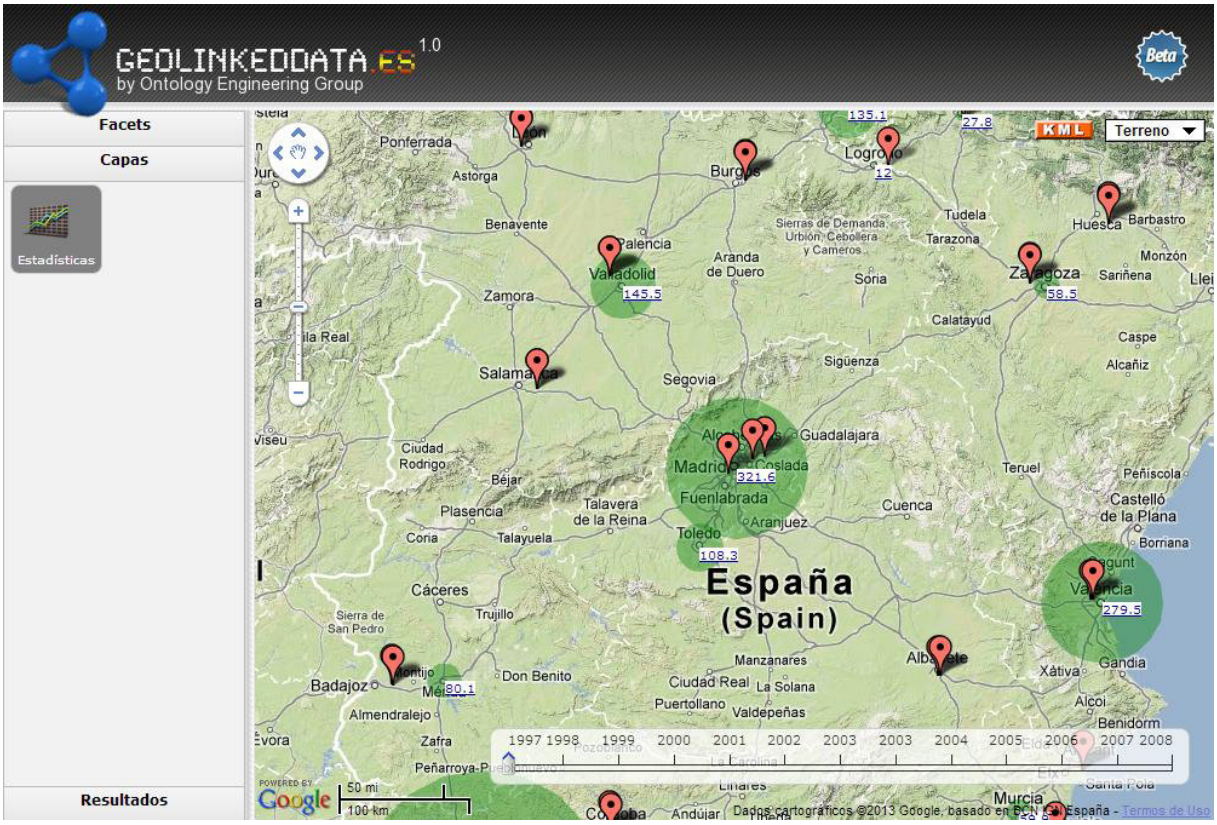


FIGURA 21 - TELA DO SITE GEO.LINKEDDATA.ES (2013)

3.6.4. GeoNames

O banco de dados *GeoNames* é um conjunto de mais de oito milhões de nomes geográficos, disponíveis para *download* gratuito. Todas as feições são categorizadas em nove classes e subcategorizadas em 645 códigos. O *site* responde a mais de 11 milhões de requisições diárias de *web services* (OGBUJI, 2008).

Apesar do GeoNames ser uma das mais ricas e mais usadas fontes de informação geoespacial em formato de dados interligados abertos, inclusive para os aplicativos do site dados.gov.br, a descrição dos metadados é bastante simples e não comporta, por exemplo, uma descrição detalhada da autoridade de quem forneceu a informação. Outra questão é que este suporta apenas dados pontuais, no Datum WGS 84. Na FIGURA 22 é ilustrado o resultado para uma pesquisa sobre a palavra “Curitiba”.

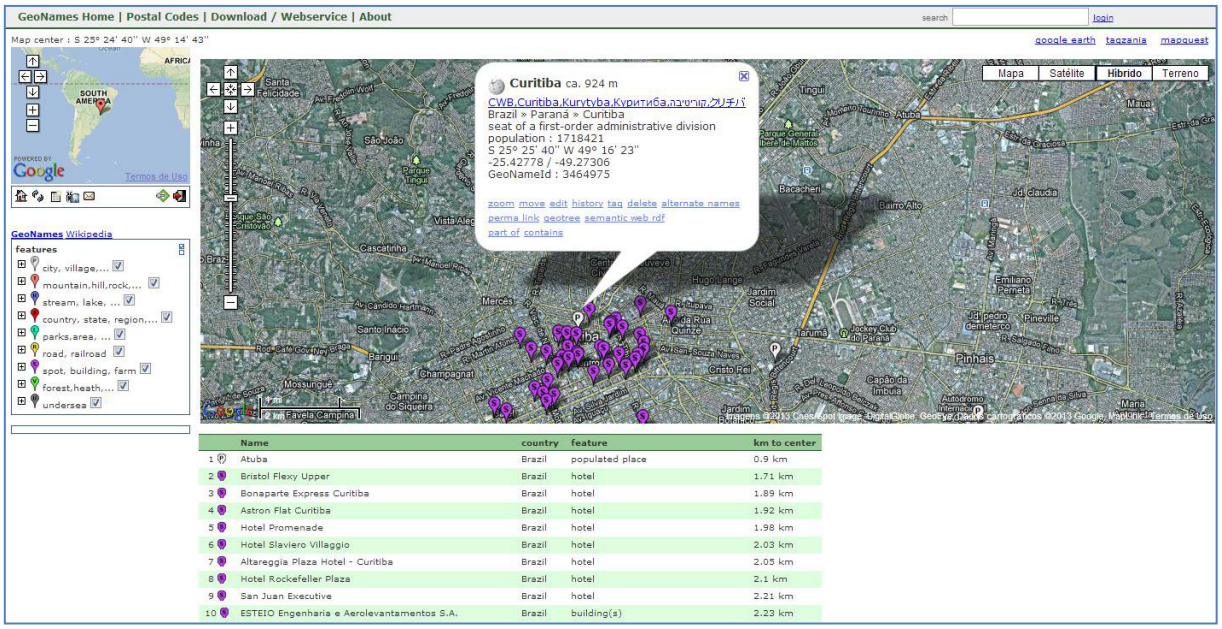


FIGURA 22 - TELA DO SITE GEONAMES.ORG

3.6.5. *LinkedGeoData (* baseado no *Open Street Map*)

O objetivo do projeto *LinkedGeoData* (LGD) é alçar o *Open Street Map* para a *Web* Semântica (STADLER *et al.*, 2012). O *Open Street Map* é um conhecido projeto no campo da informação geográfica voluntária (*Volunteered Geographic Information - VGI*), no qual dados geográficos relevantes são compilados por voluntários, salvos e disponibilizados a todos na Internet (NEIS, ZIELSTRA e ZIPF, 2011).

O *LinkedGeoData* contém mais de um bilhão de nós (dados pontuais) e 100 milhões de caminhos (dados lineares e poligonais) que resultam em, aproximadamente, 20 bilhões de triplas RDF. Os dados estão disponíveis de acordo com os princípios dos dados interligados e estão relacionados com os projetos DBpedia e GeoNames. Ainda não estão disponíveis os chamados relacionamentos, que são a forma de criação de feições complexas, como *multipolygons*.

Cada entidade no *Open Street Map* possui um identificador numérico chamado OSM ID e é descrita por um conjunto de caracteres chamado *tag*. É com base nestes *tags* padronizados que foi criada a ontologia na qual se baseia o LinkedGeoData, mostrada na FIGURA 23.

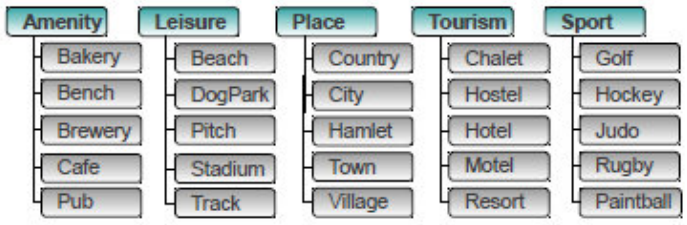


FIGURA 23 - EXTRATO DA ONTOLOGIA DO LINKEDGEODATA
FONTE: STADLER *et al.*, 2012

Para mostrar o potencial de aplicação dos dados estruturados de forma semântica, o site inclui um *browser* com o qual podem ser pesquisados valores textuais em todas as triplas publicadas. Na FIGURA 24 é ilustrado o resultado para busca pelo “Café do Paço”, em Curitiba.

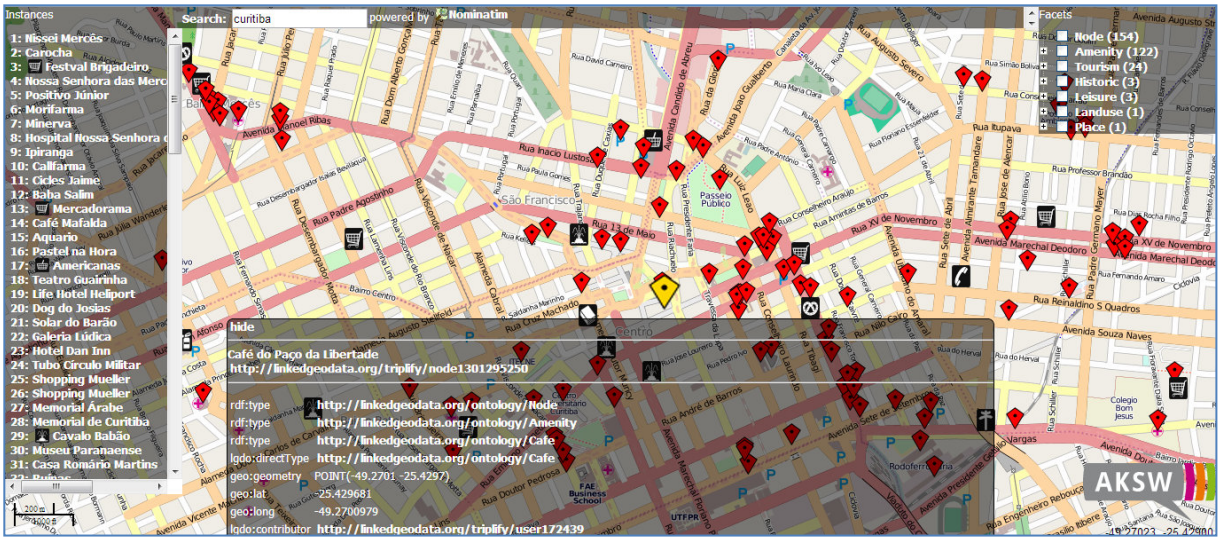


FIGURA 24 - BUSCA NO *BROWSER* DO PROJETO LINKEDGEODATA (2013)

4. INFRAESTUTURA DE DADOS ESPACIAIS

As IDEs foram estabelecidas para ampliar a disseminação e o uso de dados geoespaciais interoperáveis, usando arquitetura baseada em serviços padronizados e também arranjos legais e institucionais. Como havia sido colocado no capítulo 01, é grande a demanda por uma nova geração de IDEs, na qual os usuários tenham uma participação mais forte, via união com a informação geográfica voluntária, e também que isso seja possível, de forma mais fácil, para o usuário não especializado.

A criação da INDE-BR veio estabelecer uma política pública de ênfase no compartilhamento dos dados e definir ações para que os dados oficiais, em especial os chamados dados de referência, sejam descritos e disponibilizados de forma padronizada. As atividades já empreendidas pela iniciativa, em conjunto com os padrões homologados até esta data, são estudados neste capítulo, assim como uma análise das dificuldades de interação do usuário com as informações disponíveis.

4.1. IDEs – HISTÓRICO E CONCEITOS FUNDAMENTAIS

As Infraestruturas de Dados Espaciais (IDEs) são descritas como “o conjunto de tecnologias, políticas públicas e arranjos institucionais para facilitar a disponibilização e o acesso a dados espaciais” (NEBERT, 2004). Quando são criadas para acesso a dados oficiais de determinada nação são denominadas de Infraestruturas Nacionais de Dados Espaciais (INDEs), essas definidas como “uma rede dinâmica composta de pessoas, dados, políticas, padrões e tecnologias para utilização melhorada de dados e serviços espaciais dentro da jurisdição de um país” (GRUS *et al*., 2006).

O interesse no tópico remonta ao início dos anos 90. As IDEs pioneiras foram a americana (1994), a australiana (1996), e, na América Latina, muitas já estão oficializadas e em funcionamento, como as da México, Chile e Argentina (ALBITES, 2002; IGLESIAS, 2008; RICKERT e MACHUCA, 2008). A IDE europeia, chamada INSPIRE, foi criada por uma diretiva, em 2007, e estabelece uma infraestrutura

continental baseada nas dos 27 Estados-membro da União Europeia (CRAGLIA e ANNONI, 2007).

A interoperabilidade gerada pela arquitetura orientada a serviços, principalmente quando apoiada pelos geoserviços padronizados pelo OGC, sempre foi uma das pedras fundamentais da tecnologia na qual se baseiam as IDEs (FRIIS-CHRISTENSEN *et al.*, 2006; BARIK, SAMADDAR e SAMADDAR, 2001; KIEHLE, GREVE e HEIER, 2007; GRANELL, DIAZ e GOULD, 2010; DAVIS JR e ALVES, 2005). Uma arquitetura de IDE é frequentemente retratada por diagramas semelhantes ao da FIGURA 25. Nela, o usuário humano interage com um Geoportal para buscar e descobrir informações geoespaciais e interagir com o servidor de geoserviços. Analogamente, um software cliente também interage com o servidor de geoserviços e com um *serviço web* de Catálogo. Tanto o *serviço web* de Catálogo quanto o Geoportal se alimentam com base nos metadados publicados em um catálogo de dados e serviços.

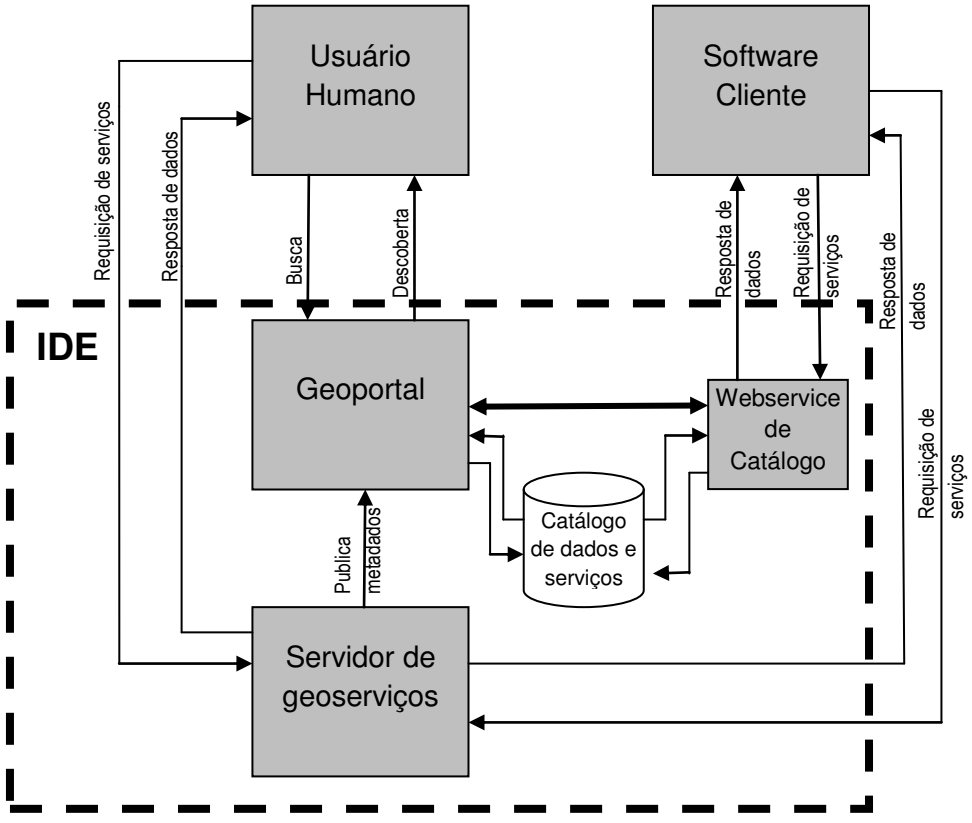


FIGURA 25 - VISÃO DE IDE ORIENTADA A SERVIÇOS.
FONTE: DAVIS JR e ALVES, 2005

4.2. INDE-BR

No Brasil, a INDE-BR foi criada através do decreto n° 6.666, de 27 de novembro de 2008, tendo a CONCAR (Comissão Nacional de Cartografia) como a sua entidade regulamentadora, conforme a FIGURA 26. A CONCAR é formada por diversos membros do governo federal, com participações dos governos estaduais, entidades da sociedade civil e academia.

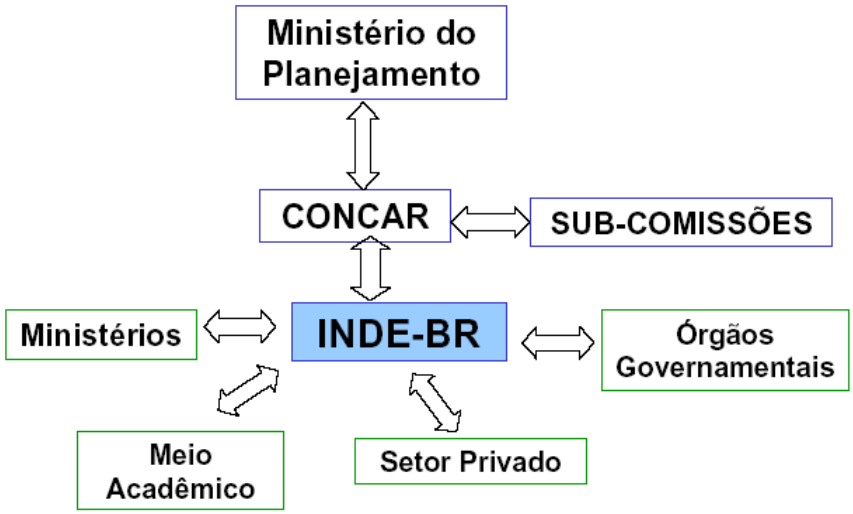


FIGURA 26 - ESTRUTURA DA INDE-BR
FONTE: AUGUSTO e FREITAS, 2007

O decreto de criação, além de declarar os objetivos da iniciativa e definir os principais conceitos envolvidos e destacar responsabilidades, também estabeleceu um prazo de 180 dias para que a CONCAR redigisse um Plano de Ação detalhando o planejamento para a implementação da INDE.

A CONCAR é um órgão colegiado do Ministério do Planejamento, atualizada conforme Decreto s/nº de 1º de agosto de 2008, descendente da antiga COCAR, instituída pelo Decreto Lei 243, de 28 de fevereiro de 1967. Ela é estruturada por uma Secretaria Executiva, cinco subcomissões e, atualmente, conta com cinco comitês especializados.

Os comitês especializados são os seguintes:

- Comitê de Estruturação da Mapoteca Nacional Digital (CEMND): seu objetivo original era propor a estruturação de uma Mapoteca Nacional Digital (MND), de uso compartilhado, na qual estariam disponibilizados os produtos do mapeamento de referência realizado pelos diversos órgãos componentes do Sistema Cartográfico Nacional (SCN). Hoje seus trabalhos são focados na normatização da ET-EDGV (Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais), que é detalhada a seguir.
- Comitê de Estruturação de Metadados Geoespaciais (CEMG): seu objetivo é trabalhar com a padronização de metadados, como o Perfil de Metadados Geoespaciais do Brasil, suas atualizações, treinamento, material de treinamento, entre outros.
- Comitê de Nomes Geográficos (CNGEO): promove a padronização de nomes geográficos dentro do território, assim como dos nomes estrangeiros que serão inseridos em produtos cartográficos nacionais.
- Comitê de Normatização do Mapeamento Cadastral (CNMC): propõe normas para mapeamento cadastral.
- Comitê de Implantação da Infraestrutura Nacional de Dados Espaciais (CINDE): desenvolve as ações e atividades necessárias para a implantação da Infraestrutura Nacional de Dados Espaciais - INDE, consoante com as propostas de linhas de ação emanadas no seu Plano de Ação.

