

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

MARCIO BIGOLIN

**Agrupamento Personalizado de Pontos em
Web Maps usando um Modelo
Multidimensional - APPWM**

Dissertação apresentada como requisito parcial
para a obtenção do grau de Mestre em Ciência
da Computação

Prof. Dr. Renata Galante
Orientador

Porto Alegre, dezembro de 2014.

CIP – CATALOGAÇÃO NA PUBLICAÇÃO

Bigolin, Marcio

Agrupamento Personalizado de Pontos em Web Maps usando um Modelo Multidimensional - APPWM / Marcio Bigolin. – 2014. 94 f.:il.

Orientador: Renata Galante;

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Instituto de Informática, Programa de Pós-Graduação em Computação, Porto Alegre, BR-RS, 2014.

1. Modelo multidimensional. 2. Agrupamento de dados. 3. Web mapping. 4. Geovisualização. I. Galante, Renata, orient. II. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos Alexandre Netto

Vice-Reitor: Prof. Rui Vicente Oppermann

Pró-Reitor de Pós-Graduação: Prof. Vladimir Pinheiro do Nascimento

Diretor do Instituto de Informática: Prof. Luís da Cunha Lamb

Coordenador do PPGC: Prof. Luigi Carro

Bibliotecária-Chefe do Instituto de Informática: Beatriz Regina Bastos Haro

AGRADECIMENTO

Agradeço a Deus, em primeiro lugar, por me dar paciência e sabedoria para ter chegado até aqui.

Ao meus pais, pelo apoio e dedicação em todos esses anos de estudo. Sem eles eu jamais chegaria até aqui.

Agradeço a minha namorada que me suportou nas crises assim como a Meg (nossa cadelinha) que ficava me lambendo os pés nas madrugadas de pesquisa.

À Universidade de Caxias do Sul, especialmente ao ISAM (Instituto de Saneamento Ambiental), assim como a Universidade Federal do Rio Grande do Sul, e ao grupo de Pesquisa do GRID, por dar a oportunidade de explorar um dos problemas e encontrar possíveis soluções.

A minha orientadora Dra Renata Galante e colaboradora Dra Helena Graziottin, pelo apoio desde o início desta jornada e que me proporcionaram muito mais que orientação para este trabalho, mas sim, um grande exemplo de autenticidade, valores éticos e amor ao estudo, que me serviram de espelho para minha própria construção intelectual. Agradeço ao valioso exemplo e amizade.

Agradeço a todos os meus professores, pela atenção a mim dedicada durante a realização deste trabalho,

Enfim, agradeço a todos que, de alguma forma, contribuíram não só para realização deste trabalho, mas para a conclusão de mais uma etapa de minha vida.

RESUMO

Com o avanço da geração de informação georeferenciada torna-se extremamente importante desenvolver técnicas que auxiliem na melhora da visualização dessas informações. Neste sentido os *web maps* tornam-se cada vez mais comuns na difusão dessas informações. Esses sistemas permitem ao usuário explorar tendências geográficas de forma rápida e sem necessidade de muito conhecimento técnico em cartografia e softwares específicos. As áreas do mapa onde ocorre um mesmo evento com maior incidência geram visualizações confusas e que não possibilitam uma adequada tomada de decisão. Essas áreas, quando representadas através de pontos (o que é bastante comum), provocará uma sobreposição massiva de dados, devido à densidade de informações. Esta dissertação propõe uma técnica que utiliza um modelo de dados multidimensional para auxiliar a exibição das informações em um *web map*, de acordo com o contexto do usuário. Esse modelo organiza os dados por níveis geográficos e permite assim uma melhor compreensão da informação exibida. Os experimentos desenvolvidos mostraram que a técnica foi considerada de fácil utilização e de uma necessidade pequena de conhecimento para a execução das tarefas. Isso pode ser visto que das 59 consultas propostas para serem geradas apenas 7 precisam de mudanças significativas para serem executadas. Esses resultados permitem comprovar que o modelo se apresenta como uma boa alternativa para a tomada de decisão sobre mapas produzidos em ambiente *web*.

Palavras-Chave: Modelo de dados, agrupamento de informação, geovisualização, tomada de decisão.

Multidimensional Model for Cluster Points in Web Maps

ABSTRACT

The advancement of generation of geo-referenced information becomes extremely important to develop techniques that help in improving the display of this information. In this sense the web maps become increasingly common in the dissemination of such information. These systems allow the user to explore geographical trends quickly and without much technical knowledge in cartography and specific software . The map areas where there is a single event with a higher incidence generate confusing views and not allow proper decision making. These areas , as represented by points (which is quite common) , will cause a massive overlay data , due to the density of information. This work proposes a technique that uses a multidimensional data model to support the display of information on a web map, according to the user's context . This model organizes data by geographical levels and thus allows a better understanding of the information displayed. Developed experiments showed that the technique was considered easy to use and a small need for knowledge to perform the tasks. It can be seen that the 59 queries proposals to be generated only 7 significant changes need to be executed. These results allow to prove that the model is presented as a good alternative for decision-making on maps produced in a web environment .

Keywords: Data model, clustering, geovisualization, decision making..

LISTA DE FIGURAS

Figura 1.1 – Registro de aves no web map I3Geo do Ministério do Meio Ambiente.....	14
Figura 1.2 – Suinocultores da região COREDE-Serra.....	15
Figura 1.3 – Pontos com registros de estiagens no Rio Grande do Sul no período de 1982 a 2013.....	15
Figura 2.1 – Dimensão achatada com nível representando a hierarquia.....	19
Figura 2.2 – Atributo de dimensão comportando como hierarquia implícita.....	20
Figura 3.1 – Exemplo de Chorem Mapa.....	23
Figura 3.2 – Mosaico de técnicas de agregação de informação.....	25
Figura 3.3 – Implementação clássica x implementação SOLAP.....	28
Figura 3.4 – Detalhe da ferramenta de zoom nível com coxcombs e nível de pontos.....	31
Figura 4.1 – Esquema de utilização do modelo multidimensional.....	33
Figura 4.2 – Modelo multidimensional.....	36
Figura 4.3 – Destaque da dimensão política.....	38
Figura 4.4 – Tabelas de dimensões hidrográficas.....	40
Figura 4.5 – Dimensão física tipo de solo.....	41
Figura 4.6 – Dimensão vegetação e seus atributos.....	42
Figura 4.7 – Dimensão de clinografia.....	43
Figura 4.8 – Taxonomia de geomorfologia.....	44
Figura 4.9 – Dimensão geomorfologia.....	44
Figura 4.10 – Dimensões modeladas para o clima.....	45
Figura 4.11 – Dimensão temporal dia e hora.....	46
Figura 4.12 – Dimensões temporais com dupla relação em relação ao ponto.....	47
Figura 4.13 – Valores indefinidos para hora.....	47
Figura 4.14 – Modelo multidimensional completo.....	51
Figura 5.1 – Ferramenta APPWM.....	53
Figura 5.2 – Legenda do mapa com funcionalidades para controlar os agrupamentos.....	55
Figura 5.3 – Visualização dos agrupamento de dados.....	57
Figura 5.4 – Manutenção dos dados das dimensões.....	58
Figura 5.5 – Geração automática de estilos.....	59
Figura 5.6 – Plano de informação duplicado.....	60
Figura 5.7 – Ícone gerado a partir de web service.....	60
Figura 5.8 – Tamanho de agrupamento.....	61
Figura 5.9 – Carregamento de dados do tipo shape.....	64
Figura 5.10 – Relações entre polígonos e pontos.....	65
Figura 5.11 – Modelo de dados para Experimento.....	66
Figura 5.12 – Múltiplos planos de informação.....	67
Figura 5.13 – Funcionalidade apenas um grupo.....	67
Figura 5.14 – Exploração temporal no APPWM.....	68
Figura 6.1 – Ajuda e treinamento com a ferramenta.....	71
Figura 6.2 – Formulário para identificação de usuário.....	73
Figura 6.3 – Categorias para a escala do SUS.....	76
Figura 6.4 – Descrição perfil usuários ISAM.....	79
Figura 6.5 – Tempo médio no experimento e tempo médio na ferramenta.....	80
Figura 6.6 – Plataforma dos usuários.....	80
Figura 6.7 – Resultados de avaliação através do questionário.....	82

Figura 6.8 – Descrição perfil dos usuários do GRID.....	85
Figura 6.9 – Tempo médio no experimento e tempo médio na ferramenta.....	86
Figura 6.10 – Plataforma dos usuários.....	86
Figura 6.11 – Distribuição das notas do experimento realizado com o GRID.....	87

LISTA DE TABELAS

Tabela 3.1 – Comparação entre as técnicas de visualização de informação.....	25
Tabela 3.2 – Tabela resumo dos recursos disponíveis em cada técnica.....	30
Tabela 3.3 – Comparação entre APPWM e outras ferramentas.....	32
Tabela 4.1 – Intervalos rígidos de Lepsch.....	43
Tabela 4.2 – Tuplas pré-carregadas no modelo.....	48
Tabela 6.1 – Questionário sobre as funcionalidades do sistema.....	74
Tabela 6.2 – Resoluções de tela dos participantes do experimento no ISAM.....	81
Tabela 6.3 – Resoluções de tela dos participantes do experimento no GRID.....	87

LISTA DE ABREVIATURAS E SIGLAS

ANA	Agência Nacional de Águas
APPWM	Agrupamento personalizado de pontos em web maps
BD	Banco de Dados
CNUDM	Convenção das Nações Unidas sobre o Direito do Mar
CRUD	<i>Create Recovery Update Delete</i> (criar, recuperar, atualizar e deletar)
CSS	Cascading Style Sheet
DW	<i>Data Warehouse</i>
ETL	<i>Extraction Transform Load</i> (Extração, Transformação e Carga)
GRID	Grupo de Gestão de Risco de Desastres
GUI	<i>Graphical User Interface</i>
HTML	<i>HyperText Markup Language</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
ISAM	Instituto de Saneamento Ambiental
KML	<i>Keyhole Markup Language</i>
MDX	<i>Multidimensional Expressions</i>
MMA	Ministério do Meio Ambiente
OLAP	<i>Online Analytical Processing</i>
OLTP	<i>Online Transaction Processing</i>
SCC	Sistemas de Classificações Climáticas
SFSSQL	<i>Simple Features Structured Query Language</i>
SGBD	Sistema Gerenciador de Banco de Dados
SIG	Sistema de Informação Geográfica
SQL	<i>Structured Query Language</i>
UCS	Universidade de Caxias do Sul
UFRGS	Universidade Federal do Rio Grande do Sul
UNCLOS	<i>United Nations Convention on the Law of the Sea</i>
WADL	<i>Web Application Description Language</i>
ZEE	Zona Econômica Exclusiva

