Ministério da
Ciência, Tecnologia
e Inovação

sid.inpe.br/mtc-m19/2013/03.29.14.28-TDI

UMA ARQUITETURA PARA GESTÃO DE CONHECIMENTO CIENTÍFICO SENSÍVEL AO CONTEXTO GEOGRÁFICO

Alaor Bianco Rodrigues

Dissertação de Mestrado do Curso de Pós-Graduação em Computação Aplicada, orientada pelos Drs. Walter Abrahão Santos, e Sidnei João Siqueira Sant'Anna, aprovada em 29 de abril de 2013.

URL do documento original:
<http://urlib.net/8JMKD3MGP7W/3DQMCRS>

INPE
São José dos Campos
2013

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE
Gabinete do Diretor (GB)
Serviço de Informação e Documentação (SID)
Caixa Postal 515 - CEP 12.245-970
São José dos Campos - SP - Brasil
Tel.:(012) 3208-6923/6921
Fax: (012) 3208-6919
E-mail: pubtc@sid.inpe.br

CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELECTUAL DO INPE (RE/DIR-204):
Presidente:
Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)
Membros:
Dr. Antonio Fernando Bertachini de Almeida Prado - Coordenação Engenharia e Tecnologia Espacial (ETE)
Drª Inez Staciarini Batista - Coordenação Ciências Espaciais e Atmosféricas (CEA)
Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)
Dr. Germano de Souza Kienbaum - Centro de Tecnologias Especiais (CTE)
Dr. Manoel Alonso Gan - Centro de Previsão de Tempo e Estudos Climáticos (CPT)
Drª Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação
Dr. Plínio Carlos Alvalá - Centro de Ciência do Sistema Terrestre (CST)
BIBLIOTECA DIGITAL:
Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)
REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:
Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)
Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)
EDITORAÇÃO ELETRÔNICA:
Maria Tereza Smith de Brito - Serviço de Informação e Documentação (SID)
Luciana Manacero - Serviço de Informação e Documentação (SID)

Ministério da
Ciência, Tecnologia
e Inovação

sid.inpe.br/mtc-m19/2013/03.29.14.28-TDI

UMA ARQUITETURA PARA GESTÃO DE CONHECIMENTO CIENTÍFICO SENSÍVEL AO CONTEXTO GEOGRÁFICO

Alaor Bianco Rodrigues

Dissertação de Mestrado do Curso de Pós-Graduação em Computação Aplicada, orientada pelos Drs. Walter Abrahão Santos, e Sidnei João Siqueira Sant'Anna, aprovada em 29 de abril de 2013.

URL do documento original:
<http://urlib.net/8JMKD3MGP7W/3DQMCRS>

INPE
São José dos Campos
2013

Dados Internacionais de Catalogação na Publicação (CIP)

Rodrigues, Alaor Bianco.
R618a Uma arquitetura para gestão de conhecimento científico sensível ao contexto geográfico / Alaor Bianco Rodrigues. – São José dos Campos : INPE, 2013.
xxvi + 139 p. ; (sid.inpe.br/mtc-m19/2013/03.29.14.28-TDI)

Dissertação (Mestrado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2013.
Orientadores : Drs. Walter Abrahão Santos, e Sidnei João Siqueira Sant'Anna.

1. gestão do conhecimento. 2. dados espaciais. 3. IDE. 4. OGC. 5. SOA. 6. GeoPortal. I.Título.

CDU 004.774.6

Esta obra foi licenciada sob uma Licença Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada.

This work is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported License.

Aprovado (a) pela Banca Examinadora em cumprimento ao requisito exigido para obtenção do Título de *Mestre* em

Computação Aplicada

Dr. Eduardo Martins Guerra

Presidente / INPE / São José dos Campos - SP

Dra. Corina da Costa Freitas

Orientador(a) / INPE / SJCampos - SP

Dr. Walter Abrahão dos Santos

Orientador(a) / INPE / São José dos Campos - SP

Dr. Sidnei João Siqueira Sant'Anna

Orientador(a) / INPE / São José dos Campos - SP

Dr. Elcio Hideiti Shiguemori

Convidado(a) / IEAv/CTA / São José dos Campos - SP

Dr. Valdir Gil Pilat

Convidado(a) / UNIVAP / São José dos Campos - SP

Este trabalho foi aprovado por:

( ) maioria simples

(X) unanimidade

Aluno (a): *Alaor Bianco Rodrigues*

“Se dois homens vêm andando por uma estrada, cada um com um pão, e, ao se encontrarem, trocarem os pães, cada um vai embora com um.
Se dois homens vêm andando por uma estrada, cada um com uma idéia, e, ao se encontrarem, trocarem as idéias, cada um vai embora com duas”.

Provérbio chinês

À minha amada esposa, meus pais e irmãos que trilharam este caminho a meu lado.

AGRADECIMENTOS

Agradeço a Deus, em quem creio e confio, pelas oportunidades e pela sabedoria em aproveitá-las, por colocar em meu caminho situações e pessoas que tornaram este momento da minha vida possível.

Agradeço a minha família, minha esposa, pais e irmãos que sempre souberam entender meus momentos de ausência e que mesmo nestes, sempre estiveram ao meu lado, seja motivando ou confortando-me.

Agradeço aos meus amigos e colegas do INPE com os quais pude dividir várias conquistas e aprendizados, e aqueles que sempre estiveram prontos para me ouvir, orientar e dividir seus conhecimentos.

Agradeço aos meus orientadores e professores que sempre foram muito mais do que professores, foram amigos e verdadeiros mestres que conduzem seus discípulos não pelo caminho mais fácil, mas pelo caminho correto e os ensinam não somente fórmulas e métodos, mas os ensinam a vencer.

Por fim, agradeço ao Instituto Nacional de Pesquisas Espaciais pela oportunidade de cursar este mestrado, preparando-me para aproveitar as oportunidades que a vida oferece ao colocar em movimento os conhecimentos adquiridos.

RESUMO

Neste trabalho é explorado o contexto de e-Science, onde o apoio computacional torna o fluxo de trabalho de pesquisa científica mais eficiente. É abordada a problemática da ingerência de uma parcela do conhecimento produzido durante um trabalho de pesquisa, a falta da visão geoespacial destes trabalhos e dos artefatos que o compõe. Uma arquitetura para a gestão dos trabalhos científicos é proposta organizando-os segundo sua componente espacial e criando mecanismos para externalizar a parcela do conhecimento que, em geral, é ignorada ou difícil de ser comunicar por meios formais. É abordada a implementação da arquitetura proposta, que em linhas gerais emprega o paradigma de componentes distribuídos numa arquitetura orientada a serviços. A arquitetura contém um componente cliente que orquestra os serviços especializados dos demais componentes da solução de forma interoperável. Para avaliar a arquitetura proposta foi conduzido um estudo de caso, considerando um trabalho de pesquisa, que utiliza e produz vários artefatos multimídias e espaciais. A região escolhida para o estudo de caso foi a do estado do Pará, especificamente a Floresta Nacional do Tapajós e arredores. Como resultado foi gerado um software aplicado à gestão de conhecimento capaz de lidar com uma gama de artefatos multimídias e geoespaciais, bem como algoritmos de processamento seguindo padrões da Open Geospatial Consortium.

An Architecture for Scientific Knowledge Management of Geospatially Enabled Artifacts

ABSTRACT

This work explores the context of e-Science, where a computational support makes scientific workflow more efficient. The problem of mismanagement of the knowledge produced during a research is addressed as well as their lack of geospatial vision. An architecture for the management of scientific works is proposed organizing them according to their spatial component and creating mechanisms to imerge part of knowledge that is generally ignored or difficult to communicate through formal means. The implementation of the proposed architecture is built upon the paradigm of distributed components in a service-oriented architecture. This solution has a client component that orchestrates specialized services from the others solution components in an interoperable way. A case study evaluates the proposed architecture using and generating various multimedia and geospatial artifacts from a previous research undergone. The territorial area chosen is in the Brazilian Pará state, specifically the Tapajós National Forest and surrounding areas. As a result a software application has been written for managing knowledge capable to deal with a variety of multimedia and geospatial artifacts as well as processing algorithms per the Open Geospatial Consortium standards.

LISTA DE FIGURAS

Pág.

Figura 3.1: Espiral do conhecimento. 16
Figura 4.1: Modelo de domínio do padrão CMIS. 21
Figura 4.2: Serviços *web* do padrão CMIS. 21
Figura 5.1: Arquitetura SOA. 32
Figura 5.2: Diagrama de sequência WMS. 38
Figura 5.3: Diagrama de Sequência WFS. 40
Figura 5.4: Diagrama de Sequência WCS. 42
Figura 5.5: SOA Geoespacial. 51
Figura 6.1: Abstração da abordagem componentizada. 55
Figura 6.2: Visão de componentes da arquitetura. 56
Figura 6.3: Teste de desempenho entre Mapserver e Geoserver. 61
Figura 7.1: Diagrama macro da solução. 66
Figura 7.2: Macro arquitetura proposta. 72
Figura 7.3: Exemplo de estrutura de uma pesquisa de mestrado. 75
Figura 7.4: Camadas especiais. 76
Figura 7.5: Representação do Serviço de Relacionamento GIS – CMS. 78
Figura 7.6: Estrutura da aplicação cliente. 83
Figura 7.7: Widget. 84
Figura 7.8: Módulo de filtragem digital de imagem. 86
Figura 7.9: Módulo de busca no GeoPortal. 87
Figura 8.1: Mapas disponíveis no IBGE. 96
Figura 8.2: Inclusão dinâmica de serviços de solos. 97
Figura 8.3: Inclusão dinâmica de serviços hidrografia. 97
Figura 8.4: Adição de dado matricial a visualização atual do mapa. 98
Figura 8.5: Busca de dados por extensão espacial. 99
Figura 8.6: Filtro de uma imagem da região de interesse. 101
Figura 8.7: Vinculando dados matriciais. 103
Figura 8.8: Vinculando dados vetoriais. 104
Figura 8.9: Vinculando os resultados. 105
Figura 8.10: Vínculo entre área de interesse e conteúdo do CMS. 106
Figura 8.11: Relacionando um conteúdo com um ponto de interesse. 107
Figura 8.12: Busca por conteúdo relacionado. 109
Figura 8.13: Recuperar itens relacionados. 110
Figura 8.14: Recuperar resultados. 110

LISTA DE TABELAS

Pág.

Tabela 3.1: Dado, informação e conhecimento. 14
Tabela 4.1: Exemplo de alguns metadados comuns. 23
Tabela 5.1: Operações WMS. 37
Tabela 5.2: Operações WFS. 39
Tabela 5.3: Operações WCS 41
Tabela 5.4: Operações CSW 43
Tabela 5.5: Operações WPS 44
Tabela 7.1: Requisitos exigidos para a solução. 66
Tabela 7.2: Operações do Serviço de Relacionamento GIS – CMIS. 78
Tabela 7.3: Serviços e seus respectivos databases 82
Tabela A.1: Serviços disponíveis no padrão CMIS. 125

LISTA DE SIGLAS E ABREVIATURAS

ANA	Agência Nacional de Águas
ANEEL	Agência Nacional de Energia Elétrica
API	*Application programming interface*
CMIS	*Content Management Interoperability Services*
CENAPAD	Centros Nacionais de Processamento de Alto Desempenho
CMS	*Content Management System*
CONCAR	Comissão Nacional de Cartografia
CPRM	Companhia de Pesquisa de Recursos Minerais
CSW	*Catalogue Service for the Web*
DBDG	Diretório Brasileiro de Dados Geoespaciais
DCMI	*Dublin Core Metadata Initiative*
DM	*Document Management*
DPI	Divisão de Processamento de Imagens
DSG	Diretoria de Serviço Geográfico do Exército Brasileiro
ECM	*Enterprise Content Management*
ECMS	*Enterprise Content Management System*
GDAL	*Geospatial Data Abstraction Library*
GIF	*Graphics Interchange Format*
GIS	*Geographic Information System*
GML	*Geography Markup Language*
GPS	*Global Positioning System*
GSDI	*Global Spatial Data Infrastructure Association*
HTTP	*Hypertext Transfer Protocol*
IBGE	Instituto Brasileiro de Geografia e Estatística
IDE	Infraestruturas de dados espaciais
INDE	Infraestrutura Nacional de Dados Espaciais
INPA	Instituto Nacional de Pesquisas da Amazônia
INPE	Instituto Nacional de Pesquisas Espaciais
INSPIRE	*Infrastructure for Spatial Information in the European*

	Community
JPEG	*Joint Photographic Experts Group*
KML	*Keyhole Markup Language*
LBA	*Large Scale Biosphere-Atmosphere Experiment in Amazonia*
MCT	Ministério da Ciência e Tecnologia
MGB	Metadados Geoespaciais do Brasil
MMA	Ministério do Meio Ambiente
NOAA	National Oceanic and Atmospheric Administration
OASIS	*Organization for the Advancement of Structured Information Standards*
OGC	*Open Geospatial Consortium*
OSGEO	*The Open Source Geospatial Foundation*
PDF	*Portable document format*
PNG	*Portable Network Graphics*
RADAR	*Radio Detection And Ranging*
REST	*Representational State Tranfer*
ROI	*Region of Interest*
SAR	*Synthetic Aperture RADAR*
SFSQL	*Simple Features for SQL*
SGBD	Sistemas Gerenciadores de Bancos de Dados
SIG	Sistema de Informação Geográfica
SINAPAD	Sistema Nacional de Processamento de Alto Desempenho
SLD	*Styled Layer Descriptor*
SOA	*Service Oriented Architecture*
SOAP	*Simple Object Access Protocol*
SQL	*Structured Query Language*
SVG	*Scalable Vector Graphics*
UDDI	*Universal Description, Discovery and Integration*
URI	*Uniform Resource Identifier*
URL	*Uniform Resource Locator*
W3C	*World Wide Web Consortium*

WCS	*Web Coverage Service*
WFS	*Web Feature Service*
WKB	*Well Known Binary*
WKT	*Well Known Text*
WMC	*Web Map Context*
WMS	*Web Map Service*
WPS	*Web Processing Service*
WS-Common	*OpenGIS Web Service Common Implementation Specification*
WSDL	*Web Service Description Language*
XML	*Extensible Markup Language*
UFRGS	Universidade Federal do Rio Grande do Sul
UFMG	Universidade Federal de Minas Gerais
UFC	Universidade Federal do Ceará
UNICAMP	Universidade Estadual de Campinas
UFRJ	Universidade Federal do Rio de Janeiro
UFPE	Universidade Federal de Pernambuco
LNCC	Laboratório Nacional de Computação Científica

SUMÁRIO

Pág.

1 INTRODUÇÃO 1
1.1. Motivação 4
1.2. Objetivo 5
1.3. Contribuições 6
1.4. Aspecto Inovador 6
1.5. Organização do trabalho 7
2 HISTÓRICO DE *E-SCIENCE* 9
2.1. Histórico de *e-Science* no Brasil 9
2.2. Trabalhos Relacionados 11
2.3. Considerações finais do capítulo 12
3 O CONHECIMENTO E SUA GESTÃO 13
3.1. Gestão do Conhecimento 13
3.2. Computação como Infraestrutura para Gestão do Conhecimento 17
3.3. Considerações finais do capítulo 18
4 SISTEMA DE GERENCIAMENTO DE CONTEÚDO 19
4.1. Sistema de Gerenciamento de Conteúdo Corporativo 19
4.2. Serviço Interoperável para Gestão de Conteúdo 20
4.3. Metadado 22
4.3.1. Dublin Core 24
4.4. Considerações finais do capítulo 27
5 ARQUITETURA DE DADOS ESPACIAIS 29
5.1. Arquitetura Orientada a Serviços 29
5.2. Interoperabilidade 32
5.3. *Open Geospatial Consortium* 34
5.3.1. *Geography Markup Language* 35
5.3.2. OGC *Web Services* 36
5.3.3. WS-Common 36
5.3.4. *Web Map Service* 37
5.3.5. *Web Feature Service* 39
5.3.6. *Web Coverage Service* 40
5.3.7. *Catalog Service for the Web* 42

5.3.8. *Web Processing Service* 44
5.4. Arquitetura de Dados Espacial 45
5.5. Infraestrutura de Dados Espaciais 46
5.6. Servidor de Mapas 48
5.7. GeoPortal 49
5.8. Considerações finais do capítulo 51
6 PROJETO ARQUITETURAL 53
6.1. Definição do Problema 53
6.2. Projeto Arquitetural 54
6.3. Infraestrutura Necessária 56
6.3.1. Postgres 57
6.3.2. GeoServer 58
6.3.3. Esri GeoPortal Server 61
6.3.4. Alfresco 63
6.4. Considerações finais do capítulo 64
7 ARQUITETURA PROPOSTA 65
7.1. Visão geral 65
7.2. Arquitetura Proposta 70
7.3. Realização dos Requisitos 72
7.4. Serviço CMS 73
7.5. Serviço de Conteúdo Geográfico 75
7.6. Serviço Relacionamento GIS – CMS 77
7.7. GeoPortal Server 80
7.8. PostGIS 81
7.9. Aplicação Cliente 82
7.9.1. Módulos 85
7.10. Considerações finais do capítulo 87
8 ESTUDO DE CASO 89
8.1. Região da Floresta Nacional do Tapajós 89
8.2. Escolha do trabalho de pesquisa para o estudo de caso 92
8.3. Planejamento do estudo de caso 93
8.4. Início de um trabalho de pesquisa 94
8.5. Descoberta de dados e análises de regiões 95
8.6. Serviço *Web* de Processamento 100

8.7. Gestão dos arquivos multimídia e dados espaciais 102
8.8. Espacialização de conteúdo 105
8.9. Recuperando o contexto de uma pesquisa 107
8.10. Considerações finais do capítulo 111
9 CONCLUSÕES 113
9.1. Considerações finais 113
9.2. Contribuições 114
9.3. Sugestões para trabalhos futuros 115
REFERÊNCIAS BIBLIOGRÁFICAS 117
APÊNDICE A – Serviços disponíveis no padrão CMIS 125
APÊNDICE B – Termos comuns utilizados na especificação OWS 129
APÊNDICE C – Artigo apresentado. 131

1 INTRODUÇÃO

A tecnologia da informação vem revolucionando o modo como diversas áreas da ciência e engenharia estão sendo conduzidas. A utilização de recursos computacionais no desenvolvimento da pesquisa beneficia o trabalho das comunidades científicas facilitando o compartilhamento de dados e serviços computacionais, além de contribuir para a construção de uma infraestrutura de dados e de uma comunidade científica distribuída (LUDASCHER et al., 2006). A computação vem se apresentando como uma terceira abordagem para a pesquisa, em complemento às tradicionais teórica e experimental. Simulações computacionais apresentam-se como um novo ator no processo científico, sendo utilizadas em diversas áreas onde a observação de fenômenos se mostra dispendiosa, como no caso de perfuração de poços de petróleo, ou invasivas, como no caso da neurociência. Em ambos os casos, o desenvolvimento da Ciência da Computação no suporte à Ciência se mostra fundamental e indiscutível (CSBC, 2012).

O contexto em que a computação se torna parte integrante e imprescindível para o sucesso na realização de pesquisas científicas das mais variadas áreas, é o contexto de *e-Science*, em que a ciência é realizada com o apoio computacional, se tornando assim, mais eficiente. O termo *e-Science* geralmente é empregado para descrever o desenvolvimento de infraestruturas de serviços de *software* capazes de prover acesso a facilidades remotas, recursos computacionais distribuídos, incluindo ontologias, armazenamento de informações em bancos de dados, bioinformática, modelagem de *workflows* e processos científicos, serviços *web* semânticos, gerência de componentes de *software*, computação em nuvem e em grade, disseminação e compartilhamento de dados, resultados e conhecimento. Em geral, um ambiente de *e-Science* permite o compartilhamento de recursos entre comunidades de indivíduos, grupos, laboratórios e instituições, possibilitando o gerenciamento cooperativo de facilidades como equipamentos, instrumentação, experimentos entre outros. Permite, ainda, análise colaborativa de produtos (dados) oriundos dessas facilidades. São características de *e-Science* o acesso

a uma vasta coleção de dados, utilização de recursos computacionais em larga escala, utilização de recursos heterogêneos e dinâmicos de múltiplas organizações e *workflows* (CARDOSO, 2007; MATOS et al., 2007).

O volume de conhecimento produzido no meio científico concede destaque para os meios de gerenciá-los. Há grandes preocupações em como armazená-lo, distribuí-lo, evidenciar e informar a sua existência. A organização deste conhecimento, ou os produtos gerados de suas análises, segundo sua componente geográfica vêm ganhando destaque a cada dia em diversas áreas. A adição da componente geográfica permite uma grande flexibilidade em possibilitar a busca por dados relevantes segundo sua localização no globo terrestre.

A proposta de organizar os diferentes dados existentes segundo sua componente geográfica remete ao uso de tecnologias que permitam o manuseio de uma grande quantidade de informações de diversas naturezas, organizando-as convenientemente a fim de permitir que diferentes interações possam ser realizadas. Tais tecnologias, que permitem organizar computacionalmente dados e informações geográficas são partes integrantes de uma disciplina emergente, o Geoprocessamento.

As ferramentas computacionais para Geoprocessamento constituem os chamados Sistemas de Informação Geográfica (SIG). Os SIGs são sistemas computacionais que permitem a captura, armazenamento, recuperação, análise e apresentação de dados referenciados geograficamente (WORBOYS, 1995). Portanto permitem realizar análises complexas, ao integrar dados do mundo real, obtidos de diversas fontes e de diferentes formatos, compondo um banco de dados geográficos.

De uma forma geral, um sistema de informação geográfica é composto por quatro componentes principais. Estes componentes são denominados por componente de captura dos dados, componente de armazenamento, componente de análise e componente de apresentação dos dados (CAMARA et al., 2001).

O componente de captura de dados é responsável pela entrada e integração de dados possibilitando a obtenção dos dados através de diversos métodos como leitura ótica, digitalização de mapas ou aquisição via meio magnético. O componente responsável pelo armazenamento e recuperação de dados geográficos provê as estruturas de dados que possibilitam a compactação de imagens, armazenamento de relacionamentos espaciais (topologia), acesso aos dados através de índices espaciais, etc. O conjunto das funções de processamento compõe o componente de análise, que é o componente que mais se diferencia de sistema para sistema, porém, existe um grande número de funções que são comuns a esses sistemas (CAMARA et al., 2001).

A maioria dos dados gerados por uma aplicação de SIG são apresentados na forma de mapas. Desta maneira, um SIG deve prover funções que são equivalentes àquelas dos sistemas de cartografia automatizada, como colocação automática de rótulos, descrição de legendas e escalas gráficas, entre outras, formando assim o componente de apresentação dos dados (CAMARA et al., 2001).

Apenas organizar os dados segundo sua componente espacial não é suficiente. Deveria haver um modo de compartilhar a maior parte possível do conhecimento gerado durante uma pesquisa. Muitos esforços têm sido empregados para compartilhar dados, análises e resultados a cerca de pesquisas. A era da informação e a internet trazem grandes facilidades para que pesquisadores possam compartilhar seus avanços e se atualizar sobre pesquisas similares ao redor do mundo, mas o formato do que é publicado restringe o quanto do conhecimento que pode ser transferido (LEITE, 2006). A publicação científica, diz respeito a uma parcela de todo o conhecimento e experiência que pode ser explicitado e comunicado por meios formais (LEITE, 2006).

Nos estudos de Leite (2006) conclui-se que boa parte do conhecimento científico explícito é fruto da externalização de uma parcela daquilo que se criou e aprendeu durante a realização de uma pesquisa. Essa parcela diz

respeito ao conhecimento tácito que é possível de ser explícita, ou seja, reduzido à informação para poder ser efetivamente e amplamente comunicado.

Muito comumente, há vários tipos de informações que são produzidas ao longo de uma pesquisa ou estudo. Estas informações ou produtos destas pesquisas e estudos encontram-se espalhadas em diversos locais, sejam eles repositórios conhecidos, agregadores ou no pior dos casos em computadores pessoais. No entanto, estes artefatos não são publicados ou compartilhados com toda, ou parte, da comunidade científica. Estes podem ser artefatos importantes para a criação do conhecimento ou compõe estruturas comunicáveis do conhecimento e que são passíveis de ser externalizadas.

Em cenários onde estas informações são compartilhadas uma mesma informação pode estar replicada em diversos locais, aumentando os esforços para mantê-la atualizada. A dificuldade em indexar ou desenvolver um mecanismo de busca que consiga cobrir todos os locais é também um desafio eminente e pode requerer muitos recursos, principalmente quando os formatos dos dados são heterogêneos, sendo compostos por planilhas eletrônicas, arquivos de texto, dados vetoriais, dados matriciais ou arquivos multimídia como fotos, áudios e vídeos.

1.1. Motivação

No início da criação de um novo conhecimento, o esforço de um pesquisador parte daquilo que foi construído anteriormente por outros pesquisadores. Isto é, geralmente recorre à literatura de sua especialidade, e, ao fim, divulga os resultados de sua pesquisa por meio dos veículos de comunicação apropriados à sua área de conhecimento (LEITE, 2006).

Percebe-se assim a importância da comunicação, informar ao mundo científico seus feitos, resultados e etc. Meadows (1999) diz que a comunicação reside no coração da ciência, sendo tão vital quanto à própria pesquisa. No entanto, apenas uma fração do que é produzido durante uma pesquisa é publicado, ou seja, é formalmente comunicado a comunidade. Braga (1985) ressalta que a

comunicação formal é responsável por apenas 20% de todas as comunicações no processo de geração do conhecimento. As demais são constituídas de processos informais, e uma grande parcela desse conhecimento encontra-se em um formato que poderia ser explícito, como anotações, planilha de resultados, registros de experimentos, resultados parciais, etc. Mesmo com o advento da era da informação e facilidades que a internet propicia, os argumentos de Braga (1985) se sustentam, pois de lá para cá o formato das publicações científicas são as mesmas e continuam representando apenas uma parcela do conhecimento, como discutido por Leite (2006).

Os recursos computacionais e o ambiente *web* muito contribuem para um cenário de compartilhamento e comunicação. Os recursos computacionais facilitam o trabalho em rede, podendo manter os conhecimentos descentralizados junto aos locais em que são criados e/ou utilizados com maior frequência e melhorando o grau de interatividade do usuário com os registros de conhecimentos (DAVENPORT; PRUSAK, 1998). A computação é efetivamente útil para a gestão do conhecimento, se for empregada utilizando-se uma interatividade humana sistemática (DAVENPORT, 2001).

É possível perceber que uma parcela do conhecimento produzido durante uma pesquisa não é publicada e uma parcela de todo esse conhecimento pode possuir relevância espacial. O uso e emprego de recursos computacionais podem auxiliar a gestão desta parcela do conhecimento e consequentemente apoiar o desenvolvimento de novas pesquisas e da própria ciência.

1.2. Objetivo

O objetivo principal deste trabalho é propor e desenvolver uma arquitetura capaz de auxiliar a organização de trabalhos científicos, provendo facilidades para armazenar e recuperar os artefatos produzidos durante as pesquisas segundo sua componente espacial.

A fim de manter viva a memória do que já foi produzido serão fornecidos meios para que a parcela do conhecimento tácito, produzido durante uma pesquisa e

que geralmente não são publicados por meios formais, sejam externalizados e posteriormente descobertos, recuperados e utilizados. Esta parcela de conhecimento pode ser externalizada sob um formato heterogêneo de dados, sendo esta arquitetura capaz de comportar dados de formatos variados provindos de diferentes fontes. Estes dados podem ser planilhas eletrônicas, arquivos de texto, dados vetoriais, dados matriciais ou arquivos multimídia como fotos, áudios e vídeos e poderão ser agrupados e pesquisados segundo sua componente geográfica.

1.3. Contribuições

As principais contribuições deste trabalho são:

- *Software* para gestão de conhecimento capaz de comportar formatos de dados variados e georreferenciá-los;
- Exemplo de orquestração de serviços distribuídos para compor uma solução de *software*;
- Exploração de Infraestruturas de Dados Espaciais e da Infraestrutura Nacional de Dados Espaciais e dos elementos que as compõem; e
- Comparação entre diversos *softwares* para Sistemas de Gestão de Conteúdo, Sistemas para Gestão de Conteúdo Geográfico, Sistemas Gerenciadores de Banco de Dados com Extensão Espacial e GeoPortais.

1.4. Aspecto Inovador

Para o cenário de uso abordado nesta dissertação, os dados geográficos podem já existir em diversas outras fontes ou nem sempre representá-los como uma geometria de ponto é suficiente. Em determinados momentos faz mais sentido representar um artefato como uma área, ou seja, através de uma geometria de polígono ou vários artefatos estão relacionados a uma mesma localidade. É também necessário considerar os dados matriciais e conseguir,

de alguma forma, relacionar estes dados a algum artefato. Em determinados momentos o conhecimento produzido pode ser a implementação de algum algoritmo de processamento e seria melhor distribuído se oferecido como um serviço de processamento, recebendo como entrada artefatos específicos, processando-os e devolvendo os resultados de interesse.

Estas especificidades inviabilizam o uso de alguma solução existente, fazendo-se necessário um estudo e implementação de uma solução que contemple os requisitos aqui abordados. O trabalho desenvolvido nesta dissertação inova ao orquestrar diversos serviços distribuídos e especialistas para compor uma solução completa para os casos supracitados.

1.5. Organização do trabalho

Esta dissertação está dividida em capítulos. No Capítulo 1 é introduzido o assunto, apresentando a motivação para seu desenvolvimento bem como seus principais objetivos. No Capítulo 2 foi apresentada uma breve revisão sobre o histórico de *e-Science* no Brasil e um levantamento de trabalhos relacionados. No Capítulo 3 é abordada a gestão do conhecimento e como a computação é capaz de auxiliá-la. O Capítulo 4 descrevendo os sistemas de gestão de conteúdo e os elementos importantes que o compõe, tais como os metadados e os serviços de interoperabilidade. O Capítulo 5 discorre sobre arquiteturas de dados espaciais, introduzindo a visão de arquiteturas orientadas a serviço, que são base para as especificações de serviços interoperáveis para dados geoespaciais, bem como os elementos que a compõe. No Capítulo 6 é apresentado o projeto arquitetural, definindo o problema a ser resolvido, propondo uma solução conceitual para o mesmo e avaliando as ferramentas disponíveis para tal. O Capítulo 7 apresenta a arquitetura concreta para o projeto arquitetural, entendendo os requisitos da solução e como os componentes da arquitetura foram organizados e se inter-relacionam para compor a solução. O Capítulo 8 apresenta o estudo de caso, conduzido para avaliar a arquitetura proposta. Por fim, o Capítulo 9 apresenta as conclusões sobre este trabalho, suas principais contribuições e uma sugestão de trabalhos futuros.

2 HISTÓRICO DE *E-SCIENCE*

O termo *e-Science* geralmente é empregado para descrever contextos em que a ciência é realizada com o apoio computacional, tornando-se assim, mais eficiente. Este termo é muito amplo, englobando desde o uso e emprego de infraestruturas de hardware de alto desempenho até aplicações específicas de *software* para apoio ao desenvolvimento da ciência. Neste capítulo é apresentado um breve histórico sobre *e-Science* no Brasil, destacando alguns movimentos na área de hardware e *software*; e em seguida são apresentados alguns trabalhos relacionados a este.

2.1. Histórico de *e-Science* no Brasil

Reconhecendo a importância da computação para o progresso científico, em 26 de julho de 2004, o Vice-Presidente da República - José Alencar Gomes Da Silva - no exercício do cargo de Presidente da República, assinou o decreto Nº 5.156 institucionalizando o Sistema Nacional de Processamento de Alto Desempenho – SINAPAD. O SINAPAD é uma rede de centros de computação de alto desempenho, geograficamente distribuídos, instituída pelo Ministério da Ciência e Tecnologia (MCT). São oito (8) unidades, denominadas Centros Nacionais de Processamento de Alto Desempenho (CENAPAD), operadas respectivamente pela UFRGS, UFMG, UFC, UNICAMP, UFRJ, UFPE, INPE e LNCC, sendo este último o coordenador do sistema, segundo delegação do MCT. Outras grandes instituições do território nacional dispõem de uma infraestrutura computacional de alto desempenho, como a Universidade Estadual Paulista (UNESP) com seu Núcleo de Computação Científica com uma arquitetura de grade de processamento distribuído.

A criação destes grandes centros de computação possibilitou que diversos projetos e pesquisas pudessem ser realizados, fornecendo uma infraestrutura de *hardware* necessária para o processamento computacional. No campo de *software*, há também grandes movimentos com intuito de fornecer agilidade e facilidades auxiliando os pesquisadores no desenvolvimento de seus trabalhos.

O Centro de Computação da Universidade de Campinas desenvolveu um projeto de um *software* de apoio à pesquisa chamado SCIENCE, cujo propósito é auxiliar os pesquisadores e grupos de pesquisas a organizar suas atividades em torno de um sítio na Internet (UNICAMP, 2011?). Esta ferramenta auxilia a geração de interfaces *web* de projetos de pesquisa, a organização e a publicação de trabalhos de grupos de pesquisa, a interação de pesquisadores, alunos e comunidade via *web* bem como o gerenciamento de projetos de pesquisa.

O crescente volume de informação produzida e disponível abre espaço para trabalhos específicos para tratar do gerenciamento do conhecimento. A indústria de produção de informação tem se desenvolvido à margem das revoluções e do crescimento industrial, absorvendo, assim, as suas características marcantes. A geração de estoques de informação adotou para si os preceitos da produtividade e da técnica como o seu mercado de trabalho. A crescente produção de informação precisa ser reunida e armazenada de forma eficiente, obedecendo a critérios de produtividade na estocagem, ou seja, o maior número de estruturas informacionais deve ser colocado em menor espaço possível dentro de limites da eficácia e custo (BARRETO, 1994).

O gerenciamento do conhecimento extrapola os limites da armazenagem dos dados. Para extrair o conhecimento dos aglomerados de dados tem-se que compreender a forma como eles se relacionam e criar um modo de descobri-lo. Neste sentido, o Núcleo de Pesquisa em Qualidade de *Software* (NPQS) da Universidade Federal de Juiz de Fora (UFJF), vem trabalhando com temas em ontologias aplicados à *e-Science* desde 2000. Como destaque, um de seus trabalhos vem empregando tecnologias como agentes, ontologias e serviços *web* semânticos para compor uma infraestrutura de apoio a ciência (PALAZZI et al., 2009). Outro trabalho desta mesma instituição, que merece destaque, emprega o uso de ontologias para busca por artefatos científicos na área de modelagem computacional, especificando e implementado uma arquitetura para tal (MENDES et al., 2008).

Outra área explorada dentro do contexto de *e-Science* é *Workflows* Científicos, que tem sido muito utilizado no apoio aos experimentos científicos. Os *workflows* científicos são abstrações que permitem a especificação destes experimentos, de maneira estruturada, através de um fluxo de programas, serviços e dados para produzir um resultado final (DEELMAN et al. 2009).

2.2. Trabalhos Relacionados

Alguns autores vêm desenvolvendo trabalhos sobre o tema de gestão de conhecimento científico e estudando ferramentas e alternativas para auxiliar e facilitar os processos envolvidos em tais atividades. Leite e Costa (2006) discutem a adequação e a aplicabilidade de repositórios institucionais como uma ferramenta para tal, abordando as peculiaridades do conhecimento científico, bem como o ambiente no qual acontecem os processos de sua criação, compartilhamento e uso.

A organização e agrupamento de dados pode se dar por meio de diversos paradigmas, por exemplo, por seu significado semântico ou contextual assim como em alguns dos trabalhos supracitados. Esta abordagem pode ser relevante e eleita como a melhor opção dentro de um determinado contexto, mas pode haver casos onde a exploração de outros paradigmas pode trazer benefícios significantes.

Em casos onde a referência geográfica é relevante, pode-se adotar a componente geográfica como o paradigma de organização e agrupamento dos dados. Fato que introduz uma flexibilidade na busca por dados numa região do globo terrestre e facilita o processo de descoberta de dados para estas regiões. Por exemplo, o uso de ferramentas de mapa e simbologia cartográfica como uma forma intuitiva de informar a existência e disponibilidade de dados para uma dada região.

Neste sentido, vale ressaltar o trabalho desenvolvido por Cañete et al. (2010) que consiste de um sistema de informações de biodiversidade baseado em

banco de dados, API do Google Maps e o sistema R[1], que permite catalogar dados a respeitos de espécimes coletadas, analisá-los e apresentá-los num mapa.

Algumas plataformas de gestão de conteúdo têm seus horizontes expandidos, quando a componente espacial é adicionada como uma das informações a serem gerenciadas. Um exemplo deste tipo de plataforma é o Cartaro (CARTARO, 2012) que se intitula GeoCMS por oferecer suporte a dados georreferenciados. O Cartaro faz uso de padrões da OGC ao usar em sua arquitetura subprodutos como PostGIS e GeoServer associado ao CMS Drupal (DRUPAL, 2012). O Cartaro, assim como algumas outras plataformas de CMS oferece suporte à componente espacial, mas quase sempre o que se tem é adição das informações de latitude e longitude a algum conteúdo no CMS. Em determinado cenário isso pode não ser suficiente, casos, por exemplo, onde um determinado conteúdo está associado a uma representação de uma área ou vários conteúdos estão relacionados a uma representação de um ponto geográfico ou a uma área (polígono).

2.3. Considerações finais do capítulo

Neste capítulo foi apresentado um breve histórico de *e-Science* no Brasil, comentando desde a institucionalização do SINAPAD e de alguns outros centros de processamento de alto desempenho privados até alguns trabalhos desenvolvidos em algumas universidades. Foram apresentados alguns trabalhos relacionados com gestão de conhecimento. Alguns com possibilidade de empregar a componente espacial como diferencial, porém aplicados a um problema específico ou ferramentas de gestão de conteúdo com extensão espacial limitada a informações de latitude e longitude de seus conteúdos.

