



**Universidade de Brasília**

Instituto de Ciências Exatas  
Departamento de Ciência da Computação

**Implementação de um geocatálogo utilizando banco  
de dados orientado a grafos para apoio à metodologia  
SERVUS**

Drausio Gomes dos Santos

Dissertação apresentada como requisito parcial para conclusão do  
Mestrado Profissional em Computação Aplicada

Orientador  
Prof. Dr. Alexandre Zaghetto

Brasília  
2016

Ficha catalográfica elaborada automaticamente,  
com os dados fornecidos pelo(a) autor(a)

SSA237 Santos, Drausio Gomes dos  
i Implementação de um geocatálogo utilizando banco  
de dados orientado a grafos para apoio à metodologia  
SERVUS / Drausio Gomes dos Santos; orientador  
Alexandre Zaghetto. -- Brasília, 2016.  
110 p.

Dissertação (Mestrado - Mestrado Profissional em  
Computação Aplicada) -- Universidade de Brasília, 2016.

1. Desenvolvimento de sistemas. 2. Geocatálogo.  
3. Infraestrutura de dados espaciais. 4. Semântica.  
I. Zaghetto, Alexandre, orient. II. Título.



# Dedicatória

Dedico este trabalho aos meus pais, Benedicto (*in memoriam*), que me ensinou a importância do saber para as nossas conquistas, e Aparecida (*in memoriam*), que me ensinou a importância do amor em família.

Dedico à minha esposa Claudia, incentivadora, companheira, parceira em todas os passos de minha vida e aos meus filhos Paulo César e Ana Claudia.

# Agradecimentos

Agradeço a Deus pela saúde necessária para que, ao completar meus 50 anos de nascimento, pudesse superar este grande desafio.

Meus agradecimentos ao meu orientador Prof. Dr. Alexandre Zaghetto pelo apoio e dedicação e, principalmente, pelas pontuais intervenções nos rumos do trabalho, proporcionando objetividade e uma melhor compreensão de um trabalho acadêmico.

Meus agradecimentos a Profa. Dra. Célia Ghedini Ralha e ao Prof. Dr. Guilherme Novaes Ramos pelas substanciais sugestões de correções na fase de qualificação.

Meus agradecimentos à Paulo César Santos, Me. Alexandre Vaz Roriz e Me. Paulo Merson, pelo auxílio nas revisões, e a Evandro Torezan, pelo apoio na camada de apresentação do produto.

Por fim, agradeço ao Tribunal de Contas da União pelo apoio institucional, em especial à Secretaria de Soluções de TI, nas figuras de Paulo André Mattos de Carvalho e Rodrigo César Santos Felisdório, anterior e atual secretários e à Secretaria de Controle Externo da Agricultura e do Meio Ambiente, na figura do atual secretário, Junnius Marques Arifa.

# Resumo

Esta pesquisa apresenta a construção de um geocatálogo que utiliza a semântica e ontologias para auxílio no processo de descoberta de recursos geográficos em serviços de uma infraestrutura de dados espaciais. Sua arquitetura incorpora um banco de dados não relacional orientado a grafos e é passível de implementação em ambiente de nuvem. Para sua validação foi utilizada a metodologia de desenvolvimento de sistemas SERVUS, especializada em arquitetura orientada a serviços geoespaciais. O estudo de caso abordou uma situação típica de auditoria governamental no tema de meio ambiente.

**Palavras-chave:** Geocatálogo, SERVUS, IDE, semântica

# Abstract

This research presents the construction of a Geo-Catalogue using semantics and ontologies to aid in the discovery process of geographic features in service of a spatial data infrastructure. Its architecture incorporates a non-relational graph database and is liable for deployment to an environment of clouds. For its validation, the methodology of development of system called SERVUS was used. It is a methodology specialized in services of geospatial oriented architecture. The study of the case discusses a typical situation of government auditing in the environmental theme.

**Keywords:** geo-catalogue, SERVUS, SDI, semantic

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Contextualização . . . . .	1
1.2	Problema de pesquisa . . . . .	4
1.3	Objetivos . . . . .	5
1.4	Metodologia de pesquisa . . . . .	6
1.5	Estrutura do trabalho . . . . .	6
<b>2</b>	<b>Revisão bibliográfica</b>	<b>8</b>
2.1	Infraestrutura de Dados Espaciais . . . . .	8
2.2	Arquitetura Orientada a Serviços (SOA) . . . . .	18
2.3	Metodologias de Análise e Projeto Orientadas a Serviço . . . . .	20
2.4	Metodologia SERVUS . . . . .	22
2.5	Geocatálogos e busca semântica . . . . .	31
2.6	Trabalhos correlatos . . . . .	39
<b>3</b>	<b>O modelo proposto</b>	<b>42</b>
3.1	Levantamento de requisitos . . . . .	42
3.2	Projeto e construção do geocatálogo . . . . .	44
<b>4</b>	<b>Validação do modelo</b>	<b>59</b>
4.1	Aplicando a metodologia SERVUS com auxílio do geocatálogo . . . . .	59
4.2	Avaliação segundo o método GQM . . . . .	68
<b>5</b>	<b>Conclusão</b>	<b>74</b>
	<b>Referências</b>	<b>76</b>
	<b>Apêndice</b>	<b>78</b>
<b>A</b>	<b>Mapa estratégico do TCU 2015 à 2021</b>	<b>79</b>

Anexo	81
I Exemplo de arquivo XSL para <i>parser</i> de XML com metadados de serviços CSW	81
II Exemplo de arquivo XML ontologia VCE	83
III Consulta para geração do <i>ranking</i>	84
IV Parâmetros de configuração de consulta para a atividade de <i>matching</i>	86
V Principais telas do geocatálogo	88
VI Pesquisa de avaliação e resultados	91

# Lista de Figuras

1.1	Diretrizes de planejamento e biênios. . . . .	2
2.1	Componentes de uma IDE. . . . .	9
2.2	Arquitetura IDE. . . . .	11
2.3	Modelo UML conceitual dos metadados básicos. . . . .	14
2.4	Princípios e objetivos de SOA. . . . .	20
2.5	Metamodelo de processos de desenvolvimento de serviços. . . . .	21
2.6	Mapeamento de requisitos para capacidades. . . . .	24
2.7	Hierarquia de modelos SERVUS. . . . .	24
2.8	Modelo conceitual SERVUS. . . . .	27
2.9	Modelo BPMN com as etapas do processo de desenvolvimento SERVUS. . . . .	28
2.10	Modelos e atividades de projeto. . . . .	30
2.11	Diagrama de atividades de <i>Discovery e Matching</i> . . . . .	31
2.12	Classificação de ontologias de Guarino. . . . .	33
2.13	Arquitetura do experimento de Lieberman. . . . .	34
2.14	Sequência de uma busca semântica. . . . .	35
2.15	Arquitetura e esquema de busca semântica. . . . .	36
2.16	Correspondência semântica. . . . .	37
2.17	Arquitetura de implementação SERVUS. . . . .	38
2.18	Processo de descoberta de informação. . . . .	40
3.1	Planejamento pelo método GQM( <i>Goal, Question, Metric</i> ). . . . .	45
3.2	Arquitetura de implementação. . . . .	47
3.3	Uma expressão básica na linguagem Cypher. . . . .	47
3.4	Modelo de dados para Recursos. . . . .	49
3.5	Rede de Recursos/Recursos semânticos. . . . .	50
3.6	Modelo de dados para Ontologias. . . . .	51
3.7	Processo de <i>harvesting</i> . . . . .	52
3.8	Estrutura de busca do catálogo. . . . .	54
3.9	Modelo de relacionamento Recurso requisitado X Recurso ofertado. . . . .	55

4.1	Modelo de domínio. . . . .	60
4.2	Metamodelo de Informação. . . . .	61
4.3	Modelo de Informação. . . . .	62
4.4	<i>Matching</i> de recursos utilizando o geocatálogo. . . . .	63
4.5	Modelo de projeto - Recurso RR_001. . . . .	64
4.6	Modelo de projeto - Recurso RR_002. . . . .	65
4.7	Visualização das camadas pela ferramenta QGIS. . . . .	66
4.8	Visualização de áreas discordantes. . . . .	67
4.9	Gráfico resultado da pesquisa. . . . .	73
V.1	Tela com a classificação por ontologia. . . . .	88
V.2	Tela com o filtro espacial. . . . .	89
V.3	Tela com a configuração de pesos. . . . .	89
V.4	Tela com a listagem do ranking. . . . .	90
VI.1	Parâmetros utilizados na consulta. . . . .	99
VI.2	Resultado da consulta. . . . .	99

# Lista de Tabelas

2.1	Parâmetros da operação <i>getCapabilities</i> . . . . .	13
2.2	Perfis WFS. . . . .	16
2.3	Interpretação de RM-ODP para RM-OA de Usländer. . . . .	26
2.4	Processo de desenvolvimento: etapas, objetivos e artefatos. . . . .	29
3.1	Tabela de pesos. . . . .	56
3.2	Cálculo do índice de relevância. . . . .	56
4.1	Extensão do caso uso. . . . .	61
4.2	Modelos da metodologia SERVUS utilizados. . . . .	68
4.3	Atividades do processo da metodologia SERVUS utilizados. . . . .	69
4.4	Valores mínimos e máximos para a faixas atribuídas para pesquisa e índice de relevância. . . . .	71
4.5	Dados consolidados da pesquisa. . . . .	72
4.6	Cenários e tempos de resposta do geocatálogo. . . . .	73
IV.1	Configuração de pesos para consulta. . . . .	86
IV.2	Configuração das caixas de seleção de filtragem. . . . .	87

# Lista de Abreviaturas e Siglas

<b>BEA-ALUI</b>	<i>BEA AquaLogic User Interface.</i>
<b>CIM</b>	<i>Computation Independent Model.</i>
<b>CONCAR</b>	Comissão Nacional de Cartografia.
<b>CPRM</b>	Companhia de Pesquisa de Recursos Minerais.
<b>CSV</b>	<i>Comma-Separated Values.</i>
<b>CSW</b>	<i>Catalogue Service Web.</i>
<b>DAMLDB</b>	<i>DARPA Agent Markup Language Database.</i>
<b>DATAGEO-SP</b>	Projeto de Dados Geográficos da Secretaria do Meio Ambiente de SP.
<b>DC</b>	<i>Dublin Core.</i>
<b>EEA</b>	Agência Européia do Ambiente.
<b>EFS</b>	Entidades Fiscalizadoras Superiores.
<b>EIS</b>	<i>Environmental Information System.</i>
<b>ET-EDGV</b>	Especificação Técnica para Aquisição de Dados Geospaciais Vetoriais.
<b>GML</b>	<i>Geography Markup Language.</i>
<b>GQM</b>	<i>Goal Question Metric.</i>
<b>GSW</b>	<i>Geospatial Semantic Web.</i>
<b>IBAMA</b>	Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis.
<b>IBGE</b>	Instituto Brasileiro de Geografia e Estatística.
<b>IDE</b>	Infraestrutura de Dados Espacial.
<b>IDEs</b>	Infraestruturas de Dados Espaciais.
<b>INDE</b>	Infraestrutura Nacional de Dados Espaciais.
<b>INDE-BR</b>	Infraestrutura Nacional de Dados Espaciais do Brasil.

<b>JSON</b>	<i>JavaScript Object Notation.</i>
<b>MDA</b>	<i>Model-Driven Architecture.</i>
<b>MDD</b>	<i>Model-Driven Development.</i>
<b>MOF</b>	<i>Meta-Object Facility.</i>
<b>OGC</b>	<i>Open Geospatial Consortium.</i>
<b>OLACEFS</b>	Organização Latinoamericana e do Caribe de Entidades Fiscalizadoras Superiores.
<b>OMT-G</b>	<i>Object Modeling Technique for Geographic Applications.</i>
<b>OWL</b>	<i>Web Ontology Language.</i>
<b>OWL-S</b>	<i>OWL Web Ontology Language for Services.</i>
<b>PIM</b>	<i>Platform Independent Model.</i>
<b>PSM</b>	<i>Platform Specific Model.</i>
<b>RDF</b>	<i>Resource Description Framework.</i>
<b>RDFS</b>	<i>Resource Description Framework Schema.</i>
<b>REST</b>	<i>Representational State Transfer.</i>
<b>RM-OA</b>	<i>Reference Model for de ORCHESTRA Architecture.</i>
<b>RM-ODP</b>	<i>Reference Model of Open Distributed Processing.</i>
<b>S-Match</b>	<i>Semantic Matching.</i>
<b>SAIL</b>	<i>Storage And Inference Layer.</i>
<b>SERVUS</b>	<i>Service and Use Case.</i>
<b>SIG</b>	Sistema de Informação Geográfica.
<b>SOA</b>	<i>Service Oriented Architecture.</i>
<b>SSDI</b>	<i>Semantic Spatial Data Infrastructure.</i>
<b>SWEET</b>	<i>Semantic Web for Earth and Environmental Terminology.</i>
<b>TCU</b>	Tribunal de Contas da União.
<b>TI</b>	<i>Tecnologia da Informação.</i>
<b>UDDI</b>	<i>Universal Description, Discovery and Integration.</i>

**UML** *Unified Modeling Language.*

**URL** *Universal Resource Locator.*

**VCE** Vocabulário do Controle Externo.

**VCGE** Vocabulário Controlado do Governo Eletrônico.

**WCS** *Web Coverage Service.*

**WFS** *Web Feature Service.*

**WMS** *Web Map Service.*

**XML** *eXtensible Markup Language.*

**XSL** *eXtensible Stylesheet Language.*

**XSLT** *eXtensible Stylesheet Language Transformations.*

# Capítulo 1

## Introdução

### 1.1 Contextualização

O Tribunal de Contas da União (TCU), também conhecido como Corte de Contas, é um tribunal administrativo que exerce o papel de entidade fiscalizadora superior, cuja principal função consiste em auxiliar o Congresso Nacional na atividade de controle externo, conforme as disposições amparadas pela Constituição Federal Art.71<sup>1</sup>.

Entende-se por controle externo a ação de “exercer a fiscalização contábil, financeira, orçamentária, operacional e patrimonial da União e das entidades da administração direta e indireta, quanto à legalidade, à legitimidade e à economicidade, assim como exercer a fiscalização da aplicação das subvenções e da renúncia de receitas”<sup>2</sup>. Para exercer suas funções o TCU dispõe de um corpo técnico de auditores que realizam auditorias, levantamentos, inspeções, acompanhamentos e monitoramentos, a fim de subsidiar as decisões de um colegiado de ministros<sup>3</sup>.

O TCU, como órgão responsável pelo controle externo da administração pública, necessita de dados e informações referentes à boa utilização dos recursos públicos e à efetividade das ações governamentais. As alocações de recursos podem ocorrer de forma regionalizada ou distribuída pelo território nacional, neste sentido, a informação espacial ou geográfica é um importante fator a ser considerado na avaliação do resultado efetivo da aplicação do dinheiro público.

Torna-se um desafio para os órgãos de controle obterem esta visão espacial do resultado das ações governamentais de forma simplificada e atualizada, uma vez que os dados são descentralizados, produzidos por fontes variadas e em momentos diferentes. Acrescido a

---

<sup>1</sup>CF-1988, [http://www.planalto.gov.br/ccivil\\_03/Constituicao/Constituicao.htm](http://www.planalto.gov.br/ccivil_03/Constituicao/Constituicao.htm)

<sup>2</sup>Lei orgânica do TCU, [http://www.planalto.gov.br/ccivil\\_03/Leis/L8443.htm](http://www.planalto.gov.br/ccivil_03/Leis/L8443.htm)

<sup>3</sup><http://portal.tcu.gov.br/institucional/conheca-o-tcu/funcionamento/>

isto, tem-se observado um aumento da disponibilização de informações na Internet sob a forma de dados abertos, associado à adoção de padrões mínimos de interoperabilidade.

Em 2015 o TCU iniciou um novo ciclo de planejamento estratégico para o horizonte de seis anos. O Apêndice apresenta o mapa estratégico, do qual destacamos três objetivos estratégicos relevantes para este trabalho: “aprimorar o uso da tecnologia da informação como instrumento de inovação para o controle”, “realizar diagnósticos sistêmicos em áreas relevantes” e “induzir a disponibilidade e a confiabilidade de informações da Administração Pública”.

A Figura 1.1 apresenta alguns dos direcionamentos dados pelas administrações do TCU nos últimos quatro biênios. Destacamos as diretrizes de adoção do Governo Eletrônico (*e-Government*), dados abertos, análise de grandes massas de dados (*big data*), auditorias preditivas e busca da efetividade no controle da administração pública.

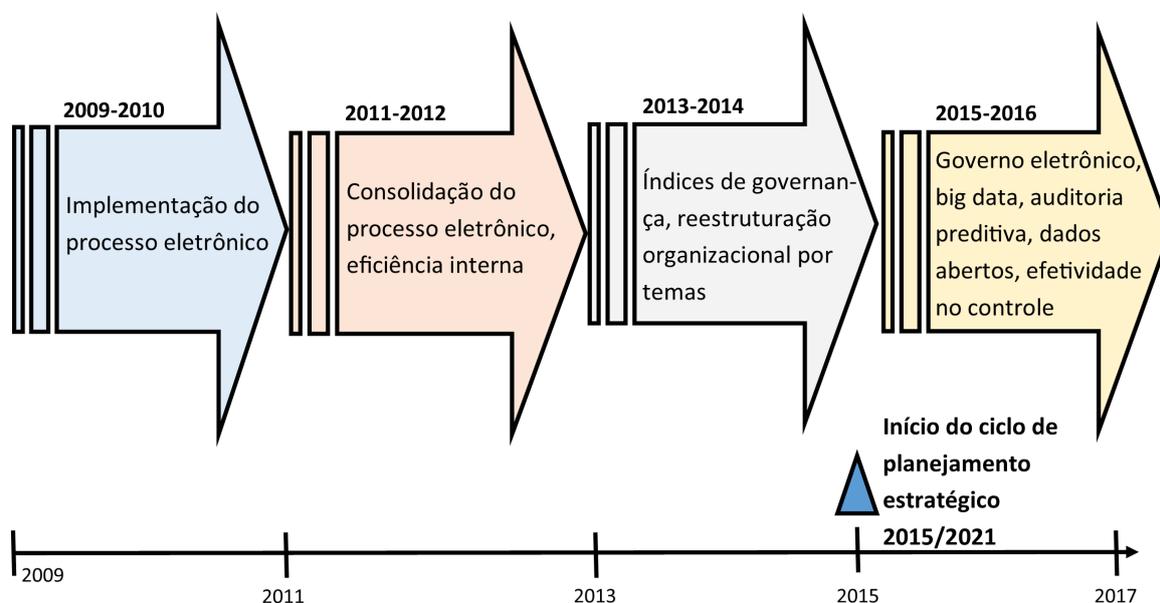


Figura 1.1: Diretrizes de planejamento e biênios.

O decreto N° 6.666, de 27 de novembro de 2008, instituiu a Infraestrutura Nacional de Dados Espaciais do Brasil (INDE-BR), que consiste numa iniciativa de criação de uma arquitetura para disseminação e regramento do uso de dados geográficos no âmbito da administração pública. Cabe ressaltar um dos objetivos definidos no decreto, o de “evitar a duplicidade de ações e o desperdício de recursos na obtenção de dados geoespaciais pelos órgãos da administração pública, por meio da divulgação dos metadados relativos a esses dados disponíveis nas entidades e nos órgãos públicos das esferas federal, estadual,

distrital e municipal”. Para concretizar esta iniciativa governamental foram criados o Diretório Brasileiro de Dados Geoespaciais (DBDG) e o Sistema de Informações Geográficas do Brasil (SIG Brasil), além do estabelecimento, em 2010, de um Plano de Ação<sup>4</sup> para implantação.

Camboim em [8] ressalta a iniciativa governamental de criação da Infraestrutura Nacional de Dados Abertos (INDA) e do Portal Brasileiro de Dados Abertos como estratégia de adoção de Dados Interligados Abertos (*Linked Open Data*). Em seu trabalho, Camboim propõe uma arquitetura capaz de tornar os dados da INDE-BR disponíveis na INDA de forma integrada. Destaca-se neste trabalho a utilização, em sua proposta de arquitetura, de camadas semânticas e de ontologias, fazendo uma ligação entre as técnicas da web semântica e os dados geoespaciais, ingressando no campo da geossemântica.

As arquiteturas das IDE atuais são, via de regra, orientadas a serviço e adotam os padrões tecnológicos estabelecidos pela *Open Geospatial Consortium* (OGC). Os serviços geográficos-espaciais, também conhecidos como geosserviços, permitem a obtenção de mapas cartográficos, mapa temáticos, dados geográficos e metadados.

Um dos componentes essenciais numa arquitetura de IDE é o catálogo de serviços. Alguns órgãos de governo tiveram iniciativas de implementar soluções de geocatálogos, no intuito de facilitar a localização de recursos por temas ou áreas geográficas específicas. Exemplos no Brasil são o Geocatálogo do Ministério do Meio Ambiente<sup>5</sup> e o Geocatálogo SEI<sup>6</sup>, ligado a Superintendência de Estudos Econômicos e Sociais da Bahia. Um exemplo de geocatálogo semântico<sup>7</sup> é o da província de Trento na Itália.

O desenvolvimento de sistemas em arquiteturas orientadas a serviços para IDE possui algumas peculiaridades que o distingue do desenvolvimento de uma arquitetura SOA padrão. Essa diferenciação é oriunda principalmente de requisitos informacionais para um grupo bem diferenciados de dados, por exemplo, mapas, dados geográficos em formato XML e GML, e por uma maior necessidade de reaproveitamento de serviços e adoção mais rígida de padrões.

A metodologia SERVUS de desenvolvimento de sistemas, proposta por Usländer em [35], feita especialmente para tratar do uso e composição de serviços geoespaciais e utilizada na construção de serviços na infraestrutura de dados espaciais da comunidade europeia, se propõe a lidar com estas peculiaridades de sistemas geográficos.

---

<sup>4</sup>[http://www.concar.gov.br/plano\\_acao.aspx](http://www.concar.gov.br/plano_acao.aspx)

<sup>5</sup><http://geocatalogo.mma.gov.br/>

<sup>6</sup><http://geocatalogo.sei.ba.gov.br/site/>

<sup>7</sup><http://www.territorio.provincia.tn.it/portal/server.pt>

## 1.2 Problema de pesquisa

O TCU para exercer a sua atribuição de auditoria externa demanda informações de variados tipos: contábeis, demográficas e, dentre outras, as espaciais ou geográficas. Acrescido a esta variedade de tipos, temos um grande número de fontes e uma heterogeneidade de modelos de informação.

O governo em sua estratégia de dados abertos tem disponibilizado o acesso a arquiteturas que centralizam estas informações em plataformas abertas, com o intuito de diminuir os custos de produção e de distribuição e aumentar a transparência de suas ações. A INDE-BR é uma destas arquiteturas e constitui-se num grande repositório aberto de dados e informações geográficas.

O TCU carece de ferramentas e metodologia para acessar estes recursos de forma integrada aos seus sistemas e aos processos de negócio. Um primeiro passo para a utilização seria conseguir localizar os recursos de forma rápida e eficiente.

Nesse contexto uma ferramenta de geocatálogo para descoberta de recursos, apoiada por uma metodologia que auxilie na construção de serviços e no consumo das informações geográficas, pode ser elemento importante no atingimento dos objetivos estratégicos do TCU.

Podemos resumir o problema de pesquisa na seguinte questão: Como possibilitar ao corpo técnico do TCU o acesso às informações geográficas disponibilizadas pelos geosserviços de infraestruturas de dados espaciais da Internet, de forma a inseri-los aos seus processos de trabalho e integrá-los a sua infraestrutura tecnológica atual ?

### Justificativas

Seguem abaixo algumas justificativas para a realização do trabalho:

1. Com a consolidação da INDE-BR<sup>8</sup>, a ampliação dos dados disponíveis em seu diretório e o conseqüente aumento dos geosserviços disponibilizados, há uma oportunidade para que o TCU use esses dados e informações geográficas para uma melhor atuação no controle externo;
2. Os produtos atuais utilizados para implementar uma infraestrutura de dados espaciais adotam padrões básicos para disponibilizar seus dados e metadados, com destaque para os padrões *Open Geospatial Consortium* (OGC)<sup>9</sup>, facilitando a interconectividade entre sistemas;

---

<sup>8</sup>Histórico da INDE, <http://www.inde.gov.br/a-inde/historico>

<sup>9</sup>Padrões OGC, <http://www.opengeospatial.org/standards>

3. Alguns órgãos públicos, como Ministério do Meio Ambiente, Agência Nacional de Águas, Fundação Nacional do Índio, Ministério do Desenvolvimento Agrário, Instituto de Pesquisa Econômica Aplicada, Instituto Brasileiro de Geografia e Estatística, já possuem arquitetura institucional para disponibilização de dados e informações geográficas<sup>10</sup>, portanto haveria a oportunidade de obter diretamente destas infraestruturas insumos necessários para as auditorias e informações referentes aos resultados destas instituições;
4. O TCU está evoluindo a sua arquitetura de sistemas em busca de possibilitar o consumo e a produção de serviços no estilo arquitetural SOA;

### 1.3 Objetivos

O objetivo geral deste trabalho é a especificação e implantação de um geocatálogo, com buscas baseadas em ontologias, como ferramenta de auxílio ao trabalho de auditoria para a localização de recursos e informações geográficas em uma infraestrutura de dados espaciais.

Para consecução do objetivo geral foram definidos os seguintes objetivos específicos:

- Especificar e implementar, no nível de protótipo, um geocatálogo baseado na proposta de ambiente de projeto da metodologia SERVUS.
- Avaliar por meio de estudo de caso a sua utilização num cenário de auditoria na área de meio ambiente e propor a sua utilização no TCU.
- Apresentar o resultado do estudo da metodologia de desenvolvimento de sistemas SERVUS.
- Apresentar o resultado do estudo de Infraestruturas de Dados Espaciais (IDEs) e os protocolos de implementação de geosserviços.
- Apresentar o resultado dos estudos de implementações de catálogos semânticos e da utilização da semântica e ontologias no processo de descoberta de geosserviços.
- Modelagem de um banco de dados orientado a grafos para utilização da metodologia SERVUS.
- Elaborar e implementar um procedimento de coleta de metadados de geosserviços e de carga em banco de dados não relacional orientado a grafos.

---

<sup>10</sup>Catálogo de Metadados INDE, <http://www.metadados.inde.gov.br/geonetwork/srv/por/main.home>

## 1.4 Metodologia de pesquisa

O trabalho foi desenvolvido conforme a seguinte metodologia:

- **Revisão de literatura:** realização de estudos com base em artigos científicos, teses de doutorado, dissertações de mestrado e livros editados, nos seguintes assuntos: Infraestrutura de Dados Espaciais (IDEs), Metodologia de desenvolvimento de sistemas orientados a serviço, Metodologia de desenvolvimento SERVUS, Banco de dados orientado a grafos, Geocatálogos e Busca semântica.
- **Levantamento de requisitos:** definição dos requisitos baseado nos geocatálogos estudados, principalmente o proposto pela metodologia SERVUS, nas necessidades das equipes de auditoria da área de meio ambiente e na infraestrutura tecnológica atual do órgão.
- **Definição da arquitetura:** elaboração de uma arquitetura de alto nível para atendimento dos requisitos considerando-se, dentre outros aspectos: fontes de dados, arquiteturas de IDEs e geosserviços da INDE-BR.
- **Desenvolvimento do protótipo:** definição das linguagens de desenvolvimentos, componentes, bibliotecas de software e ambiente de desenvolvimento .
- **Implantação:** definição do ambiente e instalação do protótipo.
- **Elaboração de um cenário de utilização:** elaboração de um cenário de utilização do geocatálogo em conjunto com a metodologia de desenvolvimento SERVUS, levando-se em consideração um cenário típico de auditoria em meio ambiente.
- **Validação do protótipo e avaliação do resultados:** validação do protótipo, avaliação dos resultados e das perspectivas de utilização da metodologia.

## 1.5 Estrutura do trabalho

Além desta introdução, este documento está organizado nos demais capítulos abaixo descritos.

O Capítulo 2 apresenta a fundamentação teórica dos temas: Infraestrutura de Dados Espacial (IDE); Arquitetura Orientada a Serviço; Metodologia de Desenvolvimento SERVUS; Geocatálogos e Busca semântica.

Quanto a IDE são tratados os conceitos, a composição, as funcionalidades e a arquitetura básica. A fim de subsidiar os temas de recuperação de metadados e descoberta de geosserviços na rede, será dado destaque ao componente tecnológico, mais especificamente aos padrões de implementação de geosserviços.

A fundamentação teórica relacionada a Arquitetura Orientada a Serviço tem foco nos princípios e objetivos. Trata também da especialização de SOA para informações geoespaciais e de metodologias de análise e projeto orientadas a serviço. É apresentada a metodologia escolhida para o nosso estudo de caso: SERVUS. São abordados o processo de desenvolvimento, as etapas, as principais atividades, modelos e artefatos.

Ainda no Capítulo 2 tratamos da semântica e de ontologias como elementos que contribuem para uma melhor interoperabilidade e descoberta de recursos em IDEs. São apresentados experimentos e catálogos semânticos, dentre eles o proposto pela metodologia SERVUS, o qual serviu de base para construção do protótipo fruto deste trabalho de pesquisa.

O Capítulo 3 apresenta o modelo proposto, incluindo os requisitos, os detalhes do projeto e da implementação de um geocatálogo baseado no ambiente de projeto da metodologia SERVUS, que utiliza um banco de dados orientado a grafos e se apoia em ontologias para o processo de busca.

O Capítulo 4 apresenta um estudo de caso de aplicação da metodologia SERVUS, apoiada pelo geocatálogo e utilizando um cenário típico de auditorias realizadas no âmbito das unidades técnicas do TCU que lidam com o tema meio ambiente e agricultura. Ao final são apresentados os resultados do trabalho científico segundo a técnica *Goal Question Metric* (GQM).

O Capítulo 5 apresenta as conclusões e os trabalhos futuros.

# Capítulo 2

## Revisão bibliográfica

Este capítulo apresenta os conceitos básicos associados a Infraestrutura de Dados Espaciais, os componentes de sua arquitetura tecnológica e os padrões para disponibilização de geosserviços, destacando a utilização prática de suas interfaces e operações.

É apresentada, também, uma breve revisão bibliográfica de SOA e de metodologias de desenvolvimento orientadas a serviço com o intuito de apresentar a metodologia SERVUS, os conceitos associados, seus modelos, seu processo de desenvolvimento e artefatos.

É feita uma revisão de literatura no tema geocatálogos e busca semântica por meio da apresentação de experimentos no campo da geosseântica e exemplos de geocatálogos. Por fim, avaliamos alguns trabalhos correlatos a este trabalho de pesquisa.

### 2.1 Infraestrutura de Dados Espaciais

Poveda e Vazquez em [3] conceituam IDE como uma infraestrutura necessária para o acesso, compartilhamento, troca, combinação e análise dos dados geográficas, de forma padrão e interoperável. Também consideram a necessidade de que estes dados estejam disponíveis em rede, por meio de um conjunto de sistemas que utilizam protocolos e interfaces padrões, propiciando a criação de aplicações que possam ser vistas pelo usuário como um único sistema.

De uma maneira simplificada apresentam uma IDE como um “SIG aberto implementado sobre a rede com tudo o que isso implica: componentes distribuídos, interfaces padrões, interoperabilidade, coordenação, acesso aos dados, capacidade de análise como objetivo, etc.” Apresentam a IDE como um elemento importante para a democratização das informações geográficas e ampliam a definição para além do tecnológico.

A principal distinção entre o conceito de Sistema de Informação Geográfica (SIG) e Infraestrutura de Dados Espacial (IDE) é que o segundo tem foco no compartilhamento de informações geográficas, estendendo sua atuação para além das fronteiras da organização.

Infraestrutura de Dados Espaciais pode ser classificada como uma estrutura de componentes tecnológicos, geográficos, sociais e políticos [3]. A Figura 2.1 ilustra estes componentes.

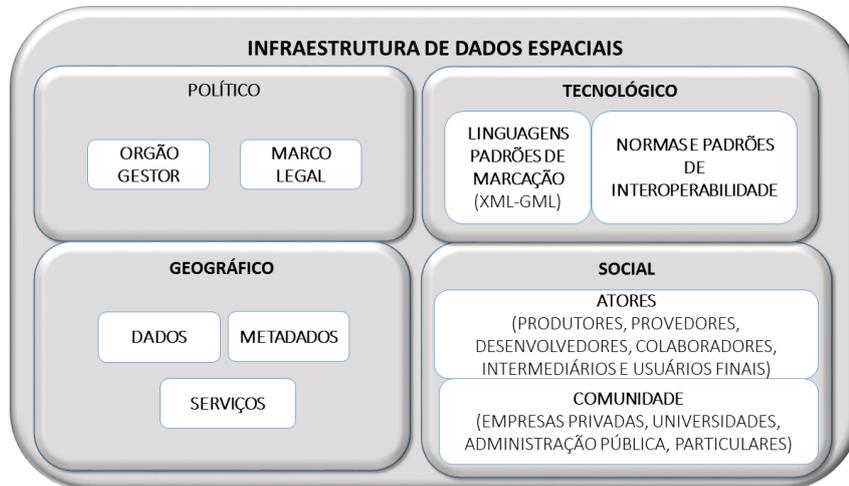


Figura 2.1: Componentes de uma IDE.

O componente tecnológico é representado por uma arquitetura baseada em padrões de interoperabilidade e capaz de compartilhar dados e informações geográficas. As linguagens de marcação XML e GML têm um papel fundamental neste componente. A definição de Davis e outros em [21] nos auxilia na compreensão do papel da interoperabilidade e das linguagens de marcação:

**IDE:** “Uma IDE pode ser entendida como a confluência entre diversos (em potencial) provedores de dados geográficos, cada qual fornecendo acesso a dados através de serviços Web específicos, aplicações cujas interfaces e conexões são expressas em XML e podem ser encontrados através de mensagens em XML”.

O componente social é representado por um conjunto de atores, dentre eles, produtores de dados, provedores de serviços, usuários, desenvolvedores de software e responsáveis por padrões e normas, além de uma grande comunidade formada por empresas privadas, governo, universidades e a sociedade em geral.

O componente geográfico é representado pelos dados, seus metadados e os geosserviços.

O componente político é representado pelas pessoas e órgão responsáveis por estabelecer o marco regulatório e as regras de funcionamento. O Brasil estabeleceu o seu marco legal pela edição do decreto 6.666/2008<sup>1</sup> que instituiu, no âmbito do poder executivo federal, a Infraestrutura Nacional de Dados Espaciais do Brasil (INDE-BR). Complemen-

<sup>1</sup>[http://www.inde.gov.br/images/inde/20@Decreto6666\char'\\_27112008.pdf](http://www.inde.gov.br/images/inde/20@Decreto6666\char'_27112008.pdf)

tando o componente político temos o Comissão Nacional de Cartografia (CONCAR)<sup>2</sup>, como principal órgão gestor e regulador da INDE.

Um marco importante para concretização da INDE no Brasil foi o estabelecimento de um plano de ação<sup>3</sup>.