SUMÁRIO

1 INTRODUÇÃO.....	12
1.1 Problema de pesquisa.....	13
1.2 Objetivos.....	16
1.3 Estrutura da dissertação.....	16
2 CONCEITOS FUNDAMENTAIS.....	17
2.1 Características da modelagem multidimensional.....	17
2.2 Informações geográficas.....	18
2.2.1 Hierarquia geográfica.....	19
2.3 Componentes de um Web Map.....	21
3 TRABALHOS RELACIONADOS.....	22
3.1 Visualização de dados geográficos agregados.....	22
3.2 Sistemas de Informação Geográficos.....	26
3.3 Modelos OLAP e SOLAP.....	27
3.4 Algoritmos de agrupamento.....	28
3.5 Requisitos do problema.....	29
3.6 Comparação com outras ferramentas.....	30
4 MODELO MULTIDIMENSIONAL PARA AGRUPAMENTO DE DADOS.....	33
4.1 Motivação.....	34
4.2 Modelagem proposta.....	35
4.2.1 Dimensão política.....	37
4.2.2 Dimensões físicas.....	38
4.2.2.1 Dimensão hidrográfica.....	39
4.2.2.2 Dimensão tipo de solo.....	40
4.2.2.3 Dimensão vegetação.....	41
4.2.2.4 Dimensão clinografia.....	42
4.2.2.5 Dimensão Geomorfologia.....	43
4.2.2.6 Dimensão clima.....	44
4.2.3 Dimensões socioeconômicas.....	45
4.2.4 Dimensões temporais.....	46
4.3 Dimensões personalizadas.....	47
4.4 Limitações do modelo.....	48
4.5 Considerações sobre o modelo.....	49
5 A FERRAMENTA PARA O AGRUPAMENTO PERSONALIZADO DE PONTOS EM WEB MAPS (APPWM).....	52
5.1 Arquitetura da APPWM.....	52
5.2 Camada de visualização.....	53
5.2.1 Implementação do cliente.....	54
5.2.1.1 Exploração dos grupos.....	56

5.2.2	Implementação do servidor.....	57
5.2.3	Gerência dos dados das dimensões.....	58
5.2.4	Gerador de ícones e estilos.....	59
5.2.5	Geração de planos de informação para os mapas.....	60
5.3	Agrupamento de dados.....	61
5.4	Camada de Implementação dos dados.....	62
5.4.1	SGBD e extensão espacial.....	62
5.4.2	Extraction Transform Load (ETL).....	63
5.5	Implementação dos experimentos.....	65
5.6	Resultados da implementação.....	66
5.7	Considerações sobre a implementação.....	68
6	EXPERIMENTOS.....	69
6.1	Configurações dos Experimentos.....	69
6.1.1	Metodologia.....	70
6.1.1.1	Plano de avaliação.....	70
6.1.1.2	Treinamento.....	70
6.1.1.3	Perfil do usuários.....	72
6.1.1.4	Avaliação pós-execução de tarefas.....	73
6.1.2	Métricas de avaliação.....	75
6.1.2.1	Casos de sucesso e casos de falha.....	75
6.2	Instituto de Saneamento Ambiental - ISAM.....	76
6.2.1	Descrição do Domínio.....	76
6.2.2	Roteiro do experimento.....	77
6.2.3	Perfil dos usuários.....	78
6.2.4	Resultados obtidos.....	81
6.3	Grupo de Gestão de Riscos e Desastres Naturais – GRID.....	83
6.3.1	Descrição do domínio.....	83
6.3.2	Roteiro do experimento.....	84
6.3.3	Perfil dos usuários.....	84
6.3.4	Resultados Obtidos.....	87
6.4	Considerações sobre os resultados obtidos.....	89
7	CONCLUSÃO.....	90
7.1	Trabalhos futuros.....	91
	REFERÊNCIAS.....	92
	APÊNDICE A - DESCRIÇÃO COMPLETA DAS TABELAS DE DIMENSÕES. 98	
	APÊNDICE B - DOCUMENTAÇÃO WEB SERVICE DE GERADOR DE ESTILOS.....	105

1 INTRODUÇÃO

Durante muitos anos, a análise espacial vem sendo instrumento de apoio à tomada de decisão. Um dos primeiros exemplos na história da utilização da análise espacial é o caso do mapa do Dr. Jonh Snow. Em 1854, havia uma grave epidemia de cólera na zona do Soho em Londres. Na época, desconhecia-se a forma de contaminação da doença. Diante disso, Snow relacionou no mapa da cidade a localização dos doentes e dos poços de captação, que nesse período representavam a principal fonte de água para os habitantes da região. Dessa forma, pôde-se identificar a causa da doença de veiculação hídrica até então desconhecida (CÂMARA; MEDEIROS, 1998). Nesse sentido, os mapas ajudam a compreender as transformações e os problemas do mundo atual representando e sintetizando informações históricas, políticas, econômicas, físicas e biológicas de diferentes lugares do mundo, pois permitem associá-los a sua localização.

Utilizando-se de mapas interativos ou de Sistemas de Informação Geográfica (SIG) em plataforma web, a distribuição e a quantidade de informações geográficas aumentam cada vez mais. Uma das características dos SIGs e dos mapas interativos é a visualização de planos de informação, ou seja, a sobreposição de camadas de informações distintas, podendo assim, em tempo real, produzir mapas diferentes. Um exemplo de plano de informação muito comum é o de pontos. Com ele é possível identificar a localização de eventos ou outros objetos geográficos (pontos telefônicos, casas, pontos comerciais, hidrantes, entre outros). Normalmente, o objetivo deste plano de informação é mostrar a distribuição dos dados no mapa (TYNER, 2010). Porém, em muitos casos o número/quantidade de pontos é relevante para a compreensão da informação, bem como a análise da frequência dos objetos e/ou eventos e a sua relação com o meio em que estão inseridos.

No caso, por exemplo, de dados de cheias, incêndios, terremotos, entre outros eventos, além da possibilidade da geração de alta quantidade de pontos (próximos ou não) existe uma outra informação associada, o tempo/período. Tal informação permite analisar a progressão de um determinado evento durante um ou mais períodos. Essa informação associada pode implicar em descobrimento de eventos cíclicos, locais mais propícios a desastres e outras conclusões pertinentes ao domínio.

Nesse sentido, um dos grandes desafios que a comunidade de Informação Geográfica enfrenta é fornecer aos tomadores de decisão ferramentas avançadas, as quais são capazes – semanticamente e visualmente – de integrar aspectos quantitativos, qualitativos e cognitivos de um domínio de interesse. Quando uma grande quantidade de dados está disponível, a síntese de informação e um resultado derivado podem resultar em uma atividade demorada e cara (CHIARA et al., 2011).

A análise Geovisual trabalha com problemas que envolvem o espaço geográfico e vários objetos, eventos, fenômenos e processos. As técnicas visuais de análise são essenciais para lidar com os conjuntos de dados atuais (que se ampliam rapidamente em tamanho e complexidade), para descobrir padrões e para otimizar e orientar processos de tomada de decisões (ANDRIENKO et al., 2011).

Em ambientes *web*, os mapas interativos têm a possibilidade de servir de subsídio, tanto para tomadores de decisão de nível técnico com conhecimento avançado em informações geográficas, assim como de usuários eventuais. Esses usuários, mesmo que eventuais, podem demandar consultas complexas, como por exemplo a agregação de um dado espacial. Cabe destacar que a utilização de softwares renomados de Sistemas de Informação Geográfica como o ArcGIS¹ e o QGIS² precisam de uma sequência de passos e/ou de um processamento mais demorado assim como de instalação de *softwares* ou *plugins* adicionais. Esses *softwares*, por sua vez, não disponibilizam essas funcionalidades de forma *online*, característica que dificulta o acesso a um usuário eventual ou a um gestor que precisa tomar uma decisão. Sendo assim, as técnicas e as funcionalidades que forem adicionadas a um *web map* tornam-se a melhor solução para esses usuários.

1.1 Problema de pesquisa

Tendo em vista os cenários atuais descritos por Andrienko et al., (2011), o elevado volume de dados plotados em um mapa pode gerar visualizações confusas e normalmente de difícil interpretação.

A Figura 1.1 ilustra um exemplo dessa situação retirado do *web map* i3Geo do Ministério do Meio Ambiente (MMA)³ onde pode-se observar o mapa da América do

¹ArcGIS – Software comercial para SIGs - <http://www.arcgis.com/features/features.html>

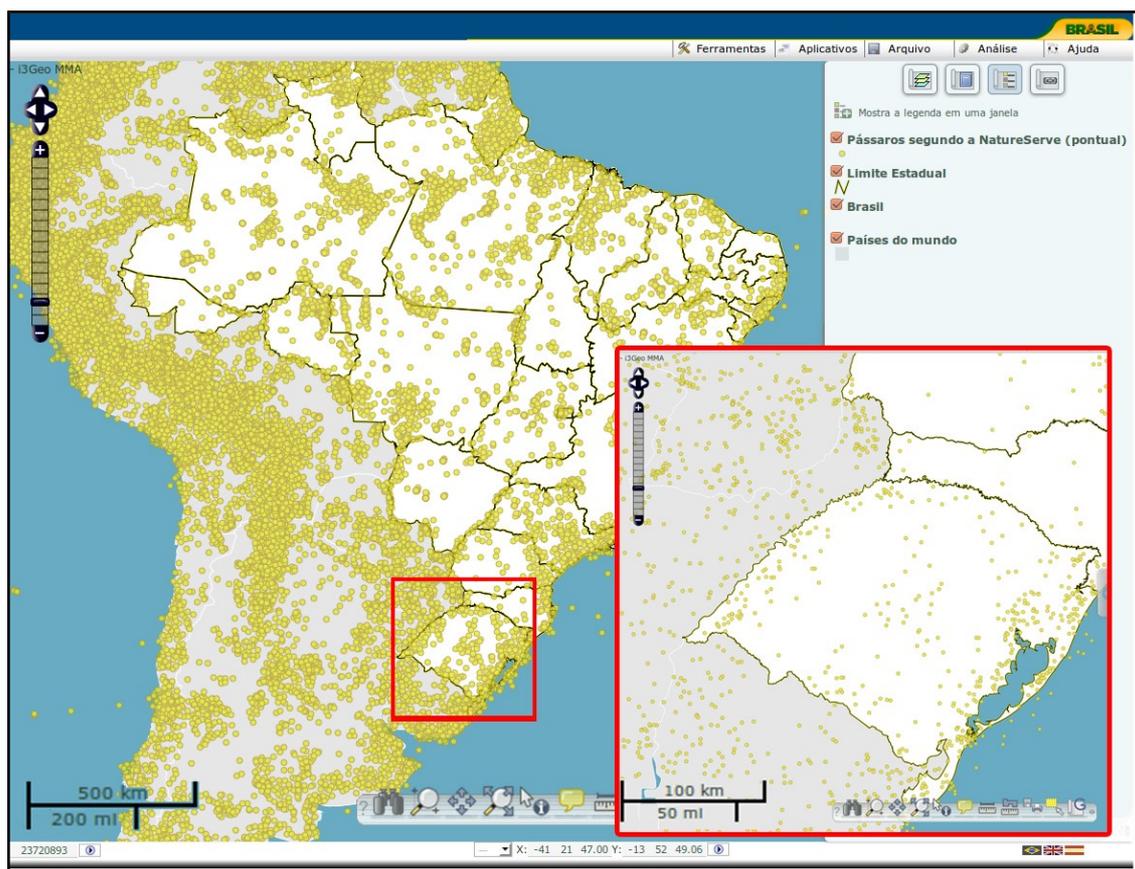
²QGIS – Software livre para SIGs - http://www2.qgis.org/pt_BR/site/