[1] Sistema R é um sistema estatístico para uso geral e muito utilizado em aplicações de geoprocessamento.

3 O CONHECIMENTO E SUA GESTÃO

Em sentido amplo, ciência (do latim *scientia*, traduzido por "conhecimento") refere-se a qualquer conhecimento ou prática sistemáticos. Em sentido estrito, ciência refere-se ao sistema de adquirir conhecimento baseado no método científico bem como ao corpo organizado de conhecimento conseguido através de tais pesquisas.

Pode-se entender que o produto essencial da ciência é o conhecimento e por isso uma preocupação em gerenciar, registrar, comunicar e disseminá-lo. Uma das questões que se coloca é de que forma as organizações podem conciliar o conhecimento que se encontra na cabeça dos seus integrantes com as informações existentes em suas bases de dados, nos papéis, planilhas e relatórios por elas gerados, transformando-os em uma ferramenta transformadora, inovadora ou geradora de novos produtos. Outra questão diz respeito a como reter esse conhecimento para que ele se torne propriedade da organização.

3.1. Gestão do Conhecimento

Gestão do conhecimento é um tema relativamente novo, multidisciplinar e muito explorado em diversas pesquisas, porém quase sempre seu foco são as organizações empresariais. No entanto há iniciativas da aplicação destes conceitos da gestão de conhecimento sob o âmbito do conhecimento científico como dissertado, principalmente, em Leite (2006).

Para compreender Gestão do Conhecimento, primeiramente deve-se conceituar dado, informação e conhecimento.

Dado: pode ser definido como um conjunto de fatos distintos e objetivos, relativos a eventos (DAVENPORT; PRUSAK, 1998). É o conjunto discreto e objetivo de fatos sobre um determinado evento ou objeto. É informação bruta, descrição exata de algo ou de algum evento. Os dados em si não são dotados de relevância, propósito e significado, mas são importantes por se tratar da matéria-prima essencial para a criação da informação.

Informação: mensagem contendo um emissor e um receptor e cujo significado envolve a interpretação de algo, baseado num conjunto de dados, ou seja, é uma mensagem com dados que fazem diferença. São dados interpretados, dotados de relevância e propósito (DRUCKER, 1999). É um fluxo de mensagens, um produto capaz de gerar conhecimento. É um meio ou material necessário para extrair e construir o conhecimento. Afeta o conhecimento acrescentando-lhe algo ou reestruturando-o (MACHLUP; MANSFIELD, 1983).

Conhecimento: deriva da informação assim como esta, dos dados. O conhecimento não é puro nem simples, mas é uma mistura de elementos. É uma mistura fluída de experiências, valores, crenças, informação contextual e intuição fornecendo insumo necessário para que uma pessoa habilitada a interprete, avalie e tome decisões acerca de casos. Segundo Davenport e Prusak (1998), o conhecimento pode ser comparado a um sistema vivo, que cresce e se modifica à medida que interage com o meio ambiente. Nonaka e Takeushi (1997) observam, ainda, que o conhecimento, diferentemente da informação, refere-se a crenças e compromisso.

Davenport (1998) realizou uma comparação dos principais componentes e características de dado, informação e conhecimento. Esta comparação é apresentada resumidamente na Tabela 3.1.

Tabela 3.1: Dado, informação e conhecimento.

Dado	Informação	Conhecimento
Simples observações sobre o estado do mundo -Facilmente estruturado -Facilmente obtido por máquinas -Frequentemente quantificado -Facilmente transferível	Dados dotados de relevância e propósito -Requer unidade de análise -Exige consenso em relação ao significado -Exige necessariamente a mediação humana	Informação valiosa da mente humana -De difícil estruturação -De difícil captura em máquinas -Frequentemente tácito -De difícil transferência

Fonte: Davenport (1998).

Nonaka e Takeuchi (1997) forneceram uma grande contribuição para o assunto, sendo suas obras as importantes referências sobre gestão de conhecimento. Estes autores realizaram uma importante distinção entre os tipos de conhecimento humano, classificando-os em conhecimento tácito e conhecimento explícito.

Conhecimentos explícitos: são os conhecimentos estruturados capazes de serem verbalizados ou transmitidos. É o que pode ser articulado na linguagem formal, inclusive em afirmações gramaticais, expressões matemáticas, especificações e facilmente transmitido, sistematizado e comunicado.

Conhecimentos tácitos: são os conhecimentos inerentes as pessoas, isto é, o conhecimento pessoal incorporado à experiência individual e envolve fatores intangíveis como, por exemplo, crenças pessoais, perspectivas, sistema de valor, *insights*, intuições, emoções e habilidades. É difícil de ser articulado na linguagem formal e transmitido por se tratar da parcela não estruturada do conhecimento.

Nonaka e Takeuchi (1997), consideram que um trabalho efetivo com o conhecimento somente é possível em um ambiente em que possa ocorrer a contínua conversão entre esses dois formatos. Segundo os mesmos autores, existe um ciclo contínuo dentro das organizações onde conhecimento explícito se transforma em tácito e vice-versa. Este ciclo é apresentado por eles como uma “espiral do conhecimento” conforme pode ser observado na Figura 3.1. A dinâmica da espiral de criação do conhecimento organizacional é uma questão fundamental da teoria apresentada. Somente por meio dos quatro modos de conversão (socialização, externalização, combinação e internalização), que se dão a partir da interação entre conhecimento tácito e explícito, é construído um novo conhecimento (NONAKA; TAKEUCHI, 1997).

Socialização: conversão de conhecimento tácito em conhecimento tácito, “é um processo de compartilhamento de experiências e, a partir daí, de criação do conhecimento tácito” (NONAKA; TAKEUCHI, 1997). Isto acontece quando as pessoas conversam ou discutem entre si.

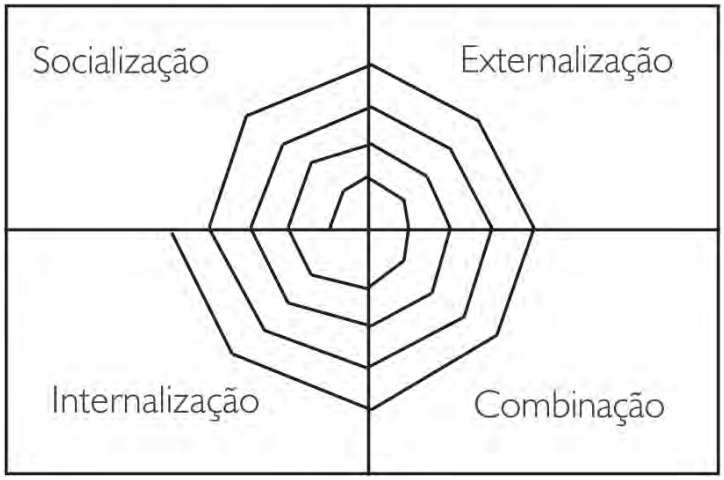

Figura 3.1: Espiral do conhecimento.
Fonte: Adaptado de Nonaka e Takeuchi, 1997.

Externalização: conversão do conhecimento tácito em conhecimento explícito, entendida como um processo relacionado à articulação do conhecimento tácito em formas explícitas e comunicáveis, “é um processo de criação do conhecimento perfeito, na medida em que o conhecimento tácito se torna explícito” (NONAKA; TAKEUCHI, 1997). Está presente no ato de escrever, registrar ou dissertar sobre um assunto.

Combinação: conversão do conhecimento explícito em conhecimento explícito, “envolve a combinação de conjuntos diferentes de conhecimento explícito” (NONAKA; TAKEUCHI, 1997).

Internalização: conversão do conhecimento explícito em conhecimento tácito, é definida como “o processo de incorporação do conhecimento explícito no conhecimento tácito”. (NONAKA; TAKEUCHI, 1997). Ocorre quando conhecimentos explícitos são colocados em prática pela pessoa, gerando assim, novos conhecimentos tácitos.

A Gestão de Conhecimento tem como objetivo macro, garantir e facilitar a existência, manutenção e crescimento destes processos de transformação. Assumindo não apenas o papel de ser uma infraestrutura para tal, mas também

criando condições ambientais e motivacionais que faça com que as pessoas vivam e reforcem estes ciclos de transformação.

3.2. Computação como Infraestrutura para Gestão do Conhecimento

O acesso livre a resultados de pesquisa tem sido visto como fator que maximiza o acesso à pesquisa propriamente dita. Assim, maximiza e acelera o impacto das pesquisas e, consequentemente, sua produtividade, progresso e recompensas (BRODY; HARNAD, 2003). Resultados de uma pesquisa realizada por Lawrence (2001), por exemplo, estão em acordo com o que preconizam Brody e Harnad, pois mostraram um crescimento de 336%, em média, nas citações a artigos disponíveis *online*, em relação a artigos publicados *off-line*, na mesma fonte. Como observa Lawrence, "para maximizar o impacto, minimizar a redundância e acelerar o progresso científico, autores e editores deveriam visar a tornar a pesquisa facilmente acessível".

Os recursos computacionais e o ambiente *web* muito contribuem para este cenário de compartilhamento e comunicação. De acordo com Davenport et al. (1998) os recursos computacionais facilitam o trabalho em rede, podendo manter os conhecimentos descentralizados junto aos locais em que são mais gerados e/ou utilizados. Por outro lado os recursos computacionais melhoram o grau de interatividade do usuário com os registros de conhecimentos, por exemplo, com as *lessons learned* (DAVENPORT; PRUSAK, 1998). A computação é efetivamente útil para a Gestão do Conhecimento, se for empregada utilizando-se uma interatividade humana sistemática (DAVENPORT, 2001).

A importância da computação torna-se inquestionável para a Gestão do Conhecimento, ficando como um novo desafio, oferecer uma forma de sistematizar, agilizar os processos de comunicação e aumentar a interação entre membros das comunidades científicas, especialmente em ambientes de acesso livre à informação.

Numa bem-sucedida sistematização da gestão do conhecimento deve considerar que o conhecimento pode existir em dois formatos, tanto na mente das pessoas, quanto em registros diversos. Portanto, a tecnologia da informação tem grande importância no acesso e na renovação dos conhecimentos, constituindo o suporte para que isso ocorra.

3.3. Considerações finais do capítulo

Neste capítulo foram apresentados conceitos importantes sobre a gestão de conhecimento e que serão necessários para compreender alguns pontos do trabalho. Foi mencionado o trabalho de Leite (2006) que discute como recursos computacionais podem favorecer a espiral do conhecimento descrito por Nonaka e Takeuchi (1997). Muito da aplicação de recursos computacionais à gestão de conhecimento, apresentada neste trabalho, se baseia nas pesquisas de Leite (2006) e pode-se perceber que muitas das ferramentas descritas por este autor podem ser encontradas em Sistemas de Gerenciamento de Conteúdo, tema do próximo capítulo.

4 SISTEMA DE GERENCIAMENTO DE CONTEÚDO

Neste capítulo são abordados os Sistemas Gerenciadores de Conteúdo, explanando como é feito o intercâmbio de dados entre sistemas deste tipo através do padrão CMIS e como os metadados são importantes para descrever os atributos destes conteúdos.

4.1. Sistema de Gerenciamento de Conteúdo Corporativo

Sistema de Gerenciamento de Conteúdo Corporativo, ou do inglês *Enterprise Content Management* (ECM) compreende as estratégias, ferramentas, processos e habilidades que uma organização precisa para gerenciar seus ativos de informação durante o seu ciclo de vida, incluindo todos os ativos digitais, como documentos, dados, relatórios e páginas da *web* (Smith e McKeen, 2003). O *Meta Group*, instituição que promove o conhecimento e seu uso aplicado ao mercado, define o ECM como a tecnologia que fornece os meios para criar, capturar, gerenciar, armazenar, publicar, distribuir, pesquisar, personalizar, apresentar e imprimir qualquer conteúdo digital, tais como imagens, texto, relatórios, vídeo, áudio, dados transacionais, catálogo, código. Sistemas de ECM concentram-se na captura, armazenamento, recuperação e disseminação de arquivos digitais para uso corporativo (WEISETH et al., 2002).

Sistema de Gerenciamento de Conteúdo Corporativo, ou simplesmente Sistema de Gerenciamento de Conteúdo (CMS da sigla em inglês) é a expressão utilizada para descrever ferramentas que promovem meios de gerenciamento, publicação e manutenção destes ativos de informação (BHAUMIK, 2011). As ferramentas disponíveis em um CMS possibilitam uma forma fácil e padronizada de publicar, gerenciar e distribuir conteúdo na *web*, promovendo uma abstração do conteúdo a ser manipulado, da forma como ele o será. Sistemas deste tipo, atualmente, além das funcionalidades supracitadas, ainda contam com algumas outras características importantes, tais como:

- Estrutura para *workflow*, que são conjunto de ações e atividades que podem ser encadeadas em uma determinada sequência obedecendo as regras pré-estabelecidas;

- Instâncias públicas e privadas para desta forma efetuar o controle de publicação de conteúdo;

- Sistema de fórum de discussões para permitir colaboração entre usuários na discussão de determinados tópicos ou assuntos;

- Carga de arquivos, para armazenar e gerenciar o ciclo de vida de um ativo informacional;

- Modelos (*templates*) para construção de páginas.

4.2. Serviço Interoperável para Gestão de Conteúdo

Serviço interoperável para Gestão de Conteúdo, do inglês *Content Management Interoperability Services* (CMIS), é um padrão aberto que visa melhorar a interoperabilidade entre Sistemas de Gerenciamento de Conteúdo, para tal define uma camada abstrata para manipular conteúdos através de protocolos *web* (OASIS, 2010). É um padrão aprovado pela *Organization for the Advancement of Structured Information Standards* (OASIS) como uma de suas especificações.

O padrão CMIS define um modelo de domínio e uma coleção de serviços *web*, que podem ser implementados utilizando WSDL e SOAP ou utilizando REST e sua convenção AtomPub, que são padrões utilizados em sistemas distribuídos orientados a serviços e serão melhor explorados na Seção 5.1. Seu modelo de domínio especifica objetos (*object* ou *content*) para representar arquivos e diretórios e suas respectivas propriedades (*property*) (OASIS, 2010). A Figura 4.1 ilustra em alto nível o modelo de domínio do padrão CMIS, onde pode ser visto que o tipo de dado *Object* pode ser especializado dando origem a novos tipos como *Document, Folder, Relationship* e *Policy* e que cada instancia de *Object* pode possuir propriedades (*property*).

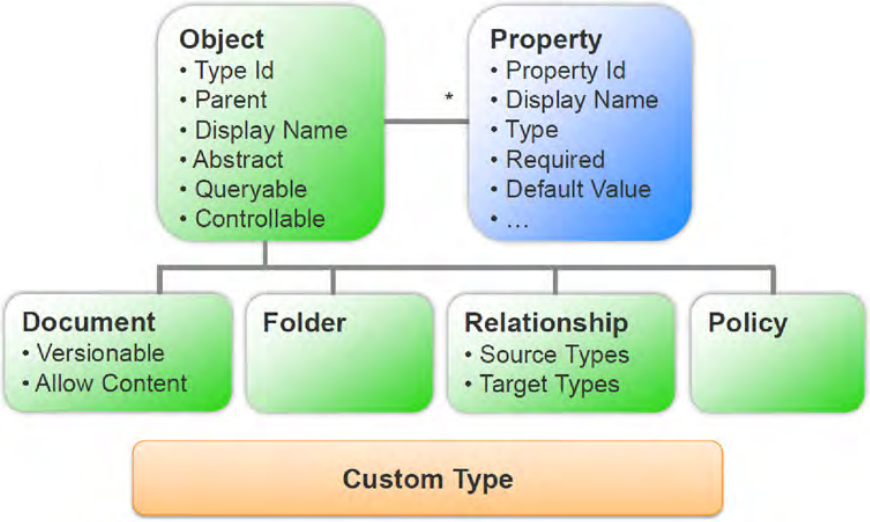

Figura 4.1: Modelo de domínio do padrão CMIS.
Fonte: Alfresco, 2011.

O conjunto de serviços *web* do padrão CMIS provê serviços para a manipulação destes objetos, adição e obtenção dos mesmos, controle de acesso, *checkout* e versionamento, e ainda a capacidade de definir relações entre os mesmos (OASIS, 2010). A Figura 4.2 ilustra os serviços *web* disponíveis no padrão CMIS, agrupando cada um dos respectivos métodos em temas. Na Tabela A.1 pode ser visto em mais detalhes estes serviços *web* e uma descrição de cada um dos métodos disponíveis em cada um deles.

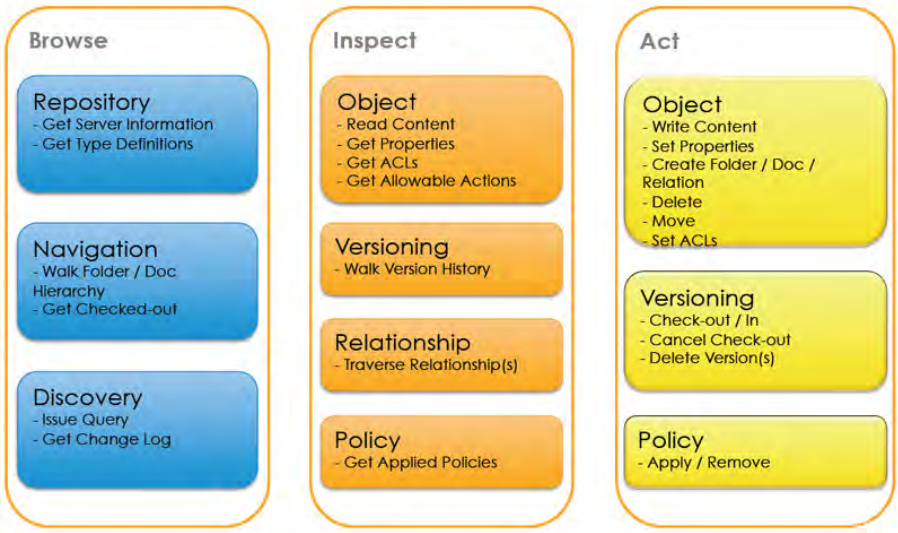

Figura 4.2: Serviços *web* do padrão CMIS.
Fonte: Alfresco, 2011.

4.3. Metadado

Metadados, ou dados sobre dados, fornecem informações sobre um determinado recurso sejam eles físicos ou digitais, promovendo a interoperabilidade, identificação, compartilhamento, integração, utilização/reutilização, gerenciamento e recuperação dos mesmos de maneira mais eficiente (BHAUMIK, 2011). Os metadados são dados descritivos que podem informar sobre o título, autor, data, publicação, palavras-chaves, descrição, localização de recursos, seus objetivos e características, mostrando como, quando e por quem o recurso foi armazenado e como está formatado.

Um dos maiores desafios na implementação de um CMS é estabelecer um conjunto de metadados eficiente. Ter um repositório de dados onde seus usuários não podem pesquisar e acessar informações de maneira fácil e eficiente não é melhor que armazenar dados em um sistema de arquivos simples.

De uma forma muito simples, metadados é um conjunto de campos e valores utilizados para descrever e categorizar o conteúdo a ser armazenado. Eles são muito importantes para os seguintes aspectos:

- **Pesquisa:** Os usuários muitas vezes desejam procurar dados associados a um conteúdo como um autor, data de publicação, palavras-chaves, tópicos, dentre outros;

- **Distribuição:** Os valores associados ao conteúdo são frequentemente usados por aplicativos para determinar quando e onde o conteúdo será distribuído ou compartilhado; e

- **Acesso:** Os valores associados ao conteúdo podem ser usados para aplicar alguma diretiva de segurança restringindo o acesso a algum conteúdo baseado em valores de metadado associados ao mesmo.

Há vários modelos pré-definidos de metadados para diversas áreas que contemplam as informações importantes sobre seus conteúdos. Alguns dos mais comuns são descritos na Tabela 4.1.

Tabela 4.1: Exemplo de alguns metadados comuns.

Área	Metadado
Arquivamento e Ciências Sociais	*Data Documentation Initiative* (DDI)
	Text Encoding Initiative (TEI)
Arquivamento	*Encoded Archival Description* (EAD)
Artes	*Categories for Description of Works* (CDWA)
	Visual Resources Association (VRA Core)
Biologia	*Darwin Core*
Indústria Editorial	*Online Information Exchange* (ONIX)
Data Warehousing	*Common Warehouse Metamodel* (CWM)
Ecologia	*Ecological Metadata Language* (EML)
Educação	*Learning Objects Metadata* (IEEE LOM)
Dados Geográficos	*Content Standard for Digital Geospatial Metadata* (CSDGM)
Governamentais / Organizacionais	*E-Government metadata Standard* (e-GMS)
	Global Information Locator Service (GILS)
	ISO/IEC 11179
Imagens	NISO MIX Z39.87
Biblioteconomia	*Machine Readable Cataloging* (MARC)
	Metadata Encoding and Transmission Standard (METS)
	Metadata Object Description Schema (MODS)

(continua)

Tabela 4.1 – Conclusão.

Mídia	PBCore
Música	*Music Encoding Initiative* (MEI)
Recursos compartilhados	*Dublin Core*
	Digital Object Identifier (DOI)

4.3.1. Dublin Core

A Iniciativa *Dublin Core* Metadados *(Dublin Core Metadata Initiative* – DCMI) é uma organização dedicada a promover a adoção de padrões de interoperabilidade de metadados e desenvolver vocabulários especializados para descrever fontes que tornem mais inteligentes sistemas de descobrimento de informações (DCMI, 2012). O *Dublin Core Resource Description* é um formato de metadados menos estruturado e mais flexível, que adota a sintaxe do *Resource Description Framework,* que é uma linguagem para representar informação na Internet endossada e recomendada pela W3C (W3C, 2004). *Dublin Core* pode ser definido como sendo o conjunto de elementos de metadados planejado para facilitar a descrição de recursos eletrônicos, sendo um dos padrões mais conhecidos e tradicionalmente adotados em sistemas gerenciadores de conteúdo.

O conjunto de metadados descrito pelo *Dublin Core* é composto de 15 elementos, os quais poderiam ser descritos como o mais baixo denominador comum para descrição de um recurso. A descrição de Hillmann (2005) para cada um destes elementos é apresentada a seguir.

1. **Título** (*Title*): O nome dado ao recurso. Normalmente, um título é um nome pelo qual o recurso é formalmente conhecido.

2. **Criador** (*Creator*): As pessoas ou organizações principais responsáveis pela criação do conteúdo intelectual do recurso sejam eles autores, fotógrafos ou ilustradores.

3. **Assunto** (*Subject*): Este elemento é relacionado ao tema do conteúdo do recurso. Tipicamente, um elemento assunto será expresso como palavras-chave, frases-chave ou códigos de classificação que descrevem o tópico do recurso. A prática recomendada é para selecionar valores ou palavras de um vocabulário padrão ou esquema de classificação formal.

4. **Descrição** (*Description*): Elemento que está diretamente ligado ao conteúdo do recurso. A descrição pode incluir, mas não está limitado a, um resumo, uma tabela de conteúdos, uma referência a uma representação gráfica de conteúdo ou um texto livre do conteúdo.

5. **Publicador** (*Publisher*): O elemento publicador é a entidade responsável por tornar o recurso disponível. Exemplos de um elemento publicador incluem uma pessoa, organização ou um serviço.

6. **Contribuidor** (*Contributor*): Uma entidade responsável por fazer contribuições ao conteúdo do recurso é denominada contribuidor. Exemplos de um Contribuidor incluem uma pessoa, organização ou um serviço.

7. **Data** (*Date*): Uma data associada a um evento no ciclo de vida do recurso. Tipicamente, o elemento Data será associada com a criação ou a disponibilidade do recurso.

8. **Tipo** (*Type*): A natureza ou gênero do conteúdo do recurso é denominada Tipo do recurso. Este elemento pode incluir termos que descrevem categorias gerais, funções, gêneros, ou níveis de agregação de conteúdo. A prática recomendada é para selecionar um valor de um vocabulário padrão.

9. **Formato** (*Format*): O elemento Formato se relaciona com a manifestação física ou digital do recurso. Normalmente, este elemento pode incluir o tipo de mídia ou dimensões do recurso. Exemplos de dimensões de um recurso incluem o tamanho e a duração. Formato

pode ser usado para determinar o *software*, *hardware* ou outro equipamento necessário para exibir ou operar o recurso.

10. **Identificador** (*Identifier*): Significa uma referência não ambígua ao recurso dentro de um determinado contexto. A prática recomendada é a identificação do recurso por meio de um elemento Identificador numérico ou textual em conformidade com um sistema formal.

11. **Origem** (*Source*): O elemento Origem é uma referência a um recurso do qual o presente recurso é derivado. O recurso presente pode ser derivado do recurso de origem, no todo ou em parte. Na prática recomenda-se referenciar o recurso por meio de um identificador.

12. **Idioma** (*Language*): O elemento Idioma define a linguagem do conteúdo intelectual adotada no recurso.

13. **Relação** (*Relation*): O elemento Relação é uma referência a outro recurso relacionado. Na prática recomenda-se referenciar o recurso por meio de um identificador.

14. **Abrangência** (*Coverage*): A Abrangência pode ser vista como a extensão ou âmbito de aplicação do conteúdo do recurso. Abrangência normalmente inclui localização espacial (um nome de lugar ou coordenadas geográficas), período temporal (período, data ou intervalo de datas) ou jurisdição (como nome de uma entidade administrativa). A prática recomendada é a seleção um valor de um vocabulário padrão.

15. **Direitos** (*Rights*): O elemento Direitos pode ser entendido como uma declaração de direito sobre a propriedade. Este elemento é um identificador que vincula a uma declaração de direito sobre a propriedade, ou um identificador que vincula a um serviço que fornece informação sobre o direito de propriedade do recurso.

Cada elemento *Dublin Core* é opcional e pode ser repetido. Estes elementos não possuem uma ordem específica e podem ser estendidos para melhor atender uma necessidade específica ou de um determinado domínio.

4.4. Considerações finais do capítulo

Neste capítulo foram apresentados os Sistemas Gerenciadores de Conteúdo e sua capacidade de gerir conteúdos, que podem ser arquivos de variados formatos, tais como planilhas, documentos de texto, fotos, áudios, vídeos, etc. Essa capacidade de lidar com formatos heterogêneos de dados faz destes sistemas uma boa opção para gerir a parcela externalizável do conhecimento produzido durante uma pesquisa. Foi apresentando o metadado Dublin Core, um dos padrões mais conhecidos e tradicionalmente adotados em sistemas gerenciadores de conteúdo, e sua flexibilidade para descrever atributos de qualquer tipo de conteúdo. Foi apresentado também o padrão CMIS que disponibiliza uma série de operações sob a forma de serviços *web* para a manipulação dos conteúdos. Estes conceitos são importantes para compreender com um arquivo multimídia qualquer, seja ele um documento de texto, uma planilha, foto ou vídeo, é armazenado em um CMS, catalogando seus atributos através de metadados e possibilitando sua manipulação através do padrão CMIS.

5 ARQUITETURA DE DADOS ESPACIAIS

Este capítulo se inicia discorrendo sobre arquiteturas de *software* orientadas a serviços e sobre interoperabilidade entre sistemas. Estes tópicos são importantes para se compreender como dados espaciais são distribuídos sob o paradigma de orientação a serviços obedecendo as especificações da *Open Geospatial Consortium* (OGC).

São apresentadas as principais especificações da OGC para distribuição, processamento e busca de dados espaciais. É apresentado também o conceito de GeoPortal e como ele, associado ao Servidor de Mapas, que provê os dados espaciais sob as especificações OGC, formam a base para Infraestruturas de Dados Espaciais. São citados exemplos de algumas infraestruturas de dados espaciais ao redor do mundo e, mais especificamente no Brasil, a Infraestrutura Nacional de Dados Espaciais.

5.1. Arquitetura Orientada a Serviços

O desenvolvimento de *software* baseado em componentes não é recente, mas tem sido alvo de muito interesse atualmente, pelo seu potencial de reduzir custos e tempo de desenvolvimento e, especialmente, devido ao interesse na instalação e atualização de sistemas distribuídos. Uma das abordagens mais interessantes nessa área é a de arquitetura orientada a serviço, *Service Oriented Architecture* – SOA (PAPAZOGLOU; GEORGAKOPOULOS , 2003).

SOA é um paradigma de desenvolvimento de sistemas distribuídos que disponibiliza funcionalidades em forma de serviços para aplicações clientes, usuários finais ou para a construção de novos serviços (COLAN, 2004). Tais serviços são autocontidos, o que significa que informações que descrevem os serviços, incluindo suas funcionalidades, interface dos métodos, comportamento, e qualidade, podem ser obtidas do próprio serviço, através de um conjunto padronizado de métodos. Serviços podem ser primários, se programados em uma linguagem de programação qualquer e independentes

de outros serviços, ou compostos, se parte do seu processamento depender de outros serviços (CURBERA et al., 2003).

Na arquitetura orientada a serviço é criado um ambiente de descoberta dinâmico e se faz o uso de serviços *web* através da rede. Serviços *web* são disponibilizados usando tecnologias como eX*tensible Markup Language* (XML), *Web Service Description Language* (WSDL), *Simple Object Access Protocol* (SOAP) e *Universal Description, Discovery and Integration* (UDDI), tecnologias estas, que habilitam a implementação e desenvolvimento da arquitetura.

A tecnologia de serviços *web* tem como principal característica de interoperabilidade a utilização de interfaces para abstrair a implementação dos serviços. Tais interfaces são definidas através de WSDL (*Web Service Description Language*) e regulam a forma como o serviço é utilizado. Para a troca de informações, uma das formas é utilizado o protocolo SOAP. Desta forma a troca de informação é realizada em forma de mensagens, dentro de ambientes distribuídos, baseado-se na tecnologia XML e consiste de três partes:

1. Um envelope que define uma estrutura para descrever o que está numa mensagem e como a processar;

2. Um conjunto de regras de codificação para descrever instâncias de tipos de dados definidos na aplicação; e

3. Uma convenção para representar as chamadas a procedimentos remotos e as respectivas respostas.

Um serviço *web* após ser publicado e disponibilizado precisa ser descoberto e para esta finalidade existe um diretório para publicação e busca por serviços, definido pelo padrão UDDI. Gunzer (2002) define UDDI como um padrão desenvolvido para fornecer um diretório de busca para os negócios e seus serviços. Tem como objetivo ser um mediador do serviço, permitindo que os clientes requisitantes encontrem um fornecedor do serviço apropriado.

Dentro da arquitetura SOA as principais operações para serviços *web* são publicação, localização, seleção, chamada e composição de serviço. A operação de publicação é realizada por um provedor de serviço, e consiste na criação de uma descrição em WSDL e sua publicação em canais de localização ou descoberta (Catalogador de Serviço), na *web*. Esses canais usam o protocolo UDDI para registrar o serviço, com metadados através dos quais usuários podem localizá-lo em catálogos ou diretórios. Ao serem publicados os serviços *web* recebem um Identificador Uniforme de Recursos (*Uniform Resource Identifier* – URI) para que sejam unicamente identificados e acessados.

Os usuários podem então selecioná-lo usando o UDDI, pela requisição do endereço em URI do serviço *web* e do seu arquivo de descrição em WSDL. Através da URI será possível usar o serviço, mas é em sua descrição em WSDL onde o cliente pode conhecer como usá-lo e quais respostas obterá.

A chamada ao serviço, por sua vez, ocorre após o conhecimento sobre os métodos publicamente disponíveis, descritos no arquivo em WSDL, e o estabelecimento de uma comunicação direta entre o *software* cliente e o provedor do serviço propriamente dito, o que em geral acontece através do protocolo HTTP (*Hypertext Transfer Protocol*). Quando uma requisição a um serviço *web* (*request*) é realizada, esta é constituída pela chamada de um método com possível passagem de parâmetros ao mesmo, todos empacotados em uma mensagem SOAP codificada em XML. A chamada normalmente termina com a recepção de uma segunda mensagem, também codificada em XML, com o resultado do processamento do serviço *web* (*response* ou *replay*).

A Figura 5.1 é uma ilustração do funcionamento de um serviço *web* dentro da arquitetura SOA. Ilustrando a comunicação entre o provedor de serviço e o consumidor de serviço através de troca de mensagens. Ilustra, ainda, a publicação do serviço *web* no catálogo (UDDI) para que o consumidor de serviço seja capaz e encontrá-lo e utilizá-lo baseado no contrato definido pelo WSDL.

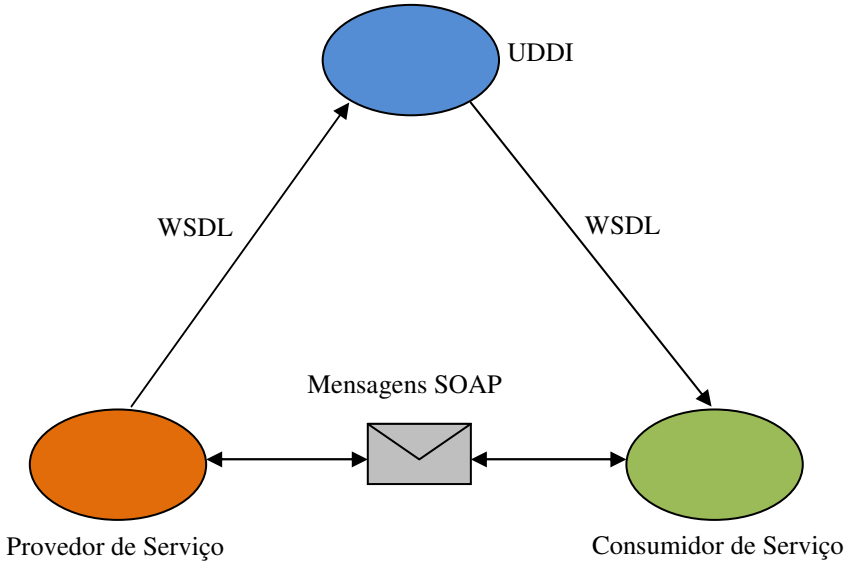

Figura 5.1: Arquitetura SOA.

5.2. Interoperabilidade

A crescente demanda por intercâmbio de informações entre sistemas distintos, desenvolvidos em uma grande diversidade de tecnologias e linguagens, e a busca por evitar a duplicações de dados e de processos levou a uma grande evolução na área de arquitetura de sistemas de informação e à maneira pela qual os *softwares* são desenvolvidos. Acrescidos da extrema necessidade de integração entre sistemas, torna-se cada vez maior a necessidade por interoperabilidade entre sistemas.

O grande passo para a interoperabilidade foi a criação e adoção de padrões, a começar pelo padrão de intercâmbio de dados genérico denominado XML. Este padrão foi desenvolvido por um consórcio chamado *World Wide Web Consortium* (W3C), que tem por objetivo desenvolver tecnologias interoperáveis para a *internet*.

O XML é uma linguagem de marcação estruturada e extensível, ou seja, não existe um conjunto pré-definido de marcadores que podem ser utilizados, estes são criados dependendo da finalidade da aplicação. Este formato foi desenvolvido para transportar e armazenar dados, permitindo também a interoperabilidade, pois os conteúdos descritos em XML podem ser

armazenados como um arquivo texto, podendo ser manipulado em qualquer editor e por qualquer linguagem de programação existente.

Como mencionado, arquitetura orientada a serviços é essencialmente um conjunto de serviços que se comunicam. Essas comunicações são feitas através de *Middleware* que abstrai a complexidade das comunicações entre os serviços, simplificando assim o seu desenvolvimento. Os serviços podem não se encontrar na mesma máquina ou sistema, e que o *software* que compõe os serviços pode ser desenvolvido em linguagens diferentes e para plataformas diferentes. Atualmente os serviços *web* são vistos como a melhor opção para *Middleware* de comunicação numa arquitetura orientada a serviços (BARRY, 2003).

Um serviço *web* é um sistema de *software*, identificado através de uma URI, na qual interfaces públicas e contratos são definidos e descritos em XML. Estas definições podem ser descobertas por outros sistemas de *software*. Estes sistemas podem então interagir com o serviço *web* em uma maneira prescrita pela sua definição, usando mensagens baseadas em XML e transportadas por protolocos da Internet (W3C, 2003).

Entre as principais vantagens do uso de serviços *web*, podem ser citadas:

- Interface abstrata: Os serviços *web* fornecem uma interface abstrata para acesso aos métodos disponibilizados, ocultando os detalhes da implementação do usuário do serviço;

- Semântica acompanha os dados: Ao invés de trafegarem somente os dados, a comunicação entre o servidor e o cliente carrega consigo metadados;

- Portabilidade: Por se tratar de um padrão aberto, baseado em XML, garante-se a portabilidade das mensagens mesmo sob diferentes plataformas de operação;

- Segurança: Opcionalmente, as informações trafegadas podem ser criptografadas; e

- Utilização de recursos: Os serviços *web* são sistemas não invasivos, pois não consomem recursos de comunicação enquanto em estado de espera.

5.3. *Open Geospatial Consortium*

O *Open Geospatial Consortium (OGC)* é um consórcio formado por companhias, agências governamentais e universidades, criado para promover o desenvolvimento de tecnologias que facilitem a interoperabilidade entre sistemas envolvendo informação espacial e localização (PERCIVALL, 2003). Os produtos do trabalho do OGC são apresentados sob a forma de especificações de interfaces e padrões de intercâmbio de dados. O consórcio surgiu em 1994 com o objetivo de garantir a interoperabilidade dos dados entre os diferentes *softwares* e conta com diversas grandes empresas da indústria de geotecnologia como consorciados.

Atualmente, o OGC estabeleceu diversas especificações que podem ser categorizados em serviços *web*, formatos de dados, conversão de coordenadas, padrões de consulta, padrões para sensores e representação de dados. Dentre eles destacam-se o *Geography Markup Language* (GML) e *Keyhole Markup Language* (KML) que são formatos para intercâmbio de dados. *Web Feature Service* (WFS), *Web Map Service* (WMS) e *Web Coverage Service* (WCS) que são serviços *web* para aquisição de dados e *Web Processing Service* (WPS) que é um serviço *web* para processamento de dados. Por fim, o padrão *Catalogue Service for the Web* (CSW) que é um serviço de busca. Existem, ainda, os padrões *Styled Layer Descriptor* (SLD), *Simple Features for SQL* (SFSQL) e *Web Map Context* (WMC) que não são abordados nesta dissertação por não serem essenciais para o desenvolvimento da mesma, mas podem ser explorados em OGC (2013).