O conceito de INDE expresso no texto do plano de ação ressalta a integração entre as esferas administrativas federais, estaduais e municipais:

**Infraestrutura Nacional de Dados Espaciais (INDE):** “conjunto integrado de tecnologias; políticas; mecanismos e procedimentos de coordenação e monitoramento; padrões e acordos, necessário para facilitar e ordenar a geração, o armazenamento, o acesso, o compartilhamento, a disseminação e o uso dos dados geoespaciais de origem federal, estadual, distrital e municipal” [11].

Segue abaixo um conjunto amplo de funcionalidades que podem compor uma IDE, segundo Poveda e Vasquez [3]:

1. Permitir a busca de informação geográfica e metadados que indiquem a zona geográfica, o formato, ano em que se produziu, a forma de acesso, a qualidade das informações, o responsável pela informação e outros considerados relevantes.
2. Permitir a visualização de mapas, ortofotos<sup>4</sup>, mapas de terreno e dados de geográficos de diferentes instituições, com sistemas de referência diferentes, com formatos distintos, propriedades heterogêneas e e com a possibilidade de sobreposição de camadas.
3. Permitir a busca de uma entidade geográfica pelo seu nome e posicioná-la em um mapa.
4. Acessar as entidades geográficas, seus atributos, coordenadas, topologia e geometria de uma forma padrão.
5. Realizar operações de análise básicas, tais como roteamento, cálculo de áreas e análise de superfície.
6. Realizar a transformação de um modelo de dados para outro, caso ambos estejam padronizados.
7. Permitir a baixa de dados para análise em um SIG, caso o serviço oferecido pela IDE não atenda às necessidades do usuário.

---

<sup>2</sup><http://www.concar.gov.br/>

<sup>3</sup><http://www.governoeletronico.gov.br/anexos/apresentacao-plano-de-acao-inde/>

<sup>4</sup>Ortofotos são fotografias aéreas corrigidas para um formato ortogonal que permitem a sua utilização como mapas, ver <http://mundogeo.com/blog/2000/12/01/ortofoto-a-imagem-que-e-um-mapa/>

As IDEs têm um importante papel na eliminação dos obstáculos que dificultam o acesso às informações geográficas, geram perda de tempo e tornam difícil e custosa a reutilização dos dados geográficos.

### 2.1.1 Arquiteturas de IDEs

As arquiteturas atuais são predominantemente arquiteturas de serviço, Davis e Alves em [21] apresentam um modelo de arquitetura IDE, ilustrada na Figura 2.2. Um elemento importante nesta arquitetura é o geoportal. Segundo os autores, “a funcionalidade esperada de um geoportal inclui (1) a descoberta de fontes de informação e conteúdo, e (2) acesso on-line a aplicações”. Destacam a incompatibilidade entre alguns padrões adotados em IDEs e padrões da WEB, como por exemplo, os utilizados nos catálogos de serviço.

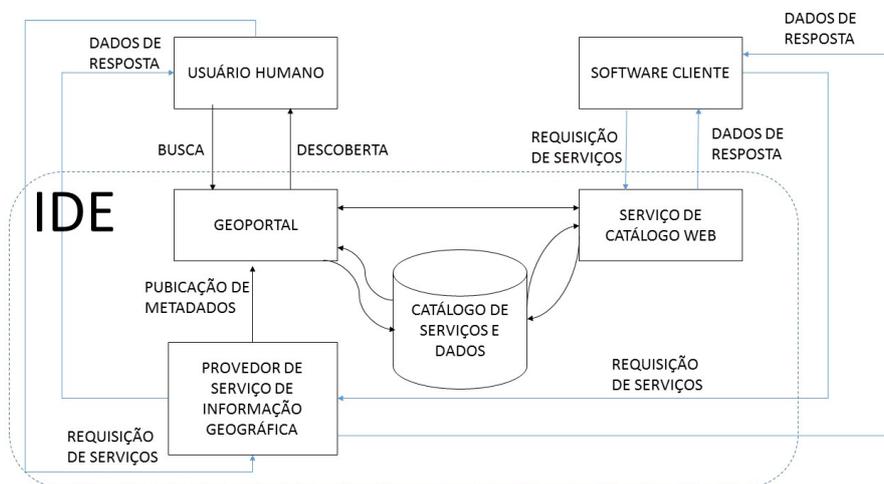


Figura 2.2: Arquitetura IDE.  
Fonte: adaptado de [21].

### 2.1.2 Padrões para geosserviços

A *Open Geospatial Consortium* (OGC) é o principal órgão padronizador para IDEs. A OGC é um consórcio internacional industrial de companhias, agências governamentais e universidades que participam do desenvolvimento e disponibilização pública de padrões de interface [29].

Segundo Davis e Alves[21] os principais padrões OGC utilizados em arquiteturas de IDEs são:

1. *Web Feature Service* (WFS): “provê uma interface para inserção, seleção, atualização e remoção de feições geográficas (objetos)”.

2. *Web Coverage Service* (WCS): “provê acesso a geocampos, da mesma maneira como no *Web Feature Service*. Entretanto, este serviço não retorna imagens dos geocampos, mas sim detalhes semânticos sobre os mesmos”.
3. *Web Gazetteer Service*: “estende o *Web Feature Service* com recursos para a implementação de interfaces para gazetteers ”.
4. *Web Registry Service* e *OpenGIS Catalog Service* (OCS ou CSW): “implementam uma funcionalidade operacional similar ao UDDI”.
5. *Web Coordinate Transformation Service* (WCTS): “provê algoritmos que convertem coordenadas de objetos espaciais entre diferentes sistemas de referência espacial”.
6. *Web Map Service* (WMS): “um serviço para produção de mapas on-line, para que sejam visualizados diretamente na Web ou em aplicativos gráficos genéricos. Nesse serviço, os mapas são renderizações (apresentações) da realidade, e não incluem, portanto, o dado geográfico atual, a partir do qual o mapa foi criado”.
7. *Web Terrain Service* (WTS): “similar ao *Web Map Service*, mas destinado à visualização tridimensional de superfícies. Ambos podem produzir apresentações em formatos de imagem ou no formato SVG (*Scalable Vector Graphics*), que é vetorial”.

### 2.1.2.1 Uma interface comum de acesso a geosserviços

O *OGC Web Services Common Standard* é um padrão que propõe uma interface comum para um conjunto de padrões OGC de geosserviços: WMS, WFS e WCS. A iniciativa busca padronizar as formas de acesso, operações, parâmetros obrigatórios e opcionais, além de estruturas de dados, a fim de diminuir os esforços de interoperabilidade. Um dos tópicos tratados na especificação é a forma de recuperação das capacidades dos geosserviços, ou melhor, dos metadados acerca dessas capacidades, para esse fim, é instituída a operação obrigatória *GetCapabilities*. Esta operação possui um conjunto de parâmetros que estão detalhados na Tabela 2.1.

Segue abaixo um exemplo de URL com parâmetros no formato chave/valor para utilização da operação *GetCapabilities*.

```
http://hostname:port/path?service=WCS&request=GetCapabilities&
acceptVersions=1.0.00.8.3&sections=Contents&updateSequence=XYZ123&
acceptFormats=text/xml&acceptLanguages=en-CA fr-CA
```

A resposta para a operação *getCapabilities* consiste em um documento com os metadados ou, em caso de erro, com a descrição e códigos de erro. Os possíveis erros são: parâmetro perdido, parâmetro inválido, falha na negociação da versão, valor inválido do parâmetro sequencial de atualização.

Tabela 2.1: Parâmetros da operação *getCapabilities*.

Nome Parâmetro	Definição	Tipo de dado e valor	Multiplicidade e uso
<i>service</i>	Identificador do tipo de serviço	String . Ex. WMS, WCS, CSW	Um - obrigatório
<i>request</i>	Nome da operação	String. Ex. <i>GetCapabilities</i>	Um - obrigatório
<i>acceptVersions</i>	Versões aceitas pelo cliente ordenadas segundo a preferência	Lista de String no formato x.y.z . Ex. 1.2.5	Zero ou um - opcional
<i>section</i>	Lista não ordenada das seções a serem apresentadas no documento retornado pelo serviço	Lista de String	Zero ou um - opcional
<i>updateSequence</i>	Número sequencial representando a versão do documento, o valor é incrementado pelo serviço toda vez que o documento de metadados sofre alterações	String	Zero ou um - opcional
<i>acceptFormats</i>	Sequência priorizada de formatos aceitos pelo cliente	String. Seguem o padrão MIME de tipos. Ex text/xml	Zero ou um - opcional
<i>acceptLanguages</i>	Lista de linguagens aceitas pelo cliente	String. Segue o padrão IETF RFC 4646 <sup>5</sup> . Ex. pt-BR	Zero ou um - opcional

Fonte: Adaptado de [38].

O documento retornado possui os campos da versão do documento (obrigatório) e da sequência de atualização (opcional), além de outro conjunto de campos organizados em sessões. As sessões são: identificação do serviço, identificação do provedor, operações disponíveis, conteúdos e linguagens.

A sessão que apresenta os conteúdos engloba as informações referentes aos tipos de dados que são fornecidos pelo serviço. O modelo UML conceitual simplificado da Figura 2.3 ilustra a organização dos campos de metadados.

As demais operações desta interface de serviço devem adotar os parâmetros *service*, *version*, *request* como obrigatórios e o *acceptLanguages* como opcional.

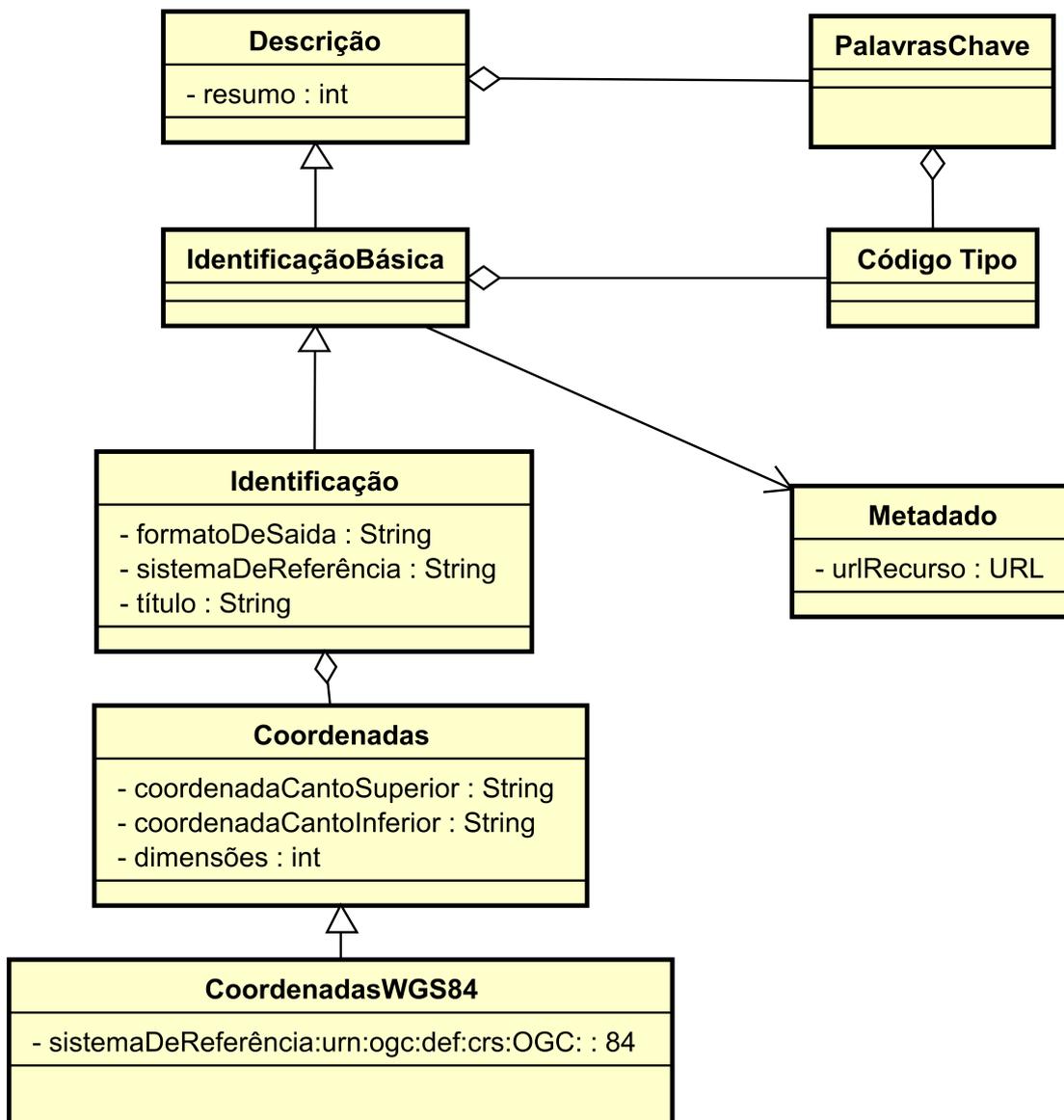


Figura 2.3: Modelo UML conceitual dos metadados básicos.

### 2.1.2.2 O serviço de mapas - WMS

Consiste num serviço que produz mapas dinâmicos, em que as informações geográficas são organizadas em camadas (*layers*). A interface do serviço oferece operações para recuperação dos metadados, recuperação do mapa em formato de imagem e recuperação de informações de camadas específicas [14].

A operação de recuperação de metadados é a *getCapabilities*, já elucidada no tópico anterior. Uma das informações adicionais retornadas pelo método *getCapabilities*, específica deste serviço, são os metadados das *layers*.

A operação de recuperação de mapas é a *getMap*. Os parâmetros de requisição são as coordenadas, largura, altura e formato da imagem, o sistema de referência das coordenadas, as camadas de informação que se deseja exibir, os estilos de formatação de cada camada, se transparente ou opaca, a cor de fundo, dentre outros [3].

Segue abaixo um exemplo de URL com parâmetros no formato chave/valor para utilização de *GetMap*.

```
http://hostname:port/path?service=WMS&request=GetMap&
version=1.3.0&crs=EPSG:25830&
bbox=420353.19,4468089.67,469858.99,4494819.23&
width=729&height=345&layers=agricultura&
style=padrao&format=image/png&bgcolor=0xFFFFFFFF&
transparent=false
```

O serviço pode implementar uma operação denominada *getFeatureInfo* que recupera metadados referentes as camadas marcadas como pesquisáveis. Esta operação permite ao usuário obter informações específicas de uma área do mapa.

Segue abaixo um exemplo de URL com parâmetros no formato chave/valor para utilização de *GetFeatureInfo*.

```
http://hostname:port/path?service=WMS&
request=GetFeatureInfo&version=1.1.1&crs=EPSG:25830&
query_layers=agricultura&feature_count=100&
x=291&y=445&info_format=text/html&
bbox=420353.19,4468089.67,469858.99,4494819.23&
width=729&height=345&layers=agricultura&
style=padrao&format=image/png&bgcolor=0xFFFFFFFF&
transparent=false
```

### 2.1.2.3 Recuperando dados geográficos com serviços WFS

Existem dois modelos para representação de informações geográficas: vetorial e *raster*. O primeiro modela do ponto de vista discreto e o segundo do ponto de vista contínuo. O serviço WFS trabalha com entidades geográficas com dados discretos ou vetoriais em formato GML, representando atributos e geometrias. O serviço proporciona a publicação, acesso e atualização de dados geográficos vetoriais.

O padrão WFS suporta quatro tipos de perfil: simples, básico, transacional, de bloqueio. A Tabela 2.2 relaciona os perfis às operações disponíveis e detalha o papel das operações conforme apresentado em [3].

Segue abaixo um exemplo ilustrando o comando *GetFeature* do perfil Simples [37]:

Tabela 2.2: Perfis WFS.

S=Simple, B=Básico, T=Transaccional, BL = de Bloqueio				Operação	Descrição
X	X	X	X	<i>GetCapabilities</i>	Retorna um documento com os metadados do serviço.
X	X	X	X	<i>DescribeFeatureType</i>	Retorna um esquema XML com a descrição dos tipos de dados.
X	X	X	X	<i>ListStoredQueries</i>	Retorna uma lista de consultas armazenadas no servidor.
X	X	X	X	<i>DescribeStoredQueries</i>	Retorna um esquema XML com os metadados das consultas armazenadas.
X	X	X	X	<i>GetFeature</i>	Retorna uma coleção de objetos geográficos em função de uma consulta armazenada.
	X	X	X	<i>GetFeature</i>	Retorna uma coleção de objetos geográficos em função de uma consulta sob demanda (Query).
	X	X	X	<i>GetPropertyValue</i>	Recupera o valor de uma propriedade de um objeto geográfico.
		X	X	<i>Transaction</i>	Executa operações de exclusão, inclusão, atualização ou substituição de um objeto geográfico.
			X	<i>GetFeatureWithLock</i>	Recupera um conjunto de objetos geográficos e bloqueia qualquer tipo de transação sobre eles.
			X	<i>LockFeature</i>	Bloqueia objetos geográficos.

```

http://www.someserver.com/wfs.cgi?SERVICE=WFS&
VERSION=2.0.0&
REQUEST=GetFeature&
STOREDQUERY_ID=urn:ogc:def:query:OGC-WFS::GetFeatureById&
ID=INWATERA_1M.1013

```

#### 2.1.2.4 Recuperando informações espaço-temporais com WCS

Segundo [2], WCS é o serviço que dá suporte a recuperação de dados espaciais como coberturas, isto é, informação geoespacial representando fenômenos que variam no tempo e no espaço.

Em [3] temos uma descrição mais detalhada do conceito de cobertura como sendo: “elementos que suportam a representação de fenômenos geográficos no domínio espacial, temporal, ou espaço-temporal, e que contêm uma série de atributos comuns a todas as

posições geográficas dentro de um dado domínio que descreve o fenômeno representado. O domínio consiste em uma coleção de posições específicas e ordenadas em um espaço coordenado que pode definir-se em no máximo três dimensões espaciais (x,y,z) e uma dimensão temporal (t)”.

As operações básicas do serviço são *GetCoverage* e *DescribeCoverage*, a primeira recupera o documento XML/GML com os dados e a segunda com os seus metadados.

Segue abaixo um exemplo do comando *GetCoverage* extraído de [3] e da resposta a requisição:

```
http://www.idee.es/wcs/IDEE-WCS-UTM30N/wcsServlet
?service=wcs&request=getcoverage&version=1.0.0&crs=epsg:23030
&bbox=470000,4130300,470200,4130500&coverage=mdt25_peninsula_
zip&resx=25&resy=25&format=asciigrid&exceptions=xml

ncols          8
nrows          8
xllcorner      470000.0
yllcorner      4130300.0
cellsize       25.0
NODATA_value   -999.0
1208.0 1213.0 1220.0 1227.0 1231.0 1231.0 1230.0 1230.0
1206.0 1211.0 1220.0 1227.0 1231.0 1231.0 1230.0 1229.0
1203.0 1210.0 1218.0 1225.0 1229.0 1231.0 1230.0 1229.0
1200.0 1208.0 1214.0 1222.0 1226.0 1229.0 1229.0 1227.0
1202.0 1209.0 1212.0 1218.0 1223.0 1227.0 1227.0 1226.0
1202.0 1209.0 1208.0 1214.0 1221.0 1224.0 1224.0 1224.0
1201.0 1201.0 1199.0 1210.0 1216.0 1221.0 1222.0 1222.0
1191.0 1191.0 1200.0 1205.0 1213.0 1219.0 1221.0 1221.0
```

Atuações típicas de serviços WCS seriam o de fornecimento de dados para acompanhamento meteorológico e previsão do tempo, geração de modelos de elevação de solo ou de superfície. Servidores WCS podem publicar e acessar conjunto de dados no formato *raster*, em malha, ou coleções de séries temporais. São exemplos de coberturas: imagens de rastreamento (*raster*) de satélite, redes irregulares trianguladas (TIN<sup>6</sup>) e coberturas de pontos e polígonos.

### 2.1.2.5 Publicando e acessando metadados de serviços com CSW

O catálogo de serviço permite a localização de dados ou serviços geográficos através de um conjunto de operações, abaixo descritas:

<sup>6</sup>Tipo específico de modelo digital de terreno, baseado em triângulos

1. *DescribeRecord*: Descreve o esquema de metadados, ou modelo de informação, utilizado pelo serviço e a lista de elementos que compõem um registro de catálogo.
2. *GetDomain*: Recupera a gama de valores preenchidos que compõem o domínio de um determinado elemento de registros de metadados.
3. *GetRecords* : Recupera um conjunto de registros de metadados.
4. *GetRecordById*: Recupera um registro específico a partir de sua identificação.
5. *Harvest*: Permite a programação de consultas periódicas a outros serviços possibilitando a incorporação de metadados.
6. *Transaction*: Permite a criação, exclusão e atualização de registros no catálogo.

Segue abaixo um exemplo de uso do comando *GetRecords*:

```
http://www.metadados.inde.gov.br/geonetwork/srv/por/csw?
request=GetRecords&
service=CSW&version=2.0.2&
typeName=csw:Record&
constraintLanguage=CQL_TEXT&
namespace=xmlns(csw=http://www.opengis.net/cat/csw/2.0.2),
  xmlns(gmd=http://www.isotc211.org/2005/gmd)&
constraint_language_version=1.1.0&
resultType=results&
startPosition=4675&
maxRecords=5000&
ElementSetName=full
```

## 2.2 Arquitetura Orientada a Serviços (SOA)

Segundo MacKenzie *et al.* em [25], SOA é “um paradigma para organizar e utilizar capacidades distribuídas que podem estar sobre o controle de diferentes domínios proprietários”. O elemento responsável por suprir estas capacidades são os serviços. Para Marks e Bell em [27] SOA é apresentada como uma arquitetura conceitual, na qual as funcionalidades do negócio são expostas como serviços compartilhados e reutilizáveis, em uma rede de tecnologia da informação. Os serviços são classificados como módulos de negócio ou como funcionalidades de uma aplicação, com suas interfaces expostas e invocados por meio de mensagens.

Clements em [10] apresenta SOA como um estilo arquitetural capaz de, em um determinado contexto, solucionar problemas específicos de arquiteturas de sistemas de *software*. Faz parte deste contexto um conjunto de consumidores que necessitam utilizar serviços

proporcionados por provedores distintos, sendo que os consumidores não possuem um conhecimento prévio da implementação desses serviços. Os problemas específicos a serem solucionados com o estilo arquitetural SOA se traduzem em como obter interoperabilidade entre componentes distribuídos em plataformas, linguagens, localizações e endereços de Internet distintos; como localizá-los e combiná-los dinamicamente e como atingir performance, disponibilidade e segurança adequados.

Uma das estratégias utilizadas pelos profissionais e cientistas da área de tecnologia da informação para disseminar novas ideias, ou paradigmas, que possam representar uma ruptura ou mudança na forma de se produzirem sistemas de informação, tem sido a publicação de manifestos. Dentre os mais difundidos na área de TI, temos o manifesto do software livre, ou manifesto GNU<sup>7</sup> e, mais recentemente, o manifesto ágil<sup>8</sup>. Uma característica destes manifestos é a definição de uma lista de valores, no qual valores do novo paradigma se sobrepõem a valores do antigo paradigma.

Utilizando a mesma estratégia, foi lançado em 2009 por dezessete cientistas e profissionais de TI, o manifesto SOA<sup>9</sup>. O manifesto apresenta SOA como o resultado da aplicação do novo paradigma da orientação a serviço. Este novo paradigma auxiliaria as empresas em agregar valor ao negócio, obter maior agilidade, melhorar a eficácia do custos investidos e alinhar as necessidades de mudança ao negócio.

Erl em [16] lista as ambições e os valores presentes no manifesto SOA, sobre a forma de priorização de valores:

- O valor do negócio sobre a estratégia técnica.
- Os objetivos estratégicos sobre os benefícios específicos do projeto.
- Uma interoperabilidade intrínseca sobre uma integração personalizada.
- Serviços compartilhados sobre implementações de fins específicos.
- Flexibilidade sobre a otimização.
- Refinamento evolutivo sobre perseguição da perfeição inicial.

Erl, também apresenta em [16] os sete objetivos de se aplicar a orientação a serviço, sendo os três últimos considerados como estratégicos:

1. Maior interoperabilidade intrínseca.
2. Aumento da federação.

---

<sup>7</sup>Escrito por Richard Stallman em 1985, critica o uso de *softwares* proprietários em detrimento de *softwares* de código aberto, <http://www.gnu.org/gnu/manifesto.pt-br.html>

<sup>8</sup>Manifesto ágil, <http://www.manifestoagil.com.br/>

<sup>9</sup>Manifesto SOA, <http://www.soa-manifesto.org/aboutmanifesto.html>

3. Maior opção de diversificação de fornecedores.
4. Aumento do alinhamento entre o domínio do negócio e o tecnológico.
5. Maior retorno sobre o investimento.
6. Maior agilidade organizacional.
7. Diminuir a carga sobre a TI.

A relação entre princípios e objetivos é apresentada na Figura 2.4.

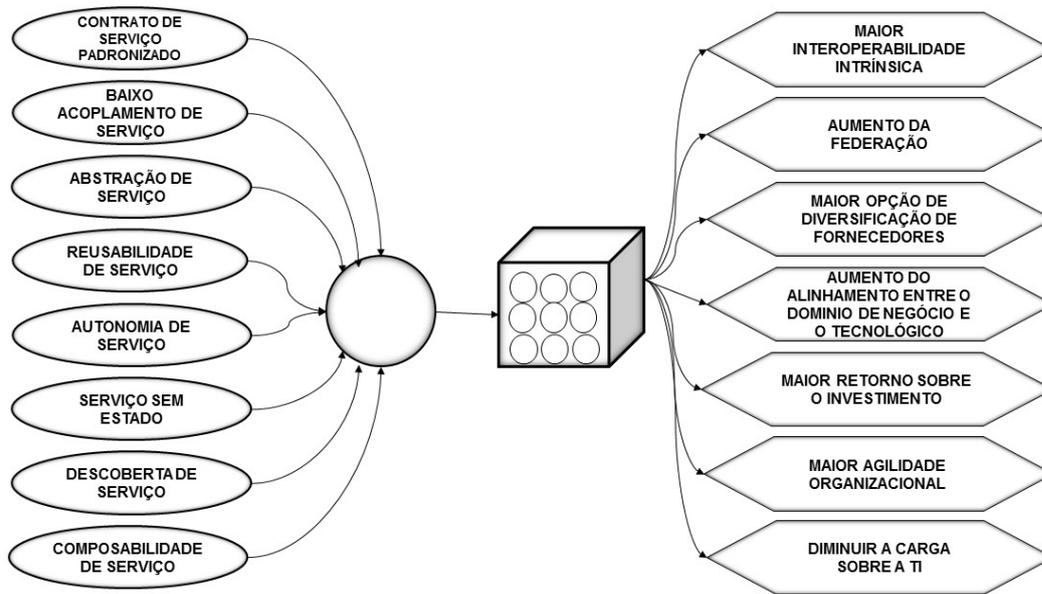


Figura 2.4: Princípios e objetivos de SOA.  
Fonte: adaptado de [16].

## 2.3 Metodologias de Análise e Projeto Orientadas a Serviço

Kohlborn *et al.* em [22] realizaram um trabalho de análise crítica e comparativo entre metodologias orientadas a serviço. Segundo os autores, haveria a necessidade de avanços relacionadas aos métodos de identificação e de análise de serviços, tanto no nível de serviços de negócio quanto no nível de serviços de *software*. Os avanços seriam no intuito de um melhor nível de alinhamento entre negócio e TI e uma maior agilidade nas mudanças dos serviços. Os autores ressaltam a necessidade de uma nova abordagem para suprir a

real necessidade de alinhamento, preconizada pela orientação a serviço, caracterizando-a como uma abordagem holística de análise de serviços.

Lane e Richardson em [23] realizaram uma revisão de literatura com o objetivo de identificar modelos de processos de desenvolvimento de serviços e, como objetivo secundário, identificar dentre eles modelos que auxiliassem na adaptação de aplicações orientadas a serviço. Um dos resultados deste trabalho foi um metamodelo *Unified Modeling Language* (UML)<sup>10</sup>, apresentado na Figura 2.5, que expressa uma tipologia, hierarquia e relacionamentos de processos.

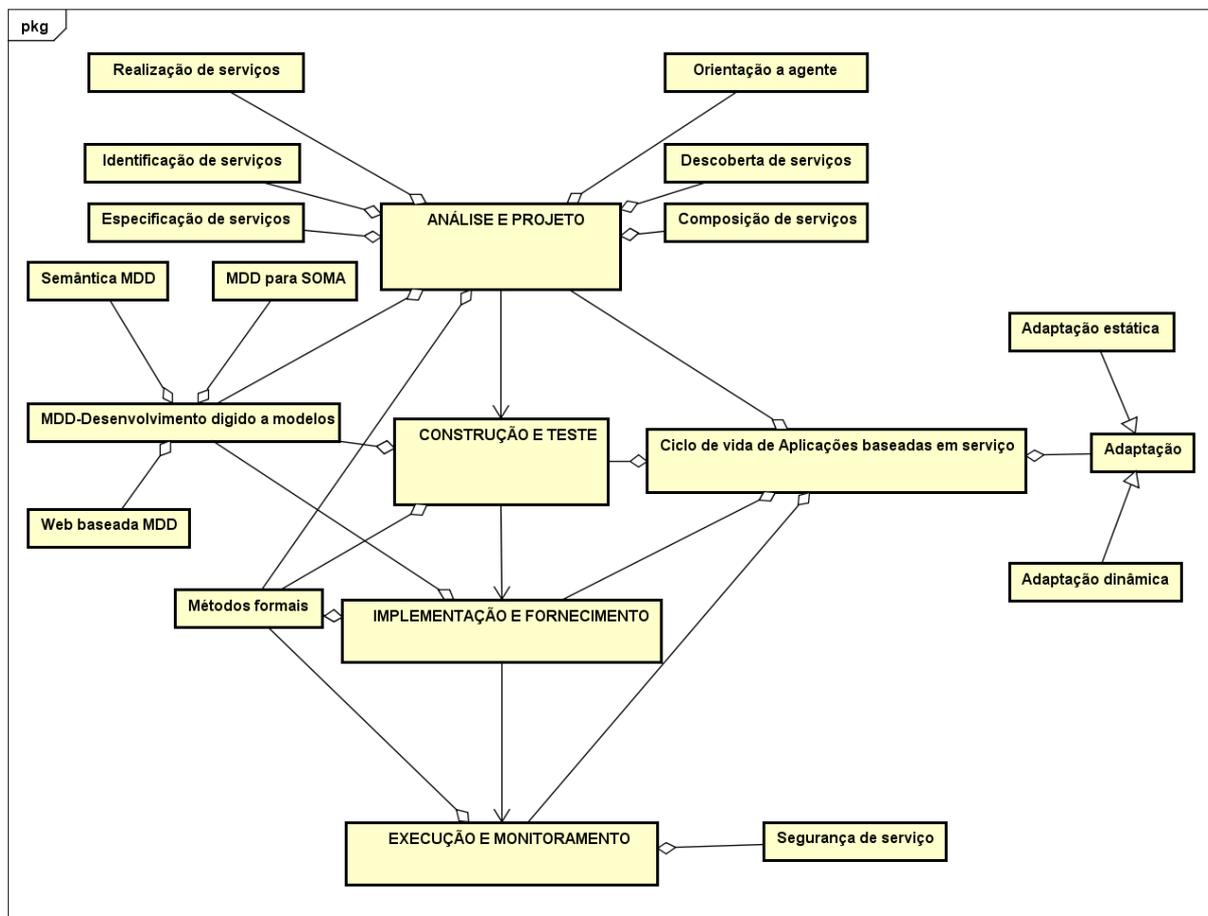


Figura 2.5: Metamodelo de processos de desenvolvimento de serviços.  
Fonte: adaptado de [23].

O metamodelo apresentado possui quatro grandes grupos de processos: análise e projeto, construção e teste, implementação e fornecimento e, por último, execução e fornecimento. As duas classes mais presentes na composição dos grupos foram os métodos formais e os baseados em desenvolvimento dirigido a modelos - MDD<sup>11</sup>. Outro achado

<sup>10</sup>Para criação de meta modelos em UML utiliza-se o padrão MOF, <http://www.omg.org/mof/>

<sup>11</sup>Estilo de desenvolvimento cujo principal artefato são os modelos. Está intimamente ligado a MDA - Arquitetura Orientada a Modelo

da pesquisa foi que havia um número relativamente pequeno de artigos que tratavam de experiências ou resultados práticos do uso dos modelos, o que refletia um pouco amadurecimento dos processos estudados.

Ameller *et al.* em [1] avaliaram o relacionamento entre desenvolvimento dirigido a modelos (MDD) e arquiteturas orientadas a serviço (SOA) no intuito de compreender quais características do primeiro contribuiriam para o desenvolvimento do segundo. Avaliações quantitativas concluíram uma predominância em MDD de abordagens *top-down* em relação a *bottom-up* e horizontais, predominância de contribuição a métodos da fase construção e grande uso de diagramas UML, principalmente diagramas de classe e de atividades.

Usländer em [35], na elaboração da metodologia SERVUS, faz uma avaliação das metodologias para análise e projeto orientadas a serviços. A avaliação foi feita de forma a levantar os aspectos não atendidos, ou abordados de forma diferenciada, por estas metodologias e que seriam essenciais em uma metodologia para o desenvolvimento de sistemas de informação ambientais orientado a serviços.

Seguem abaixo algumas considerações feitas pelo autor:

1. As metodologias atuais partem do pressupostos da existência de uma grande variedade de processos de negócio, fato que não ocorre com sistemas de informações ambientais, e valorizam os aspectos funcionais em detrimento dos informacionais.
2. O mapeamento de requisitos funcionais para capacidades dos serviços devem considerar a maneira como as meta informações estão descritas.
3. Os modelos de alto nível, como por exemplo o CIM<sup>12</sup> da MDA, são negligenciados e dificultando o atendimento aos requisitos de suporte a iteratividade no desenvolvimento conjunto de requisitos e arquitetura, e os associados a *feedback* compreensão das necessidades do usuário final.
4. Há uma deficiência de ferramentas não proprietárias para dessas metodologias.

Para suprir estas deficiências Usländer propõe uma nova metodologia, que será apresentada no próxima seção.

## 2.4 Metodologia SERVUS

A sigla SERVUS é oriunda de dois termos da engenharia de desenvolvimento de sistemas: serviço (**SERV**ice) e caso de uso (**US**e case). Consiste em uma metodologia de aná-

---

<sup>12</sup>A MDA propõe três tipos de modelos: PSM - Modelo específico de plataforma, CIM - Modelo independente de computação, PIM - Modelo independente de plataforma

lise e projeto, orientado a serviços, voltada para construção de Sistemas de Informação Ambientais (EIS).

O problema central que a metodologia se propõe a resolver é, dado um conjunto de casos de uso que atendam aos requisitos de negócio e demandem recursos, e dado uma rede de serviços geoespaciais com capacidades e que disponibilize recursos, como descobrir e associar recursos requisitados a recursos ofertados, de forma a compor uma aplicação que supra os requisitos de qualidade, funcionais e informacionais estabelecidos. Dois outros elementos compõe este cenário: descrições semânticas de recursos e condições inerentes (*side conditions*) ao desenvolvimentos de sistemas ambientais, como por exemplo, processos de descoberta de recursos, uso de padrões, combinações de recursos.