³I3GEO – Ferramenta de *web mapping* utilizado pelo MMA para mapas web disponível em: <http://mapas.mma.gov.br/i3geo>

Sul com os registros pontuais de aves. A área em destaque (bordas vermelha – no mapa com escala 500Km) mostra uma determinada área, onde existe uma grande quantidade de pontos. Isso demonstra que dependendo do nível de detalhamento ou da região de abrangência analisada, torna-se quase impossível visualizar e acessar a dispersão dos pontos. Já com uma escala mais aproximada (mapa com borda em vermelho – escala de 100Km), a visualização de pontos é facilitada.

Nesse sentido, percebe-se que a escala influencia diretamente na visualização da informação. Mesmo que a visualização da localização dos pontos em escala de 2KM possibilite uma apresentação viável de determinados pontos, não é possível analisar, por exemplo, o Brasil como um todo, nem estimar a quantidade. Para conseguir visualizar outras características (dispersão e quantidade por região), é necessário utilizar técnicas de agrupamento ou de frequência para o adequado entendimento das informações.

Figura 1.1 – Registro de aves no *web map* I3Geo do Ministério do Meio Ambiente



Esse problema também ocorre nos demais experimentos realizados nesta dissertação: experimento 1 (Figura 1.2), realizado com os suinocultores da região do COREDE-Serra; experimento 2 (Figura 1.3) realizado com os desastres naturais do Rio Grande do Sul (RS).

Figura 1.2 – Suinocultores da região COREDE-Serra

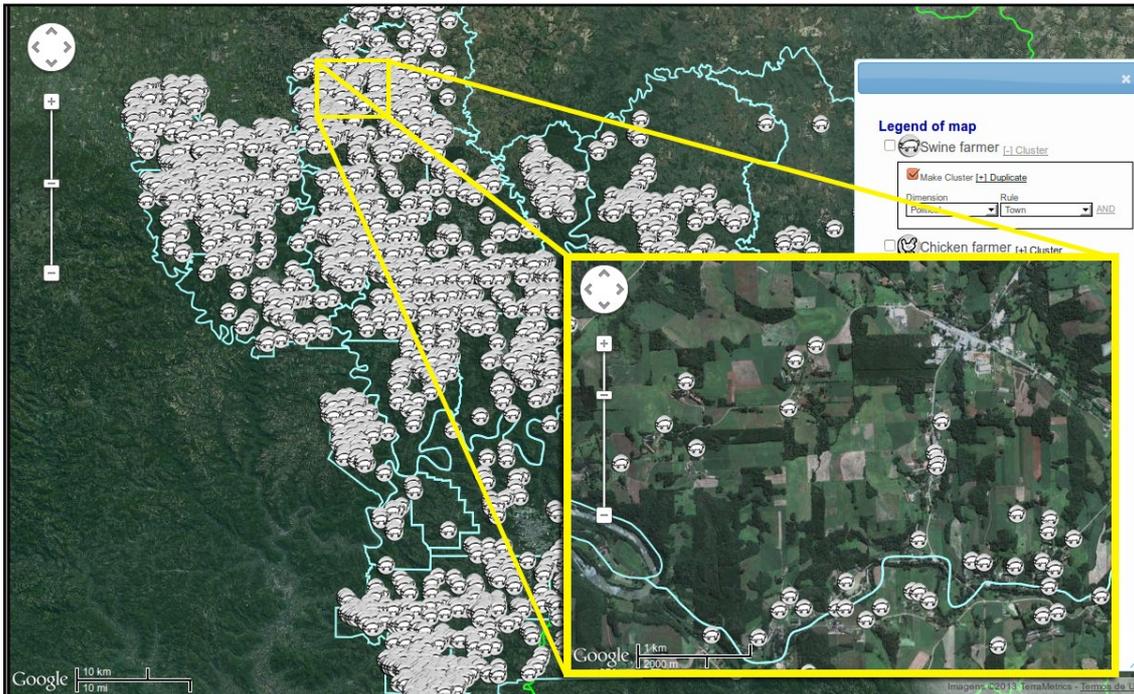
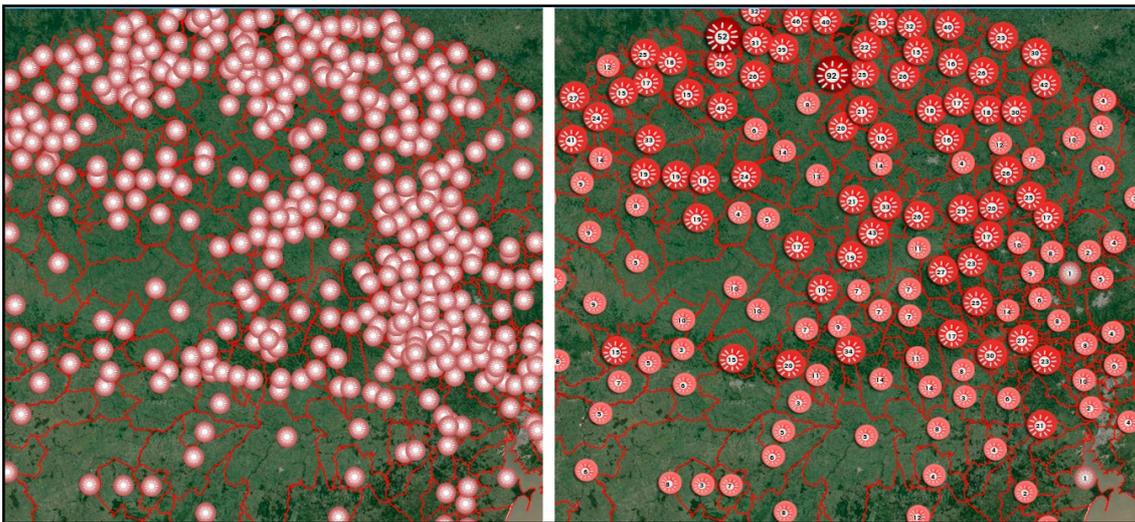


Figura 1.3 – Pontos com registros de estiagens no Rio Grande do Sul no período de 1982 a 2013



Outro problema apresentado no experimento 2 é que além da proximidade dos pontos existe uma terceira variável envolvida: o tempo. Dessa forma, os dados, além de próximos devido à escala (como os problemas anteriores), possui uma sobreposição devido ao fator tempo (repetição do evento aos longos dos anos). Esse problema é evidenciado na Figura 1.3 com o mesmo mapa com os pontos sobrepostos e as quantidades calculadas.

1.2 Objetivos

O objetivo principal deste trabalho é desenvolver uma técnica para a melhor visualização da densidade de informação em mapas *web*, utilizando um modelo multidimensional para a consulta e escolha dos critérios do usuário. Nesse sentido, pretende-se explorar o conceito de modelagem de dados geográficos, permitindo por meio de uma estrutura genérica agrupar os metadados ou metainformações associadas, que sirvam para que possa ser realizada uma análise conforme o contexto da informação.

Os demais objetivos deste trabalho são:

- selecionar e aplicar uma técnica de agrupamento de informações geográficas em um mapa *web*, explorando um algoritmo que não seja oneroso ao sistema e que permita uma melhor visualização de dados espaciais, especificamente pontos;
- explorar agregações e desagregações de múltiplos planos de informação em um mesmo *web map*, possibilitando a apresentação de diversos tipos de informação ao mesmo tempo no mesmo mapa. Como por exemplo, apresentar enchentes e secas ao mesmo tempo em um mesmo mapa;
- explorar agregações distintas para um mesmo plano de informação, o que permite avaliar dois contextos diferentes em um mesmo mapa, como por exemplo, visualizar os eventos de seca por período e por região.

1.3 Estrutura da dissertação

A dissertação está organizada da seguinte forma: o capítulo 2 apresenta um detalhamento dos conceitos fundamentais de *web maps* e modelagem dimensional para este trabalho; o capítulo 3 apresenta trabalhos relacionados que abordam possíveis soluções para o problema de densidade de pontos e abordagem de tomadas de decisões sobre mapas; o capítulo 4 apresenta o modelo dimensional desenvolvido para auxiliar na escolha dos agrupamentos de pontos; o capítulo 5 apresenta a ferramenta APPWM que foi desenvolvida com vistas a validar a aplicação de um modelo genérico; o capítulo 6, por sua vez, apresenta os experimentos com a aplicação e o modelo desenvolvido e; o capítulo 7 apresenta o fechamento do trabalho, assim como as perspectivas futuras e trabalhos em desenvolvimento.

2 CONCEITOS FUNDAMENTAIS

Este capítulo tem como objetivo apresentar as principais definições e conceitos utilizados neste trabalho para a construção do modelo multidimensional (capítulo 4) e de *web maps* (capítulo 5).

2.1 Características da modelagem multidimensional

A modelagem multidimensional consiste na organização das estruturas de informação em fatos e dimensões (Kimball e Ross, 2005). Um fato contém as medidas úteis dos processos de negócios (vendas, entregas, estoques, entre outros), enquanto uma dimensão representa o contexto (quem, onde, quando, etc) (INMON, 2005). Um evento e/ou um objeto geográfico também pode ser representado nesse modelo de fatos e dimensões (RIVEST; MARCHAND, 2001; BEDARD et al., 2007; PROULX; BÉDARD, 2008; SILVA, DA et al., 2010; SILVA; SANTOS, 2011).

O modelo multidimensional apresenta algumas características como não ser normalizado e possuir redundância de dados. Em contraponto, o modelo multidimensional é uma ferramenta amplamente usada para a tomada de decisão e seu modelo é fortemente voltado para a facilidade e velocidade de consultas (KIMBALL; ROSS, 2013). Para suprir a necessidade da qualidade das informações e a consistência da inserção, torna-se necessário, no processo de ETL (*Extraction Transform and Load*), o tratamento dos aspectos de consistência de informação que um modelo normalizado prioriza.