O CINDE é, portanto, o comitê diretamente relacionado à implantação da INDE-BR, em estreita colaboração com os demais comitês, pois todas as normas aprovadas pela CONCAR têm importância na infraestrutura. O comitê é estruturado em Grupos de Trabalho, que são os seguintes: Gestão; Normas e Padrões; Dados e Metadados; Tecnologia; Capacitação e Treinamento; e Difusão e Divulgação.

As atividades do CINDE incluem a administração de um ambiente wiki para a comunicação das atividades dos grupos de trabalho, subsídio aos acordos que caracterizam a adesão à INDE por parte das instituições interessadas e o acompanhamento da criação e atualização do Plano de Ação ao longo do tempo.

No Plano de Ação (CONCAR, 2010), foram definidos três ciclos de implementação para a INDE-BR. O Ciclo I é a fase inicial, previsto inicialmente para durar um ano, na qual todos os atores federais devem estar envolvidos. O Ciclo II, ainda não iniciado, teria duração de dois anos, e compreenderia a consolidação da infraestrutura, bem como sua extensão aos demais níveis de governo (estadual e municipal), e uma maior coordenação com iniciativas globais. O ciclo seguinte, previsto para durar cinco anos, seria de integração com os demais setores não governamentais da sociedade.

No Plano de Ação é ressaltada a valorização e investimento no indivíduo como ator central na estruturação da INDE. São definidos uma série de papéis assumidos pelos atores na INDE, desde gestão estratégica, aquisição e publicação de dados e metadados, manutenção de geoserviços e usuários em geral.

Outra definição importante no Plano de Ação é com relação aos dados de referência e temáticos que irão compor a INDE-BR, bem como um inventário detalhado de quem são seus produtores em âmbito nacional. Alguns exemplos podem ser encontrados na TABELA 4.

TABELA 4 - EXEMPLOS DE PRODUTORES E DADOS DA INDE-BR.

GRUPOS DE INFORMAÇÃO GEOESPACIAL		EXEMPLOS DE PRODUTORES
Dados de Referência	Controle Geodésico	IBGE, Observatório Nacional.
	Cartografia Terrestre Básica (Mapeamento Geográfico, Topográfico e Cadastral).	IBGE, Prefeituras, Exército/DSG.
	Cartografia Especial (Mapeamento Náutico e Aeronáutico)	Aeronáutica/ICA e Marinha/DHN.
	Subsidiários e Acessórios (Mosaicos Ortorretificados, Nomes Geográficos, Divisão Político-Administrativa, Terras Indígenas, Unidades de Conservação, Dados Fundiários, entre outros).	FUNAI, INCRA, ICMBio, IBGE, Órgãos Estaduais de Terra e Cartografia, Ministério das Relações Exteriores, INPE, entre outros.
Dados Temáticos	Meio Abiótico (Geomorfologia, Geologia, Solos, entre outros).	IBGE, Embrapa, DNPM, CPRM.
	Meio Biótico (Vegetação, Biomas, Biodiversidade, entre outros).	MMA e órgãos estaduais e municipais integrantes do SISNAMA.
	Meio Antrópico (Zoneamento Ecológico-Econômico, Riscos, Agricultura, Transportes, entre outros).	Ministério da Agricultura, Transportes.

FONTE: CONCAR, 2010

Com base na descrição acima é possível verificar a gama de atores envolvidos na produção de informação geoespacial no Brasil, mesmo considerando esta uma amostra não exaustiva de instituições diversas, que possuem cultura, processos e objetivos próprios. Adicionalmente aos produtores, ainda mais diversos são os usuários da INDE-BR, definidos no Plano de Ação como sendo “os integrantes de órgãos federais, estaduais, municipais, distritais, dos órgãos da administração direta e indireta (autárquicas ou fundacionais); da iniciativa privada, de organizações civis, de organizações não governamentais (ONGs) e associações; e cidadãos (usuário comum que não está vinculado a nenhuma organização ou entidade), que utilizam os dados e metadados fornecidos pela INDE para integrá-los em diferentes aplicações, efetuando análises temáticas e de assuntos diversos, ou simplesmente utilizando os serviços do geoportal SIG Brasil” (CONCAR, 2010). Essa descrição retrata a diversidade dos usuários envolvidos, desde especialistas em distintas áreas do conhecimento até cidadãos em geral.

4.2.1. Diretório Brasileiro de Dados Geoespaciais e Portal SIG Brasil

No âmbito da INDE-BR, a rede de servidores de dados capaz de reunir eletronicamente produtores, gestores e usuários dos dados geoespaciais envolvidos é denominada Diretório Brasileiro de Dados Geoespaciais, ou DBDG. Cada produtor de dados pode criar seu próprio servidor dentro do DBDG, constituindo um nó autônomo, ou então utilizar o nó central, que é gerenciado pelo IBGE (Instituto Brasileiro de Geografia e Estatística). Para inserção da instituição ao DBDG está prevista a assinatura de um Termo de Adesão, no qual constarão os dados, metadados e serviços a serem disponibilizados, bem como, o planejamento da manutenção do nó, caso seja executado pela própria instituição (CONCAR, 2010).

A FIGURA 27 mostra o diagrama conceitual do DBDG, com uma camada de servidores das diversas instituições, uma camada intermediária na qual funciona o Portal SIG Brasil, que congrega o catálogo de metadados de dados e serviços, visualizadores e o módulo de administração. E na camada do usuário, ficam as aplicações que podem ser customizadas utilizando serviços da INDE-BR.

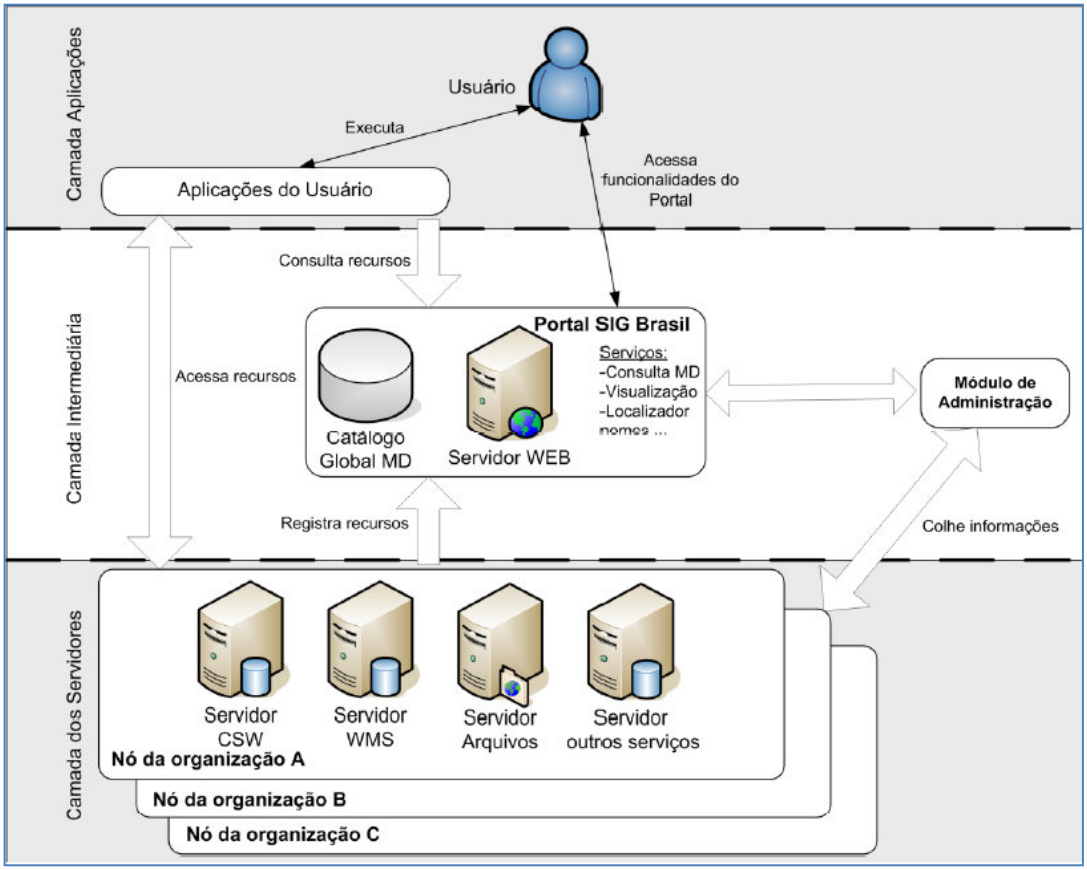


FIGURA 27 - DIAGRAMA CONCEITUAL DO DBDG
FONTE: CONCAR, 2010

Os serviços disponíveis no portal inde.gov.br são os seguintes:

- CPRM – WMS: mapas geológicos e geoambientais (escala 1:1.000.000) , imagens SRTM sombreadas (relevo do Brasil), áreas de relevantes interesses minerais e estatísticas de blocos e modelos geomáticos para gemas, metais e minerais industriais.
- IBGE – WMS: este serviço inclui, entre outros, a base do Brasil ao milionésimo, bancos de dados geodésicos, várias publicações temáticas e a divisão político-administrativa brasileira.
- IBGE – Metadados: catálogo de metadados IBGE.
- MMA – WMS: inclui dados ambientais, como o Zoneamento Ecológico Econômico dos estados, mas também dados de base cartográfica.
- MMA – Metadados: catálogo de metadados MMA.

4.2.2. NORMAS E PADRÕES VIGENTES DA INDE

4.2.2.1. ESPECIFICAÇÃO TÉCNICA PARA ESTRUTURAÇÃO DE DADOS GEOESPACIAIS VETORIAIS (ET-EDGV)

O Mapeamento Sistemático Brasileiro é executado no Brasil pelo IBGE (Instituto Brasileiro de Geografia e Estatística) e pela DSG (Diretoria de Serviço Geográfico do Exército). A Política Cartográfica Nacional foi estabelecida em 1967, através do Decreto-Lei n° 243, que define as escalas-padrão do Sistema Cartográfico Nacional: 1:25.000, 1:50.000, 1:100.000, 1:250.000, 1:500.000 e 1:1.000.000 - e dá competência para execução de Normas Técnicas para Cartografia Terrestre ao IBGE, à DSG e à COCAR (predecessora da CONCAR). O decreto de criação da INDE-BR transfere o papel normalizador à CONCAR.

O primeiro esforço em classificação das informações cartográficas existiu como forma de planejamento para padronização da simbologia e legendas no Manual Técnico T-34-700 (DELOU, 2006). Com o advento da cartografia digital e sua utilização nos órgãos produtores da cartografia sistemática nacional, esse interesse foi canalizado para padronização de níveis de informação no *software* utilizado na época. Só a partir da MTD (Mapoteca Topográfica Digital), e posteriormente da MDN (Mapoteca Nacional Digital) e da ET-EDGV o esforço foi dedicado em forma de modelagem conceitual de banco de dados, finalmente separando a representação lógica no banco da representação visual das feições.

A ET-EDGV tem por finalidade “padronizar estruturas de dados que viabilizem o compartilhamento de dados, a interoperabilidade e a racionalização de recursos entre os produtores e usuários de dados e informação cartográfica” (CONCAR, 2007), em conjunto com diversos parceiros especialistas em cada área de conhecimento.

Para cada uma dessas categorias foi criado um Diagrama de Classes e a Relação de Classes e Objetos, baseado na notação OMT-G para descrever as informações espaciais e semânticas das classes de objetos. O OMT-G (BORGES,

DAVIS e LAENDER, 2001) é um modelo de dados orientado a objeto desenvolvido especialmente para dados geográficos.

Para cada classe, foi determinada uma descrição, código, domínio e primitiva geométrica, bem como seus relacionamentos, conforme o exemplo da FIGURA 28 e TABELA 5. Neste exemplo, está retratada parte da categoria Pontos de Referência. A classe Ponto_Cotado_Altimétrico, vinda da categoria Relevo, é um dado de estrutura vetorial do tipo ponto, coincidente tanto com a classe Pto_Ref_Geod_Topo, quanto com a classe Pto_controle. Estas são por sua vez subclasses da classe Pto_Geod_Topo_Controle.

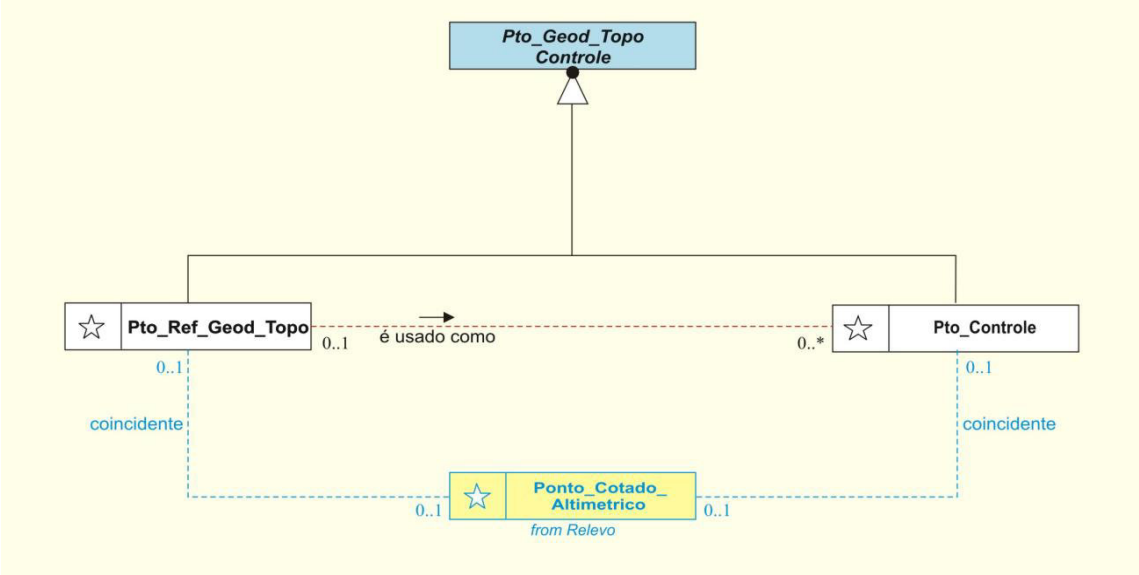


FIGURA 28 - DIAGRAMA DA CLASSE PONTO COTADO ALTIMÉTRICO
FONTE: CONCAR, 2007

TABELA 5 - DESCRIÇÃO DE DADOS DA FEIÇÃO PONTO COTADO ALTIMÉTRICO

CLASSE	DESCRIÇÃO			CÓDIGO	PRIMITIVA GEOMÉTRICA		
Ponto_Cotado_Altimetrico	**Ponto com cota de altitude conhecida.**			**2.05**			
Atributo	**Tipo**	**Tam**	**Descrição**	**Domínio**	**Descrição**	**Requisito**	**Foto**
geometriaAproximada	Boleano	-	Indica que a geometria adquirida é aproximada, em relação a escala prevista para o produto cartográfico.	Sim	-	**NÃO NULO**	-
				Não	-		
cotaComprovada	Boleano	-	Indica se a cota no ponto é comprovada	Sim	Altitudes determinadas por 1)Nivelamento geométrico; 2)Rastreamento de satélites; 3)Nivelamento trigonométrico; 4)Nivelamento barométrico.	**NÃO NULO**	-
				Não	Altitudes determinadas por métodos fotogramétricos e a inserção dos mesmos, na carta, objetiva complementar o traçado das curvas de nível na representação do relevo.		-
cota	Real	**-**	Indica o valor da altitude ortométrica do ponto, em metros.	A ser preenchido	-	**NÃO NULO**	-

FONTE: CONCAR, 2007

Como ressaltado em estudo de Tiainen e Carlson (2006) a gestão de vocabulários como é a definição de feições da ET-EDGV apresenta duas dificuldades principais:

- A atualização da contínua evolução do conhecimento é crucial e de difícil gerenciamento.
- A responsabilidade pela definição das correspondências, atributos e feições recai principalmente no produtor dos dados, podendo acarretar um distanciamento do universo do usuário.

4.2.2.2. PERFIL MGB - PERFIL DE METADADOS GEOESPACIAIS DO BRASIL

O perfil brasileiro (CONCAR,2009) para a norma ISO 19115:2003 foi homologado em dezembro de 2009. Nele, além da tradução para o português, foram criadas listas de domínio e exemplos de preenchimento adequados à realidade brasileira. É com base neste perfil que foi construído o catálogo de metadados da INDE-BR, criado com a plataforma livre *Geonetwork*, e que está disponível no endereço metadados.inde.gov.br.

5. INTEGRAÇÃO DADOS INTERLIGADOS e IDEs

5.1. DIFERENÇAS CONCEITUAIS E VISÕES SOBRE A INTEGRAÇÃO

As diferenças entre as abordagens conceituais dos serviços, nas quais se baseiam as IDEs e os dados interligados abertos têm gerado controvérsia na comunidade de informação geoespacial. Numa polêmica postagem no seu *blog,* em 2011, Thierry Gregorius declara sua mudança de posição com relação ao assunto:

> Como muitos na indústria geoespacial, eu cresci com a noção de que o mundo estava ali para ser abstraído, estruturado, ordenado e modelado com uma grande acurácia. Quando eu entrei na indústria no final dos anos 1990s, SIG e bancos de dados relacionais eram o estado da arte. Pessoas falavam em como as Infraestruturas de Dados Espaciais iam criar representações virtuais de tudo que existe no mundo. O nirvana digital estava perto. Quando o nirvana finalmente chegou, ele não parece exatamente com o que algumas pessoas pensaram. No lugar do 'Mundo de Lego' que muitos esperavam, ou seja, uma pilha de tijolos, construídos organizadamente desde o solo – ele parecia mais como uma pilha de massinha: amorfo, gelatinoso e bagunçado (GREGORIUS, 2011, tradução livre da autora).

Em resposta, Carl Reid, da OGC afirma:

> Não é um OU outro. Há requerimentos para os dois, sejam usados de forma independente ou em concepções mistas. Cientistas, pesquisadores, analistas militares e outros continuarão a precisar analisar a resolução, proveniência, acurácia e outras medidas de qualidade de dados espaço-temporais e sua aptidão para o uso. Ao mesmo tempo, eles, junto com muitos outros (negócios, redes sociais, etc.) estão gratos pelo 'gelatinoso" Big Data e suas ferramentas inovadoras para fazer inferências e descobrir tendências (citado em ARCTUR, 2011, tradução livre da autora).

Outros pesquisadores encontram complementaridade nas visões, como Schade e Cox (2010) que dizem que os dados interligados podem ser a solução para desafios das IDEs como a heterogeneidade de sistemas legados e a variedade de padrões existentes.

Haar (2009) faz uma interessante comparação entre os *mashups* e as IDES, como representado na FIGURA 29. De um lado temos o extremo do conteúdo informal, gerado pelo usuário em uma comunidade aberta e, no outro, temos os serviços *web*, dados oficiais e taxonomias rígidas. O autor sugere uma convergência entre as abordagens. Das coisas que as IDEs podem aprender dos *mashups*, o

autor cita: melhores formas de aproveitar o conhecimento local, melhoramento na coleta e atualização dos dados, criação de serviços para necessidades de nichos de mercado, expansão de mercado graças ao surgimento de novas aplicações, e o reconhecimento dos usuários como produtores de dados. Mas algumas questões permanecem a ser pesquisadas, como o controle de qualidade de dados, segurança, adaptação dos papéis e modelos de negócio existentes, e dificuldades com direitos de propriedade intelectual. Sweeney (2010) também coloca preocupações desta ordem, com a existência de dados autoritativos (como os dados de referência em uma IDE, publicados por uma instituição oficial) em um ambiente aberto. Para destacar o diferencial desta base oficial de dados com relação a outras, o mesmo autor sugere que os provedores publiquem seus dados livremente e de forma fácil na IDE, e providenciem os metadados que ressaltem sua autoridade, da mesma forma que aceitem o retorno da comunidade. Desta forma, o fornecedor não só manteria sua autoridade, como poderia, inclusive, aumentar a confiança junto aos usuários.