5.3.1. *Geography Markup Language*

A especificação *Geography Markup Language* (GML) tem como objetivo disponibilizar um esquema comum de intercâmbio de dados georreferenciados entre diferentes *softwares*. É uma especificação baseada no padrão XML para definir uma forma de codificar dados geográficos. Ela foi desenvolvida para permitir codificar não somente as feições geográficas, mas também os seus atributos textuais sendo, portanto, o formato ideal para o armazenamento e transporte de dados vetoriais georreferenciados.

O objetivo da GML é oferecer um conjunto de regras com as quais um usuário pode definir sua própria linguagem para descrever seus dados. A GML em sua versão 3.0 inclui esquemas que contêm os modelos de geometria, feições (*features*) e superfícies. Os esquemas são modelos para descrever algum tipo de dados e estão publicados nas especificações do OGC (COX et al., 2003) sendo que os principais são os seguintes:

- ***BasicTypes*:** engloba uma série de componentes simples e genéricos para representação arbitrária de atributos, nulos ou não.

- ***Topology*:** especifica as definições do esquema geométrico dos dados, bem como sua descrição.

- ***Coordinate Reference Systems*:** para sistemas de referência de coordenadas.

- ***Temporal Information and Dynamic Feature*:** estende aos elementos características temporais dos dados geográficos e suas funções dinamicamente definidas.

- ***Definitions and Dictionaries*:** definições das condições de uso dentro de documentos com certas propriedades ou informações referentes à propriedade padrão.

- ***Metadata*:** utilizado para definir as propriedades dos pacotes de dados que podem ser utilizados através de outros dados já existentes.

5.3.2. OGC *Web Services*

As especificações do OGC baseiam-se em um *framework* arquitetural, chamado de *OpenGIS Services Framework* (PERCIVALL, 2003), que especifica o escopo, objetivos e comportamento de uma série de componentes. Neste sentido, o *framework* representa uma arquitetura de referência para desenvolvimento de aplicações geográficas, no espírito da ISO 19119.

A definição dos componentes segue o paradigma de serviços *web* e, portanto, está sujeita a restrições conceituais e de implementação. As restrições conceituais endereçam funcionalidade e incluem orientação a serviços, auto-descrição dos serviços e operação sem estados persistentes (*stateless operation*). As restrições de implementação endereçam questões relativas à interoperabilidade, incluindo a adoção de formatos de intercâmbio em XML, utilização de protocolos comuns à Internet (DAVIS JR et al., 2005).

O emprego do paradigma de serviços *web* aplicado na distribuição de dados georreferenciados, muito contribuiu para a ampliação de sua disseminação e utilização. A seguir será descrito, de forma sucinta, o funcionamento e o propósito dos principais serviços *web* OGC (OWS), porém, primeiro será abordado o funcionamento genérico de um OWS descrito na especificação de implementação *OpenGIS Web Service Common Implementation Specification* (*WS-Common*).

5.3.3. WS-Common

Esta especificação trata sobre os aspectos comuns existentes em outras diversas especificações de implementação de OWS, referindo-se aos serviços *Web Map Service* (WMS), *Web Feature Service* (WFS) e *Web Coverage Service* (WCS).

Em geral este padrão define parâmetros, estruturas de dados e codificações usadas nas requisições de serviços pelos clientes e as respostas enviadas pelos servidores. As especificações de cada OWS devem, portanto, conter apenas as suas especificidades. Alguns dos termos comuns utilizados nesta especificação e que são essenciais para a completa compreensão do funcionamento de todos os serviços OGC são detalhados no Apêndice B.

5.3.4. *Web Map Service*

A especificação *OpenGIS Web Map Service* (WMS) define um serviço para a produção de mapas na Internet. A WMS fornece uma representação visual dos dados na forma de uma imagem digital. Os mapas produzidos são representações geradas em formatos de imagem, como PNG, GIF e JPEG, ou em formatos vetoriais, como o SVG. O mapa resultante pode combinar várias camadas de informação, permite que seja definido o estilo de exibição para cada camada, além de poder informar os atributos de uma determinada feição, passando como parâmetro as coordenadas do ponto desejado.

O cliente ao requisitar um mapa utilizando este serviço, deve fornecer um conjunto de parâmetros que informa ao servidor: as camadas desejadas, os estilos que devem ser aplicados sobre as camadas, a área de cobertura do mapa, a projeção ou sistema de coordenadas geográficas, o formato da imagem gerada e também o seu tamanho.

Este serviço suporta as seguintes operações, descritas na Tabela 5.1.

Tabela 5.1: Operações WMS.

Operação	Descrição
`GetCapabilities`	Retorna os metadados do serviço, descrevendo quais as interfaces suportadas pelo servidor, quais parâmetros são aceitos, quais as camadas de informação que podem ser servidas e em quais formatos de imagem ele pode servi-las.

(continua)

Tabela 5.1 – Conclusão.

`GetMap`	Permite que os clientes solicitem imagens correspondentes aos mapas que podem ser servidos.
`GetFeatureInfo`	Fornece maiores informações aos clientes a respeito de pontos específicos na imagem retornada pelo servidor, ou seja, um elemento (*feature*) particular de um mapa.

O diagrama básico de sequência de execução de um serviço WMS é apresentado na Figura 5.2, demonstrando as operações invocadas pelo cliente e as respostas emitidas pelo servidor.

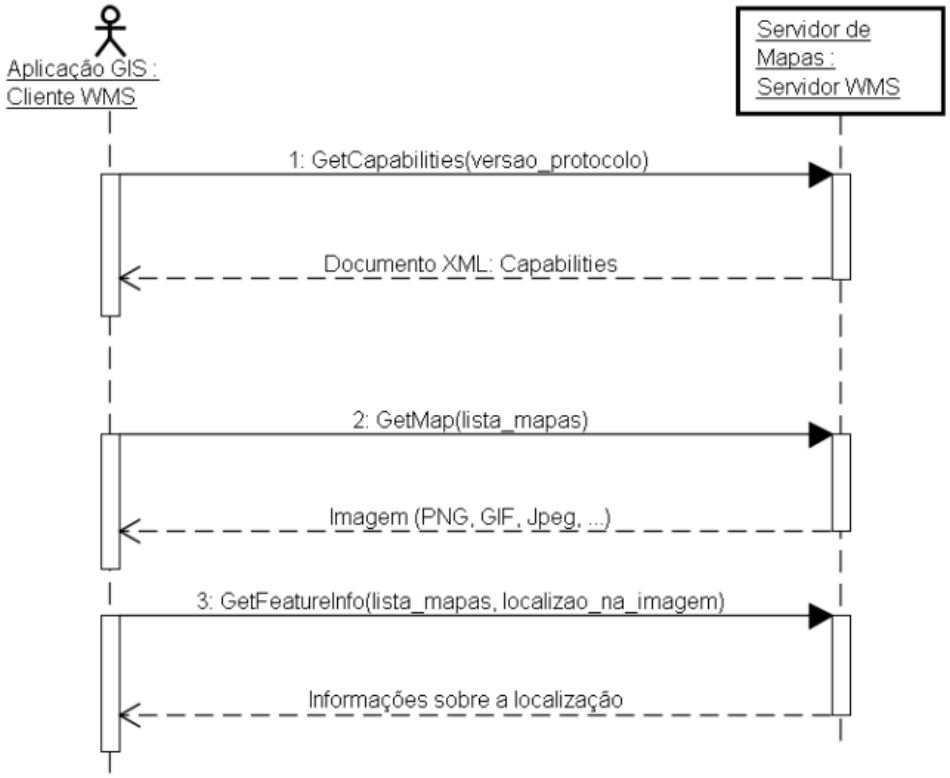

Figura 5.2: Diagrama de sequência WMS.

Fonte: Queiroz, 2007a.

5.3.5. *Web Feature Service*

A especificação *OpenGIS Web Feature Service* (WFS) define um serviço para que clientes possam recuperar objetos espaciais em formato GML, ou seja, feições georreferenciadas e seus atributos, de servidores WFS. O serviço pode ser implementado pelo servidor em duas versões: a básica e a transacional. Na versão básica apenas operações de consulta ficam disponíveis. Na versão transacional é implementado o serviço completo, que inclui operações de inserção, exclusão, atualização, consulta de objetos geográficos (DAVIS JR et al., 2005).

Este serviço suporta as seguintes operações, descritas na Tabela 5.2. Enquanto que o diagrama básico de sequência de execução de um serviço WFS é apresentado na Figura 5.3, demonstrando as operações invocadas pelo cliente e as respostas emitidas pelo servidor.

Tabela 5.2: Operações WFS.

Operação	Descrição
`GetCapabilities`	Descreve as características do servidor.
`DescribeFeatureType`	Descreve a estrutura dos tipos de objeto que podem ser servidos.
`GetFeature`	Retorna instâncias dos objetos disponíveis na base de dados. O cliente pode selecionar quais objetos deseja por critérios espaciais ou não.
`Transaction`	Utilizado para a execução de operações de modificação dos objetos (inserção, exclusão e atualização).
`LockFeature`	Bloqueia uma ou mais instâncias durante uma transação.

(continua)

Tabela 5.2 – Conclusão.

`GetFeatureWithLock`	Similar a *GetFeature* no entanto, indica ao servidor WFS para bloquear as instâncias selecionadas, pois possivelmente haverá uma atualização em operações futuras.
`GetGmlObject`	Esta operação permite recuperar feições e elementos através do ID.

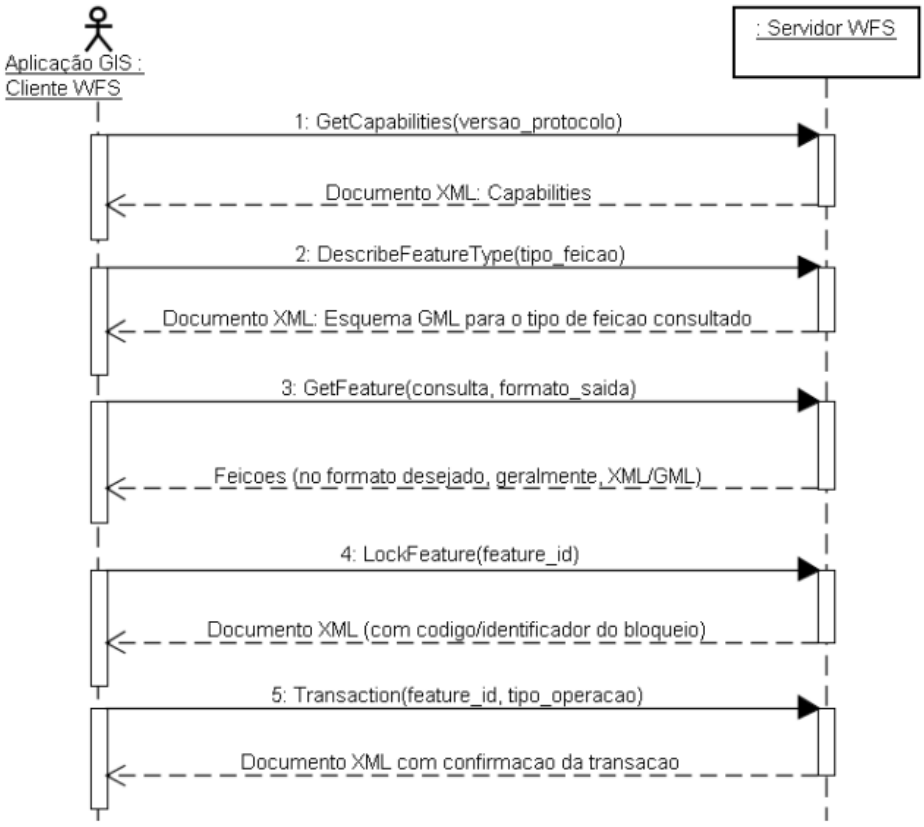

Figura 5.3: Diagrama de Sequência WFS.
Fonte: Queiroz, 2007b.

5.3.6. *Web Coverage Service*

Para o acesso a dados de cobertura (*coverages*), ou seja, que representam fenômenos com variação contínua no espaço, o OGC criou a especificação

OpenGIS Web Coverage Service (WCS). O serviço é específico para o tratamento de dados modelados como geo-campos, em complementação ao serviço WFS, que trata de dados modelados como geo-objetos, isto é, que representam entidades espaciais discretas e bem definidas (DAVIS JR et al., 2005). Geo-campos representam a distribuição espacial de uma variável geográfica sobre uma região da superfície terrestre enquanto que geo-objetos representam as entidades individualizáveis do domínio geográfico, que possuem atributos descritivos e que são usualmente armazenados num banco de dados convencional podendo possuir várias representações geométricas.

Este tipo de serviço não retorna, simplesmente, as imagens das *coverages* como resposta das requisições, mas sim dados sobre a semântica dos fenômenos representados. Como exemplos deste tipo de dados pode-se citar imagens de satélite, modelo numérico de elevação e mapas de temperatura.

Este serviço suporta as seguintes operações, descritas na Tabela 5.3.

Tabela 5.3: Operações WCS.

Operação	Descrição
`GetCapabilities`	Retorna os metadados do serviço e dos *layers* (*coverages*) disponíveis.
`DescribeCoverage`	Retorna o detalhamento da *coverage* em formato Xml .
`GetCoverage`	Retorna o dado propriamente dito, no formato especificado.

O diagrama básico de sequência de execução de um serviço WCS é apresentado na Figura 5.4, demonstrando as operações invocadas pelo cliente e as respostas emitidas pelo servidor.

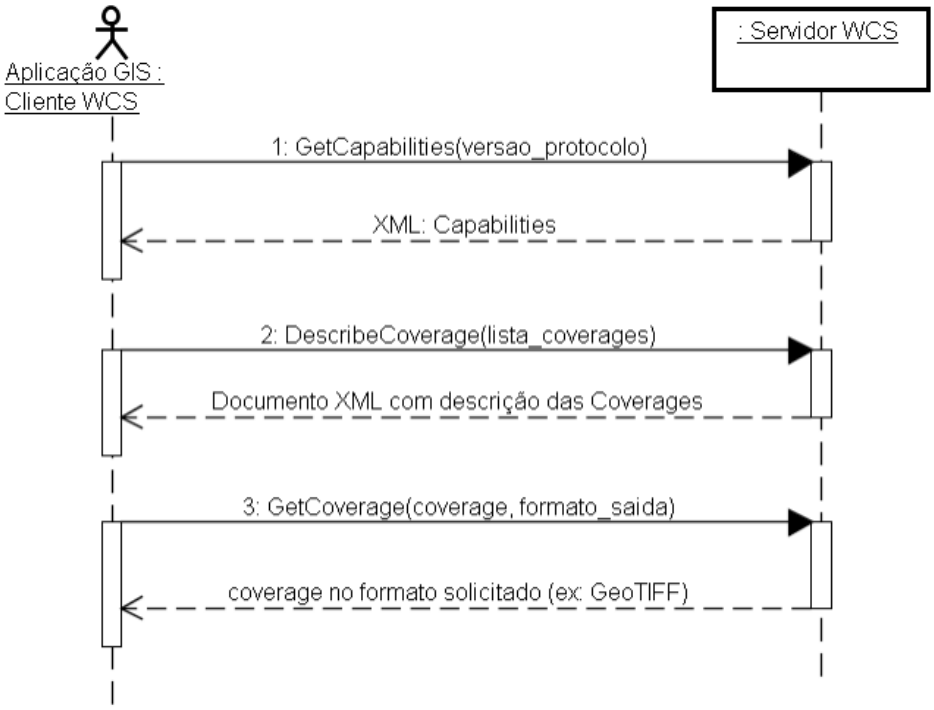

Figura 5.4: Diagrama de Sequência WCS.
Fonte: Queiroz et al., 2007.

5.3.7. *Catalog Service for the Web*

A especificação *OpenGIS Catalog Services for the Web* (CSW) diz respeito as interfaces de comunicação entre clientes e catálogos de serviços para a publicação e busca em coleções de informações descritivas (metadados) de dados espaciais, serviços e outros objetos relacionados. Os metadados de um catálogo representam as características dos recursos que podem ser pesquisados e apresentados para a avaliação e processamento tanto de homens quanto de máquinas (DAVIS JR et al., 2005).

Na especificação do CSW, é definida uma linguagem de consulta comum a todas as interfaces do serviço, a *Common Catalog Query Language*, que possui uma sintaxe semelhante a uma cláusula *"Where"* do SQL. É proposto um esquema de metadados básico para facilitar a interoperabilidade dos catálogos. Desta forma, uma mesma consulta pode ser executada em

diferentes catálogos. Um conjunto de atributos mínimo é definido para as consultas e também para as respostas das consultas.

Este serviço suporta as seguintes operações, descritas na Tabela 5.4.

Tabela 5.4: Operações CSW.

Operação	Descrição
`GetCapabilities`	Permite que um cliente recupere metadados descrevendo as características do servidor.
`DescribeRecord`	Permite um cliente descobrir elementos suportados por um catálogo ou que alguns ou todos elementos sejam descritos.
`GetRecords`	Reúne as operações *search* e *present*, que respectivamente são responsáveis pelo resultado dos elementos que atendem a busca e os metadados dos elementos afetados pela busca.
`GetDomain`	Retorna o domínio (tipos de valores possíveis) de um atributo.
`GetRecordById`	Recupera um registro pelo seu identificador.
`Transaction`	Utilizada para que um cliente solicite um conjunto de ações de inserção, remoção ou alteração de itens do catálogo.
`Harvest`	Operação que tenta recuperar recursos de uma localização específica e que pode ser reprocessada de tempos em tempos, inserindo ou atualizando o catálogo.

5.3.8. *Web Processing Service*

A especificação *Web Processing Services* (WPS) permite que serviços de processamento de dados, sejam eles algoritmos ou modelos computacionais que operem em dados espaciais, sejam disponibilizados e acessados por meio de serviços *web*. Na WPS não se especifica quais processos podem ser implementados, e sim, um mecanismo genérico para implementar qualquer processamento de dados geoespacial. Também não são especificados quais são os dados de entrada necessários e produzidos pelo processo, mas uma forma de descrever as entradas e saídas do processo.

Este serviço suporta as seguintes operações, descritas na Tabela 5.5.

Tabela 5.5: Operações WPS.

Operação	Descrição
`GetCapabilities`	Permite que o cliente receba metadados do servidor, tais como identificação e atividades que podem ser executadas.
`DescribeProcess`	Permite que o cliente obtenha informações mais detalhadas sobre uma ou mais operações que o servidor pode executar, tais como parâmetros de entrada e de saída.
`Execute`	Permite que o cliente efetivamente execute o processamento desejado, desde que forneça os parâmetros necessários, recebendo a resposta solicitada.

5.4. Arquitetura de Dados Espacial

Desde a adoção de sistemas de computação para atender a demanda de representação cartográfica e manipulação de dados georreferenciados, diversas arquiteturas foram propostas e desenvolvidas, cada qual respeitando as limitações computacionais de sua época. Em Câmara et al. (2001) e Queiroz (2003), pode-se ver as várias arquiteturas possíveis para armazenamento e manipulação deste tipo de dados, ditos não convencionais. Estas arquiteturas vão desde abordagens baseadas em armazenamento em arquivos binários proprietários, que só podem ser manipulados pelo *software* que o gerou, até esquemas mais sofisticados de armazenamento de dados em Sistemas Gerenciadores de Bancos de Dados (SGBD) Objeto-Relacionais, por meio de extensões espaciais.

Duas das arquiteturas descritas em Câmara et al. (2001), que vigoraram por muito tempo e que ainda hoje resistem em muitos nichos de mercado são as arquiteturas tradicional e dual. Ambas armazenam a geometria das feições espaciais e os atributos em locais separados, sendo que a primeira armazena em arquivos e a segunda armazena em SGBDs de fato.

Uma arquitetura mais robusta que tem dominado os sistemas geográficos atualmente é a baseada em SGBDs Relacionais com a capacidade de armazenar dados georreferenciados de maneira integrada (LISBOA FILHO; IOCHPE, 2001). Tal solução consegue fornecer aos sistemas, além da representação espacial das feições georreferenciadas, todas as vantagens que um SGBD pode proporcionar, como controle de redundâncias, compartilhamento dos dados, segurança, interfaceamento, esquematização, controle de integridade, cópias de segurança e restauração facilitados.

Estas funcionalidades geográficas em SGBDs são geralmente conhecidas como extensões espaciais de um SGBD, que como o nome sugere, adiciona tais funcionalidades a um SGBD convencional. Vários *softwares* livres e proprietários como `Oracle`, `PostgreSQL` e `MySQL` implementam essa arquitetura, baseando-se no padrão SFSSQL da OGC.

5.5. Infraestrutura de Dados Espaciais

Infraestruturas de Dados Espaciais (IDE) são vistas como um novo enfoque para a criação, distribuição e uso da informação geográfica, com ênfase na interoperabilidade (SHEKHAR; XIONG, 2008). IDEs procuram ir além da simples distribuição de mapas e dados cartográficos previamente existentes. Elas têm o potencial de se tornarem elementos fundamentais para a compreensão do espaço, disseminando dados e informação espacial juntamente com metadados sobre origem, qualidade e descrições semânticas.

De fato, uma IDE é composta de políticas, tecnologias e padrões que interconectam uma comunidade de usuários de informação espacial e atividades de suporte correlatas voltadas para a produção e gerenciamento de informação espacial (PHILLIPS et al., 1999). Esta abordagem evita esforços redundantes e reduz o custo de produção para bancos de dados novos ou existentes, através do compartilhamento de recursos. IDEs são vistas como hierarquias de fontes de dados, conforme definidas por Rajabifard et al. (2000), nas quais a informação é consolidada a partir do nível corporativo e local, até chegar aos níveis regional e global (JACOBY et al., 2002). Podem ser implementadas por meio do encadeamento de serviços de informação (ALAMEH, 2003) ou pela integração de componentes de *software* (GRANELL et al. 2005) que podem ser encontrados a partir de geoportais (MAGUIRE; LONGLEY, 2005). O termo "Infraestrutura de Dados Espaciais" designa a acumulação de tecnologias, normas e planos institucionais, que promove a disponibilidade e o acesso padronizado à dados geoespaciais (PARMA, 2006).

Enquanto uma IDE materializa uma oferta de dados geográficos, a partir de provedores que publicam serviços *web* específicos, um geoportal provê meios para que usuários humanos possam dispor de acesso interativo aos dados incluindo visualizadores e ferramentas de descoberta baseadas em metadados. A implementação de IDEs tem sido realizada com clara preferência pelo uso de arquiteturas orientadas por serviços em particular usando padrões definidos pelo OGC, tais como o WFS, o WMS e vários outros (KLOPFER, 2005).

O Brasil encontra-se no processo de elaboração de sua Infraestrutura Nacional de Dados Espaciais (INDE), seguindo assim uma tendência mundial. Em 27 de novembro de 2008, o Decreto Presidencial nº6.666 instituiu, no âmbito do Poder Executivo Federal, a Infraestrutura Nacional de Dados Espaciais – INDE. A INDE tem por objetivo, dentre outros, a de organização dos metadados dos produtos geoespaciais. De acordo com o Decreto, os metadados são informações descritivas sobre os dados, incluindo as características do seu levantamento, produção, qualidade e estrutura de armazenamento. São essenciais para promover a sua documentação, integração e disponibilização, bem como possibilitar a sua busca e exploração dos produtos geoespaciais.

A INDE nasce com o propósito de catalogar, integrar e harmonizar dados geoespaciais existentes nas instituições do governo brasileiro, produtoras e mantenedoras desse tipo de dado, de maneira que possam ser facilmente localizados, explorados e acessados para os mais diversos usos, por qualquer cliente que tenha acesso à Internet. Os dados geoespaciais serão catalogados através dos seus respectivos metadados, publicados pelos produtores e mantenedores desses dados.

A disponibilização, o compartilhamento e o acesso a dados e informações geoespaciais, bem como aos serviços relacionados, serão viabilizados, na INDE, através de uma rede de servidores integrados à Internet, que reunirá produtores, gestores e usuários destas informações na *web*. Esta rede de servidores denomina-se Diretório Brasileiro de Dados Geoespaciais, ou DBDG.

As seguintes instituições são nós potenciais do DBDG:

- Agência Nacional de Transportes Aquaviários – ANTAQ;
- Agência Nacional de Transportes Terrestres – ANTT;
- Centro Gestor e Operacional do Sistema de Proteção da Amazônia – CENSIPAM;

- Serviço Geológico do Brasil – CPRM;
- Diretoria de Hidrografia e Navegação – DHN;
- Departamento Nacional de Infraestrutura de Transportes – DNIT;
- Diretoria de Serviço Geográfico do Exército Brasileiro – DSG;
- Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA;
- Instituto Brasileiro de Geografia e Estatística – IBGE;
- Instituto de Cartografia Aeronáutica – ICA;
- Instituto Nacional de Colonização e Reforma Agrária – INCRA;
- Instituto Nacional de Pesquisas Espaciais – INPE;
- Ministério das Cidades;
- Ministério da Fazenda;
- Ministério do Meio Ambiente – MMA;
- Ministério dos Transportes; e
- Secretaria do Patrimônio da União – SPU.

5.6. Servidor de Mapas

O servidor de mapas é a camada de *software* responsável pela intermediação entre os *softwares* clientes e os dados. É o servidor de mapas quem provê os serviços *web*, de acordo com os padrões OGC, garantindo, dessa forma, a interoperabilidade entre diferentes *softwares*.

Nos dias atuais existem vários *softwares* para esse fim que possuem certificação OGC, tanto soluções proprietárias, como por exemplo, `ArcGis for Server`, `Cubewerx`, `Oracle Application Server`, `MapViewer` e

Geomatica WebServer Suite, quanto soluções livres tais como Mapserver, Geoserver e Deegree. Existem, ainda, outros *softwares* provedores de serviços *web* de dados geoespaciais em outros formatos muito difundidos na *web*, mas que ainda não implementam as especificações OGC, dentre estes pode-se citar produtos como FeatureServer e Mapfish Server.

5.7. GeoPortal

A atual tecnologia da informação geográfica está totalmente voltada para conectar e integrar todos os esforços individuais de um Sistema de Informações Geográficas (SIG), no intuito de fornecer uma base SIG global e beneficiar toda a sociedade com suas potencialidades. Vislumbra-se esta base SIG como uma ampla rede de distribuição de serviços SIG na Internet, que englobaria a descrição e a modelagem de todo o conhecimento que se tem sobre o planeta Terra. Esta base suportaria mapas, publicação de dados, catalogação de metadados e pesquisa de serviços geográficos para uso em uma série de aplicações regionais, nacionais e globais (DANGERMOND, 2005).

Os desafios para se alcançar uma base SIG global são muitos. Pode-se começar pelo intercâmbio de dados e serviços, interoperabilidade entre diferentes sistemas e plataformas, passando pelos processos de descoberta e catalogação destes recursos entre outros.

Neste sentido, muito tem-se falado sobre Infraestrutura de Dados Espaciais, com atuação muito forte da *Global Spatial Data Infrastructure Association* (GSDI), que é uma associação de organizações, agências, empresas e indivíduos que tem como propósito promover cooperação e colaboração internacional para desenvolvimento de Infraestruturas de Dados Espaciais.

Atualmente, o portal é a tecnologia mais popular e amplamente utilizada para implementar uma infraestrutura de dados espaciais, seja local, regional, nacional ou em nível internacional. Não há, porém, consenso sobre a definição

de portal. Neste contexto dois termos são geralmente utilizados pelas comunidades especialistas: GeoPortal e GIS Portal. Geoportal pode ser visto como um ponto único de acesso à informação espacial, independentemente da localização, formato ou estrutura da fonte de dados. OGC (2004) define um geoportal como uma interface humana a uma coleção de recursos de informação geoespaciais *online*, incluindo conjuntos de dados e serviços. Maguire e Longley (2005) definem um geoportal como um *gateway* na *Web* que organiza o conteúdo e serviços, tais como diretórios, ferramentas de busca, informações da comunidade, recursos de suporte de dados e aplicações. Tait (2005) define um geoportal como um sítio *web*, um ponto de entrada para conteúdo geográfico na *web* ou, mais simplesmente, um sítio *web* onde o conteúdo geográfico pode ser descoberto.

Os Geoportais usam os metadados para gestão da geoinformação, porque estes incluem informações suficientes para permitir a localização, interpretação, uso e a aplicação apropriada e correta dos dados. Esta gestão se torna possível através de especificações sobre a origem, histórico, conteúdo, representação, qualidade e demais características relevantes dos dados.

As IDEs atualmente existentes estão focadas principalmente nas escalas regionais e nacionais, como por exemplo, no Canadá (geodiscover.cgdi.ca), Estados Unidos (gos2.geodata.gov), Europa (inspire-geoportal.eu), França (geoportail.fr), Portugal (snig.igeo.pt/Portal), Noruega (www.geonorge.no), Austrália (http://asdd.ga.gov.au/asdd), Índia (gisserver.nic.in), entre outros, que construíram, através de vários acordos colaborativos, modelos de dados espaciais padronizados para compartilhar geoinformação e contribuir com o conhecimento geográfico e para tal, fazem uso dos Geoportais (HÜBNER; OLIVEIRA, 2008). Os Geoportais também são utilizados por outros setores não necessariamente ligados as IDEs, mas com a mesma finalidade de divulgar e facilitar o acesso à provedores de dados, serviços e aplicativos geográficos para uso multifinalitário.

5.8. Considerações finais do capítulo

Neste capítulo foi abordada a temática da distribuição de dados espaciais sob a forma de serviços, seguindo as especificações recomendadas pela OGC. Foi abordada questões sobre interoperabilidade e, considerando o crescimento das infraestruturas de dados espaciais e, no Brasil, a institucionalização da Infraestrutura Nacional de Dados Espaciais, a interoperabilidade é um fator essencial. É a interoperabilidade quem garante que todos os sistemas deste ecossistema complexo falarão a mesma linguagem e serão capazes de intercambiar informações.

Os serviços *web* especificados pela OGC possuem um método chamado `GetCapabilities` que retornam um contrato sobre as operações que eles suportam, similar ao WSDL. Estes serviços *web* são publicados em um Servidor de Mapas, tal como o GeoServer, e são catalogados pelo GeoPortal. Observando a Figura 5.5 pode-se fazer uma analogia com a arquitetura orientada a serviços, onde há a presença de um consumidor de dados, um provedor de serviços e um diretório catalogador destes serviços. O consumidor de serviços pode realizar uma busca no diretório de serviços e, dado que encontrou determinado serviço, poderá invocar o método `GetCapabilities` e recuperar o contrato deste serviços, ficando assim, ciente das operações disponíveis no mesmo.

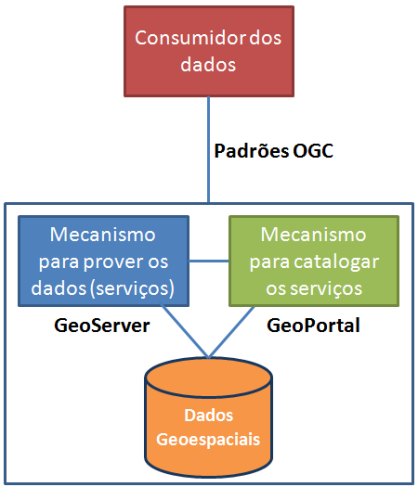

Figura 5.5: SOA Geoespacial.

As buscas no GeoPortal são realizadas através do padrão CSW e o amadurecimento e adoção deste, se apoia e fornece apoio para alguns movimentos atuais de democratização de acesso a informação (CINDE, 2010) e compartilhamento de dados. Na Europa, o INSPIRE[2] é um bom exemplo disso e, mais recentemente no Brasil, a INDE[3] compromete-se a desempenhar este papel.

[2] http://inspire.jrc.ec.europa.eu/
[3] http://www.inde.gov.br/

6 PROJETO ARQUITETURAL

Neste capítulo é definido o problema a ser resolvido e a proposição de uma arquitetura abstrata como solução. São avaliados quais os requisitos que a arquitetura deve realizar, identificando os componentes que a compõe e suas respectivas responsabilidades. Em seguida são avaliadas as ferramentas necessárias para construção desta arquitetura e a justificativa de escolha de cada uma delas.

6.1. Definição do Problema

Conforme Leite (2006), no início da criação de um novo conhecimento, o esforço de um pesquisador parte daquilo que foi construído anteriormente por outros pesquisadores, ou seja, recorre à literatura de sua especialidade, aos trabalhos anteriores e trabalhos relacionados. Em algumas situações vários detalhes dos trabalhos anteriores, que seriam usados como base para iniciar um novo trabalho, não estão disponíveis. Quase sempre o que se tem disponível são publicações formais em formato de uma dissertação, tese ou artigos acessíveis através de uma biblioteca ou anais de eventos. Este tipo de publicação possui um formato específico, onde alguns tipos de conteúdos são resumidos ou ignorados.

Muitos trabalhos de pesquisa podem possuir relevância espacial, ou seja, pode estar relacionados com alguma localização no globo terrestre ou diz respeito a alguma área ou local. Em geral, é difícil responder perguntas do tipo “Quais dados existem para esta região?” ou “Quais pesquisas foram conduzidas nesta região?”. Muitos trabalhos podem gerar diversos produtos, tais como imagens filtradas e classificadas, levantamentos em campo ou até mesmo algoritmos. Uma publicação formal nem sempre é o melhor formato para representar e distribuir este tipo de conteúdo. Há ainda uma série de artefatos usados como apoio durante algumas etapas destes tipos de trabalhos, tais como planilhas de resultados, medições em campo, delimitação de regiões de interesse (amostras de determinadas classes), fotos, arquivos de áudio ou vídeo e que são passíveis de distribuição.

Nota-se que muitos destes artefatos, produzidos durante um trabalho de pesquisa, constitui uma parcela do conhecimento que não está sendo gerida ou não é de fácil acesso. Resumidamente o problema atacado aqui é a falta ou má gestão dos artefatos produzidos durante um trabalho de pesquisa, que constitui uma parcela do conhecimento produzido e que pode ou não ter relevância espacial. O acesso ou a gestão de tais artefatos são essenciais para a reprodução do contexto de determinadas pesquisas ou ainda, a melhor distribuição destes artefatos pode contribuir para facilitar e enriquecer o desenvolvimento de novas pesquisas.

6.2. Projeto Arquitetural

O principal objetivo da arquitetura como solução para o problema exposto neste capítulo é fornecer suporte para a gestão dos variados artefatos produzidos durante um trabalho de pesquisa e fornecer facilidades para que a parcela tácita do conhecimento seja externalizada.

Neste trabalho, os artefatos produzidos durante uma pesquisa foram classificados em três tipos. O primeiro tipo compreende os arquivos multimídias, tais como os documentos de texto, planilhas, fotos, áudio e vídeos. O segundo tipo são os dados espaciais, tais como imagens de diversos sensores, classes de feições de variados temas ou dados de cobertura representando alguma grandeza física. Por fim, o terceiro tipo compreende a implementação de algoritmos de processamento (o que se convencionou chamar de artefato de processamento), que foram utilizados na metodologia ou desenvolvimento do trabalho de pesquisa e que, em geral, opera sobre algum dos outros dois tipos de artefatos. Outra responsabilidade desta arquitetura é relacionar os artefatos não espaciais a uma localização geográfica, permitindo que os mesmo sejam encontrados através desta.

A abordagem escolhida para o projeto desta arquitetura é orientação a serviços, com componentização de cada serviço segundo suas responsabilidades específicas. Para orquestrar estes serviços haverá um componente cliente, que se comunicará com os demais componentes através

de padrões interoperáveis. Uma abstração conceitual desta abordagem pode ser vista na Figura 6.1, onde há a presença de um componente cliente se comunicando com os demais componentes, sendo que todas estas comunicações são realizadas através de padrões interoperáveis.

Figura 6.1: Abstração da abordagem componentizada.

A principal vantagem desta abordagem arquitetural é o fraco acoplamento entre os componentes. O fraco acoplamento entre os componentes e o fato das comunicações serem através de padrões interoperáveis permite a distribuição dos componentes e a escalabilidade individual dos mesmos. Permite que cada um dos componentes sejam substituídos por outros, desde que implementem ou suportem o mesmo padrão de comunicação. Esta abordagem permite ainda, a definição clara das responsabilidades de cada componente, o que, em geral, diminui a complexidade de entendimento, manutenção e evolução da arquitetura.

Na Figura 6.2 é apresentada a visão dos componentes do projeto arquitetural, onde é possível observar a existência de um componente cliente, responsável

por orquestrar os serviços providos pelos demais componentes. Um Componente de Gestão de Artefatos Multimídias responsável pela gestão de todos os artefatos que não possuem componente espacial, dos quais se destacam documentos de texto, planilhas, fotos, áudio, vídeos etc. Um Componente de Gestão de Artefatos Espaciais, responsável pela gestão de todos os artefatos que possuem componente espacial, tais como imagens de sensores ópticos e de RADAR, dados vetoriais, dados de cobertura terrestre e outros que também possua componente espacial. Um Componente de Gestão de Artefatos de Processamento, responsável por gerir as implementações de algoritmos de processamento. Por fim, um Componente de Gestão de Relacionamento Espacial, responsável por gerir o relacionamento entre os elementos multimídias a uma localização geográfica.