O processo de descoberta deve levar em consideração a existência de descrições dos recursos, na forma de metadados ou anotações semânticas, assim como as associações entre estes recursos, na forma de uma rede de recursos semânticos. Os requisitos de interoperabilidade e de reuso, e os padrões de geosserviços devem ser fortemente considerados no processo de descoberta de capacidades na rede de geosserviços.

Segundo Uslander, “SERVUS compreende o projeto de EIS<sup>13</sup> como uma atividade iterativa de descoberta e associação: capacidades disponíveis são descobertas e associadas a requisitos do usuário formulados como casos de uso”.

A Figura 2.6 ilustra o mapeamento entre requisitos e capacidades.

A metodologia é composta por uma linguagem de modelagem, um processo de desenvolvimento e uma arquitetura de referência. A linguagem de modelagem é apoiada por um modelo conceitual de referência. Durante o processo de desenvolvimento são criados artefatos tais como: um modelo de domínio, um modelo de processos/casos de uso e um modelo de projeto. O modelo de projeto é o principal deles, ele expressará os requisitos e as capacidades sob a forma de recursos. Os modelos gerados são baseados nas camadas de abstração propostas por Bieberstein *et al.* em [4].

A Figura 2.7 exhibe o mapeamento entre requisitos e capacidades, os modelos produzidos e a associação com as camadas propostas Bieberstein.

### 2.4.1 O modelo de domínio

O modelo de domínio está associado a camada de abstração empresarial e prioriza os aspectos informacionais do negócio em detrimento dos aspectos organizacionais. A metodologia não prescreve notações, linguagens ou nível de formalismo para o modelo de domínio. Dentre as opções de modelagem de domínio temos as linguagens de ontologia<sup>14</sup> e a UML. Em alguns projetos o modelo de domínio pode simplesmente ser importado

---

<sup>13</sup>EIS são sistemas de informações ambientais que lidam com serviços e informações geoespaciais

<sup>14</sup>Segundo Gruber, ontologia é uma especificação explícita de uma conceituação [20]

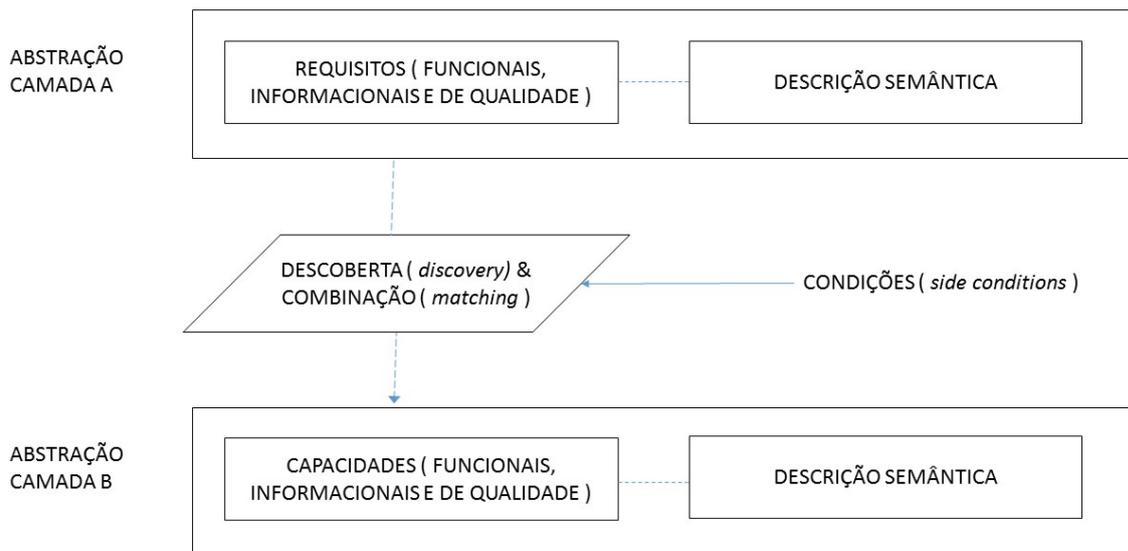


Figura 2.6: Mapeamento de requisitos para capacidades.  
 Fonte: adaptado de [35].

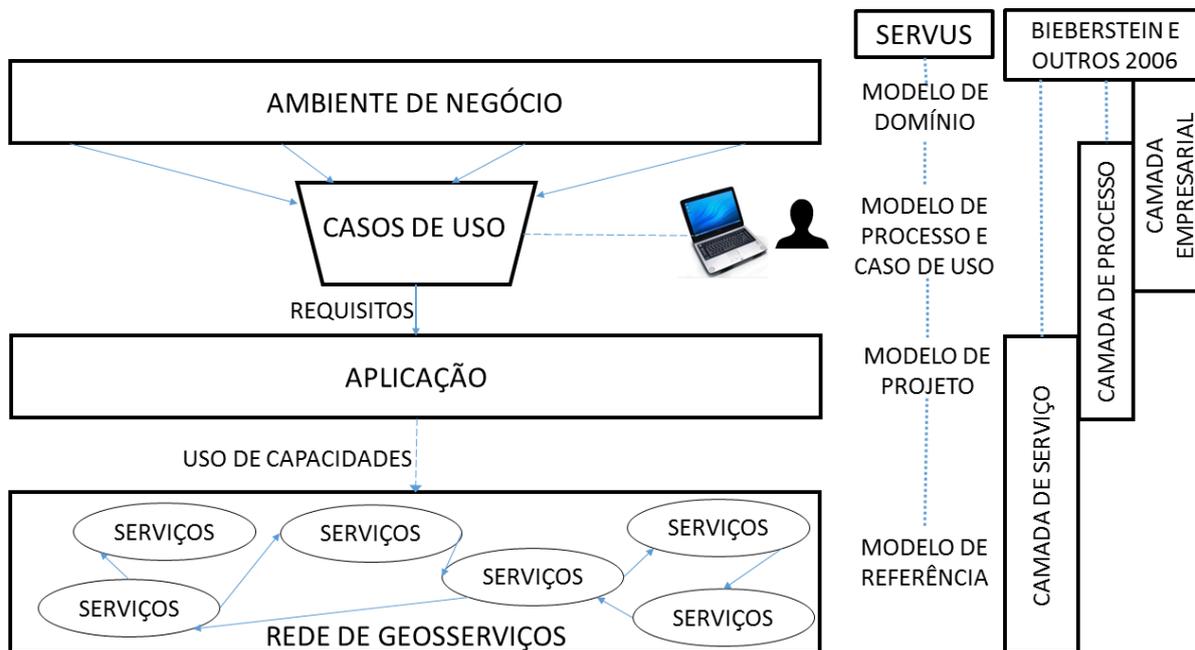


Figura 2.7: Hierarquia de modelos SERVUS.  
 Fonte: adaptado de [35].

de forma integral, em outros ele poderá ser construído ou incrementado no decorrer do processo de desenvolvimento.

Nas palavras de Usländer em [35] temos que “o modelo de domínio da metodologia SERVUS representa o domínio temático para qual o problema pertence. Ele formalmente define a parte do mundo que compreende o universo de discurso entre o usuário e o projetista do sistema, isto é, compreende o conhecimento compartilhado a cerca do domínio da aplicação. Tipicamente tais conhecimentos compartilhados são representados pela especificação de uma ontologia”.

Um exemplo típico de modelagem de domínio em UML é a Especificação Técnica para Aquisição de Dados Geoespaciais Vetoriais (ET-EDGV)<sup>15</sup> da Infraestrutura Nacional de Dados Espaciais (INDE) do Brasil. Um exemplo de modelo de domínio em ontologia são as produzidas pela *Semantic Web for Earth and Environmental Terminology* (SWEET) e apresentadas em [31].

Modelos de domínio são tipicamente conceituais e podem seguir o padrão *General Feature Model* proposto pelo OGC, cuja unidade básica de modelagem, ou principal classe abstrata, é a feição geográfica (*geographic feature*). Em [35] feições são apresentadas como abstrações do mundo real com localizações na Terra, ou seja georreferenciadas, e possuem atributos como tema, localização, tempo e podem associar-se entre si.

## 2.4.2 Modelos de processo e de caso de uso

Os modelos de processo e caso de uso são o resultado da atividade de análise de requisitos e estão relacionados a camada de processo. Adota-se a definição de Jacobson e Pan-Wei Ng [35, apud] de que um caso de uso modela o comportamento de um sistema e consiste numa sequência de ações realizada pelo sistema e expressam requisitos funcionais, informacionais e de qualidade. SERVUS propõe a utilização de uma extensão de caso de uso que reforce os aspectos informacionais do caso de uso. A extensão de caso uso considera as ações do sistema como demandadoras de recursos em serviços externos ou como provedoras de recursos informacionais. A metodologia sugere a utilização de uma descrição textual estruturada seguindo um *template* proposto. Um exemplo de *template* poderia ser composto pelo seguintes campos: escopo, ator principal, pessoas envolvidas (*stakeholders*), pré-condições, pós condições, requisitos principais e extensões.

## 2.4.3 O modelo de referência - RM-OA

O modelo de referência “especifica os principais conceitos, termos e relacionamentos que são essenciais para a arquitetura” [35]. SERVUS utiliza um modelo de referência próprio denominado ORCHESTRA ou *Reference Model for de ORCHESTRA Architecture*

---

<sup>15</sup><http://www.concar.ibge.gov.br/detalheDocumentos.aspx?cod=94>

(RM-OA). Segundo Usländer, a definição de um modelo de referência é um aspecto importante para um co-desenvolvimento entre requisitos e arquitetura. O modelo de referência é composto por um *framework* arquitetural, responsável pelas orientações e regras de como especificar o sistema, e por um modelo conceitual com as principais entidades e relacionamentos.

O *framework* arquitetural incorpora o conceito de visão arquitetural, presente no modelo de referência *Reference Model of Open Distributed Processing* (RM-ODP) para sistemas distribuídos, realizando uma pequena adaptação para os conceitos de orientação a serviço. A Tabela 2.3 apresenta esta adaptação entre RM-ODP e RM-OA proposta por Usländer.

Tabela 2.3: Interpretação de RM-ODP para RM-OA de Usländer.

Ponto de vista RM-ODP	Interpretação RM-OA(aplicada a SERVUS)
Empresarial	Reflete a fase de análise em termos do sistema e requisitos do usuário, assim como a avaliação da tecnologia. Inclui regras que governam atores e grupos de atores e seus papéis.
Informação	Especifica a abordagem de modelagem de todas categorias de informação, incluindo suas características temáticas, espaciais, temporais, assim como seus metadados.
Computacional	(No RM-OA e na metodologia SERVUS refere-se ao <b>ponto de vista de serviço</b> ). Especifica as interfaces e os tipos de serviço.
Tecnologia	Especifica as escolhas tecnológicas da plataforma, suas características e questões operacionais
Engenharia	Especifica o mapeamento das especificações de serviço e modelos de informação para a plataforma escolhida. Considera as características e princípios para a rede de geosserviços.

Fonte: Adaptado de [35].

O modelo conceitual de SERVUS é um metamodelo segundo os preceitos de MDA e MOF, com seus metamodelos e metaclasses. As principais metaclasses são: feição(*feature*), interface, serviço e recurso. O modelo conceitual ilustrado na Figura 2.8 é composto por três subconjuntos de metaclasses, o primeiro está associado visão de serviço, o segundo visão de informação e o terceiro à visão de recurso.

O modelo conceitual tem as seguintes características:

- Feições possuem propriedades que podem ser atributos descrevendo suas características.
- Operações e papéis podem ser outros tipos de atributos de feições.

- Feições podem associar-se entre si e estas associações podem ser consideradas como feições.
- Serviços possuem interfaces, que possuem operações compostas por métodos e métodos possuem parâmetros de entrada ou de saída.
- Recursos possuem representações, métodos e ligações com outros recursos.

No proposta de modelo de recurso são incluídas as classes Recurso Semântico e Ligação Semântica para melhor expressar as questões semânticas. Recurso e Recurso semântico serão a base para a criação de uma rede de recursos.

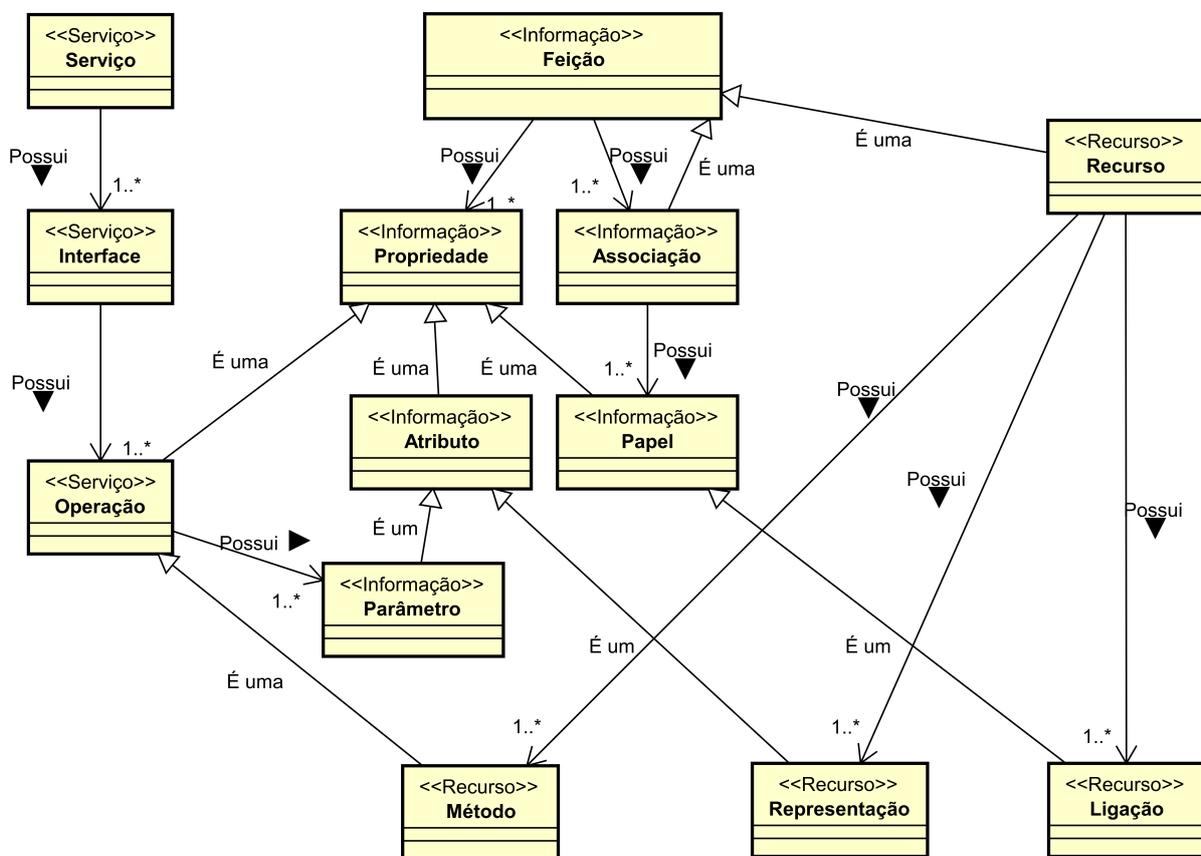


Figura 2.8: Modelo conceitual SERVUS.

#### 2.4.4 O modelo de projeto

O modelo de projeto é o principal da metodologia e participa tanto da camada de processo, quanto da de serviço. O modelo de projeto pode ser visto como uma composição de outros três modelos: o de requisitos, o de capacidades e o de mapeamento entre os dois primeiros.

A base para a criação dos modelo de requisitos (REQ's) e de capacidades(CAP's) é modelo conceitual de recursos apresentado anteriormente. Ou seja, requisitos e capacidades devem ser traduzidos para recursos e recursos semânticos. Os componentes do modelo de capacidades podem ser tipos ou instâncias de capacidades. O modelo de mapeamento (REQ2CAP) pode ser representado por um grafo interligando os nós de requisitos e os nós de capacidades.

## 2.4.5 O processo de desenvolvimento

A Figura 2.9 apresenta um modelo em BPMN simplificado com as principais etapas do processo de desenvolvimento. A Tabela 2.4 apresenta os principais objetivos e os artefatos gerados em cada etapa.

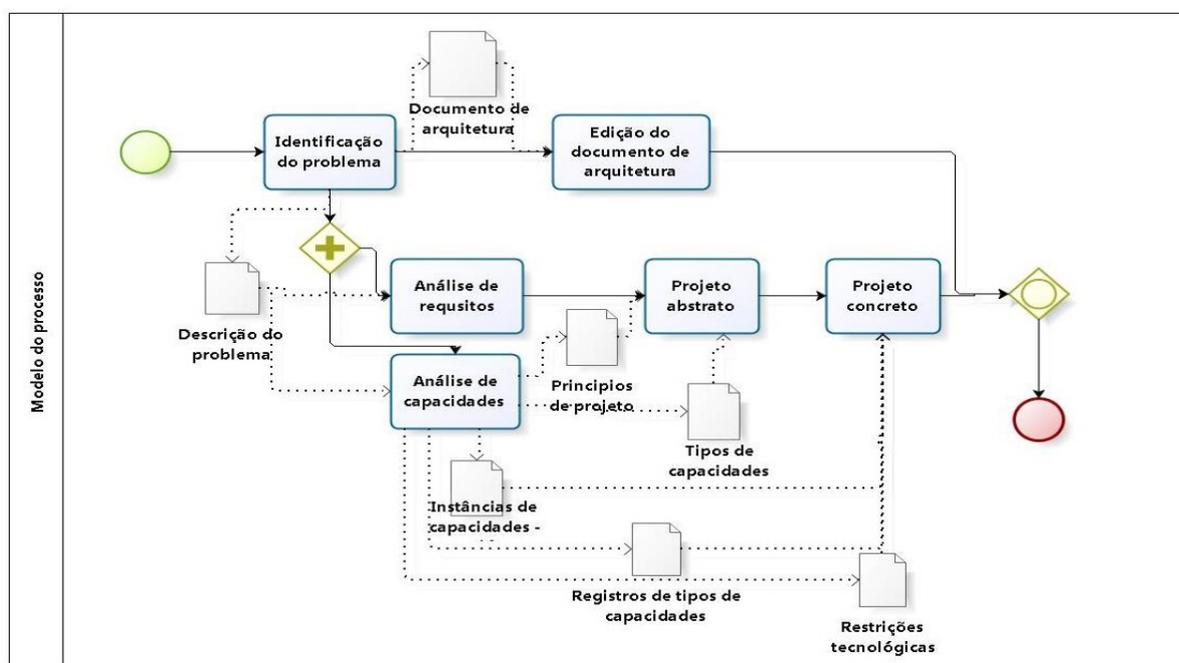


Figura 2.9: Modelo BPMN com as etapas do processo de desenvolvimento SERVUS.

Segue abaixo uma descrição das principais atividades do processo de desenvolvimento. A Figura 2.10 apresenta um quadro geral dos modelos e atividades da etapa de projeto.

1. *Publishing*: atividade, manual ou automatizada, de busca de capacidades na rede de serviços geoespaciais e de incorporação ao modelo de capacidades. As capacidades são traduzidas para recursos ofertados, ou seja, os disponíveis na rede de geosserviços.

Tabela 2.4: Processo de desenvolvimento: etapas, objetivos e artefatos.

<b>Etapa</b>	<b>Objetivos</b>	<b>Artefatos</b>
Identificação do problema	Análise e descrição do problema a ser resolvido.	Descrição do problema e a versão inicial do documento de arquitetura
Edição do documento de arquitetura	Construção de documento de arquitetura com as visões propostas pela RM-OA.	Versões intermediárias e versão final
Análise de capacidades	Mapeamento dos tipos e instâncias de capacidades disponíveis na rede de geosserviços.	Arquitetura de referência - RM-OA, padrões geoespaciais, tecnologias e a rede de geosserviços como artefatos de entrada. Princípios de projeto, restrições tecnológicas, registros de tipos e instâncias de capacidades como artefatos de saída
Análise de Requisitos	Refinamento da descrição do problema e tradução para requisitos do usuário e do sistema (informativos, de qualidade e funcionais)	Documento de arquitetura atualizado com a visão arquitetural empresarial, modelo de domínio
Projeto abstrato	Elaboração de projeto independente de plataforma. Construção de um modelo de informação e de serviços baseado nos tipos de capacidades.	Tipos de capacidades e princípios de projeto como entrada. Documento de arquitetura atualizado com as visões arquiteturais de informação e de serviço e decisões de projeto como artefatos de saída
Projeto concreto	Especificação de uma arquitetura de implementação e de uma plataforma concreta de serviços	Instâncias de capacidades e restrições tecnológicas como entrada. Documento de arquitetura atualizado com as visões arquiteturais de tecnologia e engenharia como saída

Fonte: Adaptado de [35].

Nesta etapa podem ser incorporadas ao modelo de projeto os tipos de capacidades, seus metadados ou a própria instância das capacidades. Geralmente a atividade é auxiliada por um procedimento automatizado ou manual de consulta a catálogos de serviço. A etapa de publicação pode ser feita em dois passos: a publicação dos tipos de capacidades e a publicação das instâncias das capacidades. Tanto os tipos quanto as instâncias podem ser associadas a conceitos de ontologias.

2. *Rephrasing*: Atividade responsável por traduzir os requisitos, de forma a extrair um conjunto de recursos necessários para realização dos casos de uso. Inicia-se com a



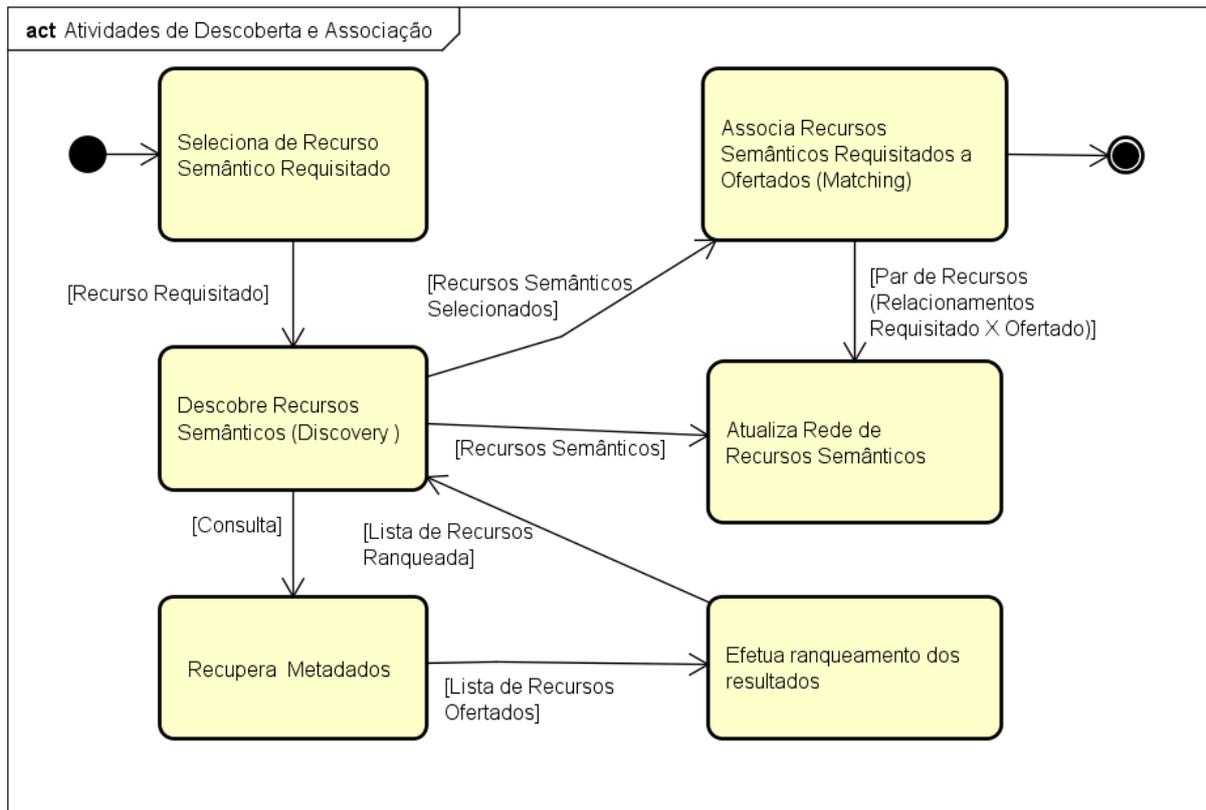


Figura 2.11: Diagrama de atividades de *Discovery e Matching*.  
Fonte: adaptado de [35].

5. *Grounding*: atividade que disponibiliza uma nova capacidade na rede de serviços geoespaciais.

## 2.5 Geocatálogos e busca semântica

Um dos aspectos essenciais para o bom funcionamento das Infraestruturas de Dados Espaciais (IDEs) é a interoperabilidade. Embora a disseminação dos padrões da OGC tenham proporcionado um avanço nesta questão, a ampliação da rede de geosserviços, a necessidade de análises do espaço geográfico envolvendo múltiplas feições (*features*) e com informações oriundas de produtores diversos, cada qual com seu domínio de conhecimento e vocabulários específicos, fez com que os especialistas na área de geoinformática ampliassem a discussão para o campo da organização dos metadados e da utilização da semântica.

Vaccari *et al.* em [36] realizaram análises referentes à interoperabilidade entre infraestruturas de dados espaciais. Elencam aspectos a serem considerados no âmbito dos dados geográficos e de geosserviços, a fim de que sejam minimizadas as deficiências de busca e

seleção de recursos e se atinja uma integração eficiente entre os serviços, superando os que autores classificaram como heterogeneidade semântica.

Em relação aos dados geográficos, expõem questões relacionadas às diferenças nas sintaxes dos dados, na estrutura ou representação dos dados e, principalmente, na semântica dos dados. Segundo os autores conflitos ocorrem quando são atribuídos os mesmos nomes (homônimos) à classes ou tipos de atributos que possuem diferentes significados ou então quando nomes distintos (sinônimos) são utilizados em elementos semanticamente idênticos.

Este desafio de lidar com a heterogeneidade semântica está presente na tarefa de se construir serviços responsáveis pela descoberta e seleção de dados geográficos e pela composição de geosserviços numa arquitetura SOA. A arquitetura SOA possui três componentes básicos: o cliente de serviço, o provedor de serviço e o catálogo de serviços. Este último tem um significativo papel de grande repositório de informações para localização dos serviços na rede.

Uma das possibilidades para lidar com a resolução de problemas semânticos seria o uso de ontologias. Segundo Gruber em [5, apud] “Ontologia é uma especificação formal e explícita de uma conceitualização compartilhada”. Segundo o esquema de classificação de ontologias de Guarino em [5, apud], ilustrado na Figura 2.12, elas podem ser classificadas quanto a generalidade como: de nível superior, de domínio, de tarefas e de aplicação. Ao lidarmos com a heterogeneidade semântica estamos também tratando de diferentes ontologias e em vários níveis.

Os geocatálogos semânticos a serem apresentados buscaram suprir as deficiências de interoperabilidade e de busca anteriormente colocados, principalmente pela utilização formal de ontologias.

### 2.5.1 O experimento de *Lieberman*

Um dos marcos da utilização de ontologias em IDEs foi o experimento de *Lieberman* [24] realizado no âmbito da OGC<sup>16</sup>. Seu objetivo era avançar no estabelecimento de uma Web Semântica Geográfica, centrada na descoberta, busca e consumo de informações geoespaciais e baseada em especificação semântica formal. Dentre os principais resultados alcançados podemos destacar: o desenvolvimento de geo-ontologias, a elaboração de uma proposta arquitetural de busca semântica e a construção de consultas semânticas em serviços web distribuídos.

---

<sup>16</sup><http://www.opengeospatial.org>

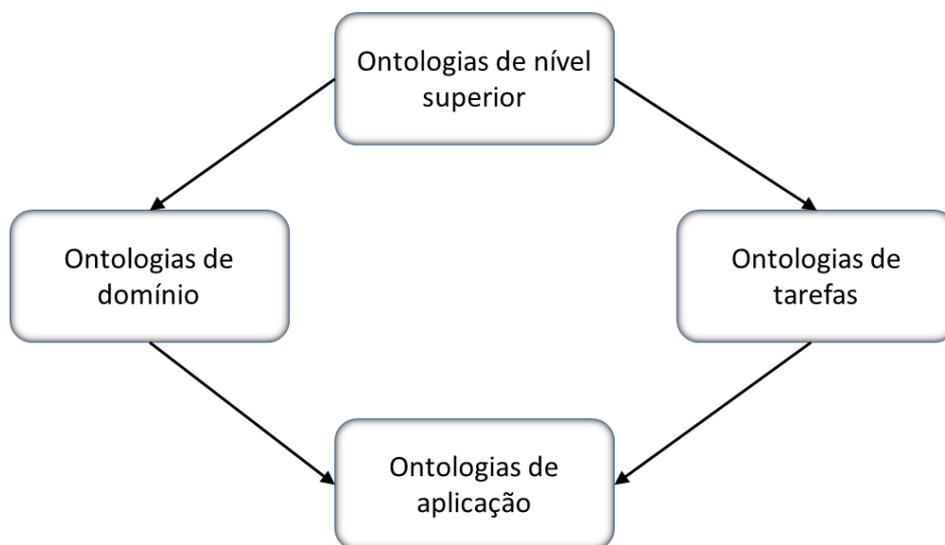


Figura 2.12: Classificação de ontologias de Guarino.

As geo-ontologias criadas foram traduzidas para a linguagem OWL<sup>17</sup> e representavam conjuntos de dados da navegação aeroviária (aeronaves, aeroportos, rotas e localizações geográficas).

As atividades desenvolvidas no experimento foram:

1. Construção de ontologias na linguagem OWL representando os conjuntos de dados fornecidos por serviços de feições geográficas WFS<sup>18</sup>.
2. Construção de uma camada de serviços web (OWL-S)<sup>19</sup> possibilitando buscas nas ontologias criadas.
3. Utilização e refinamento de uma linguagem de busca para atuar sobre a camada OWL-S.
4. Realização de busca não distribuída em serviço WFS.
5. Implementação e teste das capacidades da camada OWL-S.
6. Realização de buscas semânticas envolvendo requisições remotas para serviços de metadados.

<sup>17</sup>Linguagem de marcação criada pela W3C para a construção de ontologias, suas classes, propriedades e indivíduos. Foi baseada na linguagens DAML e OIL. <http://www.w3.org/TR/owl2-syntax>

<sup>18</sup><http://www.opengeospatial.org/standards/wfs>

<sup>19</sup><http://www.w3.org/Submission/OWL-S/>

A Figura 2.13 ilustra a arquitetura criada para o experimento. O bloco arquitetural representado por Sesame consiste em um servidor web, o bloco GSW SAIL é responsável por interpretar os documentos OWL e gerenciar as consultas, o bloco DAMLDB SAIL faz a interface com o banco de grafos DAMLDB.

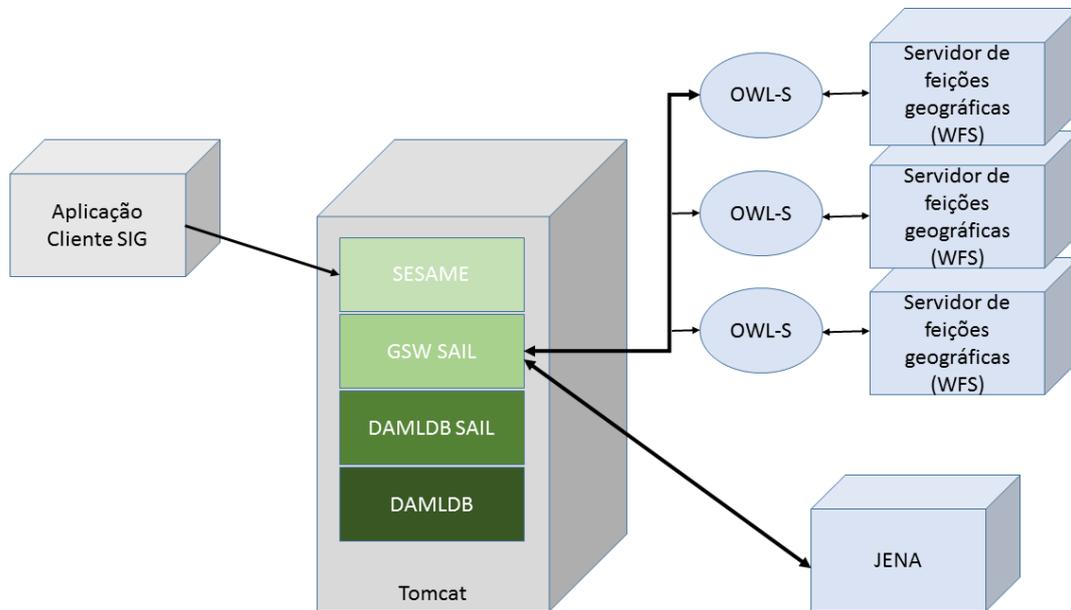


Figura 2.13: Arquitetura do experimento de Lieberman.  
Fonte: adaptado de [24].

O protótipo do experimento utilizou-se da arquitetura Sesame<sup>20</sup>[6], concebida para armazenamento e busca em grandes quantidades de metadados no formato RDF<sup>21</sup> e RDFS<sup>22</sup>. Para um melhor manuseio no formato OWL foi incorporado o framework denominado JENA<sup>23</sup>[9].

A figura 2.14 ilustra uma sequência de busca semântica e os componentes associados.

Para que possa ser atingido o objetivo final de recuperação do recurso em atendimento à busca inicial, ocorrem procedimentos de enriquecimento da consulta por um raciocinador de ontologias e por um cliente de busca. Cabe ao raciocinador recuperar informações de conceitos ou vocabulários controlados associados aos termos de busca em um repositório de ontologias. O cliente de busca, após as solicitações ao raciocinador, recupera informações junto ao serviço de feições e ao catálogo de metadados para, por fim, obter o recurso solicitado pelo usuário.

<sup>20</sup><http://rdf4j.org/>

<sup>21</sup><http://www.w3.org/RDF/>

<sup>22</sup><http://www.w3.org/TR/rdf-schema/>

<sup>23</sup><https://jena.apache.org/documentation/ontology/>

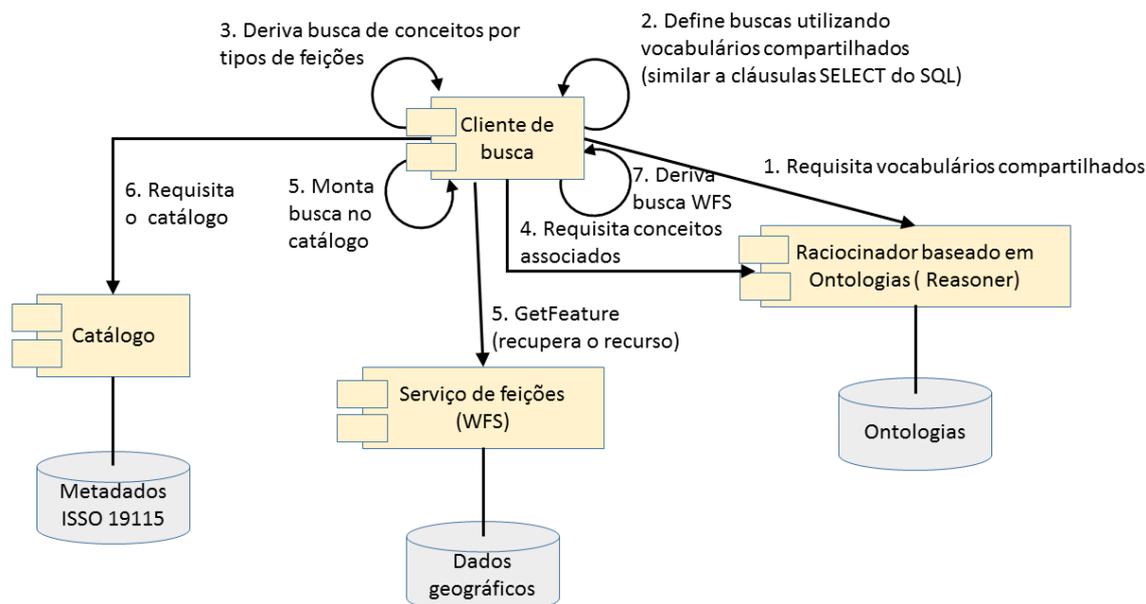


Figura 2.14: Sequência de uma busca semântica.  
Fonte: adaptado de [24].