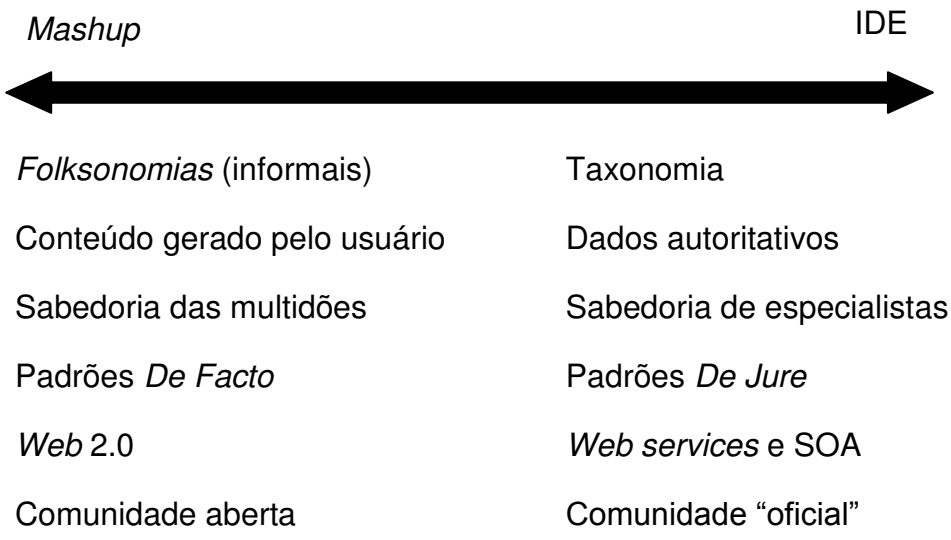


FIGURA 29 - CONTÍNUO ENTRE OS *MASHUPS* E AS IDES
FONTE: HAAR, 2009

Mesmo no Brasil, a integração da abordagem dos *mashups* com uma nova visão para INDE-BR já é sugerida. Borba *et. al* (2012) propõem uma infraestrutura colaborativa com a união das IDEs de diversos níveis, conceitos de *web* 2.0, *cloud computing* e a multidão (*crowd sourcing)* atuando como fornecedor voluntário de informação geoespacial.

A TABELA 6 vem de uma extensa análise de Granell, Schade, e Hobona (2010) detalhando as principais características da estruturação das IDEs e dos dados interligados. Este levantamento sumariza muitas características abordadas nos capítulos anteriores, chamando atenção, por um lado, do potencial dos dados interligados, mas que atualmente ainda não contam com interfaces amigáveis para o tratamento e a exploração dos dados e, por outro, das IDEs, com interfaces mais estruturadas, mas também muito dependentes de aplicações e padrões exclusivos.

TABELA 6 - COMPARAÇÃO IDES E DADOS INTERLIGADOS

Parâmetros	**IDE**	**Dados Interligados**
	MODELO DE DADOS	
Base Lógica	Espaços de dados comunitários distribuídos (em possível cooperação)	Um único espaço de dados global
Unidade de abstração	Serviço	Recurso
Identificação	*Endpoints* de serviços	URIs
Representação	Alguns modelos de dados detalhados (GML,...) Principalmente interpretável por máquina Metadados e dados separados	Modelo RDF Interpretável por máquina Metadados e dados unidos
É interligado?	Sem *links* explícitos	*Links* implícitos
	ACESSO AOS DADOS	
Base Lógica	Acesso a metadados via serviço de catálogo e acesso aos dados via serviços especializados	Acesso a recursos via identificador único
Mecanismo	Acesso não tão simples, mas padronizado Mecanismo de acesso detalhado com muitos parâmetros Conexões padronizadas, HTTP GET, POST-XML, SOAP	Acesso simples e padronizado Interface uniforme HTTO GET/POST/PUT/DELETE URL permite acesso a HTTP e URIs
Clientes	Múltiplos clientes (*desktop*, *web*, móvel, *mashups*)	Alguns clientes *web* semânticos e *browsers* (ainda não muito amigáveis)
É interligado?	Sem *links* explícitos entre fontes de IDEs Metadados e recursos podem ser ligados pelo descritor "Recurso online", se existir GML permite a ligação de elementos do modelo de dados geoespacial usando XLINK (usado raramente)	Representação do recurso explicitamente descreve as ligações para os recursos relacionados
	DESCOBERTA	

Parâmetros	IDE	Dados Interligados
Base Lógica	Metadados de dados e serviços em repositórios centralizados	Recursos com possibilidade de busca na *web* em si
Mecanismo	Consultas espaciais via filtro OGC através de serviço de catálogo Busca Simples (palavras-chave, títulos) e Avançada Consultas distribuídas e *harvesting* em catálogos remotos	Consultas baseadas em SPARQL Requer alto nível de especialidade Consultas se expandem sobre muitas fontes de dados distribuídas se beneficiando da estrutura grafo do RDF.
Clientes	Muitos clientes de catálogo	Clientes proprietários para buscas SPARQL diretas
É interligado?	Dados e serviços isolados. Somente *links* entre serviços de acesso aos dados e os conjuntos de dados associados Palavras-chave comuns são potencialmente uma forma de "ligação"	Fontes de dados interligadas As consultas SPARQL exploram de forma natural a noção de ligação entre fontes de dados.
PUBLICAÇÃO		
Base Lógica	Preencher catálogo de metadados e exposição dos dados via serviços. Sincronização complexa.	Transformação de dados legados em triplas RDF
Mecanismo	Editores de metadados e ferramentas de publicação complexas (ainda desconectados) Nível de especialista requerido	Soluções em negociação de conteúdo traduzem dados legados em modelos RDF Sincronização complexa, uso especializado
Clientes	Editores de metadados as vezes unificados com catálogos geoespaciais	Necessário cliente para estabelecimento dos *links*
É interligado?	Sem *links* explicitamente digitados	Quem publica tem que explicitar os *links* de seus dados estruturados com outras fontes de dados.

FONTE: GRANELL, SCHADE e HOBONA, 2010

Portanto, a demanda por integração existe e tem sido tratada principalmente como um desafio duplo: a estruturação do conhecimento existente nas IDEs para seu acesso via dados interligados, e a exposição direta dos dados geoespaciais em formato de recursos.

5.2. A ESCOLHA DO USO DO GML COMO INTEGRADOR

Neste trabalho, a escolha de usar o formato GML como integrador é apoiada em estudos como o de Schade e Cox (2010), que ressaltam que o GML, em sua criação, foi fortemente influenciado pelo RDF. Os pesquisadores argumentam que a abordagem dos dados interligados não é uma novidade para a comunidade geoespacial, uma vez que os princípios de estruturas isomórficos já tinham sido introduzidos em padrões de IDEs há mais de uma década. Eles propõem combinar o GML com o conceito de negociação de conteúdo como uma solução simples para colocar as IDEs atuais no caminho dos dados interligados, conceito que também foi adotado para esta pesquisa. O estudo revela que a parte básica do mapeamento entre GML e RDF é simples:

- xlink:href = rdf:resource
- gml:identifier = rdf:about

E, portanto, o GML é convertido de forma direta em RDF. Esta vantagem, combinada à previsão do formato na especificação recente do GeoSPARQL, mostra-se bastante promissora e é usada como base do serviço *web* desenvolvido nesta pesquisa.

5.3. ABORDAGENS DE INTEGRAÇÃO BASEADA EM SERVIÇOS

Granell, Schade e Hobona (2010) acreditam que a abordagem de dados interligados pode ser projetada nas IDEs para ser usada como uma forma de aproximá-las das informações geográficas voluntárias (VGI). No trabalho dos autores são sugeridas três formas de suporte a dados interligados nas IDEs:

- Orientação a recursos completa através de repositórios RDF nativos: a aplicação direta de dados interligados para IDEs procura converter todos os modelos e conjuntos de dados e metadados disponíveis no formato RDF. Apesar de ser um cenário ótimo em termos de ligar os elementos disponíveis, ele pode ser irrealista em termos da quantidade de dados. Além disso, apesar de tecnicamente plausível,

toda a combinação de políticas pública, acordos e consenso entre todos os envolvidos poderia resultar num trabalho infindável.

- Semi-orientação a recursos através de reforço: uma alternativa é aumentar os elementos existentes da IDE através de recursos externos. Nessa abordagem, a comunidade da IDE tem que aumentar seus horizontes e criar interfaces de serviços que se adequem às necessidades dos dados interligados. Isso seria possível com o fornecimento de RDF, em referência cruzada com dados espaciais já estabelecidos como GML e disseminados através das atuais interfaces de geoserviços. Isso implicaria numa abordagem híbrida da atual estrutura de orientação a serviços com a orientação a recursos. Um dos benefícios seria que os dados interligados seriam dinâmicos, com o conteúdo gerado sob demanda. Uma das limitações seria a estrutura, sendo alterada dinamicamente, poderia afetar o acesso e a consulta aos dados.
- Semi-orientação a recursos através de serviços mediadores: outra alternativa é oferecer a "negociação de conteúdo" através da transformação de dados para que estes sejam adequados ao uso em várias aplicações. Esses serviços podem se tornar uma ponte entre os dados interligados e as IDEs. A diferença principal com relação à abordagem de reforço é que o reforço requer uma modificação nos modelos atuais de informação geoespacial, quando os mediadores podem atuar sem essa necessidade.

Foi escolhida para utilização neste trabalho a terceira linha de integração, com base em vários exemplos na literatura que usaram este tipo de abordagem para os objetivos semelhantes aos pretendidos para esta pesquisa. Um deles é o trabalho de Gomes Jr e Medeiros (2007), no qual criaram um serviço para buscas de dados ecológicos que envia o resultado de busca SPARQL como argumento de um filtro para um serviço WFS, que então retorna um GML. Outro serviço, chamado de CSW2LD, foi proposto em trabalho de Lopez-Pellicer *et al.* (2011), o qual publica catálogo de metadados no formato OGC CSW para o formato de dados interligados.

Também foi um serviço intermediário a solução adotada por Tschirner, Scherp e Staab (2011) propondo a conexão da INSPIRE com dados interligados. Para atingir esse objetivo, os autores destacam dois passos: a criação de ontologias comuns e a criação de consultas usando SPARQL. Há necessidade de criação de ontologias para modelar a conversão de GML para OWL, auxiliar na armazenagem

das informações estruturais do GML, facilitar a referência com outros dados e preparar a base para o alinhamento com ontologias superiores da *Web* Semântica. Esta solução é baseada em serviços *web,* evita a armazenagem de dados temporários e mantém a estrutura existente nas instituições.

Dentre os desafios deste modelo estão a coordenação da infraestrutura semântica dos dados INSPIRE de referência e uma instância de gestão das ontologias. Esta preocupação permeia também outros trabalhos, como o de Lacasta *et al.* (2007) que propõem um *Web Ontology Service* (WOS) específico para facilitar a gestão e o uso de Ontologias. Expandindo esta visão, Janowicz *et al.* (2010) sugerem também um *Web Reasoning Service* (WRS) para trazer capacidades de inferências às IDEs.

Finalmente, Schade e Granell (2010) também propõem duas alternativas para integração. Segundo eles, as IDEs já contém uma série de recursos interligados e que, para serem acessados, podem ser utilizadas soluções existentes com poucas alterações em termos de padrões e tecnologias. Novamente, é identificado que o GML possui todas as capacidades necessárias para a integração. O conceito de “negociação de conteúdo” permite que sejam mantidas as estruturas clássicas das IDEs, para serem usadas em algumas aplicações especialistas, enquanto dados em RDF podem ser usados por uma comunidade mais ampla. Estas estruturas podem ser usadas para dados ou metadados. Assim, os autores propuseram dois cenários de integração.

Em um primeiro cenário é sugerido o uso de *links* embutidos nos metadados de um determinado recurso da IDE. Este recurso pode ser acessado via um serviço WMS, ou um serviço de *download*, dependendo do caso. No segundo cenário, além da possibilidade de seguir *links* dentro das IDEs, a proposta é descortinar dados escondidos em serviços de acesso aos dados através da geração de dados RDF sob demanda. Nesse caso, a negociação de conteúdo é feita no nível da feição, e dependendo do cliente o WFS poderia oferecer os dados no GML clássico, ou em RDF, ou mesmo em HTML.

A negociação de conteúdo é uma forma para que, na arquitetura decrita nesta pesquisa, o serviço *web* proposto seja colocado como alternativa de acesso para clientes que utilizam dados interligados, enquanto os serviços WMS, WFS e *download* da INDE-BR continuariam a dispor conteúdo para clientes específicos.

6. METODOLOGIA

6.1. DIAGRAMA DE CASOS DE USO DA SOLUÇÃO

Para cumprir os objetivos de integração de bases de dados diversas, a arquitetura proposta incluiu diversos casos de uso executados pelos atores da solução. O Diagrama de Casos de Uso em formato UML (*Unified Modeling Language)* está representado na FIGURA 30, identificando os principais atores envolvidos na solução e os principais grupos de interações envolvidos, chamados de casos de uso. Alguns deles são executados atualmente, como a publicação de dados em formato RDF pela comunidade de Dados Interligados. Outros estão previstos mas ainda não implementados totalmente, como a publicação de dados em RDF pelos atores da INDA e de WFS pelos atores da INDE.

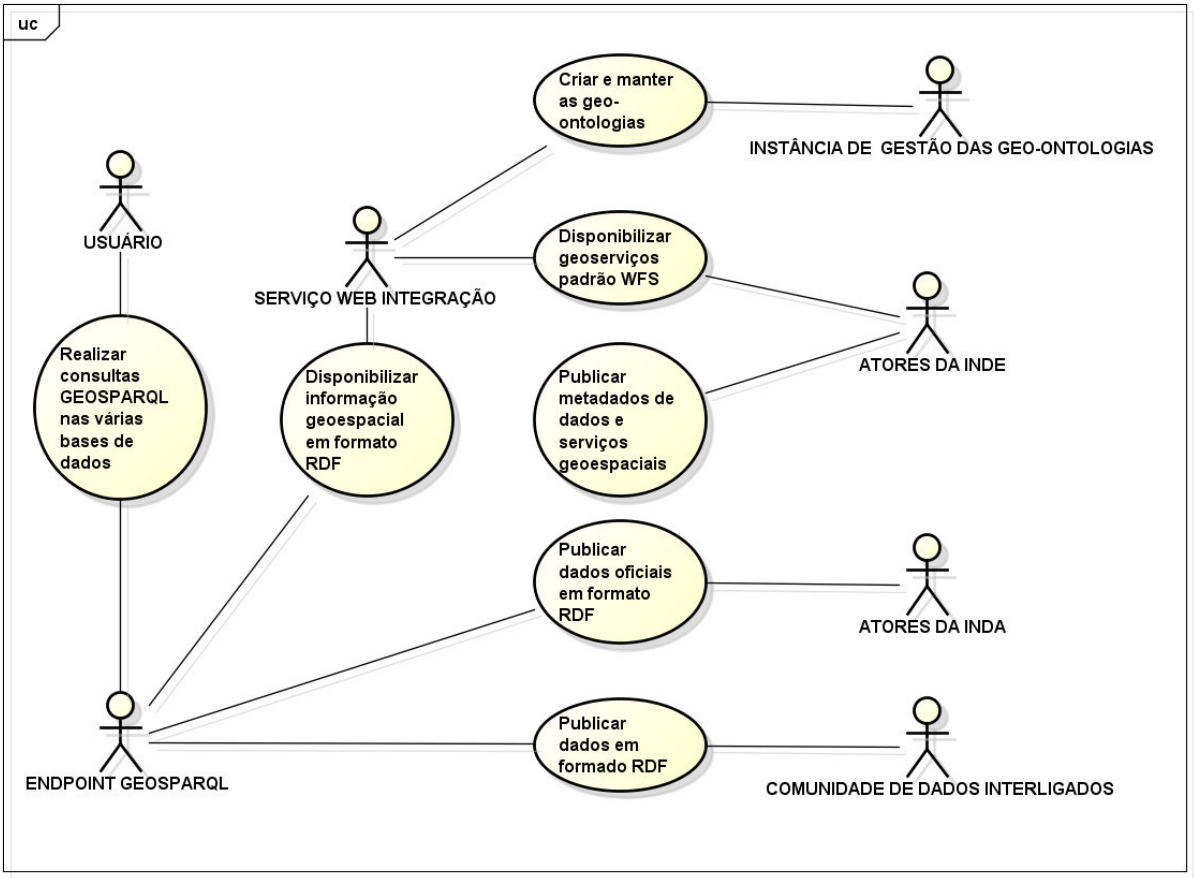


FIGURA 30 – DIGRAMA DE CASOS DE USO DA ARQUITETURA PROPOSTA
FONTE: A AUTORA, 2013

Há um importante ator identificado que atualmente não está proposto nas instâncias gestoras da CONCAR, que seria a responsável pela criação e gestão de ontologias relacionadas às informações geoespaciais da INDE, com atenção especial aos dados de referência. A criação e atualização de ontologias é essencial para aplicações semânticas que utilizem estes dados.

Outra ação fundamental nesta arquitetura é a operação de um serviço *web* mediador que possibilite a integração dos dados publicados em WFS automaticamente com os dados em RDF, cuja operação será detalhada a seguir neste trabalho. E, finalmente, um *endpoint* GEOSPARQL que permita ao usuário interagir com as informações. Esta *interface* poderia ser beneficiada por consultas pré-definidas, mapas interativos, tutoriais, entre outras formas de facilitar a navegação.

Neste trabalho foi desenvolvido o serviço *web* de integração e a arquitetura total demonstrada através da exposição das consultas GEOSPARQL propostas no capítulo seguinte.

6.2. EFEITOS DA ARQUITETURA NAS ETAPAS DE PUBLICAÇÃO DE DADOS INTERLIGADOS GEOESPACIAIS

Os principais passos para publicação de dados interligados são descritos por Corcho e Pérez (2010) e ilustrado na parte superior da linha tracejada da FIGURA 31. Na parte inferior o desenho mostra como ficaria o processo com a utilização da arquitetura proposta. Primeiramente, são selecionadas as bases de dados que serão publicadas. Depois, o desenvolvimento de vocabulários ou ontologias adequados para os temas. Em seguida viriam em dois passos separados a criação e publicação dos dados em formato RDF. Segundo a proposta deste trabalho, estas duas etapas são substituídas pela integração *online* proporcionada pelo serviço *web,* e com isso traz integridade na atualização dinâmica dos dados originais. Seria originalmente necessário na sequência um processo de limpeza de eventuais erros ou redundâncias, etapa que também seria eliminada na solução proposta. É importante ressaltar, no entanto, que a falta desta etapa de limpeza *a posteriori* exige observar com ênfase a existência de dados originais na qualidade suficiente para aplicação

desejada e que o processo de desenvolvimento das ontologias seja feito com muito cuidado.

Os processos finais demonstram a fase de ligação com outras bases disponíveis e a disponibilização ferramentas de busca e acesso aos dados ao usuário.

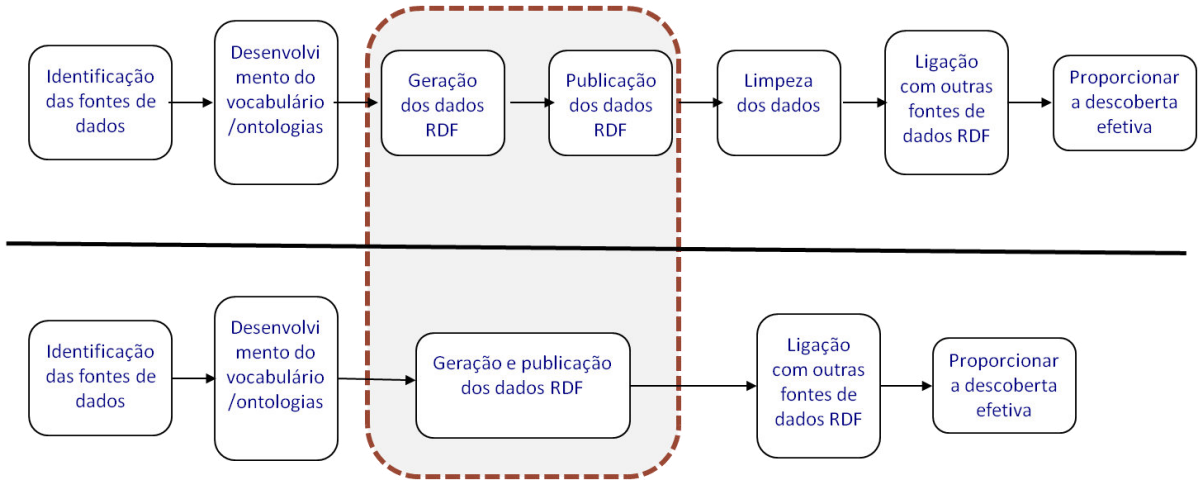


FIGURA 31 - PASSOS PARA DISPONIBILIZAÇÃO DE DADOS INTERLIGADOS.
FONTE: ADAPATADO DE CORCHO e PÉREZ, 2010

Dentro desta estrutura, as principais etapas para criação do serviço *web* de integração foram: definição do tema de trabalho, criação da ontologia, definição da arquitetura, desenvolvimento e testes.

6.3. PREMISSAS PARA A CRIAÇÃO DO SERVIÇO *WEB*

A primeira decisão tomada foi a utilização de um serviço mediador, como exposto anteriormente. Foi escolhida esta opção por aproveitar a estrutura existente para INDE-BR, garantido a utilização de recursos humanos e de *software* preparados até o momento. O serviço WFS, usado como exemplo no caso, não está disponível ainda na INDE-BR, mas todo *software* gerador de WMS também tem potencial para fornecer WFS.