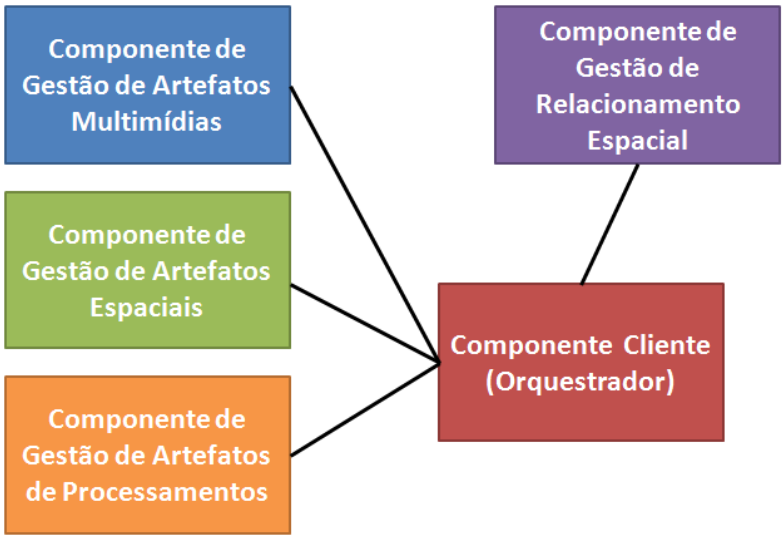

Figura 6.2: Visão de componentes da arquitetura.

6.3. Infraestrutura Necessária

Nesta seção são apresentadas as ferramentas necessárias para materializar o projeto arquitetural proposto neste capítulo. Será apresentada cada uma das ferramentas necessárias, os critérios de sua avaliação e justificativa de sua escolha. No Capítulo 7 será abordado como estas ferramentas foram efetivamente utilizadas para compor a arquitetura concreta.

O Componente de Gestão de Artefatos Multimídias é composto por um Sistema Gerenciador de Conteúdo e o eleito para isso foi o *software* chamado `Alfresco`. O Componente de Gestão de Artefatos Espaciais e o Componente de Gestão de Artefatos de Processamento são compostos por dois tipos de *softwares*, um gerenciador de conteúdo geográfico e um geoportal, sendo os eleitos o `GeoServer` e o `Esri GeoPortal Server`, respectivamente. O Componente de Gestão de Relacionamento Espacial foi integralmente desenvolvido para esta finalidade e, por fim, o Componente Cliente também foi desenvolvido utilizando as interfaces de programação do `ArcGIS for Flex`. Para os componentes que requerem persistência de dados, foi utilizado o sistema gerenciador de banco de dados `Postgres`.

6.3.1. Postgres

O `PostgreSQL` é um Sistema Gerenciador de Banco de Dados (SGBD) objeto-relacional robusto, que além das funcionalidades tradicionais de um SGBD, como por exemplo indexação, visões das tabelas (*views*), é extensível por meio de funções (*functions*), com controle de concorrência multi-versão, gatilhos (*triggers*), e que atende aos padrões SQL92 e SQL99 da *American National Standards Institute* (ANSI, 2012). Além das funcionalidade tradicionais de um SGBD, ainda possui uma extensão espacial chamada `PostGIS`. É considerado objeto-relacional, pois embora não seja um SGBD orientado a objetos, possui algumas características como herança, tipos de dados definidos pelo usuário, entre outras.

O `PostGIS`, foi lançada em 2001 pela empresa canadense chamada *Refractions Research*, que é a principal mantenedora, mas é mantido pela comunidade de usuários e desenvolvedores espalhados pelo mundo. Atende às especificações OGC e inclusive é certificado pela mesma.

6.3.1.1. Opção pelo PostgresSQL

Dentre os SGBDs disponíveis convém ressaltar alguns com amplo uso, mesmo que proprietários, como o caso do `Oracle Spatial`, `IBM DB2` com `Spatial`

Extender e Microsoft SQL Server. Dentre os *softwares* livres avaliados estão o Postgres com extensão espacial PostGIS e MySQL Spatial.

MySQL é um *software* livre muito popular e que teve grande expansão com o seu uso em aplicações voltadas para *web*, pois apresenta excelente desempenho, confiabilidade e facilidade de uso. O MySQL implementa especificações da OGC para o armazenamento de dados espaciais nos formatos *Well Known Binary* (WKB) e *Well Known Text* (WKT), atende aos padrões de consulta SQL92 e implementa algumas funções espaciais usando SFSSQL. No entanto não atende a todas as especificações da OGC, pois não possui as tabelas de metadados especificadas por este padrão.

Fatores decisórios para a opção pelo Postgres foi o suporte a dados matriciais disponíveis em sua extensão espacial (PostGIS) e ser compatível com as especificações OGC. Aliados a estas características ainda há o fato do Postgres possuir um alto rendimento em base de dados de grande volume e possuir indexação de dados espaciais avançada usando *Generalized Search Tree* (GiST) e R-Tree, o que torna seu mecanismo de busca mais eficiente e com melhor desempenho.

6.3.2. GeoServer

Segundo seu próprio sitio *web*, GeoServer é um servidor de código aberto escrito em Java que permite aos usuários compartilhar e editar dados geoespaciais. Projetado para a interoperabilidade, publica dados de qualquer fonte de dados espaciais utilizando os principais padrões abertos. Sendo um projeto conduzido pela comunidade, o GeoServer é desenvolvido, testado e suportado por um grupo diversificado de indivíduos e organizações de todo o mundo. GeoServer é a implementação de referência dos padrões WFS e WCS do OGC, sendo também certificado por alta performance no padrão WMS. GeoServer constitui um componente geoespacial para a *web* (GEOSERVER, 2012).

Quanto ao modelo de desenvolvimento do GeoServer, Holmes et al. (2005) afirmam que o projeto teria iniciado em 2001, por uma pequena empresa chamada *The Open Planning Project*. As funcionalidades deste *software* foram progressivamente desenvolvidas fruto do esforço de programadores e colaboradores individuais e de empresas (das quais destaca a *Refractions Research*). Graças a fontes de financiamento externo de empresas e organizações os desenvolvedores foram incorporando as funcionalidades de que necessitavam ao código fonte geral, beneficiando todos os utilizadores.

Entrada de Dados (Fontes de Dados)

O GeoServer oferece suporte a diversas fontes de dados de entrada, oferecendo, ainda, facilidade para se conectar a elas e publicar seus dados. Algumas das fontes de dados passíveis de serem utilizadas no GeoServer são:

- SGBDs: IBM DB2, MySQL, Oracle, PostGIS e SQL Server;
- Arquivos de dados vetoriais nos formatos: ESRI ShapeFile e MapInfo;
- Arquivos de dados *rasters*: ArcGrid (.asc), WorldImage (.png), GeoTIFF (.tiff) e Gtopo30 (.dem). Os formatos MrSID, ECW, JPEG2000, DTED, Erdas Imagine e NITF podem ser acessados, se adicionado suporte à biblioteca GDAL.

Saída de Dados

O GeoServer implementa as especificações OGC, sendo ainda certificado por esta. Este servidor é capaz de oferecer dados nos seguintes formatos:

- WMS
    - imagem: png, geotiff, geotiff8, gif, jpeg, png8, svg, svg xml, svg+xml, tiff, tiff8.
    - outros formatos: kml, kmz, openlayers, rss, GeoRSS e PDF.

- WFS e WFS-T
    - Formatos de saída: GML2, GML2-GZIP, GML3, SHAPE-ZIP, JSON.

6.3.2.1. Opção pelo GeoServer

O processo de escolha do servidor de mapas foi norteado pelos seguintes pontos: ser um *software* com licença livre, implementar as especificações OGC (WMS, WFS, WCS e WPS) preferencialmente com certificação pela OGC e possuir uma boa documentação, com uma comunidade de usuários e desenvolvedores local.

A princípio, os *softwares* proprietários (`ArcGis for Server`, `Cubewerx`, `Oracle Application Server MapViewer`, `Geomatica`) foram descartados por não serem de licença livre. Enquanto que os *softwares* `FeatureServer` e `Mapfish Server`, não se enquadram, pois não implementam os padrões OGC.

Foram, então, avaliados: `MapServer`, `Deegree`, e `GeoServer`. O `Deegree` é pioneiro na implementação de novos padrões, como por exemplo WPS e serviços de autenticação e segurança para serviços *web* OGC. É, atualmente, um *software* incubado na *The Open Source Geospatial Foundation* (OSGEO), mas tem seu desenvolvimento concentrado na Alemanha, onde conta com grande número de participantes. Em território nacional é pouco difundido e usado, dando preferência aos *softwares* `MapServer` e `GeoServer`.

Foram avaliados quesitos de desempenho, avaliando principalmente resultados obtidos por Aime e Deoliveira (2008). Em relação à desempenho após testes, Aime e Deoliveira (2008) consideraram os *softwares* `GeoServer` e `MapServer` equivalentes, embora em seus resultados percebe-se uma ligeira vantagem de performance para o *software* `GeoServer`, principalmente quando utilizado em conjunto com o banco de dados `PostGIS` (`Postgres`), conforme pode ser visto na Figura 6.3. Na Figura 6.3 pode ser observar o desempenho (tempo de

resposta) entre diferentes versões do padrão GML na medida em que mais requisições concorrentes são realizadas.

O uso em conjunto com o `PostGIS` é o que se aplica neste trabalho, que com os argumentos supracitados e o fato de que `GeoServer` atualmente é o único *software* desta categoria a ser citado no Plano de Ação para Implantação da Infraestrutura Nacional de Dados Espaciais (CINDE, 2010) foi o que sacramentou a sua escolha.

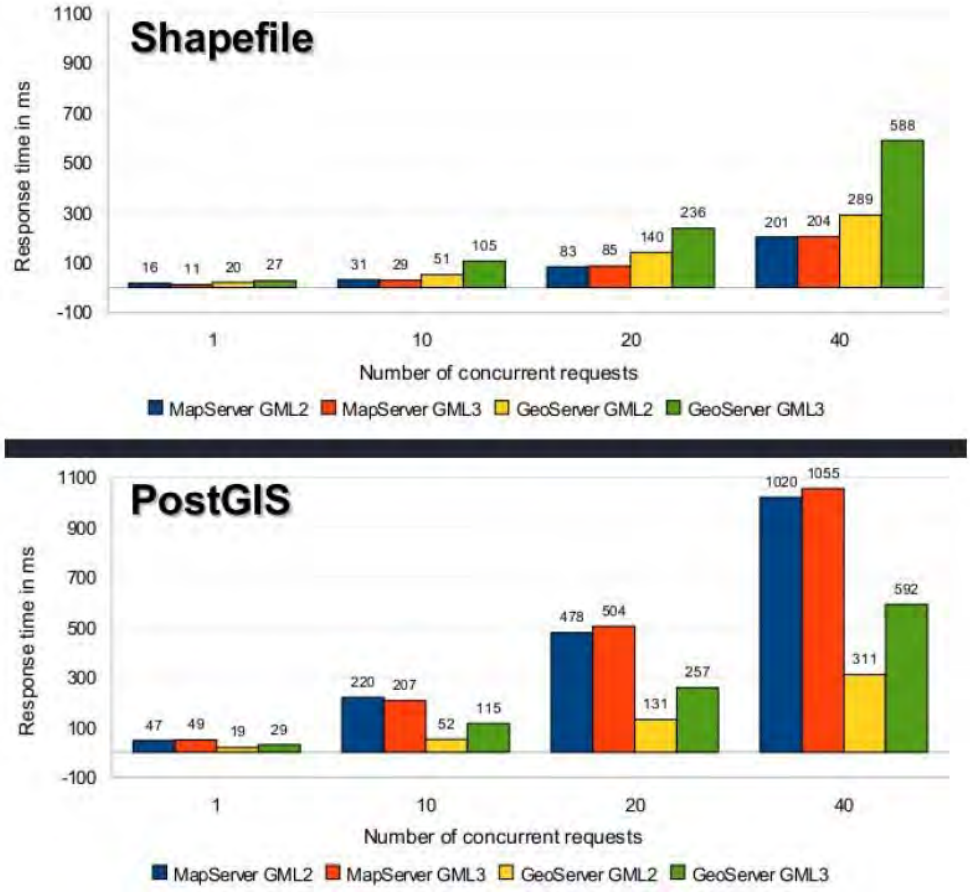

Figura 6.3: Teste de desempenho entre Mapserver e Geoserver.
Fonte: Aime e Deoliveira, 2008.

6.3.3. Esri GeoPortal Server

Esri Geoportal Server é um produto de código aberto, desenvolvido em Java, que permite a descoberta e utilização de recursos geoespaciais, incluindo

conjuntos de dados (*datasets*), dados matriciais e serviços *web*. Ele auxilia as organizações a gerenciar e publicar metadados para os recursos geoespaciais de modo que possibilite os usuários descobrir e se conectar a esses recursos. Os usuários são capazes de interagir com o catálogo do servidor através de um portal *web* de interface personalizável, e também através de uma variedade de clientes que se conectam a ele, tais como CSW, REST e OpenSearch. Isso permite aos usuários procurar por recursos não apenas com as ferramentas da família ArcGIS, mas com qualquer ferramenta personalizada. O geoportal não duplica os recursos, mas armazena e cataloga os metadados dos recursos, bem como as informações para acessá-los.

6.3.3.1. Opção pelo Esri GeoPortal

A opção pelo Esri GeoPortal se deu considerando alguns fatores, começando pelas funcionalidades disponíveis, tais como:

- Catalogação publicação e descoberta de dados;
- Aderência às especificações OGC, principalmente CSW 2.0.2;
- Suporte a catalogação de *datasets*, ferramentas de análise e serviços *web;*
- Disponibilidade de *Application Program Interface* (API) para criação de aplicações clientes e integração entre aplicações; e
- Possibilidade de execução de buscas federadas.

O fato do GeoPortal ser um projeto aberto também foi um dos fatores decisivos que acarretaram a sua escolha. Vale, ainda, ressaltar que o GeoPortal é mantido principalmente pela Esri, uma das maiores empresas no setor de geotecnologias, e que como um produto já possui um tempo considerável de maturidade, estando disponível desde 2003.

6.3.4. Alfresco

Alfresco é um sistema de gerenciamento de conteúdo coorporativo (ECMS) de porte corporativo para atender desde instituições pequenas até as de grande porte, com variado grau de sofisticação e detalhes que cada uma requer. Todo o *software* é licenciado em *software* livre e o Java é sua plataforma tecnológica (ALFRESCO, 2012).

A empresa Alfresco que patrocina o desenvolvimento do *software* foi formada em 2005 por um grupo de profissionais de `Documentum`, `Business Objects`, `FileNet/IBM` e `Hewlett-Packard`. A solução Alfresco é a solução mais completa e profissional no segmento de ECMS *open source* devido à variedade de necessidades de gestão de documentos atendidas e pela experiência da equipe desenvolvedora com esta tecnologia.

Alfresco é desenvolvido e distribuído em vários módulos, dentre os quais o módulo DM (*Document Management*) é um dos principais elementos da arquitetura desta solução. O módulo DM é o coração do Alfresco, ele é o repositório de documentos sendo constituído pelo banco de dados relacional, oferecendo suporte ao `Oracle`, `MySql`, `SqlServer` ou `Postgres` para armazenar os metadados e pela área de *storage* onde são armazenados os documentos em formato acessível ao servidor Alfresco. Os documentos são expostos aos usuários somente conforme as permissões definidas no sistema e o repositório de documentos, apesar de poder ser acessado por uma variedade de protocolos, sempre exibe ao usuário somente os documentos que estão em sua alçada de permissão.

6.3.4.1. Opção pelo Alfresco

Como plataforma de CMS, além do Alfresco foi avaliado o produto Nuxeo (NUXEO, 2012), que também é código fonte aberto e atendia aos requisitos da solução. Porém a opção pelo Alfresco se ocorreu por sua facilidade em criar e estender novos tipos de conteúdos e por seu bom suporte e aderência ao padrão CMIS. Em testes preliminares algumas buscas realizadas via CMIS API

na plataforma do Nuxeo não foram executadas com sucesso. Visto que o correto funcionamento da CMIS API é vital para a arquitetura desenvolvida e considerando as facilidades que o Alfresco oferece em configurar e estender suas funcionalidades fizeram deste o eleito final.

6.4. Considerações finais do capítulo

Neste capítulo foi discutido o problema a ser solucionado e o projeto arquitetural para o problema proposto. Foi explicada a abordagem para a construção da solução, sendo escolhido o paradigma de orientação a serviços como principal padrão arquitetural. Em seguida foi descrito em alto nível os componentes principais do projeto e as ferramentas escolhidas para a construção da solução, discutindo a responsabilidade de cada uma. Foram realizadas comparações entre algumas alternativas das ferramentas e foi apresentado os critérios e justificativa de escolha de cada uma. O próximo capítulo irá discutir como estas ferramentas foram utilizadas para criar um *software* baseado no projeto arquitetural apresentado neste capítulo.

7 ARQUITETURA PROPOSTA

Neste capitulo é apresentada a arquitetura proposta como solução para o problema proposto no Capítulo 6. Detalha-se cada elemento que compõe a arquitetura proposta como solução e como cada um destes elementos se inter-relacionam.

7.1. Visão geral

Foram eleitos dois usuários da Divisão de Processamento de Imagens do INPE para representarem o grupo de usuários do *software* desenvolvido neste trabalho. Foram realizadas reuniões, envolvendo os dois usuários eleitos, o autor deste trabalho e seus orientadores, sendo um dos produtos destas reuniões um diagrama macro da solução agrupando os diversos itens e dificuldade encontrados durante o desenvolvimento de um trabalho de pesquisa. Neste diagrama foi construída uma visão das características de um trabalho de pesquisa e elementos comuns, presentes em uma boa parte deles. Foram identificados os tipos de artefatos envolvidos (mídias) e os formatos em que se encontram bem como algumas capacidades que a solução deveria possuir. Este diagrama é apresentado na Figura 7.1, onde é possível visualizar os grandes temas que devem ser considerados na solução e quais são os elementos menores que o compõem. Os grandes temas considerados na solução englobam os **Requisitos** que a solução deve atender e as **Restrições** quanto ao seu desenvolvimento, que são os itens que nortearão a sua construção. Outro grande tema considerado neste diagrama são os **Trabalhos Produzidos** que visa agrupar as principais características, fases e produtos gerados em grande parte dos trabalhos de pesquisa. Por fim, outro grande tema são os **Artefatos Envolvidos** que visa descrever alguns dos principais possíveis artefatos e formatos que a solução precisa ser capaz de gerir.

A partir deste diagrama macro da solução, foram extraídos os requisitos que a solução deveria contemplar e estes requisitos, juntamente com algumas premissas e restrições nortearam o desenvolvimento desta solução. Os requisitos extraídos são resumidamente apresentados na Tabela 7.1.

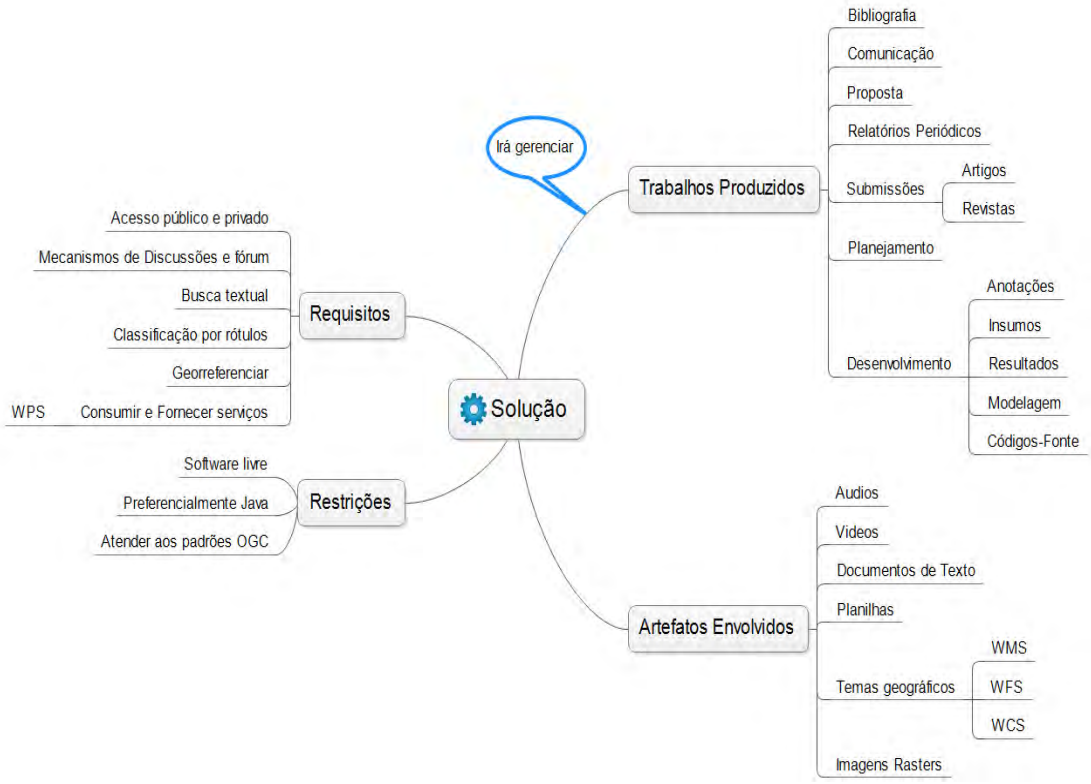

Figura 7.1: Diagrama macro da solução.

Tabela 7.1: Requisitos exigidos para a solução.

ID	Requisito
RQ01	A solução deve contemplar um mecanismo de fórum.
RQ02	A solução deve contemplar um mecanismo de listas de discussões.
RQ03	A solução deve ser de acesso público, mas com recursos de restrições de acesso a determinados conteúdos.
RQ04	A solução deve contemplar mecanismo de armazenamento de arquivos multimídias.
RQ05	A solução deve ser suficientemente configurável de modo que possam ser definidos quais metadados são importantes para cada tipo de conteúdo.
RQ06	A solução deve possuir mecanismo de busca por todo o conteúdo textual.
RQ07	A solução deve contemplar mecanismo de classificação de conteúdo por rótulos.
RQ08	A solução deve prover conteúdo geográfico segundo padrões abertos OGC (WMS, WFS, WCF, WPS).
RQ09	A solução deve usar produtos de *software* livre, preferencialmente de código fonte aberto e na linguagem Java.

(continua)

Tabela 7.1 – Conclusão.

RQ10	A solução deve contemplar a manipulação, armazenamento e recuperação de imagens vetoriais e matriciais.
RQ11	A solução deve contemplar o agrupamento e o relacionamento de conteúdos.
RQ12	A solução deve contemplar o referenciamento geográfico dos conteúdos.
RQ13	A solução deve contemplar a plotagens dos elementos georreferenciados no mapa.
RQ14	A solução deve ser capaz de consumir serviços *web* de geolocalização, WPS e mapas base disponíveis na Internet.

Antes de iniciar o desenvolvimento da solução, alguns pontos e restrições já eram conhecidos e foram adotas as seguintes premissas:

a) Os padrões OGC estão em constante amadurecimento, porém o paradigma de orientação a serviço impõe restrições de desempenho que serão aceitas; e

b) A solução pode ser composta por diversos subsistemas distribuídos, que serão integrados através do uso de padrões abertos, deixando a responsabilidade de orquestrar os serviços envolvidos para a aplicação cliente.

De certa forma, no cenário descrito tem-se a perda de uma parcela do conhecimento produzido e isto influenciou este trabalho a investigar e pesquisar assuntos relacionados à gestão do conhecimento. Nonaka e Takeuchi (1997) forneceram uma grande contribuição para o assunto, sendo suas obras as maiores referências atualmente. Estes autores realizaram uma importante distinção entre os tipos de conhecimento humano, classificando-os em conhecimento tácito e conhecimento explícito. Os conhecimentos explícitos são aqueles estruturados e capazes de serem verbalizados, facilmente transmitido, sistematizado e comunicado. Enquanto os conhecimentos tácitos são aqueles inerentes às pessoas, isto é, o conhecimento pessoal incorporado à experiência individual, crenças e valores. É difícil de ser articulado na

linguagem formal e transmitido por se tratar da parcela não estruturada do conhecimento.

Nonaka e Takeuchi (1997), ainda, consideram que um trabalho efetivo com o conhecimento somente é possível em um ambiente em que possa ocorrer a contínua conversão entre esses dois formatos. Segundo estes autores são 4 os processos de conversão entre os dois tipos de conhecimento: socialização, externalização, combinação e internalização.

Gestão do conhecimento é um tema relativamente novo, multidisciplinar e muito explorado em diversas pesquisas, mas quase sempre seu foco são as organizações empresariais. No entanto, há iniciativas da aplicação destes conceitos da gestão de conhecimento sob o âmbito do conhecimento científico como dissertado, principalmente, em Leite (2006).

Alguns autores vêm desenvolvendo trabalhos sobre o tema de gestão de conhecimento científico e estudando ferramentas e alternativas para auxiliar e facilitar os processos envolvidos em tais atividades. Leite e Costa (2006) discutem a adequação e aplicabilidade de repositórios institucionais como uma ferramenta para tal, abordando as peculiaridades do conhecimento científico, bem como o ambiente no qual ocorrem os processos de sua criação, compartilhamento e uso.

Partindo dos resultados de Leite e Costa (2006) observou-se como as infraestruturas computacionais poderiam contribuir para a gestão do conhecimento e facilitar a contínua conversão entre os formatos de conhecimento discutidos em Nonaka e Takeuchi (1997). Tais infraestruturas devem oferecer funcionalidades que auxiliam e estimulam a dinâmica da espiral de criação do conhecimento, conforme discorrido a seguir:

1. Estimula o processo de **socialização** do conhecimento visto sob a ótica de algumas características básicas. Primeiro, a diversidade de formatos em que as informações podem existir, tais como: artigos científicos, livros eletrônicos, capítulos de livros, relatórios técnicos, textos para

discussão, teses, dissertações, anotações de campo, apresentações, fotos, vídeos, imagens de satélite, planilhas etc. A possibilidade em combinar estas mídias no processo de comunicação aumenta as condições favoráveis à assimilação do conhecimento. O resultado é uma transferência da informação e do conhecimento mais efetiva, pois muito do conhecimento científico gerado por um pesquisador não é possível de ser comunicado por meios formais como os artigos em periódicos ou livros e transforma parte do conhecimento que antes era puramente tácito e transferido apenas informalmente em conhecimento explícito.

Ainda sob a ótica da socialização, é estimulada a interação informal entre pesquisadores interessados em um mesmo assunto. Fato que possibilita discussões; compartilhamento de ideias e esboços; coleta de sugestões e comentários enriquecendo as pesquisas e intensificando a troca de experiências.

2. Instrumento de **externalização** do conhecimento tácito que, segundo Nonaka e Takeuchi (1997), trata-se do processo de criação do conhecimento perfeito, ao fornecerem a possibilidade de armazenar múltiplos formatos desse conhecimento. As publicações científicas são formais e desta forma formatam o conhecimento e de certa forma limitam seus horizontes. Uma infraestrutura capaz de armazenar os conhecimentos informais aproxima os demais pesquisadores aos elementos que compõe o estado do conhecimento de seu autor. Neste cenário, parte do conhecimento tácito é transformado em uma estrutura comunicável permitindo que esta seja processada, armazenada e recuperada.

3. Permite a transformação de um determinado conjunto de conhecimento explícito, por meio de agrupamento, acréscimo, categorização e classificação, criando um novo conjunto de conhecimento ou criando e/ou acrescentando um novo conhecimento, constituindo, assim, o processo de **combinação**.

4. Facilita o processo de **internalização** por criar condições favoráveis para que o conhecimento explícito armazenado seja convertido em conhecimento tácito do indivíduo.

Estas funcionalidades estão alinhadas aos requisitos já mencionados, principalmente os requisitos RQ01, RQ02, RQ04, RQ07 e RQ11. Parte destas funcionalidades é encontrada em Sistemas de Gerenciamento de Conteúdo.

Os Sistemas de Gerenciamento de Conteúdo são ferramentas que promovem meios de gerenciamento, publicação e manutenção dos ativos de informação. Esta categoria de sistemas ainda incluem funcionalidades de fórum, listas de discussões, *workflows*, controle de acesso, associações, classificação e categorização, o que cria um ambiente propício para gestão do conhecimento uma vez que facilitam a existência, manutenção e crescimento dos processos de transformação citados em Nonaka e Takeuchi (1997).

A adoção de uma ferramenta de CMS conseguiu elucidar uma parte dos requisitos, conseguindo gerir os conteúdos de diversos formatos e provendo uma série de ferramentas sobre os mesmos. No entanto, ainda havia alguns requisitos que não eram contemplados por esta ferramenta. Neste caso os requisitos que fogem ao escopo de atuação das ferramentas de CMS são, em essência, relacionados ao contexto espacial. Desta forma, estes requisitos foram tratados fora do CMS, incluindo na arquitetura da solução um elemento de gerenciamento de conteúdo geográfico.

Tendo em vista a necessidade de relacionar um artefato a algum ponto no globo terrestre, esta abordagem implica em uma integração entre a ferramenta de CMS e a ferramenta de gestão de conteúdo geográfico. Esta integração foi realizada através do padrão aberto *Content Management Interoperability Services*.

7.2. Arquitetura Proposta

A macroarquitetura proposta para a solução do problema descrito no Capítulo 6 pode ser vista na Figura 7.2. Ela foi concebida para manipular conteúdos sejam

eles georreferenciados ou não. A arquitetura proposta é constituída por 3 subsistemas distribuídos e organizados segundo o paradigma de orientação a serviços. O primeiro subsistema é o que se convencionou chamar de Serviço de CMS (1), que neste caso foi disponibilizar as funcionalidades do *software* Alfresco como serviços *web*. Este subsistema está identificado pelo número 1 na Figura 7.2. Ele é responsável por armazenar diversos tipos de conteúdos e mídias, a citar os documentos de texto, planilhas, fotos, vídeos, áudios etc. Este subsistema oferece todo o ferramental de um CMS sobre tais conteúdos, como classificação, agrupamento, buscas textuais, comentários, discussões, compartilhamento e o controle de restrição de acesso por usuário.

O segundo subsistema compõe o que se chamou de SOA geoespacial, constituído do Serviço de Conteúdo Geográfico (2), o GeoPortal (3) e seus provedores de dados (6) e serviços externos (7). O GeoPortal, aqui, desempenha o papel de catalogador de serviços, sejam eles conteúdos (WMS, WFS, WCS) ou processamentos (WPS), análogo ao UDDI na arquitetura SOA. O Servidor de Conteúdo Geográfico (2) desempenha papel análogo ao Provedor de Serviço na arquitetura SOA, fornecendo os dados e processamentos registrado no GeoPortal sob as especificações da OGC WMS, WFS, WCS e WPS.

O cliente de serviços Drikka (5) é uma aplicação cliente capaz de consumir dados e serviços nos padrões fornecidos tanto pelo Servidor de Conteúdo Geográfico (2) quanto por provedores externos (6 e 7) registrados no GeoPortal (3). O Drikka desempenha papel semelhante ao Consumidor de Serviços da arquitetura SOA. Como o conteúdo do CMS e os conteúdos georreferenciados existentes no Serviço de Conteúdo Geográfico (2) e nos provedores externos (6 e 7) estão separados e possuem relacionamento, há a necessidade de um serviço capaz de relacioná-los. Este relacionamento é realizado pelo terceiro subsistema da arquitetura, chamado Serviço de Relacionamento GIS – CMS (8), possibilitando relacionamento do tipo muitos para muitos entre as partes. Por fim, tem-se o banco de dados Postgres 9.1 com sua extensão espacial PostGIS 2.0 servindo aos demais componentes da arquitetura.

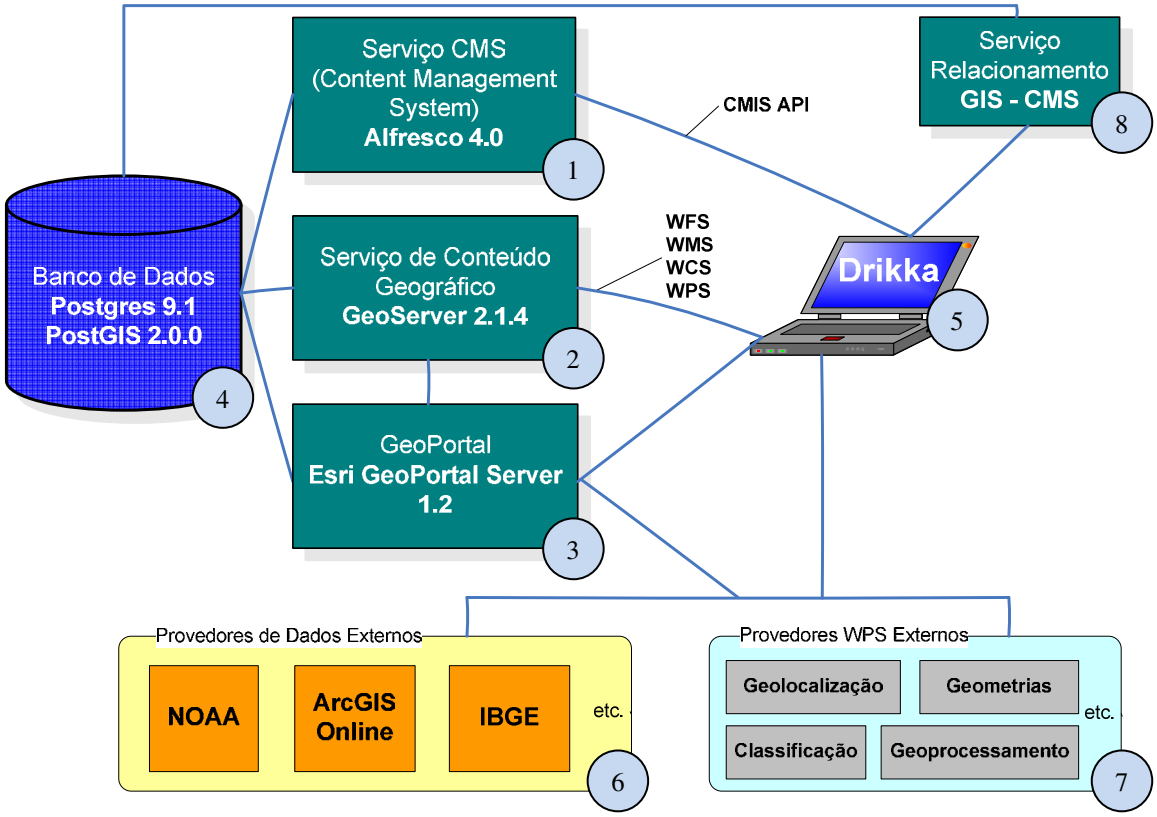

Figura 7.2: Macro arquitetura proposta.

7.3. Realização dos Requisitos

A adoção de um *software* de CMS introduz muitas facilidades a solução, dentre elas o mecanismo de fórum, lista de discussões, controle de acesso de usuários e o armazenamento de arquivos multimídias. O *software* de CMS é utilizado para realizar os requisitos RQ01, RQ02, RQ03, RQ04, RQ05 e RQ07. O requisito RQ06, para os conteúdos multimídias é também atendido pelo CMS, bastando ativar a busca por *full-text* do *software* escolhido. Para os artefatos espaciais ele é atendido através do mecanismo de busca do GeoPortal, que realiza uma busca textual pelos metadados dos artefatos. Os requisitos RQ08, RQ10 e RQ14 são elucidados através do GeoSever. O requisito RQ09 é um requisito arquitetural e cumprido na íntegra. Para a realização dos requisitos RQ11 e RQ12 foi desenvolvido o Serviço de Relacionamento GIS – CMS, melhor discutido na Seção 7.6. Por fim, o requisito RQ13 é cumprido ao desenvolver o componente cliente da solução

baseada numa aplicação de mapas, plotando os elementos num mapa segundo suas características espaciais.

7.4. Serviço CMS

O Serviço de CMS é provido pelo *software* Alfresco, em sua versão 4.0. Alfresco é desenvolvido em Java e requer o Java *runtime* 1.7 para funcionar corretamente. Para a publicação da aplicação há diversos servidores *web* disponíveis e compatíveis, porém foi escolhido o Apache TomCat 7 (TOMCAT, 2012) por ser recomendado pelo fabricante (ALFRESCO, 2012) e por ser um servidor *web* compatível com os demais elementos da arquitetura.

O Alfresco possui uma versão *community*, que é de código aberto e de distribuição gratuita, podendo ser adquirida no sítio do principal mantenedor `http://www.alfresco.com/latest-community`. Há versões disponíveis para os ambientes Windows, Linux e Mac.

Na estrutura do Alfresco todo conteúdo é representado por um tipo específico de dado denominado *content* (cm:content). Um *content* pode representar um arquivo multimídia como um arquivo de texto, foto, vídeo ou outro qualquer, mas pode também representar um conjunto ou agrupamento de arquivo análogo a um diretório (*folder*). Desta forma cada um dos conteúdos existentes no Alfresco é um *content* e possui um identificador único.

O tipo de dado *content* pode ser especializado num esquema de herança e possuir atributos e comportamentos específicos. Um tipo especializado de *content* é o tipo *folder* (cm:folder), que assim como mencionado, possibilita agrupar um conjunto de outros *contents* e ainda possuir atributos específicos.

Foi criado um novo tipo de conteúdo, herdando de *content*, para representar os arquivos multimídias chamado `drikkaDoc` (dk:drikkaDoc). Este tipo representa diversos tipos de arquivos tais como documentos de texto, planilhas, apresentações multimídias, áudio, vídeo e fotos. Para descrever cada um destes arquivos foi escolhido um conjunto de atributos do metadado *Dublin Core*. À medida que surgir determinado conteúdo que precisar ser manipulado

pelo CMS outros tipos podem ser criados herdando de `drikkaDoc` e adicionando atributos específicos ou utilizando adicionalmente outro conjunto de metadados. Podem ser acrescentados tantos atributos quanto forem necessários para descrever ou representar um dado ou conteúdo no CMS, sejam estes atributos pertencentes a algum conjunto de metadados conhecido ou não.