## 2.5.2 O geocatálogo semântico de Trento

O geocatálogo semântico foi concebido para dar suporte a administração de recursos geográficos para a região de Trento na Itália. Seguiu aos padrões estabelecidos para a Infraestrutura de Dados Espaciais da Comunidade Europeia (INSPIRE), dentre eles, a recomendação de utilização de catálogos de metadados que adotassem o padrão CSW da *Open Geospatial Consortium* (OGC) [34].

A estratégia utilizada na implementação foi a de aproveitar a estrutura de dados espaciais já existente e incorporar recursos que suprissem as deficiências de busca semântica, de usabilidade e de multilinguagem. Sua arquitetura, apresentada na Figura 2.15, seguiu o padrão de três camadas: camada de apresentação, lógica de negócio e acesso a dados.

A camada de apresentação foi incorporada a um portal geográfico<sup>24</sup> já existente, que utilizava a plataforma BEA-ALUI<sup>25</sup>. O serviço de descoberta de recursos geográficos foi incorporado ao portal utilizando-se da tecnologia de *portlets*. Os *portlets* respondiam pelas funções de gerenciamento de metadados, gerenciamento de usuários e grupos de usuários e configuração do sistema. A camada de acesso a dados é realizada pelos serviços do *Geonetwork*<sup>26</sup> responsáveis pelo acesso aos metadados e aos dados geográficos. A camada

<sup>24</sup><http://www.territorio.provincia.tn.it/portal/server.pt>

<sup>25</sup>[https://docs.oracle.com/cd/E13174\char'\\_01/alui/](https://docs.oracle.com/cd/E13174\char'_01/alui/)

<sup>26</sup><http://geonetwork-opensource.org/>

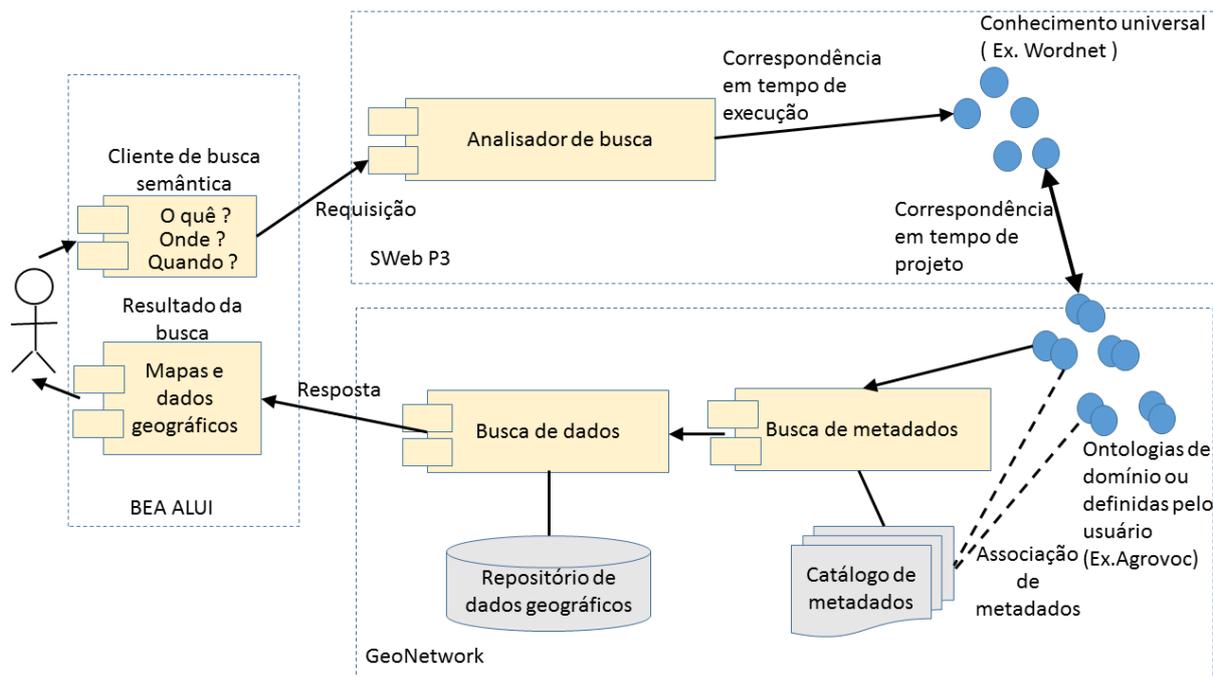


Figura 2.15: Arquitetura e esquema de busca semântica.  
 Fonte: adaptado de [34].

de lógica de negócio é realizada pelo componente SWeb P3 que atua sobre os serviços do *Geonetwork* e funciona como uma extensão para buscas semânticas.

Para a realização das buscas semânticas o componente SWeb P3 utiliza o S-Match<sup>27</sup> [19]. O S-Match é um framework responsável pela execução das operações de correspondência semântica. A principal tarefa do S-Match pode ser resumida no estabelecimento de correspondências entre duas estruturas de árvore representando ontologias ou vocabulários controlados. A utilização de um *framework* de correspondência semântica auxilia na resolução de um esquema de busca semântica envolvendo buscas diretas, indiretas ou por alinhamento [12], conforme ilustrado na figura 2.16.

Seguem abaixo os procedimentos de busca semântica no geocatálogo de Trento :

1. O usuário efetua a busca no portal geográfico, segundo os parâmetros de pesquisa expressando “o quê” esta pesquisando, “onde”, ou em que área específica a consulta se refere, e o “quando”, ou seja, para qual período de tempo a consulta está limitada.
2. Em tempo de execução, o analisador de consulta avalia os parâmetros de pesquisa e busca correspondências num vocabulário controlado, no caso o Wordnet<sup>28</sup>.

<sup>27</sup><http://semanticmatching.org/s-match.html>

<sup>28</sup><https://wordnet.princeton.edu/wordnet/>

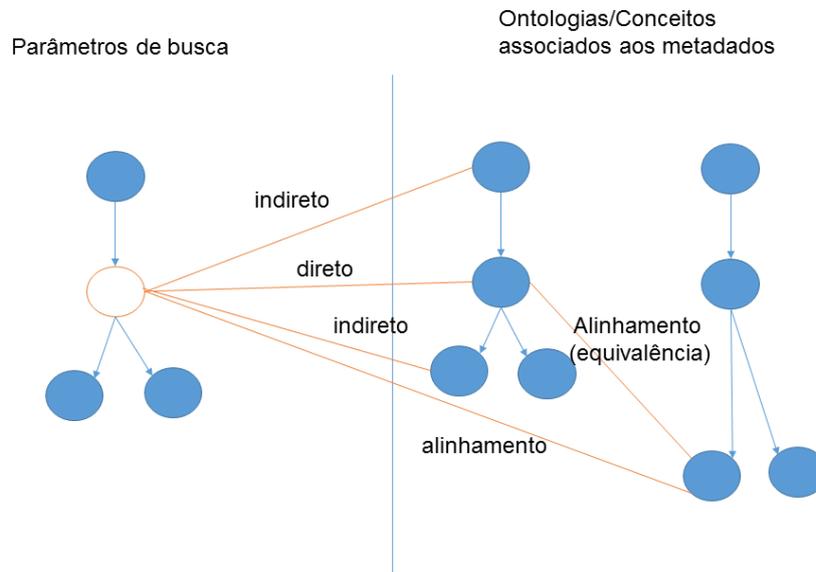


Figura 2.16: Correspondência semântica.

3. Uma vez identificados os termos correspondentes, direta ou indiretamente, são localizadas as associações efetuadas em tempo de projeto que relacionam as ontologias aos termos do vocabulário controlado.
4. Os termos das ontologias serão utilizados para as consultas nos catálogos de metadados e os registros que melhor representem os termos pesquisados serão repassados para o elemento responsável pela busca dos recursos geográficos.
5. Os recursos geográficos serão recuperados e encaminhados como resposta para o cliente.

Um aspecto importante do geocatálogo semântico de Trento é a utilização de múltiplas ontologias, de domínio ou elaboradas pelo usuário, ou, por um outro ponto de vista, uma única ontologia com várias facetadas. Ontologias com facetadas são ontologias divididas em subárvores, denominadas facetadas, cada qual codificando uma diferente dimensão ou aspecto do domínio do conhecimento [17]. As principais facetadas representadas na ontologia foram a hidrografia da região e a formação geológica com suas elevações e depressões naturais.

### 2.5.3 O catálogo semântico da metodologia SERVUS

A metodologia de desenvolvimento de sistemas SERVUS propõe a criação de um catálogo semântico para compor uma arquitetura de implementação. Esta arquitetura possui

quatro grandes blocos: interface de usuário, ambiente de projeto, serviços geoespaciais e sistemas fontes. O catálogo semântico está inserido no ambiente de projeto, conforme ilustrado na Figura 2.17.

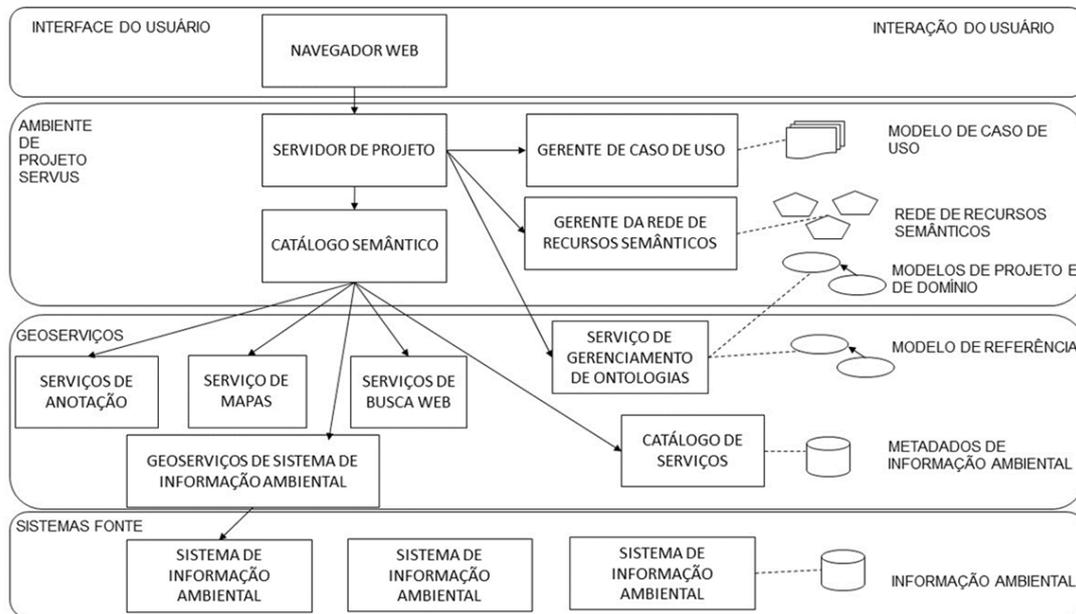


Figura 2.17: Arquitetura de implementação SERVUS.

Fonte: adaptado de [35].

O papel do catálogo semântico é o de se comunicar com os geosserviços da rede e processar as consultas por recursos. Assim como o geocatálogo de Trento, funciona como uma extensão semântica de catálogos padrão CSW<sup>29</sup>. A extensão semântica possibilita as consultas baseadas em ontologias e a avaliação da proximidade semântica dos resultados obtidos. Também possibilita ampliar as fontes de pesquisa de metadados, através de um modelo flexível que não está atrelado a esquemas de metadados específicos. Essa abordagem permite incluir pesquisas em máquinas de busca da internet (google, yahoo) e em gerenciadores de ontologias, além dos catálogos de metadados CSW e UDDI<sup>30</sup>, serviços WMS<sup>31</sup> e WCS<sup>32</sup>. No que se refere ao suporte a metodologia, o catálogo semântico responde pelas atividades de *harvesting*, que consiste na coleta de metadados dos serviços, e de *publishing*, publicação dos recursos disponíveis para uma rede de recursos semânticos.

<sup>29</sup><http://www.opengeospatial.org/standards/cat>

<sup>30</sup><http://uddi.xml.org/>

<sup>31</sup><http://www.opengeospatial.org/standards/wms>

<sup>32</sup><http://www.opengeospatial.org/standards/wcs>

## 2.6 Trabalhos correlatos

Ao tratarmos informações geográficas, lidamos com uma grande quantidade de variáveis, sejam representações variadas para a mesma informação, sejam escalas diferentes, sejam sistemas de referenciamento ou fronteiras geográficas de maior ou menor extensão, todas elas dificultam uma seleção exata da informação requisitada. Uma das formas de lidar com esta questão é o estabelecimento de um ranqueamento dos itens selecionados baseados num índice de relevância da informação. Em catálogos semânticos este índice reflete um nível de proximidade semântica entre os parâmetros de pesquisa e os dados e metadados da informação.

Andrade e Baptista em [13] trilharam este caminho que incluía busca semântica, listagem de grande quantidade de informações oriundas de fontes diversas, estabelecimento de métricas e ranqueamento de recursos disponibilizados numa IDE. Eles partiram do pressuposto de que o uso de ontologias seria o caminho para tornar mais precisas as buscas e facilitar o processo de automatização.

No que se refere a definição da métrica, optaram por uma métrica combinada de dois fatores. O primeiro fator levava em consideração o quanto o tema pesquisado seria similar a informação geográfica e o segundo fator considerava a importância relativa do tema dentro do serviço que provê a informação. Para cálculo da similaridade foi utilizado o recurso de construção de uma rede semântica que incorporou todos os conceitos definidos na ontologia e os relacionamentos semânticos existentes entre eles. Nesta rede os nós representaram os conceitos e as ligações os relacionamentos semânticos. A Figura 2.18 ilustra o esquema de busca semântica utilizado.

Daltio e Carvalho em [12] propuseram a construção de um *framework* para a recuperação de imagens de satélite, utilizando ontologias para integrar aspectos semânticos e espaciais. O *framework* era baseado em um processo de anotação semântica dos recursos geográficos e na utilização de um serviço de gerenciamento de ontologias, o Aondê. Cabe ressaltar neste trabalho a importância dada ao processo de seleção das ontologias. As ontologias selecionadas foram a AGROVOC<sup>33</sup>, SWEET [32], VCGE<sup>34</sup> e OnLocus [15]. Na implementação do *framework* foi utilizado como repositório de dados um banco de dados relacional e criada uma camada de recuperação de dados acessada via *Interface Web*.

Uma das contribuições do trabalho de Daltio e Carvalho foi o estabelecimento de uma heurística utilizada na consulta, que leva em consideração associações entre termos de pesquisa e as anotações semânticas. As associações podem ocorrer de forma direta, indireta ou por alinhamento. Nas conclusões os autores ressaltam a importância da precisão

---

<sup>33</sup><http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus>

<sup>34</sup><http://www.governoeletronico.gov.br/acoes-e-projetos/e-ping-padres-de-interoperabilidade/vcge>

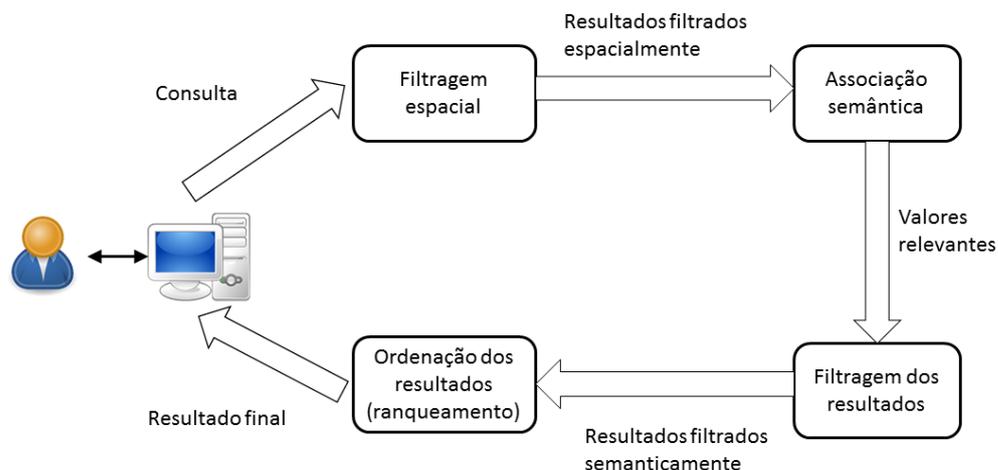


Figura 2.18: Processo de descoberta de informação.  
 Fonte: adaptado de [13].

das anotações semânticas e da especificidade da ontologia como fatores que influenciam a assertividade na recuperação dos dados.

Gimenez *et al.* em [18] apresentam uma proposta de integração semântica para a INDE-BR usando geo-ontologias. Segundo os autores, trabalhos anteriores adotaram dois modelos de incorporação da semântica em IDEs. Um seria denominado modelo puro que propõe um *Semantic Spatial Data Infrastructure* (SSDI). Num SSDI serviços semânticos são incorporados a IDE. Um segundo modelo seria o de uma IDE com um camada semântica.

A proposta de Gimenez *et al.* segue o segundo modelo. Nesta proposta cada provedor de informações geográficas elaboraria a sua geo-ontologia de domínio baseada na Especificação Técnica para Aquisição de Dados Geoespaciais Vetoriais (ET-EDGV)<sup>35</sup>. A ET-EDGV consiste numa ampla especificação técnica e modelagem conceitual para classificação dos dados vetoriais da INDE-BR. A proposta seria composta de três etapas: a preparação semântica das informações, a busca semântica e a integração semântica.

Na etapa de preparação ocorreria a revisão e atualização das ontologias de domínio, o mapeamento dos metadados para estas ontologias e a criação de um catálogo semântico das geoinformações catalogadas. Na etapa de busca semântica, com base nas geo-ontologias, poderiam ser realizadas consultas e descobertas de geoinformações. O

<sup>35</sup><http://www.concar.ibge.gov.br/detalheDocumentos.aspx?cod=94>

resultado desta etapa seria uma lista semanticamente ordenada de geoinformações. Por fim, a etapa de integração possibilitaria a composição de uma geoinformação integrada, considerando-se as restrições temporais e geocartográficas.

Por fim, Camboim em [8], ao estruturar sua proposta de integração entre a INDE-BR e a INDA, passou pela necessidade de criação de ontologias que tomaram como base a ET-EDGV. Para isso traduziu parte de uma modelagem de classes geográficas OMT-G relacionadas a níveis administrativos e territoriais em linguagem OWL.

# Capítulo 3

## O modelo proposto

### 3.1 Levantamento de requisitos

Para a etapa de levantamento de requisitos foram considerados como fontes os experimentos, mecanismos de busca e geocatálogos apresentados no Capítulo 2, em especial o geocatálogo semântico da metodologia SERVUS, as necessidades da área de auditoria ambiental, no que diz respeito a ferramentas de auxílio a auditoria, e a infraestrutura tecnológica atual do TCU e suas restrições tecnológicas.

As principais informações e recursos geográficos disponibilizados pelos geosserviços são mapas temáticos, mapas físicos, dados georreferenciados e metadados, que detalham os recursos e indicam a sua localização na Internet. Com vistas a possibilitar o acesso a esses geosserviços foi definido o seguinte **requisito informacional**:

**Requisito 1:** O sistema deverá proporcionar acesso a mapas temáticos e físicos em formato de imagens(jpeg,gif,tiff), assim como permitir o *download* de dados georreferenciados em formato XML/GML e possibilitar a recuperação de metadados dos recursos, dentre eles, *hiperlinks* para acesso a visualizadores, quando existentes.

Os geocatálogos estudados têm em comum um modelo de busca que se propõe a atender três perguntas básicas: “O quê”, “Onde” e “Quando”. Um cenário típico de busca é expresso por uma breve descrição do item procurado, algumas palavras-chaves, uma delimitação territorial, ou espacial, e uma faixa de tempo. Um exemplo de cenário de busca poderia ser o de recuperação de recursos geográficos relacionados a reservas florestais, localizadas entre quatro coordenadas geográficas, cujos dados refiram-se ao período de 2011 a 2014.

Outro ponto em comum, bastante presente no experimento de Liberman, no geocatálogo da metodologia SERVUS e no geocatálogo de Trento, de forma bastante expressiva, é a utilização da semântica e de ontologias em auxílio ao procedimento de busca. Em geral,

os procedimentos de busca auxiliados por ontologias são complementados pela utilização de um índice de relevância para fins de ranqueamento.

No que se refere a expectativa da equipe de auditoria, espera-se que o produto atue com uma ferramenta de trabalho, que facilite o acesso a recursos geográficos que possam servir de insumo para fiscalizações.

Considerando estas características foram propostos os seguintes **requisitos funcionais**:

**Requisito 2:** O sistema deverá utilizar parâmetros de pesquisa que possibilitem a busca pelos conceitos: “O quê”, “Onde” e “Quando”.

**Requisito 3:** O sistema deverá permitir a pesquisa por palavras-chave, descrição, classificação semântica (Ontologias).

**Requisito 4:** O sistema deverá efetuar o ranqueamento avaliando o nível de proximidade semântica entre os parâmetros de busca e os metadados dos recursos .

**Requisito 5:** O sistema deverá atuar como ferramenta de trabalho no auxílio a auditorias .

Nielsen em [28] estabelece algumas heurísticas que servem de parâmetro para classificação de aspectos de usabilidade da aplicação. Uma delas define um tempo de resposta de até 10 segundos como aceitável pelo usuário, ou seja, um tempo padrão de *feedback* do sistema suficiente para manter a atenção do usuário.

As expectativas da equipe de auditoria são de uma boa usabilidade e operações simples para a localização de recursos geográficos. Por outro lado, as expectativas das equipes de tecnologia são de facilidade de manutenção, rápida recuperação de falhas e flexibilidade na implantação.

Com vistas ao exposto, foram definidos os seguintes **requisitos de qualidade**:

**Requisito 6 :** Performance - o protótipo deve retornar buscas simples (até três termos ou palavras-chaves) em até 10 segundos e buscas complexas (mais de três termos e com classificações por ontologias) em até 25 segundos.

**Requisito 7:** Usabilidade - facilidade de operação e localização dos recursos geográficos de maior relevância na primeira página da listagem. Fonte: equipe de auditoria.

**Requisito 8:** Tecnologia - aplicação Web, acessível via navegador, com flexibilidade de implantação em ambiente de nuvem. Fonte: área de tecnologia e desenvolvimento.

### 3.1.1 Planejamento de avaliação e validação dos resultados da pesquisa

Para avaliação dos objetivos gerais da pesquisa e do produto geocatálogo foi utilizado o método *Goal Question Metric* (GQM). Basile *et al* em [7] apresentam esta abordagem de avaliação que se baseia no estabelecimento de medições com foco em objetivos, que podem ser aplicadas ao ciclo de vida de um produto, processo ou recursos. O método propõe o estabelecimento de uma estrutura *top down* de objetivos, questões e métricas. Os objetivos caracterizam-se por um objeto de análise, um ponto de vista, um propósito, um foco e um ambiente. A ideia geral é que a resolução das questões, mediante obtenção das métricas, defina o grau de atingimento dos objetivos. O planejamento para avaliação do modelo proposto é apresentado na Figura 3.1.

## 3.2 Projeto e construção do geocatálogo

Para linguagem de desenvolvimento da camada de apresentação foi selecionado HTML5 e javascript, por se tratarem de linguagens padrões na Internet, de fácil prototipação e com grande gama de componentes para atendimento dos requisitos propostos. Para a camada de negócio foi selecionada a linguagem Java, por se tratar da linguagem adotada pela equipe de tecnologia da instituição e atender aos requisitos de aplicação web.

Para ambiente de desenvolvimento foram selecionados o Eclipse e controle de versão de código no GitHub, por se tratarem de ferramentas gratuitas e comporem o ambiente padrão da instituição.

O banco de dados NoSql foi escolhido em virtude de requisitos de desempenho. Um banco orientado a grafos foi escolhido por melhor se adequar aos conceitos de rede de recursos e hierarquia de ontologias, além de proporcionar uma maior riqueza e flexibilidade no estabelecimento de relacionamentos. O banco Neo4J foi escolhido por contar com uma versão para a comunidade, permitir a instalação em ambiente de nuvem via GrapheneDB e possuir interface padrão REST.

Quanto às ontologias classificadoras de busca foram selecionadas: o Vocabulário do Controle Externo (VCE), por se tratar da ferramenta institucional para auxílio na produção de material oficial, o Vocabulário Controlado do Governo Eletrônico (VCGE), por se tratar da proposta de classificação do governo eletrônico e o Especificação Técnica para Aquisição de Dados Geoespaciais Vetoriais (ET-EDGV), por se tratar de proposta de padronização da INDE-BR.

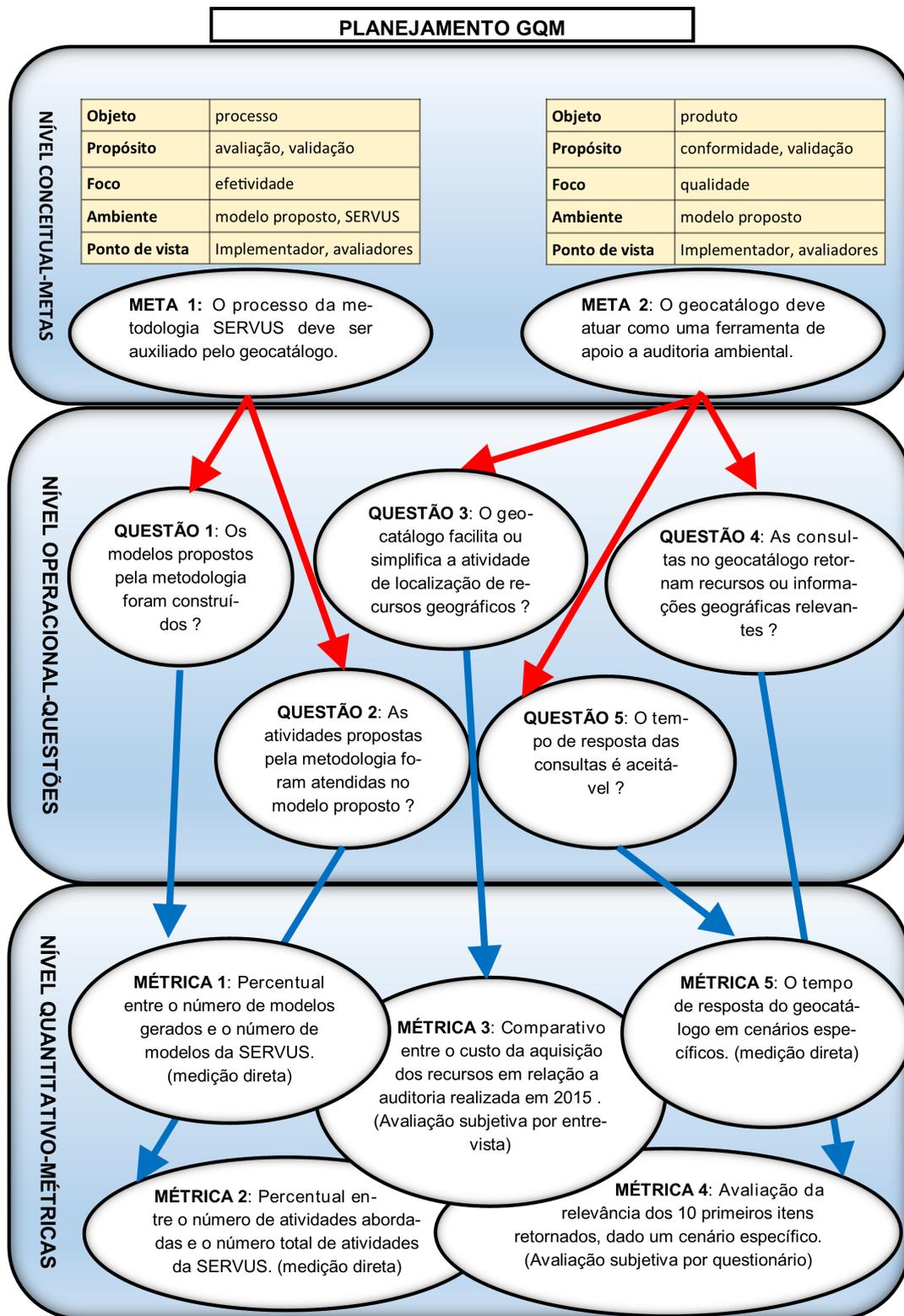


Figura 3.1: Planejamento pelo método GQM (*Goal, Question, Metric*).

Foram selecionados como fonte de dados os geosserviços dos servidores oficiais da INDE-BR e do IBGE, em virtude da grande quantidade recursos disponíveis, e os geosserviços do IBAMA, em virtude de afinidade com o tema meio ambiente.

### 3.2.1 A arquitetura

A arquitetura escolhida para implementação do protótipo do geocatálogo semântico foi a de uma aplicação web de três camadas. A camada de apresentação foi desenvolvida em HTML5 utilizando as bibliotecas javascript Bootstrap 3.3.5<sup>1</sup>, JQuery 2.1.0<sup>2</sup>, GoogleMaps 3.0<sup>3</sup> e o componente bootstrap-slider.js<sup>4</sup>. A camada de negócio foi implementada em Java versão 1.7 e a aplicação empacotada no formato de *web archive - war*. As requisições da camada de apresentação são feitas por http e acessam um servidor padrão REST<sup>5</sup> que retorna documentos no formato JSON<sup>6</sup>. A camada de negócio foi hospedada num serviço de nuvem EC2 da Amazon<sup>7</sup>. A camada de persistência foi implementada através de um banco de dados NoSQL de grafos, o Neo4J<sup>8</sup> 2.3.1. A hospedagem da camada de persistência foi realizada no sítio GRAPHENEDB<sup>9</sup> que implementa os bancos Neo4J num ambiente de nuvem. A comunicação entre camada de negócio e camada de persistência também ocorre por meio de interfaces REST e no formato JSON. A Figura 3.2 representa a arquitetura implementada.

#### 3.2.1.1 A linguagem Cypher

Cypher é uma linguagem de consulta a banco de dados orientado a grafos, tem como um de seus pontos fortes a facilidade de expressar os modelos de grafos, segundo Robinson *et al.* em [33], ela foi projetada para ser facilmente compreendida por desenvolvedores, profissionais de banco de dados e *stakeholders*. A forma das expressões lembram diretamente diagramas de grafos. A Figura 3.3 representa um diagrama de grafo básico e a expressão em Cypher representando este diagrama.

---

<sup>1</sup><http://getbootstrap.com/>

<sup>2</sup><https://jquery.com/>

<sup>3</sup><https://developers.google.com/maps/>

<sup>4</sup><https://github.com/seiyria/bootstrap-slider>

<sup>5</sup><https://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>

<sup>6</sup><http://www.json.org/>

<sup>7</sup><https://aws.amazon.com/pt/ec2/>

<sup>8</sup><http://neo4j.com/>

<sup>9</sup><http://www.graphenedb.com/>

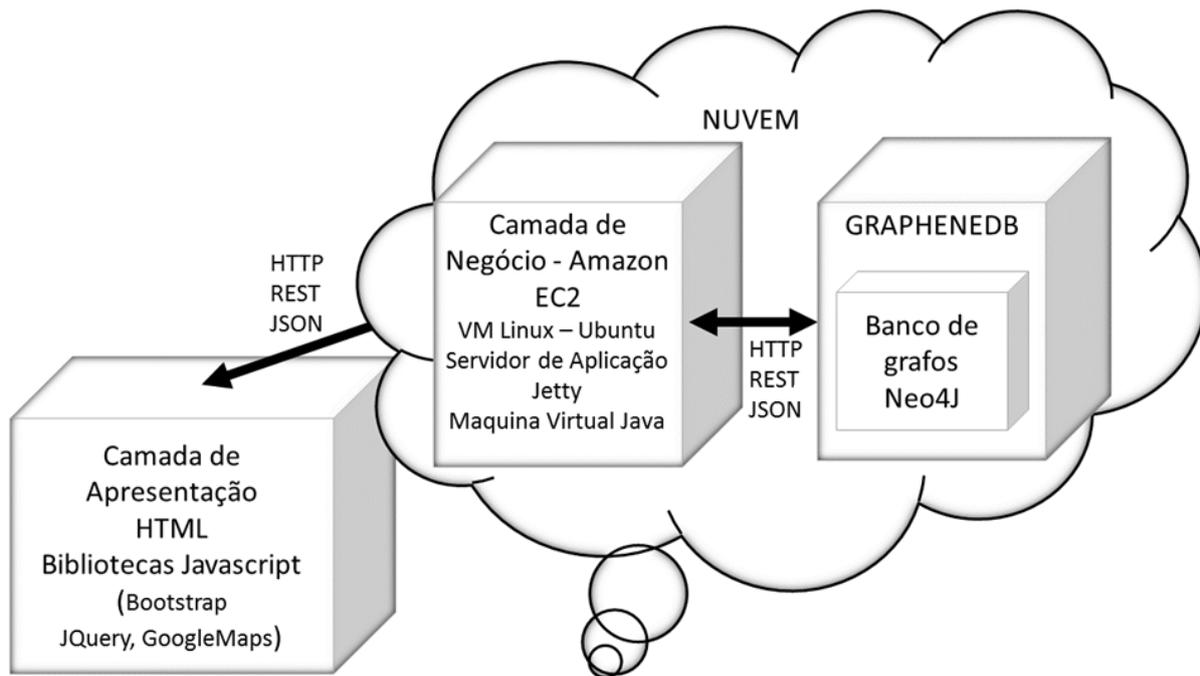
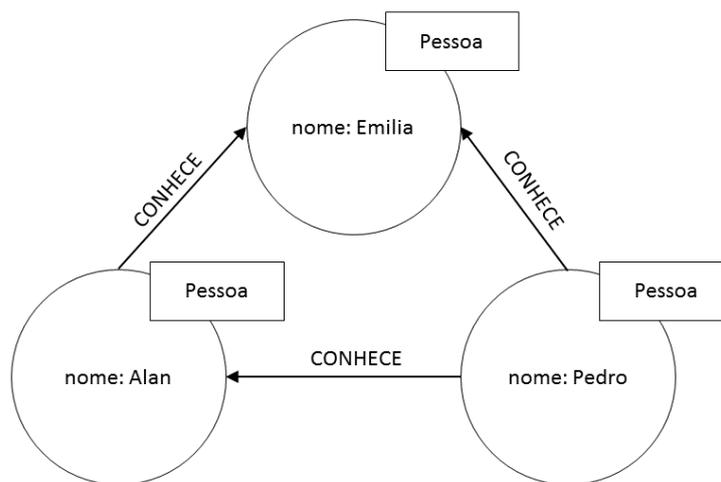


Figura 3.2: Arquitetura de implementação.