Adicionalmente, caso fosse escolhida a publicação total dos dados geoespaciais em formato RDF, seria perdida a conexão com a base oficial, forçando

uma repetição do processo a cada nova versão dos dados. Alguns temas que formam os dados de referência da INDE-BR, como a divisão político-administrativa brasileira, sofrem alterações constantes decorrentes de natureza legal ou judicial, ajustes e refinamentos cartográficos, utilização continuada de melhores técnicas e insumos de produção, entre outros (IBGE, 2012). Portanto, é importante que a ligação entre os dados RDF publicados e a base oficial seja dinâmica e automática para evitar problemas de entendimento sobre qual a versão mais atual disponível.

A FIGURA 32 apresenta a publicação dos dados em formato RDF a partir de seus bancos de dados originais. As diversas triplas RDF podem apontar para definições e objetos originários de outras fontes.

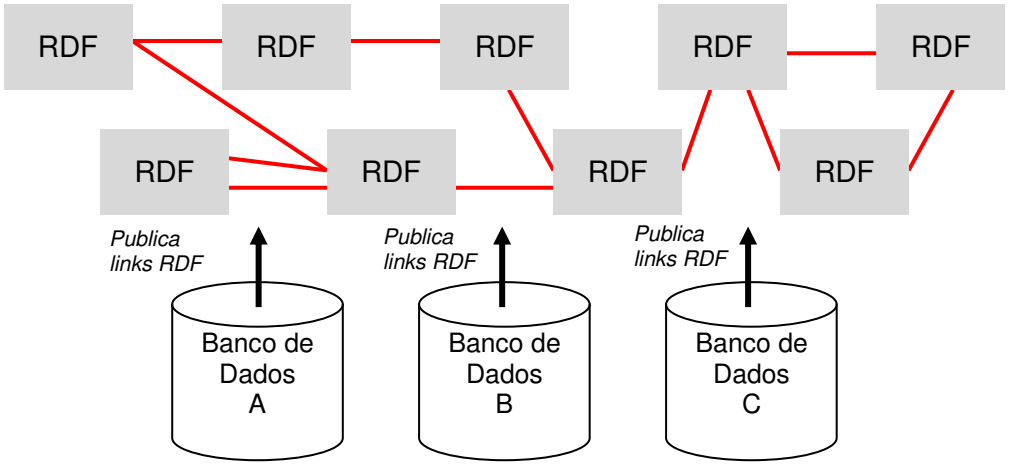


FIGURA 32 - PUBLICAÇÃO DE RDF
FONTE: ADAPTADO DE BREITMAN, 2009

Atualmente, a INDE-BR tem o potencial de disponibilizar informação através de serviços interoperáveis, conforme ilustrado na FIGURA 33.

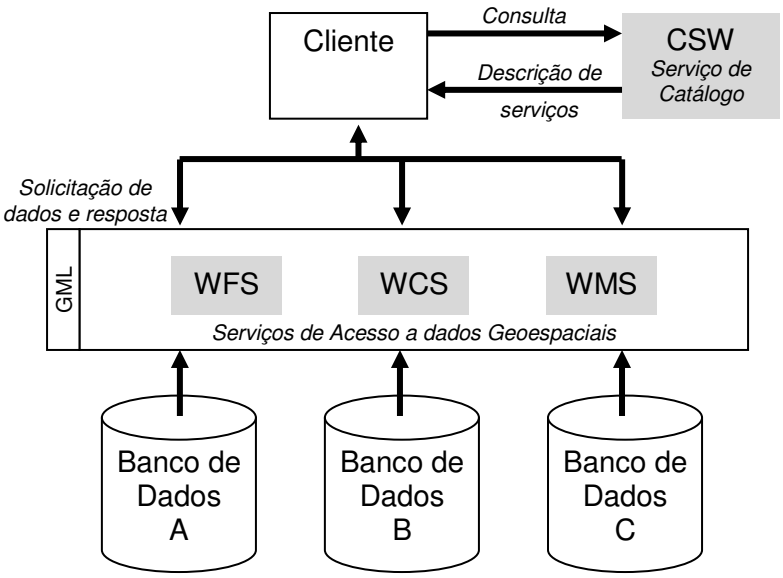


FIGURA 33 - PUBLICAÇÃO DE DADOS ESPACIAIS ATRAVÉS DE GEOSERVIÇOS
FONTE: ADAPTADO DE NERBRET, 2004 E ZHAO, 2009

A arquitetura proposta aqui permite a intermediação automática entre dados já publicados em formato OGC WFS, através de um serviço *web*, que os tornem disponíveis em RDF, mas sem necessidade de conversões *offline,* como pode ser observado no esquema da FIGURA 34.

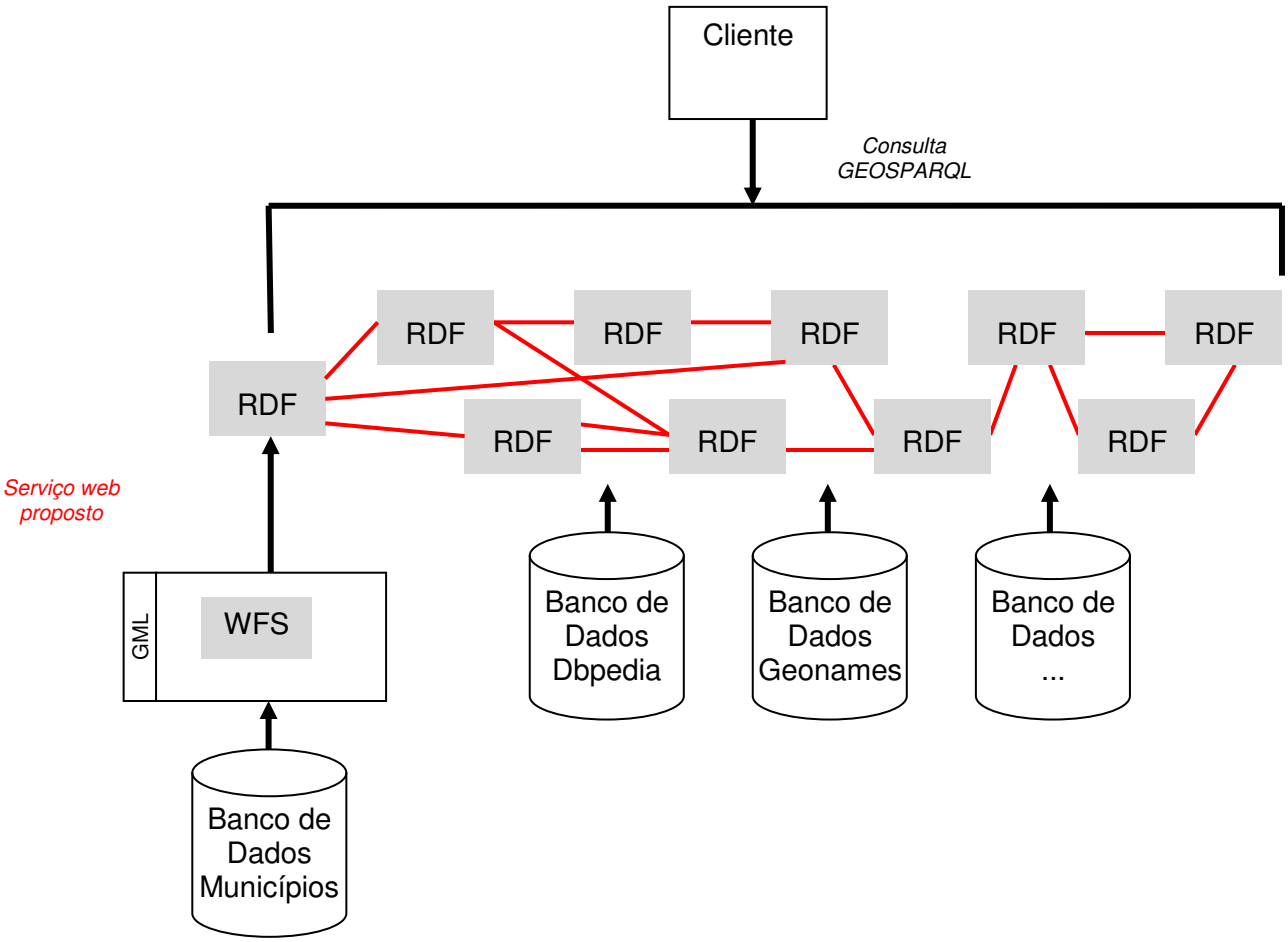


FIGURA 34 – ARQUITETURA PROPOSTA.
FONTE: A AUTORA, 2013

6.4. ESCOLHA DO TEMA PARA A IMPLEMENTAÇÃO

O tema escolhido para a aplicação desenvolvida com base na arquitetura foi a Divisão Político-Administrativa Brasileira, em especial a malha municipal do Estado do Paraná, com 399 municípios, na versão publicada para referência do Censo 2010. Os municípios são, segundo a Constituição da República Federativa do Brasil de 1988, os entes federativos de menor nível hierárquico. Este é, portanto, um tema que interage com diversas informações de caráter administrativo e que tem potencial para congregar diversas bases de dados interligados.

A escolha do tema de divisão político-administrativa como prioritário para criação de aplicações semânticas e ontologias é citada em várias referências, tais como, Gantner (2011) sobre a definição de ontologia espaço-temporal para as unidades administrativas na Suíça; Albornos *et al.* (2008) criaram uma ontologia de unidades administrativas para os Estados europeus, incluindo dados históricos da

Grã-Bretanha, Suécia e Estônia; também Lopez-Pellicer *et al.* (2008) relataram a experiência da criação das ontologias que resultaram na publicação das unidades administrativas espanholas no GeoLinkedData.

A Divisão Territorial Brasileira pode ser considerada uma das mais dinâmicas dos países ocidentais, principalmente no que se refere à criação de municípios e à alteração de suas divisas (IBGE, 2012), o que a reforça como um tema particularmente interessante para aplicações dinâmicas com ligação direta à malha mais recente disponível.

6.5. CRIAÇÃO DA ONTOLOGIA

O segundo passo para permitir a criação do RDF foi a criação da Ontologia em formato OWL. Para tanto, iniciou-se uma pesquisa sobre os conceitos relacionados, encontrados em Lima *et al.* 2002, que elenca na divisão territorial vigente:

- Grandes Regiões - há cinco Grandes Regiões no Brasil: Norte, Nordeste, Sudeste, Sul e Centro-Oeste.
- Unidades da Federação - o Brasil conta com 27 Unidades da Federação, sendo 26 Estados e 1 Distrito Federal.
- Mesorregiões Geográficas - as Mesorregiões Geográficas são conjuntos de municípios contíguos pertencentes à mesma Unidade da Federação.
- Microrregiões Geográficas - as Microrregiões Geográficas são conjuntos de municípios contíguos e partes das mesorregiões que apresentam especificidades quanto à organização do espaço.
- Regiões Metropolitanas - são conjuntos de municípios de uma mesma Unidade da Federação reunidos, por legislação estadual, segundo critérios estabelecidos pelo próprio Estado.
- Municípios - são entes federativos de menor nível hierárquico, sendo sua criação feita por legislação estadual conforme critérios estabelecidos pelo próprio estado.

- Cidade - é a sede do município e quem lhe dá o nome. É constituída pela área urbana do distrito sede e delimitada pelo perímetro urbano estabelecido por lei municipal.
- Distritos - são unidades administrativas municipais criadas por lei municipal; podem também ser criados pela mesma lei estadual que criou o município. Todo município possui, no mínimo, um distrito, que é denominado distrito sede e que abriga a cidade.
- Vila - é a sede do distrito que não abriga a cidade. É a vila que empresta seu nome ao distrito. É delimitada pelo perímetro urbano definido, por lei municipal, como a área urbana do distrito que não a sede do município.
- Subdistritos - são subdivisões intra-urbanas existentes em alguns municípios, criadas por legislação municipal com fins administrativos.
- Bairro – são divisões intra-urbanas criadas, identificadas e delimitadas por legislação municipal específica.
- Setor Censitário - é a unidade espacial de coleta de informações utilizada pelo IBGE para a realização das pesquisas. É estabelecido a partir de critérios que associam o número de domicílios à adequação da área a ser percorrida pela pessoa que realiza a coleta de dados. Seus limites obedecem a todos os limites hierarquicamente superiores: bairro, vila, distrito, cidade, município e Unidade da Federação, e são sempre divididos em área rural e urbana (IBGE,2011), de acordo com a legislação municipal de perímetro urbano.

De acordo com os dados do Censo 2010 do IBGE (IBGE, 2011), na data de referência do Censo existiam no Brasil 27 Unidades da Federação (sendo 26 Estados e 1 Distrito Federal), 5.565 municípios, 10.283 distritos, 662 subdistritos, 14.402 bairros e 316.574 Setores Censitários, dos quais 240.382 são urbanos e 76.192 rurais.

Para a criação da ontologia, foram buscadas todas as inter-relações implícitas entre as classes envolvidas. Existe uma diversidade grande de conceitos, como por exemplo, determinada cidade, que é a sede de um município, faz parte também de um distrito e subdistrito, mas simultaneamente é subdividida em bairros e este em setores censitários, aos quais estão ligadas todas as tabelas de divulgação dos dados do Censo desagregados por setor.

Como subsídio, também foi usado o modelo conceitual das categorias Limites e Localidades da ET-EDGV conforme FIGURAS 35 e 36. Nestes, observa-se que as classes que formam a ontologia estão descritas no modelo, sendo Unidade_Federacao formada por agregação de Municípios, estes formados por Distritos e estes por subdistritos, além de Regiao_Admnistrativa e Bairro, fazer parte de uma super classe chamada Area_Politico_Administrativa. As cidades, vilas e localidades são classes da categoria Localidades contidas nas respectivas Areas_Politico_Admnistrativas. Os setores censitários não fazem parte do modelo, mas seus limites fazem parte da classe Limite_Operacional. Portanto, é possível observar a estrutura das classes da ET-EDGV, mas a complexidade de seus relacionamentos fica delineada de forma mais apropriada para aplicações semânticas na forma de ontologia.

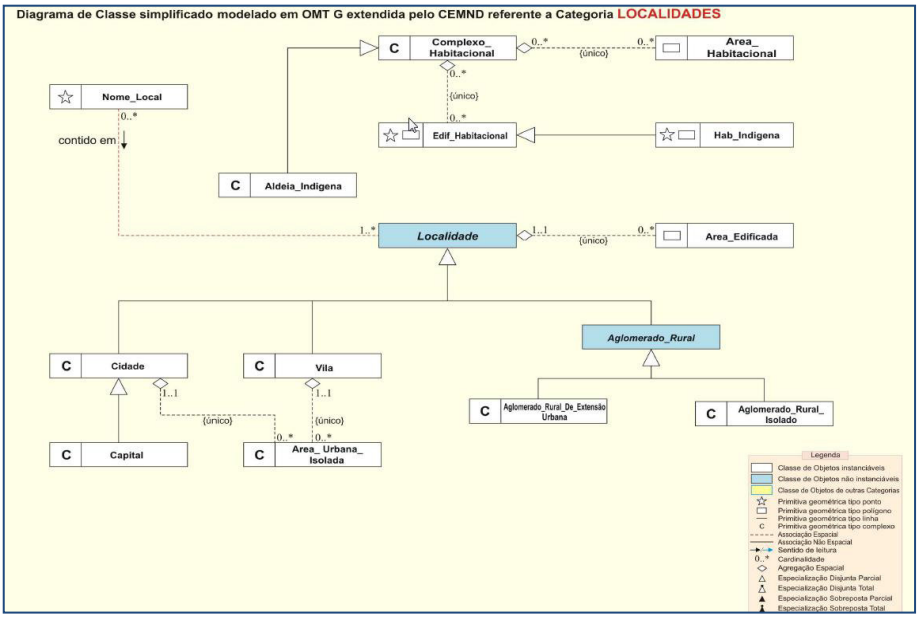


FIGURA 35 - CATEGORIA LOCALIDADES
FONTE: CONCAR, 2007

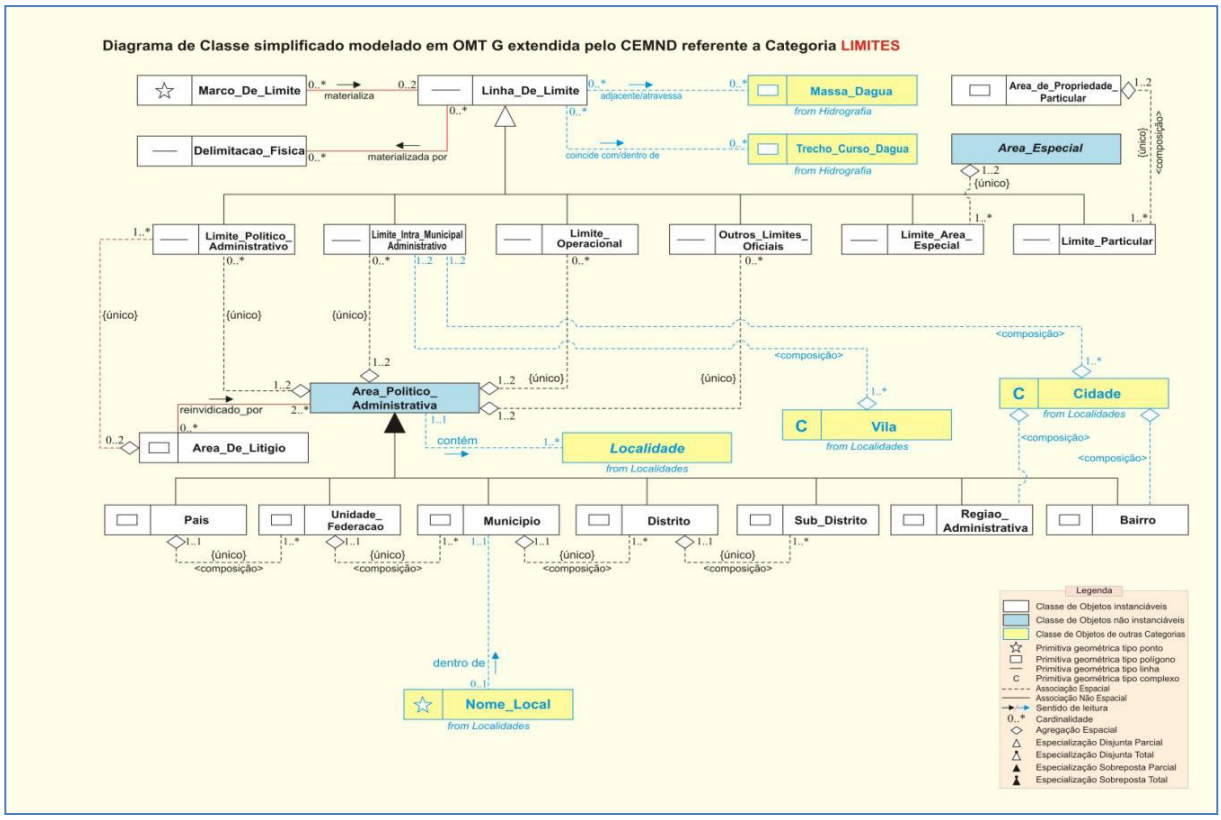


FIGURA 36 - CATEGORIA LIMITES
FONTE: CONCAR, 2007

Essa diversidade de níveis administrativos e territoriais precisou ter seus relacionamentos mapeados a fim de criar a ontologia em formato OWL, usando o *software* livre Protégé, versão 4.2.0, desenvolvido pela Universidade de Stanford.

A primeira classe é a da Grande Região, na qual está contida a classe Unidade da Federação. A esta, podem estar relacionadas duas classes: Estado e Distrito Federal. Agrupados de forma diferente, a classe Município pode pertencer à classe Regiao_metropolitana e, simultaneamente, pertencer à classe Microrregiao, e esta, à classe Mesorregiao. Associada à classe Municipio tem-se a classe Distrito e Subdistrito. A classe Bairro está ligada ao Subdistrito. A mais desagregada das classes, a de Setor_Censitario, que pode pertencer a um Bairro, Subdistrito, ao Distrito Federal diretamente e às classe de Area_Urbana e Area_Rural. À classe de Area_Urbana, por sua vez, encontram-se associadas as classes de Area_urbana_isolada, Nucleo, Povoado, Vila e Cidade, a qual também está relacionada à classe Bairro. Na FIGURA 37 pode-se observar o resultado em formato gráfico.