Um segundo tipo criado foi o tipo `drikkaPesquisa` (dk:drikkaPesquisa), herdando de *folder*. Este tipo representa uma pesquisa de pós-graduação. Partindo-se do princípio de que em uma pesquisa de pós-graduação são gerados diversos tipos de dados, artefatos ou conteúdos faz todo o sentido que esta seja capaz de agrupar uma série de outros conteúdos. Este tipo também possui o conjunto de atributos do metadado *Dublin Core* e assim como o tipo `drikkaDoc` pode ser herdado ou adicionado atributos que venham a fazer sentido.

Para fins de exemplo considere a situação onde uma nova pesquisa de mestrado é iniciada, haverá um diretório virtual utilizando o tipo `drikkaPesquisa` com os atributos sobre a mesma. Haverá dados sobre o autor, assunto, descrição e quais outros dados fizerem sentido. Dentro de `drikkaPesquisa` poderá haver diversos conteúdos (`drikkaDoc`) ou mesmo diretórios (folder).

Esta situação pode ser melhor explicada considerando a Figura 7.3, onde o conteúdo do tipo `drikkaPesquisa` (1) é apresentado, logo a direita são apresentados os seus atributos (3) mais importantes e abaixo os rótulos (*tags*) ou palavras-chave sobre a pesquisa (4). Dentro do diretório da pesquisa (1) ainda há diversos diretórios (2) para organizar os artefatos que a compõem.

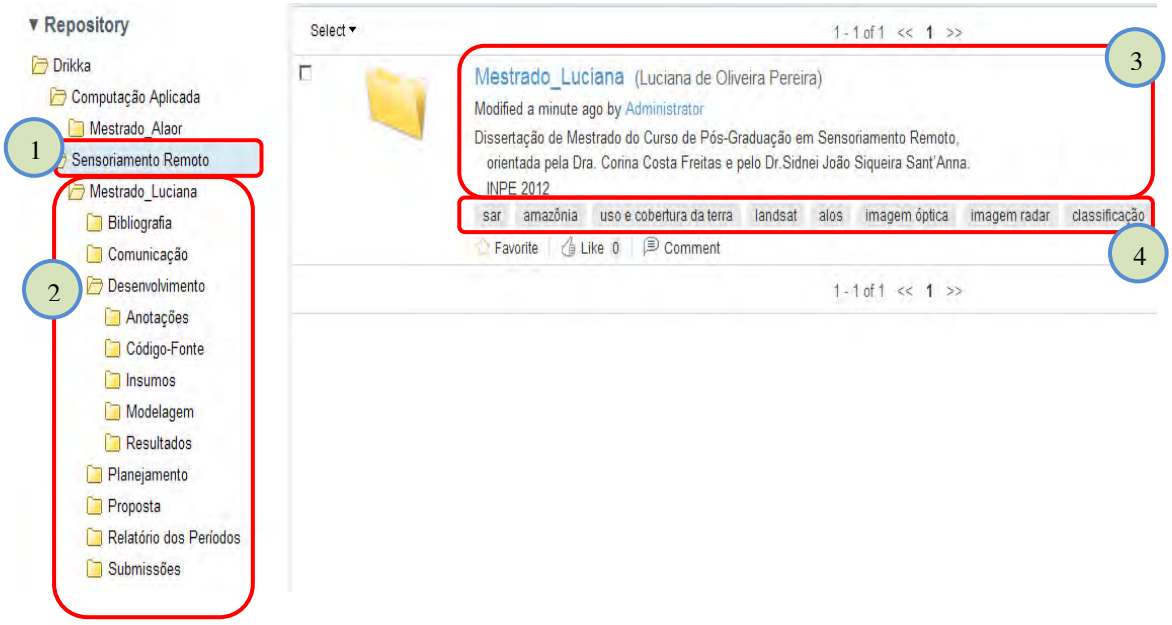

Figura 7.3: Exemplo de estrutura de uma pesquisa de mestrado.

A árvore de diretórios apresentada na Figura 7.3 em (2) representa os diretórios padrão quando se cria um novo diretório virtual do tipo `drikkaPesquisa` para um novo aluno ou usuário. Estes diretórios formam uma estrutura básica que se entendeu como sendo recorrente em grande parte das pesquisas. Nenhum destes diretórios é obrigatório conter algum conteúdo, mas recomenda-se que a convenção seja respeitada para que intuitivamente todos saibam onde encontrar determinada informação sobre uma pesquisa uma vez que já usou a solução ou conhece a convenção.

7.5. Serviço de Conteúdo Geográfico

O Serviço de Conteúdo Geográfico é provido pelo *software* GeoServer, em sua versão 2.1.4. O GeoServer é um *software* desenvolvido em Java e neste cenário foi publicado no servidor *web* TomCat 7.

A principal responsabilidade do GeoServer neste cenário é prover os dados nos padrões OGC WFS, WMS e WCS e serviços de processamento segundo o padrão WPS. O GeoServer é capaz de fornecer dados a partir de várias fontes distintas como *shapefiles*, banco de dados geográficos, cascatear dados WFS

e WMS de outros servidores *web*. Neste projeto, uma grande parte dos dados se encontram no banco de dados geográfico PostGIS.

Há duas camadas particulares registradas no GeoServer, são elas: `Pontos_Interesse` e `Area_Interesse`. Estas são as camadas dos dados geográficos que se relacionarão com os conteúdos do CMS e serão discutidas adiante na Seção 7.6, ou seja, um conteúdo no CMS pode estar relacionado a uma ou mais feições nestas camadas. Na Figura 7.4 pode-se verificar que estas camadas, sendo representadas como tabelas no banco de dados geográfico, são muito simples, ambas possuem um atributo identificador único chamado "GIS_ID" que é um Identificador Único Universal (*Globally Unique Identifier* - GUID) e um campo "*Geometry*" que é do tipo *Geometry*, sendo *Geometry Polygon* para a camada Area_Interesse e *Geometry Point* para Ponto_Interesse.

Figura 7.4: Camadas especiais.

No GeoServer foi criada uma *workspace* chamada `drikka`, para ser exclusiva para a solução. Uma *workspace* no GeoServer é geralmente utilizada para agrupar camadas ou temas e ao se criar uma nova fonte de dados deve-se escolher esta *workspace*. Como estas camadas especiais são tabelas geográficas no PostGIS, foi criado uma fonte de dados (*store*) no GeoServer, com o nome de `DrikkaPostGIS`, conectando-se ao PostGIS e publicadas as duas camadas, Area_Interesse e Ponto_Interesse, como serviço de feição (WFS).

7.6. Serviço Relacionamento GIS – CMS

Todos os conteúdos ou artefatos estão armazenados no CMS, ou seja, todos os documentos de texto, planilhas eletrônicas, fotos, áudio, vídeos etc. Estes artefatos podem estar relacionados a algum ponto ou área no globo terrestre. Determinado aluno pode desenvolver sua pesquisa e esta, como um todo, pode estar relacionada a uma área da Floresta Nacional do Tapajós e deseja-se representar este relacionamento. Algumas fotos podem estar relacionadas ao cultivo de milho numa área menor ou algumas medições podem estar relacionadas a algum equipamento específico num determinado ponto geográfico.

Alguns artefatos possuem informação de sua componente espacial, caso de fotos digitais obtidas a partir de equipamentos que possuem GPS embutido e conseguem inserir as informações (metadados) de latitude e longitude (*GeoTagging*), mas alguns formatos de mídias não possuem esta capacidade e mesmo assim suas informações possuem relevância geográfica. Neste caso, podem-se citar como exemplos, anotações de campo, planilhas de medições etc.

O Serviço de Relacionamento GIS - CMS é capaz de relacionar um elemento existente no CMS com um elemento existente no Gerenciador de Conteúdo Geográfico, ou melhor dizendo, é capaz de relacionar vários elementos no CMS com vários elementos no Gerenciador de Conteúdo Geográfico. Isso é possível devido ao fato de um elemento no CMS possuir um identificador único e um elemento geográfico, no Gerenciador de Conteúdo Geográfico, também o possuir.

O Serviço de Relacionamento consiste de um serviço *web*, descrito em WSDL e desenvolvido em Java com o auxílio do *framework* JavaWS. Este serviço *web* disponibiliza algumas operações para relacionar elementos do CMS aos elementos geográficos, presentes no GeoServer ou acessíveis pelo GeoPortal. Na Figura 7.5 é possível observar as operações disponíveis no serviço *web*, sendo as operações `getRelatedToCms`, `getRelatedToLayer`,

`getRelatedToPoint` e `getRelatedToPolygon` as operações de recuperação dos relacionamentos e as operações `relateToLayer`, `relateToPoint` e `relateToPolygon` as operações que realizam o relacionamento entre os elementos do GIS e do CMS.

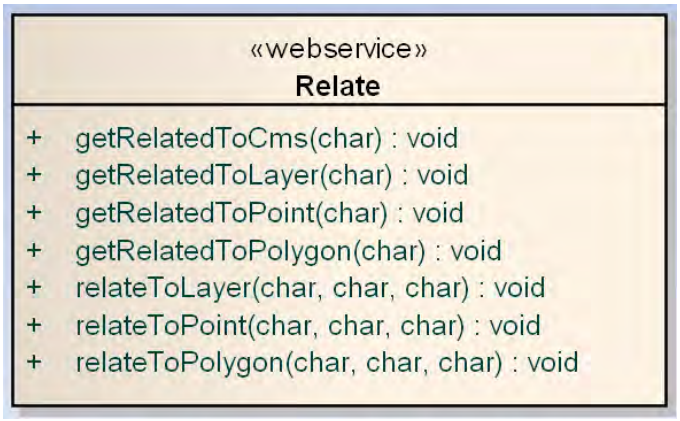

Figura 7.5: Representação do Serviço de Relacionamento GIS – CMS.

Na Tabela 7.2 é possível verificar as operações disponíveis no Serviço de Relacionamento GIS – CMS, bem como sua descrição e parâmetros necessários para execução.

Tabela 7.2: Operações do Serviço de Relacionamento GIS – CMIS.

Operação	Descrição	Parâmetros
`getRelatedToCms`	Obtém a lista de elementos relacionados ao elemento do CMS passado como parâmetro.	**cmsUri**: do tipo texto representando a identificação do elemento do CMS.
`getRelatedToLayer`	Obtém a lista de elementos relacionados a camada de tema geográfico passado como parâmetro.	**gisUri**: do tipo texto representando a identificação da camada do tema geográfico.

(continua)

Tabela 7.2 – Conclusão.

`getRelatedToPoint`	Obtém a lista de elementos relacionados a feição de ponto passada como parâmetro.	**gisUri**: do tipo texto representando a identificação da feição de ponto.
`getRelatedToPolygon`	Obtém a lista de elementos relacionados a feição de polígono passada como parâmetro.	**gisUri**: do tipo texto representando a identificação da feição de polígono.
`relateToLayer`	Relaciona um elemento do CMS a uma camada de tema geográfico.	**cmsType**: do tipo texto representando o tipo de elemento do CMS, sendo "CONTENT" ou "FOLDER". **cmsUri**: do tipo texto representando a identificação do elemento do CMS. **gisUri**: do tipo texto representando a identificação da feição (ponto ou polígono).
`relateToPoint`	Relaciona um elemento do CMS a uma feição de ponto.	
`relateToPolygon`	Relaciona um elemento do CMS a uma feição de polígono.	

7.7. GeoPortal Server

Antes da inclusão do GeoPortal, não existia na solução um mecanismo de descoberta dos dados geográficos. Ainda, a inclusão ou uso de dados que não estivessem publicados no Gerenciador de Conteúdo Geográfico era trabalhosa, pois era necessário replicá-la. Considerando o fato de diversos trabalhos já terem sido realizados e terem gerado um volume considerável de dados e informações e estes já estarem disponíveis como serviço, faz todo sentido usá-los sem ter que replicar ou duplicá-los. Supondo também que as agências e órgãos nacionais, tais como IBGE, ANA, ANEEL e entre outras, produzem dados e informações geográficas e os disponibilizem, deveria haver uma forma de descobrir estes dados e usá-los a partir de sua própria origem. Uma forma eficiente de fazer isso é adotar um Catálogo de Dados Geográficos, preferencialmente usando padrões abertos. No caso da arquitetura proposta foi adotado o padrão CSW da OGC.

Para desempenhar este papel de catálogo de dados dentro da solução, foi adotado o `Esri GeoPortal Server`. Cada um dos provedores de dados e informações geográficas que possuírem um catálogo e implementar o padrão CSW poderá ser adicionado como uma fonte de busca dentro do `GeoPortal`. Inicialmente foram adicionados os provedores `IBGE`, `ArcGIS.com` e `NOAA`, mas vale ressaltar que qualquer provedor pode ser adicionado, desde que implemente os padrões OGC.

Quando se adiciona uma nova fonte de busca têm-se duas opções para realizar as buscas no mesmo. A primeira é fazer uso da operação `Harvest` do padrão CSW e criar uma base local dos metadados disponíveis nesta fonte de dados. A segunda opção é realizar buscas federadas nas fontes de dados. Para os provedores adicionados inicialmente optou-se pela segunda opção para que a busca sempre reflita a realidade naquele momento, uma vez que o sincronismo pela operação `Harvest` é agendada e em dado momento pode estar desatualizada ou não refletir a realidade.

A adoção do `GeoPortal` não se limita apenas aos momentos de consumir dados e informações externas, mas também adiciona a possibilidade e capacidade de ser também um provedor de dados, ou seja, distribuir os dados existentes no Gerenciador de Conteúdo Geográfico através das funcionalidade de um catálogo de dados geográfico e do padrão CSW. Esta infraestrutura é a base tecnológica para a criação e participação em uma Infraestrutura de Dados Espacial, seja como consumidor ou provedor de dados. O padrão de metadado adotado é o Perfil de Metadados Geoespaciais do Brasil (Perfil MGB) homologado pela CONCAR e parte das diretivas para a elaboração da Infraestrutura Nacional de Dados Espaciais.

A adoção do `GeoPortal` traz, ainda, benefícios para efetuar buscas textuais ou por um envelope geográfico. As buscas textuais são realizadas através dos atributos descritivos dos metadados e o motor de busca utilizado é o `Lucene`, que indexa o conteúdo textual em arquivos no *filesystem* para tornar as buscas mais rápidas e eficientes. É possível também criar um serviço de ontologia e configurar o motor de buscas para considerá-lo.

7.8. PostGIS

Vários elementos da arquitetura proposta utilizam um banco de dados relacional para persistirem seus dados. Este é o caso do Serviço de CMS, Serviço de Conteúdo Geográfico, GeoPortal Server e o Serviço de Relacionamento GIS – CMS. O caso do Serviço de Conteúdo Geográfico, além de seus dados administrativos, alguns temas geográfico também são armazenados em banco de dados relacional (Postgres) através da extensão espacial PostGIS.

Todos estes serviços podem compartilhar um mesmo banco de dados, mas para manter certa independência adotou-se a estratégia de criar um *Database*[4] para cada um deles, conforme descrito na Tabela 7.3.

[4] Um database no Postgres é uma base de dados que existe dentro de uma mesma instalação do mesmo, podendo haver várias para uma mesma instalação dentro de um mesmo servidor.

Tabela 7.3: Serviços e seus respectivos *databases.*

Database	Serviço
postgis	Serviço de Conteúdo Geográfico
alfresco	Serviço de CMS
geoportal	GeoPortal Server
relationship	Serviço de Relacionamento GIS – CMS

7.9. Aplicação Cliente

A aplicação cliente foi desenvolvida em tecnologia `Adobe Flex`, com linguagem de programação *Actionscript* 3 (ADOBE, 2011). Uma importante decisão de projeto foi quanto ao desenvolvimento da aplicação cliente. Optou-se por desenvolver uma aplicação cliente com certo grau de inteligência (*smart client*), neste caso essa inteligência é justamente a capacidade de se comunicar com os demais elementos da arquitetura e ser o ponto de integração entre alguns deles, conforme exposta na Figura 7.2, onde se pode observar que os serviços não se comunicam entre si, mas sim através da aplicação cliente.

Para tornar a solução adotada mais generalista, com capacidade e possibilidade de se substituir facilmente alguns dos subsistemas ou de escalá-los individualmente, optou-se pela abordagem de sistemas distribuídos, tratando-os como subsistemas independentes. Estes subsistemas são distintos e com responsabilidades bem definidas, mas em dado momento eles precisam colaborar entre si, como por exemplo, o relacionamento entre um elemento geográfico e um conteúdo no CMS.

Para este cenário, optou-se por uma aplicação cliente que seja capaz de servir-se dos subsistemas através de protocolos e padrões abertos e também que seja capaz de orquestrar a integração entre eles, como é o caso do relacionamento entre o elemento geográfico e o conteúdo no CMS. A aplicação cliente é capaz de fazer uso dos padrões OGC WMS, WFS e WCS para recuperar os elementos geográficos e fazer uso do padrão CMIS para recuperar os conteúdos no CMS. De posse dos identificadores de cada um

destes elementos é possível relacioná-los através de um serviço específico para tal. As decisões de como realizar esta operação e como orquestrar os serviços para colaborarem fica a cargo desta aplicação cliente.

Na Figura 7.6 pode ser observada estrutura básica da aplicação cliente. Pode ser visto os subsistemas em (1) ofertando seus serviços especialistas, a aplicação cliente em (2) e cada um dos seus módulos (3) que fazem uso dos serviços dos subsistemas.

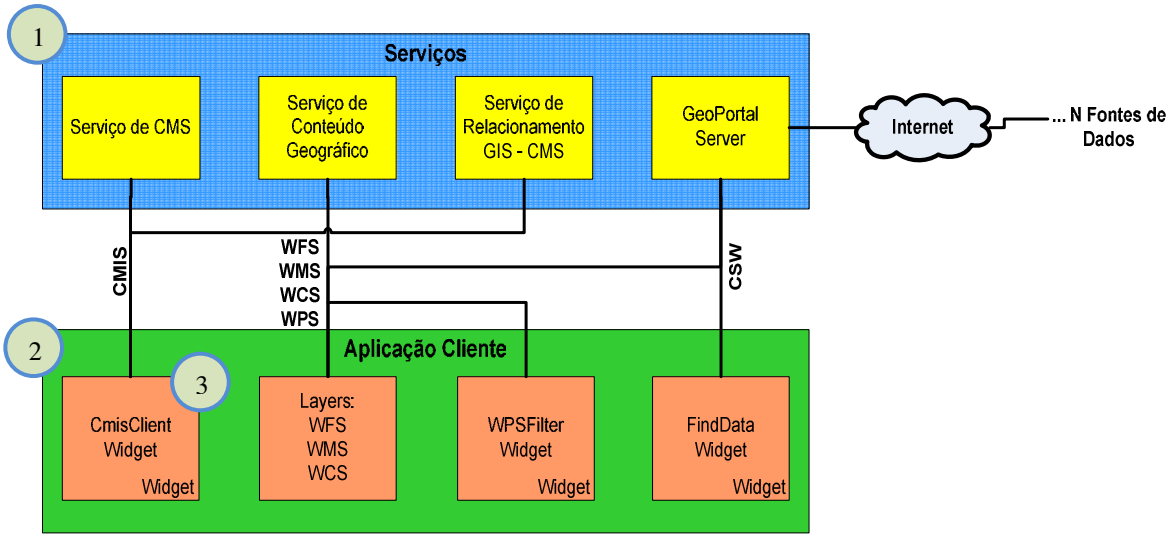

Figura 7.6: Estrutura da aplicação cliente.

Uma aplicação desenvolvida em `Flex` gera um arquivo de extensão swf, que é um arquivo binário a ser baixado pelo navegador do usuário, quando este inicializar a aplicação, é executado através de um *plugin* disponibilizado pela Adobe, o *Flash Plugin*. Para aplicações *web* devem-se ter alguns cuidados, tais como o tamanho que este arquivo swf pode atingir, já que o mesmo será baixado para a máquina do usuário assim que o usuário acessar a aplicação.

A medida que a aplicação crescer e novas funcionalidades forem adicionadas este arquivo pode aumentar de tamanho e demorar para ser baixado pelo navegador do usuário. Para evitar este problema a aplicação foi dividida em módulos. Os módulos são arquivos SWF que podem ser carregados e descarregados por um aplicativo. Eles não podem ser executados

independentemente de uma aplicação, mas qualquer número de aplicativos podem compartilhar os módulos. Módulos permitem dividir a aplicação em várias peças ou módulos. A aplicação principal pode carregar dinamicamente outros módulos assim que necessite dele e não é necessário carregar todos os módulos quando inicializa a aplicação, nem se faz necessário carregar todos os módulos, evitando assim que se carregue um módulo cujo qual o usuário não irá interagir. Quando o aplicativo não precisar mais de um módulo, ele pode descarregá-lo para liberar memória e recursos (ADOBE, 2011).

Outro benefício importante da adoção de módulos é o fato de não precisar recompilar e publicar toda a aplicação para adicionar um novo módulo. Estes podem ser adicionados e usados pela aplicação principal bastando indicar a URL de onde o mesmo está disponível e possíveis configurações específicas que este venha a possuir.

Figura 7.7: *Widget*.

Os módulos da aplicação foram desenvolvidos em formato de *widgets*. *Widgets* são componentes de uma interface gráfica de usuário e neste caso são janelas que provêm alguma funcionalidade. Na Figura 7.7 pode se ver um *widget* em (1) que provê as funcionalidades de integração com o CMS, listando e

interagindo com seus conteúdos. Este *widget* é também representado na Figura 7.6 que fornece uma visão da organização interna dos módulos e a interação de cada um com os serviços da arquitetura proposta, através do item (3) como `CmisClientWidget`.

7.9.1. Módulos

A aplicação cliente conta com mais alguns módulos desenvolvidos como *widgets,* descritos a seguir.

`CmisClientWidget`

Este módulo é responsável pelas interações com o CMS, sendo capaz de listar seus conteúdos navegando pela estrutura hierárquica de diretórios ou realizando buscas textuais por estes conteúdos. É possível recuperar as informações detalhadas de cada um dos conteúdos e seus metadados bem como baixar o conteúdo original do CMS.

Todas estas operações são realizadas através do padrão CMIS e para tal foi utilizada a biblioteca `flex-cmis-client`[5] como base com algumas funcionalidades estendidas para atender as necessidades do projeto.

`WPSFilterWidget`

Este módulo é responsável por executar um processamento de filtragem de uma imagem através de um serviço desenvolvido em Java e acessível através do padrão WPS da OGC. O serviço de filtragem desenvolvido é realizado através do filtro da média da vizinhança e disponibilizado no GeoServer usando a versão 1.0.0 da especificação WPS. A filtragem é realizada sobre uma banda da imagem por vez.

Na Figura 7.8 pode ser observado o *widget* deste módulo, sendo (1) a URL de uma imagem a ser processada, (2) o tamanho da janela de filtro, (3) a imagem original e (4) a imagem após o processo de filtragem.

[5] http://code.google.com/p/flex-cmis-client/

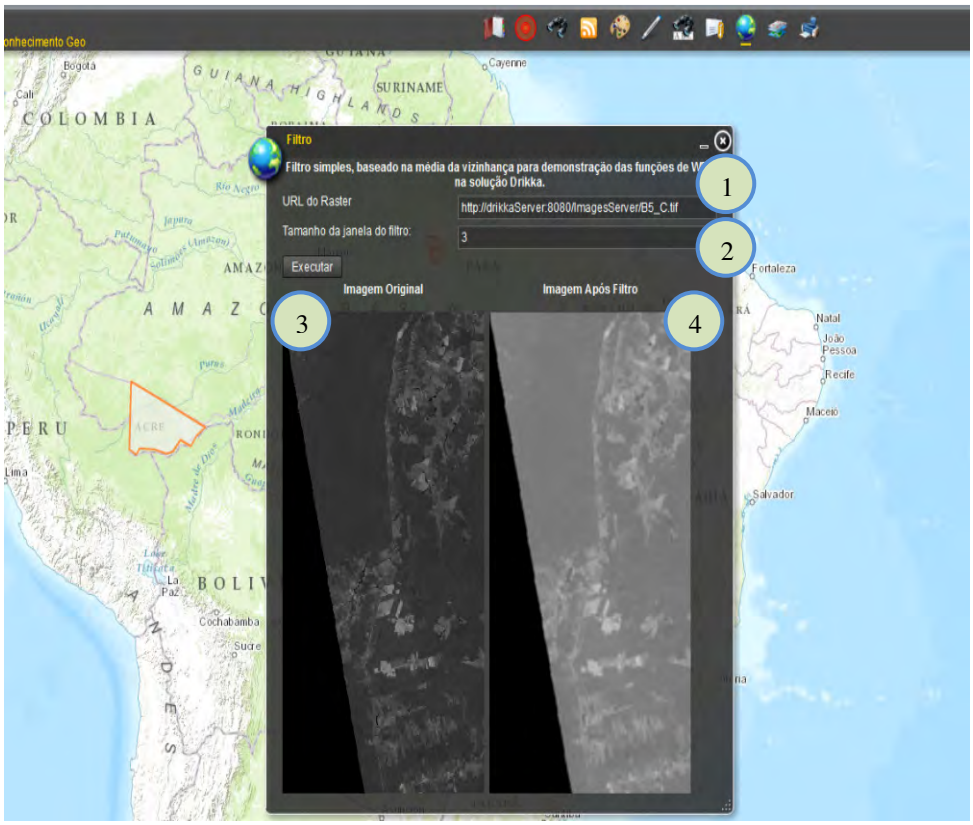

Figura 7.8: Módulo de filtragem digital de imagem.

`FindDataWidget`

Este módulo é responsável por executar as buscas no `GeoPortal` através do padrão CSW da OGC. Este módulo é visualizado na Figura 7.9, onde pode ser observada a caixa de entrada de textos (1) na qual são inseridos os termos para as buscas, a caixa de seleção do catálogo de dados onde a busca deverá ser executada (2) e alguns filtros espaciais que podem ser usados na busca (3). Já o retângulo em vermelho (4) apresentado na Figura 7.9 representa a extensão espacial de um resultado de uma busca.

O `GeoPortal` permite que os catálogos CSW sejam cascateados e que buscas sejam realizadas de forma federada nestes catálogos. Este módulo disponibiliza estas funcionalidades, por esta razão é apresentada a caixa de seleção para a escolha do catálogo que se deseja utilizar para a busca. Quanto aos filtros espaciais disponíveis é possível realizar uma busca textual que

inclua também aspectos espaciais, como o critério de que o resultado da busca deve estar contido ou deve se interseccionar com a visualização atual do mapa (envelope).

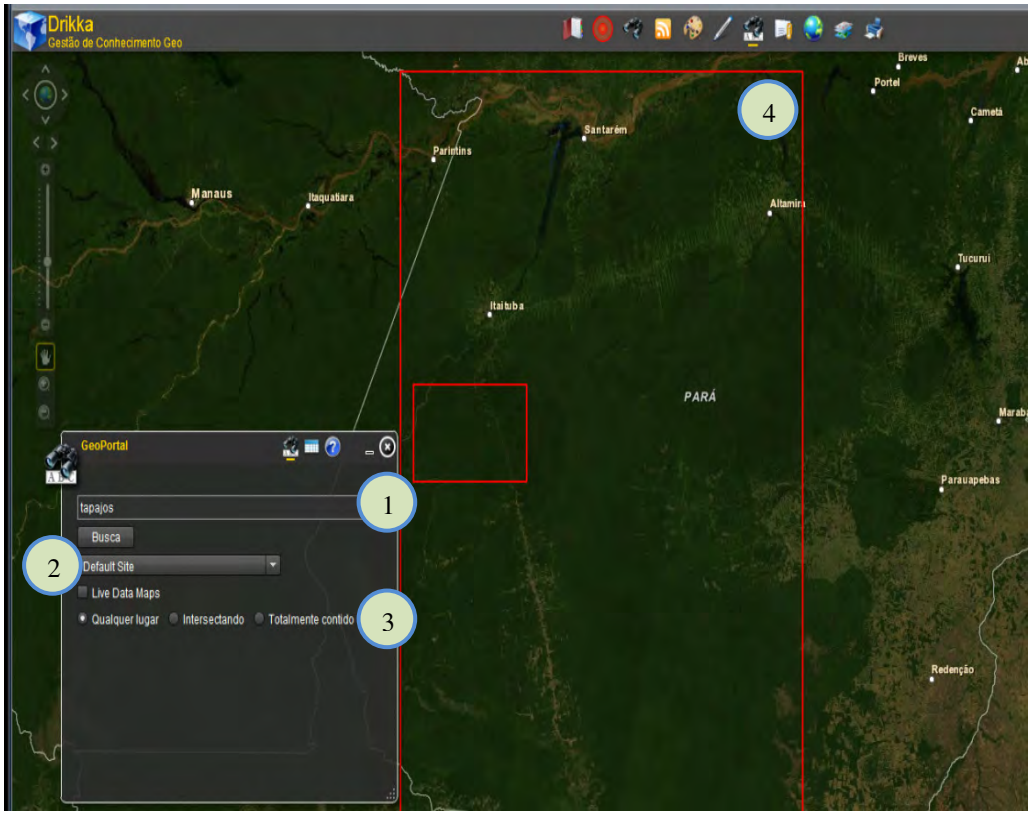

Figura 7.9: Módulo de busca no `GeoPortal`.

7.10. Considerações finais do capítulo

Neste capítulo foi apresentada a arquitetura construída como solução do problema proposto. Foi discutido cada um dos elementos que a compõe e como estes elementos se inter-relacionam. Foram apresentados os requisitos levantados juntamente com os usuários e como a solução desenvolvida endereça cada um deles.

A arquitetura da solução proposta provê um grande poder de adaptação e um fraco acoplamento entre os principais componentes, isso se deve em grande parte por fazer uso do paradigma de orientação a serviço. Os principais componentes interagem através de padrões e protocolos abertos e

interoperáveis, o que permite que sejam facilmente substituídos ou facilmente escalados, inclusive escalados individualmente.

Todos os componentes da solução são de código-fonte aberto, o que permitiu que adaptações pontuais fossem realizadas e é o que possibilita que a solução seja evoluída e melhorada. Todo o código-fonte escrito está disponível no GitHub[6] e pode ser facilmente obtido e estendido.

A organização e desenvolvimento da aplicação cliente de forma modular, permite que sejam carregados apenas os módulos que realmente serão usados, mas, introduz também uma facilidade para se criar e adicionar novos módulos e funcionalidades. Isso é importante principalmente pelas pretensões futuras de disponibilizar os algoritmos de tratamento de imagens via Serviço de Processamento *Web* (WPS).

Os padrões OGC estão amadurecendo e alguns deles já estão amplamente em uso, caso do *Web Map Service*, *Web Feature Service* e o *Catalog Service for Web*. Há varias implementações de código-fonte aberto, inclusive várias implementações de referência, como é o caso do `GeoServer` com o padrão WMS. Com o amadurecimento das Infraestruturas de Dados Espaciais o padrão CSW está cada vez mais sendo utilizado e difundido. A adoção de padrões quase sempre representa um *trade-off* entre alguns fatores. Durante o desenvolvimento deste trabalho puderam-se experimentar dificuldades de desempenho ao usar o padrão WPS para processamento de imagens. O WPS é um padrão recente, ainda em sua versão 1.0.0 e dependendo de sua aplicação pode representar o movimento de um grande volume de dados entre nós de uma rede, como o caso de enviar uma imagem para ser processada, o que em ambiente *web* pode ser um grande desafio. Em contrapartida representa uma grande evolução no compartilhamento de conhecimento sob a forma de algoritmos implementados e prontos para serem utilizados e aplicados em um problema real.

[6] www.github.com/alaorbianco

8 ESTUDO DE CASO

Este capítulo aborda o estudo de caso conduzido para avaliar a aplicação da arquitetura proposta e desenvolvida neste trabalho. Para este estudo de caso foi escolhido um trabalho de pesquisa referente a uma região da Floresta Nacional do Tapajós, no estado do Pará. O trabalho de pesquisa escolhido para o estudo de caso utiliza vários artefatos multimídias e espaciais, produzindo outros artefatos como resultados. O que se deseja avaliar é a capacidade que a arquitetura proposta possui de gerir a heterogênea variedade de artefatos utilizados ou gerados durante esta pesquisa e a facilidade de acessá-los e utilizá-los posteriormente por outros pesquisadores.

Neste capítulo será apresentada a região da Floresta Nacional do Tapajós, sua importância para o âmbito científico e alguns dos trabalhos conduzidos nesta região. Será descrito o trabalho de pesquisa utilizado para este estudo de caso, qual foi a dinâmica de execução deste trabalho quanto aos artefatos que ele utiliza e produz. Em seguida é apresentado um breve planejamento das atividades do estudo de caso em dois momentos distintos, o primeiro é o que um pesquisador faria enquanto desenvolve sua pesquisa e o segundo é o que um pesquisador faria para recuperar o contexto desta pesquisa. Por fim, a execução do estudo de caso propriamente dito e as considerações finais deste capítulo.

8.1. Região da Floresta Nacional do Tapajós

O ano de 1998 foi o marco inicial de um experimento mantido por um acordo de cooperação internacional chamado *Large Scale Biosphere-Atmosphere Experiment in Amazonia* (LBA). Em setembro de 2007, de experimento, o LBA tornou-se um programa de governo, renovando a agenda de pesquisas iniciada em 1998, sob a coordenação científica do Instituto Nacional de Pesquisas da Amazônia (INPA). É uma das maiores experiências científicas do mundo na área ambiental; somando 156 projetos de pesquisa (100 deles já finalizados), desenvolvidos por 281 instituições nacionais e estrangeiras (LBA, 2011).

O programa teve importante papel na formação de recursos humanos formando mais de 500 mestres e doutores brasileiros. Mais de 150 pesquisas da chamada "ciência de ponta", em parcerias com cerca de 281 instituições nacionais e estrangeiras, realizadas por 1400 cientistas brasileiros e outros 900 pesquisadores de países amazônicos, de 8 nações européias e de instituições americanas, visaram estudar e entender as mudanças climáticas e ambientais em curso, para favorecer um processo de desenvolvimento sustentável na Amazônia (LBA, 2011).

Entre 2000 e 2006, cotas de bolsas científicas possibilitaram o ingresso de 211 jovens pesquisadores nas atividades do Programa LBA sendo que 41% destas bolsas foram destinadas ao estado do Pará (LBA, 2011). Estas bolsas, juntamente com diversas outras iniciativas beneficiaram uma das áreas mais importantes da bacia amazônica, a região da Floresta Nacional de Tapajós e arredores, na cidade de Santarém.

A Floresta Nacional do Tapajós, situada no noroeste do Pará, é um ambiente único, composto por um conjunto de ecossistemas complexos, heterogêneos e que com o passar dos anos vem sofrendo muito com as pressões de uso e ocupação do solo (LBA, 1995). Esta condição elege este território como um grande candidato a diversas pesquisas e estudos.

Na Floresta Nacional do Tapajós já foram alcançados importantes resultados em pesquisas com o uso da tecnologia de sensoriamento remoto. Hernandez Filho et al. (1993) conseguiram estimar o estoque madeireiro da floresta com o uso de uma intensa base de dados de inventários florestais, imagens do satélite TM/Landsat e fotografias aéreas.

Shimabukuro et al. (1998) encontraram forte relação entre a geomorfologia e vegetação através da integração das imagens dos satélites TM/Landsat e RADARSAT. Yanasse et al. (1997) conseguiram estimar as fases sucessionais da vegetação secundária em uma porção da Floresta Nacional do Tapajós (faixa de 50 km x 23 km, ao longo da BR-163), com o uso de uma imagem SAR SIR-C (bandas C e L, polarizações HH e HV com 8 *looks*) e um acervo de

imagens TM/Landsat de dez anos (1984 a 1993). Verona (2002) identificou e quantificou a flutuação espectral da vegetação ocasionada pelas variações fisiológicas da floresta utilizando uma série de imagens multitemporais TM/Landsat e dados de índice de área foliar.

Diversos pesquisadores do INPE vêm desenvolvendo trabalhos nesta região gerando diferentes tipos de dados, Moura (2005) trabalhou com modelos digitais de superfície e modelo digitais de terreno gerados por interferometria nas bandas P e X. Santos et al. (2004) estimou a biomassa florestal, utilizando um modelo digital de elevação gerados de dados interferométricos provindos de dados SAR em bandas X e P. Servello et al. (2009) realizou classificações de uso e ocupação do solo usando dados SAR polarimétricos orbitais e trabalhando com os dados complexos vetoriais e matriciais. Freeman et al. (1999) mapearam os ciclos de inundação da bacia amazônica entre as épocas de alta e baixa inundação gerando uma sequência de aproximadamente 5000 quadros.

Para a realização dos trabalhos supramencionados, diversos trabalhos de campo foram realizados. Os dados destes trabalhos encontram-se em planilhas, fotos, vídeos, anotações de campo etc. sendo que muitos destes dados não são de fácil acesso a comunidade científica.

De acordo com o exposto pode-se notar que o grande volume de pesquisas, estudos e projetos realizados na região da Floresta Nacional do Tapajós, geraram grandes quantidades de dados, dos mais variados formatos, compondo um acervo heterogêneo e distribuído em diversos locais. Dada esta grande quantidade heterogênea de informação gerada desta região, que está dispersa em vários lugares, seria interessante ter um modo de prover acesso a toda essa informação e que permita realizar processos com os dados de outros pesquisadores.

8.2. Escolha do trabalho de pesquisa para o estudo de caso

O estudo de caso foi desenvolvido usando como base o trabalho de pesquisa em sensoriamento remoto de Pereira (2012) do INPE. A região de interesse deste trabalho de pesquisa é a Floresta Nacional do Tapajós e arredores. Este trabalho teve como objetivo principal avaliar diferentes formas de integrar dados (fusões e combinações) de diferentes sensores imageadores (SAR e ópticos). No trabalho visa-se identificar a forma de integração mais adequada para o mapeamento do uso e cobertura da terra de uma região de fronteira agrícola da Amazônia brasileira, utilizando como insumos para tal, algumas imagens TM/LANDSAT5 e ALOS/PALSAR e trabalhos em campo.