`(emilia:Pessoa {nome='Emília'})<-[CONHECE]-(pedro:Pessoa {nome='Pedro'})-[CONHECE]->(alan:Pessoa {nome='Alan'})-[CONHECE]->(emilia)`

Figura 3.3: Uma expressão básica na linguagem Cypher.  
Fonte: adaptado de [33].

### 3.2.2 O modelo de dados

O modelo de dados segue os padrões de modelagem de grafos. Segundo Robinson *et al.* em [33], “grafo é apenas uma coleção de vértices e bordas ou, em uma linguagem menos intimidadora, um conjunto de nós e os relacionamentos que os conectam. Grafos representam entidades como nós e as maneiras pelas quais estas entidades se relacionam com o mundo como relacionamentos”.

Os modelos utilizados são grafos de propriedades rotuladas compostos por *nodes* (nós), relacionamentos, rótulos e propriedades. *Nodes* e relacionamentos podem possuir propriedades. Propriedades são estruturas do tipo chave-valor, sendo a chave do tipo *String* e o valor qualquer tipo primitivo ou uma coleção. Os relacionamentos são binários, ou seja, ligam um nó inicial a um nó final, possuem um nome e são direcionais.

Os modelos de dados baseados em grafos são comumente utilizados quando os relacionamentos possuem um valor de negócio preponderante em relação às entidades. Numa resolução de um problema de busca de recursos a informação mais importante está localizada nos relacionamentos entre recursos requisitados e recursos ofertados.

A Figura 3.4 ilustra o modelo de dados para recursos. Possui uma entidade do metamodelo denominada Recurso que tem relacionamentos do tipo `[:INSTANCIA]` com `(n:Recurso:Requisitado)` e `(m:Recurso:Ofertado)`. O relacionamento nomeado `[:LOCALIZA]` liga recursos ofertados e requisitados. Os nós `(m:Recurso:Ofertado)` possuem relação do tipo `[:CONNECTA]` com `(o:RecursoSemantico:Ofertado)`. Os atributos do nó `(n:Recurso:Requisitado)` representam os parâmetros de consulta utilizados, já os atributos do nó `(o:RecursoSemantico:Ofertado)` representam os atributos dos metadados do recurso no padrão *Dublin Core* (DC)<sup>10</sup> e os atributos do nó `(n:Recurso:Requisitado)` representam o recurso geográfico, suas *Universal Resource Locator* (URL)s de acesso e origem.

A Figura 3.5 ilustra uma rede de recursos no modelo proposto, visualizada pela interface de navegação do banco Neo4J.

Para a modelagem das Ontologias optou-se por um modelo simplificado capaz de expressar estruturas em árvore por meio de auto relacionamentos. Os nós `(n:Ontologia)` se conectam por meio de relacionamentos `[:EH_UM]`, a fim de refletir estruturas do tipo **filhos -> pai**, e `[:TERMO_ESPECIFICO_DE]`, para refletir conexões do tipo **termo específico -> termo geral**. A Figura 3.6 ilustra o modelo de dados proposto para as Ontologias.

---

<sup>10</sup><http://dublincore.org/documents/dces/>

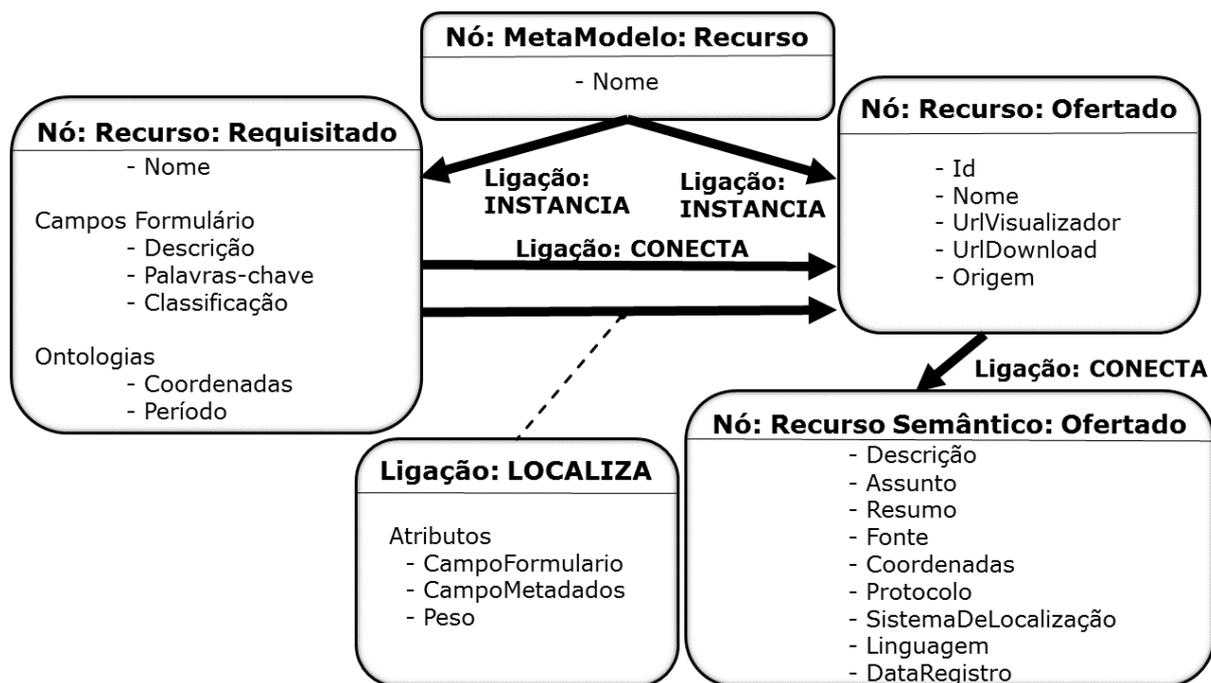


Figura 3.4: Modelo de dados para Recursos.

### 3.2.2.1 Carregando os recursos: *harvesting*

Segundo Marienfeld *et al.* em [26], “um desafio para portais de dados abertos é a agregação de metadados a partir de vários catálogos de dados (em diferentes níveis administrativos ou a partir de diferentes campos de aplicação), também conhecida como *metadata harvesting*”. Portanto o contexto que utilizaremos para *metadata harvesting* é o de agregação e coleta de metadados oriundos de catálogos de serviço disponibilizados numa rede de geosserviços na Internet.

Para a construção do protótipo de geocatálogo semântico utilizamos um processo de *harvesting* manual apoiado por *scripts*, conforme ilustrado na Figura 3.7 e descrito em passos a seguir:

1. Recuperação dos arquivos XML de metadados :

Os arquivos XML contendo os metadados são recuperados através de requisições http aos serviços CSW e WFS por meio de comandos *GetRecords* e *GetCapabilities*, respectivamente.

Exemplo de conexão a servidor CSW:

```
http://www.metadados.inde.gov.br/geonetwork/srv/por/csw?
request=GetRecords&service=CSW&version=2.0.2&
typeName=csw:Record&constraintLanguage=CQL_TEXT&
namespace=xmlns(csw=http://www.opengis.net/cat/csw/2.0.2),
```

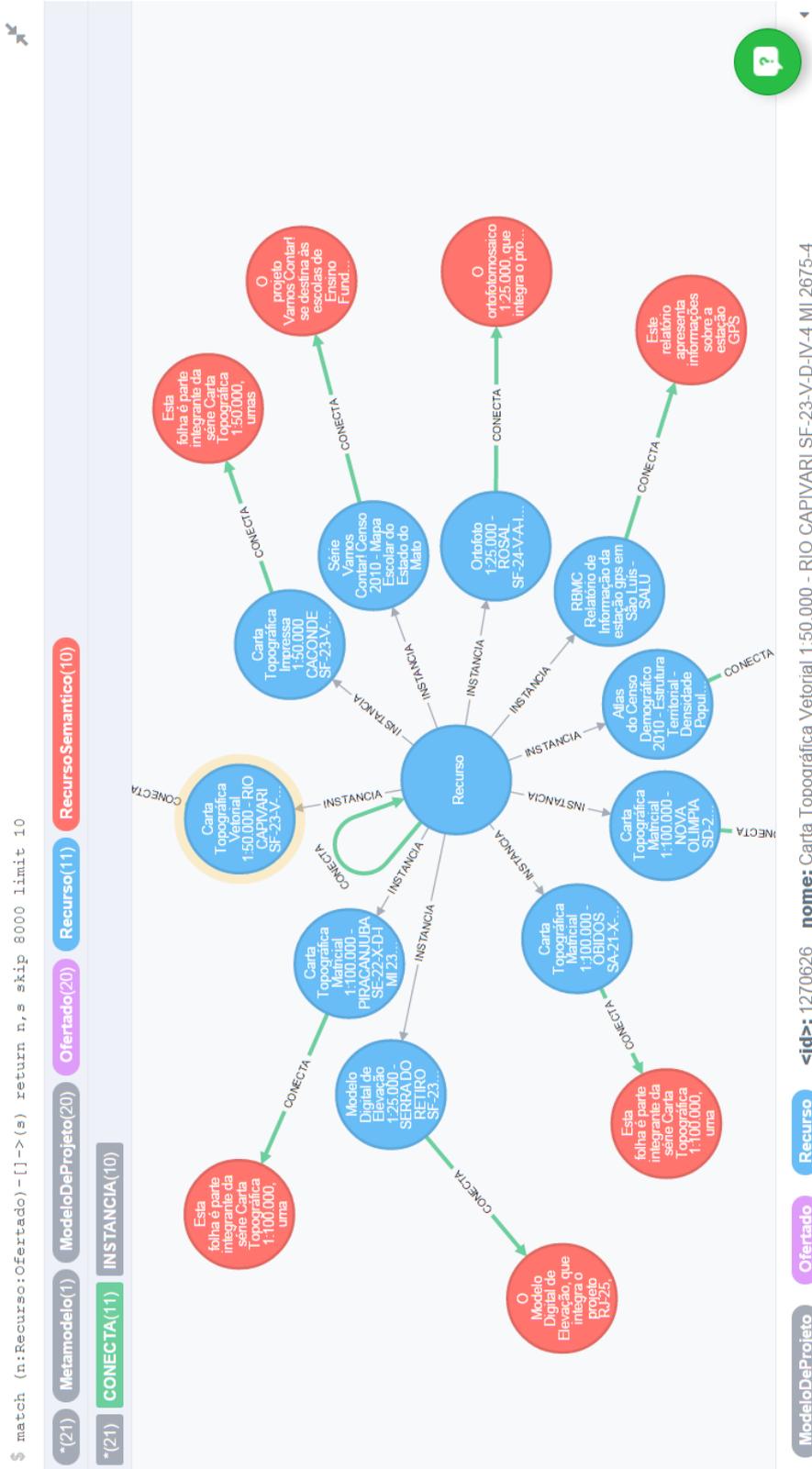


Figura 3.5: Rede de Recursos/Recursos semânticos.

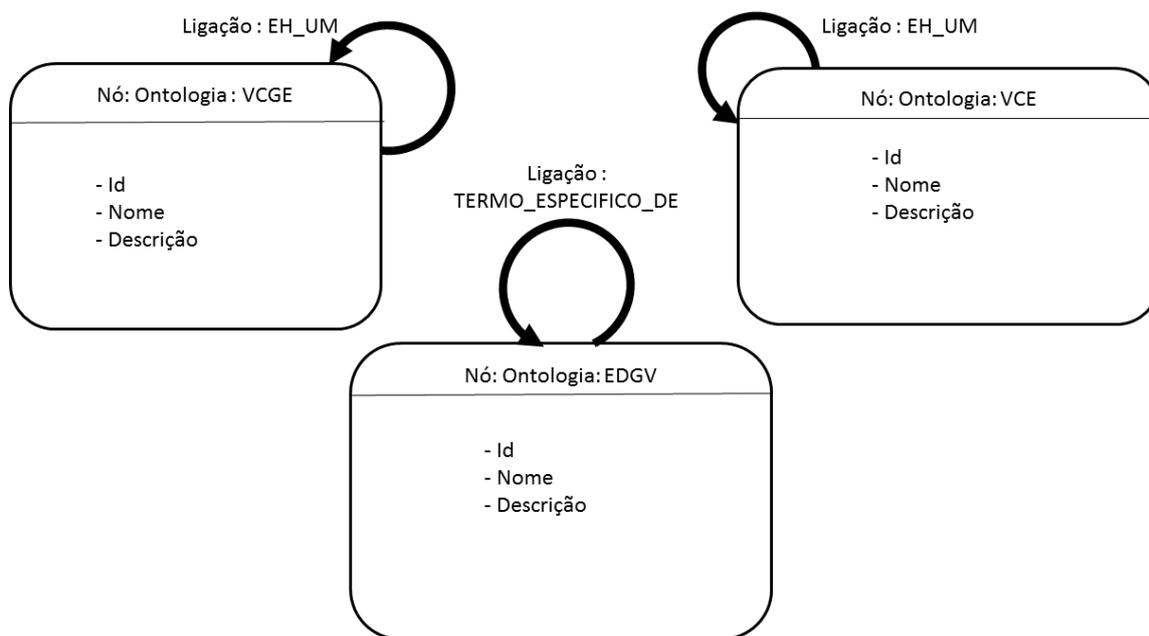


Figura 3.6: Modelo de dados para Ontologias.

```

xmlns(gmd=http://www.isotc211.org/2005/gmd)&
constraint_language_version=1.1.0&resultType=results&
startPosition=4675&maxRecords=5000&ElementSetName=full
  
```

Exemplo de conexão a servidor WFS:

```

http://www.geoservicos.ibge.gov.br/geoserver/wfs?
request=GetCapabilities&service=WFS&
acceptFormats=application/xml
  
```

## 2. Conversão dos arquivos XML para CSV:

São gerados arquivos em formato CSV com informações extraídas dos arquivos XML de metadados. Para extração é utilizada o processador XSLT Apache xalan 2.7.1<sup>11</sup> e arquivos no formato XSL.

Exemplo de comando para execução do *parser*:

```

java -cp "xalan-2.7.1.jar;serializer-2.7.1.jar;
xml-apis-xerces-2.7.1.jar" org.apache.xalan.xslt.Process
-IN inde_1_4502.xml -XSL foo.xsl -OUT foo.out
  
```

Exemplo de arquivo XSL que realiza o *parser* pode se visto no Anexo I

<sup>11</sup><https://xml.apache.org/xalan-j/>

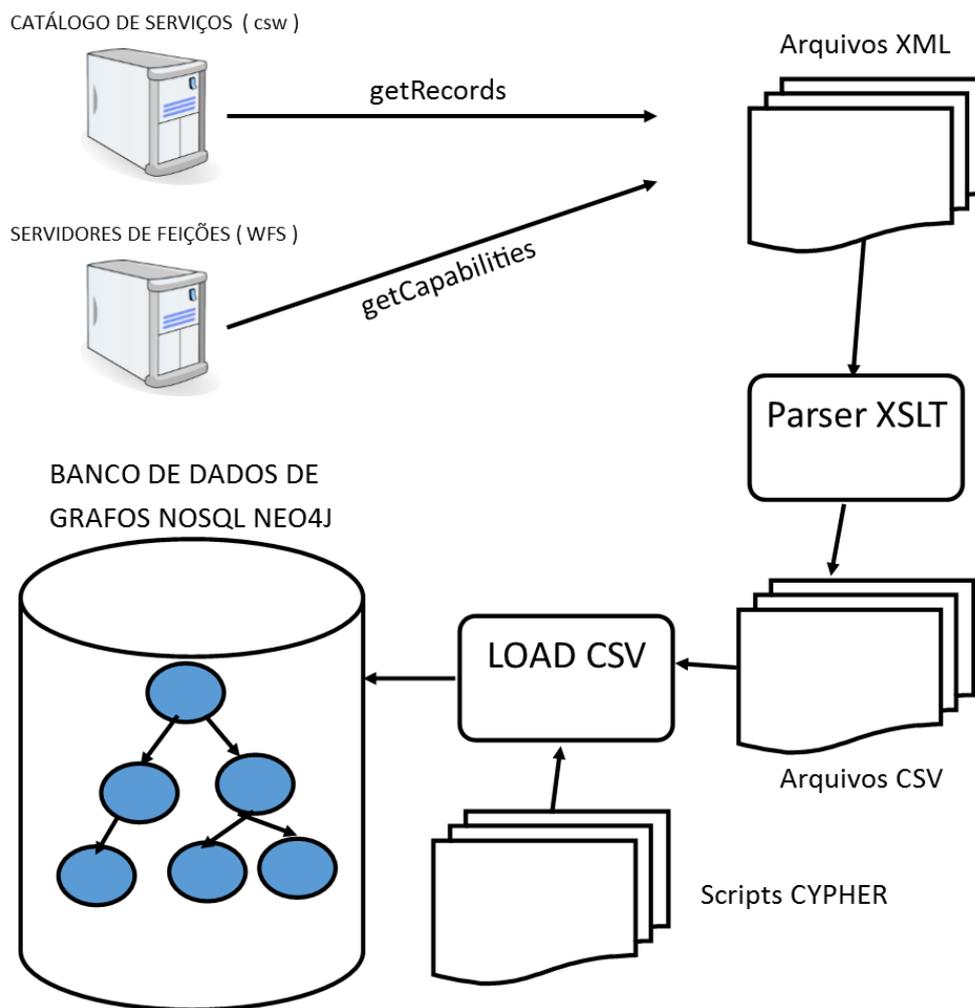


Figura 3.7: Processo de *harvesting*.

3. Tratamento dos arquivos CSV para extração de caracteres inválidos.
4. Execução de scripts na linguagem Cypher para carga do banco de grafos.

A carga no banco de grafos é realizada através do comando LOAD CSV da linguagem, conforme exemplo abaixo:

```

1 // Cria recursos ofertados WFS a partir do CSV
2 MATCH (r:Metamodelo {nome:'Recurso'})
3 LOAD CSV WITH HEADERS FROM "file:d:/desv/wfs_ibge.csv" AS
  csvLine
4 // Cria recurso ofertado
5 CREATE (r)-[:INSTANCIA]->(
  ro:Recurso:ModeloDeProjeto:Ofertado
6 {idRecurso: csvLine.ID, nome: csvLine.title, link:
  csvLine.link,

```

```

7   urlDownload: csvLine.download, protocolo: csvLine.protocol,
8   fonte: csvLine.source})
9   // Cria recurso semantico
10  CREATE (ro)-[:CONECTA]->(
      rs:RecursoSemantico:ModeloDeProjeto:Ofertado
11  {linguagem: csvLine.language,
12  resumo: csvLine.abstract, source: csvLine.source,
13  coordenadaInferior: csvLine.lowerCorner,
14  coordenadaSuperior: csvLine.upperCorner,
15  protocolo: csvLine.protocol, palavrasChave:
      csvLine.subject});

```

Na linha 2 é selecionado um nó já existente com o rótulo (*label*) Metamodelo e com o atributo 'nome' igual a 'Recurso'. A linha 3 solicita a carga do arquivo CSV localizado em disco, já com a linha de cabeçalho, e as linhas do arquivo CSV serão identificadas pelo apelido csvLine. Da linha 5 à linha 8 estão os comandos para criação do recurso ofertado e do relacionamento do tipo **[:INSTANCIA]** com o nó Recurso do metamodelo. Da linha 10 à linha 15 são criados o recurso semântico e o relacionamento do tipo **[:CONECTA]** com o recurso ofertado.

### 3.2.2.2 Carregando as Ontologias

As ontologias selecionadas como classificadores para a busca semântica foram: Vocabulário Controlado do Governo Eletrônico (VCGE), Vocabulário do Controle Externo (VCE), Especificação Técnica para Aquisição de Dados Geoespaciais Vetoriais (ET-EDGV).

Para carga do VCGE foi utilizado arquivo vcge-2-0-3.csv disponível no sítio do governo eletrônico<sup>12</sup> e utilizados comandos na linguagem cypher conforme abaixo:

```

1   // Carrega ontologia VCGE
2   CREATE (a:Ontologia:VCGE{id: '0', nome: 'VCGE'
3   , descricao: 'Vocabulario Controlado do Governo Eletronico'});
4   // Cria os node VCGE
5   LOAD CSV WITH HEADERS FROM "file:d:/desv/vcge-2-0-3.csv" AS
      csvLine
6   CREATE (a:Ontologia:VCGE{id: csvLine.CODIGO, nome:
      csvLine.TERMO,
7   descricao: csvLine.NOTAS});
8   // Cria os relacionamentos
9   LOAD CSV WITH HEADERS FROM "file:d:/desv/vcge-2-0-3.csv" AS
      csvLine
10  MATCH (a:Ontologia:VCGE{id: csvLine.CODIGO}),

```

<sup>12</sup><http://www.governoeletronico.gov.br/acoes-e-projetos/e-ping-padres-de-interoperabilidade/vcge>

```

11 (b:Ontologia:VCGE{id: csvLine.PAI_CODIGO})
12 CREATE (a)-[:EH_UM]->(b);

```

A carga da ontologia VCE foi realizada a partir de arquivo XML, exibido no Anexo II, realizado *parser* para arquivos CSV e carregados para o banco com comandos cypher de forma similar a ontologia VCGE. Para o protótipo foram carregados apenas os conceitos referentes às áreas de agricultura, pecuária e meio-ambiente.

Para ontologia ET-EDGV modelos de dados no formato OMT-G foram traduzidos para a linguagem OWL utilizando-se a ferramenta Protege 5.0 <sup>13</sup>. A tradução foi realizada para as categorias de informação de hidrografia e de vegetação e somente para os relacionamentos de herança e composição. Em seguida foi realizado *parser* para CSV e carregado para o banco de dados via comandos Cypher LOAD CSV.

### 3.2.3 Camada de negócio

A camada de negócio é responsável por receber as requisições REST da camada de apresentação, montar as consultas, submeter requisições REST para o banco de dados e enviar as respostas no formato JSON.

A Figura 3.8 apresenta a estrutura de busca do catálogo.

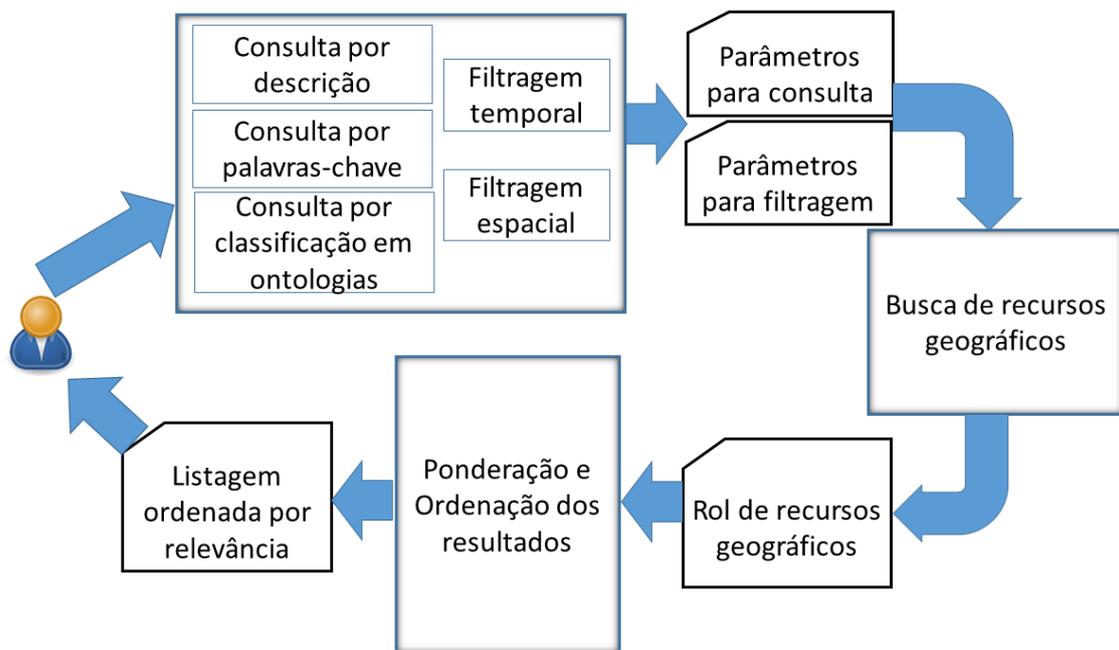


Figura 3.8: Estrutura de busca do catálogo.

<sup>13</sup><http://protege.stanford.edu/>

### 3.2.3.1 O modelo de busca

As consultas podem ser realizadas por descrição (texto livre), palavras-chave e por termos das ontologias. Para cada um dos itens anteriores as frases ou termos são divididos em *tokens* que serão os termos utilizados na busca. Cada termo de busca será pesquisado nos campos dos nós de recurso e de recurso semântico. Podem ser realizadas filtragens por coordenadas geográficas (filtro espacial) e por ano (temporal). Para cada par, composto pelo termo e o campo pesquisado, é atribuído um peso que será a base para o procedimento de ranqueamento. A Figura 3.9 apresenta um modelo que exemplifica o processo de ponderação e ranqueamento.

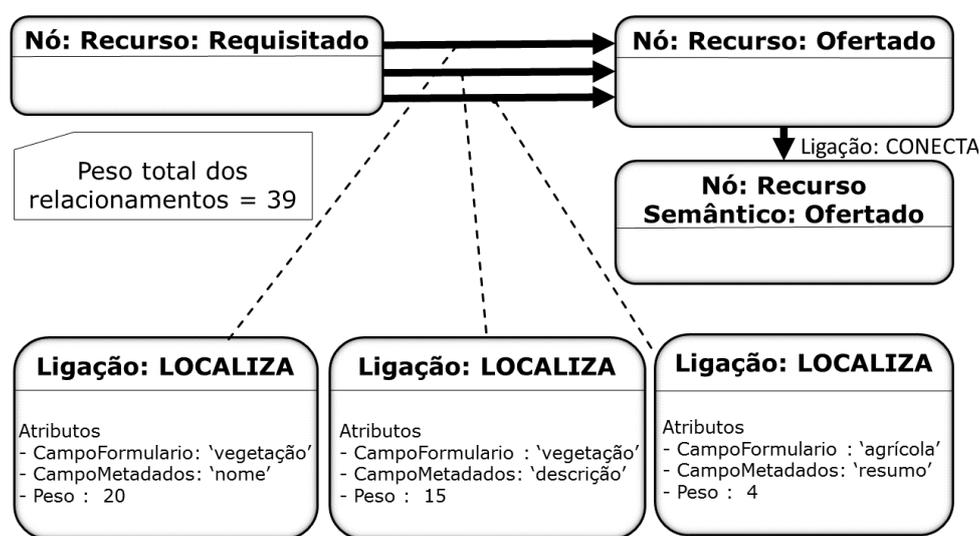


Figura 3.9: Modelo de relacionamento Recurso requisitado X Recurso ofertado.

### 3.2.3.2 Ranqueamento e índice de relevância

O resultado efetivo de uma busca depende da qualidade dos metadados e do valor de negócio atribuído a cada campo. Para lidar com este aspecto foi elaborado um esquema de ponderação dos campos de pesquisa e dos campos de metadados que afetam diretamente o cálculo do peso total de um relacionamento entre recursos requisitados e ofertados, que obedece a Fórmula 3.1 .

Os pesos dos campos de pesquisa e de metadados podem variar de 0 a 10 e PT (Peso Total dos relacionamentos), N (Número de relacionamentos), PP (Peso do campo de pesquisa) e PM (Peso do campo de metadados).

$$PT = \sum_{k=1}^N (PP * PM) \quad (3.1)$$

Para melhor expressar a relevância dos relacionamentos foi criado um índice que varia 0 a 1 que permitiria comparar consultas que geram quantidades diferentes de relacionamentos e com pesos diferentes para os campos envolvidos. O índice é uma relação entre o peso total dos relacionamentos e somatória das pontuações dos relacionamentos caso os mesmos fossem localizados em todos os campos dos metadados. As Tabela 3.1, 3.2 e a Fórmula 3.2 demonstram o cálculo do índice de relevância.

Na Fórmula 3.2 temos IR (Índice de relevância), PT (Peso total dos relacionamentos), N (número de relacionamentos), M (Número total de campos de metadados), PM (Peso do campo metadado) e PP (Peso do campo de pesquisa).

$$IR = PT / \sum_{k=1}^N ((\sum_{i=1}^M PM) * PP) \quad (3.2)$$

Tabela 3.1: Tabela de pesos.

<b>Campo termo pesquisa</b>	<b>Peso termo pesquisa</b>	<b>Campo metadado</b>	<b>Peso metadado</b>
Descricao	8	Nome	8
Palavra-chave	5	Descrição	6
Ontologia VCE (termo raiz)	6	Assunto	5
Ontologia VCE (termo pai ou filho)	4	Resumo	5
Ontologia VCGE (termo raiz)	6	Fonte	3
Ontologia VCGE (termo pai ou filho)	4		
Ontologia EDGV (termo raiz)	8	Total	27
Ontologia EDGV (termo pai ou filho)	6		

Tabela 3.2: Cálculo do índice de relevância.

<b>Termos de pesquisa</b>	<b>Campo com termo localizado</b>	<b>Peso</b>	<b>Peso Máximo</b>
Descrição : 'vegetação' - (peso 8)	Nome - (peso 8)	64	216
Ontologia VCE (termo raiz): 'vegetação' - (peso 6)	Descrição - (peso 6)	36	162
Palavra-chave : 'agrícola' - (peso 5)	Resumo - (peso 5)	25	135
	Peso Total	125	513
	Índice de relevância calculado (125/513)	0,24	

### 3.2.3.3 Recuperação dos recursos

Os registros dos recursos são recuperados mediante requisições REST para o banco de dados remoto compostas pelos comandos na linguagem Cypher. Os principais tipos de interação são: consulta a ontologias, criação de relacionamentos entre os recursos e ranqueamento dos recursos.

O exemplo abaixo apresenta uma recuperação da árvore representando a ontologia VCGE.

```
1 // Carrega arvore ontologia VCGE
2 MATCH (n:Ontologia:VCGE)-[ ]->(r:Ontologia:VCGE)
3 RETURN n.id as id ,n.nome as nome, n.descricao as descricao ,
4         r.id, r.nome order by n.nome;
```

O exemplo abaixo apresenta a criação de um relacionamento entre recursos a partir de termo de pesquisa e peso.

```
1 // Cria um relacionamento entre recursos
2 MATCH (n:Recurso:Ofertado)-[ ]->(a:RecursoSemantico)
3 WHERE a.descricao=~ toString('( ?i).*agricultura.*')
4 WITH n
5 MATCH (r:Recurso{codRecurso:'RecursoRequisitado_001'})
6 MERGE
   (r)-[z:LOCALIZA{peso:30, termo:'agricultura', campo:'descricao'}]->n
```

Na linha 2 e 3 são localizados recursos ofertados que tenham relacionamentos com recursos semânticos cujo campo de metadado 'descrição' possua o termo 'agricultura'. Na linha 4 os nós localizados são disponibilizados para o comando seguinte pela cláusula WITH. Na linha 5 é selecionado o recurso requisitado. Na linha 6 é criado o relacionamento entre os recursos com os atributos peso, termo e campo. A cláusula MERGE é utilizada no lugar de CREATE para evitar duplicidades de relacionamentos com os mesmos termos e campos.

Por fim, o exemplo abaixo apresenta a criação do *ranking* de recursos ofertados. No Anexo III é apresentada uma consulta mais completa incluindo aspectos de paginação em banco e filtro espacial.

```
1 // Criacao de um ranking
2 MATCH
   (a:Recurso{codRecurso:'RecursoRequisitado_001'})-[p:LOCALIZA]->
3   (b:Recurso:Ofertado)-[r]->(c:RecursoSemantico) where p.peso >
   0
4 RETURN sum(p.peso) as peso , b.nome, b.urlVisualizador,
   b.urlDwonload,
5   b.origem, id(b)
```

### 3.2.4 Camada de apresentação

As funcionalidades presentes na camada de apresentação são :

1. Pesquisas por campo texto e palavras-chave
2. Pesquisa por classificação em ontologias : Vocabulário do Controle Externo (VCE), Vocabulário Controlado do Governo Eletrônico (VCGE), Especificação Técnica para Aquisição de Dados Geoespaciais Vetoriais (ET-EDGV)
3. Filtragem do resultado por coordenadas espaciais
4. Listagem paginada de recursos geográficos com *links* para visualização do recurso, *download* do recurso e visualização dos metadados
5. Parametrização da consulta pela definição de pesos para os campos de pesquisa e os campos de metadados.

No Anexo V são exibidas as principais telas do sistema. Na Figura V.1 é apresentada seleção de ontologias, na Figura V.2 é exibida a filtragem espacial com apoio de mapa, na Figura V.3 o estabelecimento de pesos para os parâmetros de consulta e campos de metadados e a Figura V.4 a listagem final com os *links* para visualização, *download* e metadados.

# Capítulo 4

## Validação do modelo

### 4.1 Aplicando a metodologia SERVUS com auxílio do geocatálogo

Para validação de uso do geocatálogo semântico em conjunto com as etapas da metodologia SERVUS, utilizaremos como base um problema relacionado a uma auditoria realizada em 2015 pelo Tribunal de Contas da União (TCU), em conjunto com outros onze Entidades Fiscalizadoras Superiores (EFS) pertencentes Organização Latinoamericana e do Caribe de Entidades Fiscalizadoras Superiores (OLACEFS), denominada “Auditoria Coordenada - Áreas Protegidas - América Latina”[30].

Uma das atividades desta auditoria foi o levantamento das áreas de proteção ambiental da América Latina. Nesta capítulo reproduziremos a tarefa de recuperação de levantamento das áreas brasileiras, serão criadas algumas visualizações através de ferramenta QGIS<sup>1</sup> e ressaltaremos um achado de auditoria.

Pela metodologia SERVUS o ponto inicial se dá pela definição do problema:

**Definição do problema:** Como obter os dados necessários, a nível Brasil, para o levantamento da totalização de áreas de proteção ambiental, utilizando como base os geosserviços disponíveis na rede de infraestruturas de dados espaciais nacional, de instituições e órgãos públicos.

#### 4.1.1 Modelo de domínio

O modelo de domínio é apresentado na Figura 4.1 e será composto pelos conceitos associados ao problema e ao tema área proteção ambiental. Estes conceitos serão obtidos pelas classificações segundo as ontologias VCE, VCGE e ET-EDGV. Os relacionamentos

---

<sup>1</sup>Ferramenta SIG livre e de código aberto disponível em <http://www.qgis.org/en/site/>

consistem basicamente de relações instanciação da classe do metamodelo, de herança por tipo e de composição. As duas feições preponderantes são área de proteção ambiental e limites geográficos.