A modelagem multidimensional é amplamente utilizada em *data warehouses*(DW), que consistem para Inmon (2005, p. 25) em “uma coleção de dados orientada por temas, integrado, variante no tempo e não-volátil, que dá apoio às decisões da administração”. Essas características podem ser entendidas da seguinte forma:

- **orientado por tema** – lida com informações estreitamente relacionadas ao negócio da empresa (os fatos de um modelo multidimensional). Por exemplo, em uma empresa atacadista, o tema principal seria vendas;
- **integrado** – armazena informações oriundas de diversos sistemas fontes. Essas informações podem ter a mesma semântica, porém assumem valores distintos

em cada sistema. É papel do DW integrar as informações em um único ambiente, para que possam assumir um valor uniforme;

- **variante no tempo** – os dados armazenados são relativos a determinado período de tempo;
- **não-volátil** – ao contrário dos sistemas OLTPs (*Online Transaction Processing* ou Processamento de Transações em Tempo Real), nos quais os dados antigos são constantemente atualizados e até mesmo excluídos, um DW armazena informações históricas da organização. Por isso, apenas inserções e consultas são permitidas nesse ambiente.

Um DW é, portanto, um banco de dados multidimensional modelado a partir dos conceitos de fatos, medidas, dimensões e hierarquia de dimensões. O fato é o objeto de análise do DW, enquanto as medidas são atributos dos fatos normalmente na forma de variáveis numéricas.

No ambiente de DW são necessárias ferramentas que implementem funcionalidades específicas para lidar com a abordagem multidimensional. Essas ferramentas são denominadas ferramentas OLAP (*Online Analytical Processing*)(Codd et al., 1993).

As ferramentas OLAP possuem operações que permitem aos usuários navegarem e explorarem dimensões e suas hierarquias, viabilizando a análise das informações em vários níveis de detalhamento ou granularidade. Dentre as operações OLAP, duas delas têm relação muito próxima à visualização de escalas diferentes em mapas, as quais são:

- *Roll-up* – realizar essa operação significa ir de um nível mais baixo a um nível mais alto na hierarquia de determinada dimensão. O resultado dessa operação é uma informação mais agregada (ou sumariada).
- *Drill-down* – a operação tem efeito oposto à *Roll-up*, ou seja, o usuário tem uma informação agregada e sobre a qual deseja descobrir detalhes.

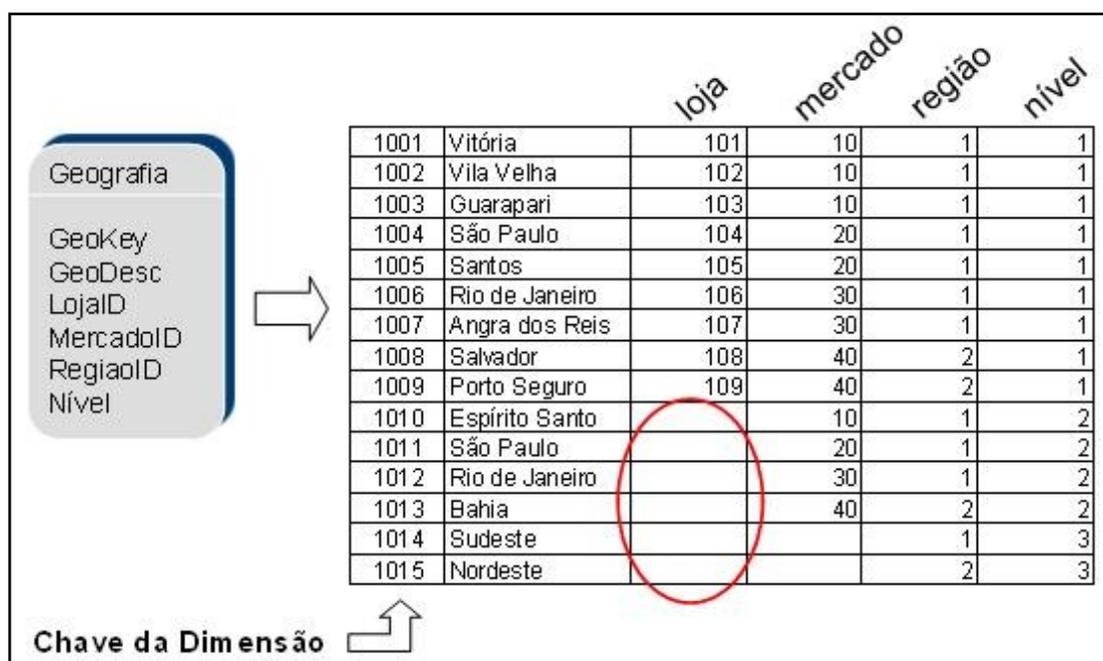
2.2 Informações geográficas

Em bancos de dados, as informações geográficas podem ser pequenas descrições de endereços, nomes de cidades, países, e também podem representar informações mais complexas, como coordenadas, polígonos georreferenciados, entre outras formas geográficas e representações (CÂMARA, 2005).

2.2.1 Hierarquia geográfica

As informações geográficas possuem uma hierarquia intrínseca, como por exemplo um estado localizado dentro de um país, que está contido em um continente. A hierarquia em um modelo dimensional recomenda ser achatada em apenas uma dimensão, isso evita a necessidade de junções típicas em modelos normalizados (ROSS, 2009). Para os casos de achatamento da hierarquia em uma dimensão, para Nardi (2007) recomenda-se utilizar um indicador de nível, possibilitando o controle sobre a granularidade/hierarquia em uma consulta. Essa situação é apresentada na Figura 2.1.

Figura 2.1 – Dimensão achatada com nível representando a hierarquia.



Fonte: Nardi (2007)

O identificador de granularidade para o modelo é necessário, por exemplo, na montagem do menu de seleção das agregações da aplicação desenvolvida. Essa opção permite verificar até que nível hierárquico é possível restringir o agrupamento. Outra utilidade do nível de granularidade no ETL é no caso das dimensões físicas e políticas, pois esse identificador limita até que nível da dimensão o algoritmo do ETL vai procurar, ou seja, verificando se o ponto encontra-se no limite dos polígonos.

Ainda a respeito de hierarquia, as dimensões podem se dividir em implícitas e explícitas. As hierarquias explícitas são caracterizadas por uma sequência de entidades interligadas, nas quais os relacionamentos, entre cada par de entidades é N:1, na sequência. Um exemplo pode ser visto na dimensão política, na qual o município está

subordinado a um estado, que por sua vez está subordinado a um país. Já as hierarquias implícitas representam as hierarquias embutidas nos atributos das dimensões. Um exemplo disso é quando categorizamos um atributo de uma dimensão, no caso, a classe de declividade de um ponto na dimensão clinografia. Normalmente, esse atributo pode vir a se tornar uma dimensão própria (Figura 2.2).

Figura 2.2 – Atributo de dimensão comportando como hierarquia implícita

Transformação de hierarquia implícita em hierarquia explícita	
	
<p>Se analisada a dimensão física hidrografia (dim_fisica_hidrografia), percebe-se que ela possui uma hierarquia implícita em seu campo “bacia_comite”, ou seja, o ponto pode pertencer a um ou a outro comitê de bacia.</p>	<p>Para simplificar a implementação, tanto em termos de construção de interface, como manter o padrão para fazer uma consulta de agregação, as hierarquias implícitas tornam-se uma nova tabela de dimensão.</p>

Como visto na Figura 2.2, uma hierarquia implícita pode ser representada como um atributo de uma dimensão. Há dois problemas de realizar essa representação. O primeiro caracteriza-se pela inviabilidade de ter o controle sobre os valores possíveis desses atributos, uma vez que, para isso, torna-se necessária a normalização das dimensões. O segundo problema ocorre devido ao fato de que, para cada nível da hierarquia explícita da dimensão necessita-se da geração de uma combinação de todos os valores possíveis. No exemplo da Figura 2.2 é preciso realizar a combinação dos (n_1) níveis e as (n_2) comitês. Essa combinação (n_1) x (n_2) pode tanto ser pequena quanto gerar números gigantescos, e para cada novo n_1 ou n_2 deverá ser criada uma nova e grande quantidade de combinações. O caso mais comum para as combinações tenderem ao infinito é a existência de várias hierarquias em uma mesma dimensão, somadas a uma grande quantidade de entradas para um determinado campo. Um exemplo disso é

caso de agrupar em uma mesma dimensão tempo dia e hora: cada dia representado deverá possuir 24 entradas, caso necessite representar os minutos será necessário 1440 entradas para cada dia, e se for representar os segundos serão 86400 entradas para cada dia.

Para o modelo proposto nesta dissertação, as situações de hierarquia implícitas foram subdivididas em duas ou mais dimensões, sendo que o nome delas foi padronizado da seguinte forma: nome da dimensão inicial acrescido de um sufixo que caracteriza a hierarquia, conforme apresentado na Figura 2.2.

2.3 Componentes de um *Web Map*

O termo *web map* ou *web mapping* é o processo de concepção, execução, geração e entrega de mapas na *World Wide Web*, o qual lida principalmente com questões tecnológicas. Enquanto isso, a cartografia *web* estuda adicionalmente aspectos teóricos: o uso de *web maps*, a avaliação e otimização de técnicas e fluxos de trabalho, a possibilidade de utilização, aspectos sociais e muito mais. Já o WebGIS ou InternetGIS está relacionado ao mapeamento web, mas com ênfase em análise, processamento de dados geográficos específicos do projeto, bem como aspectos exploratórios. Muitas vezes, o *WebGIS* e *Web Map* são usados como sinônimos, mesmo que eles não signifiquem a mesma coisa. Na verdade, a fronteira entre Web maps e webGIS é embaçada. *Web mapping* são, muitas vezes, um meio de apresentação dos WebGIS e os *web maps* estão ganhando cada vez mais capacidades analíticas (NEUMANN, 2008).

Para este trabalho, utilizam-se ferramentas típicas de *web maps* como a biblioteca OpenLayers que, segundo a definição em (OPENLAYERS, 2013) “é uma biblioteca *Open Source Java Script*, para carregar, mostrar e renderizar mapas com a utilização de múltiplas fontes da *web*”.

A biblioteca OpenLayers, assim como outras bibliotecas de *web maps*, como a GoogleMaps Api⁴, fornecem funcionalidades para alterar o *zoom* (aproximar ou afastar o mapa, que em cartografia significa modificar a escala). Outra funcionalidade típica dessas bibliotecas é a incorporação de dados sobre o mapa montado em tempo real, tornando-o um ambiente de exploração cartográfico dinâmico e de ampla utilização na *web* na atualidade.