A dinâmica do trabalho de pesquisa, objeto do estudo de caso, se inicia com a aquisição das imagens ópticas e SAR da área de interesse, seguida de trabalhos em campo para coletar dados das diferentes áreas referentes às classes que se deseja classificar. São coletados, em campo, diversos tipos de informações relativas aos dados a serem classificados. Estas informações são utilizadas para o treinamento e a validação da classificação e que inclui fotos das áreas de interesse e anotações sobre as mesmas. Ao final do trabalho, espera-se ter as imagens com as áreas classificadas e uma matriz de confusão, contendo os resultados de eficiência da metodologia de classificação utilizada.

Este trabalho de pesquisa foi escolhido para ser o objeto deste estudo de caso por apresentar uma variedade heterogênea de artefatos como insumos e resultados. Como insumos são utilizados dados matriciais, provindos de diferentes sensores, dados vetoriais coletados em campo, dados de apoio como fotos da região e anotações de campo. Os resultados são dados matriciais, fruto da classificação, e uma matriz de confusão. Além destes artefatos, ainda há algoritmos de processamento de dados, que realizam deste a filtragem e preparação dos dados até a classificação das imagens em si.

8.3. Planejamento do estudo de caso

O estudo de caso conduzido neste trabalho possui dois grandes momentos. O primeiro momento é avaliar a capacidade de gerir os diferentes artefatos produzidos ou utilizados durante uma pesquisa. Estes artefatos, em geral, representam uma parcela do conhecimento difícil de ser comunicada ou difundida em meio formais e o que se pretende aqui é torná-las externalizáveis.

O segundo momento é avaliar a capacidade de recuperar o contexto de uma pesquisa. Recuperar desde os insumos utilizados, os documentos que descrevem a pesquisa em si ou a metodologia aplicada, os materiais utilizados como apoio e o resultado propriamente dito. Adicionalmente é demonstrada a capacidade de execução de algoritmos sobre os artefatos, neste caso em específico, a execução de um algoritmo de filtragem aplicado sobre uma imagem usada como insumo. Este algoritmo de filtragem é apenas um dos processamentos aplicados no trabalho original, mas todos estes elementos são essenciais para reproduzir com um mínimo de fidelidade o contexto de uma pesquisa.

O primeiro momento do estudo de caso descreve desde como seria o início de uma nova pesquisa, passando por descoberta e análises preliminares em regiões de interesse. Este momento é importante para a descoberta de dados segundo sua componente geográfica e possíveis levantamentos bibliográficos para uma dada região. Para isso será demonstrada a interação com o GeoPortal. Em seguida será demonstrada a capacidade de gerir os artefatos existentes em uma pesquisa, sejam eles espaciais ou não. Os artefatos em questão são os dados matriciais dos diferentes sensores, os dados vetoriais coletados em campo, os documentos de texto, planilhas, fotos etc. É demonstrado como vincular os dados multimídias aos dados espaciais. O vínculo dos dados multimídias aos dados espaciais permite que os dados multimídias sejam encontrados a partir dos dados espaciais e que os dados espaciais sejam encontrados a partir dos dados multimídias. Além disso, é possível espacializar artefatos que não possuem informação espacial, ou seja, relacionar um artefato multimídia qualquer a uma localização geográfica. Por

fim é explorada a capacidade de compartilhar algoritmos de processamento de dados como um serviço, através do exemplo da execução do algoritmo de filtragem aplicado a uma imagem utilizada no trabalho de pesquisa original.

No segundo momento do estudo de caso esperar comprovar a possibilidade de recuperar o contexto de uma pesquisa, ou seja, recuperar os dados importantes para reproduzir a pesquisa em questão. Para isso deve ser possível recuperar os dados de insumo da pesquisa, que são as imagens dos diferentes sensores e os dados coletados em campo. Deve ser possível ainda recuperar os dados multimídias referentes a metodologia e resultados.

8.4. Início de um trabalho de pesquisa

No início do trabalho de pesquisa é criado um espaço no CMS, que por sugestão terá alguns diretórios pré-definidos, similar ao mostrado na Figura 7.3. São preenchidos alguns dados como o autor, título e palavras-chaves para facilitar futuras buscas de conteúdos relacionados.

A estrutura pré-definida de diretórios auxilia o pesquisador a separar os conteúdos segundo sua classificação. Os diretórios ajudam também aos demais usuários quando realizarem uma busca, pois por analogia e convenção o objeto de interesse deverá estar em um diretório específico. Por exemplo, se o objeto de interesse for uma amostra de uma dada classe (uma região de interesse) por convenção ela se encontrará no diretório de Desenvolvimento\Insumos.

Apesar de existirem mecanismo em um ECMS para auxiliar na obrigatoriedade de preenchimento de artefatos ou diretórios, como por exemplo, *triggers* e *workflows* este mecanismos não foram explorados e nem adotados, o que faz da estrutura de diretórios mais uma convenção do que uma regra propriamente dita. Esta foi uma decisão de projeto, tomada em conjunto com os usuários chaves participantes da fase de especificação, uma vez que a natureza das pesquisas são diversas e a obrigatoriedade poderia ser uma barreira para a criatividade ou para os aspectos inovadores.

8.5. Descoberta de dados e análises de regiões

Os resultados de estudos conduzidos em uma determinada área ou região podem sofrer influência dos meios externos e das características físicas do local, tais como umidade, relevo, composição do solo, entre outras. Muitos dados sobre estas características físicas são disponibilizadas por diversos órgãos federais ou instituições responsáveis, das quais se podem citar o Instituto Brasileiro de Geografia e Estatística (IBGE), Agência Nacional de Águas (ANA), Ministério do Meio Ambiente (MMA) e Companhia de Pesquisa de Recursos Minerais (CPRM).

Estes dados quase sempre são abundantes, mas nem sempre fáceis de serem encontrados ou utilizados. Em geral estão espalhados nas infraestruturas de cada uma das organizações, mas este cenário começa a mudar com a institucionalização da INDE e muitas organizações começam a disponibilizar seus dados de forma organizada e metódica sob a forma de catálogos, adotando principalmente os padrões da OGC para tal.

Um dos componentes desta arquitetura proposta como solução é o `GeoPortal`, que consegue realizar buscas nos catálogos de dados destas organizações e permite criar e compor mapas com diversos dados de diferentes origens. Na Figura 8.1 pode-se verificar o componente de busca (1) realizando uma busca e retornando alguns dados disponíveis no catálogo de dados do IBGE. Os dados retornados são serviços de mapa no padrão WMS com os seguintes temas: Solos, Amazônia, Pedologia, Geomorfologia, Biomas, Geologia, Censo e Uso da Terra.

À medida que a INDE amadurecer e mais nós forem acrescentados ou mais organizações disponibilizarem seus dados de modo que possam ser compartilhados e utilizados, mais completas podem ser as análises, levando em consideração ou cruzando dados de diversas naturezas. Pode-se citar a título de exemplo o cruzamento de dados censitários, arruamento, geológicos, hidrológicos, financeiros, criminalísticos etc. A INDE possui várias instituições

que são nós potenciais no Diretório Brasileiro de Dados Geoespaciais, conforme pode ser visto na Seção 5.5.

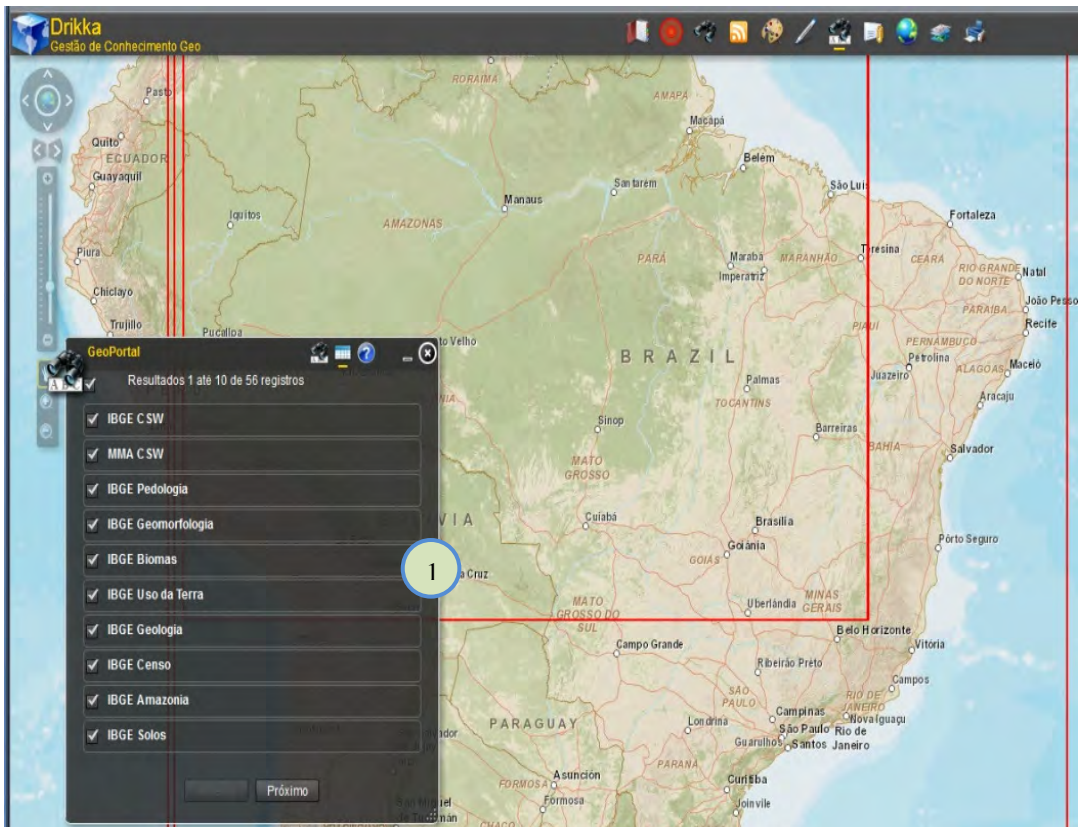

Figura 8.1: Mapas disponíveis no IBGE.

Os resultados destas buscas em catálogos de dados podem trazer dados de variados formatos. Caso os dados retornados estejam nos formatos WFS ou WMS eles podem ser incluídos na visualização corrente do mapa, conforme visto nas Figura 8.2 e Figura 8.3. Na Figura 8.2 é visto o resultado de uma busca simples por Solos no catálogo de dados do IBGE retornando um serviço WMS (1) do tema de Solos, que foi adicionado à visualização corrente do mapa e sua respectiva legenda pode ser vista em (2). Enquanto que na Figura 8.3 é visto um serviço WMS com o tema de referencial hidrográfico do Brasil, resultado de uma busca por hidrografia (1) realizada no catálogo do IBGE e adicionado a visualização corrente do mapa. Pode ainda ser vista a legenda do serviço de hidrografia em (2).

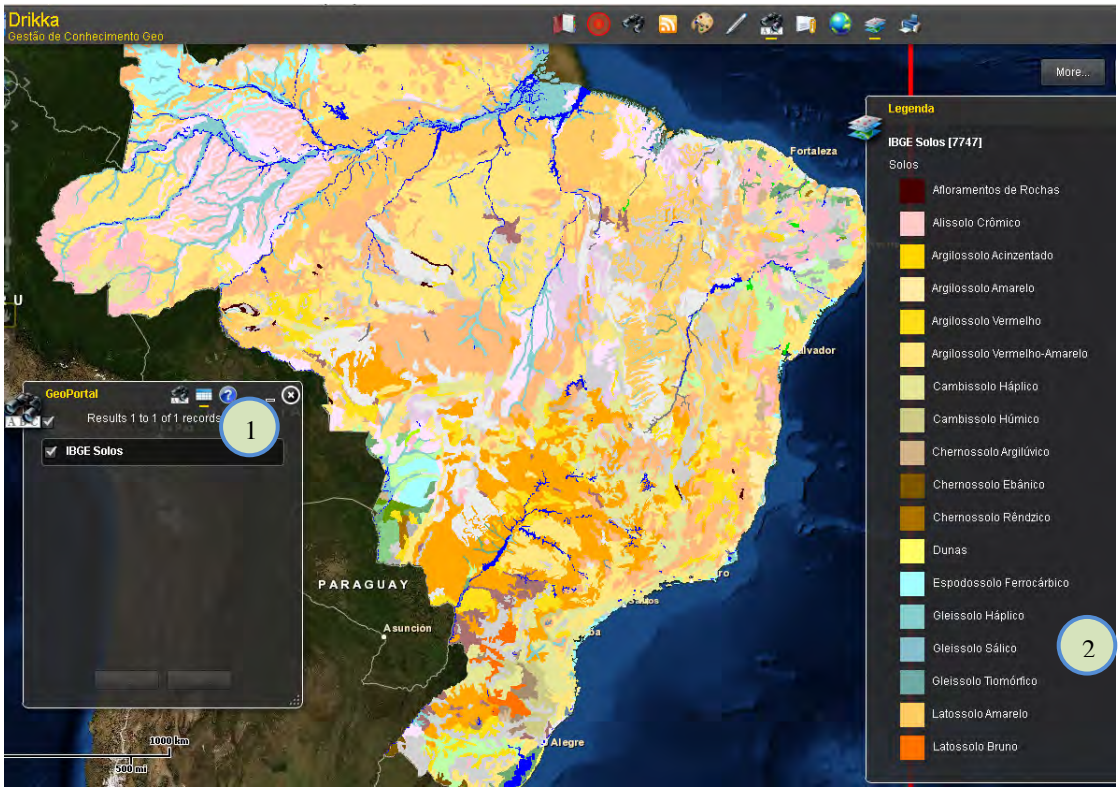

Figura 8.2: Inclusão dinâmica de serviços de solos.

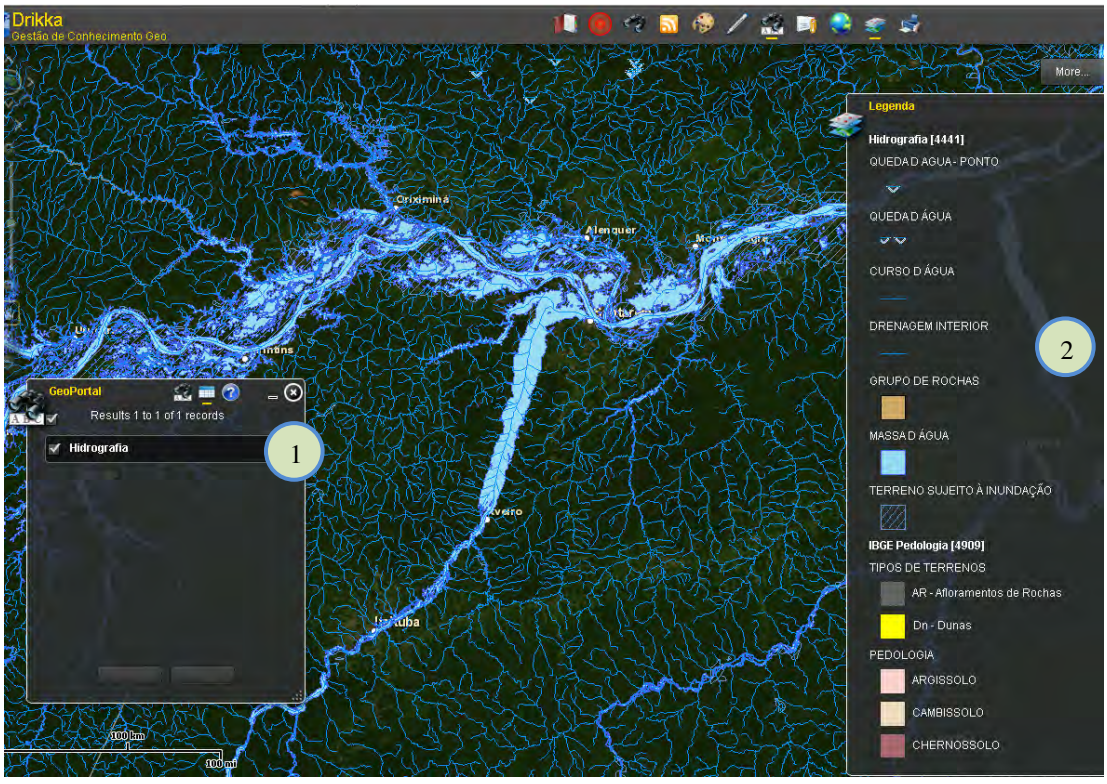

Figura 8.3: Inclusão dinâmica de serviços hidrografia.

Estas buscas são realizadas pelo módulo de buscas no `GeoPortal`, melhor detalhado na Seção 7.9.1 e ilustrado na Figura 7.9.

É possível e também interessante, adicionar uma representação WMS de dados matriciais na visualização corrente do mapa e avaliá-la contra o mapa base (*basemap*) para se ter uma noção melhor da sua região de referência. Como mencionado, quando os resultados das buscas no `GeoPortal` retornar um serviço WMS ou WFS é possível adicioná-lo dinamicamente à visualização atual do mapa. A Figura 8.4 ilustra um dado matricial sendo adicionado à visualização atual do mapa e contrastando com o *basemap* de uma imagem óptica da região.

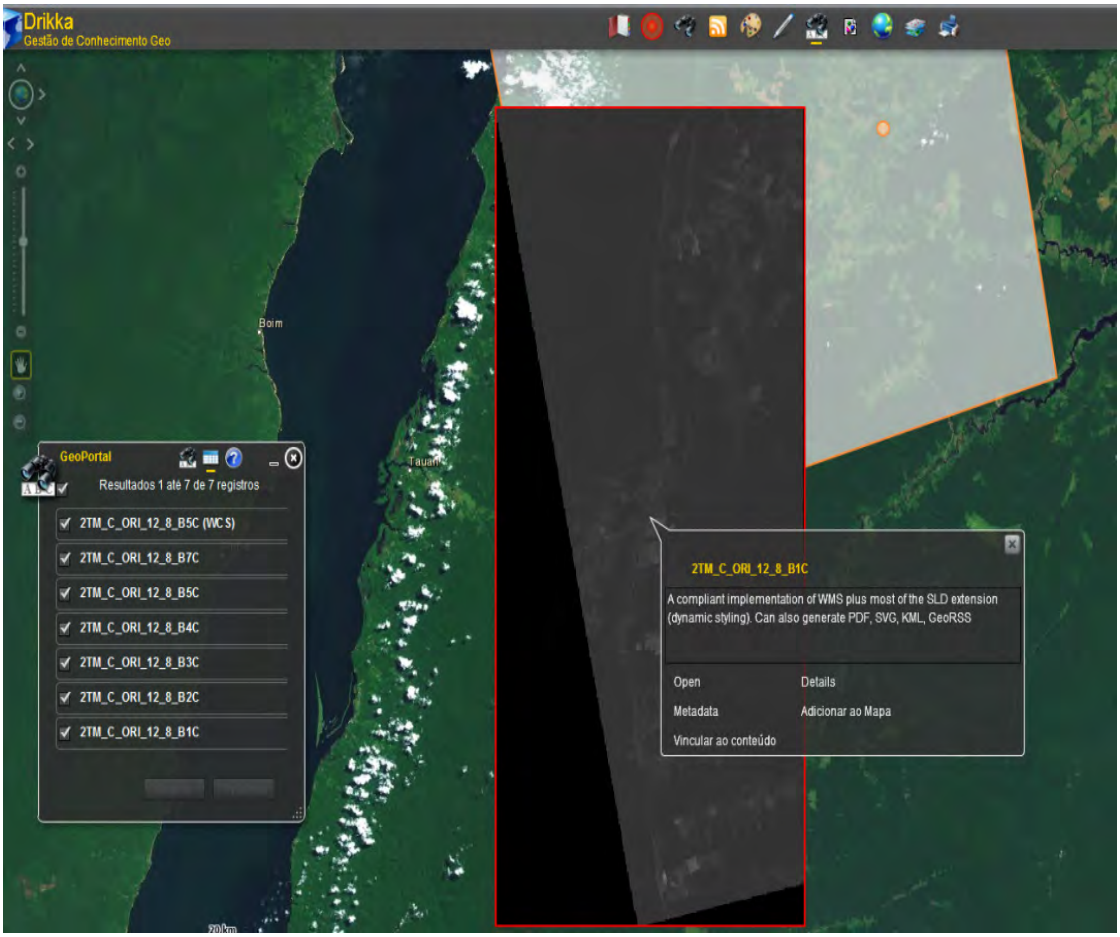

Figura 8.4: Adição de dado matricial a visualização atual do mapa.

Uma importante etapa durante o desenvolvimento de um trabalho de pesquisa é a revisão bibliográfica, o que inclui levantar artefatos que já foram produzidos antes em outras pesquisas, entender suas contribuições e escopo. O levantamento de um acervo de dados pode ser importante mesmo que não relacionado diretamente a uma revisão bibliográfica.

A capacidade de realizar buscas por estes dados segundo sua referência espacial pode ser pertinente, imagina-se, por exemplo, querer saber quais dados matriciais estão disponíveis para uma dada região e isso é possível através do `GeoPortal`. Na Figura 8.5 é possível verificar uma busca de todos os dados (definidos pelo caractere *) completamente contidos (*fully within*) (1) na visualização atual do mapa. Pode-se ainda escolher o operador topológico de intersecção (*Intersecting*) com a extensão espacial atual para realizar a busca ou buscar em qualquer extensão espacial (*Anywhere*).

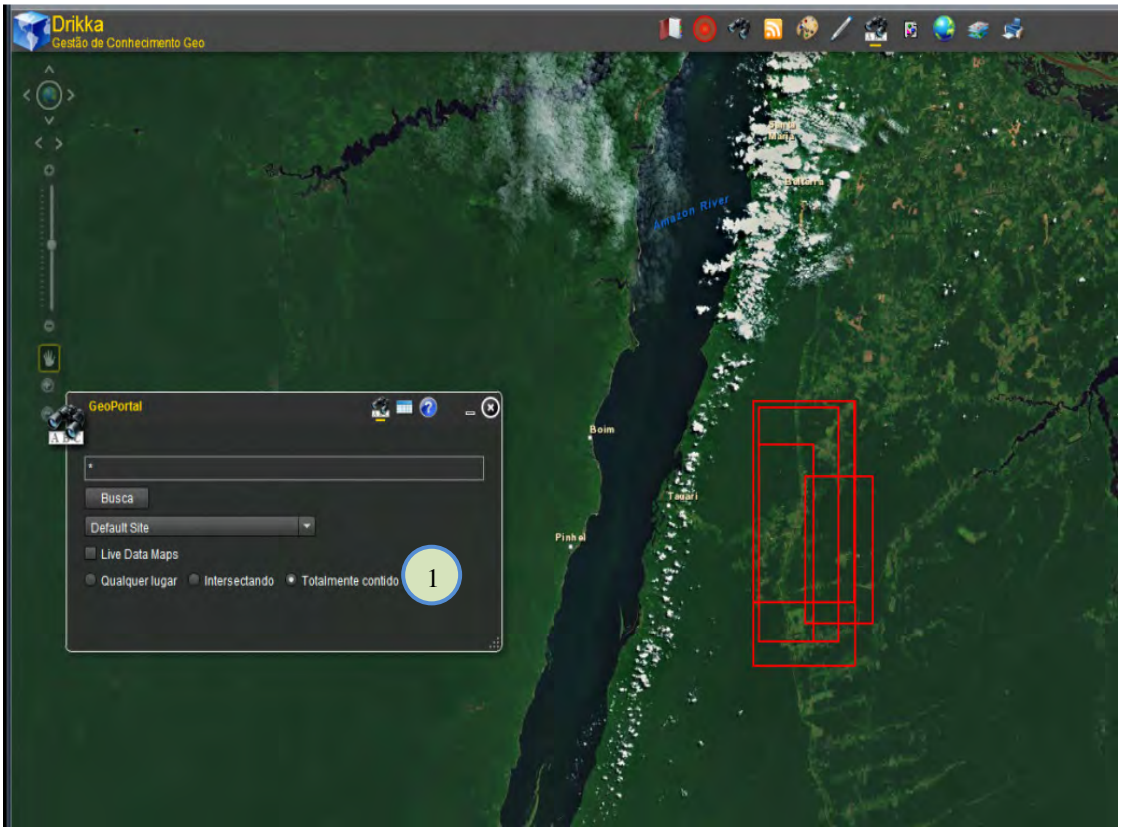

Figura 8.5: Busca de dados por extensão espacial.

8.6. Serviço *Web* de Processamento

Dados matriciais para serem utilizados, quase sempre, precisam passar por algum tipo de processamento, seja ele filtro, classificação, fusão ou outro qualquer ou mesmo uma combinação de vários. No INPE, mais especificamente na Divisão de Processamento de Dados (DPI) e Sensoriamento Remoto (DSR), tem-se uma experiência e um legado muito grande quanto ao tratamento de dados matriciais e é um desejo disponibilizar alguns algoritmos como um serviço. Estes algoritmos de processamento de imagens representam uma parcela do conhecimento, desta forma, a arquitetura proposta como solução contempla esta situação. A disponibilização de um serviço foi realizada através do uso do padrão WPS da OGC, disponibilizando um serviço *web* de processamento registrado e hospedado no `GeoServer`.

O padrão WPS é recente e ainda em fase de maturidade, estando em sua versão 1.0.0. Ele opera em ambiente *web*, o que por si já impõe certas limitações, por exemplo, quanto ao fluxo de tráfego de dados e desempenho, principalmente se comparado aos programas de mesa. Desconsiderando momentaneamente estas limitações, o padrão WPS é muito versátil, permitindo vários tipos de entradas e saídas, inclusive com possibilidade de cascatear serviços *web.* Isto é, pode-se desenvolver um serviço *web* que recebe como entrada diretamente o resultado de outro serviço *web*.

A arquitetura proposta como solução conta com um serviço *web* síncrono de filtro para processamento de dados matriciais, onde a entrada pode ser uma URL de um dado matricial. Na Figura 8.6 pode-se observar o módulo responsável por executar este serviço *web*, processar a imagem e exibir seu resultado. Ainda na Figura 8.6, pode-se verificar a URL da imagem em (3), que pode ser uma URL direta ou indireta, em (4) o tamanho da janela de processamento em *pixels*, em (1) uma representação da imagem original e em (2) a imagem após a aplicação do filtro. O serviço de filtro é realizado através do filtro de média.

Considerando uma dada aplicação, a imagem pode ser obtida através de um serviço *web* do padrão WCS e passar por uma sequência de processamentos. Este serviço apresentado na Figura 8.6 demonstra apenas uma parcela deste trabalho, automatizando o filtro a partir de uma aquisição de imagem interna ou externa a solução. Vale ressaltar que o foco principal deste trabalho é fornecer a infraestrutura necessária para que estes processamentos sejam oferecidos como serviços. O serviço de filtro desenvolvido é uma prova de conceito da possibilidade e viabilidade de disponibilizar algoritmos de tratamento de imagens como serviços *web*, considerando e respeitando as limitações do ambiente *web* atual. A partir desta infraestrutura e do modelo utilizado, mais serviços podem ser desenvolvidos e disponibilizados.

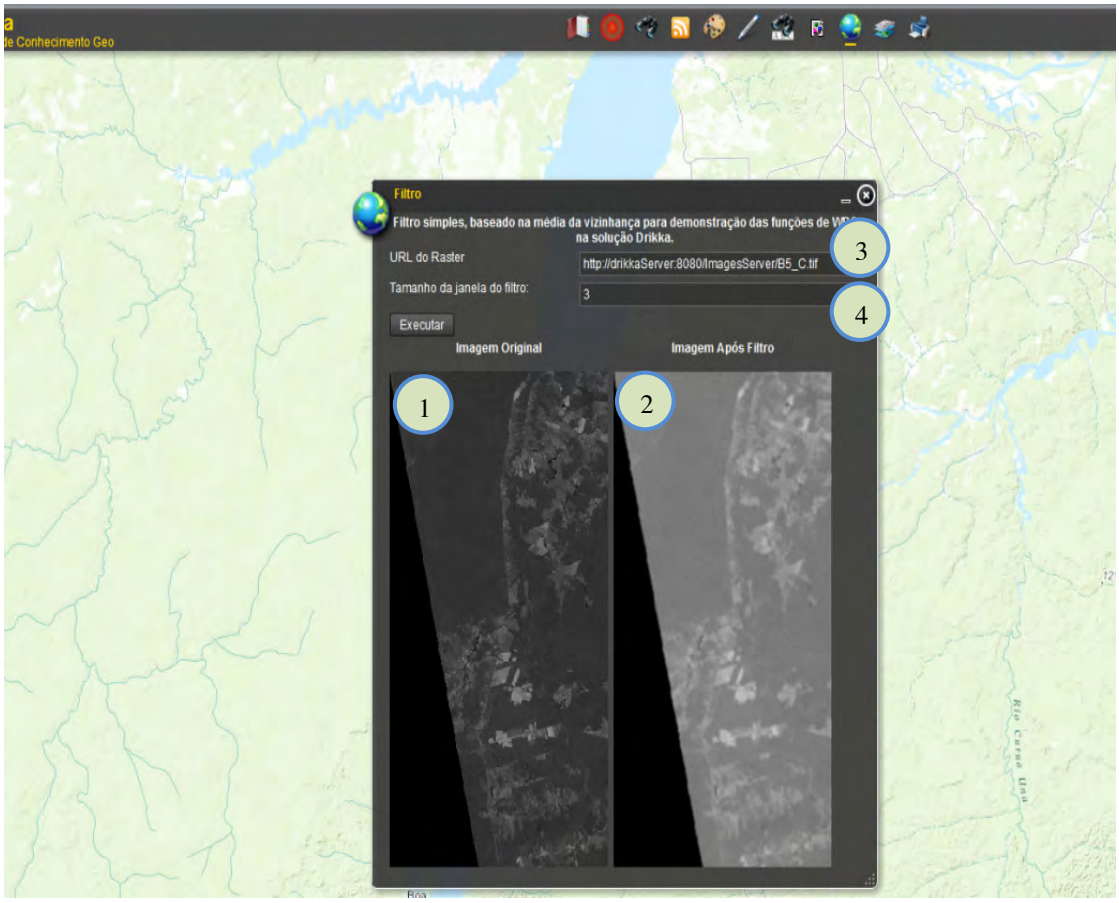

Figura 8.6: Filtro de uma imagem da região de interesse.

8.7. Gestão dos arquivos multimídia e dados espaciais

Durante o desenvolvimento de uma pesquisa são produzidos diversos tipos de arquivos em variados formatos. Uma das grandes preocupações durante a produção deste trabalho foi o suporte e gestão destes diferentes arquivos ao longo do ciclo de vida de uma pesquisa. Dependendo da natureza da pesquisa, estes arquivos são gerados em diferentes ambientes e diferentes momentos. No cenário deste estudo de caso, há vários trabalhos realizados em campo, que significa visitar a região de interesse *in loco* e conduzir uma série de atividades. Estas atividades incluem medições, anotações, fotos, entre outras.

A adoção de uma ferramenta de CMS atenua muito este problema, uma vez que oferece suporte a muitos tipos de arquivos. Grande parte do que é produzido em campo é passível de serem geridos pelo CMS, como documentos de texto, planilhas de medições, fotos e apresentações.

Outro tipo de dados gerados geralmente é de origem espacial, por exemplo, as delimitações da região de interesse, para treinamento ou validação de um dado classificador. Estes dados podem ser armazenados em formato *ShapeFile*, que até poderia ser armazenado no CMS, mas que faz pouco sentido pois desta forma sua natureza espacial pode ser pouco explorada.

Os dados em *ShapeFiles* podem então ser diretamente disponibilizados pelo `GeoServer` ou serem primeiro importados no `PostGIS` e a partir daí serem disponibilizados pelo `GeoServer`. Após estas etapas eles serão indexados pelo `GeoPortal` e aparecerão como resultado nas buscas realizadas através do mesmo.

Neste estudo de caso, foi criado um arquivo de texto no CMS descrevendo os dados vetoriais e matriciais utilizados como insumos para o desenvolvimento da pesquisa. Este arquivo descreve a origem dos dados, seu conteúdo e a metodologia de sua aquisição e tratamento. Os dados espaciais, em si, foram publicados no `GeoServer` e disponibilizados a partir de lá. Na Figura 8.7 é mostrado como os dados matriciais foram vinculados ao arquivo no CMS. No

diretório “Insumos” (2) da estrutura de diretório do CMS foi criado o arquivo “DadosMatriciais.doc” (1) que descreve os dados matriciais utilizados e em (3) é apresentada a lista de dados matriciais a serem vinculados ao conteúdo do CMS, que ao serem selecionados são detalhados e poderão ser vinculados ao conteúdo do CMS através do acionamento do botão “Vincular ao Conteúdo” (4).

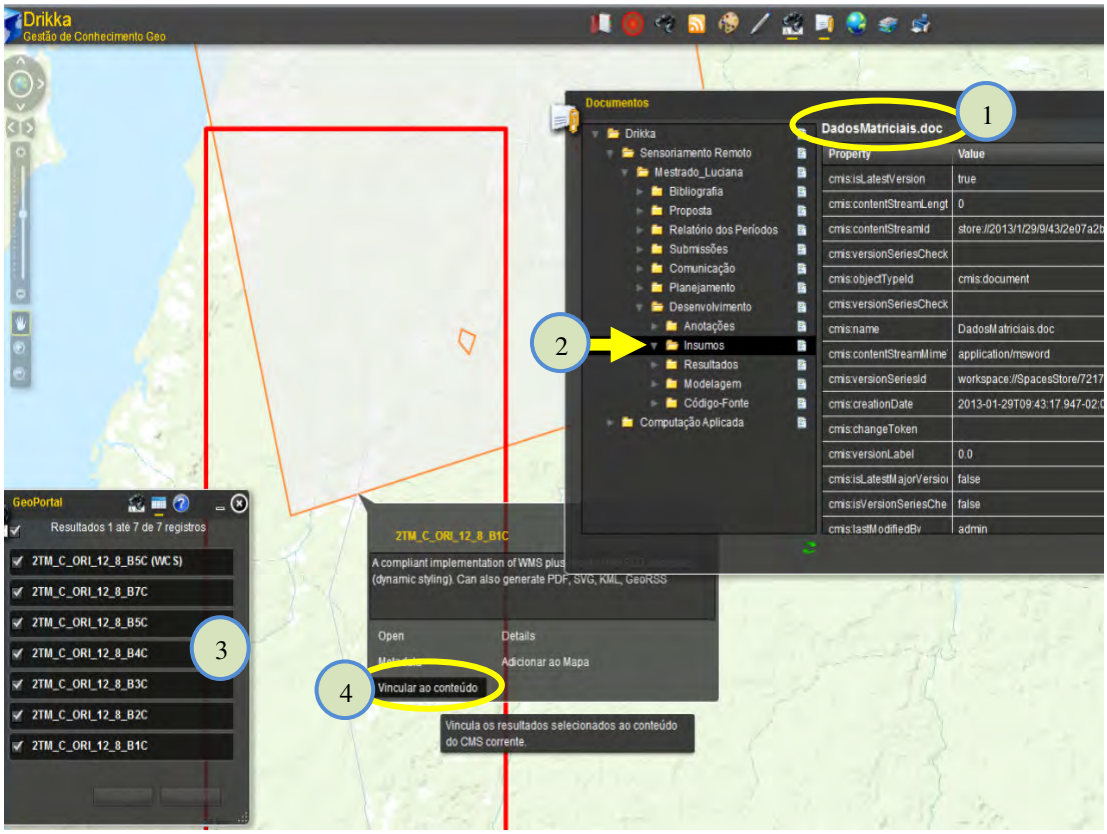

Figura 8.7: Vinculando dados matriciais.

Algo semelhante ocorre com os dados vetoriais, conforme ilustrado na Figura 8.8, onde os dados coletados em campo são transformados em dados vetoriais e são apresentados na lista de resultados em (3), delimitando pequenas áreas para treinamento e validação do classificador para cada uma das 13 classes de interesse do trabalho de pesquisa e vinculados ao documento “DadosVetoriais.doc” (1) no CMS através do botão “Vincular ao Conteúdo” em (4).

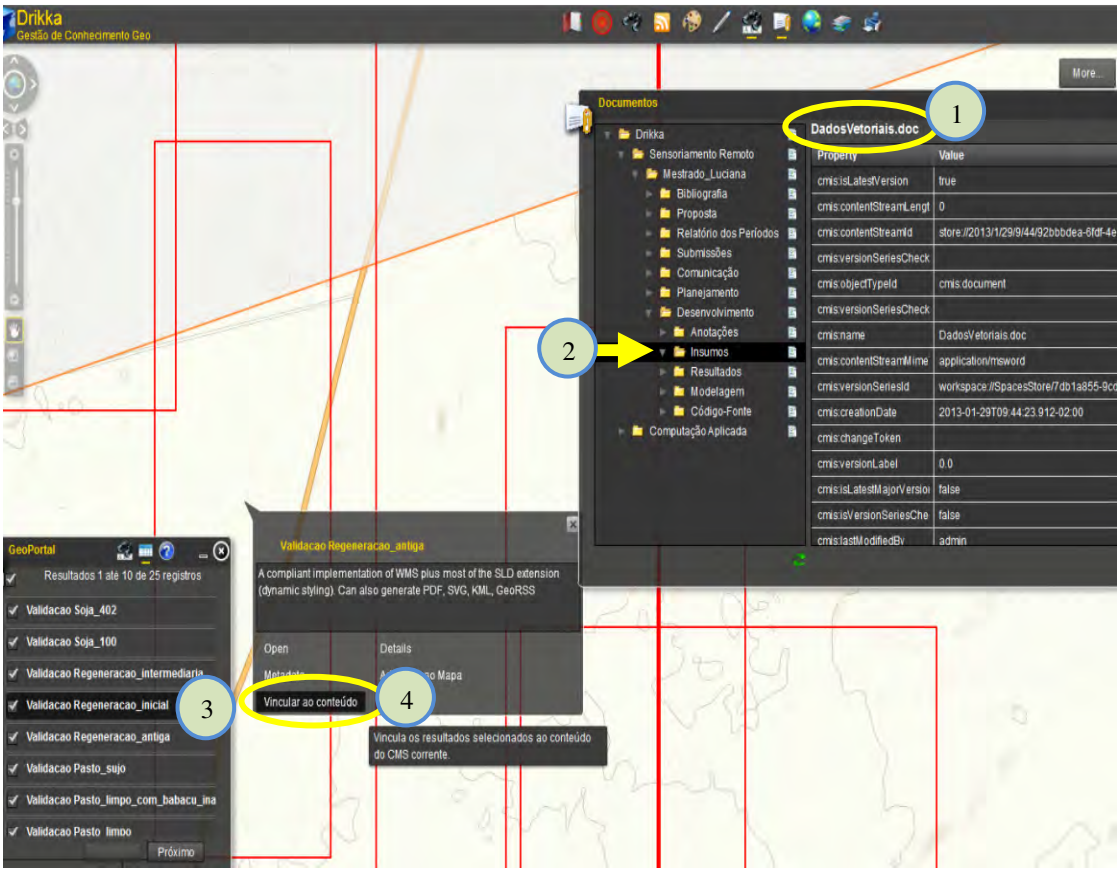

Figura 8.8: Vinculando dados vetoriais.