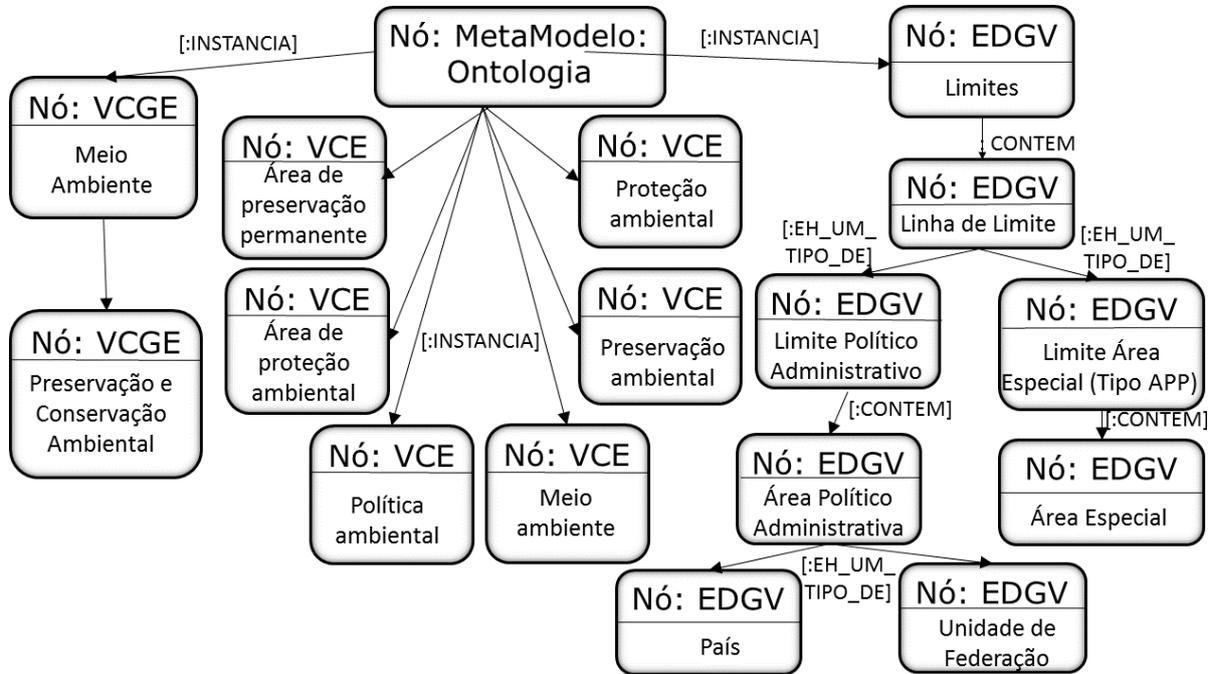


Figura 4.1: Modelo de domínio.

#### 4.1.2 Modelo de Processo/Caso de Uso

Para a produção do modelo de processo/caso de uso, ou modelo de informação, foi elaborado o metamodelo apresentado na Figura 4.2. Os elementos básicos do modelo são os casos de uso, cenários e ações. As conexões entre as ações de caso de uso e o recurso requisitado serão geradas durante a atividade de *rephrasing*.

A metodologia SERVUS propõem duas extensões para os casos de uso que viabilizem a atividade de *rephrasing* [35]. A primeira indica que para cada ação de caso de uso possam ser associados um ou mais tipos e instâncias de recursos, a segunda sugere que as descrições textuais possam ser anotadas com conceitos do modelo de domínio.

Seguem abaixo as definições do caso de uso:

**Processo:** Avaliação de possíveis irregularidades.

**Descrição do processo:** Avaliação de possíveis distorções entre os dados apresentados e a situação de fato.

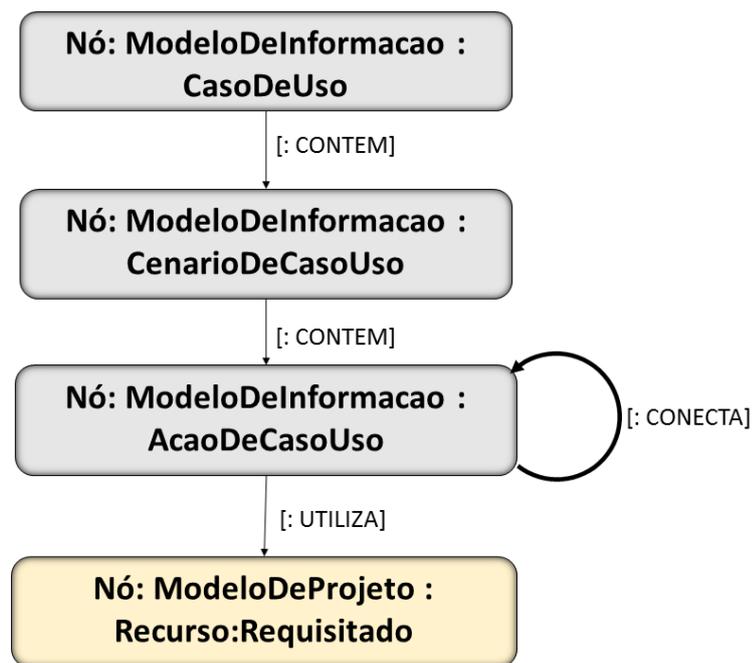


Figura 4.2: Metamodelo de Informação.

**Caso de uso:** Levantamento das áreas protegidas.

**Descrição do caso de uso:** Processo de levantamento das áreas de proteção ambiental no Brasil e sua principais características: localização, área, etc...

**Cenário:** Visualização das áreas de proteção ambiental.

**Descrição cenário:** Recuperação dos dados a serem utilizados na construção de camadas de visualização das informações necessárias para a avaliação.

A Tabela 4.1 apresenta as extensões do cenário de caso de uso.

Tabela 4.1: Extensão do caso uso.

<b>Ação</b>	<b>Descrição recurso</b>	<b>Tipo de serviço</b>	<b>Classificação por ontologias</b>	<b>Palavras-chave</b>
Recuperar limites geográficos do Brasil	Mapa do Brasil com seus estados	WFS	Nenhuma	cartografia
Recuperar limites das áreas de proteção ambiental	Limites de áreas de proteção ambiental	WFS	VCGE: Preservação e Conservação Ambiental, VCE: Proteção ambiental, VCE: Sustentabilidade ambiental	APA

A Figura 4.3 apresenta o modelo de processo/caso de uso (Modelo de informação), resultado do processo de *rephrasing* que se efetiva com a tradução dos requisitos necessários para execução das ações do caso de uso em recursos requisitados.

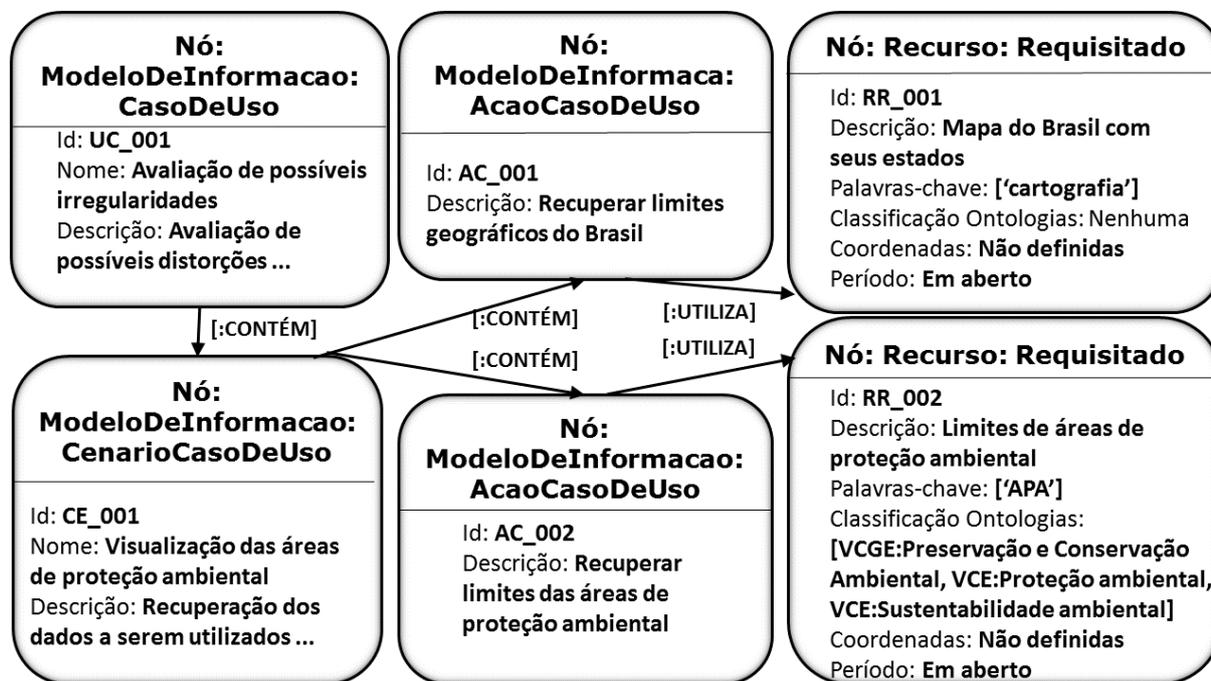


Figura 4.3: Modelo de Informação.

### Associando recursos requisitados a recursos ofertados (*matching*)

Para a atividade de (*matching*) é utilizado como ferramenta o geocatálogo semântico. Os atributos do recurso requisitado são transferidos para os campos do formulário de pesquisa. Após a submissão das pesquisas foram obtidos os resultados e selecionados os recursos ofertados que melhor atenderam aos requisitos. As configurações (pesos) utilizadas estão apresentada no Anexo IV.

Para o recurso RR\_001 foram selecionados os seguintes itens :

1. Grandes Regiões do Brasil

(CGEO:AP\_T\_Grandes\_Regioes): Posição - 2, Índice de relevância - 0,07 , Peso - 108, Tipo - WFS, Fonte - IBGE.

2. Base Cartográfica do Brasil Contínua ao Milionésimo

(CGEO:ANMS2010\_09 \_SCN\_BCIM): Posição - 7, Índice de relevância - 0,05 , Peso - 51, Tipo - WFS, Fonte - IBGE.

Para o recurso RR\_002 foram selecionados os seguintes itens :

1. Unidades de Proteção Integral

(csr:unidade\_protecao\_integral): Posição - 1, Índice de relevância - 0,05 , Tipo - WFS, Peso - 108, Fonte - IBAMA.

2. BC250 Unidade de Proteção Integral (Polígono)

(CCAR:BC250\_Unidade\_Protecao\_Integral\_A): Posição - 3, Índice de relevância - 0,05 , Peso - 108, Tipo - WFS, Fonte - IBGE.

3. Áreas especiais de proteção e conservação da biodiversidade

(CGEO:ANMS2010\_03\_apcb): Posição - 13, Índice de relevância - 0.04 , Peso - 81, Tipo - WFS, Fonte - IBGE.

A Figura 4.4 ilustra o uso do geocatálogo semântico para a seleção dos recursos ofertados para o RR\_002.

The screenshot shows the 'Geocatálogo semântico' interface. At the top, there are navigation buttons for 'VCGE', 'VCE', 'EDGV', 'Mapa', and 'Configurar'. Below this is a search form with a 'Descrição:' field containing 'Limites de áreas de proteção ambiental' and 'Palavras-chave (separadas por vírgula):' field containing 'APA'. There are also 'De (ANO)' and 'Até (ANO)' fields. Below the search form are 'Pesquisar' and 'Limpar' buttons. The search results are displayed in a table with columns for 'Índice', 'Peso', 'Recurso', 'Fonte', and 'Tipo'. The results include indicators of sustainable development, units of integral protection, and biochemical oxygen demand (BOD) in a specific region.

Índice	Peso	Recurso	Fonte	Tipo
0.07	144	Indicadores de Desenvolvimento Sustentável 2015 - Classe de periculosidade ambiental(CGEO:IDS_03_Comerc_de_Agrotoxicos_e_Afins_Total_1_2)	IBGE	WFS
0.05	108	Unidades de Proteção Integral(csr:lim_unidade_protecao_integral_a)	IBAMA	WFS
0.05	108	BC250 Unidade de Proteção Integral (Ponto)(CCAR:BC250_Unidade_Protecao_Integral_P)	IBGE	WFS
0.05	108	BC250 Unidade de Proteção Integral (Polígono)(CCAR:BC250_Unidade_Protecao_Integral_A)	IBGE	WFS
0.05	105	Indicadores de Desenvolvimento Sustentável 2015 - Demanda Bioquímica de Oxigênio (DBO) - Região Sul(CGEO:IDS_14_Demanda_Bioqui_de_Oxig_Regiao_Sul_2_2)	IBGE	WFS

Figura 4.4: *Matching* de recursos utilizando o geocatálogo.

### 4.1.3 Modelo de Projeto

O modelo de projeto é apresentado na Figura 4.5 e Figura 4.6 e expressa a associação entre os recursos ofertados e requisitados resultantes da atividade de *matching*, formando uma rede de recursos.

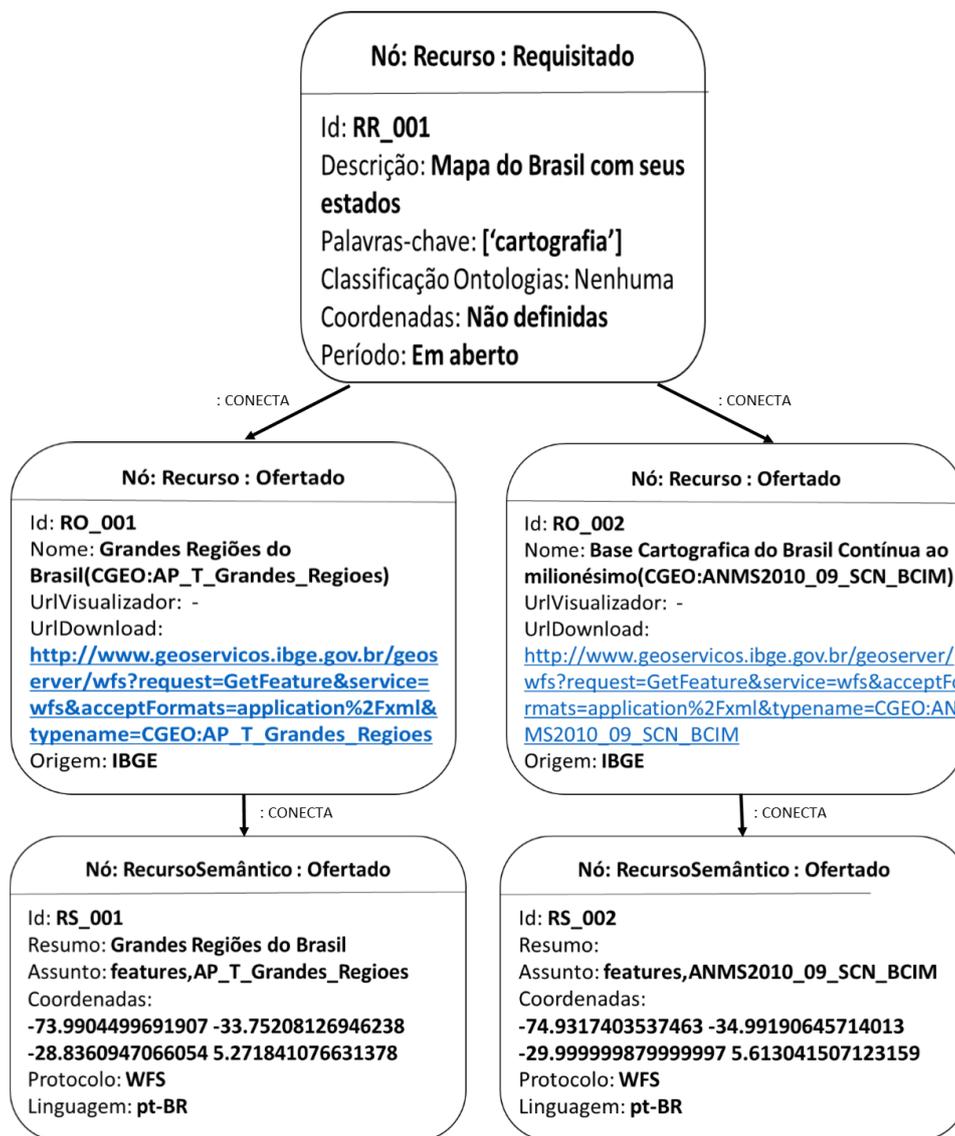


Figura 4.5: Modelo de projeto - Recurso RR\_001.

#### 4.1.3.1 Visualizando os dados

Para visualização dos dados os arquivos XML/GML dos recursos oferecidos, oriundos dos serviços WFS, foram importados para ferramenta QGSI versão 2.12.1 - Lyon com Grass 6.4.3, através da funcionalidade “Adicionar camada vetorial”. Foram criadas cinco camadas, uma para cada recurso, conforme descrito abaixo:

1. RO\_001 - Camada com os limites geográficos do Brasil e suas unidades federativas.
2. RO\_002 - Camada com a base cartográfica do Brasil.

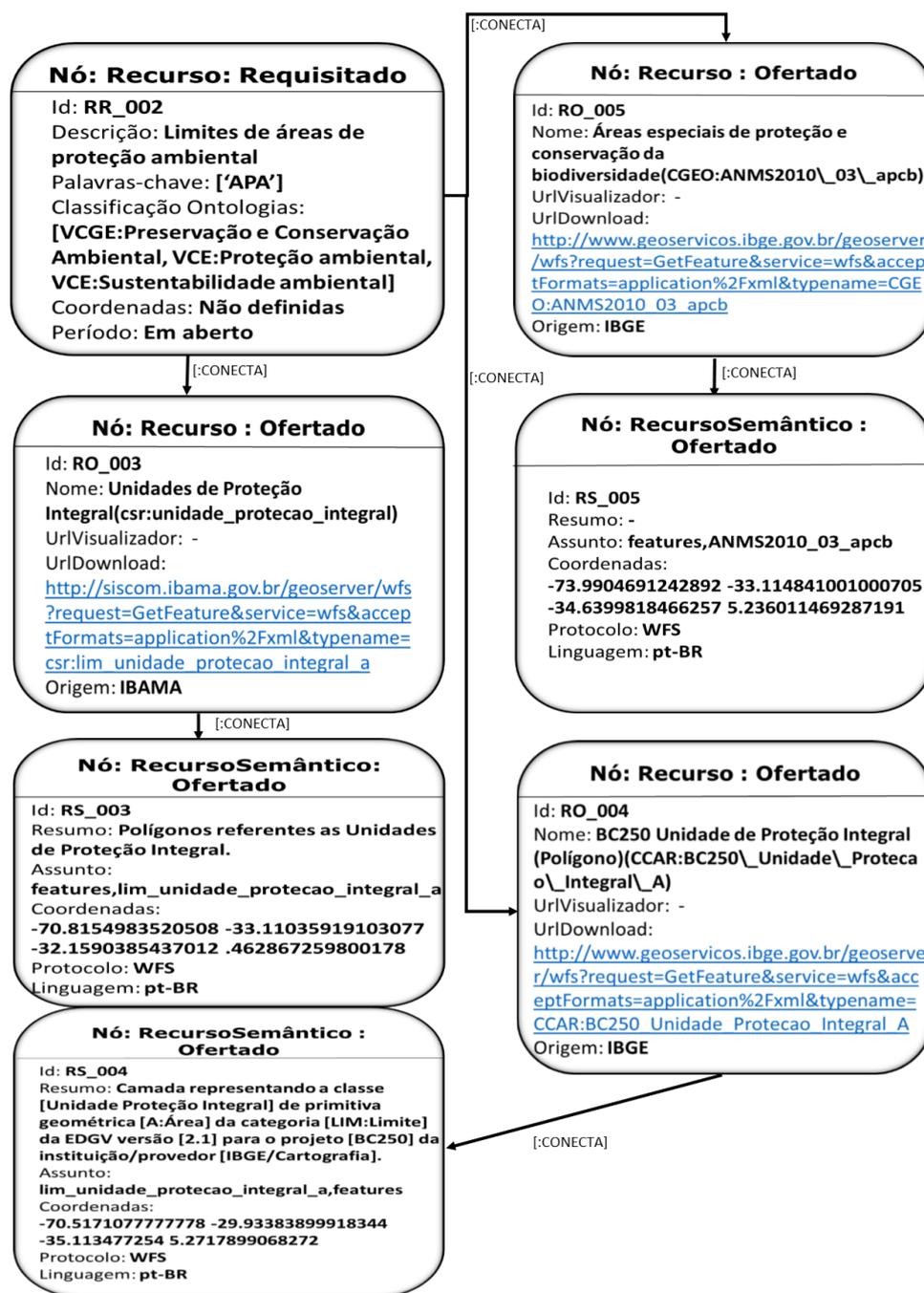


Figura 4.6: Modelo de projeto - Recurso RR\_002.

3. RO\_003 - Camada com os limites da unidades de proteção integral - Fonte IBAMA, contendo 666 áreas delimitadas.
4. RO\_004 - Camada com os limites da unidades de proteção integral - Fonte IBGE, contendo 362 áreas delimitadas.

5. RO\_005 - Camada com os limites das áreas especiais de proteção e conservação da biodiversidade - Fonte IBGE, contendo 1722 áreas delimitadas.

A Figura 4.7 exibe as camadas importadas na ferramenta QGIS.

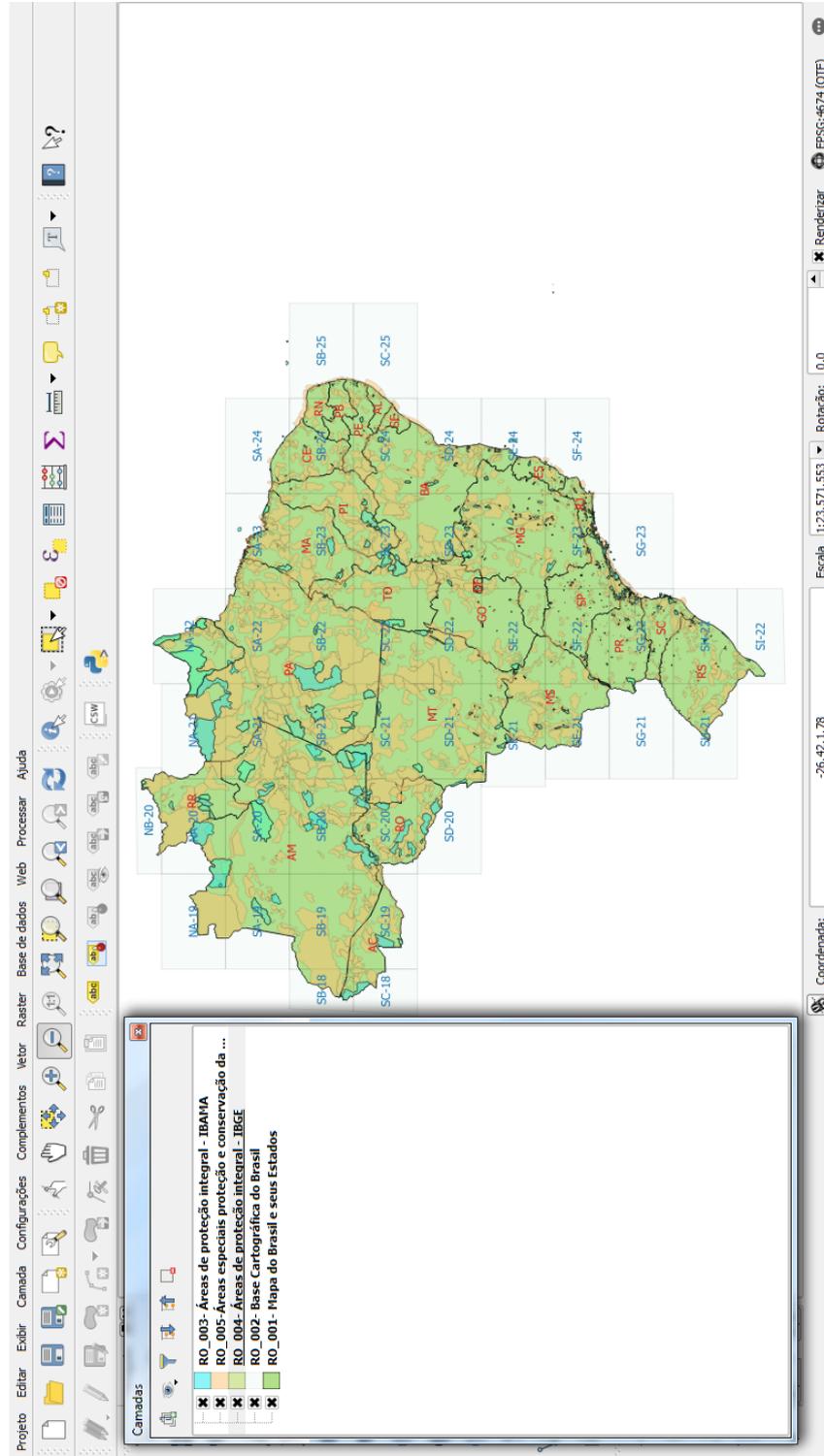


Figura 4.7: Visualização das camadas pela ferramenta QGIS.

#### 4.1.3.2 Avaliação de possíveis distorções

A verificação de possíveis distorções pode ser realizada pelo aumento da escala de visualização da ferramenta. Como exemplo, na Figura 4.8 apresentamos uma região da carta SC-22 que engloba o Parque Nacional do Araguaia e o Parque Estadual do Cantão. Verificamos a existência de duas áreas discordantes: uma presente apenas na camada RO\_003 e outra presente apenas na RO\_004.

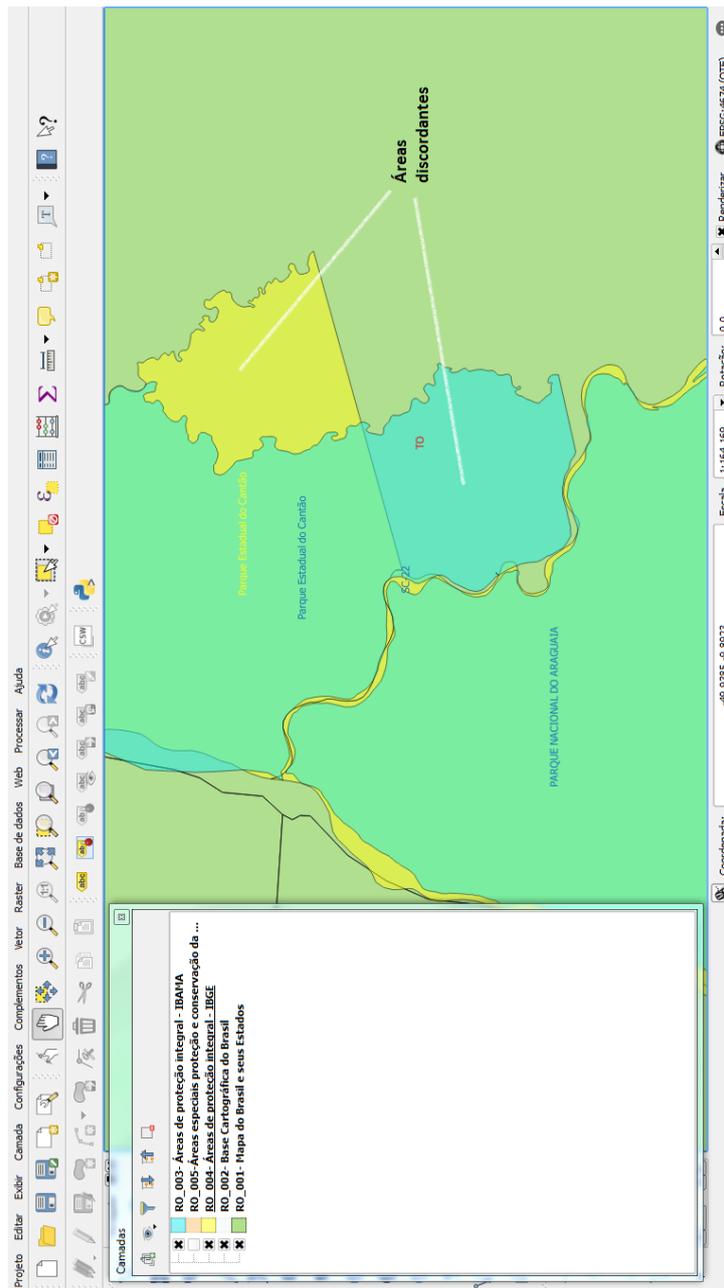


Figura 4.8: Visualização de áreas discordantes.

## 4.2 Avaliação segundo o método GQM

### 4.2.1 Meta 1: O processo da metodologia SERVUS deve ser auxiliado pelo geocatálogo.

#### 4.2.1.1 Questão 1: Os modelos propostos pela metodologia foram construídos?

Métrica 1: Percentual entre o número de modelos gerados e o número de modelos da SERVUS.

A Tabela 4.2 apresenta os modelos que compõem a metodologia e a indicação de quais foram produzidos no âmbito do modelo de pesquisa apresentado no Capítulo 3.

Tabela 4.2: Modelos da metodologia SERVUS utilizados.

Nome do modelo	Utilizado (SIM ou Não)
Modelo de domínio	SIM
Modelo de referência	NÃO
Modelo de informação (Processo e Caso de Uso)	SIM
Modelo de projeto - Recursos requisitados - REQs	SIM
Modelo de projeto - Capacidades ofertadas - CAPs	SIM
Modelo de projeto - Recursos Ofertados/Requisitados - REQ2CAP	SIM

O valor calculado para esta métrica é de 83,3%.

#### 4.2.1.2 Questão 2: As atividades propostas pela metodologia foram atendidas no modelo proposto?

Métrica 2: Percentual entre o número de atividades abordadas e o número total de atividades da SERVUS.

A Tabela 4.3 apresenta as atividades que compõem a metodologia e a indicação de quais foram executadas no âmbito do modelo de pesquisa apresentado no Capítulo 3.

O valor calculado para esta métrica é de 75%.

Tabela 4.3: Atividades do processo da metodologia SERVUS utilizados.

Nome da atividade	Utilizado (SIM ou Não)
Modelagem do problema	SIM
Modelagem do domínio	SIM
Publicação	SIM
Reformulação	SIM
Descoberta	SIM
Combinação	SIM
<i>Feedback Generation</i>	NÃO
<i>Grounding</i>	NÃO

## 4.2.2 Meta 2: O geocatálogo deve atuar como uma ferramenta de apoio a auditoria ambiental.

### 4.2.2.1 Questão 3: O geocatálogo facilita ou simplifica a atividade de localização de recursos geográficos ?

Métrica 3: Comparativo entre o custo da aquisição dos recursos em relação a auditoria realizada em 2015.

A auditoria utilizada como referência foi a ‘Auditoria Coordenada de Áreas Protegidas da América Latina’[30]. Para avaliação desta métrica foi elaborada uma entrevista com um dos membros da equipe de auditoria realizada em 2015, a fim de serem levantados aspectos das dificuldades enfrentadas na aquisição de dados e recursos geográficos.

Segue abaixo a entrevista:

Para fins de avaliação do geocatálogo semântico, produto da dissertação de mestrado "Implementação de um geocatálogo utilizando bancos de dados orientado a grafos para apoio à metodologia SERVUS", a ser implantado no TCU, solicito o obséquio de responder as questões abaixo:

A auditoria coordenada de áreas protegidas da América Latina, realizada em 2015, necessitou em uma de suas fases do levantamento da localização, do quantitativo e das características das denominadas APAs( Áreas de Proteção Ambiental).

1) Descreva de forma resumida os passos adotados para obtenção dos recursos e informações geográficas.

“À época de realização dos trabalhos da auditoria em unidades de conservação da Amazônia, não havia na SecexAmbiental conhecimento a respeito dos sistemas

de informações geográficas. As primeiras tentativas de se trabalhar com mapas foram em programas de edição gráfica, tais como Adobe Illustrator e Photoshop. Houve necessidade de obtenção de mapas das unidades de conservação brasileiras, e isso poderia ser obtido no site do Ministério do Meio Ambiente. O formato em que os arquivos eram disponibilizados (shape), contudo, não poderia ser aberto por programas de edição gráfica. Aí, descobriram-se os sistemas de informação geográfica. Os arquivos com camadas da divisão administrativa do Brasil e das unidades de conservação estavam concentrados no site do MMA, o que facilitou sua obtenção. Contudo, para avançarmos na análise da efetividade das unidades de conservação na proteção do meio ambiente, decidimos analisar o desmatamento e os fluxos de carbono nas UCs da Amazônia. As informações necessárias foram, entre outras, os usos da terra na região em pelo menos dois momentos históricos (para fins de comparação e as taxas de desmatamento dentro de cada UC e fora delas também. Os dados sobre desmatamento foram extraídos do sistema Deter (Detecção de Desmatamento em Tempo Real), enquanto os dados sobre usos da terra foram trazidos do IBGE. Por falhas no arquivo, as informações de usos da terra não puderam ser abertas no software utilizado pela equipe (QGIS), o que levou à utilização dos mapas em PDF. A combinação da utilização dos mapas em PDF com mapas exportados pelo QGIS permitiu, por fim, a execução das análises desejadas.”

Além disso, foram também utilizados dados georreferenciados estaduais, obtidos pela combinação da pesquisa pela própria equipe do TCU com o envio de dados complementares pelos Tribunais de Contas Estaduais participantes da auditoria coordenada.

2) Estime quais foram as necessidades em termos de tempo e recursos humanos utilizados.

“O processo de obtenção dos recursos geográficos foi concomitante ao desenvolvimento de outros trabalhos da auditoria. Contudo, exclusivamente para a obtenção, pode-se estimar um tempo líquido de duas semanas. Quanto aos recursos humanos, apenas uma pessoa se dedicou a essa tarefa.”

3) Baseado no vídeo demonstrativo em anexo e nas apresentações que você participou, cite alguns benefícios da utilização de um geocatálogo.

“Grande parte das informações geográficas utilizadas na auditoria de unidades de conservação foi encontrada com base em conversas com colegas, e esse processo levou um tempo considerável. No geocatálogo, é possível reunir essas informações de forma muito mais rápida. Ainda que o acesso à informação não esteja disponível,

ao menos já se sabe da existência de tal informação, tornando-se possível buscá-la em contato direto com a fonte.”

“Outra vantagem é a facilidade de visualizar os metadados, que, em geral, não são de fácil acesso para quem não tem conhecimento específico da área de informações geográficas. De forma geral, o geocatálogo torna a busca de informação geográficas um processo mais acessível aos leigos no assunto.”

Para a obtenção dos dados, verificação da pertinência, importação para ferramenta QGIS e montagem das camadas, neste trabalho de pesquisa, foram utilizadas cinco horas de trabalho. Embora os dados obtidos representem um percentual dos obtidos na auditoria de 2015, estima-se que com a utilização do geocatálogo esta fase de obtenção dos recursos possa ser reduzida de duas semanas para três dias.

#### 4.2.2.2 Questão 4: As consultas no geocatálogo retornam recursos ou informações geográficas relevantes ?

Métrica 4: Avaliação da relevância dos 10 primeiros itens retornados, dado um cenário específico. Para resolução desta questão foi elaborada uma pesquisa *on line* direcionada a profissionais que atuam em auditoria ambiental e profissionais de TI que atuam em sistemas de busca.

O objetivo da pesquisa é comparar, dado um cenário específico, a relevância atribuída na pesquisa com o índice de relevância fornecido pelo geocatálogo. A pesquisa é apresentada no Anexo VI e consolidada na Figura 4.9 e na Tabela 4.5. A Tabela 4.4 apresenta as faixas atribuídas para a pesquisa e para os índices de relevância. As faixas foram distribuídas proporcionalmente, considerando o limite máximo de 5 pontos atribuído a cada questão e o índice de relevância máximo de 0,09 obtido na consulta exibida no Anexo VI. Foram considerados os dez primeiros resultados, cujos arquivos encontravam-se disponíveis para *download*.