⁴API do Google Maps disponível em: <https://developers.google.com/maps/?hl=pt-br>

3 TRABALHOS RELACIONADOS

Como método na busca das soluções para o problema elencado neste trabalho, foram realizadas duas revisões bibliográficas. A primeira revisão teve como intuito o levantamento das técnicas para o agrupamento de informações. Os principais trabalhos encontrados referentes às técnicas são apresentados na seção 3.1.

Em um segundo momento, após definido o sistema de visualização mais adequado para o tipo de informação levantada, pesquisaram-se trabalhos que envolvem a manipulação das informações geográficas para a tomada de decisão, os quais são apresentados na seção 3.2 os sistemas de informação geográfica e na seção 3.3 os sistemas SOLAPs, respectivamente.

Para encerrar o levantamento dos trabalhos relacionados, foi elencado o estado da arte quanto a algoritmos de agrupamentos (seção 3.4) e foram discutidos quais os requisitos que a proposta iria abordar (seção 3.5). Por último, foi feita uma comparação dos trabalhos com os objetivos a serem atingidos por este trabalho (seção 3.6).

3.1 Visualização de dados geográficos agregados

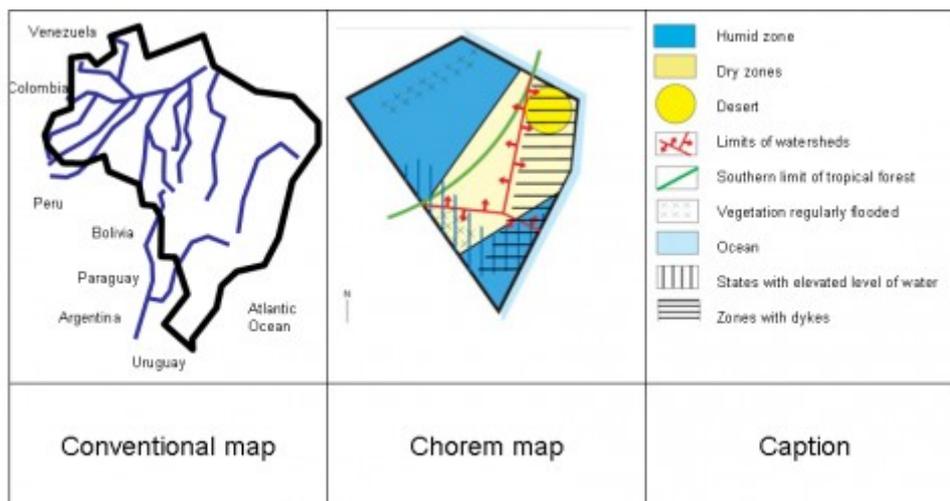
Há complexidade no meio ambiente que nos cerca, caracterizando pela riqueza em detalhes. Os mapas trabalham estrategicamente diminuindo essa quantidade de detalhes, agrupando fenômenos e cortando o que é supérfluo. Somente dessa forma pode-se transmitir as informações que se quer .

Os processos geográficos são dependentes de escala. A mudança da percepção do espaço geográfico está relacionada com a mudança de escala. Como consequência, ao variar as escalas, os mesmos objetos podem ser representados de maneira diferente (BORGES, 2002). Ao construir mapas tanto estáticos quanto dinâmicos, os objetos geográficos possuem, além das informações que representam (área de um município, trecho de um rio, localização de um capital), uma geometria. Na generalização cartográfica, a geometria do objeto pode ser simplificada ou alterada, e um novo objeto pode passar a ser síntese de um conjunto de objetos, alguns dos quais podem ser preservados enquanto outros são eliminados.

No contexto de eliminação de detalhes, é possível ver a técnica de *Chorem* que foi introduzida inicialmente pelo geógrafo francês Brunet como uma representação esquemática de território que elimina os detalhes que não são úteis para a compreensão

de um mapa (CHIARA et al., 2011). O *Chorem* simplifica limites e escalas, a fim de apresentar apenas um conjunto de informações abstratas (LAURINI et al., 2009), como mostra a Figura 3.1. Dessa forma, o tomador de decisão foca-se na informação e não em detalhes mais complexos. O *Chorem*, portanto, é uma representação esquematizada do território, que elimina qualquer detalhe não necessário para a compreensão do mapa.

Figura 3.1 – Exemplo de Chorem Mapa



Fonte: Chiara et al. (2011).

Brunet (1986), afirma que apenas sete formas são suficientes para descrever os modelos que representam os *chorems*:

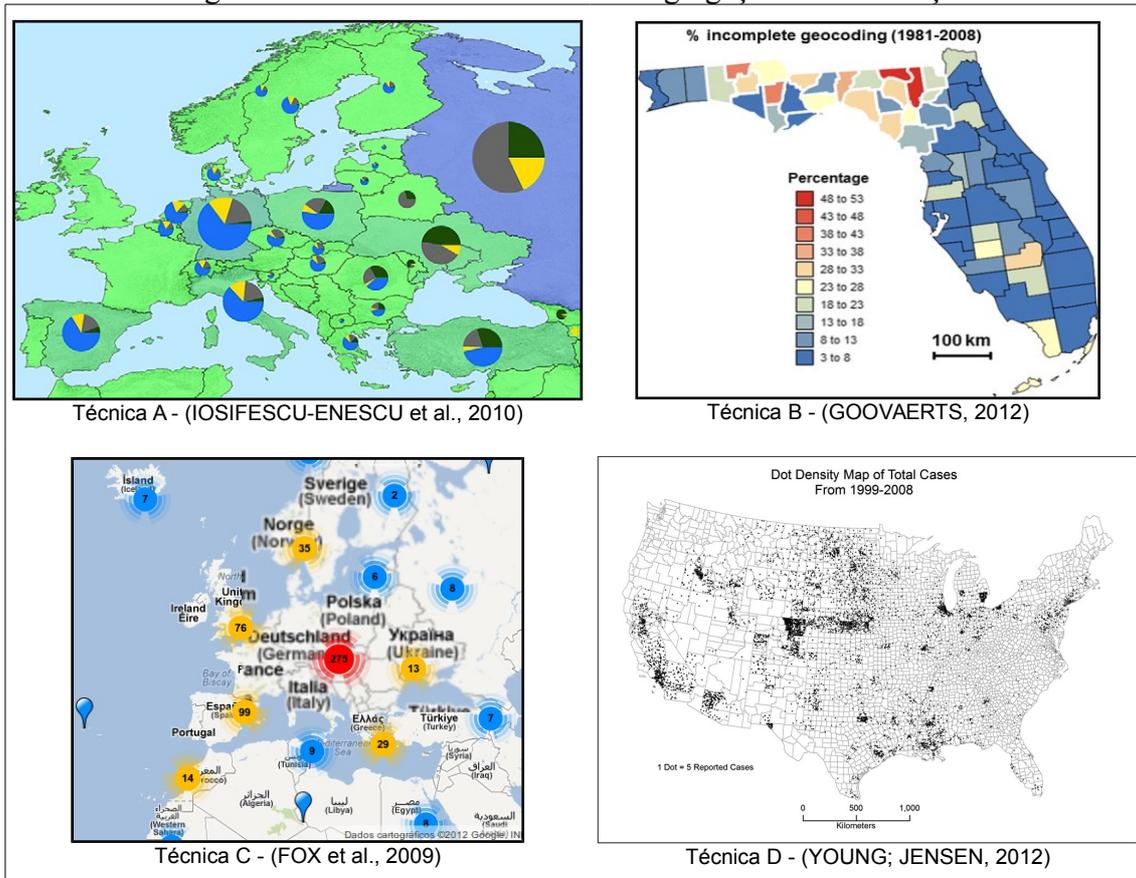
1. *point* (ponto), a intenção de representar lugares e pontos;
2. *line* (linha), a intenção de representar caminhos, quebras, fronteiras e relações;
3. *area* (área), a intenção de representar extensões territoriais e formas;
4. *flow* (fluxo), a intenção de representar a dinâmica, os movimentos, a simetria, e a intensidade;
5. *passage* (passagem), a intenção de representar pontes, passagens, túneis e bifurcações;
6. *polarization* (polarização), a intenção de representar foco e dinâmica;
7. *gradient* (gradiente), a intenção de representar a dissimetria, a atração e a repulsão.

Portanto, a simplificação ou generalização pode ser uma forma de auxiliar na tomada de decisão e na hora de interpretar informações de um mapa. Essas simplificações podem ser usadas para gerar algoritmos de agregações de outras formas geográficas diferentes de pontos.

Ainda em se tratando de agregação, existem diversas técnicas para o agrupamento das informações em mapas, as quais estão listadas a seguir, e ilustradas na Figura 3.2:

- a técnica A, Símbolos Graduados, consiste em desenhar as informações com símbolos ou círculos graduados. Os valores quantitativos são agrupados em classes. Dentro de uma classe, todos os recursos são desenhados com o mesmo símbolo. Um exemplo disso pode ser visto nos trabalhos de (IOSIFESCU-ENESCU et al., 2010; AUER et al., 2011; MACEACHREN et al., 2011);
- a técnica B, Mapas Coropléticos, consiste em uma forma de representar grande quantidade de dados usando a abstração de cores. Sendo assim, a representação de quantidades em um mapa é construída variando as cores (MACIEJEWSKI et al., 2011; GOOVAERTS, 2012) ;
- a técnica C, *cluster* ou agrupamento, consiste em agrupar os pontos próximos conforme o *zoom* é modificado no mapa, como a biblioteca MarkerCluster (FOX et al., 2009);
- a técnica D, *Dot Density Maps*, consiste em agrupar as informações usando uma relação simples: 1 ponto plotado no mapa equivale a 5 pontos e a diferença em relação à técnica de *cluster* é que essa relação se mantém estática por todo o mapa.

Figura 3.2 – Mosaico de técnicas de agregação de informação



Dentre os trabalhos estudados na busca por artigos que tratassem de geovisualização e densidade de informação, a Tabela 3.1 apresenta os principais trabalhos estudados e as técnicas por eles utilizadas.

Tabela 3.1 – Comparação entre as técnicas de visualização de informação

	<i>Símbolos graduados</i>	<i>Mapas coropléticos</i>	<i>Dot Density Maps</i>	<i>Agrupamento</i>
Silva e Santos (2012)				X
Goovaerts (2012)		X		
Young e Jensen (2012)			X	
Auer et al. (2011)	X	X		
Krygier e Wood (2011).		X		
Maciejewski et al. (2011)	X	X		
Fox et al. (2009)				X
Maciejewski et al. (2008)		X		

No trabalho de Auer et al. (2011), é abordado a solução de mapas coropléticos, além de mapas de símbolos graduados. Os autores elencam 3 problemas para esse tipo de solução de símbolos graduados que afetam principalmente *web maps*:

1. definição de um algoritmo para traduzir os valores em tamanhos ou símbolos;

2. tamanho padrão do símbolo, ou seja, quão grande e quão pequeno deve ser o símbolo em função da quantidade de faixas de frequência;

3. valores extremos.