Com o auxílio de uma ferramenta como esta, é possível que os profissionais em campo consigam registrar os dados coletados em campo no mesmo ambiente em que outros profissionais trabalharão nas análises, diminuindo o tempo de espera entre estas duas etapas. Todos estes artefatos, em geral, eram perdidos ou não são de fácil acesso a outros profissionais, por que na maioria das vezes, o que se tem como fruto de uma pesquisa são os artefatos formais como uma tese, dissertação ou artigos.

O acesso aos dados reais utilizados como insumos ou aos resultados facilita a reprodução e evolução dos trabalhos de pesquisas. Os resultados também são geridos e disponibilizados pela infraestrutura implementada. Neste estudo de caso, os resultados são dados matriciais, produto da classificação, contendo as classes de cada tema de interesse e a matriz de confusão contendo os índices de eficiência do classificador.

A maneira como os resultados são relacionados com a pesquisa é semelhante ao modo como foi realizado com os insumos. A Figura 8.9 ilustra como os resultados foram vinculados, sendo criado o documento “Resultados.doc” (1), descrevendo a metodologia adotada para sua obtenção, disponível no diretório “Resultados” (2) e vinculados os dados matriciais correspondentes, retornados de uma pesquisa no `GeoPortal` (3). Nota-se ainda a existência do conteúdo referente à matriz de confusão no formato de uma planilha eletrônica sob o nome de KAPPA.xlsx em (4).

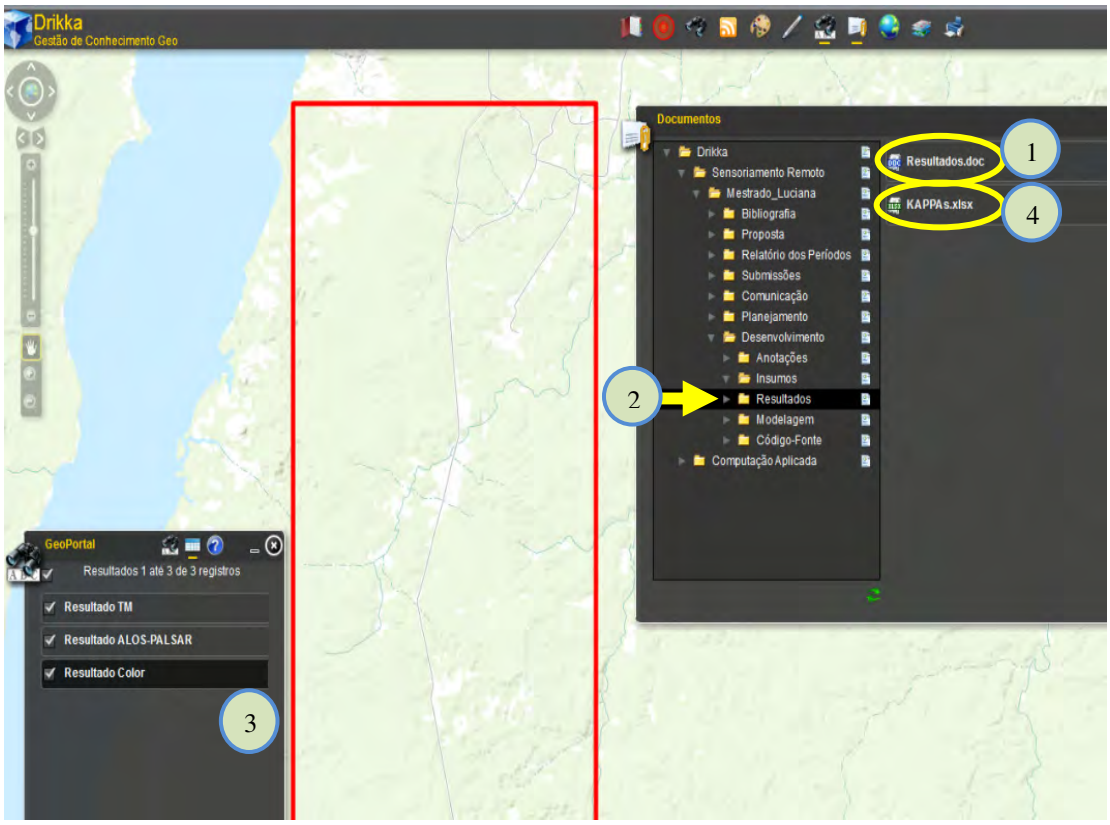

Figura 8.9: Vinculando os resultados.

8.8. Espacialização de conteúdo

Determinados tipos de conteúdos não possuem, necessariamente, natureza espacial, caso de uma dissertação, por exemplo. Uma dissertação não necessariamente possui uma referência espacial, mas o fato de estar relacionada a uma determinada área de interesse torna convidativa a possibilidade de relacioná-la a este lugar no espaço.

Para estes cenários a solução desenvolvida contempla a capacidade de relacionar qualquer conteúdo no CMS a qualquer feição de ponto ou área de interesse. Na Figura 8.10 é ilustrado como a área representada por um polígono (1) é relacionada ao diretório “Mestrado_Luciana” no CMS (2) através da função “Vincular ao Conteúdo” (3). Em linhas gerais, esta referência espacial pode ser interpretado como: “o trabalho de mestrado da Luciana está relacionado a esta área”.

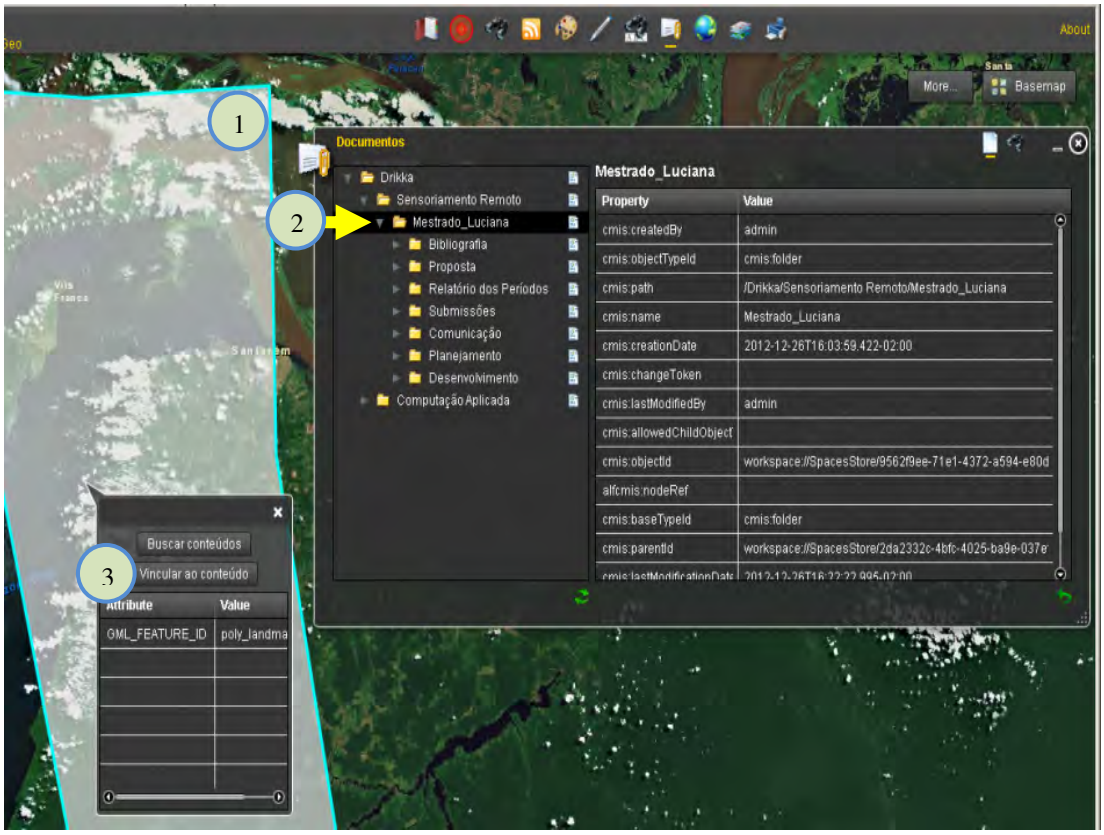

Figura 8.10: Vínculo entre área de interesse e conteúdo do CMS.

No caso apresentado na Figura 8.10 todo o diretório de pesquisa esta sendo relacionado a uma determinada área, mas pode haver casos onde um único conteúdo pode se relacionar com um determinado ponto. Na Figura 8.11 um conteúdo de foto de nome “RIMG0004.JPG” (1) está sendo relacionado com o ponto de interesse (2). Ainda na Figura 8.11 pode-se observar em (3) a existência de um editor de feições, que permite que sejam adicionadas novas

áreas ou pontos de interesse. Observa-se também em (4) a existência de um botão chamado “Buscar Conteúdos” que realiza uma busca por todos os conteúdos relacionados a uma área ou ponto de interesse. É possível verificar na Figura 8.11 em (5) um espaço destinado a apresentar um *preview* do conteúdo corrente, sendo esta funcionalidade disponível para alguns formatos de dados, neste caso, conteúdos de fotos.

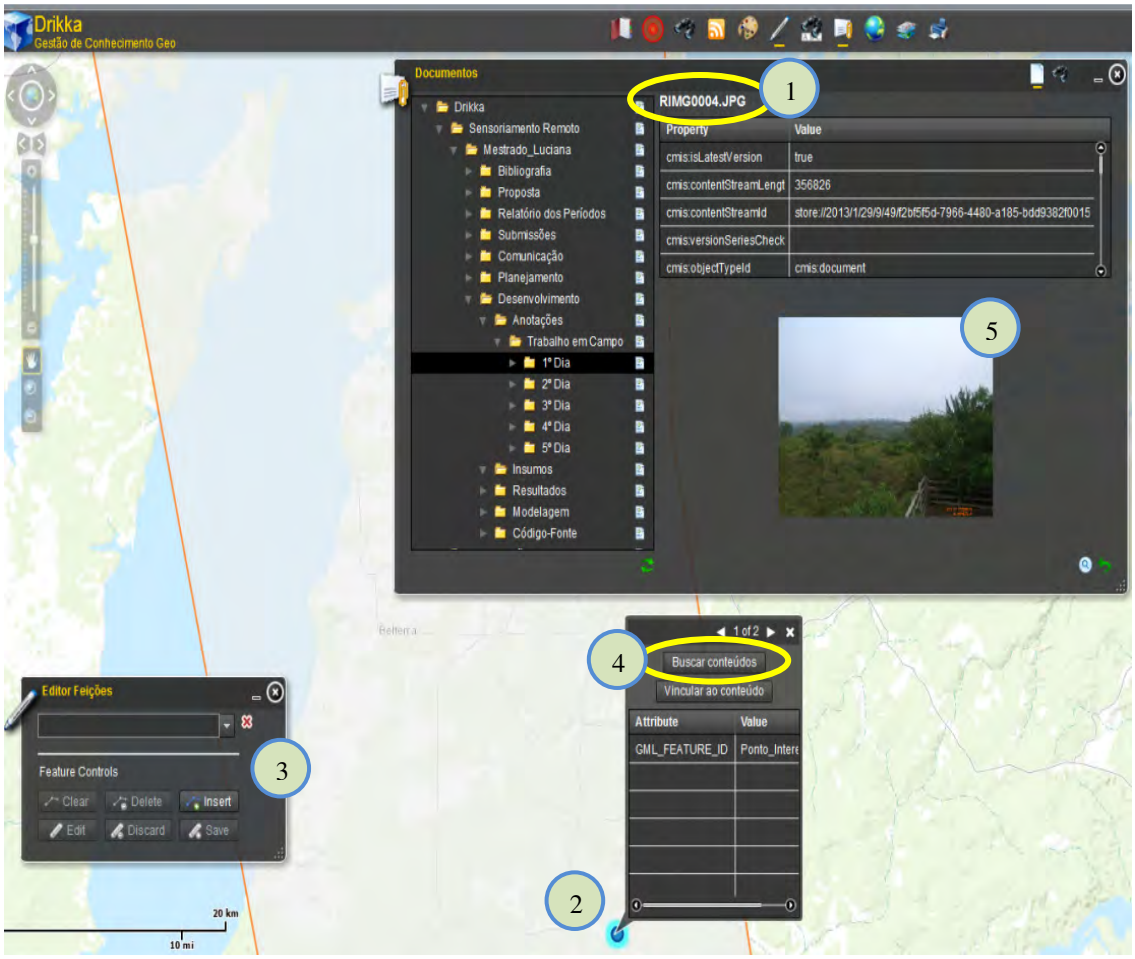

Figura 8.11: Relacionando um conteúdo com um ponto de interesse.

8.9. Recuperando o contexto de uma pesquisa

Como já mencionado anteriormente, o produto final de uma pesquisa de pós-graduação, quase sempre se resume a uma dissertação ou tese. Produtos disponíveis numa biblioteca ou por outros meios digitais, ou seja, apenas um artefato formal. Considerando que um novo pesquisador se interessou por um

dado trabalho de pesquisa e gostaria de reproduzir seus feitos para avaliar se sua metodologia proposta é mais eficiente para o mesmo cenário poderá encontrar certas dificuldades.

Considere que este novo pesquisador se interessou pelo trabalho apresentado neste estudo de caso, mas que a arquitetura proposta como solução ainda não estivesse disponível, como ele obteria as mesmas imagens de RADAR ou ópticas utilizadas? Como seriam obtidos os mesmos dados das amostras de treinamento e de validação do classificador ou como seriam obtidos os resultados? Talvez o único destes artefatos passível de publicação pelos meios formais seja a matriz de confusão, mas seria esta suficiente para que este novo pesquisador conseguisse reproduzir os mesmo resultados a partir dos mesmos insumos? Teria este novo pesquisador a necessidade de desenvolver ou implementar todos os algoritmos envolvidos? Por fim, como descobrir a existência de tais dados para uma área de interesse qualquer?

A partir da navegação pelo mapa da aplicação desenvolvida é possível identificar áreas que possua trabalhos de pesquisa relacionados. A partir de uma busca no catálogo do `GeoPortal` é possível recuperar algum conteúdo e a partir deste descobrir a qual trabalho de pesquisa o mesmo está relacionado. Considerando a primeira opção, a Figura 8.12 ilustra como obter os detalhes de um trabalho de pesquisa relacionado a uma determinada área, através da opção “Buscar conteúdos” (1). Esta opção irá realizar uma busca no CMS e obter os conteúdos relacionados a mesma, conforme pode ser observado ainda na Figura 8.12 em (2).

Dado o interesse deste novo pesquisador por este trabalho de pesquisa, o mesmo pode navegar pela estrutura de diretórios que compõe o trabalho, descobrindo quais são os insumos utilizados, qual a metodologia aplicada, quais algoritmos utilizados e quais resultados obtidos. Adicionalmente, ainda pode obter anotações de campo, fotos da área e algumas observações que possam vir a ser importante em seu novo trabalho.

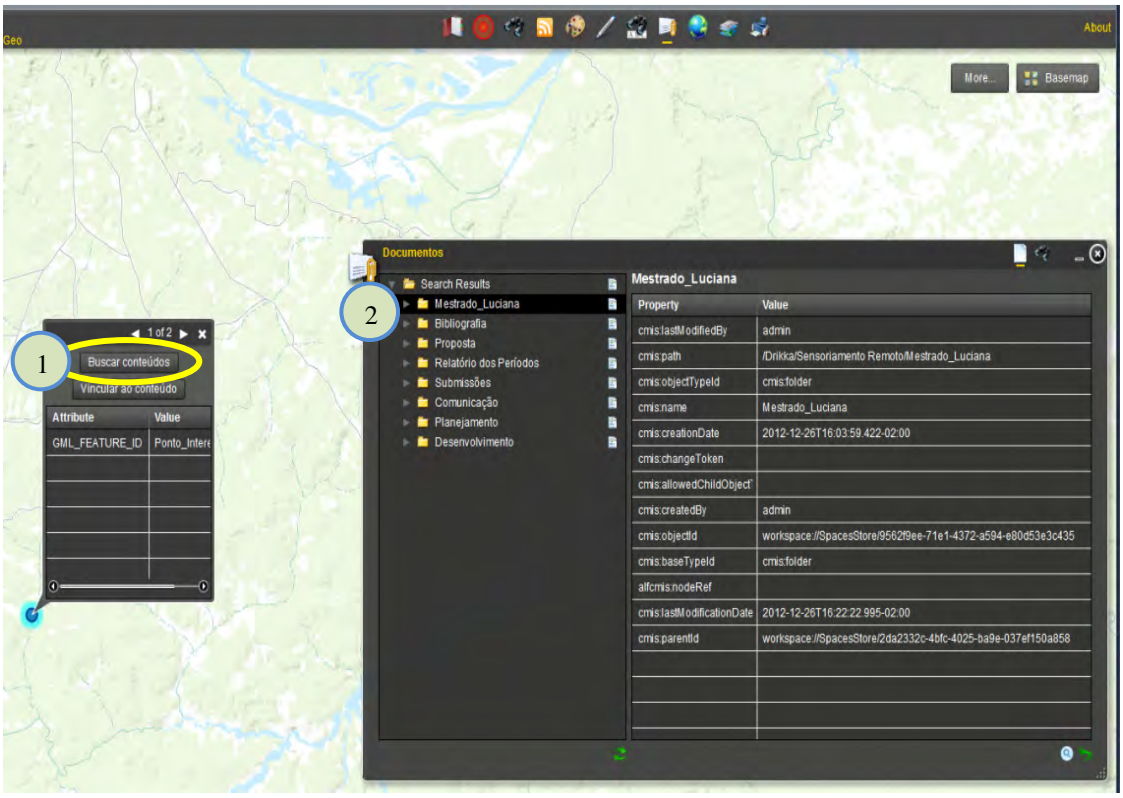

Figura 8.12: Busca por conteúdo relacionado.

O pesquisador pode verificar a existência dos documentos “DadosVetoriais.doc” e “DadosMatriciais.doc” navegando até o diretório de “Insumos” e a partir destes documentos é possível recuperar os dados espaciais utilizados. A Figura 8.13 ilustra a obtenção dos dados matriciais através da opção “Recuperar itens relacionados” (1) e estes itens relacionados são exibidos na lista de resultados em (2). O mesmo pode ser feito com os dados vetoriais que representam os dados utilizados para treinamento e validação do classificador.

Semelhante procedimento ocorre com os resultados, que podem ser obtidos no diretório “Resultados” (2) no documento “Resultados.doc” (1) que descreve a metodologia e está vinculado aos dados matriciais produto da classificação (3), conforme pode ser visto na Figura 8.14. Adicionalmente ainda está disponível a matriz de confusão, neste mesmo diretório. Na Figura 8.14, pode ser vista também a representação da composição colorida do resultado da classificação em (4).

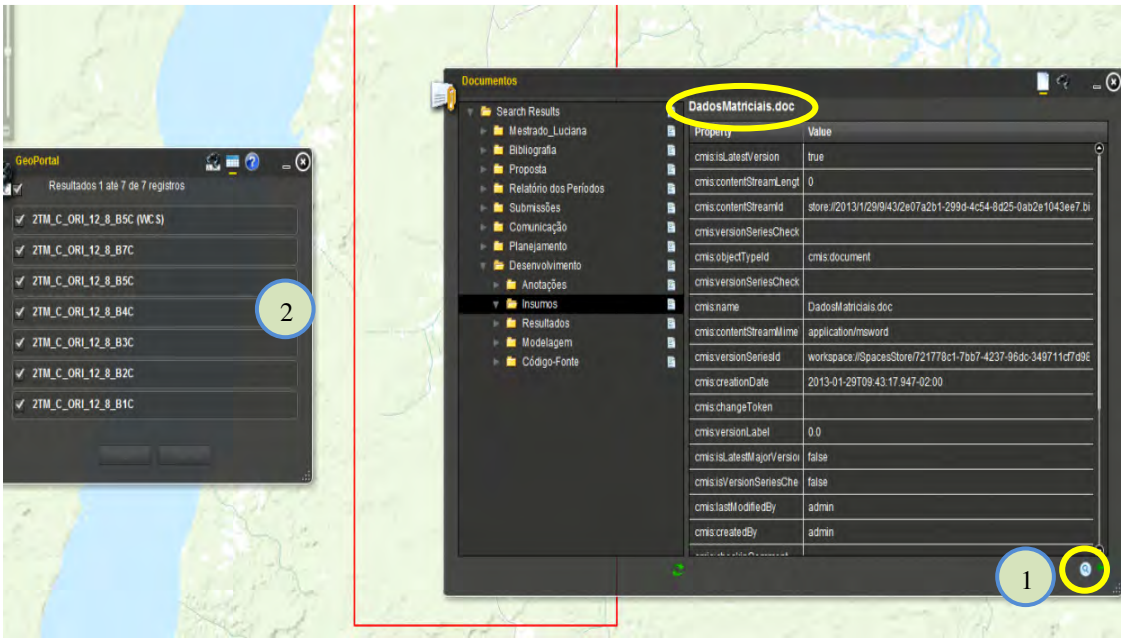

Figura 8.13: Recuperar itens relacionados.

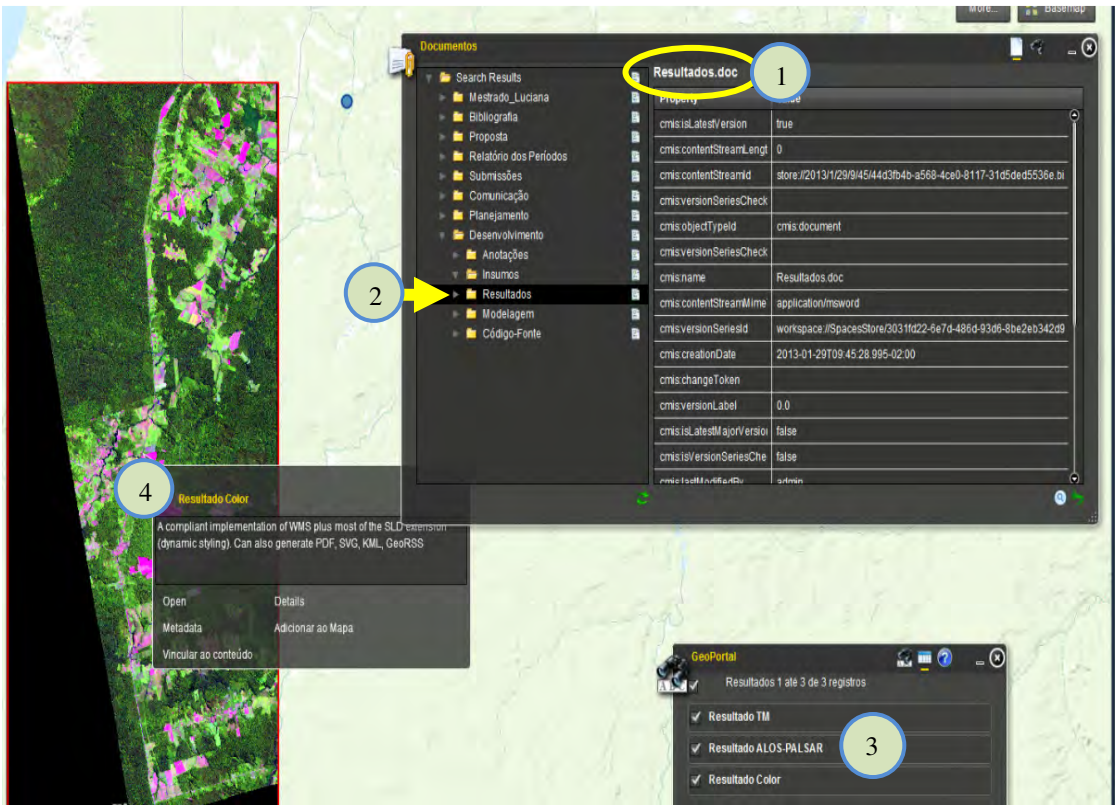

Figura 8.14: Recuperar resultados.

Desta forma, através da arquitetura desenvolvida como solução, é possível o acesso não apenas ao artefato formal, mas também aos insumos e resultados de seu desenvolvimento. Permitindo a reprodução integral da metodologia de trabalho e que por ventura seus resultados possam ser comparados aos de novas pesquisas e metodologias.

8.10. Considerações finais do capítulo

Neste capítulo foi apresentado o estudo de caso conduzido neste trabalho. O estudo de caso foi dividido em dois momentos, sendo o primeiro momento a interação de um pesquisador durante o desenvolvimento de sua pesquisa. Neste momento a aplicação permite que ele explore a área de interesse de sua pesquisa, procurando por dados em diversas instituições e dados sobre outras pesquisas relacionadas à mesma região. Esperava-se que a arquitetura proposta comportasse e gerisse uma variedade de artefatos, espaciais ou não. Isso foi realizado com sucesso, os dados multimídias foram geridos pelo componente de CMS da aplicação e os dados espaciais foram geridos pelo conjunto GeoServer e GeoPortal. O componente cliente foi capaz de possibilitar a visualização, descoberta e relacionamento entre os artefatos multimídias e espaciais através da orquestração dos serviços especialistas oferecidos por cada componente da arquitetura.

O segundo momento consiste no uso da aplicação por alguma outra pessoa ou pesquisador na intenção de recuperar o contexto de uma pesquisa já realizada. Esperava-se que a aplicação permitisse que as pesquisas fossem encontradas e os seus detalhes fossem recuperados, entenda por detalhes os insumos, resultados e artefatos importantes para sua compreensão e reprodução metodológica. Isso foi realizado com sucesso, foi possível recuperar todo o contexto de uma pesquisa navegando pelo mapa e identificando pontos ou áreas com pesquisas relacionadas ou, outra abordagem foi, navegar pela estrutura de diretórios do CMS e a partir deste recuperar os elementos espaciais associados.

Foi demonstrado também que algoritmos de processamento podem ser publicados e utilizados, como o exemplo do algoritmo de filtragem de imagens desenvolvido neste trabalho. Os artefatos que foram incluídos e manipulados neste estudo de caso, em geral, compreendem uma parcela do conhecimento que antes não era publicada ou não era de fácil acesso. Desta forma é possível disponibilizar todos os artefatos que foram utilizados como insumo em um trabalho de pesquisa, toda sua teoria e metodologia, seus algoritmos e processamentos e, por fim, os resultados. Isso torna mais fácil reproduzir na integra um resultado de uma pesquisa.

9 CONCLUSÕES

Neste capitulo final da dissertação, são apresentadas as principais contribuições deste trabalho, considerações finais e sugestões para trabalhos futuros.

9.1. Considerações finais

Mesmo com o advento da internet e grandes facilidades para comunicação e compartilhamento de dados, os principais modos de comunicação científica continuam sendo as publicações em meios formais. Os meios formais formatam o conhecimento e limitam o compartilhamento do conhecimento para apenas uma parcela de todo o conhecimento produzido durante um trabalho de pesquisa. Adicionalmente, existe o fato de que muitos trabalhos possuem uma relevância espacial, ou seja, estes trabalhos estão relacionados a alguma localização no globo terrestre. Estes foram os principais motivadores para se propor e desenvolver uma arquitetura capaz de auxiliar a gestão dos artefatos produzidos durante uma pesquisa, que em geral representam uma parcela do conhecimento e que não era possível de serem externalizados, e organizá-los de forma a conseguir recuperá-los segundo sua componente espacial.

Para avaliar a arquitetura proposta para cenários como o supracitado, foi implementada uma solução e conduzido um estudo de caso. No estudo de caso pôde ser comprovado que a arquitetura proposta é capaz de manipular e gerir artefatos multimídias tais como documento, fotos, vídeos e qualquer outro tipo de arquivo. Isso é possível através do componente de CMS adotado para compor a solução. Pôde-se, ainda, manipular e gerir os dados espaciais através dos componentes de gerenciamento de conteúdo geográfico e o geoportal. Estes dois elementos são os elementos fundamentais para a construção de infraestruturas de dados espaciais e possibilitaram referenciar e incorporar dados de instituições externas como o IBGE, ou resumidamente permitem que sejam fornecidos ou consumidos dados de outras infraestruturas de dados espaciais. Outra parcela do conhecimento é representada pelos algoritmos de processamento que, em geral também são difíceis de serem

compartilhados e publicados. Foi desenvolvido um serviço de processamento de imagem demonstrando as possibilidades de distribuir e compartilhar uma parcela do conhecimento sob a forma de algoritmos implementados, constituindo serviços de processamento.

Desta forma, revisitando a Seção 1.2, pode-se concluir que os objetivos propostos foram alcançados. Quanto a algum trabalho de pesquisa, é possível além de descobri-lo através de sua componente geográfica, recuperar os artefatos utilizados como insumo, os documentos que constituem a teoria e metodologia, os resultados e os processamentos necessários para reproduzi-los na integra. Tudo isso é realizado sob o paradigma de componentes distribuídos numa arquitetura orientada a serviços, onde todas as interações entre os componentes são realizadas através de padrões interoperáveis.

9.2. Contribuições

A principal contribuição deste trabalho é o desenvolvimento de um *software* aplicado a gestão de conhecimento, capaz de comportar artefatos multimídias e geoespaciais, e que está disponível publicamente. Adicionalmente, dada a multidisciplinaridade dos temas abordados, este trabalho pode ser um referencial importante para trabalhos relacionados, principalmente sobre os temas geoespaciais e arquitetura de *software* com orquestração de serviços. Em relação às ferramentas utilizadas para compor a solução aqui apresentada, foram realizadas comparações entre diversos *softwares*, o que também é uma contribuição valiosa e pode auxiliar a produção de novos trabalhos.

Quanto às publicações, até o presente momento foram produzidos os seguintes artigos:

- RODRIGUES, A. B.; SANTOS, W. A.; SANT'ANNA, S. J. S.; FREITAS, C. C. Proposta de infraestrutura para a gestão de conhecimento científico sensível ao contexto geográfico. In: WORKSHOP DOS CURSOS DE COMPUTAÇÃO APLICADA DO INPE, 12. (WORCAP), 2012, São José dos Campos. **Anais...** São José dos Campos: INPE, 2012.

- RODRIGUES, A. B.; SANTOS, W. A.; SANTANNA, S. J. S.; FREITAS, C. C. Proposta de infraestrutura para a gestão de conhecimento científico sensível ao contexto geográfico. In: SIMPÓSIO BRASILEIRO DE GEOINFORMÁTICA, 13. (GEOINFO), 2012, Campos do Jordão. **Anais...** São José dos Campos: INPE, 2012. p. 140-145. On-line, CD-ROM. ISSN 2179-4820.

9.3. Sugestões para trabalhos futuros

Quanto aos objetivos proposto neste trabalho, todas as funcionalidades foram desenvolvidas, contudo há alguns pontos de extensão e evolução. A primeira, e talvez mais importante sugestão, seja a exploração de ontologias no componente de geoportal. Um serviço de ontologia possibilitaria aos usuários realizarem buscas não apenas por termos específicos, mas também por termos relacionados de acordo com a definição da ontologia. O `Esri GeoPortal Server` possui um serviço de ontologia opcional que pode ser uma alternativa inicial.

Avaliando os custos de infraestrutura, sejam eles diretos e indiretos, e o movimento por parte de algumas instituições para disponibilizarem suas aplicações em ambiente de nuvem, esta também pode ser uma alternativa para esta solução. A arquitetura foi projetada para ser distribuída e independente, o que facilitaria a migração para ambiente de computação em nuvem. Isso seria interessante principalmente para os serviços de processamentos custosos que podem requerer escalabilidade muito grande para um pequeno espaço de tempo.

Na Divisão de Processamento de Imagens do INPE há um grande legado de algoritmos de processamento de imagens desenvolvidos em linguagem `Envi IDL` (*Interactive Data Language*) e que poderiam ser disponibilizados como um serviço OGC WPS. A linguagem IDL possui um aporte tecnológico que permite compartilhar informações com diferentes linguagens, por exemplo entre IDL e Java. Esta tecnologia chama-se *bridge* e a `IDL-Java bridge` permite incorporar o poder de processamento da linguagem IDL em objetos Java. No entanto este assunto não foi explorado e pode representar o enriquecimento

dos serviços de processamento WPS disponíveis na solução e o compartilhamento destes conhecimentos já produzidos.

REFERÊNCIAS BIBLIOGRÁFICAS

ADOBE. **Sítio web da Adobe**. 2011. Disponível em: <http://www.adobe.com/devnet/actionscript.html>. Acesso em: 17 set. 2011.

AIME, A.; DEOLIVEIRA, J. Comparing the Performance of Open Source Web Map Servers. In: FREE AND OPEN SOURCE FOR GEOSPATIAL CONFERENCE (FOSS4G 2008). 2008, Cape Town, South Africa. **Proceedings…** Cape Town, Osgeo, 2008.

ALAMEH, N. Chaining geographic information Web services. **IEEE Internet Computing**, v. 7, n. 5, p. 22-29, 2003.

ALFRESCO. **Apache Chemistry: The Alfresco Open Source Implementation of CMIS.** 2011. 46 transparências. Disponível em: <http://www.slideshare.net/alfresco/apache-chemistrywebinar/>. Acessado em: 11 jun. 2012.

ALFRESCO. **Sítio web do Alfresco**. 2012. Disponível em: <www.alfresco.com>. Acesso em: 17 set. 2012.

ANSI. **Sítio web da American National Standards Institute**. 2012. Disponível em: <http://www.ansi.org/>. Acesso em: 17 set. 2012.

BARRETO, A. A. A questão da Informação. **Revista São Paulo em Perspectiva**. Fundação Seade, v. 8, n. 4, 1994.

BARRY, D. K. **Web Services and service-oriented architectures**: the savvy manager's guide. San Francisco, CA: Morgan Kaufmann, 2003. 245 p. ISBN (1558609067).

BHAUMIK, S. **Alfresco 3 Cookbook**. Birmingham, UK: Packt Publishing, 2011. 380 p. ISBN 184951108X.

BRAGA, G. M. Informação, ciência da informação: breves reflexões em três tempos. **Ciência da Informação**, v. 24, n. 1, p. 84-88, 1985.

BRASIL. **Decreto nº 6666, de 27 de novembro de 2008.** Institui no âmbito do Poder Executivo Federal, a Infra-Estrutura Nacional de Dados Espaciais – INDE, e dá outras providências. Brasília, 2008.

BRODY, T.; HARNAD, S. **The research impact cycle.** 2003. 56 transparências. Disponível em: <http://opcit.eprints.org/feb19oa/harnad-cycle.ppt>. Acesso em: 01 jun. 2012.

CAMARA, G.; DAVIS, C. **Arquitetura de sitemas de informação geográfica**. In: CÂMARA, Gilberto; DAVIS, Clodoveu; MONTEIRO, Antônio Miguel Vieira (Ed.). Introdução à ciência da geoinformação. São José dos Campos: INPE, 2001. p. 35. (INPE-8078-PRE/3893). Disponível em:

<http://www.dpi.inpe.br/gilberto/livro/introd/cap3-arquitetura.pdf>. Acesso em: 01 mai. 2012.

CAÑETE, S. C.; TAVARES, D. L. M.; ESTRELA, P. C.; FREITAS, T. R. O.; HENKIN R.; GALANTE, R., FREITAS, C. M. D. S. Integrando visualização e análise de dados em sistema de gerenciamento de dados de biodiversidade. In: IV BRAZILIAN E-SCIENCE WORKSHOP. 2010. Belo Horizonte, Brasil. **Anais...** Belo Horizonte, SBC, 2010.

CARDOSO, J. **Semantic Web Services**: Theory Tools and Applications. EUA: IGI Global, 2007. 372 p. ISBN (978-1-59904-047-9).

CARTARO. **Sítio web do Cartaro**. 2012. Disponível em: <http://cartaro.org>. Acesso em: 07 set. 2012.

CINDE - Comitê de Planejamento da Infraestrutura Nacional de Dados Espaciais. **Plano de Ação para Implantação da INDE – Infraestrutura Nacional de Dados Espaciais.** Rio de Janeiro, 2010. Disponível em: <http://www.concar.ibge.gov.br/arquivo/PlanoDeAcaoINDE.pdf>. Acesso em: 01 mai. 2012.

COLAN, M. **Service-Oriented Architecture expands the vision of Web services, Part 1: Characteristics of Service-Oriented Architecture**. IBM, 2004. Disponível em: <http://www.ibm.com/developerworks/library/ws-soaintro/index.html>. Acesso em: 03 out. 2012.

COLLINS, H. M. Tacit knowledge, trust and the Q of sapphire. **Social Studies of Science**, n. 31, p. 71-85, 2001.