Tabela 4.4: Valores mínimos e máximos para a faixas atribuídas para pesquisa e índice de relevância.

Faixa	Mínimo e Máximo da pesquisa	Mínimo e Máximo do índice de relevância
Extrema relevância	4,50 - 5,00	0,081 - 0,090
Alta relevância	3,50 - 4,49	0,063 - 0,080
Relevante	2,50 - 3,49	0,045 - 0,062
Pouco relevante	1,50 - 2,49	0,027 - 0,044
Nenhuma relevância	1,00 - 1,49	0,018 - 0,026

Tabela 4.5: Dados consolidados da pesquisa.

Questões	Índice de relevância	Classificação pelo índice de relevância	Média da pesquisa	Classificação pela pesquisa	DP(Desvio padrão) das questões da pesquisa
A	0,090	Extrema relevância	4,10	Alta relevância	0,96
B	0,090	Extrema relevância	4,27	Alta relevância	0,78
C	0,090	Extrema relevância	3,27	Relevante	0,98
D	0,060	Alta relevância	3,63	Relevante	1,13
E	0,060	Alta relevância	3,43	Relevante	1,03
F	0,060	Alta relevância	4,03	Alta relevância	0,93
G	0,060	Alta relevância	4,00	Alta relevância	1,08
H	0,060	Alta relevância	4,13	Alta relevância	0,68
I	0,050	Relevante	3,80	Alta relevância	0,81
J	0,040	Pouco relevante	4,20	Alta relevância	0,89

Considerando-se as faixas estabelecidas na Tabela 4.4, a totalidade dos dez primeiros itens foram classificados como relevantes ou de alta relevância para o contexto, ou cenário de auditoria, apresentado. A classificação da pesquisa se mostrou coincidente com a do índice de relevância para 40% dos itens e se mostrou igual ou mais positiva para 50% dos itens.

#### 4.2.2.3 Questão 5: O tempo de resposta das consultas é aceitável ?

Métrica 5: O tempo de resposta do geocatálogo em cenários específicos.

A Tabela 4.6 apresenta os cenários com os tipos de pesquisa possíveis, o grau de complexidade e o tempo médio de resposta verificado.

Do total de nove cenários levantados, considerando-se o nível de complexidade e o Requisito 6, apenas três excederam ao tempo limite. O tempo médio, considerando-se todos os cenários, foi de 18 segundos, já para os cenários simples e complexos foram de 11 segundos e 23 segundos, respectivamente.

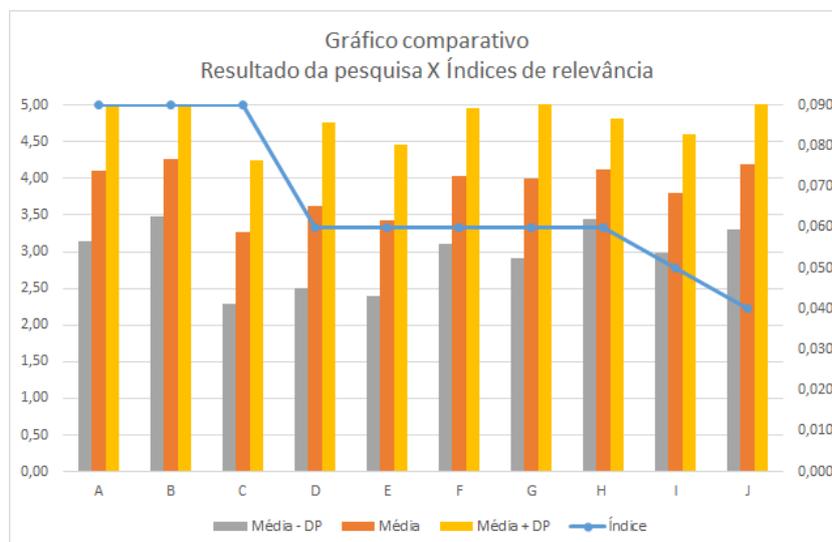


Figura 4.9: Gráfico resultado da pesquisa.

Tabela 4.6: Cenários e tempos de resposta do geocatálogo.

Cenário	Complexidade	Tempo de resposta em segundos
Consulta por descrição (Termo pesquisado: 'Área de proteção ambiental')	simples	13
Consulta por palavra-chave (Termo pesquisado: 'conservação')	simples	5
Consulta por classificação em uma ontologia (Termo pesquisado: 'Água')	simples	7,5
Consulta por descrição e palavra-chave	simples	18
Consulta por descrição, palavra-chave e classificação em uma ontologia	complexo	23
Consulta por descrição, palavra-chave e classificação em duas ontologias	complexo	25
Consulta por descrição, palavra-chave e classificação em três ontologia	complexo	31
Consulta por descrição e palavra-chave com filtragem espacial	complexo	19
Consulta por descrição e palavra-chave com filtragem temporal	complexo	19

# Capítulo 5

## Conclusão

Esta dissertação abordou a implementação de um geocatálogo que possibilita a utilização de ontologias como parâmetros de pesquisa. As consultas são realizadas sobre os metadados de geosserviços de infraestruturas de dados espaciais. Foram abordados aspectos da implantação tais como: arquitetura e modelo de dados. No que se refere a arquitetura, o geocatálogo foi implementado como uma aplicação Web de três camadas, passível de implantação em ambiente de nuvem. O modelo de dados foi implementado em banco de dados orientado a grafos.

Com base na montagem do cenário de auditoria e nos resultados do planejamento GQM apresentados no Capítulo 4, nos quais as etapas e modelos da metodologia SERVUS foram construídos, podemos afirmar que o geocatálogo se mostrou viável como ferramenta de apoio a metodologia SERVUS, possibilitando a atuação nas etapas de *publishing* e *matching*. Podem ser consideradas como contribuições deste trabalho a utilização da linguagem Cypher no manuseio de rede de recursos, ontologias e metamodelos, de forma integrada, em banco de dados não relacionais orientado a grafos. O estabelecimento de mecanismos de parametrização das consultas, mediante ponderação de parâmetros e campos de metadados e o estabelecimento de um índice de relevância, são contribuições para o campo da busca em bases de metadados.

No que se refere a utilização do geocatálogo como ferramenta de auxílio em auditorias, ele proporcionou o acesso a uma base consolidada de 19.000 recursos geográficos, oriundos da Infraestrutura Nacional de Dados Espaciais do Brasil (INDE-BR), Instituto Brasileiro de Geografia e Estatística (IBGE), Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA), Projeto de Dados Geográficos da Secretaria do Meio Ambiente de SP (DATAGEO-SP) e Agência Europeia do Ambiente (EEA) e Companhia de Pesquisa de Recursos Minerais (CPRM). Os recursos consistem de mapas temáticos, mapas de base, dados em formatos XML ou *shapefile*, além de acesso a visualizadores *online*. Na resolução dos questionamentos do planejamento GQM obtivemos resultados

satisfatórios quanto a relevância das informações apresentadas pelo geocatálogo e a perspectiva de melhora da produtividade na obtenção dos recursos e dados geográficos por parte da equipe de auditoria.

Como trabalhos futuros podemos considerar a automatização do processo de *harvesting* e a integração com uma infraestrutura de dados espaciais corporativa, de forma que o geocatálogo atue conjuntamente com geoportais. Outro ponto a se considerar é a utilização da metodologia SERVUS para o desenvolvimento de soluções corporativas em auditorias na área de meio ambiente.

# Referências

- [1] David Ameller, Xavier Burgués, Oriol Collell, Dolors Costal, Xavier Franch, e Mike P Papazoglou. Development of service-oriented architectures using model-driven development: A mapping study. *Information and Software Technology*, 62:42–66, 2015. 22
- [2] Peter Baumann. Ogc wcs 2.0 interface standard- core: Corrigendum, 2012. 16
- [3] Miguel Á Bernabé-Poveda e Carlos M López-Vázquez. *Fundamentos de las Infraestructuras de Datos Espaciales (IDE)*. BibliotecaOnline SL, 2012. 8, 9, 10, 15, 16, 17
- [4] Norbert Bieberstein. *Service-oriented architecture compass: business value, planning, and enterprise roadmap*. FT Press, 2006. 23
- [5] Karin Breitman. *Web Semântica: A Internet do futuro*. LTC-Livros Técnicos e Científicos Editora S.A, 2005. 32
- [6] Jeen Broekstra, Arjohn Kampman, e Frank Van Harmelen. Sesame: A generic architecture for storing and querying rdf and rdf schema. In *The Semantic Web—ISWC 2002*, pages 54–68. Springer, 2002. 34
- [7] VRBG Caldiera e H Dieter Rombach. The goal question metric approach. *Encyclopedia of software engineering*, 2(1994):528–532, 1994. 44
- [8] Silvana Philippi Camboim. Arquitetura para integração de dados interligados abertos à INDE-BR. Tese de doutorado, Universidade Federal do Paraná - UFPR, 2013. 3, 41
- [9] Jeremy J Carroll, Ian Dickinson, Chris Dollin, Dave Reynolds, Andy Seaborne, e Kevin Wilkinson. Jena: implementing the semantic web recommendations. In *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters*, pages 74–83. ACM, 2004. 34
- [10] Paul C Clements. *Software architecture in practice*. Tese (Doutorado), Software Engineering Institute, 2002. 18
- [11] CONCAR. Plano de ação para implantação da infraestrutura nacional de dados espaciais., 2010. 10

- [12] J Daltio e C A Carvalho. Um framework para recuperação semântica de dados espaciais. In *Embrapa Gestão Territorial-Artigo em anais de congresso (ALICE)*. In: Brazilian Symposium on Geoinformatics, 13., 2012, Campos do Jordão. Proceedings... São José dos Campos: MCTI/INPE, 2012. p. 60-65., 2013. 36, 39
- [13] Fabio G de Andrade e Cláudio de S Baptista. Using semantic similarity to improve information discovery in spatial data infrastructures. *Journal of Information and Data Management*, 2(2):181, 2011. 39, 40
- [14] Jeff de la Beaujardiere. Opengis web map server implementation specification, 2006. 14
- [15] Karla Albuquerque de Vasconcelos Borges. Uso de uma ontologia de lugar urbano para reconhecimento e extração de evidências geo-espaciais na web. 2006. 39
- [16] Thomas Erl, Clive Gee, Pethuru Raj Chelliah, Jürgen Kress, Hajo Normann, Berthold Maier, Leo Shuster, Bernd Trops, Torsten Winterberg, Clemens Utschig, et al. *Next Generation SOA: A Concise Introduction to Service Technology & Service-Oriented Architecture*. Pearson Education, 2014. 19, 20
- [17] Feroz Farazi, Vincenzo Maltese, Fausto Giunchiglia, e Alexander Ivanyukovich. A faceted ontology for a semantic geo-catalogue. In *The Semantic Web: Research and Applications*, pages 169–182. Springer, 2011. 37
- [18] Paulo JA Gimenez, Asterio K Tanaka, e Fernanda Baião. Uma proposta de integração semântica para a infraestrutura nacional de dados espaciais usando geo-ontologias. 2013. 40
- [19] Fausto Giunchiglia, Aliaksandr Autayeu, e Juan Pane. S-match: an open source framework for matching lightweight ontologies. *Web Intelligence and Agent Systems: An International Journal 1 (2010) 1–9 1*, 2010. 36
- [20] Thomas R Gruber. Toward principles for the design of ontologies used for knowledge sharing? *International journal of human-computer studies*, 43(5):907–928, 1995. 23
- [21] Clodoveu Augusto Davis Júnior, Alves, e Leonardo Lacerda. Infra-estrutura de dados espaciais: potencial para uso local. *Revista Informática Pública. Belo Horizonte*, 8(1):65–80, 2006. 9, 11
- [22] Thomas Kohlborn, Axel Korthaus, Taizan Chan, e Michael Rosemann. Service analysis: A critical assessment of the state of the art. 2009. 20
- [23] Stephen Lane e Ita Richardson. Process models for service-based applications: A systematic literature review. *Information and Software Technology*, 53(5):424–439, 2011. 21
- [24] Joshua Lieberman. Geospatial semantic web interoperability experiment report. *OGC Discussion Paper*, 2006. 32, 34, 35

- [25] C Matthew MacKenzie, Ken Laskey, Francis McCabe, Peter F Brown, Rebekah Metz, e Booz Allen Hamilton. Reference model for service oriented architecture 1.0. *OASIS Standard*, 12, 2006. 18
- [26] Florian Marienfeld, Ina Schieferdecker, Evanela Lapi, e Nikolay Tcholtchev. Meta-data aggregation at govdata. de: an experience report. In *Proceedings of the 9th International Symposium on Open Collaboration*, page 21. ACM, 2013. 49
- [27] Eric A Marks e Michael Bell. *Service Oriented Architecture (SOA): a planning and implementation guide for business and technology*. John Wiley & Sons, 2008. 18
- [28] Jakob Nielsen. *Usability engineering*. Elsevier, 1994. 43
- [29] OGC. Ogc reference model, version 2.1, 2011. 11
- [30] OLACEFS/COMTEMA. *Resumo Executivo - Áreas protegidas América Latina : Auditoria Coordenada*. TCU-Tribunal de Contas da União, 2015. 59, 69
- [31] Rob Raskin e Michael Pan. Semantic web for earth and environmental terminology (sweet). In *Proc. of the Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data*, 2003. 25
- [32] Robert G Raskin e Michael J Pan. Knowledge representation in the semantic web for earth and environmental terminology (sweet). *Computers & geosciences*, 31(9):1119–1125, 2005. 39
- [33] Ian Robinson, Jim Webber, e Emil Eifrem. *Graph databases*. "O'Reilly Media, Inc.", 2013. 46, 47, 48
- [34] Pavel Shvaiko, Alexander Ivanyukovich, Lorenzo Vaccari, Vincenzo Maltese, e Feroz Farazi. A semantic geo-catalogue implementation for a regional sdi. 2010. 35, 36
- [35] Thomas Usländer. *Service-oriented design of environmental information systems*, volume 5. KIT Scientific Publishing, 2010. 3, 22, 24, 25, 26, 29, 30, 31, 38, 60
- [36] Lorenzino Vaccari, Pavel Shvaiko, e Maurizio Marchese. A geo-service semantic integration in spatial data infrastructures. *International Journal of Spatial Data Infrastructures Research*, 4:24–51, 2009. 31
- [37] Panagiotis (Peter) A. Vretanos. Opengis web feature service 2.0 interface standard, 2010. 15
- [38] Arliss Whiteside e Jim Greenwood. Ogc web services common standard, 2010. 13

# **Apêndice A**

## **Mapa estratégico do TCU 2015 à 2021**

MAPA ESTRATÉGICO DO TCU - PET 2015-2021

<b>MISSÃO</b>	Aprimorar a Administração Pública em benefício da sociedade por meio do
<b>VISÃO</b>	Ser referência na promoção de uma Administração Pública efetiva, ética, ágil e responsável

**RESULTADOS**

Aprimorar a governança e a gestão em organizações e políticas públicas

Coibir a má gestão dos recursos públicos

Fomentar a Administração Pública transparente

**PROCESSOS INTERNOS — CONTROLE EXTERNO**

Intensificar a atuação com base em análise de risco

Realizar diagnósticos sistêmicos em áreas relevantes

Induzir a disponibilidade e a confiabilidade de informações da Administração Pública

Induzir o aperfeiçoamento da gestão de riscos e controles internos da Administração Pública

Estimular o monitoramento e a avaliação do desempenho pela Administração Pública

Induzir o aperfeiçoamento da gestão das políticas públicas descentralizadas

Condenar efetiva e tempestivamente os responsáveis por irregularidades e desvios

**PROCESSOS INTERNOS — RELACIONAMENTO INSTITUCIONAL**

Intensificar o intercâmbio nacional e internacional para compartilhamento de melhores práticas para o controle

Intensificar parcerias com outros órgãos de controle

Estreitar o relacionamento com o Congresso Nacional

Aperfeiçoar a comunicação do TCU com as partes interessadas

**GESTÃO E INOVAÇÃO**

Aprimorar a gestão do conhecimento

Promover integração e sinergia no Tribunal

Promover a melhoria da governança e da gestão da estratégia no TCU

Aprimorar o uso da TI como instrumento de inovação para o controle

Desenvolver capacidade organizacional ampla para trabalhar com recursos tecnológicos

Aperfeiçoar a gestão do desempenho, o reconhecimento e o desenvolvimento profissional

Assegurar suporte de bens e serviços adequado às necessidades do TCU

# Anexo I

## Exemplo de arquivo XSL para *parser* de XML com metadados de serviços CSW

```
1
2 <?xml version="1.0" encoding="UTF-8"?>
3 <xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
4   xmlns:csw="http://www.opengis.net/cat/csw/2.0.2"
5   xmlns:dc="http://purl.org/dc/elements/1.1/"
6   xmlns:str="http://exslt.org/strings"
7   xmlns:ows="http://www.opengis.net/ows" version="2.0">
8 <xsl:output method="text" />
9 <xsl:template match="/csw:GetRecordsResponse">
10 <xsl:text>ID, language , date , title , abstract , description , source , crs ,
11 lowerCorner , upperCorner , link , download , thumbnail , large_thumbnail ,
12 protocol , subject&#13;</xsl:text>
13 <xsl:apply-templates select="//csw:Record"/>
14 </xsl:template>
15 <xsl:template match="csw:Record" >
16 <xsl:text>|!!!</xsl:text>
17 <xsl:value-of select="dc:identifier"
18   /><xsl:text>|!!!, </xsl:text>
19 <xsl:value-of select="dc:language" /><xsl:text>|!!!, </xsl:text>
20 <xsl:text>|!!!</xsl:text>
21 <xsl:value-of select="dc:date" /><xsl:text>|!!!, </xsl:text>
22 <xsl:text>|!!!</xsl:text>
23 <xsl:value-of select="dc:title" /><xsl:text>|!!!, </xsl:text>
24 <xsl:text>|!!!</xsl:text>
25 <xsl:value-of select="dc:abstract"/><xsl:text>|!!!, </xsl:text>
```

```

26 <xsl:text>|!!!|</xsl:text>
27 <xsl:value-of select="dc:description"
    /><xsl:text>|!!!|,</xsl:text>
28 <xsl:text>|!!!|</xsl:text>
29 <xsl:value-of select="dc:source" /><xsl:text>|!!!|,</xsl:text>
30 <xsl:text>|!!!|</xsl:text>
31 <xsl:value-of select="ows:BoundingBox/@crs" />
32 <xsl:text>|!!!|,</xsl:text>
33 <xsl:text>|!!!|</xsl:text>
34 <xsl:value-of select="ows:BoundingBox/ows:LowerCorner" />
35 <xsl:text>|!!!|,</xsl:text>
36 <xsl:text>|!!!|</xsl:text>
37 <xsl:value-of select="ows:BoundingBox/ows:UpperCorner" />
38 <xsl:text>|!!!|,</xsl:text>
39 <xsl:text>|!!!|</xsl:text>
40 <xsl:value-of
41   select="dc:URI[@protocol='WWW:LINK-1.0-http--link']" />
42 <xsl:text>|!!!|,</xsl:text>
43 <xsl:text>|!!!|</xsl:text>
44 <xsl:value-of
45   select="dc:URI[@protocol='WWW:DOWNLOAD-1.0-http--download']" />
46 <xsl:text>|!!!|,</xsl:text>
47 <xsl:text>|!!!|</xsl:text>
48 <xsl:value-of select="dc:URI[@name='thumbnail']" />
49 <xsl:text>|!!!|,</xsl:text>
50 <xsl:text>|!!!|</xsl:text>
51 <xsl:value-of select="dc:URI[@name='large_thumbnail']" />
52 <xsl:text>|!!!|,</xsl:text>
53 <xsl:text>|!!!|</xsl:text>
54 <xsl:value-of
55   select="dc:URI[contains(@protocol,'OGC')]/@protocol" />
56 <xsl:text>|!!!|,</xsl:text>
57 <xsl:for-each select="dc:subject">
58 <xsl:value-of select="."/>
59 <xsl:if test="(following-sibling::dc:subject)">
60 <xsl:text>,</xsl:text></xsl:if>
61 </xsl:for-each>
62 <xsl:text>]!!!|</xsl:text>
63 <xsl:text>&#13;</xsl:text>
64 </xsl:template>
65 </xsl:stylesheet>

```

# Anexo II

## Exemplo de arquivo XML ontologia VCE

```
<?xml version="1.0"encoding="ISO-8859-1"?>
  <THESAURUS>
    <CONCEPT>
      <NON-DESCRIPTOR>AAE</NON-DESCRIPTOR>
      <USE>Avaliação ambiental estratégica</USE>
    </CONCEPT>
    <CONCEPT>
      <DESCRIPTOR>Abastecimento de água</DESCRIPTOR>
      <UP>Oferta de água</UP>
      <UP>Sistema de abastecimento de água</UP>
      <DEF>Abastecimento através de rede geral ou outra forma (água proveniente de
chafariz, bica ou mina, poço particular, caminhão-pipa, cursos d'água etc.). (Fonte: IBGE.
Atlas de saneamento. IBGE. Glossário. Rio de Janeiro : IBGE, 2004. P. 145. Disponível
em: http://www.ibge.gov.br/home/estatistica/populacao/atlas\_saneamento; Acesso em:
13 mar 2013). </DEF>
      <TR>Água</TR>
      <TR>Água canalizada</TR>
      <TR>Aquífero</TR>
      <TR>Equipamento urbano</TR>
    </CONCEPT>
  </THESAURUS>
```

## Anexo III

### Consulta para geração do *ranking*

```
1 MATCH
2     (a:Recurso{codRecurso:'RecursoRequisitado_001'})-[p:LOCALIZA]->
3     (b:Recurso:Ofertado)-[r]->(c:RecursoSemantico)
4 WHERE p.peso > 0
5 AND ((toFloat(split(replace(c.coordenadaSuperior,',',','.'),
6     ''')[1])
7     >= -11.29908892914917
8 AND toFloat(split(replace(c.coordenadaInferior,',',','.'), ''')[1])
9     <= -11.29908892914917)
10 OR (toFloat(split(replace(c.coordenadaSuperior,',',','.'), ''')[1])
11     >= -24.158782883092716
12 AND toFloat(split(replace(c.coordenadaInferior,',',','.'), ''')[1])
13     <= -24.158782883092716)
14 OR (toFloat(split(replace(c.coordenadaSuperior,',',','.'), ''')[1])
15     <= -11.29908892914917
16 AND toFloat(split(replace(c.coordenadaInferior,',',','.'), ''')[1])
17     >= -24.158782883092716) )
18 AND NOT ( toFloat(split(replace(c.coordenadaSuperior,',',','.'),
19     ''')[1])
20     >= -11.29908892914917
21 AND toFloat(split(replace(c.coordenadaInferior,',',','.'), ''')[1])
22     <= -11.29908892914917
23 AND toFloat(split(replace(c.coordenadaSuperior,',',','.'), ''')[1])
24     >= -24.158782883092716
25 AND toFloat(split(replace(c.coordenadaInferior,',',','.'), ''')[1])
26     <= -24.158782883092716)
27 AND ( (toFloat(split(replace(c.coordenadaInferior,',',','.'),
28     ''')[0])
29     >= -38.7158203125
30 AND toFloat(split(replace(c.coordenadaSuperior,',',','.'), ''')[0])
31     <= -38.7158203125)
```

```
28 OR (toFloat(split(replace(c.coordenadaInferior,',',','.''), ''))[0])
29 >= -55.2392578125
30 AND toFloat(split(replace(c.coordenadaSuperior,',',','.''), ''))[0])
31 <= -55.2392578125)
32 OR (toFloat(split(replace(c.coordenadaInferior,',',','.''), ''))[0])
33 <= -38.7158203125
34 AND toFloat(split(replace(c.coordenadaSuperior,',',','.''), ''))[0])
35 >= -55.2392578125))
36 AND NOT ( toFloat(split(replace(c.coordenadaSuperior,',',','.''),
    ''))[0])
37 <= -38.7158203125
38 AND toFloat(split(replace(c.coordenadaInferior,',',','.''), ''))[0])
39 >= -38.7158203125
40 AND toFloat(split(replace(c.coordenadaSuperior,',',','.''), ''))[0])
41 <= -55.2392578125
42 AND toFloat(split(replace(c.coordenadaInferior,',',','.''), ''))[0])
43 >= -55.2392578125 )
44 RETURN sum(p.peso) as peso , b.nome, b.urlVisualizador,
45 b.urlDownload, b.origem,c.protocolo, id(b)
46 ORDER BY sum(p.peso) desc SKIP 0 LIMIT 50;
```

## Anexo IV

# Parâmetros de configuração de consulta para a atividade de *matching*

Tabela IV.1: Configuração de pesos para consulta.

<b>Campo termo pesquisa</b>	<b>Peso termo pesquisa</b>	<b>Campo metadado</b>	<b>Peso metadado</b>
Descricao	9	Nome	9
Palavra-chave	8	Descrição	5
Ontologia VCE (termo raiz)	5	Assunto	4
Ontologia VCE (termo pai ou filho)	2	Resumo	3
Ontologia VCGE (termo raiz)	5	Fonte	3
Ontologia VCGE (termo pai ou filho)	2		
Ontologia EDGV (termo raiz)	5		
Ontologia EDGV (termo pai ou filho)	2		

Tabela IV.2: Configuração das caixas de seleção de filtragem.

<b>Campo termo pesquisa : filtros</b>	<b>situação</b>
Caixa de seleção exibir recursos para visualização	desmarcado
Caixa de seleção exibir recursos para <i>download</i>	marcado
Caixa de seleção exibir recursos com metadados	desmarcado

# Anexo V

## Principais telas do geocatálogo

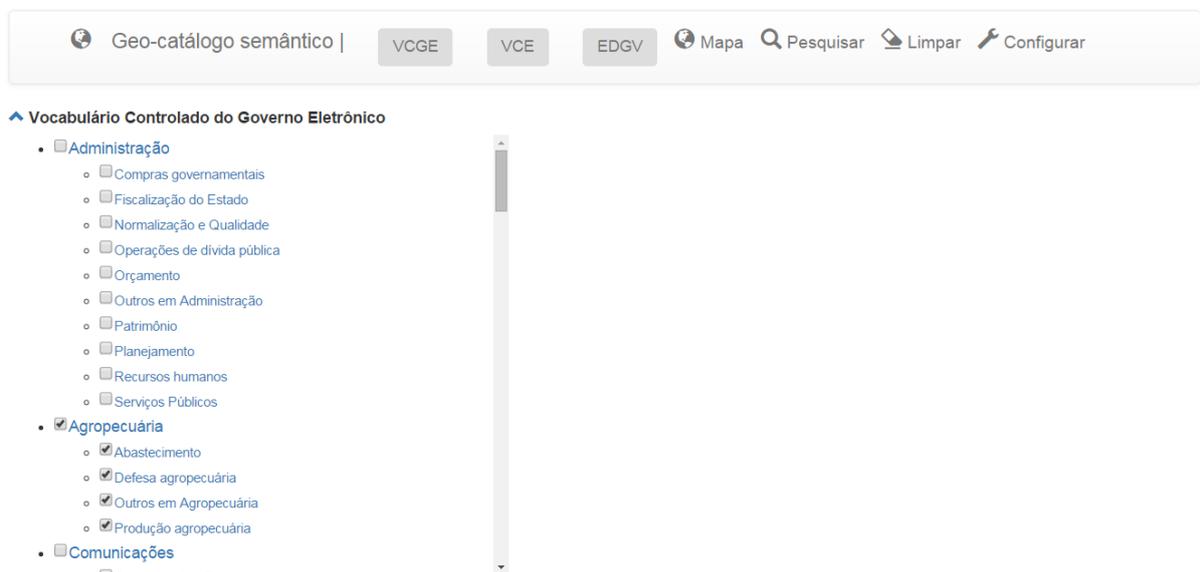


Figura V.1: Tela com a classificação por ontologia.

Geo-catálogo semântico | VCGE VCE EDGV Mapa Pesquisar Limpar Configurar

LATITUDE-NORTE	-12.23082482846048
LATITUDE-SUL	-20.610468916797126
LONGITUDE-OESTE	-52.954101520590484
LONGITUDE-LESTE	-41.660156208090484

Descrição: Palavras-chave(separadas por vírgula):

De (ANO) Até (ANO)

Resultados por página: 50 Página corrente: 1

Figura V.2: Tela com o filtro espacial.

Geo-catálogo semântico | VCGE VCE EDGV Mapa Pesquisar Limpar Configurar

Configuração Salvar

Campos de pesquisa:

Peso descrição	9	Peso palavras-chave	8
Peso VCGE (matching direto)	5	Peso VCE (matching direto)	5
Peso EDGV (matching direto)	5	Peso VCGE (matching indireto)	2
Peso VCE (matching indireto)	2	Peso EDGV (matching indireto)	2

Campos dos metadados:

Peso nome	3	Peso descrição	5	Peso assunto	4
Peso resumo	3	Peso fonte	3		

Figura V.3: Tela com a configuração de pesos.

Geo-catálogo semântico | VCGE VCE EDGV Mapa Pesquisar Limpar Configurar

Descrição: Vegetação reserva ambiental Palavras-chave(separadas por vírgula): floresta, cultivos, agricultura, sustentável

De (ANO) Até (ANO)

← Anterior Resultados por página: 50 Página corrente: 1 Próxima →

#	Índice	Peso		Recurso	Fonte	Tipo
1	0.21	297		Zoneamento Ecológico-Econômico - Aptidão para Proteção e Preservação Ambiental no município de Goiânia	INDE	WMS
2	0.20	287		Vegetação da Amazônia Legal em 2002 (1:2.500.000)	INDE	WMS
3	0.18	248		Indicadores de Desenvolvimento Sustentável 2015 - Classe de periculosidade ambiental(CGEO:IDS_03_Comerc_de_Agrotoxicos_e_Afins_Total_1_2)	IBGE	WFS
4	0.15	207		Mapa de Vegetação do Brasil (2002)	INDE	WMS
5	0.14	202		Mapa de Vegetação da Folha SD.22 – Goiás	INDE	WMS
6	0.14	202		Mapa de Vegetação da Amazônia (2002)	INDE	WMS
7	0.14	202		Pontos de Vegetação da Folha SA 23 - São Luís	INDE	WMS

Figura V.4: Tela com a listagem do ranking.

## Anexo VI

### Pesquisa de avaliação e resultados

# Pesquisa acadêmica - Mestrado UNB - Um projeto de geocatálogo semântico para o TCU.

Prezado colega,

Gostaria de contar com sua colaboração para o preenchimento desta pesquisa acadêmica.

O tempo estimado para conclusão é de 10 minutos.

Há 10 perguntas neste questionário

## Avaliação dos resultados de um busca.

**Considere o seguinte cenário: um auditor necessita de recursos e informações geográficas como insumo para a sua fiscalização. A auditoria trata de áreas de proteção ambiental, havendo a necessidade de levantamento de quantitativos, características e localização das áreas.**

---

**Foi utilizado um geocatálogo, que possibilita a busca de recursos geográficos, com os seguintes parâmetros de pesquisa :**

**Descrição :** "Áreas de proteção ambiental ou unidades de conservação ambiental"

**Palavras-chaves:** "APA"

**Classificação no VCE(Vocabulário de Controle Externo) :** "Preservação ambiental"

**Classificação no VCGE(Vocabulário Controlado do Governo Eletrônico) :** "Preservação e Conservação Ambiental"

---

**Avalie o grau de relevância dos resultados. Considere o valor 5 como maior relevância e valor 1 como menor relevância.**

**1**

**Proporção das áreas de unidades de conservação de uso sustentável.**

**Dados contendo os limites geográficos dos estados da federação e o percentual de áreas de conservação de uso sustentável em relação à área total do estado.**

**\***

Favor escolher apenas uma das opções a seguir:

- 1
- 2
- 3
- 4
- 5

**(1 - Nenhuma relevância, 2 - Pouco relevante, 3 - Relevante, 4 - Alta relevância, 5 - Extrema relevância)**

**2**

**Pelotões da Polícia Militar Ambiental para Unidades de Conservação 2015.**

**Dados contendo as unidades de conservação do estado de São Paulo que possuem pelotões da Polícia Militar Ambiental. \***

Favor escolher apenas uma das opções a seguir:

- 1
- 2
- 3
- 4
- 5

**(1 - Nenhuma relevância, 2 - Pouco relevante, 3 - Relevante, 4 - Alta relevância, 5 - Extrema relevância)**

**3****Unidades de proteção integral.****Dados fornecidos pelo IBAMA contendo áreas de proteção integral, nome, sigla, área oficial, perímetro, ano de criação, decreto que instituiu e órgão coordenador. \***

Favor escolher apenas uma das opções a seguir:

- 1
- 2
- 3
- 4
- 5

**(1 - Nenhuma relevância, 2 - Pouco relevante, 3 - Relevante, 4 - Alta relevância, 5 - Extrema relevância)****4****Unidades de Conservação de Proteção Integral 2013.****Dados contendo as unidades de conservação de proteção integral organizadas por estados e municípios. \***

Favor escolher apenas uma das opções a seguir:

- 1
- 2
- 3
- 4
- 5

**(1 - Nenhuma relevância, 2 - Pouco relevante, 3 - Relevante, 4 - Alta relevância, 5 - Extrema relevância)**

**5****Áreas especiais de proteção e conservação da biodiversidade.****Dados contendo as áreas especiais de proteção e conservação da biodiversidade, informações de planejamento, bioma pertencente, relevância e priorização. \***

Favor escolher apenas uma das opções a seguir:

- 1  
 2  
 3  
 4  
 5

**(1 - Nenhuma relevância, 2 - Pouco relevante, 3 - Relevante ,4 - Alta relevância, 5 - Extrema relevância)****6****Densidade de focos de calor por 1000 km<sup>2</sup> - unidades de conservação e terras indígenas.****Dados contendo focos de calor localizados em unidades de conservação ambiental e terras indígenas.****\***

Favor escolher apenas uma das opções a seguir:

- 1  
 2  
 3  
 4  
 5

**(1 - Nenhuma relevância, 2 - Pouco relevante, 3 - Relevante ,4 - Alta relevância, 5 - Extrema relevância)**

**7****Unidades de Conservação de Proteção Integral 2013.****Dados contendo as unidades de conservação de proteção integral com os decretos que as instituíram e as instituições ou órgãos responsáveis. \***

Favor escolher apenas uma das opções a seguir:

- 1
- 2
- 3
- 4
- 5

**(1 - Nenhuma relevância, 2 - Pouco relevante, 3 - Relevante ,4 - Alta relevância, 5 - Extrema relevância)****8****Indicadores de Desenvolvimento Sustentável 2015 - Demanda Bioquímica de Oxigênio (DBO) - Região Sul.****Demanda Bioquímica de Oxigênio (DBO) dos pontos de monitoramento em rios da Região Sul - último ano da informação, em alguns trechos selecionados de rios desta região. \***

Favor escolher apenas uma das opções a seguir:

- 1
- 2
- 3
- 4
- 5

**(1 - Nenhuma relevância, 2 - Pouco relevante, 3 - Relevante ,4 - Alta relevância, 5 - Extrema relevância)**

**9****Indicadores de Desenvolvimento Sustentável 2015 - Classe de periculosidade ambiental****Dados contendo índices por estado da federação, graduando o nível de periculosidades de comercialização de agrotóxicos e produtos afins. \***

Favor escolher apenas uma das opções a seguir:

- 1
- 2
- 3
- 4
- 5

**(1 - Nenhuma relevância, 2 - Pouco relevante, 3 - Relevante ,4 - Alta relevância, 5 - Extrema relevância)****10****Áreas especiais de proteção e conservação da biodiversidade.****Dados contendo as áreas especiais de proteção e conservação contendo nome, sigla, UF, bioma. \***

Favor escolher apenas uma das opções a seguir:

- 1
- 2
- 3
- 4
- 5

**(1 - Nenhuma relevância, 2 - Pouco relevante, 3 - Relevante ,4 - Alta relevância, 5 - Extrema relevância)**

Obrigado pela participação !