O *web mapping* permite uma interação mais rica que mapas estáticos. A técnica de círculos/símbolos graduados, por exemplo, apresenta o seguinte problema: a cada nível de *zoom* os símbolos graduados devem ser refeitos. Ao mudar a relação tamanho quantidade poderá confundir o usuário do mapa que vai ter uma nova legenda. No caso dos símbolos, não sofrerem alteração em *zooms* mais próximos as frequências altas deixariam de existir tornando a análise complexa, por existir uma sobrecarga de símbolos não utilizados.

As soluções para a visualização de informação em mapas computadorizados iterativos ou web mapas normalmente utilizam as técnicas de cartografias de mapas estáticos. As técnicas aprimoradas para mapas iterativos advém de SIGs que são apresentados na seção 3.2.

3.2 Sistemas de Informação Geográficos

Os primeiros sistemas de informação geográficos, computadorizados datam de 1962 (ESRI, 2012). Portanto sistemas de informação são bastante difundidos e utilizados para a produção de mapas computadorizados.

Conforme proposto por (DAVIS; NETO, 2001), pode-se indicar que numa visão abrangente um SIG tem os seguintes componentes:

- Interface com usuário – define a maneira pela qual o sistema interage e se comunica com o usuário;
- Entrada e Integração de Dados – o SIG deve prover mecanismos de importação e/ou integração entre dados coletados e cadastrados por instituições;
- Consulta e Análise Espacial – interage com o módulo de gerência de dados espaciais para extrair as informações desejadas pelo usuário;
- Visualização e Plotagem – permitir análises interativas, visualizar resultados de consultas e manipulações, gerar relatórios e mapas;
- Gerência de Dados Espaciais – armazenamento e recuperação de dados organizados sob a forma de um banco de dados geográficos.

3.3 Modelos OLAP e SOLAP

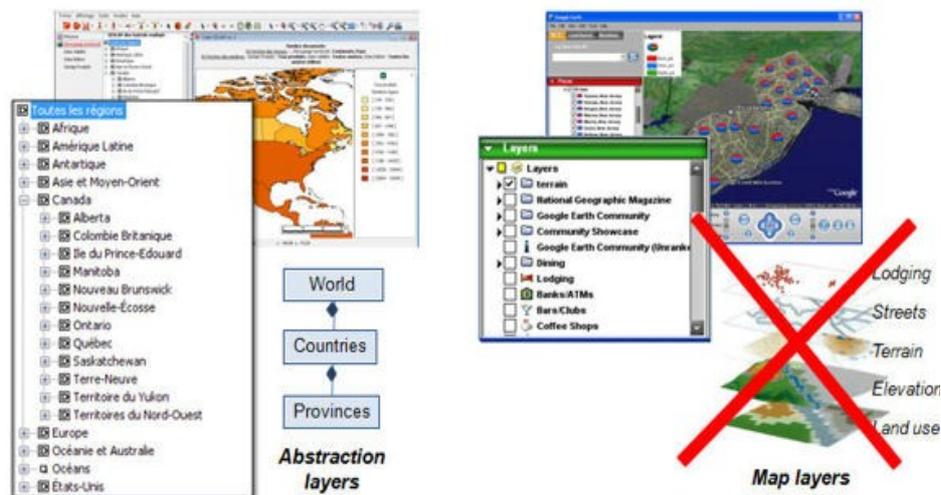
A junção de ferramentas de análise de dados com a criação de mapas gera resultados de maior relevância. Entretanto, quando se trabalha com informações pontuais e nas ocasiões em que o número de informações é denso, a visualização pode se tornar confusa (MACIEJEWSKI et al., 2008). Percebe-se, portanto, a necessidade de ferramentas que auxiliem a tomada de decisão para não-técnicos em processamento de informação geográfica. Por outro lado, a tomada de decisão utilizando-se de modelos OLAP (*Online Analytical Process*) é amplamente difundida e possui diversas ferramentas (comerciais e inclusive gratuitas como o Mondrian (PENTAHO, 2012)) implementadas para a esse fim.

Para análise de informações geográficas, pode-se verificar a existência de ferramentas denominadas *Spatial OLAP* (SOLAP), que consistem, segundo Bedard et al. (2007), em um sistema que inclui:

- I. visualização de dados - engloba as seguintes formas: mapa, tabelas e gráficos. Possibilita a construção de mapas temáticos adequados, a visualização de informação contextual e legenda interativa;
- II. exploração de dados - inclui a navegação multidimensional em qualquer forma de visualização de dados, métricas calculadas e filtragem pelos atributos das dimensões; e
- III. estrutura dos dados - abrange o suporte de diferentes formas geométricas e múltiplas representações para diferentes escalas.

O modelo clássico de construção de mapas interativos utiliza a sobreposição de planos de informação. Essa abordagem permite uma rápida construção, porém, não é possível obter informações de um objeto geográfico a partir de outro. A Figura 3.3 mostra, do lado esquerdo, um exemplo de informação agregada e, do lado direito, outra separada (PROULX; BÉDARD, 2008).

Figura 3.3 – Implementação clássica x implementação SOLAP



Fonte: Bedard et al. (2007)

Embora pesquisas em DW espaciais sejam pioneiros desde 1997 (BEDARD et al., 2007), quando o termo SOLAP foi cunhado, só nos últimos anos tem sido investigado a integração de GIS e tecnologias OLAP. Os principais temas que têm sido investigados são as arquiteturas, os modelos de dados, e os operadores e algoritmos necessários para expressar e otimizar as consultas SOLAP (GOLFARELLI et al., 2013).

3.4 Algoritmos de agrupamento

O processo de *clustering* visa agrupar as informações de forma que: (i) exista uma forte similaridade entre os elementos pertencentes ao mesmo grupo; (ii) exista uma fraca similaridade dos elementos pertencentes a grupos diferentes (ZAIANE et al., 2002). Os algoritmos de agrupamento podem agrupar, por proximidade ou por alguma outra característica, os pontos ou informações em um mapa. Tendo como objetivo construir essa similaridade através de atributos dos próprios dados, é proposto um modelo multidimensional que visa, além de agrupar por características, aproveitar a possibilidade de hierarquia intrínseca a dados geográficos (RIVEST; MARCHAND, 2001), gerando uma análise mais exploratória em um *web map*.

Silva e Santos (2011) e Silva et al. (2012) propõem o agrupamento de dados para densidade espacial utilizando o algoritmo DBScan. Uma nova representação para cada *cluster* deve ser gerada, diminuindo o número de objetos espaciais que precisam ser colocados no mapa (SILVA; SANTOS, 2011). Caso haja uma visualização de dados não espaciais, os mesmos devem também ser agregados mantendo a sincronização entre

o mapa e uma visualização tabular (RIVEST; MARCHAND, 2001; BEDARD et al., 2007; SILVA; SANTOS, 2011).

Em se tratando de *web maps* percebe-se a tendência da utilização de algoritmos do tipo grade (MAHE; BROADFOOT, 2010; HABIBULLAYEVICH et al., 2013). Esses algoritmos criam uma grade sobre o mapa e a partir de cada célula desta grade é gerada um grupo. O algoritmo normalmente agrupa marcadores de acordo com a sua distância a um centro de um grupo. Quando o marcador é adicionado, este pesquisa a sua posição em todos os grupos. Caso não seja colocado, em nenhum grupo será criado um novo grupo com o marcador.

A geração de grupos com algoritmos do tipo grade apresenta problemas na qualidade dos grupos gerados em contrapartida apresentam uma performance muito alta. Portanto, para o caso de geração de grupos constantemente (alteração do *zoom*) a alternativa é considerada uma das mais viáveis (MAHE; BROADFOOT, 2010).

3.5 Requisitos do problema

Os requisitos elencados, tendo como base a necessidade verificada nos protótipos apresentados no capítulo 1 (Figura 1.2 e Figura 1.3), assim como os requisitos já elencados em trabalhos como o de Auer et al,(2011), voltados para a geração de um mapa na *web* com grande densidade de informação são: mostrar densidade aparente da informação (em uma primeira visão, obter onde há mais ou menos informação); mostrar a densidade exata da informação (apresentar, quantos eventos estão agrupados); permitir visualizar granularidades diferentes (ver grupos conforme a escala do mapa); e manter a mesma abstração em escala diferentes (para diminuir a carga cognitiva, ou seja, manter a mesma forma de analisar o mapa independentemente do *zoom*, por exemplo, um círculo graduado que em um determinado nível de *zoom* represente uma quantidade e em outro nível um círculo de mesmo diâmetro represente quantidade diferente). Esses quatro requisitos elencados foram tabulados comparativamente na Tabela 3.2.

Tabela 3.2 – Tabela resumo dos recursos disponíveis em cada técnica.

<i>Requisitos</i>	<i>Símbolos graduados</i>	<i>Mapas coropléticos</i>	<i>Dot Density Maps</i>	<i>Agrupamento</i>
Mostrar a densidade aparente de informação	✓	✓	✓	✓
Mostrar a densidade exata da informação	X	X	✓	✓
Permitir visualizar mais que um plano de informação	✓	X	X	✓
Manter mesma abstração em escala/ <i>zoom</i> diferente	X	✓*	✓*	✓

* Não recomendado

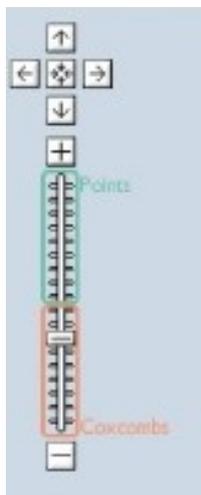
Para atender os requisitos elencados na Tabela 3.2 assim como os requisitos de permitir uma análise de acordo com os critérios do usuário, é necessário adicionar uma estrutura de dados intrinsecamente ligada ao modelo hierárquico. Nesse sentido o modelo proposto no capítulo 4, visa resolver esse problema e para comprovar a aplicação do modelo foi proposto a criação da ferramenta APPWM (Agrupamento Personalizado de Pontos em *Web Maps*). A comparação das funcionalidades que a ferramenta propõe são apresentadas na seção 3.6.