COX, S.; Daisey, P.; Lake, R.; Portele, C.; Whiteside, A. **OpenGIS Geography Markup Language (GML) Implementation Specification**. Open Geospatial Consortium, Inc. 2003.

CSBC – Congresso da Sociedade Brasileira de Computação. **VI Workshop de e-Science**. Curitiba, 2012. Disponível em: <http://www.imago.ufpr.br/csbc2012/anais_csbc/eventos/e-science/index2.html>. Acesso em: 12 set. 2012.

CURBERA, F.; KHALAF, R.; MUKHI, N.; TAI, S.; WEERAWARANA, S. The next step in Web services. **Communications of the ACM**, n. 46, p. 29-34, 2003.

DANGERMOND, J. GIS Helping manage our world. **ESRI ARCNEWS**, n. 3, v. 27, 2005. Disponível em: <http://www.esri.com/software/arcgis/geoportal>. Acesso em: 20 set. 2012.

DAVENPORT, T. H., PRUSAK, L. **Conhecimento empresarial**. Rio de Janeiro: Campus, 1998. 256 p. ISBN (8535203524).

DAVENPORT, T. **Ecologia da informação**: porque só a tecnologia não basta para o sucesso na era da informação. São Paulo: Futura, 1998. 316 p. ISBN (85-86082-72-4).

DAVENPORT, T. H. Data to knowledge to results: building an analytic capability. **California Management Review**, v. 43, n. 2, p. 117-138, 2001.

DAVIS JR, C. A.; BORGES, K. A. V.; SOUZA, L. A.; CASANOVA, M. A.; LIMA JÚNIOR, P. O. **Banco de Dados Geográficos**. c. 11. Curitiba: MundoGEO, 2005. Disponível em: <http://www.dpi.inpe.br/livros/bdados/cap11.pdf> Acesso em: 03 set. 2012.

DEELMAN, E., GANNON, D., SHIELDS, M., TAYLOR, I. Workflows and e-Science: An overview of workflow system features and capabilities. **Future Generation Computer Systems**, v. 25, n. 5, p. 528-540, 2009.

DCMI. **Sitio web da Dublin Core Metadata Initiative**. 2012. Disponível em: <http://dublincore.org/>. Acesso em: 11 out. 2012.

DRUCKER, P. **Desafios gerenciais para o século XXI**. São Paulo: Pioneira, 1999. 168 p. ISBN (8522102007).

DRUPAL. **Sítio web do Drupal**. 2012. Disponível em: <http://drupal.org/>. Acesso em: 07 set. 2012.

FREEMAN, A.; CHAPMAN, B.; HESS, L.; SIQUEIRA, P.; HOLT, J.; DUTRA, L. V. **Mapping inundation of the Amazon Basin**. 1999.

GEOSERVER. **Sítio web do GeoServer.** 2012. Disponível em: <http://br.geoserver.org/>. Acesso em: 07 set. 2012.

GRANELL, C. ET AL. Service Composition for SDIs: integrated components creation. In: II INTERNATIONAL WORKSHOP ON GEOGRAPHIC INFORMATION MANAGEMENT (GIM'05). 2005. Copenhagen, Denmark. **Proceedings…** Copenhagen, IEEE, 2005.

GUNZER, H. **Introduction to Web Services.** Universidade Federal de Pernambuco, 2002. Disponível em: <http://www.cin.ufpe.br/~redis/middleware/gunzer-webservices02.pdf >. Acesso em: 14 ago. 2012.

HERNANDEZ FILHO, P.; SHIMABUKURO, Y. E.; LEE, D. C. L.; SANTOS FILHO, C. P.; ALMEIDA, R. R. **Relatorio Final do projeto de inventario florestal na Floresta Nacional do Tapajos**. Sao Jose dos Campos: INPE, 1993. 121 p. (INPE-5422-PRP/170).

HILLMANN; D. **Using Dublin Core - The Elements**. Dublin Core, 2005. Disponível em: <http://dublincore.org/ documents/usageguide/elements.shtml>. Acesso em: 25 jun. 2012.

HOLMES, C.; DOYLE, A.; WILSON, M. Toward a Free and Open Source (FOSS) Spatial Data Infrastructure. In: FROM PHARAOHS TO GEOINFORMATICS - FIG WORKING WEEK 2005 AND GSDI-8. 2005, Cairo, Egypt. **Proceedings…** Cairo, FIG, 2005.

HÜBNER, C. E.; OLIVEIRA, F. H. Gestão da Geoinformação em Implementações Multiusuários. In: CONGRESSO BRASILEIRO DE CADASTRO TÉCNICO MULTIFINALITÁRIO (COBRAC 2008). 2008, Florianópolis, Brasil. **Anais...** Florianópolis, COBRAC, 2008.

JACOBY, S. ET AL. Developing a common spatial data infrastructure between State and Local Government - an Australian case study. **International Journal of Geographic Information Science**, n. 16, p. 305-322, 2002.

KLOPFER, M. **Interoperability & Open Architectures**: an analysis of existing standardisation processes & procedures. Open Geospatial Consortium, 2005. 26 p.

LBA – Large Scale Biosphere-Atmosphere Experiment in Amazonia. **The ecological component of an integrated Amazon study**: the effect of conversion. Manaus, 1995. 55 p. Disponível em: <http://lba.cptec.inpe.br/lba/index.html>. Acesso em: 14 mai. 2011.

LBA – Large Scale Biosphere-Atmosphere Experiment in Amazonia. **Sítio web.** 2011. Disponível em: <http://lba.cptec.inpe.br/lba/site/>. Acesso em: 28 mar. 2011.

LAWRENCE, S. Free online availability substantially increases a paper´s impact. **Nature Web Debates**. 2001. Disponível em: <http://www.nature.com/nature/debates/e-access/Articles/lawrence.html>. Acesso em: 10 jun. 2012.

LEITE, F. C. L. **Gestão do conhecimento científico no contexto acadêmico: proposta de um modelo conceitual**. 2006. 240 p. Dissertação (Mestrado em Ciência da Informação) – Universidade de Brasília, Brasília, 2006.

LEITE, F. C. L.; COSTA, S. Repositórios institucionais como ferramentas de gestão do conhecimento científico no ambiente acadêmico. **Perspectivas em Ciências da Informação**, v. 11, n. 2, p. 206-219, 2006.

LISBOA FILHO, J.; IOCHPE, C. Modelagem de Bancos de Dados Geográficos. In: XX CONGRESSO BRASILEIRO DE CARTOGRAFIA. 2001, Porto Alegre, Brasil. **Anais...** Porto Alegre, Sociedade Brasileira de Cartografia, 2001.

LUDASCHER, B.; ALTINTAS, I.; BERKLEY, C.; HIGGINS, D.; JAEGER-FRANK, E.; JONES, M.; LEE, E.; TAO, J.; ZHAO, Y. Scientific workflow management and the Kepler system. **Concurrency and Computation: Practice and Experience**, n. 18, p. 1039–1065, 2006.

MACHLUP, F.; MANSFIELD, U. **The study of information**: Interdisciplinary messages. New York: Wiley, 1983. ISBN (047188717X).

MAGUIRE, D. J.; LONGLEY, P. A. The Emergence of Geoportals and Their Role in Spatial Data Infrastructures. **Computers, Environment and Urban Systems**, v. 29, n. 1, p. 3-14, 2005.

MATOS, E.; NOCELLI, C.; BRAGA, R.; CAMPOS, F. MathWS: Broker de Serviços Web para e-Science. In: I E-SCIENCE WORKSHOP (SBES 2007). 2007, João Pessoa, Paraíba, Brasil. **Anais...** João Pessoa, SBES, 2007.

MEADOWS, A. J. **A comunicação científica**. Brasília: Briquet de Lemos, 1999. 268 p. ISBN (85-85637153)

MENDES, L. F., BRAGA, R. M., CAMPOS, F. C. A. Busca, Inovação e Registro de Serviços Web. In: IADIS CONFERÊNCIA IBERO-AMERICANA WWW/INTERNET (CIAWI 2008). 2008, Lisboa, Portugal. **Anais...** Lisboa, CIAWI, 2008.

MOURA, P. **Calibração dos modelos digitais de elevação gerados por interferometria SAR nas bandas PEX**. 2005. 238 p. (INPE-15241-TDI/1328). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2005.

NONAKA, I.; TAKEUCHI, H. **Criação do conhecimento nas empresas**: Como as empresas japonesas geram a dinâmica da inovação. Rio de Janeiro, 1997. 358 p. ISBN (8535201777).

NUXEO. **Sítio web do Nuxeo.** 2012. Disponível em: <http://www.nuxeo.com/en>. Acesso em: 07 jan. 2012.

OASIS. **Content Management Interoperability Services (CMIS) Version 1.0.** OASIS Standard, 2010.

OGC. **Geospatial Portal Reference Architecture, A Community Guide to Implementing Standards-Based Geospatial Portals.** Open Geospatial Consortium, Inc., 2004.

OGC. **Padrões OGC.** OGC, 2013. Disponível em: <http://www.opengeospatial.org/standards>. Acessado em: 01 fev. 2013.

PALAZZI, D.; SILVA, L.; MENDES, L. F.; GASPAR, W.; MATOS, E.; CAMPOS, F.; BRAGA, R. **Uso de Ontologias em Projetos de e-Science**. 2009.

PAPAZOGLOU, M. P.; GEORGAKOPOULOS, D. Service-Oriented Computing. **Communications of the ACM**, n. 46, 2003.

PARMA, G. C. Mapas cadastrais na Internet: servidores de mapas. In: CONGRESSO BRASILEIRO DE CADASTRO TÉCNICO MULTIFINALITÁRIO

(COBRAC 2006). 2006, Florianópolis, Brasil. **Anais...** Florianópolis, COBRAC, 2006.

PERCIVALL, G. **OpenGIS Reference Model**. Open Geospatial Consortium, Inc., 2003.

PEREIRA, L. O. **Avaliação de métodos de integração de imagens ópticas e de Radar para a classificação do uso e cobertura da terra na Região Amazônica**. 2012. 270 p. (sid.inpe.br/mtc-m19/2012/08.30.12.50-TDI). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2012.

PHILLIPS, A. ET AL. Spatial Data Infrastructure concepts. **The Australian Surveyor**, n. 44, p. 20-28, 1999.

QUEIROZ, G. R. **Algoritmos geométricos para banco de dados geográficos**: da teoria à prática na TerraLib. 2003. 145 p. (INPE-12150-TDI/971). Dissertação (Mestrado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2003.

QUEIROZ, G. R. **Web Map Service – WMS**. Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2007a. 27 transparências. Disponível em: <http://www.dpi.inpe.br/~gribeiro/apresentacoes/wms_2007.pdf>. Acesso em: 17 set. 2012.

QUEIROZ, G. R. **Web Feature Service – WFS**. Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2007b. 26 transparências. Disponível em: <http://www.dpi.inpe.br/~gribeiro/apresentacoes/wfs_2007.pdf>. Acesso em: 17 set. 2012.

QUEIROZ, G. R.; XAVIER, E.; OLIVEIRA, V.; CRUZ, S. A. B. **OGC Web Services**. Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2007. 325 transparências Disponível em: <http://www.dpi.inpe.br/~gribeiro/apresentacoes/ogc_ows_2007.pdf>. Acesso em: 17 set. 2012.

RAJABIFARD, A. ET AL. From Local to Global SDI Initiatives: a pyramid of building blocks. In: 4th Global Spatial Data Infrastructure Conference (GSDI 200). 2000, Cape Town, South Africa. **Proceedings…** Cape Town, GSDI, 2000.

RODRIGUES, A. B.; SANTOS, W. A.; SANT'ANNA, S. J. S.; FREITAS, C. C. Proposta de infraestrutura para a gestão de conhecimento científico sensível ao contexto geográfico. In: WORKSHOP DOS CURSOS DE COMPUTAÇÃO APLICADA DO INPE, 12. (WORCAP), 2012, São José dos Campos. **Anais...** São José dos Campos: INPE, 2012.

RODRIGUES, A. B.; SANTOS, W. A.; SANTANNA, S. J. S.; FREITAS, C. C. Proposta de infraestrutura para a gestão de conhecimento científico sensível ao

contexto geográfico. In: SIMPÓSIO BRASILEIRO DE GEOINFORMÁTICA, 13. (GEOINFO), 2012, Campos do Jordão. **Anais...** São José dos Campos: INPE, 2012. p. 140-145. On-line, CD-ROM. ISSN 2179-4820.

SANTOS, J. R.; NEEFF, T.; ARAUJO, L. S.; GAMA, F. F.; DUTRA, L. V.; SOUSA JÚNIOR, M. A. Interferometria Sar (Bandas X E P) na estimativa de biomassa florestal. In: SIMPÓSIO DE CIÊNCIAS GEODÉSICAS E TECNOLOGIAS DA GEOINFORMAÇÃO, 1. Recife. **Anais...** 2004. On-line. (INPE-11419-PRE/6855).

SERVELLO, E. L.; KUPLICH, T. M.; SHIMABUKURO, Y. E.; SANO, E. E. **Imagens polarimétricas RADARSAT-2 na discriminação de coberturas da terra na Floresta Nacional do Tapajós**. 2009.

SHEKHAR, S.; XIONG, H. **Spatial Data Infrastructure**. Encyclopedia of GIS. New York: Springer, 2007. 1359 p. ISBN (038730858X).

SHIMABUKURO, Y. E.; AMARAL, S.; AHERN, F. J.; PIETSCH, R. W. Land cover classification from RADARSAT data of the Tapajós National Forest, Brazil. **Canadian Journal of Remote Sensing**, n. 24, p. 393-401, 1998.

SMITH, H. A.; MCKEEN, J. D. Developments in practice VIII: enterprise content management. **Communications of the Association for Information Systems**, v. 11, n. 1, p. 647-659, 2003.

TAIT, M. G. Implementing Geoportals: Applications of Distributed GIS. **Computers, Environment and Urban Systems**, n. 29, p. 33–47, 2005.

TOMCAT. **Sítio web do TomCat.** 2012. Disponível em: <http://tomcat.apache.org/>. Acesso em: 07 jan. 2012.

UNICAMP – Universidade de Campinas. **Sítio Web do Centro de Computação.** UNICAMP, 2011?. Disponível em: <http://www.unicamp.br/e-science/>. Acesso em: 26 de mar. 2011.

VERONA, J. D. **Classificação e monitoramento fenológico foliar da cobertura vegetal na região da floresta Nacional do Tapajós - Pará, utilizando dados multitemporais do sensor "thematic mapper" (TM) do Landsat**. 2002. 159 p. (INPE-9556-TDI-832). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2002.

W3C – World Wide Web Consortium. **Web Services Architecture**. W3C, 2003. Disponível em <http://www.w3.org/TR/2002/WD-ws-arch-20021114>. Acesso em: 11 set. 2012.

W3C – World Wide Web Consortium. **Resource Description Framework**. W3C, 2004. Disponível em < http://www.w3.org/RDF/>. Acesso em: 11 set. 2012.

WEISETH, P. E.; OLSEN, H. H.; TVEDTE, B.; KLEPPE, A. **Statoil´s eCollaboration Strategy for 2002-2004**. Trondheim/Stavanger, Noruega. 2002.

WORBOYS, M. F. **GIS**: A Computing Perspective. London: Taylor and Francis, 1995. ISBN (0748400648/0748400656).

YANASSE, C. C. F.; SANT'ANNA, S. J. S.; FRERY, A. C.; RENNÓ, C. D.; SOARES, J. V.; LUCKMAN, A. J. Exploratory study of the relationship between tropical forest regeneration stage and SIR-C L and C data. **Remote Sensing of Environment**, n. 59, p. 180-190, 1997.

APÊNDICE A – Serviços disponíveis no padrão CMIS

A Tabela A.1 descreve os serviços *web* disponíveis no padrão CMIS e fornece uma descrição de cada um dos métodos disponíveis em cada um destes serviços *web*.

Tabela A.1: Serviços disponíveis no padrão CMIS.

Serviço	Método	Descrição
Repository Services	getRepositories	Obtém a lista de repositórios que podem ser acessados através deste *endpoint*.
	getRepositoryInfo	Obtém as informações sobre determinado repositório, como o nome do produto, versão do produto, versão suportada do padrão CMIS e quais funcionalidades do CMIS disponíveis.
	getTypeChildren	Obtém a lista de tipos de objetos descendentes de um tipo específico definido para o repositório.
	getTypeDefinition	Obtém a definição de um tipo específico com todas as definições de suas propriedades.
Navigation Services	getChildren	Obtém os objetos filhos que estão contidos no diretório especificado.
	getDescendants	Obtém os objetos filhos que estão contidos no diretório especificado e em todos os subdiretórios.
	getFolderTree	Obtém os diretórios e subdiretórios filhos de um diretório específico.
	getCheckedoutDocs	Retorna os conteúdos que estão com a opção de *checkout* para o usuário corrente.
	getObjectParents	Retorna os diretórios pai de um conteúdo.
	getFolderParent	Retorna o diretório pai de um diretório.
Object Services	createDocument	Cria um documento com o tipo especificado no diretório especificado.
	createDocument-FromSource	Cria um novo documento copiando um existente.

(continua)

Tabela A.1 – Continuação.

Object Services	createFolder	Cria um diretório com o tipo especificado no diretório especificado.
	createRelationship	Cria um relacionamento de um tipo especificado.
	createPolicy	Cria um objeto de política de um tipo especificado.
	getAllowableActions	Lista as ações permitidas pelo usuário sobre um conteúdo especifico.
	getObject	Lê um objeto do repositório e retorna suas propriedades, relacionamentos, políticas, listas de controle de acesso (*Access Control List* - ACL) e ações permitidas.
	getProperties	Obtém a lista de propriedades de um objeto.
	getObjectByPath	Obtém um objeto através do caminho amigável até ele. Por exemplo: /diretorio/subdiretorio/documento.gif
	getContentStream	Obtém o formato digital do conteúdo informacional (*raw data*).
	getRenditions	Obtém as *renditions* associadas ao documento.
	updateProperties	Atualiza as propriedades de um objeto específico.
	moveObject	Move um conteúdo de um diretório para outro.
	deleteObject	Exclui um objeto específico pelo seu identificador.
	deleteTree	Exclui o diretório específico e todos seus filhos.
	setContentStream	Atualiza o conteúdo (*raw data*) de um documento específico.
	deleteContentStream	Exclui o conteúdo de um documento específico.
Multifiling Services	addObjectToFolder	Adiciona um documento existente a um diretório.
	removeObjectFromFolder	Remove um documento de um diretório ou de todos.

(continua)

Tabela A.1 – Conclusão.

Discory Services	Query	Executa uma busca baseada na linguagem CMIS SQL.
	getContentChanges	Retorna a lista de alterações de um objeto.
Versioning Services	checkout	Executa a operação de *checkout* de um documento.
	cancelCheckout	Cancela uma operação de *checkout* de um documento.
	checkIn	Libera um documento de uma operação de *checkout* (*check in*).
	getObjectOfLastVersion	Obtém a última versão de um documento.
	getPropertiesOfLastVersion	Obtém as propriedades da última versão de um documento.
Relationship Services	getObjectRelationships	Obtém os relacionamentos de um objeto específico.
Policy Services	getAppliedPolicies	Obtém a lista de políticas aplicáveis ao objeto especificado.
	applyPolicy	Aplica uma política a um objeto.
	removePolicy	Remove uma política de um objeto.
ACL Services	getACL	Retorna a ACL corrente para o conteúdo especificado.
	applyACL	Aplica uma nova ACL adicionando ou removendo uma entrada de controle de acesso (*Access Control Entry)*.

APÊNDICE B – Termos comuns utilizados na especificação OWS

***Bounding Box*:** retângulo envolvente ou espaço delimitado numa representação do espaço, que possui um limite inferior e um limite superior em cada dimensão do sistema de referência de coordenadas. Pode ser usado para especificar restrições espaço-temporais em uma consulta, ou para delimitar uma localização aproximada de um objeto ou conjunto de objetos (feições).

***Capabilities*:** descrição das funcionalidades em XML - metadados do serviço codificados no formato XML.

***Metadata*:** metadados do serviço que descrevem as operações e dados georreferenciados disponíveis em um servidor.

***Operation*:** especificação de um processamento, operação ou consulta que um determinado objeto pode executar.

***Server*:** instância de um serviço, *software* que disponibiliza serviços.

***Service*:** funcionalidade sob a forma de serviço provida por uma entidade por meio de interfaces. Esta funcionalidade é disponibilizada por um provedor de serviços a seus usuários.

***Request*:** invocação de uma operação por um cliente, é uma requisição de uma operação.

***Response*:** resultado de uma operação, retornada de um servidor ao cliente, isto é, uma resposta do servidor a um *request*.

***Client*:** componente que solicita (requer) uma funcionalidade (operação) disponível num servidor, isto é, um *software* cliente.

***Interface*:** conjunto de operações que caracterizam o comportamento de uma entidade.

***Parameter*:** variável parâmetro cujo nome e valor são incluídos em uma operação de requisição ou de resposta do servidor.

***Resource*:** unidade de informação ou serviço que possui um endereço, tais como arquivos, imagens, programas e resultados de consultas. No contexto de um OWS, um recurso deve possuir um endereço referenciado por uma URI (*Uniform Resource Identifier*).

***Version*:** um padrão de OWS evolui com o tempo; a cada mudança é atribuída uma versão. Ao prover-se um OWS deve-se informar a versão em que foi implementada.

APÊNDICE C – Artigo apresentado.

Artigo apresentando no XIII Brazilian Symposium on GeoInformatics, de 25 a 27 de Novembro de 2012.

Proposta de infraestrutura para a gestão de conhecimento científico sensível ao contexto geográfico

Alaor Bianco Rodrigues[1], Walter Abrahão dos Santos[1], Sidnei J. Siqueira Santanna[2], Corina da Costa Freitas[2]

[1]Laboratório de Matemática e Computação Aplicada, LAC – Instituto Nacional de Pesquisas Espaciais (INPE) - São José dos Campos, SP – Brasil

[2]Divisão de Processamento de Imagens, DPI – Instituto Nacional de Pesquisas Espaciais (INPE) - São José dos Campos, SP – Brasil

alaorbianco@yahoo.com.br, walter.abrahao@lac.inpe.br, {sidnei, corina}@dpi.inpe.br

Abstract. *This work discuss how the area of e-Science has been exploited to develop an infrastructure capable of helping the management of scientific knowledge produced in the Image Processing Division at INPE, focusing on but not limited to, geospatial artifacts, applying a case study using as inputs several studies conducted by researchers at INPE in the area of the Tapajos National Forest..*

Resumo. *Neste trabalho é abordado como a área de e-Science foi explorada para o desenvolvimento de uma infraestrutura capaz de auxiliar a gestão do conhecimento científico produzido na Divisão de Processamento de Imagens do INPE, com foco em, mas não limitado a, artefatos sensíveis ao contexto geográfico, aplicado um estudo de caso usando como insumos diversos trabalhos realizadas por pesquisadores do INPE na região da Floresta Nacional do Tapajós.*

1. Introdução

No início da criação de um novo conhecimento, o esforço de um pesquisador parte daquilo que foi construído anteriormente por outros pesquisadores, ou seja, recorre à literatura de sua especialidade, e, ao fim, divulga os resultados de sua pesquisa por meio dos veículos de comunicação apropriados à sua área de conhecimento.

Percebe-se assim a importância da comunicação, informar ao mundo científico seus feitos, resultados e etc. Meadows (1999) diz que a comunicação reside no coração da ciência, sendo tão vital quanto a própria pesquisa. No entanto apenas uma fração do que é produzido durante uma pesquisa é publicado, ou seja, é formalmente comunicado a comunidade. Braga (1985) ressalta que a comunicação formal é responsável por apenas 20% de todas as comunicações no processo de geração do conhecimento. Sendo que as demais são constituídas de processos informais, e uma grande parcela desse conhecimento encontra-se em um formato que poderia ser explícito, como anotações, planilha de resultados, registros de experimentos, resultados parciais, etc.

Os recursos computacionais e o ambiente *Web* muito contribuem para um cenário de compartilhamento e comunicação. Os recursos computacionais facilitam o trabalho em rede, podendo manter os conhecimentos descentralizados junto aos locais em que são mais gerados e/ou utilizados (DAVENPORT et al., 1998) e melhorando o grau de interatividade do usuário com os registros de conhecimentos (DAVENPORT et al., 1998). A computação é efetivamente útil para a gestão do conhecimento, se for empregada utilizando-se uma sistemática interferência (interatividade) humana (DAVENPORT, 2001).

2. Gestão do Conhecimento

Gestão do conhecimento é um tema relativamente novo, multidisciplinar e muito explorado em diversas pesquisas, mas quase sempre seu foco são as organizações empresariais. No entanto há iniciativas da aplicação destes conceitos da gestão de conhecimento sob o âmbito do conhecimento científico como dissertado, principalmente, em (LEITE, 2006).

Nonaka e Takeuchi (1997) forneceram uma grande contribuição para o assunto, sendo suas obras as maiores referências atualmente. Estes autores realizaram uma importante distinção entre os tipos de conhecimento humano, classificando-os em conhecimento tácito e conhecimento explícito. Sendo os

conhecimentos explícitos aqueles estruturados capazes de serem verbalizados, facilmente transmitido, sistematizado e comunicado. Já os conhecimentos tácitos são aqueles inerentes às pessoas, isto é, o conhecimento pessoal incorporado à experiência individual, crenças e valores. É difícil de ser articulado na linguagem formal e transmitido por se tratar da parcela não estruturada do conhecimento.

Nonaka e Takeuchi (1997), ainda, consideram que um trabalho efetivo com o conhecimento somente é possível em um ambiente em que possa ocorrer a contínua conversão entre esses dois formatos. Segundo estes autores são 4 os processos de conversão entre os dois tipos de conhecimento: socialização, externalização, combinação e internalização.

3. Gerenciamento de Conteúdo

O conceito de *Enterprise Content Management* (ECM) compreende "as estratégias, ferramentas, processos e habilidades que uma organização precisa para gerenciar seus ativos de informação durante o seu ciclo de vida", incluindo todos os ativos digitais, como documentos, dados, relatórios e páginas da *web* (SMITH e MCKEEN 2003). O *Meta Group* o define como a tecnologia que fornece os meios para criar, capturar, gerenciar, armazenar, publicar, distribuir, pesquisar, personalizar, apresentar e imprimir qualquer conteúdo digital (imagens, texto, relatórios, vídeo, áudio, dados transacionais, catálogo, código). Estes sistemas se concentram na captura, armazenamento, recuperação e disseminação de arquivos digitais para uso corporativo. (WEISETH et al. 2002).

Enterprise Content Management System (ECMS), ou simplesmente *Content Management System* (CMS) é a expressão utilizada para descrever ferramentas que promovem meios de gerenciamento, publicação e manutenção destes ativos de informação. Esta categoria de sistemas ainda incluem funcionalidades de fórum, listas de discussões, *workflows*, controle de acesso, associações, classificação e categorização, o que cria um ambiente propício para gestão do conhecimento uma vez que facilitam a existência,

manutenção e crescimento dos processos de transformação citados em Nonaka e Takeuchi (1997). Assumindo não apenas o papel de uma infraestrutura para tal, mas também criando condições ambientais e motivacionais que faça com que as pessoas vivam e reforcem estes ciclos de transformação, por:

1) Estimular o processo de **socialização** do conhecimento uma vez que a diversidade de formatos em que as informações podem existir criam condições favoráveis à assimilação do conhecimento. O resultado é uma transferência da informação e do conhecimento mais efetiva, pois muito do conhecimento científico gerado por um pesquisador não é possível de ser comunicado por meios formais e transforma parte do conhecimento que antes era puramente tácito em conhecimento explicito. Ainda sob a ótica da socialização, é estimulada a interação informal entre pesquisadores interessados em um mesmo assunto, possibilitando discussões e compartilhamento de ideias e esboços para coleta de sugestões e comentários enriquecendo as pesquisas e intensificando a troca de experiências.

2) Ser instrumento de **externalização** do conhecimento tácito que, segundo Nonaka e Takeuchi (1997), trata-se do processo de criação do conhecimento perfeito, ao fornecerem a possibilidade de armazenar múltiplos formatos desse conhecimento. As publicações científicas são formais e desta forma formatam o conhecimento e de certa forma limita seus horizontes. Uma infraestrutura capaz de armazenar os conhecimentos informais aproxima os demais pesquisadores aos elementos que compõe o estado do conhecimento de seu autor. Neste cenário, parte do conhecimento tácito é transformado em uma estrutura comunicável permitindo que esta seja processada, armazenada e recuperada.

3) Permitir a transformação de um determinado conjunto de conhecimento explícito, por meio de agrupamento, acréscimo, categorização e classificação, criando um novo conjunto de conhecimento ou criando e/ou acrescentando um novo conhecimento, constituindo, assim, o processo de **combinação**.

4) Facilitar o processo de **internalização** por criar condições favoráveis para que o conhecimento explícito armazenado seja convertido em conhecimento tácito do indivíduo.

4. Revisão de Literatura

Alguns autores veem desenvolvendo trabalhos sobre o tema de gestão de conhecimento científico e estudando ferramentas e alternativas para auxiliar e facilitar os processos envolvidos em tais atividades. Leite e Costa (2006) discutem a adequação e aplicabilidade de repositórios institucionais como uma ferramenta para tal, abordando as peculiaridades do conhecimento científico, bem como o ambiente no qual se dão os processos de sua criação, compartilhamento e uso.

Contexto semelhante foi explorado por Cañete et al. (2010) ao desenvolver um sistema de informações de biodiversidade baseado em banco de dados, API do Google Maps e o sistema R, que permite catalogar dados a respeitos de espécimes coletadas, analisá-los e apresentá-los num mapa.

Este trabalho se diferencia por adotar plataformas abertas e consolidadas no mercado, reduzindo customizações e sendo muito aderente a padrões existentes. Visa não apenas ser um repositório de dados, mas uma plataforma que permita que os processos de transformação do conhecimento ocorram e sejam incentivados. Ainda, por manusear dados matriciais (*raster*) e prover um barramento de serviços sobre estes.

5. Metodologia

Na fase de levantamento, foram realizadas entrevistas com alguns usuários da Divisão de Processamento de Imagens do INPE (Pesquisadores) que representavam os demais usuários. Foram elencadas suas necessidades e criando uma lista de requisitos, conforme pode ser observado na Tabela 1.

ID	Requisito
RQ001	A solução deve contemplar um sistema de fórum.
RQ002	A solução deve contemplar um sistema de listas de discussões.
RQ003	A solução deve ser de acesso público, mas com recursos de restrições de acesso a determinados conteúdos caso pertinente.
RQ004	A solução deve contemplar mecanismo de armazenamento de arquivos multimídias.
RQ005	A solução deve ser suficientemente configurável de modo que possam ser definidos quais metadados importantes para cada tipo de conteúdo.
RQ006	A solução deve possuir mecanismo de busca por todo o conteúdo textual.
RQ007	A solução deve contemplar mecanismo de classificação de conteúdo por rótulos.
RQ008	A solução deve prover conteúdo geográfico segundo padrões abertos OGC (WMS, WFS, WCF, WPS).
RQ009	A solução deve usar produtos de *software* livre, preferencialmente de código fonte aberto e na linguagem Java.
RQ010	A solução deve contemplar a manipulação, armazenamento e recuperação de imagens vetoriais (*raster*).
RQ011	A solução deve contemplar o agrupamento e o relacionamento de conteúdos.
RQ012	A solução deve contemplar o referenciamento geográfico dos conteúdos.
RQ013	A solução deve contemplar a plotagens dos elementos georreferenciados no mapa.
RQ014	A solução deve ser capaz de consumir serviços *web* de geolocalização, GeoRSS, WPS e BaseMaps.

Tabela 1: Requisitos da solução

Analisando os requisitos, foi possível perceber que grande parte dos requisitos eram elucidados por uma ferramenta de CMS, caso dos requisitos RQ001, RQ002, RQ003, RQ004, RQ005, RQ006, RQ007 e RQ011. Mas ainda assim, havia alguns requisitos que não eram contemplados por esta. Neste caso os requisitos que fogem ao escopo de atuação das ferramentas de CMS são, em essência, relacionados ao contexto espacial. Sendo assim, estes requisitos foram tratados fora do CMS, incluindo na arquitetura da solução um elemento de gerenciamento de conteúdo geográfico.

Esta abordagem implica em integração entre o CMS e o gerenciador de conteúdo geográfico. Esta integração foi realizada utilizando o padrão CMIS -

Content Management Interoperability Services, que é um padrão aberto criado para facilitar a interoperabilidade entre sistemas CMSs.

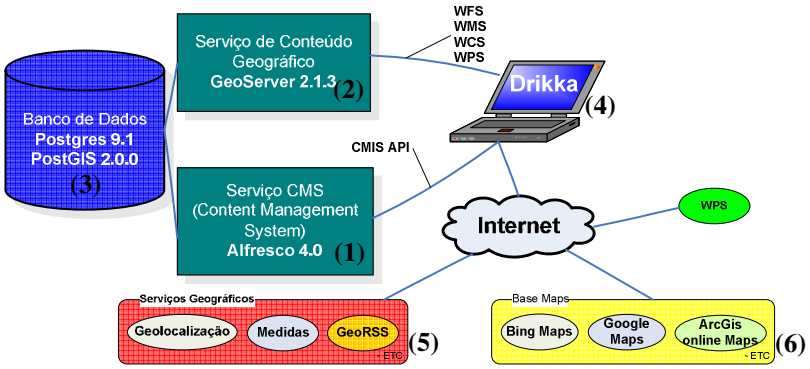

Figura 1: Arquitetura da Solução.

A Figura 1 esboça a macro-arquitetura adotada para esta solução. Foi escolhida a ferramenta Alfresco como CMS (1), que é uma ferramenta desenvolvida em Java e possui uma versão *Community* que é de código aberto. A ferramenta Alfresco é utilizada em várias instituições no mundo, e no Brasil, um grande exemplo de seu emprego é na Dataprev (Empresa de Tecnologia e Informações da Previdência Social) onde é utilizada para facilitar a gestão de documentos durante o processo de compras na empresa.

Como gerenciador de conteúdo geográfico (2), adotou-se o GeoServer, que é um *Software* livre, de código aberto, mantido pelo *Open Planning Project* e que é capaz de integrar diversos repositórios de dados geográficos com simplicidade e alta performance. O GeoServer é um servidor de *Web Map Service* (WMS), *Web Coverage Service* (WCS) e de *Web Feature Service* (WFS) completamente funcional que segue as especificações da *Open Geospatial Consortium* (OGC), além disso ainda provê um barramento de serviços *Web Processing Service* (WPS), outro padrão OGC para serviços de processamento de dados.

Tanto o Alfresco quanto o GeoServer utilizam um banco de dados relacional (3) para persistência e, uma boa solução que atende a ambos, é o Postgres com a extensão espacial PostGIS em sua versão 2.0. A aplicação cliente foi desenvolvida em Flex e consome os dados tanto do CMS via CMIS quanto do

gerenciador de conteúdo geográfico via WMS, WFS e WCS. Além destes serviços, a aplicação cliente ainda consome serviços on-line (*web*) para geolocalização, GeoRSS e medidas de feições (5) desenvolvidos como *plugins* e facilmente extensíveis. O plano de fundo do mapa (*Basemap*) (6) é um outro exemplo de serviço *web* consumido pela aplicação. Hoje é possível incluir Base Map do Google, Bing e ArcGis On-Line.

Muitos dos artefatos a serem manipulados por esta solução são imagens matriciais (*raster*) que são armazenadas no PostGIS. Para armazená-los no banco de dados é utilizada a função raster2pgsql, carregando-as em uma tabela. Cada dado *raster* é carregado em uma tabela própria. Após o carregamento dos dados *raster*, é criado uma representação vetorial (polígono) de sua área de extensão, isso é feito utilizando a função: "SELECT ST_AsBinary(ST_Buffer(ST_Union(rast::geometry), 0.000001)) FROM raster_table". A estratégia de criar um representação vetorial para os dados matriciais é, principalmente, por questões de performance em buscas e não causa efeito colateral, uma vez que visualmente uma imagem matricial propicia poucas informações relevantes.

O modelo de dados para os dados geográficos é muito simples, a princípio têm-se apenas classes de feição para pontos de interesse e áreas (polígonos) de interesse, com identificador, descrição simples e um atributo específico para relacionar o elemento geográfico a um conteúdo no CMS. Como supramencionado, a estratégia de separar o conteúdo geográfico do CMS implica em uma integração entre estes dados. No CMS são cadastrados os metadados de cada um dos conteúdos e cada conteúdo armazenado no CMS possui um identificador único. Caso este conteúdo tenha componente geográfica a ferramenta possibilita que seja criada ou selecionada uma geometria de ponto ou polígono que se relacionará com o conteúdo no CMS através de seu identificador único.

Os conteúdos multimídias, tais como arquivos de texto, planilhas, vídeos, fotos e etc, são armazenados no CMS e utilizam o modelo de metadados Dublin Core. Dublin Core pode ser definido como sendo o conjunto de elementos de

metadados planejado para facilitar a descrição de recursos eletrônicos, sendo um dos padrões mais conhecidos e tradicionalmente adotados em sistemas gerenciadores de conteúdo. Os dados matriciais utilizam um versão estendida do Dublin Core, adicionando alguns atributos específicos para tal. Os usuários do sistema optaram por especificar quais atributos são importantes para os dados matriciais ao invés de usar algum metadado existente, como FGDC (*Federal Geographic Data Committee*) ou ISO 19115, por entender que estes modelos apresentam uma quantidade muito grande de atributos que nem sempre são utilizados e que em geral representa um desincentivo ao uso.

6. Trabalhos Futuros

Como continuidade deste trabalho, está sendo desenvolvido um barramento de serviços WPS que de início proverá uma série de algoritmos de classificação para os dados matriciais. Contará ainda com melhorias no mecanismo de buscas geográficas para os conteúdos vinculados ao CMS.