Enviar questionário

Obrigado por ter preenchido o questionário.

Geocatálogo semântico | VCGE VCE EDGV Mapa Configurar Pesquisar Limpar

Descrição: Áreas de proteção ambiental ou unidades de conservação ambiental

Palavras-chave separadas por vírgula: APA

De (ANO) Até (ANO)

Fonte:

VCGE: Preservação e Conservação Ambiental | VCE: Preservação ambiental

Figura VI.1: Parâmetros utilizados na consulta.

← Anterior Resultados por página: 25 Página corrente: 1 Próxima →

#	Índice	Peso		Recurso	Fonte	Tipo
1	0.09	243		Proporção das áreas de unidades de conservação de uso sustentável(CGEO:IDS_19_Uni_de_Conserv_Uso_Sustent_Prop_Area_1_1)	IBGE	WFS
2	0.09	243		Pelotões da Polícia Militar Ambiental para Unidades de Conservação 2015(datageo:PelotaoPoliciaAmbientaUCs2015)	DATAGEO-SP	WFS
3	0.09	243		Percentual de unidades de conservação de proteção integral(CGEO:IDS_18_Uni_de_Conserv_Protecao_Int_Prop_Area_1_1)	IBGE	WFS
4	0.06	171		Indicadores de Desenvolvimento Sustentável 2015 - Classe de periculosidade ambiental(CGEO:IDS_03_Comerc_de_Agrotoxicos_e_Afins_Total_1_2)	IBGE	WFS
5	0.06	162		Unidades de conservação de proteção integral - área(CGEO:IDS_18_Uni_de_Conserv_Protecao_Int_Area_1_2)	IBGE	WFS
6	0.06	162		Áreas especiais de proteção e conservação da biodiversidade(CGEO:ANMS2010_03_apcb)	IBGE	WFS
7	0.06	162		Unidades de Conservação de Proteção Integral 2013(CGEO:IDS_18_Uni_de_Conserv_Protecao_Int_Num_2_2)	IBGE	WFS
8	0.06	162		Unidades de Conservação de Proteção Integral 2013(CGEO:IDS_17_Uni_de_Conserv_Protecao_Int_2_2)	IBGE	WFS
9	0.06	162		área das unidades de conservação de uso sustentável(CGEO:IDS_19_Uni_de_Conserv_Uso_Sustent_Area_1_2)	IBGE	WFS
10	0.06	162		Densidade de focos de calor por 1000 km? - unidades de conservação e terras indígenas(CGEO:IDS_08_Dens_Focos_Calor_UC_1_1)	IBGE	WFS
11	0.05	135		Unidades de Proteção Integral(csr:lim_unidade_protecao_integral_a)	IBAMA	WFS
12	0.04	117		Indicadores de Desenvolvimento Sustentável 2015 - Demanda Bioquímica de Oxigênio (DBO) - Região Sul(CGEO:IDS_14_Demanda_Bioqui_de_Oxig_Regiao_Sul_2_2)	IBGE	WFS

Figura VI.2: Resultado da consulta.



Filtros gerais



Filtros de resposta



Estadísticas



Resultados

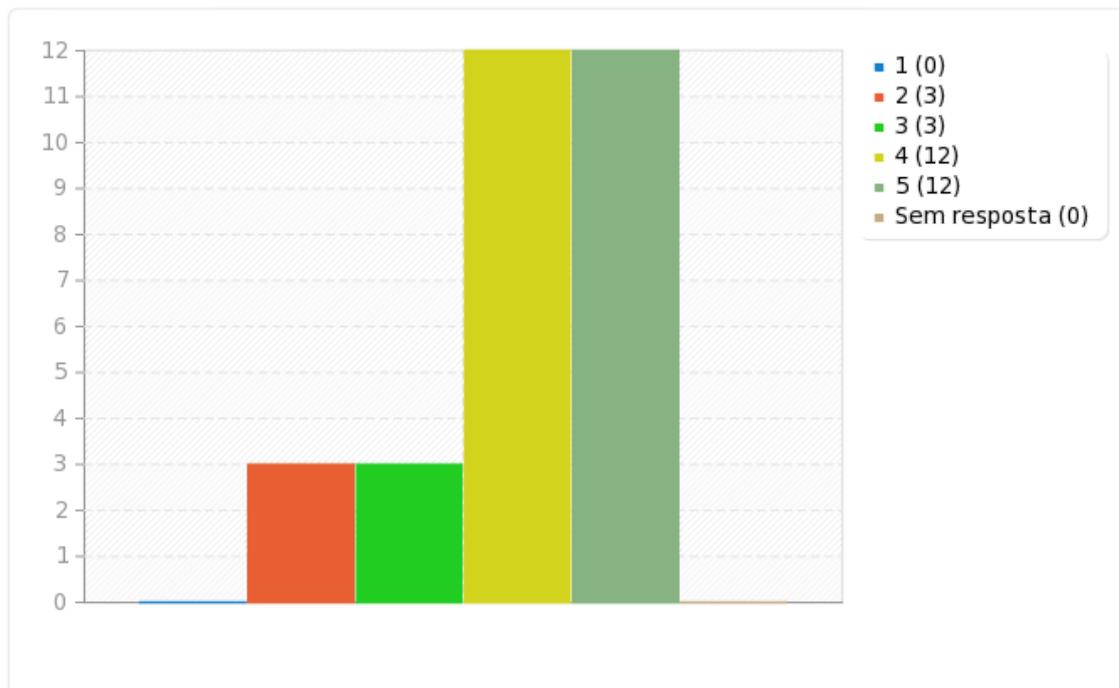
Número de registros nesta consulta: 30  
 Total de registros no questionário: 30  
 Percentagem do total: 100.00%

Navegar

Sumário dos campos para A

Proporção das áreas de unidades de conservação de uso sustentável. Dados contendo os limites geográficos dos estados da federação e o percentual de áreas de conservação de uso sustentável em relação à área total do estado.

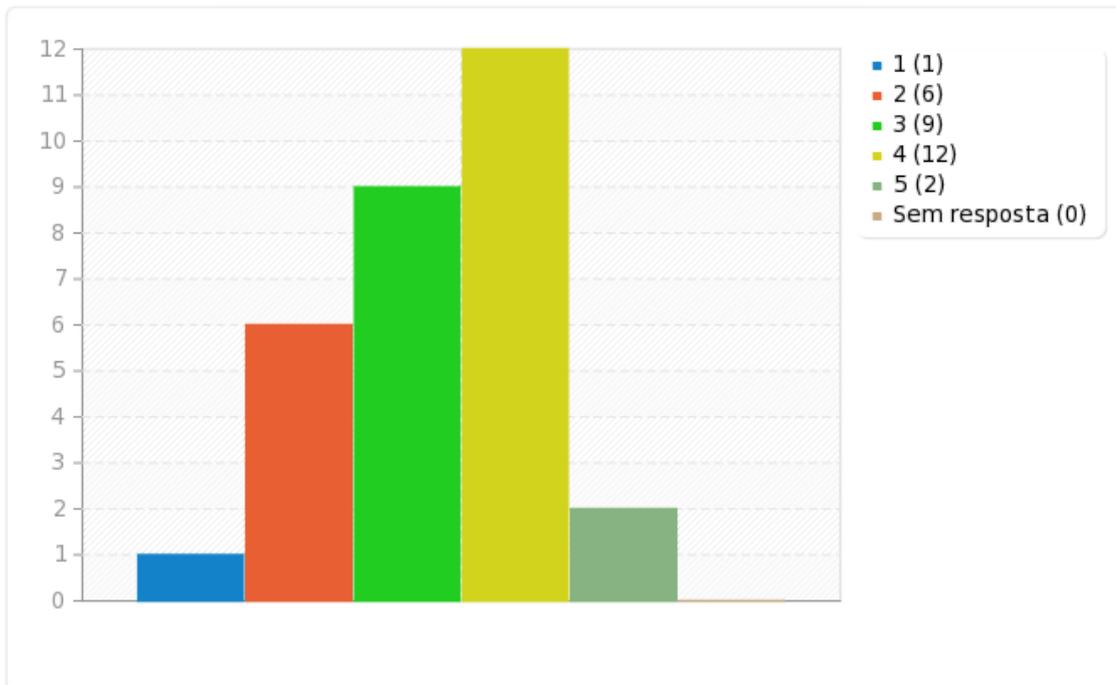
Resposta	Contagem	Percentagem	Soma
1 (1)	0	0.00%	10.00%
2 (2)	3	10.00%	
3 (3)	3	10.00%	10.00%
4 (4)	12	40.00%	
5 (5)	12	40.00%	80.00%
<b>Soma (Respostas)</b>	<b>30</b>	<b>100.00%</b>	<b>100.00%</b>
Número de casos	30	100.00%	
Sem resposta	0	0.00%	
Significado aritmético		4.1	
Desvio padrão		0.96	



### Sumário dos campos para C

**Pelotões da Polícia Militar Ambiental para Unidades de Conservação 2015. Dados contendo as unidades de conservação do estado de São Paulo que possuem pelotões da Polícia Militar Ambiental.**

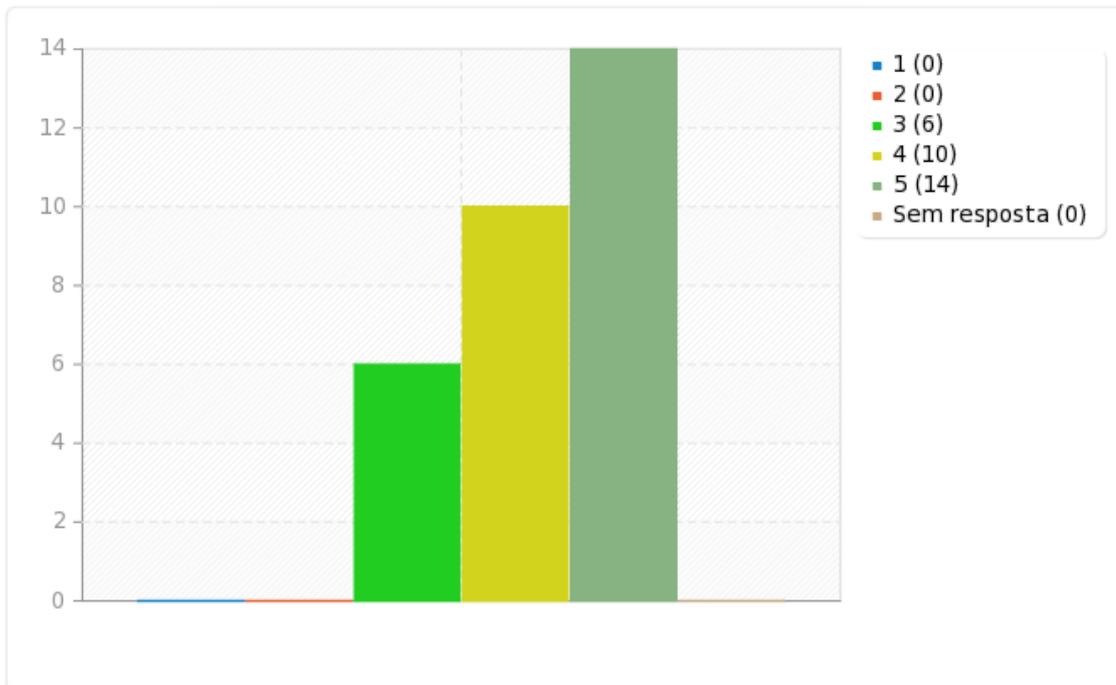
Resposta	Contagem	Porcentagem	Soma
1 (1)	1	3.33%	23.33%
2 (2)	6	20.00%	
3 (3)	9	30.00%	30.00%
4 (4)	12	40.00%	
5 (5)	2	6.67%	46.67%
<b>Soma (Respostas)</b>	<b>30</b>	<b>100.00%</b>	<b>100.00%</b>
Número de casos	30	100.00%	
Sem resposta	0	0.00%	
Significado aritmético		3.27	
Desvio padrão		0.98	



### Sumário dos campos para B

Unidades de proteção integral. Dados fornecidos pelo IBAMA contendo áreas de proteção integral, nome, sigla, área oficial, perímetro, ano de criação, decreto que instituiu e órgão coordenador.

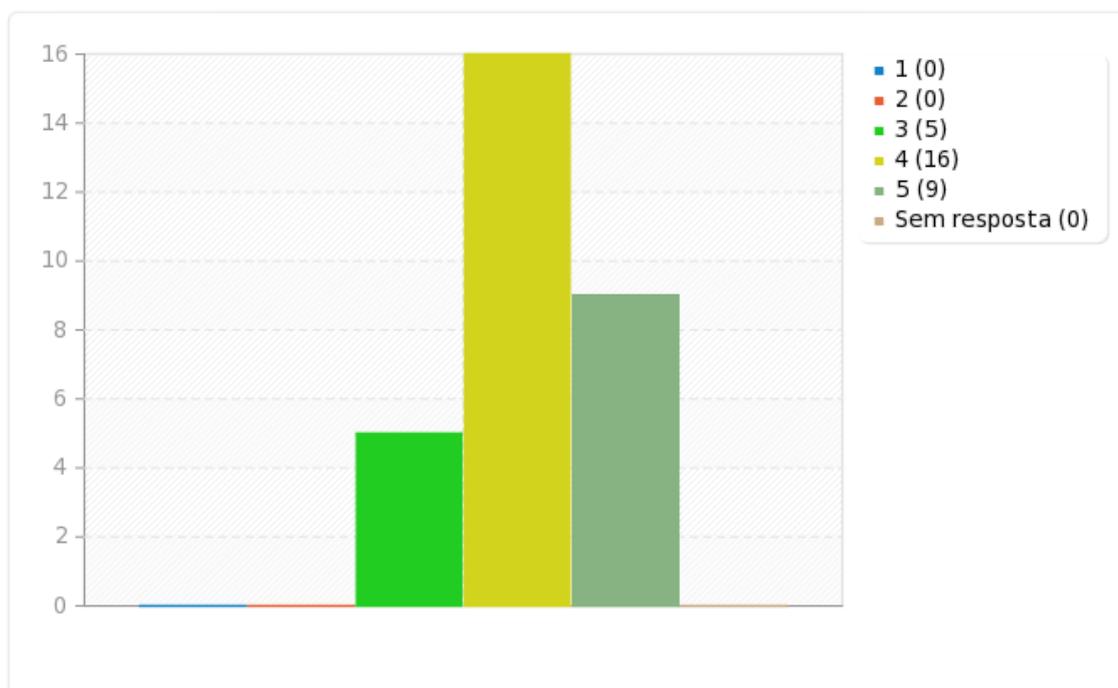
Resposta	Contagem	Porcentagem	Soma
1 (1)	0	0.00%	0.00%
2 (2)	0	0.00%	0.00%
3 (3)	6	20.00%	20.00%
4 (4)	10	33.33%	33.33%
5 (5)	14	46.67%	46.67%
<b>Soma (Respostas)</b>	<b>30</b>	<b>100.00%</b>	<b>100.00%</b>
Número de casos	30	100.00%	
Sem resposta	0	0.00%	
Significado aritmético		4.27	
Desvio padrão		0.78	



### Sumário dos campos para H

Unidades de Conservação de Proteção Integral 2013. Dados contendo as unidades de conservação de proteção integral organizadas por estados e municípios.

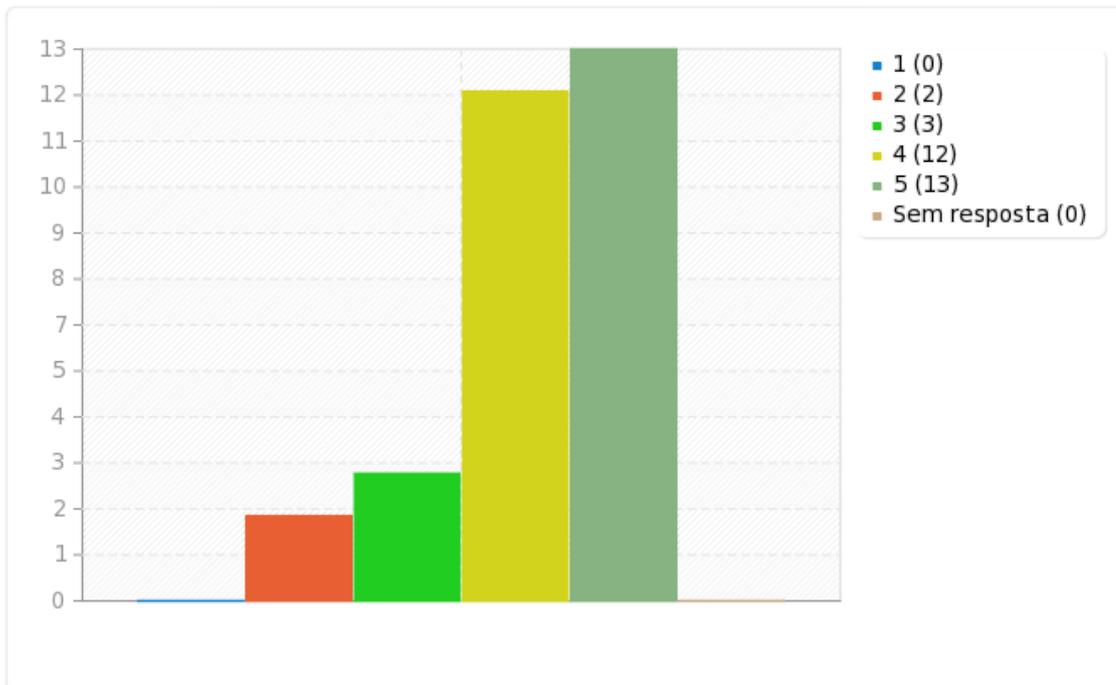
Resposta	Contagem	Porcentagem	Soma
1 (1)	0	0.00%	0.00%
2 (2)	0	0.00%	
3 (3)	5	16.67%	16.67%
4 (4)	16	53.33%	
5 (5)	9	30.00%	83.33%
<b>Soma (Respostas)</b>	<b>30</b>	<b>100.00%</b>	<b>100.00%</b>
Número de casos	30	100.00%	
Sem resposta	0	0.00%	
Significado aritmético		4.13	
Desvio padrão		0.68	



### Sumário dos campos para J

Áreas especiais de proteção e conservação da biodiversidade. Dados contendo as áreas especiais de proteção e conservação da biodiversidade, informações de planejamento, bioma pertencente, relevância e priorização.

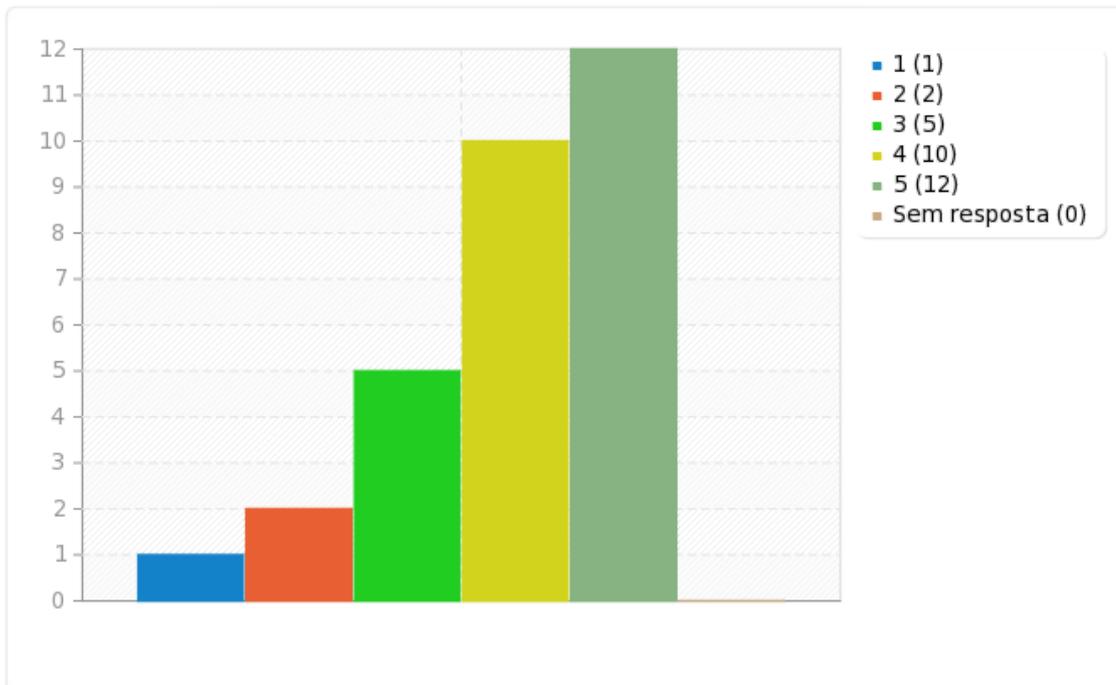
Resposta	Contagem	Porcentagem	Soma
1 (1)	0	0.00%	6.67%
2 (2)	2	6.67%	
3 (3)	3	10.00%	10.00%
4 (4)	12	40.00%	
5 (5)	13	43.33%	83.33%
<b>Soma (Respostas)</b>	<b>30</b>	<b>100.00%</b>	<b>100.00%</b>
Número de casos	30	100.00%	
Sem resposta	0	0.00%	
Significado aritmético		4.2	
Desvio padrão		0.89	



### Sumário dos campos para G

Densidade de focos de calor por 1000 km<sup>2</sup> - unidades de conservação e terras indígenas. Dados contendo focos de calor localizados em unidades de conservação ambiental e terras indígenas.

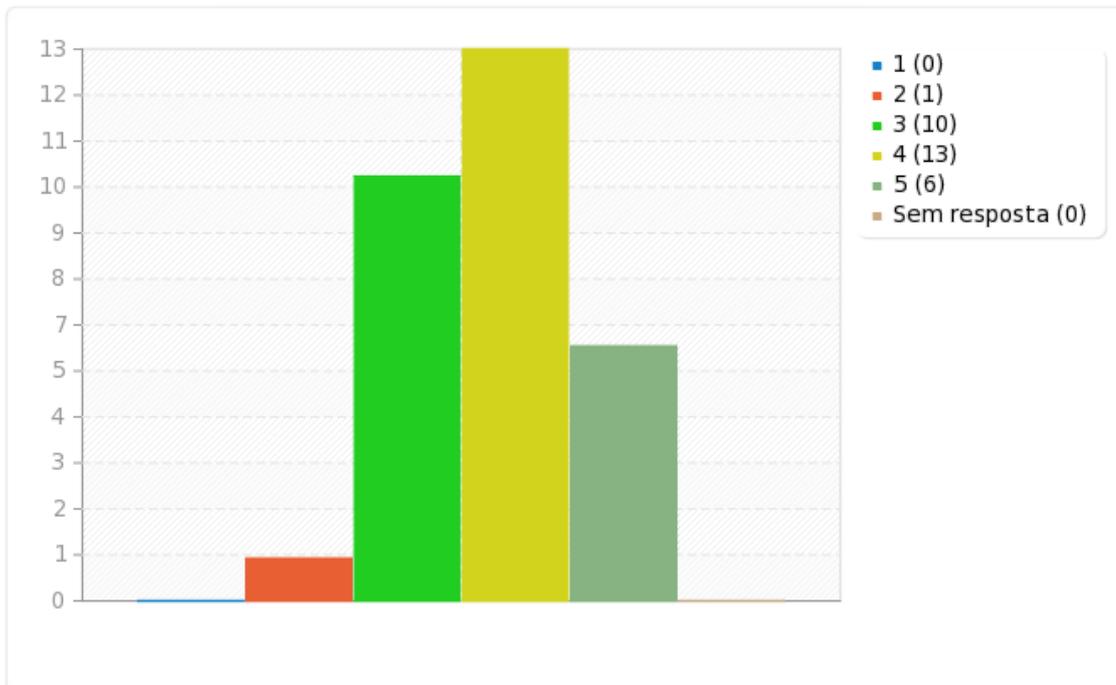
Resposta	Contagem	Porcentagem	Soma
1 (1)	1	3.33%	10.00%
2 (2)	2	6.67%	16.67%
3 (3)	5	16.67%	73.33%
4 (4)	10	33.33%	100.00%
5 (5)	12	40.00%	140.00%
<b>Soma (Respostas)</b>	<b>30</b>	<b>100.00%</b>	<b>100.00%</b>
Número de casos	30	100.00%	
Sem resposta	0	0.00%	
Significado aritmético		4	
Desvio padrão		1.08	



### Sumário dos campos para I

**Unidades de Conservação de Proteção Integral 2013. Dados contendo as unidades de conservação de proteção integral com os decretos que as instituíram e as instituições ou órgãos responsáveis.**

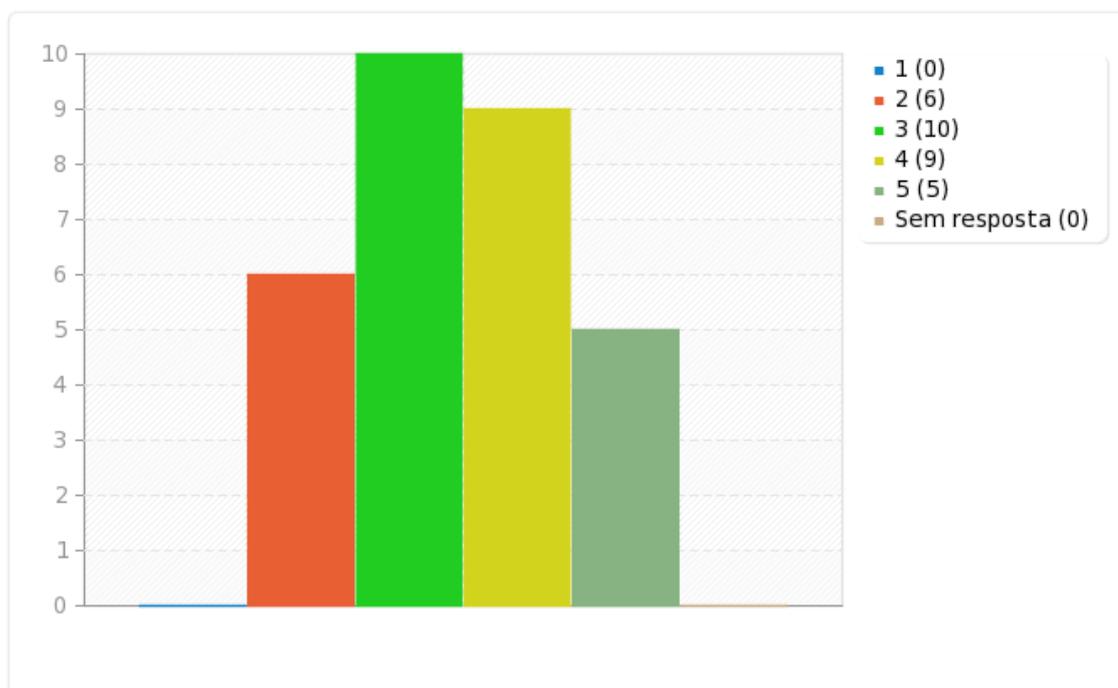
Resposta	Contagem	Porcentagem	Soma
1 (1)	0	0.00%	3.33%
2 (2)	1	3.33%	
3 (3)	10	33.33%	33.33%
4 (4)	13	43.33%	
5 (5)	6	20.00%	63.33%
<b>Soma (Respostas)</b>	<b>30</b>	<b>100.00%</b>	<b>100.00%</b>
Número de casos	30	100.00%	
Sem resposta	0	0.00%	
Significado aritmético		3.8	
Desvio padrão		0.81	



### Sumário dos campos para E

**Indicadores de Desenvolvimento Sustentável 2015 - Demanda Bioquímica de Oxigênio (DBO) - Região Sul. Demanda Bioquímica de Oxigênio (DBO) dos pontos de monitoramento em rios da Região Sul - último ano da informação, em alguns trechos selecionados de rios desta região.**

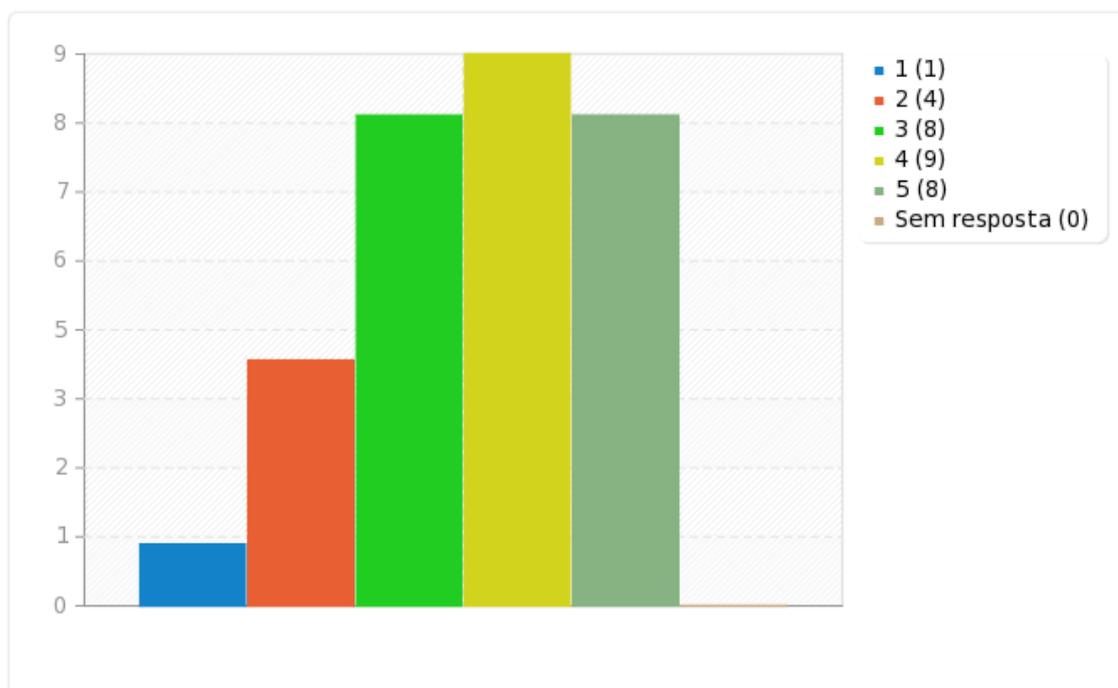
Resposta	Contagem	Porcentagem	Soma
1 (1)	0	0.00%	20.00%
2 (2)	6	20.00%	
3 (3)	10	33.33%	33.33%
4 (4)	9	30.00%	
5 (5)	5	16.67%	46.67%
<b>Soma (Respostas)</b>	<b>30</b>	<b>100.00%</b>	<b>100.00%</b>
Número de casos	30	100.00%	
Sem resposta	0	0.00%	
Significado aritmético		3.43	
Desvio padrão		1.01	



### Sumário dos campos para D

Indicadores de Desenvolvimento Sustentável 2015 - Classe de periculosidade ambiental Dados contendo índices por estado da federação, graduando o nível de periculosidades de comercialização de agrotóxicos e produtos afins.

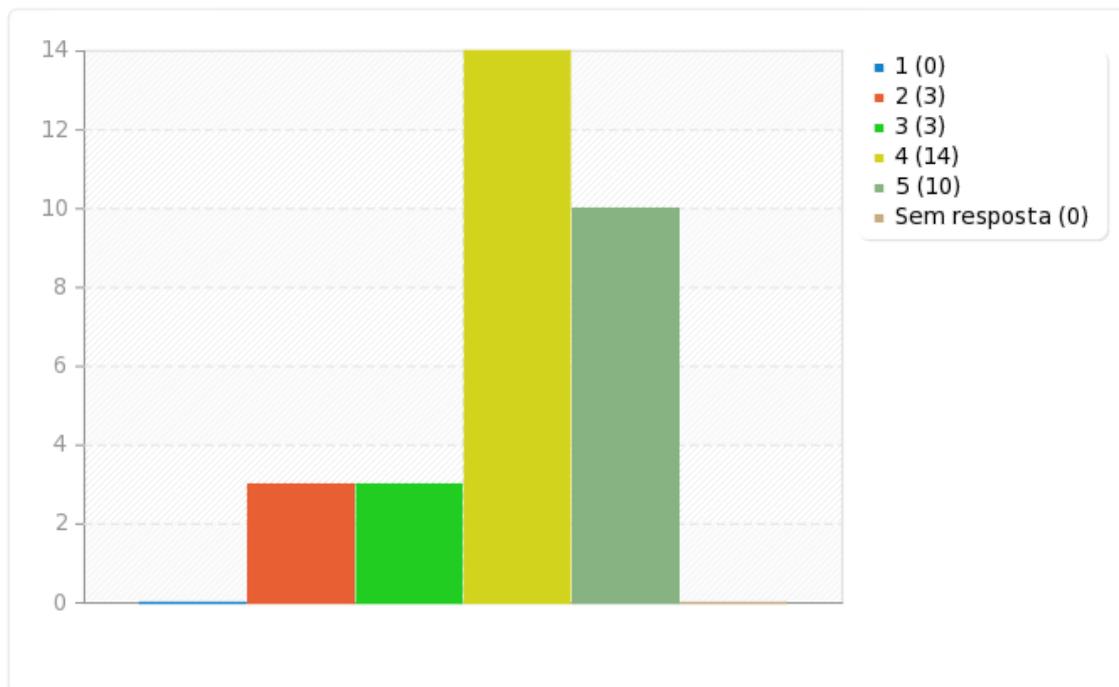
Resposta	Contagem	Porcentagem	Soma
1 (1)	1	3.33%	16.67%
2 (2)	4	13.33%	
3 (3)	8	26.67%	26.67%
4 (4)	9	30.00%	
5 (5)	8	26.67%	56.67%
<b>Soma (Respostas)</b>	<b>30</b>	<b>100.00%</b>	<b>100.00%</b>
Número de casos	30	100.00%	
Sem resposta	0	0.00%	
Significado aritmético		3.63	
Desvio padrão		1.13	



### Sumário dos campos para F

Áreas especiais de proteção e conservação da biodiversidade. Dados contendo as áreas especiais de proteção e conservação contendo nome, sigla, UF, bioma.

Resposta	Contagem	Percentagem	Soma
1 (1)	0	0.00%	10.00%
2 (2)	3	10.00%	
3 (3)	3	10.00%	10.00%
4 (4)	14	46.67%	
5 (5)	10	33.33%	80.00%
<b>Soma (Respostas)</b>	<b>30</b>	<b>100.00%</b>	<b>100.00%</b>
Número de casos	30	100.00%	
Sem resposta	0	0.00%	
Significado aritmético		4.03	
Desvio padrão		0.93	





LimeSurvey  
Versão 2.00+ Build 130929