3.6 Comparação com outras ferramentas

Independente da solução SOLAP com ou sem SIG, os sistemas de tomada de decisão em mapas tendem a migrar para web devido a flexibilidade e velocidade com que os dados podem ser consumidos. Na busca de trabalhos que permitissem a análise espacial e a integração com modelos multidimensionais, e agrupamento de ponto alguns trabalhos recebem destaque:

- HerbariaVIZ - (AUER et al., 2011) apresenta formas de visualização diferenciada para diferentes níveis de *zoom*. A grande quantidade de dados pode ser percebida pelo tamanho da fatia do *coxcomb* (Gráfico estatístico (NIGHTINGALE, 1859; COHEN, 1984)). Porém ao se aproximar essas frequências diminuem e a visualização pode ser por pontos. A Figura 3.4 apresenta o detalhe da ferramenta de *zoom* do HerbariaViz.

Figura 3.4 – Detalhe da ferramenta de *zoom* nível com *coxcombs* e nível de pontos



Fonte: Recorte de Auer et al. (2011)

- GeoWOLAP é uma ferramenta SOLAP voltada para *web* que implementa o modelo de dados multidimensional GeoCube(BIMONTE et al., 2006), esse modelo por possuir uma relação N:M entre os fatos e dimensões necessita a utilização de funções *ad-hoc* definidas pelo usuário para a realização de agregação nos fatos.
- GeoMDQL (SILVA, DA et al., 2010) é uma variação da linguagem de consulta multidimensional MDX, estendida para prover novos operadores que lidam com dados espaciais. A sua GUI (*Graphical User Interface*) foi desenvolvida para ambiente *desktop* e o trabalho comenta sobre uma interface *web*, porém ambas necessitam da utilização da linguagem. Apesar da linguagem MDX ser simples linguagens normalmente são voltadas para usuários mais avançados do que para usuários esporádicos.
- Mapwarehouse (CARDOSO, 2009) ferramenta SOLAP, desenvolvida com um conceito de *framework* que permite que usuários finais definam seus esquemas de *Data Warehouses* (DW) Espacial e realizem consultas aos DW Espaciais criados a partir destes esquemas. Para isso, a ferramenta dispõe de linguagens gráficas para definição de esquemas e consultas, implementadas na forma de um conjunto de interfaces gráficas de alto nível de abstração.
- SOLAP+ (SILVA; SANTOS, 2012) é um sistema genérico que depende uma arquitetura de três níveis composto pelo SGBD Oracle (suporte a dados espaciais), um servidor SOLAP. Nessa camada são aplicados os algoritmos de

agrupamento o DBSCAN e P-DBSCAN, e um cliente apresenta o acoplamento do OLAP com mapas.

A comparação, apresentada na Tabela 3.3, é realizada com base nos trabalhos citados apresentados em relação ao APPWM. Os requisitos foram elencados a partir de: nos métodos de visualização que foram elencados na seção 3.5; e os objetivos deste trabalho.

Tabela 3.3 – Comparação entre APPWM e outras ferramentas

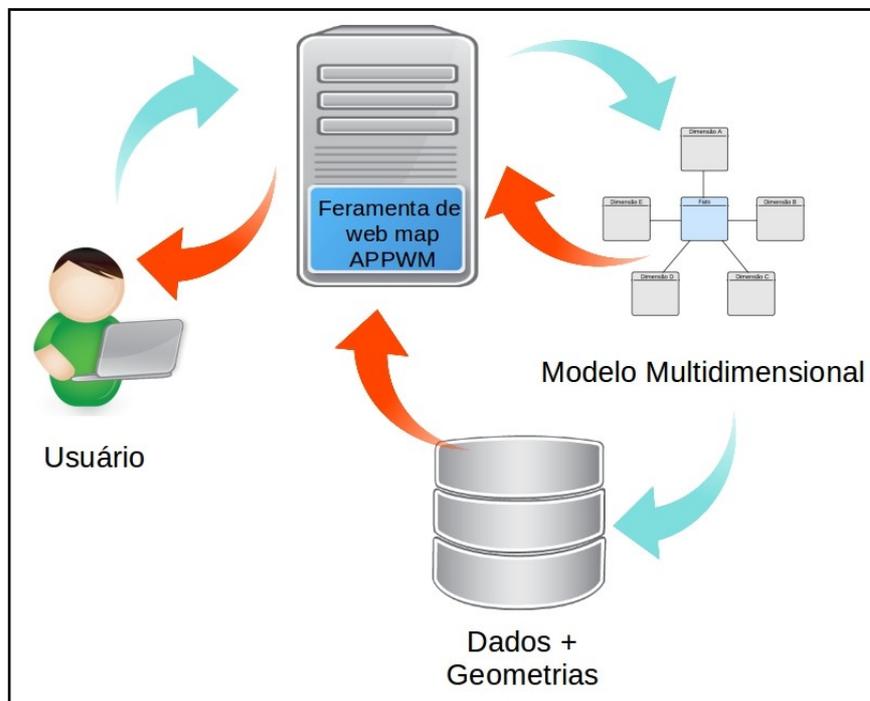
<i>Requisitos</i>	<i>HerbariaVIZ</i>	<i>Geo WOLAP</i>	<i>Geo MDQL</i>	<i>Map warehouse</i>	<i>SOLAP+</i>	<i>APPWM</i>
Interação com um modelo multidimensional		✓✓	✓✓	✓✓	✓✓	✓✓
Permitir agrupamento de pontos com a técnica de agrupamento	✓	SF	✓	✓	✓✓	✓✓
Permitir visualizar mais que um plano de informação agrupado		SF	✓✓	✓	✓✓	✓✓
Plataforma Web	✓✓	✓✓		✓✓	✓✓	✓✓
Integração com mapas externos			SF	SF	SF	✓✓
Geração de mapa em tempo Real	✓✓		✓✓	✓✓	✓✓	✓✓
Processamento de agrupamento no cliente			✓✓			✓✓
Independência de softwares proprietários	✓	✓	✓✓	✓✓	✓	✓✓
Legenda: ✓ (um ✓) atende parcialmente ✓✓ (dois ✓) atende completamente SF Sem informação para afirmar						

Neste contexto a ferramenta desenvolvida nesse trabalho visa suprir as necessidades apresentadas pelas demais. O foco principal do APPWM é servir de subsidio para o modelo multidimensional, porém das características elencadas é importante ressaltar a geração do mapa em tempo real, além de o processamento ser dado no cliente. Essa última característica tem grande impacto na velocidade de tomada de decisão ao alterar o *zoom*, este permite que não seja solicitados novos grupos para o servidor e simplesmente com base nas informações já contidas os dados sejam reagrupados.

4 MODELO MULTIDIMENSIONAL PARA AGRUPAMENTO DE DADOS

O objetivo deste capítulo é apresentar o modelo multidimensional proposto para especificar os critérios de agrupamento do usuário. O modelo proposto foi avaliado como o componente principal e fundamental na proposta de agrupamento das informações do *web map*. Para compreender esse processo, é apresentado, na Figura 4.1, o esquema de funcionamento do modelo quando embutido na ferramenta implementada. O usuário acessa a hierarquia proposta no modelo através da ferramenta APPWM (Agrupamento Personalizado de Pontos em *Web Maps*), apresentada em detalhes no capítulo 5. A ferramenta acessa e utiliza o modelo multidimensional, implementado em esquema estrela, realizando as consultas, recuperando e retornando os dados e as geometrias que possui armazenado no banco de dados. Com os dados necessários, a ferramenta APPWM agrupa e permite ao usuário uma exploração voltada para o seu contexto.

Figura 4.1 – Esquema de utilização do modelo multidimensional



Este capítulo está dividido em 5 grandes seções. A seção 4.1 apresenta a justificativa da escolha do modelo multidimensional e quais as características motivaram a utilização dessa modelagem. A seção 4.2 apresenta o modelo multidimensional completo e as dimensões trabalhadas nos experimentos. A seção 4.3

propõe a expansão do modelo através de dimensões personalizadas, para que este venha a ser utilizado com outros domínios e necessidades diferentes. A seção 4.4 apresenta algumas limitações e as soluções que devem ser implementadas em uma aplicação. A seção 4.5 conclui o capítulo pontuando as características mais importantes do modelo e como ele pode ser utilizado.

4.1 Motivação

A proposta de construir um *web map* utilizando o modelo multidimensional possibilita a análise de grande quantidade de dados, de forma simples e respeitando o contexto em diversos domínios específicos. O modelo multidimensional, por ser voltado prioritariamente para a consulta, apresenta um bom desempenho na tarefa de geração de agrupamentos. Os *web maps* necessitam de uma velocidade maior e uma menor quantidade de transferência de dados que os seus semelhantes em ambientes *desktop*. Por esse motivo a união entre um modelo multidimensional e um *web map* torna-se uma alternativa viável de exploração.

A necessidade de um modelo para servir de base na implementação dos critérios de seleção dos dados, no agrupamento deles em um mapa, levou à definição de um modelo genérico que pudesse ser usado para implementação em qualquer aplicação com pontos em um *web map*. Porém, a construção de um modelo genérico pode deixar o sistema “engessado”, ou seja, pode não suprir a necessidade do tomador de decisão e, para isso, o modelo não pretende ser definitivo, nem completo, mas sim, extensível. O modelo pretende ser um modelo base para que outros usuários possam melhorá-lo e personalizá-lo de acordo com a necessidade de cada domínio. Sendo assim, o modelo somado à metodologia de visualização permitirá uma maior facilidade na tomada de decisão e interpretação e organização de dados geográficos densos em diferentes escalas.

Os *web maps* dinâmicos podem gerir informações que vêm de diversos servidores e, nessa etapa, é necessário consolidar os dados em um banco de dados histórico e/ou em um servidor para a tomada de decisão, papel clássico dos *data warehouses*. Na literatura, como visto no capítulo 3, existem diversas propostas, ferramentas e linguagens para a utilização de modelos dimensionais e/ou OLAP com a produção de mapas (*online* ou não). O modelo, portanto, tem como meta ser uma base rápida para dar início à implementação. Além disso, ele pode agir como um norteador

para uma tomada de decisão em diferentes níveis de *zoom* e de agregação de informação geográfica.

O modelo serve, então, de base para a rápida formação de grupos de dados de acordo com o critério e o contexto do usuário de forma mais intuitiva e sem a necessidade de grande processamento de dados. Essas características são fundamentais a uma ferramenta apropriada para um *web map*.

As dimensões e as suas propriedades são a compilação das necessidades evidenciadas em ambos os experimentos. Nesse aspecto, tanto as dimensões quanto as propriedades não foram buscadas de forma exaustiva. Mesmo assim o modelo apresentado foi organizado e projetado visando ser genérico e extensível. Para a extensão do modelo, pode-se usar diversas classificações. Um ponto de partida para obter possíveis classificações é o CONCLA⁵ (Comissão Nacional de Classificação), que disponibiliza as classificações estatísticas nacionais, para alguns temas selecionados, usadas no sistema estatístico e nos cadastros administrativos do Brasil. O CONCLA também apresenta em seu site as classificações internacionais que serviram de base para a classificação brasileira.