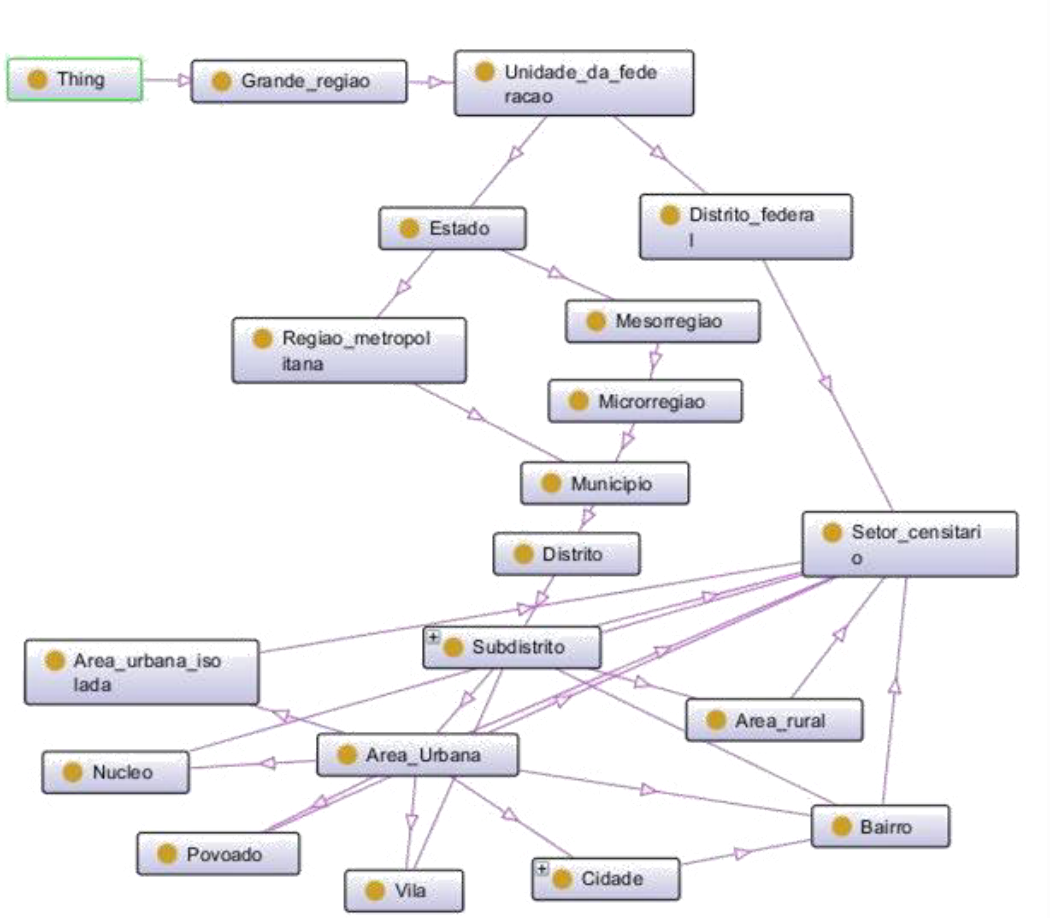


FIGURA 37 - ONTOLOGIA OWL PARA DIVISÃO TERRITORIAL BRASILEIRA
FONTE: A AUTORA, 2013

A ontologia proposta abrange mais conceitos do que os usados no serviço *web*, mas seu processo de criação fez parte do estudo das recomendações para a criação e gestão de ontologias para a INDE-BR.

Adicionalmente, para a hierarquização dos arquivos e criação das URIs onde seriam gravados os arquivos RDFs correspondentes aos municípios, foi utilizada a codificação para os entes da Divisão Territorial Brasileira já estabelecida pelo IBGE e adotada em todo o processo de aquisição e divulgação de dados da instituição.

São dois dígitos para a Unidade da Federação (sendo que o primeiro indica a Grande Região), cinco para os Municípios, dois para o Distrito, dois para o Subdistrito e quatro para o Setor Censitário, formando um geocódigo de quinze dígitos no total (IBGE, 2011), como pode ser observado no exemplo da FIGURA 38.

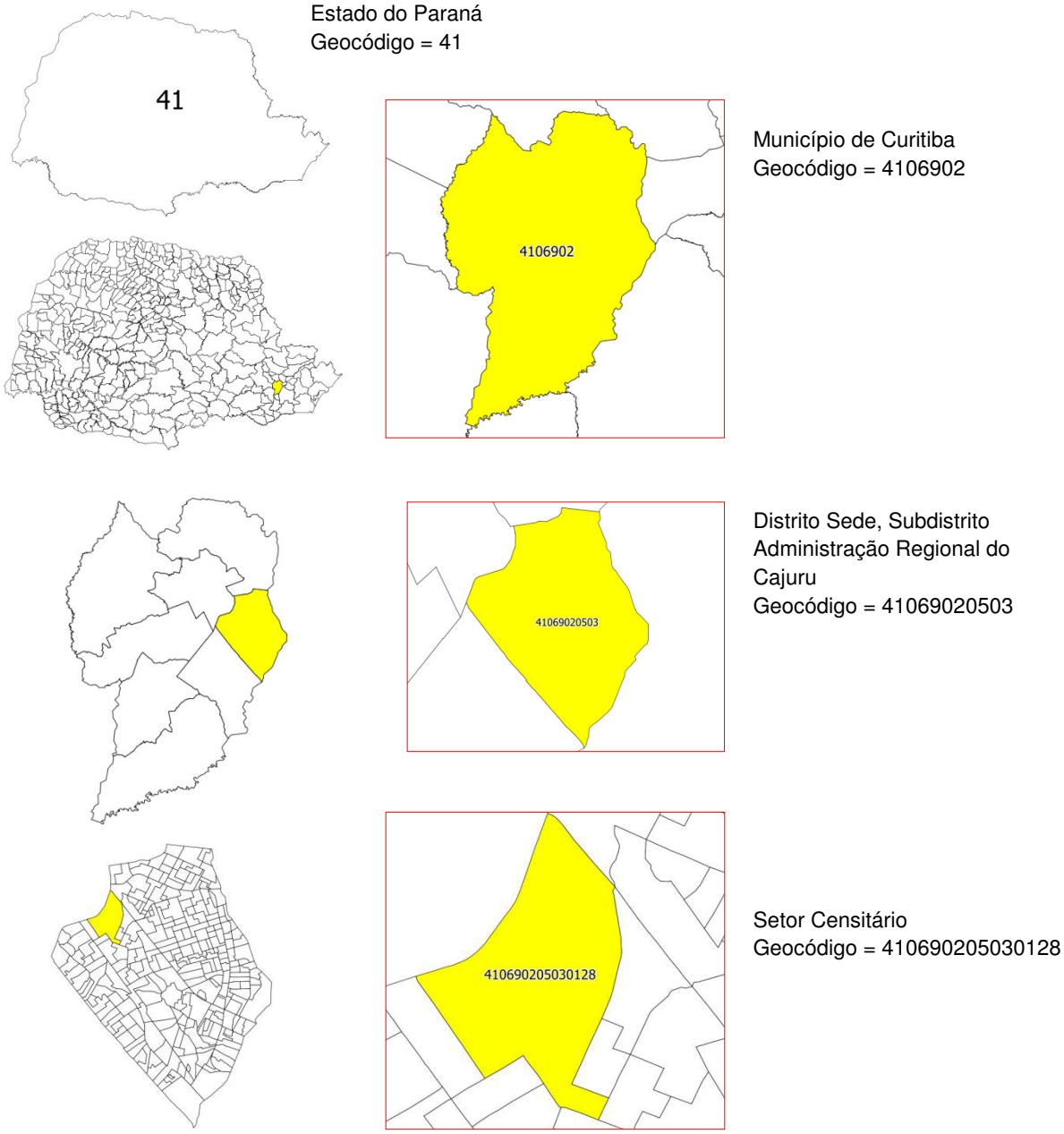


FIGURA 38 - EXEMPLO DA HIERARQUIA DE GEODCÓDIGOS.
FONTE: A AUTORA, 2013

A proposta de estrutura básica é o armazenamento dos conceitos no servidor, através de um área de armazenamento “admin”, para armazenar a definição das ontologias, que foi implementada neste trabalho apenas nos conceitos envolvidos no caso de uso (municípios), e o armazenamento sugerido dos RDFs de acordo com o nível hierárquico, como no seguinte exemplo de URIs:

`http://NOSSO_SERVIDOR/admin/41` ➔ URI do Estado do Paraná

`http://NOSSO_SERVIDOR/admin/4106902` ➔ URI do Município de Curitiba

`http://NOSSO_SERVIDOR/admin/41069020503` ➔ URI do Subdistrito Administração Regional do Cajuru

Foi então criada a estrutura dos arquivos RDF que serão usados no serviço *web.* Os RDFs seguem este modelo e são gerados sob demanda a cada solicitação. O exemplo a seguir é o modelo para o município de Curitiba, sendo que a parte marcada como “SUBSTITUIR” é a fornecida pelo serviço WFS.

```
PREFIX adm: <http://NOSSO_SERVIDOR/admin#>
PREFIX geo: http://www.opengis.net/geosparql#
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
adm:Curitiba a adm:Municipio;
rdfs:label "Curitiba";
geo:hasGeometry adm:Polygon1 .
adm:Polygon1 a geo:Polygon;
geo:asGML
"<gml:Polygon
     srsName=\"http://www.opengis.net/def/crs/OGC/1.3/CRS84\"
     xmlns:gml=\"http://www.opengis.net/gml\">
   <gml:pos>SUBSTITUIR</gml:pos>
 </gml:Polygon>"^^gml:gmlLiteral
```

6.6. DESENVOLVIMENTO DO SERVIÇO MEDIADOR

O serviço mediador funciona de acordo com as etapas especificadas na FIGURA 39.

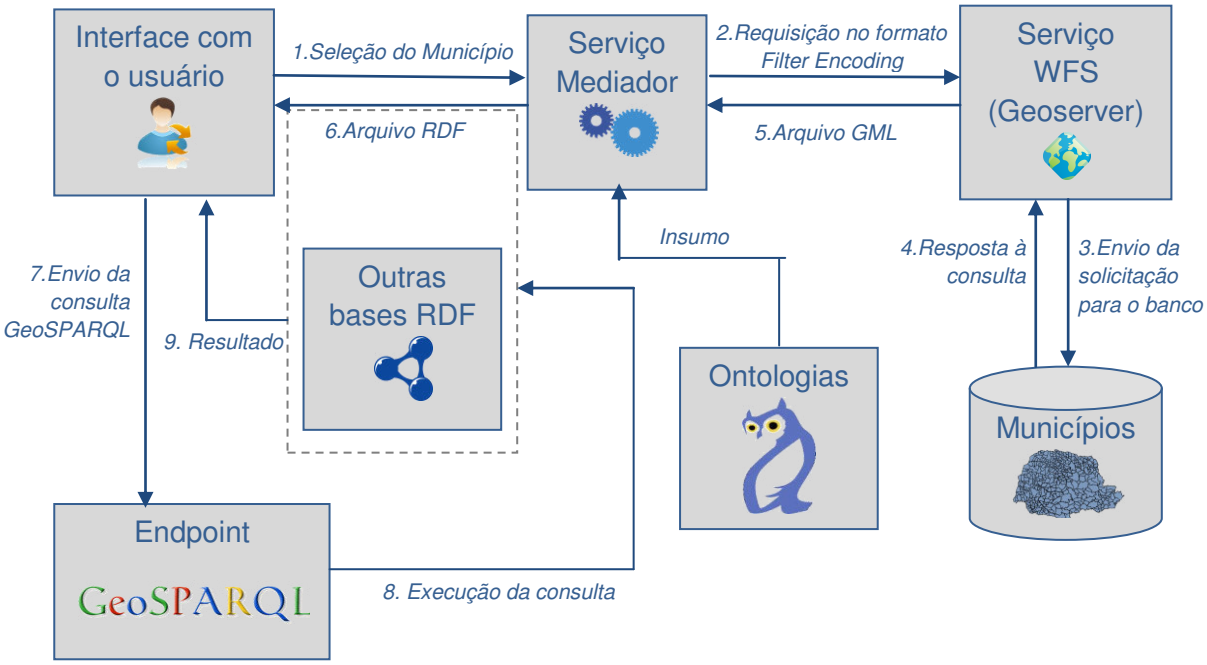


FIGURA 39 - ARQUITETURA DO SERVIÇO *WEB.*
FONTE: A AUTORA, 2013

Primeiramente, pela falta de um serviço WFS disponível na INDE-BR, este foi criado usando a malha municipal disponibilizada pelo IBGE. Para tanto, foi instalado um servidor de mapas com o *software* livre Geoserver, versão 2.4.4. Neste servidor, foi definida uma *workspace* e adicionado como fonte de dados o arquivo no formato vetorial *shapefile* da divisão municipal do Estado do Paraná. Em seguida executou-se a ativação do serviço WFS. A interface de criação do serviço é a indicada na FIGURA 40.

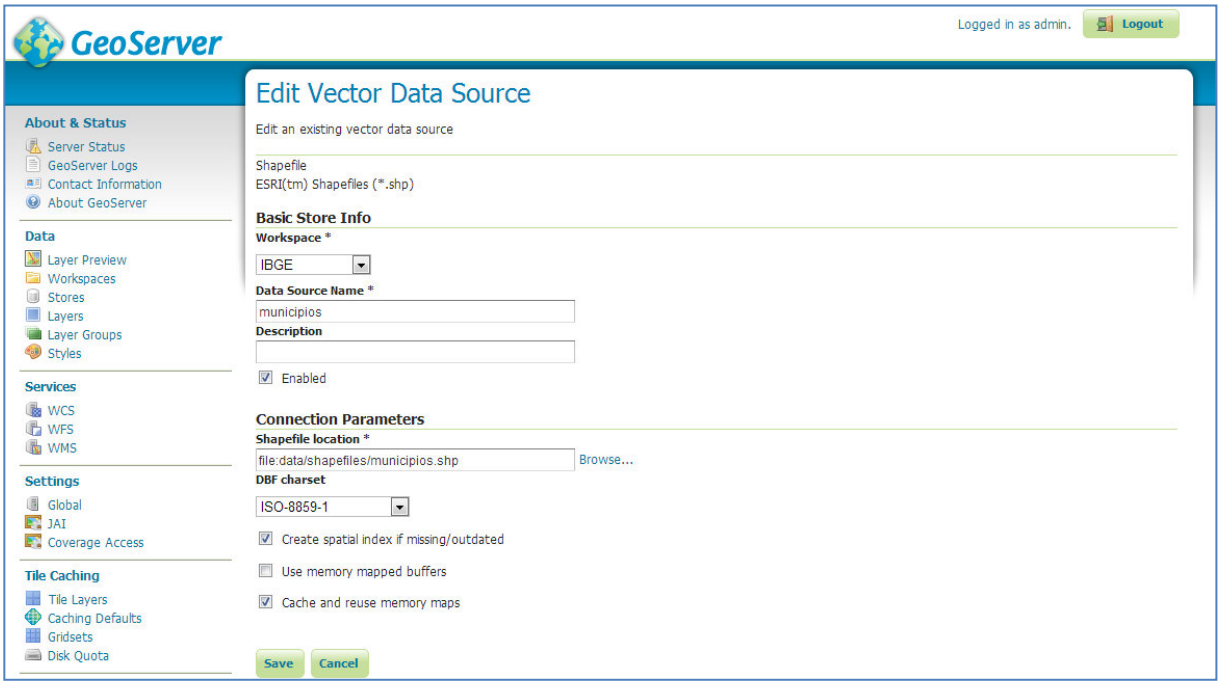


FIGURA 40 - INTERFACE DO SERVIDOR GEOSERVER.
FONTE: A AUTORA, 2013

O serviço web foi desenvolvido com a utilização da linguagem *Javascript*. No primeiro passo indicado na arquitetura, que corresponde à interface do serviço *web* com o usuário é aqui representada pela seleção de um município a partir de uma lista com os nomes dos municípios do Paraná, disponibilizada um *combo-box*, como mostrado na FIGURA 41.

FIGURA 41: CAIXA DE SELEÇÃO NA INTERFACE.
FONTE:A AUTORA, 2013

Com a ativação do botão “Recuperar Coordenadas”, é executado o código que foi desenvolvido para enviar uma requisição para o servidor WFS que segue o seguinte modelo, utilizando a o padrão *Filter Enconding* com o parâmetro de busca do código do município escolhido.

REQUISIÇÃO PARA O SERVIÇO WFS

```
http://NOSSO_SERVIDOR/geoserver/wfs?request=getfeature&
version=1.0.0&
outputformat=GML2
&service=wfs
&typename=servidor:municipios
&filter=<Filterxmlns="http://www.opengis.net/ogc">
<PropertyIsEqualTo>
<PropertyName>GEOCODIGO</PropertyName>
<Literal>4106902</Literal>
</PropertyIsEqualTo>
</Filter>
```

No terceiro passo, o WFS envia esta solicitação ao banco de dados geográfico e, em seguida, o banco retorna ao *Geoserver* a informação geoespacial que é encaminhada no formato de arquivo GML ao serviço. A parte seguinte do código desenvolvida recupera as coordenadas descritas entre as *tags* <gml:coordinates> no arquivo GML de entrada. Estas coordenadas são as dos pontos pertencentes ao perímetro do polígono que representa o município escolhido. Na sequência, o programa insere estas coordenadas no modelo de RDF existente no servidor, que tinha sido criado com a ontologia definida nas etapas anteriores. O resultado é a criação dinâmica, através do serviço *web*, do arquivo desejado em formato RDF, como mostrado na tela da interface na FIGURA 42.

FIGURA 42 - INTERFACE COM O USUÁRIO COM O RDF CRIADO DINAMICAMENTE.
FONTE: A AUTORA, 2013

Em seguida o arquivo em formato RDF criado pode ser usado para a realização de consultas usando o padrão GeoSPARQL, além de diversas aplicações usando dados interligados abertos, sendo que como ele é gerado sob demanda, sempre estará retratando a realidade do banco de dados geográfico oficial no momento da consulta.

7. RESULTADOS

Como resultado da arquitetura proposta, o serviço *web* descrito no capítulo da Metodologia permite a execução de consultas usando o padrão GEOSPARQL em bases de dados publicadas no formato RDF e informações geográficas disponibilizadas através de um serviço WFS. Em seguida serão apresentados três exemplos que demonstram as potencialidades deste tipo de consulta.

Em função do pouco tempo de existência do padrão, ainda não foi possível encontrar *endpoints* GeoSPARQL totalmente operacionais que funcionem com qualquer repositório RDF escolhido. Os exemplos aqui retratados utilizam a sintaxe de busca definida pelo padrão e o resultados foram simulados com os dados em formato de arquivo vetorial *shapefile* no *software* QuantumGIS Lisboa 1.8.0.

7.1. CONSULTA GEOSPARQL EM DUAS BASES DE DADOS: GEONAMES (RDF) e LIMITES MUNICIPAIS (WFS)

O primeiro exemplo seleciona pontos publicados na base Geonames usando o critério espacial de que os mesmos devem estar contidos em um polígono da Divisão Político Administrativa publicada em WFS e disponibilizada em formato RDF pelo serviço *web* de integração. Cabe destacar a facilidade que o padrão GeoSPARQL proporciona para consultas em bases distribuídas, bastando a declaração das URIs das mesmas, além de colocar à disposição do usuário toda série de relacionamentos espaciais.

Foram utilizados os pontos do Geonames, publicado em formato RDF, com o vocabulário W3C geo. Para compatibilizar este vocabulário com o GeoSPARQL é necessária uma etapa adicional na consulta:

```
PREFIX w3cGeo: <http://www.w3.org/2003/01/geo/wgs84_pos#>

PREFIX geo: <http://www.opengis.net/#geosparql>

SELECT (STRDT(CONCAT("POINT(",?long," ",?lat,")"),
geo:wktLiteral) AS ?wktLit)
WHERE { ?point w3cGeo:long ?long .

        ?point w3cGeo:lat ?lat }
```

Para acessar os dados RDF, foi utilizado o arquivo chamado *dump* do site Geonames, para que fosse convertido para o formato CSV e então adicionado através das coordenadas latitude e longitude, no *software* QuantumGIS. A consulta em si seleciona os pontos Geonames que se encontram dentro do polígono do município de Curitiba. O resultado da consulta pode ser observado na FIGURA 43 e na TABELA 7.

Sintaxe da Consulta:

```
PREFIX admin: <http://NOSSO_SERVIDOR/admin/>
PREFIX gn: < http://sws.geonames.org/>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
PREFIX w3cGeo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
SELECT ?f
WHERE { admin:4106902 admin:hasExactGeometry ?aGeom .
?aGeom geo:asGML ?aGML .

     SELECT (STRDT(CONCAT("POINT(",?long," ",?lat,")"),
     geo:wktLiteral) AS ?wktLit)
     WHERE { ?f w3cGeo:long ?long .
     ?f w3cGeo:lat ?lat }
?f gn:hasExactGeometry ?fGeom .
?fGeom geo:asGML ?fGML .
FILTER (geof:sfContains(?aGML, ?fGML) &&
!sameTerm(?aGeom, ?fGeom))
```

TABELA 7 - RESULTADO DA CONSULTA

geonameid	name	latitude	longitude	feature co	pop	dem
3444832	Xaxim	-25,4833	-49,2667	PPL	0	909
3445819	Umbará	-25,5672	-49,2853	PPLL	0	919
3446688	Tatuquara	-25,5536	-49,3097	PPLL	0	933
3450231	Santa Felicidade	-25,3936	-49,3500	PPLL	0	985
3464975	Curitiba	-25,4278	-49,2731	PPLA	1718421	924
3465089	Cruzeiro	-25,5167	-49,3333	PPL	0	884
3467786	Campo Comprido	-25,4425	-49,3425	PPLL	0	960
3468546	Cachimba	-25,6167	-49,3500	PPL	0	881
3469218	Boqueirão	-25,4837	-49,2502	PPL	0	885

3471324	Atuba	-25,4333	-49,2333	PPL	0	919
6300603	Curitiba	-25,4051	-49,2320	AIRP	0	926
6316472	Bacacheri	-25,3997	-49,2417	PPLL	0	940
6317794	Novo Mundo	-25,4792	-49,2939	PPLL	0	935
6317963	Pinheirinho	-25,5286	-49,3017	PPLL	0	928
6318286	Santa Quitéria	-25,4689	-49,2983	PPLL	0	928
6322752	Curitiba	-25,5040	-49,2908	ADM2	1746896	917
6465723	Alta Reggia Plaza	-25,4166	-49,2500	HTL	0	912
6495285	Pestana Curitiba Hotel	-25,4403	-49,2818	HTL	0	929
6500424	Hotel Rockefeller Plaza	-25,4456	-49,2600	HTL	0	892
6500553	Slaviero Palace Hotel	-25,4339	-49,2731	HTL	0	910
6501119	Hotel Executive Curitiba	-25,4362	-49,2668	HTL	0	897
6507507	Vernon Palace	-25,4166	-49,2500	HTL	0	912
6508687	Bristol Ambassador Flat	-25,4166	-49,2500	HTL	0	912
6508951	Novotel Sao Bento Do Sul	-25,4166	-49,2500	HTL	0	912
6511918	Grand Hotel Rayon	-25,4166	-49,2500	HTL	0	912
6519128	Hotel Deville Curitiba	-25,4166	-49,2500	HTL	0	912
6520289	Comfort Hotel Curitiba	-25,4166	-49,2500	HTL	0	912
6521262	Slaviero Rockefeller Hotel	-25,4166	-49,2500	HTL	0	912
6521336	Slaviero Full Jazz	-25,4166	-49,2500	HTL	0	912
6526702	Quality Hotel Curitiba	-25,4166	-49,2500	HTL	0	912
6527551	Bristol Saint Emilion	-25,4166	-49,2500	HTL	0	912
6528110	Bristol Dom Ricardo Hotel	-25,4166	-49,2500	HTL	0	912
6547122	Rio Belém	-25,3806	-49,2664	WTRC	0	927
6948435	Shopping Barigui	-25,4361	-49,3172	MALL	0	895
6948438	Carrefour	-25,4369	-49,3187	MALL	0	896
7279606	Altareggia Plaza Hotel - Curitiba	-25,4307	-49,2606	HTL	0	903
7279627	Astron Flat Curitiba	-25,4321	-49,2598	HTL	0	899
7279684	Bonaparte Express Curitiba	-25,4334	-49,2598	HTL	0	900
7279691	Bourbon Batel Express Hotel	-25,4443	-49,2875	HTL	0	935
7279780	Bristol Flexy Upper	-25,4267	-49,2542	HTL	0	909
7279789	Bristol Dobly Brasil 500 Hotel	-25,4416	-49,2778	HTL	0	931
7279818	Bristol Upper Residence	-25,4166	-49,2500	HTL	0	912
7279832	Caravelle Palace Hotel	-25,4304	-49,2753	HTL	0	910
7279930	Hotel Crowne Plaza Curitiba	-25,4260	-49,2708	HTL	0	911
7279963	Duomo Park Hotel	-25,4360	-49,2771	HTL	0	921
7280043	Four Points by Sheraton Curitiba	-25,4426	-49,2793	HTL	0	948
7280080	Granville Park Hotel	-25,4302	-49,2770	HTL	0	912
7280157	Hotel Centro Europeu Tourist	-25,4329	-49,2751	HTL	0	918
7280167	Hotel Tulip Inn Batel	-25,4389	-49,2791	HTL	0	924
7280203	Hotel Promenade	-25,4343	-49,2609	HTL	0	895
7280280	Lancaster Hotel e Business	-25,4343	-49,2730	HTL	0	910
7280333	Mabu Parque Resort	-25,5153	-49,3205	HTL	0	923
7280401	Monte Castello Hotéis	-25,4332	-49,2824	HTL	0	915
7280795	Roochelle Park Hotel	-25,4294	-49,2640	HTL	0	905
7280828	San Juan Executive	-25,4363	-49,2634	HTL	0	897

7280829	Hotel San Juan Johnscher	-25,4332	-49,2678	HTL	0	901
7280830	San Juan - Royal	-25,4202	-49,2689	HTL	0	912
7280843	Hotel Savoy	-25,4341	-49,2648	HTL	0	900
7280859	Hotel Slaviero Slim	-25,4250	-49,2604	HTL	0	920
7280889	Hotel Slaviero Villaggio	-25,4351	-49,2615	HTL	0	898
7280960	Transamerica Flat Curitiba	-25,4433	-49,2898	HTL	0	929
7281098	Bonaparte Express Park Barigui	-25,4325	-49,3130	HTL	0	902
7281101	Bristol Multy Ambassador Flat	-25,4336	-49,2794	HTL	0	919
7281832	Bristol Sthay Metropolitan Flat	-25,4349	-49,2749	HTL	0	918
7282280	Mabu Royal & Premium Hotel	-25,4289	-49,2666	HTL	0	909
7282410	San Juan Charm Curitiba Hotel	-25,4332	-49,2678	HTL	0	901
7283786	Hotel Deville Express Curitiba	-25,4279	-49,2654	HTL	0	906
7283788	Hotel Deville Rayon	-25,4347	-49,2763	HTL	0	910
7284063	Hotel Tulip Inn Santa Felicidade	-25,4070	-49,3253	HTL	0	962
7290282	Universidade Federal do Paraná?	-25,4285	-49,2665	UNIV	0	906

FONTE: Geonames

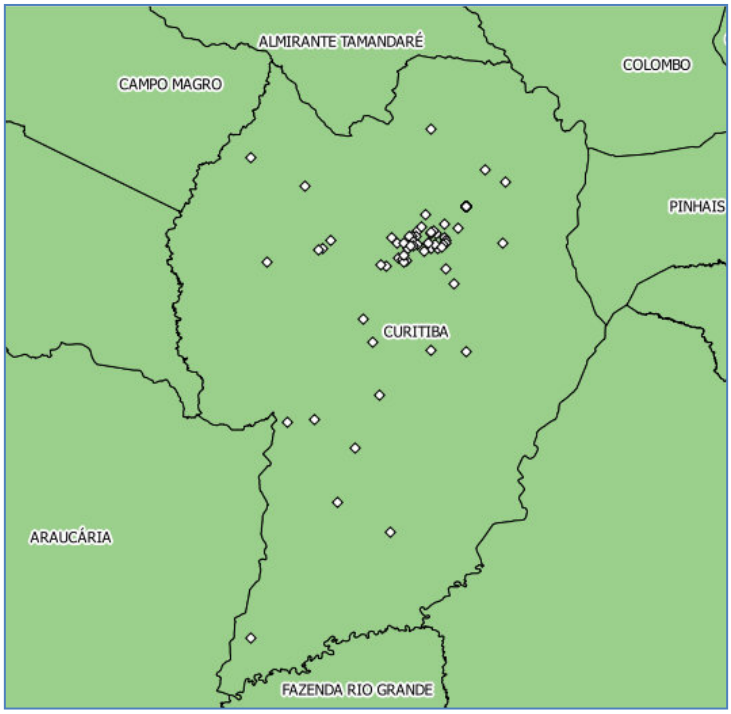


FIGURA 43 - PONTOS DO PROJETO GEONAMES DENTRO DO POLÍGONO DO MUNICÍPIO DE CURITIBA
FONTE: A AUTORA, 2013

7.2. CONSULTA GEOSPARQL EM UMA MESMA BASE DE DADOS: LIMITES MUNICIPAIS (WFS)

Este segundo exemplo de consulta GeoSPARQL mostra como a linguagem pode ser usada para realização de consultas em camadas disponibilizadas em WFS nativamente e integradas com o formato RDF pela arquitetura proposta. É portanto uma demonstração de que mesmo não envolvendo bases diferentes, com esta arquitetura temos a chance de usar pesquisas semânticas dentro de camadas de dados não publicados em formatos de dados interligados.

Esta consulta consiste, portanto, na seleção dentro de uma mesma camada, neste caso, dos limites municipais, usando o critério espacial de adjacência. São selecionados os polígonos adjacentes ao polígono escolhido, no exemplo o município de Curitiba. O resultado apresentado na FIGURA 44 é proveniente do processamento da requisição detalhada a seguir.

Sintaxe da Consulta:

```
PREFIX admin: <http://NOSSO_SERVIDOR/admin/#>

PREFIX geo: <http://www.opengis.net/ont/geosparql#>

PREFIX geof: <http://www.opengis.net/def/function/geosparql/>

SELECT ?f

WHERE { ?f admin:hasExactGeometry ?fGeom .
?fGeom geo:asGML ?fGML .
admin:4106902 admin:hasExactGeometry ?aGeom .
?aGeom geo:asGML ?aGML .
FILTER (geof:sfTouches(?fGML,?aGeom))
}
```

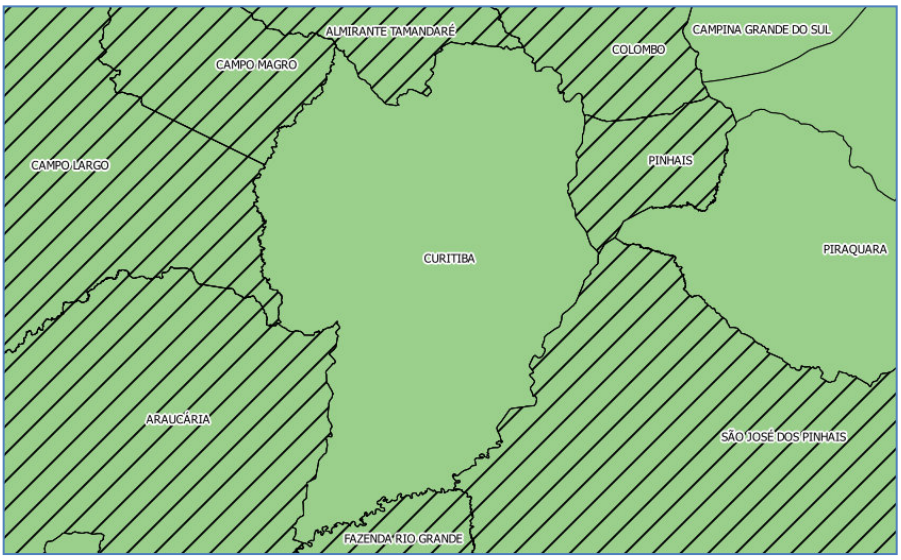


FIGURA 44 - RESULTADO DA SEGUNDA CONSULTA
FONTE: A AUTORA, 2013

7.3. CONSULTA GEOSPARQL EM DUAS BASES DE DADOS: DBPEDIA (RDF) e LIMITES MUNICIPAIS (WFS)

Uma terceira consulta utiliza novamente bases distribuídas, neste exemplo dados publicados em formato RDF pelo projeto DBpedia, que serão selecionados através do critério de uma distância definida a um polígono da camada WFS intermediada via serviço *web*. Será usada a classe *Geographic Coordinates* do DBpedia, que também está publicado usando o vocabulário W3C geo. Neste exemplo, foi especificada uma distância determinada (3°) do município de Curitiba; Os resultados podem ser observados na TABELA 8 e FIGURA 45.

Sintaxe da Consulta:

```
PREFIX admin: <http://NOSSO_SERVIDOR/admin/#>

PREFIX geo: <http://www.opengis.net/ont/geosparql#>

PREFIX geof: <http://www.opengis.net/def/function/geosparql/>

PREFIX ogcunit: http://www.opengis.net/def/uom/OGC/1.0/#>

SELECT ?f

WHERE { ?f admin:hasExactGeometry ?fGeom .
?fGeom geo:asGML ?fGML .
admin:4106902 my:hasExactGeometry ?aGeom .
?aGeom geo:asGML ?aGML .
```

```
FILTER (geof:buffer (?aGeom, 3,Ogcunit:degree): ogc:geomLiteral))
}
```

FIGURA 45 - RESULTADO DA TERCEIRA CONSULTA
FONTE: A AUTORA, 2013

TABELA 8 - RESULTADO DA TERCEIRA CONSULTA

DBPEDIA	LATITUDE	LONGITUDE
<http://dbpedia.org/resource/Balneario_Barra_do_Sul>	-26,3622	-48,6627
<http://dbpedia.org/resource/Fraiburgo>	-27,0261	-50,92138
<http://dbpedia.org/resource/Santa_Catarina_island>	-27,5516	-48,4788
<http://dbpedia.org/resource/Hercilio_Luz_Bridge>	-27,5938	-48,5661
<http://dbpedia.org/resource/Campos_Novos_Dam>	-27,6038	-51,3263
<http://dbpedia.org/resource/Morro_da_Igreja>	-28,1250	-49,4744
<http://dbpedia.org/resource/Pico_Parana>	-25,2497	-48,8130

FONTE: Dbpedia

7.4. POTENCIAL PARA UTILIZAÇÃO DE CONSULTAS GEOSPARQL INTEGRANDO DADOS DA INDE-BR

As consultas descritas nos exemplos anteriores são tais que poderiam ser executadas atualmente com dados já existentes publicados em formato de dados interligados abertos. Outras aplicações podem ser previstas caso outras informações, atualmente não disponíveis, viessem a ser fornecidas no formato RDF, como prevê o planejamento de implantação da INDA.

Mantendo a ênfase dada na camada de Divisão Política Administrativa, parte importante dos dados de referência da INDE-BR, existem muitas aplicações potenciais, das quais detalharemos duas.

Numa primeira aplicação seria necessário que as obras realizadas pelo Governo Federal fossem publicadas em formato RDF, com coordenadas. Neste caso, o usuário escolheria um município, o serviço *web* disponibilizaria o limite municipal em formato RDF vindo diretamente do WFS da INDE-BR, e através de uma consulta GeoSPARQL, seria possível selecionar as obras dentro do mesmo. O interessante é observar que, dentro do conceito de negociação de conteúdo, o RDF pode ser utilizado para execução da consulta e um WMS já formatado com a simbologia definida pode ser usada para exibir os resultados em forma de mapa. Portanto, para necessidades diferentes, a mesma informação armazenada em um único banco de dados espaciais pode ser fornecida em formatos diversos de acordo com a utilização.

Adicionalmente, com o crescimento de dados interligados publicados e com a adesão de novos atores à INDE-BR, usando padrão WFS, muitas novas aplicações mais complexas da arquitetura podem ser construídas. Um exemplo envolvendo dados de diversas fontes seria o esquema da FIGURA 46.

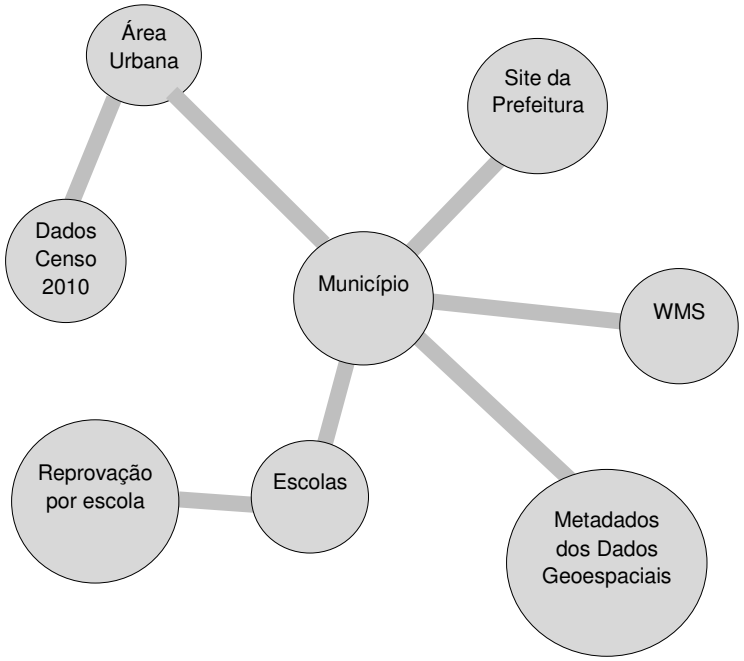


FIGURA 46 - APLICAÇÃO DE DADOS INTERLIGADOS ABERTOS DE DIVERSAS FONTES INTEGRADOS À INDE-BR
FONTE: A AUTORA, 2013

Nesta visão, imagina-se uma série de dados publicados no formato RDF. Relacionado a determinado município, é possível acessar a sua área urbana e dentro desta área, por exemplo, qual a população na faixa etária escolar proveniente do Censo 2010. Uma consulta espacial pode trazer os pontos de escolas situados neste município e, através desses, os dados do Censo Escolar do Ministério da Educação. Adicionalmente, *links* podem levar ao site da prefeitura, ao serviço WMS do mapa municipal, e aos metadados em formato ISO 19115/Perfil Brasileiro de Metadados Geoespaciais, onde detalhes da data do levantamento que originou os limites municipais, além do detalhamento do relatório de qualidade dos dados, podem estar disponíveis.

Há uma série de informações com potencial de serem interligadas, de forma a concretizar uma Infraestrutura Nacional de Dados Abertos reunindo dados dispersos em diversas bases, com a informação geoespacial oficial oriunda da INDE-BR como uma das formas mais claras de promover esta integração. Com esta abordagem implementada, ao se fazer uma pesquisa como a descrita no primeiro capítulo deste trabalho, que hoje exige uma série de passos para ser executada, passariam a ser utilizadas interfaces inovadoras com tecnologias semânticas com base em integrações como a arquitetura proposta neste trabalho.

8. CONCLUSÕES

8.1. AVALIAÇÂO DA SOLUÇÃO PROPOSTA

A contribuição deste trabalho foi demonstrar que é possível a integração de informações da INDE-BR com o formato de dados interligados abertos e assim fomentar uma série de aplicações semânticas. Com a arquitetura proposta, esta integração pode estar em funcionamento rapidamente de forma a suprir uma dificuldade já presente nas aplicações existentes no site <dados.gov.br>. O uso de um serviços mediador na arquitetura se mostrou uma solução eficiente e prática, ao menos até que novos padrões e plataformas tecnológicas para publicação de dados RDF estejam disponíveis. Desta forma também é evitada republicação de bases de dados *off-line* o que pode causar inconsistências, principalmente em temas de atualização constante.

A atualização contínua da base é uma tendência crescente, em consequência da dinâmica territorial e de legislação, associada à maior disponibilidade tecnológica de ferramentas para execução das atividades necessárias. O uso do serviço WFS como fornecedor dos dados também se mostra importante, por este já estar previsto no rol dos padrões tecnológicos da INDE-BR e estar incluso nas soluções de servidores de mapas livres e proprietárias.

Foi aproveitada a estrutura do formato GML, que possui diversas semelhanças estruturais com o formato RDF. Os demais padrões utilizados estão atualmente definidos, como GeoSPARQL e com isso não há necessidade de aguardar até que alterações nos demais padrões estejam consolidadas.

8.2. RECOMENDAÇÕES

Como visto na introdução, aplicações usando outras fontes de dados geoespaciais já estão sendo criadas usando fontes dados não oficiais, como o *Geonames*, sendo, portanto recomendável que a disponibilização de dados geoespaciais oficiais para a criação de aplicação com dados interligados aconteça com celeridade, ao menos para temas prioritários, entre os quais se incluem a

divisão político-administrativa, o banco de nomes geográficos e o cadastro de localidades, todos administrados pelo IBGE.

Apesar da tecnologia para integração estar atualmente disponível, como pôde ser observado no serviço *web* criado, evoluções em termos de adaptação de padrões, arranjos institucionais, treinamento e novas pesquisas e aplicações serão necessárias. Outro aspecto relevante é a interação das instituições brasileiras no ISO/TC211 e OGC para entre outras funções, auxiliar na transição dos padrões para *web* semântica, e a participação na criação e adaptação de padrões da área à realidade nacional.

Uma necessidade é adaptar as atribuições da CONCAR/CINDE para incluir a criação e gestão de ontologias. A criação da ontologia para a divisão político-administrativa brasileira neste trabalho permitiu observar os desafios em termos da gestão de conhecimento que decorrem deste processo. Pela dificuldade em termos de diversidade e abrangência de tal tarefa, é interessante a proposição de uma plataforma específica para gestão de ontologias. Neste aspecto, é recomendável a integração com iniciativas nacionais e internacionais que estejam em andamento, em especial o VCGE - Vocabulário Controlado do Governo Eletrônico e que seja contemplada em conjunto a gestão de URIs, com padronização de nomes, hierarquias e responsáveis, de acordo com as definições a serem estabelecidas pela INDA nas suas atribuições.

Como ponto de partida, contamos hoje com o detalhado e completo modelo da ET-EDGV, que hoje já está disponível no site do vocabulário do governo eletrônico em formato de Schema XML – XSD. No entanto, a especificação técnica não é uma ontologia, mas um modelo de dados específico para o mapeamento topográfico. Para isso, ela precisaria ser enriquecida em termos de sinônimos, termos regionais, relacionamento com outros vocabulários, atividade para a qual é imprescindível a presença de uma gama variada de atores que possam contribuir com o levantamento e documentação destes novos relacionamentos.

Seguindo os exemplos citados de outros países que foram estudados durante o desenvolvimento deste trabalho, é recomendável a criação de um site geo.dados.gov.br, para informações, dados e aplicações geoespaciais em formato aberto. A própria plataforma do site pode ser usada como distribuição de material de

ensino para a área, divulgando principalmente tutoriais para orientar a publicação de dados e a criação de aplicativos com os mesmos.

8.3. PESQUISAS FUTURAS

Os esforços para a integração de dados das IDEs no formato de dados interligados são relativamente recentes e ainda existem muitas demandas de pesquisas que são relevantes para a construção da base de conhecimento que possibilite a plena utilização destas visões de forma integrada. Orientar este processo pela preocupação com a visão do usuário é fundamental para se construir uma nova geração de IDEs que responda às várias críticas aos modelos centrados em dados e processos que foram o padrão para tais iniciativas no passado. Por exemplo, ferramentas para criação e gestão de ontologias de forma colaborativa também serão necessárias para apoiar estes processos.

Será importante testar formas de otimização na implementação do serviço *web* em um cenário real, pois, com muitos usuários, é possível existir sobrecarga por demanda excessiva, fato que atualmente leva os provedores de dados da INDE-BR a disponibilizar apenas WMS e não WFS.

Outro fator limitador das aplicações atuais é a falta de disponibilidade de *endpoints* que permitam buscas GeoSPARQL em múltiplas bases distribuídas, ferramenta que seria de muita utilidade para adição em um futuro portal semântico de informação geoespacial em formato de dados interligados aberto no Brasil. Adicionalmente, seria interessante abordar a de transformação de coordenadas via serviço *web*. Neste trabalho, todos os dados são referenciados ao Datum SIRGAS 2000, mas o GeoSPARQL tem a capacidade de trabalhar com a declaração do sistema de referência, e a arquitetura ficaria mais completa ao trabalhar com esta opção.

Imaginando um futuro acesso facilitado a diversas fontes interligadas, surgirá a necessidade de métodos que possibilitem a avaliação da qualidade dos dados obtidos de produtores diversos. Este tipo de método exigirá alguma forma ligação de metadados geoespaciais padronizados com a sua fonte de dados e permitirá o surgimento de aplicações que permitam a negociação de conteúdo e o fornecimento a cada aplicação cliente de acordo com o uso, a informação em formato RDF, WMS,

WFS, *download*, ou outra. Neste ambiente, será possível avaliar a aplicabilidade de soluções baseadas em dados geoespaciais interligados em diferentes áreas de conhecimento, como análise de risco ambiental, estudos demográficos, entre outros.

8.4. CONSIDERAÇÕES FINAIS

A principal expectativa é que a INDE-BR no futuro se consolide como o nó geoespacial da INDA, trazendo informação oficial de qualidade e procurando estender a *expertise* das instituições nacionais que trabalham há anos com o conhecimento geográfico do território, para que este possa ser usufruído pelo maior número de usuários de forma intuitiva e amigável.

Com seu papel fortalecido, a INDE-BR pode gerar sensibilização na administração pública para executar os investimentos que são tão necessários para construir uma base cartográfica atualizada e de qualidade. Os trabalhos semânticos podem proporcionar ganhos em termos de sinergia com atividades como definições a respeito de metadados, modelagem, tecnologia e capacitação, portanto podendo permear as atividades de todos os grupos de trabalho já estabelecidos do CINDE/CONCAR. A arquitetura de integração proposta neste trabalho nos mostra que alçar a INDE-BR à realidade da web semântica é uma tarefa que pode ser empreendida com tecnologia atualmente disponível e desta forma fazer com que ela seja mais amigável, colaborativa e relevante para a sociedade.

REFERÊNCIAS

ABADIE, N.; MUSTIERE, S. **Constitution et exploitation d'une taxonomie geographique a partir des specifications de bases de donnees**. Revue internationale de geomatique, v. 20, n. 2, p. 145-174, jun 2010. ISSN: 12605875 Disponivel em: <http://rig.revuesonline.com/article.jsp?articleId=14874>.

AGUNE, R. M.; GREGORIO FILHO, A. S.; BOLLIGER, S. **Governo aberto SP: disponibilização de bases de dados e informações em formato aberto**. CONGRESSO CONSAD DE GESTÃO PÚBLICA. Brasília 2010.

AHLQVIST, O. **Using uncertain conceptual spaces to translate between land cover categories**. International Journal of Geographical Information Science, v. 19, n. 7, p. 831, 2005. ISSN: 1365-8816 DOI: 10.1080/13658810500106729. Disponivel em: <http://www.informaworld.com/10.1080/13658810500106729>.

ALBITES, F. H. **Towards the Mexican National Spatial Data Infrastructure**. Procedeeings of GSDI 6. Budapeste, 2002.

ALBORNOS, L., AUCOTT, PAULA, ANDREN, P., GELFGREN, S., KOPPEL, K., MULRENIN, B., PALM, F., SOUTHALL, HUMPHREY, SVENSSON, P. AND TÜRNA, T. **QVIZ-query and context based visualization of time-spatial cultural dynamics: final activity report**, Project Report. Umeå University, Umeå, Sweden, 2008.

ALVESTRAND, H.; LIE, H. W. **Development of core Internet standards: the work of IETF and W3C**. In: BYGRAVE, L. A.; BING, J. Internet Governance: Infrastructure and Institutions. Oxford University Press, 2009. Cap. 4, p. 262.

ARCTUR, D. **Big Data vs SDI? It's not an either/or**. The OGC Blog, 2011. Disponivel em: <http://www.opengeospatial.org/blog/1492>. Acesso em: 31 jan. 2013.

ATHANASIS, N.; KALABOKIDIS, K.; AITIS, M. e SOULAKELLIS, N. **Towards a semantics-based approach in the development of geographic portals**. Computers & Geosciences, v. 35, n. 2, p. 301-308, feb 2009. ISSN LCCN: 0023 DOI: 10.1016/j.cageo.2008.01.014.

AUGUSTO, M. J. D. C.; FREITAS, A. L. B. D. **Esforços em Desenvolvimento para a Constituição da INDE – Brasil**. I Reunión Técnica conjunta de las comisiones del IPGH. 2007.

BARIK, R. K.; SAMADDAR, A. B.; SAMADDAR, S. G. **Service Oriented Architecture based SDI Model for Geographical Indication Web Services**. International Journal of Computer Applications, v. Volume 25– No.4, p. 42-49, Julho 2001.

BATTLE, R.; KOLAS, D. **Enabling the Geospatial Semantic Web with Parliament and GeoSPARQL.** Semantic Web, Volume 3, Number 4, p. 355-370, 2012

BECKER, C. **DBpedia – Extracting structured data from Wikipedia**. Wikimania 2009. Buenos Aires: 2009.

BENNETT, B. **What is a Forest? On the Vagueness of Certain Geographic Concepts**. Topoi 20, 2001. 189-201.

BERNARD, L.; CRAGLIA, M. **SDI-from spatial data infrastructure to service driven infrastructure**. Research Workshop on Cross-Learning Between Spatial Data Infrastructures and Information Infrastructures. Enschede, The Netherlands: 2005.

BERNERS-LEE, T. **Linked Data—Design Issues**. W3C, 2006. Disponivel em: <http://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 31 jan. 2013.

BERNERS-LEE, T**. Semantic Web on XML**. XML 2000. Washington DC: [s.n.]. 2000. Diponível em: http://www.w3.org/2000/Talks/1206-xml2k-tbl.

BERNERS-LEE, T. **Universal Resource Identifiers in WWW: A Unifying Syntax for the Expression of Names and Addresses of Objects on the Network as used in the World-Wide Web**. 1994. Disponível em: http://www.hjp.at/doc/rfc/rfc1630.html.

BERNERS-LEE, T.; FIELDING, R.; MASINTER, L. **Uniform Resource Identifier URI: Generic Syntax**, 2005. Disponivel em: <http://www.ietf.org/rfc/rfc3986.txt>. Acesso em: 31 jan. 2013.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. **The Semantic Web**. Scientific American, Maio 2001. 29-37.

BERNERS-LEE, T.; SHADBOLT, N. **Our manifesto for government data**. The Guardian, jan 2010. Disponivel em: <http://www.guardian.co.uk/news/datablog/2010/jan/21/timbernerslee-government-data>. Acesso em: 31 jan. 2013.

BISHR, Y. **Overcoming the Semantic and Other Barriers to GIS Interoperability**. International Journal of Geographical Information Science, v. 12, p. 299-314, 1998. Disponivel em: <http://www.informaworld.com/smpp/content~content=a713811417~db=all~tab=content~order=page>.

BIZER,C; LEHMANN,J; KOBILAROV,G ;AUER,S. ;BECKER,C ;CYGANIAK,R. e HELLMANN,S. **DBpedia - A crystallization point for the Web of Data.** Web Semantics: Science, Services and Agents on the World Wide Web, v. 7, n. 3, p. 154-165, set. 2009. ISSN ISSN: 1570-8268 LCCN: 0544 DOI: 10.1016/j.websem.2009.07.002. Disponivel em: <http://dx.doi.org/10.1016/j.websem.2009.07.002>.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. **Linked Data - The Story So Far**. Special Issue on Linked Data, International Journal on Semantic Web and Information Systems (IJSWIS), 2009. 22.

BORBA, R.L.R.; STRAUCH, J.C.M.; ESTEVES, M.G.P.; SOUZA, J.M**. INDE - Co: Infraestrutura Nacional de Dados Espaciais Colaborativa**. Em: Anais de 1º

Seminário de Metodologia do IBGE e a XI Reunião IASI sobre Estatística Pública, Rio de Janeiro, RJ. 2012

BORGES, K.A.V., DAVIS, C.A. & LAENDER, A.H.F., 2001. **OMT-G: An Object-Oriented Data Model for Geographic Applications**. Geoinformatica, 5(3), 221-260.

BRADNER, S. **The Internet Standards Process -- Revision 3**. IETF Request for Comments, 1996. Disponivel em: <http://tools.ietf.org/html/rfc2026.html>. Acesso em: 31 jan. 2013.

BRASIL. **Padrões de Interoperabilidade de Governo Eletrônico**, 2012. Disponivel em: <http://www.governoeletronico.gov.br/acoes-e-projetos/e-ping-padroes-de-interoperabilidade>. Acesso em: 31 jan. 2013.

BRASIL. **Plano de Ação da Infraestrutura Nacional de Dados Abertos**, 2012a. Disponivel em: <http://notas.dados.gov.br/planodeacaoinda/>. Acesso em: 31 jan. 2013.

BRASIL. **Portal Brasileiro de Dados Abertos**, 2012b. Disponivel em: <http://dados.gov.br/>. Acesso em: 31 jan. 2013.

BRASIL. **Plano de Ação do Governo Brasileiro Parceria para o Governo Aberto**, 2011. Disponivel em: <http://www.acessoainformacao.gov.br/acessoainformacaogov/acesso-informacao-mundo/governo-aberto/ogp-brazil-actionplan.pdf>. Acesso em: 31 jan. 2013.

BRATT, S. **Semantic Web, and Other Technologies to Watch**. W3CTechSemWeb, 2007.

BULEN, A.; CARTER, J. J.; VARANKA, D. E. **A Program for the Conversion of The National Map Data from Proprietary Format to Resource Description Framework (RDF)**. U.S. Geological Survey. Reston, Virginia, p. 18. 2011.

CAI, G., 2007. **Contextualization of Geospatial Database Semantics for Human--GIS Interaction**. Geoinformatica, 11(2), 217-237.

CÂMARA, G**. Representação Computacional de Dados Geográficos**. In: CASANOVA, M., *et al*. Bancos de Dados Geográficos. Curitiba: MundoGEO, 2005.

CLARKIN, L.; HOLMES, J. **Enterprise Mashups**. Architecture Journal, v. 13, Outubro 2007.

CONCAR. **Especificações Técnicas Para Estruturação de Dados Geoespaciais Digitais Vetoriais – ET-EDGV** v 2.0. p. 213. 2007.

CONCAR. **Perfil de Metadados Geoespaciais do Brasil** (Perfil MGB), 2009. Disponivel em: <http://www.concar.ibge.gov.br/arquivo/Perfil_MGB_Final_v1_homologado.pdf>. Acesso em: 31 jan. 2013

CONCAR. **Plano de Ação para Implantação da Infraestrutura Nacional de Dados Espaciais**. p. 205. 2010.

CONCAR. **Apresentação**. Comisssão Nacional de Cartografia, 2012. Disponivel em: <www.concar.ibge.gov.br>. Acesso em: 31 jan. 2013.

CORCHO, O.; PÉREZ, A. G. **Mini-curso sobre Linked Data**. 3º OntoBras 2010. Florianópolis: 2010.

CRAGLIA, M.; ANNONI, A. **INSPIRE: an innovative approach to the development of spatial data infrastructures in Europe**. In: ONSRUD, H. Research and Theory in Advancing Spatial Data Infrastructure Concepts. Redlands, CA: ESRI Press, 2007. p. 93-105.

CROMPVOETS, J. *et al*. **Assessing the worldwide developments of national spatial data clearinghouses**. International Journal of Geographical Information Science, v. 18, n. 7, p. 665-689, 2004.

CYGANIAK, R.; JENTZSCH, A. The Linking Open Data cloud diagram, 2011. Disponivel em: <http://richard.cyganiak.de/2007/10/lod/>. Acesso em: 31 jan. 2013.

DAVIS JR, C. A.; ALVES, L. L. **Local Spatial Data Infrastructures based on a service-oriented architecture**. BRAZILIAN SYMPOSIUM ON GEOINFORMATICS. Proceedings. [S.l.]: [s.n.]. 2005. p. 30-45.

DE SOUSA, L. G.; LEITE, J. C. **Geração de dicionários explicativos para melhorar a interação entre um SIG e seus usuários**. VII Simpósio Brasileiro de Geoinformática. Campos do Jordão: [s.n.]. 2005. p. 301-317.

DELOU, A. **Estruturação de Bases Cartográficas Para SIG: Um Estudo Comparativo dos Modelos do IBGE e da DSG**. Dissertação de Mestrado em Engenharia de Computação, Programa de Pós-Graduação em Engenharia de Computação. Área de Concentração: Geomática, UERJ. 2006.

DEVOGELE, T.; PARENT, C.; SPACCAPIETRA, S. **On spatial database integration**. International Journal of Geographical Information Science, v. 12, n. 4, p. 335-352, 1998. ISSN: 1365-8816 DOI: 10.1080/136588198241824. Disponivel em: <http://www.tandfonline.com/doi/abs/10.1080/136588198241824>.

DILLI, R.M. **Explorando Modelos Semânticos para Descoberta de Recursos na Computação Ubíqua.** Universidade Católica de Pelotas - Centro Politécnico - Programa de Pós-Graduação em Informática. 2009

DING,L.; LEBO,T.; ERICKSON,J.S.; DIFRANZO,D.; GRAVES,A.; WILLIAMS,G.T.; LI,X.; MICHAELIS,J.; ZHENG,J.; SHANGGUAN,Z.; FLORES,J.; MCGUINNESS,D.L. e HENDLER,J.A. **TWC LOGD: A portal for linked open government data ecosystems**. Web Semantics: Science, Services and Agents on the World Wide Web, v. 9, n. 3, p. 325-333, 2011. ISSN ISSN: 1570-8268 LCCN: 0025 DOI: 10.1016/j.websem.2011.06.002. Disponivel em: <http://www.sciencedirect.com/science/article/pii/S1570826811000382>.

DINIZ, V. **Como conseguir dados governamentais abertos**. III Congresso Consad de Gestão Pública. Brasília: 2009.

DOLBEAR, C.; HART, G. **Combining spatial and semantic queries into spatial databases**. Athens, Georgia: Nov, 2006.

DOLBEAR, C., HART, G. & GOODWIN, J.. **What OWL has done for geography and why we don't need it to map read**. Em Proceedings of the OWL Experiences and Directions Workshop. Athens, Georgia, 2006.

EAVES, D. **The Three Laws of Open Government Data**, 2009. Disponivel em: <http://eaves.ca/2009/09/30/three-law-of-open-government-data/>. Acesso em: 31 jan. 2013.

EVANS, J. D. **A Geospatial Interoperability Reference Model (G.I.R.M.)**. FGDC Geospatial Applications and Interoperability Working Group, 2003. Disponivel em: <http://www.fgdc.gov/participation/coordination-group/meeting-minutes/2003%20meeting%20minutes/may/GIRM_FGDC-CWG_2003-05-06.ppt>. Acesso em: 31 jan. 2013.

FALLAHI, G.; MESGARI, M. S.; TEHRAN, I. **Resolving Semantic Ambiguities of Geo Services as New Method for Linking GIS and Environmental Models.** Proceedings of GSDI 9. Santiago, Chile, 2006

FEIGENBAUM, E.; MCCORDUCK, P. **The fifth generation: artificial intelligence and Japan's computer challenge to the world**. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, 1983

FERNÁNDEZ, T. D.; CASTELLANOS, E. **Towards user-driven spatial data infrastructures**. An approach oriented to sustainable development. Proceedings of GSDI 9. Santiago, Chile: 2006.

FERNÁNDEZ, T. D.; FERNÁNDEZ, J. L. C. **Semantic Spatial Data Infrastructures**. Proceedings of ICC 2009. Santiago, Chile: 2009.

FERREIRA, K.R.; CASANOVA, M.A.; DE QUEIROZ, G.R.; DE OLIVEIRA, O.F. **Arquiteturas e linguagens**. In: CASANOVA, M.A.; CÂMARA, G.; DAVIS JR., C.; VINHAS, L.; DE QUEIROZ, G.R. (Ed.) Bancos de Dados Geográficos. Curitiba: Mundogeo, 2005.

FIELDING, R.; GETTYS, J.; MOGUL, J; FRYSTYK, H.; MASINTER, L; LEACH, P. e BERNERS-LEE, T. **Hypertext Transfer Protocol -- HTTP/1.1**. W3C Protocols, 1999. Disponivel em: <http://www.w3.org/Protocols/rfc2616/rfc2616.html>. Acesso em: 31 jan. 2013.

FONSECA, F.; SHETH, A. **Research Priorities: The Geospatial Semantic Web** – The University Consortium for Geographic Information Science, 2002. Disponivel em: <http://www.ucgis.org/priorities/research/2002researchPDF/shortterm/e_geosemantic_web.pdf>. Acesso em: 31 jan. 2013.

FRIIS-CHRISTENSEN, A.; BERNARD, L.; KANELLOPOULOS, I.; NOGUERAS-ISO, J.; PEEDELL, S.; SCHADE, S. e THORNE, C. **Building Service Oriented**

Applications on top of a Spatial Data Infrastructure - A Forest Fire Assessment Example. 9th AGILE Conference on Geographic Information Science. Visegrád, Hungary: 2006. p. 119-127.

FROZZA, A. A.; MELLO, R. S. **Um Método para Determinar a Equivalência Semântica entre Esquemas GML**. Geoinfo. 2006. p. 283-293.

GANTNER, F. **A Spatiotemporal Ontology for the Administrative Units of Switzerland**. UNIVERSITY OF ZURICH, DEPT. OF GEOGRAPHY, GIS UNIT, 2011.

GEOCONNECTIONS. **Filter Encoding** , 2012. Disponivel em: <http://geoconnections.nrcan.gc.ca/991>. Acesso em: 31 jan. 2013.

GEONAMES. **About GeoNames**, 2013. Disponivel em: <http://www.geonames.org/about.html>. Acesso em: 31 jan. 2013.

GEOSPATIAL PLATFORM. **About Geospatial Platform. Geospatial Platform Website**, 2011. Disponivel em: <http://www.geoplatform.gov/home/>. Acesso em: 31 jan. 2013.

GOMES JR, L. C.; MEDEIROS, C. B. **Ecologically-aware Queries for Biodiversity Research**. Proceedings of Geoinfo, 2007.

GOODCHILD, M. F. **Citizens as sensors: the world of volunteered geography**. GeoJournal 69, 2007. 211-221.

GOODCHILD, M. F. ESTES, J. E.; BEARD, K. e FORESMAN, T. **Multiple Roles for GIS in US Global Change Research**. REPORT OF THE SECOND SPECIALIST MEETING. Santa Fe, New Mexico: 1996.

GOODCHILD, M. F.; GLENNON, J. A. **Crowdsourcing geographic information for disaster response: a research frontier**. International Journal of Digital Earth, v. 3, n. 3, p. 231-241, 2010. DOI: 10.1080/17538941003759255. Disponivel em: <http://www.tandfonline.com/doi/abs/10.1080/17538941003759255>.

GOODWIN, J. **Location and Linked Data**. AGI Evening Lecture on Linked Data. 2012.

GOODWIN, J.; DOLBEAR, C.; HART, G. **Geographical Linked Data: The Administrative Geography of Great Britain on the Semantic Web**. Transactions in GIS, Volume 12, Issue Supplement s1, p. 19–30, December 2008

GRANELL, C.; DIAZ, L.; GOULD, M. **Service-oriented applications for environmental models: Reusable geospatial services**. Environmental Modelling & Software, v. 25, n. 2, p. 182-198, feb 2010.

GRANELL, C.; SCHADE, S.; HOBONA, G. **Linked Data: Connecting Spatial Data Infrastructures and Volunteered Geographic Information**. In: ZHAO, P.; DI, L. Geospatial Web Services: Advances in Information Interoperability. [S.l.]: IGI Global, 2010. p. 189-226.

GREGORIUS, T. **From Lego to Play-Doh: I plead guilty at the altar of Big Data**. Georeferenced, 2011. Disponivel em: <http://georeferenced.wordpress.com/>. Acesso em: 31 jan. 2013.

GRUS, L., BREGT, A. e CROMPVOETS, J. **Defining National Spatial Data Infrastructures as Complex Adaptive Systems**, Proceedings GSDI-9 Conference, 6–10 November, Santiago, Chile, 2006.

GUARINO, N. **Formal Ontology and Information Systems**. Proceedings of FOIS'98, Trento, Italy. Amsterdam: IOS Press. 1998. p. 3-15.

HAAR, P. T. **Mashing up the SDI An NMA's Perspective**. Proceedings of GSDI 11. Rotterdam, The Netherlands: 2009.

HART, G.; JOHNSON, M.; DOLBEAR, C. **Rabbit: Developing a Control Natural Language for Authoring Ontologies**. Proceedings of the 5th European Semantic Web Conference. Tenerife, Spain: 2008.

HE, H. **What Is Service-Oriented Architecture**, 2003. Disponivel em: <http://www.xml.com/pub/a/ws/2003/09/30/soa.html>. Acesso em: 31 jan. 2013.

HESS, G. N.; IOCHPE, C.; CASTANO, S. **An algorithm and implementation for geoontologies integration**. 2006.

IBGE. **Resultados do Censo 2010**, 2011. Disponível em <http://censo2010.ibge.gov.br/> Acesso em: 31 jan. 2013.

IBGE. **Mapeamento das Unidades Territoriais**, 2012. Disponivel em: <http://www.ibge.gov.br/home/geociencias/cartografia/default_territ_int.shtm?c=3>. Acesso em: 31 jan. 2013.

IBM. **IBM What is big data? - Bringing big data to the enterprise**. IBM, 2012. Disponivel em: <http://www-01.ibm.com/software/data/bigdata/>. Acesso em: 31 jan. 2013.

IGLESIAS, C. A. **SDI in Chile - National System of Coordination of Territorial Information (SNIT)- State of the art and Projections**. Proceedings of the GSDI 10. Trinidad y Tobago: [s.n.]. 2008.

ISO. **The International Organization for Standardization**, 2012. Disponivel em: <http://www.iso.org/iso/home.htm>. Acesso em: 31 jan. 2013.

ISO/IEC. **ISO 19501:2005 Information technology -- Open Distributed Processing -- Unified Modeling Language (UML) Version 1.4.2**. ISO/IEC, 2005

ISO/IEC. **ISO 9075-1:2008 Information technology -- Database languages -- SQL -- Part 1: Framework (SQL/Framework)**. ISO/IEC, 2008

ISO/TC211. **ISO 19115:2003 Geographic information -- Metadata**. 2003.

ISO/TC211. **ISO 19128:2005 Geographic information -- Web map server interface.** ISO/TC211, 2005

ISO/TC211. **ISO 19136:2007 Geographic information -- Geography Markup Language (GML).** ISO/TC211, 2007.

ISO/TC211. **ISO 19139:2007 Geographic information -- Metadata -- XML schema implementation**. ISO/TC211, 2007a.

ISO/TC211. **ISO 19115:2-Geographic information – Metadata – Part 2: Extensions for imagery and gridded data**. ISO/TC211, 2009.

ISO/TC211. **Standards Guide - ISO/TC211 Geographic Information/Geomatics**. ISO/TC211. p. 98. 2009a.

ISO/TC211. **ISO 19142:2010 Geographic information -- Web Feature Service. ISO/TC211**, 2010.

ISO/TC211. **ISO 19143:2010 Geographic information -- Filter encoding**. ISO/TC211, 2010b.

ISO/TC211. **ISO/TS 19150-1:2012 Geographic information -- Ontology -- Part 1: Framework**. ISO/TC211, 2012c.

ISO/TC211. **About ISO/TC211**, 2012. Disponivel em: <http://www.isotc211.org/>. Acesso em: 31 jan. 2013.

JANOWICZ, K.; SCHADE, S.; BRORING, A.; KESSLER, C.; MAUE, P. e STASCH, C. **Semantic enablement for spatial data infrastructures**. Transactions in GIS, v. 14, n. 2, p. 111-129, 2010.

JANOWICZ, K.; WILKES, M. **SIM-DLA: A Novel Semantic Similarity Measure for Description Logics Reducing Inter-concept to Inter-instance Similarity**. In: AROYO, L., *et al*. The Semantic Web: Research and Applications. Springer Berlin Heidelberg, v. 5554, 2009. p. 353-367.

KIEHLE, C.; GREVE, K.; HEIER, C. **Requirements for Next Generation Spatial Data Infrastructures-Standardized Web Based Geoprocessing and Web Service Orchestration**. Transactions in GIS, v. 11, n. 6, p. 819-834, 2007.

KLEIN, M. **XML, RDF, and relatives**. Intelligent Systems, IEEE, v. 16, n. 2, p. 26-28, mar-apr 2001. ISSN ISSN: 1541-1672 DOI: 10.1109/5254.920596.

KUHN, W. **Semantic reference systems**. International Journal of Geographical Information Science, v. 17, n. 5, p. 405-409, 2003. DOI: 10.1080/1365881031000114116.

KULK, S.; VAN LOENEN, B. **Brave New Open Data World?** SSRN Electronic Journal, 2012

KWAN, M.-P.; LEE, J. **Emergency response after 9/11: the potential of real-time 3D GIS for quick emergency response in micro-spatial environments.** Computers, Environment and Urban Systems, v. 29, n. 2, p. 93-113, 2005.

LACASTA, J.; NOGUERAS-ISO, J.; BEJAR, R.; MURO-MEDRANO, P. R. e ZARAZAGA-SORIA, F. J. **A Web Ontology Service to facilitate interoperability within a Spatial Data Infrastructure: Applicability to discovery**. Data Knowl. Eng., Amsterdam, The Netherlands, The Netherlands, v. 63, n. 3, p. 947-971, dec 2007.

LAITURI, M.; KODRICH, K. **On line disaster response community: People as sensors of high magnitude disasters using Internet**. Sensors, 8(5), p. 3037-3055, 2008

LEITE JR., F.L.; BAPTISTA, C.de S.; SILVA, P. de A. e DA SILVA, E. R. **WS-GIS: Towards a SOA-Based SDI Federation**. In: DAVIS JR., C. A.; MONTEIRO, A. M. V. Advances in Geoinformatics: VIII Brazilian Symposium on GeoInformatics, GEOINFO 2006. Springer Berlin Heidelberg, 2007. p. 247-265.

LIMA, M.H.P.; RODRIGUES, C.M.; SILVA, J.K.T.; MARTINS, P.C.; TERRON, S. L. e SILVA, R.L. de S.. **DIVISÃO TERRITORIAL BRASILEIRA**. IBGE. 2002.

LINDROOS, R. *et al*. **ONKI-Paikka: An ontology service for geographical data. unpublished**. Available: http://www. seco. tkk. fi, 2007. ISSN LCCN: 0002.

LOPEZ-PELLICER, F. J. *et al*. **Publishing standard geospatial catalogues in theWeb of Data**. Actas de CAEPIA 2011. San Cristóbal de La Laguna, Tenerife: 2011.

LOPEZ-PELLICER, F. J. *et al*. **The Delft Report: Linked Data and the challenges for geographic information standardization**. Revista Catalana de Geografia IV època / volum XVII / núm. 44 / febrer 2012, fev. 2012. Disponivel em: <http://oa.upm.es/9742/>.

LOPEZ-PELLICER, F. J..; FLORCZYK, A. J.; LACASTA, J.; ZARAZAGA-SORIA, F. J. e MURO-MEDRANO, P. R. **Administrative Units, an Ontological Perspective**. In: SONG, I.-Y., *et al*. Advances in Conceptual Modeling - Challenges and Opportunities. Springer Berlin Heidelberg, 2008. p. 354-363.

LUTZ, M. *et al*. **Overcoming semantic heterogeneity in spatial data infrastructures**. Comput. Geosci., Tarrytown, NY, USA, v. 35, n. 4, p. 739-752, #apr# 2009. ISSN ISSN: 0098-3004 DOI: 10.1016/j.cageo.2007.09.017. Disponivel em: <http://dx.doi.org/10.1016/j.cageo.2007.09.017>.

MASSER, I. **Changing Notions of a Spatial Data Infrastructure**. In: VAN LOENEN, B.; BESEMER, J. W. J.; ZEVENBERGEN, J. A. SDI Convergence - Research, Emerging Trends, and Critical Assessment. Delft, the Netherlands: Netherlands Geodetic Commission, 2009. p. 219-228.

MIGUEL, J. L. **Contributions to the problem of knowledge management in Spatial Data Infrastructures** - PhD Dissertation. University of Zaragoza. p. 237. 2009.

MUSTIERE, S. *et al*. **GeOnto: Enrichissement d'une taxonomie de concepts topographiques**. Spatial Analysis and GEOmatics Sageo 2009. Paris, France: Nov, 2009.

NEBERT, D.D. **Developing Spatial Data Infrastructures: The SDI Cookbook, version 2.0**. GSDI http://gsdi. org/docs2004/Cookbook/cookbookV2. 0. pdf , 2004

NEIS, P.; ZIELSTRA, D.; ZIPF, A. **The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany** 2007- 2011. Future Internet, v. 4, n. 1, p. 1-21, 2011. ISSN ISSN: 1999-5903 DOI: 10.3390/fi4010001. Disponivel em: <http://www.mdpi.com/1999-5903/4/1/1>.

OGBUJI, U. **Real Web 2.0: Open, geographic information systems at Geonames.org**. developerWorks, 2008. Disponivel em: <http://www.ibm.com/developerworks/web/library/wa-realweb7/>. Acesso em: 31 jan. 2013.

OGC. **About OGC**, 2012. Disponivel em: <http://www.opengeospatial.org/ogc/vision>. Acesso em: 31 jan. 2013.

OGC. **GeoSPARQL - A Geographic Query Language for RDF Data**. OGC. [S.l.]. 2012a.

OGC. **OGC History**. Open Geospatial Consortium, p. http://www.opengeospatial.org/ogc/history, 2012b. Disponivel em: <http://www.opengeospatial.org/ogc/historylong>. Acesso em: 31 jan. 2013.

OGC. **Simple Feature Access - Part 1: Common Architecture**, 2006. Disponivel em: <http://www.opengeospatial.org/standards/sfa>. Acesso em: 31 jan. 2013.

OGC. **Simple Feature Access - Part 2: SQL Option**, 2006a. Disponivel em: <http://www.opengeospatial.org/standards/sfs>. Acesso em: 31 jan. 2013.

OGC, **Web Feature Service.** 2010. WFS Disponivel em: < http://www.opengeospatial.org/standards/wfs>. Acesso em: 31 jan. 2013.

OPEN GOVERNMENT PARTNERSHIP. **Open Government Declaration**, 2012. Disponivel em: <http://www.opengovpartnership.org/open-government-declaration>. Acesso em: 31 jan. 2013.

OPEN GOVERNMENT WORKING GROUP. **8 Principles of Open Government Data**. OpenGovData, 2007. Disponivel em: <http://www.opengovdata.org/home/8principles>. Acesso em: 31 jan. 2013.

OPENSTAND. **Principles. OpenStand The Modern Paradigm for Standards**, 2013. Disponivel em: <http://open-stand.org/principles/>. Acesso em: 31 jan. 2013.

PETERS, M. T.; HERON, T. E. **When the Best is not Good Enough: An Examination of Best Practice**. The Journal of Special Education, v. 26, n. 4, p. 371-385, 1993. ISSN DOI: 10.1177/002246699302600403. Disponivel em: <http://sed.sagepub.com/content/26/4/371.abstract>.

POSTEL, J.; REYNOLDS, J. **RFC 959 - File Transfer Protocol, 1985**. Disponivel em: <http://tools.ietf.org/html/rfc959>. Acesso em: 31 jan. 2013.

PRUD'HOMMEAUX, E.; SEABORNE, A. **SPARQL Query Language for RDF**. W3C. 2008. Disponível em: http://www.w3.org/TR/2006/CR-rdf-sparql-query-20060406/.

RAMAGE, S & REED, C. **OGC and ISO: Complementary Standards.** OGC, 2012. Disponível em: http://www.slideshare.net/Ramages/june-2012-ogc-and-iso

RAJABIFARD, A. & WILLIAMSON, I.P.. **Spatial data infrastructures: concept, SDI hierarchy and future directions**. Proceedings of GEOMATICS, 80, 2001

RAY, E. T. **Learning XML**. 2nd. ed.: O'Reilly Media, 2009. 418 p. ISBN: 9780596004200 Disponivel em: <http://books.google.com.br/books?id=Zilck1_0c5QC>.

REED, C. & REICHARDT, M. **From Spatial Data Infrastructure to the Geospatial Web: Advancing the SDI Vision in a Changing Technology Environment**. In Proceedings of GSDI 10. St. Augustine, Trinidad, 2008.

RICKERT, J. E.; MACHUCA, J. H. **Argentina, Geographic Information System on the Internet (PROSIGA)**. Proceedings of GSDI 10. Trinidad y Tobago: 2008.

SADEGHI-NIARAKI, A. *et al*. **Ontology-based SDI to Facilitate Spatially Enabled Society**. Proceedings of GSDI 12. Singapore: 2010.

SANTOS, D. S. M.; SOARES, I. P.; MATOS, R. S. **Seminário Sobre Web Semântica: Semantic Web Reasoning**. UNIVERSIDADE FEDERAL DE MINAS GERAIS (UFMG). Belo Horizonte. 2008.

SCHADE, S.; COX, S. **Linked data in sdi or how gml is not about trees**. Proceedings of the 13th AGILE International Conference on Geographic Information Science-Geospatial Thinking. 2010.

SCHADE, S.; GRANELL, C.; D. **Augmenting SDI with linked data**. Workshop On Linked Spatiotemporal Data, in conjunction with the 6th International Conference on Geographic Information Science (GIScience 2010). Zurich, 14th September, 2010.

SHERIDAN, J.; TENNISON, J. **Linking UK Government Data. Linked Data on the Web (LDOW2010)**. Raleigh, North Carolina: [s.n.]. 2010.

SIGNER, B. **Semantic Web - Web Information Systems**, 2012. Disponivel em: <http://www.academia.edu/2225308/Semantic_Web_-_Web_Information_Systems_WE-DINF-11912_>. Acesso em: 31 jan. 2013.

SMITH, M.K., WELTY, C. & MCGUINNESS, D.L.. **OWL Web Ontology Language Guide**, 2004. Disponível em: http://www.w3.org/TR/owl-guide/ [Acessado Fevereiro 2, 2010].

STADLER, C. *et al*. **LinkedGeoData: A core for a web of spatial open data**. Semantic Web, 2012. DOI: 10.3233/SW-2011-0052. Disponivel em: <http://dx.doi.org/10.3233/SW-2011-0052>.

STOCK, K. **Ontology-Schmology: In search of informal semantics**. Workshop on Geographic Ontologies Grounding and Vagueness. Ilkley, West Yorkshire, UK: 2008.

SWEENEY, K. **Authoritative and Open**. New Zealand Geospatial Strategy Blog, 2010 Disponivel em:< http://www.geospatial.govt.nz/authoritative-and-open-part-two . Acesso em: 31 jan. 2013.

TIAINEN, E & CARLSON, E. **Spatial semantics for geoinformatics**. Em: The Nordic GIS conference Helsinki, 2 -4 October, 2006

TIE-JUN, P. et al. Combining web services toward innovative design of agile virtual enterprise supported by web 3.0. WTOC, Stevens Point, Wisconsin, USA, v. 8, n. 1, p. 81-91, jan 2009. ISSN: 1109-2742. Disponivel em: <http://dl.acm.org/citation.cfm?id=1540566.1540575>.

TSCHIRNER, S.; SCHERP, A.; STAAB, S. **Semantic access to INSPIRE**. Terra Cognita Workshop, 2011.

TSOU, M.-H. **An Operational Metadata Framework for Searching, Indexing, and Retrieving Information Services on the Internet**, In: Second International Giscience 2002. Springer. 2002. p. 313-332.

UITERMARK, H. T. *et al*. **Ontology-based integration of topographic data sets. International Journal of Applied Earth Observation and Geoinformation**, v. 7, n. 2, p. 97-106, aug 2005.

UNICODE. **The Unicode® Standard: A Technical Introduction**, 2012. Disponivel em: < http://www.unicode.org/standard/principles.html> Acesso em: 31 jan. 2013.

USERY, E. L.; VARANKA, D. **Design and development of linked data from The National Map**. Semantic Web Volume 3, Number 4 / 2012, v. 3, p. 371-384, 2012. ISSN ISSN 1570-0844.

USGS. **Introduction to Geospatial Semantics and Technology Workshop Handbook**. U.S. Geological Survey. Reston, Virginia, p. 116. 2012.

VARANKA, D. **National Topographic Modeling, Ontology-Driven Geographic Queries**. Em: The Context of the U.S. Geological Survey's The National Map. Em Park City, Utah, 2008.

VARANKA, D. **A Topographic Feature Taxonomy for a U.S. National Topographic Mapping Ontology**. Em: Em Proceedings of ICC 2009, Santiago, Chile , 2009

VARANKA, D. **Landscape Features, Standards, and Semantics**. Em: U.S. National Topographic Mapping Databases. Cancun, Mexico, 2009a

VIEGAS, R.; SOARES, V. **Querying a Geographic Database using an Ontology-Based Methodology**. 2006.

VILCHES-BLÁZQUEZ, L. *et al*. **An Approach to Publish Spatial Data on the Web: The GeoLinked Data Case**. Workshop on Linked SpatioTemporal Data 2010 in conjunction with the 6th International Conference on Geographic Information Science. Zurich: 2010.

W3C. **Basic Geo (WGS84 lat/long) Vocabulary**. W3C Semantic Web Interest Group, 2003. Disponivel em: <http://www.w3.org/2003/01/geo/>. Acesso em: 31 jan. 2013.

W3C. **About W3C**, 2012. Disponivel em: < http://www.w3.org/Consortium/>. Acesso em: 31 jan. 2013.

WACHE, H. *et al*. **Ontology-based integration of information-a survey of existing approaches**. Em: IJCAI-01 Workshop: Ontologies and Information Sharing. 2001. p. 108-117.

WEI, M. *et al*. **A conceptual design towards semantic geospatial data access**. Em: 5th International Conference on Geographic Information Science Proceedings, Park City, Utah, 2008.

WHITE HOUSE. **Big Data is a Big Deal**. Office of Science and Technology Policy, 2012. Disponivel em: <http://www.whitehouse.gov/blog/2012/03/29/big-data-big-deal>. Acesso em: 31 jan. 2013.

WINTER, S. **Ontology: buzzword or paradigm shift in GI - science?** International Journal of Geographical Information Science, 15(7), 587, 2001.

WORBOYS, M. F.; DEEN, S. M. **Semantic heterogeneity in distributed geographic databases**. SIGMOD Rec., New York, NY, USA, v. 20, n. 4, p. 30-34, dec 1991.

WORLD ECONOMIC FORUM. **Big Data, Big Impact: New Possibilities for International Development**. WEF. p. 10. 2012. Disponível em: http://www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf.

YUAN, M. **Development of a global conceptual schema for interoperable geographic information**. INTEROP '97 International Conference on Interoperating Geographic Information Systems. Santa Barbara, CA: 1997.

ZHAO, P. *et al*. **Semantic Web-based geospatial knowledge transformation.** Computers & Geosciences, v. 35, n. 4, p. 798-808, 2009.