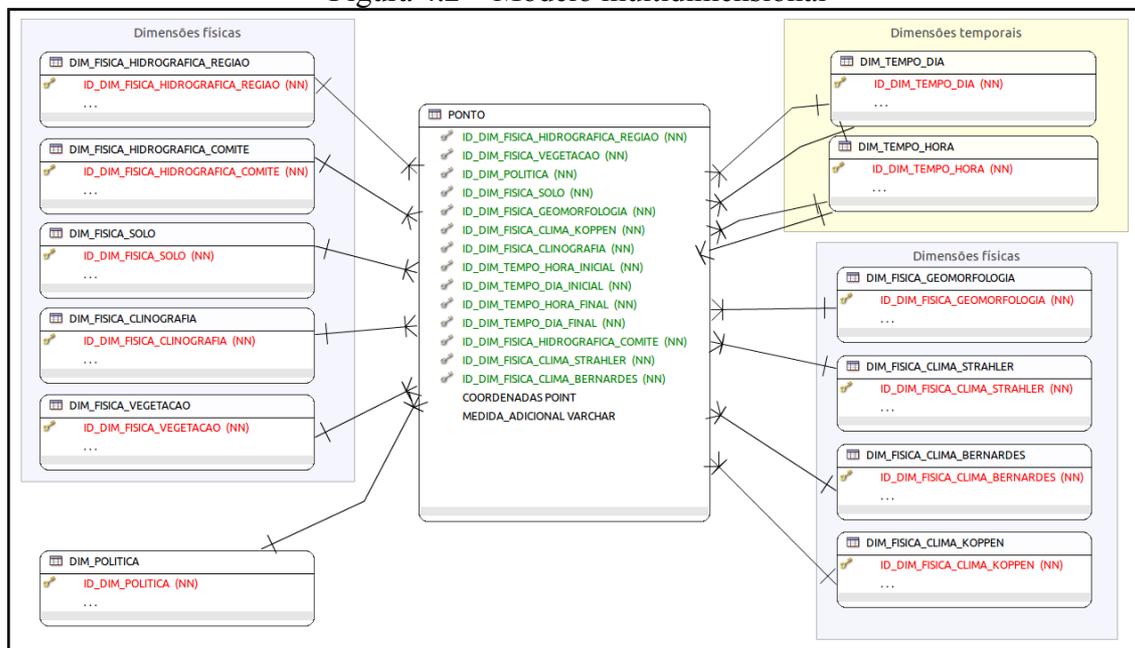
4.2 Modelagem proposta

A modelagem multidimensional pode ser implementada sobre bancos de dados relacionais de acordo com diferentes esquemas. O esquema utilizada no modelo é o *star schema* (KIMBALL; ROSS, 2013). O modelo proposto possui 4 dimensões bases (essa primeira divisão tem o intuito de organizar o modelo), a fim otimizar a geração dos grupos: a dimensão “dim_política” – auxilia na organização dos grupos por divisões políticas, como municípios, estados e países apresentada na seção 4.2.1; as dimensões físicas (prefixo das tabelas – dim_física) – auxiliam na organização dos grupos por meio de características físicas comuns, como bacia hidrográfica, região hidrográfica, formação vegetal, tipo de solo, entre outras, apresentadas na seção 4.2.2; as dimensões socioeconômicas – agrupam os dados por características econômicas e sociais apresentadas na seção 4.2.3; e a dimensão “dim_tempo” – que armazena características temporais referentes ao ponto, como estado ou validade da informação apresentada na seção 4.2.4.

⁵ <http://concla.ibge.gov.br> – Site da comissão nacional de classificação

Essas dimensões são a base para um agrupamento mais organizado, levando em consideração o contexto das aplicações. A Figura 4.2 mostra a relação dessas dimensões com o fato Ponto, as mesmas são descritas na sequencia.

Figura 4.2 – Modelo multidimensional



As dimensões apresentadas são classificadas em “obrigatoriamente existente” e “não obrigatoriamente existente”. As dimensões obrigatoriamente existente são dimensões que vão existir indiferente da utilização; por exemplo, o clima, indiferente da localização de um ponto, vai ter uma relação com alguma classe climática independente da localização no globo. Por sua vez, as dimensões não obrigatoriamente existente ocorrem quando um ponto está localizado em um local que não possui uma relação com a dimensão; por exemplo um ponto a partir do momento que estiver localizado no oceano não terá uma relação com algum município e/ou tipo de vegetação e relevo.

Um aspecto relevante de um modelo multidimensional é o fato de, em suas dimensões, armazenar, além dos dados relevantes para agregações e restrições em consultas, apelidos para rótulos de tabelas e relatórios e descrições de dados. É comum existir dimensões com diversas colunas, facilmente chegando a 50 delas (KIMBALL; ROSS, 2005). Para a modelagem, nessa sessão foram apresentadas as colunas mais significativas, e com a hierarquia achatada, sendo que a definição completa de cada dimensão e suas respectivas colunas encontram-se no apêndice A - Descrição completa das tabelas de dimensões.

4.2.1 Dimensão política

O contexto político em um mapa apresenta os limites propostos pelo homem para delimitar uma unidade administrativa. Essa unidade é determinado por fatores geográficos ou políticos. Essas unidades subdividem-se em unidades menores e possuem uma hierarquia pré-determinada.

A dimensão política começa em seu mais alto nível: o continente, quando se trata de um ponto em terra. Para pontos que estão em alto mar (região do mar sem direitos de nenhum país), as divisões políticas tornam-se desnecessárias.

A região considerada alto mar é regida pela *United Nations Convention on the Law of the Sea (UNCLOS* ou Convenção das Nações Unidas sobre o Direito do Mar – CNUDM), assim como a Zona Econômica Exclusiva (ZEE) – porção marinha entre 12 e 200 milhas náuticas da costa de um país; área sobre a qual ele possui direitos exclusivos de exploração econômica, apesar de não possuir direito exclusivo de passagem. A porção marinha de até 12 milhas da costa é considerada água territorial do país, sujeita, por isso, à sua plena soberania (UNITED NATIONS, 1982).

O próximo nível de uma determinada região retrata a divisão geográfica de acordo com as diferentes legislações preconizadas em cada país. O Brasil iniciou o mapeamento das malhas territoriais estaduais e municipais a partir do primeiro recenseamento realizado ainda no período Imperial, em 1872, passando pelos mapas políticos do Brasil reconfigurados ao longo do século XX, chegando até a divisão territorial de 2000 e 2010.

O Brasil é uma República Federativa formada por regiões as quais abrangem 26 estados e um Distrito Federal. As regiões brasileiras são divididas de acordo com suas características, conforme alguns critérios adotados pelo Instituto Brasileiro de Geografia e Estatística (IBGE), sendo esses aspectos físicos – clima, vegetação e relevo – ou socioeconômicos.

Os estados, por sua vez, são formados por microrregiões, que de acordo com o IBGE, são agrupamentos de municípios. O agrupamento de microrregiões forma mesorregiões.

Em se tratando do estado do Rio Grande do Sul, outra forma para agrupar os municípios são os COREDES (Conselhos Regionais de Desenvolvimento) – criados pela Lei nº 10.283, de 17 de outubro de 1994 (RIO GRANDE DO SUL, 1994). Esses conselhos têm por objetivo a promoção do desenvolvimento regional, harmônico e

sustentável, através da integração dos recursos e das ações de governo na região (FEE, 2013). A dimensão política com todos os campos discutidos é apresentada na Figura 4.3.

Figura 4.3 – Destaque da dimensão política

DIM_POLITICA	
PK	ID_DIM_POLITICA INT (NN)
	CONTINENTE VARCHAR
	PAIS VARCHAR (NN)
	REGIAO VARCHAR
	ESTADO VARCHAR
	MICROREGIAO VARCHAR
	MESOREGIAO VARCHAR
	COREDES VARCHAR
	CIDADE VARCHAR
	BAIRRO VARCHAR
	NIVEL INT

4.2.2 Dimensões físicas

O aspecto físico de um ponto são as características desse ponto em relação às suas características geográficas locais. Um ponto, quando em terra possui um relevo, está localizado em uma classe climática, na área de uma bacia hidrográfica, inserido em um bioma que possui um tipo de vegetação e solos distintos.

Sendo assim, a dimensão física também é uma unidade de gestão possível para se avaliar um fato. Como na dimensão política, as dimensões físicas podem estar classificadas de acordo com legislações e diretrizes de acordo com cada país. Para simplificar a dimensão física, ela foi dividida em cinco outras, e cada uma das dimensões auxilia na modelagem e permite a construção de dimensões com menos linhas, pois a divisão reduz o produto combinatorial produzido pelas inúmeras classificações decorrentes do aspecto físico.

As cinco divisões criadas foram: a dimensão hidrográfica (seção 4.2.2.1); a dimensão tipo de solo (seção 4.2.2.2); a dimensão vegetação (seção 4.2.2.3); a dimensão clinografia (seção 4.2.2.4); a dimensão relevo (seção 4.2.2.5); e a dimensão clima (seção 4.2.2.6). Essas dimensões são apresentadas nas seções seguintes. Para cada uma dessas seções vai ser descrito qual foi a referência para montar a tabela de dimensão e uma breve descrição de que informações podem ser retiradas de cada dimensão.

4.2.2.1 Dimensão hidrográfica

Uma bacia hidrográfica é caracterizada pela área de drenagem para um recurso hídrico. Entende-se por bacia hidrográfica aquela cuja área é drenada parcial ou totalmente por um ou vários cursos d'água (ANA, 2001), e por região hidrográfica a área que engloba uma ou mais bacias hidrográficas (RIO DE JANEIRO, 1986). A delimitação da bacia hidrográfica leva em consideração o recurso hídrico que se quer analisar.

Existem diversas propostas metodológicas para a delimitação de bacias e a mais aceita no Brasil é a divisão de Otto Pfafstetter⁶ (1989) apud Verdin (1997, p. 1) que trata de um método hierárquico que tem como base a topografia do terreno, permitindo um detalhamento do sistema hídrico. A importância dessa metodologia é evidente, tendo em vista a sua adoção por diversas instituições e órgãos governamentais (RUPERT, 2000). A codificação de bacias hidrográficas, apresenta a metodologia descrita a seguir:

A metodologia de Otto Pfafstetter é aplicada inicialmente para o continente e consiste em agrupar a rede da drenagem em três classes: aquelas que drenam diretamente para o mar (podendo ser uma bacia ou uma região hidrográfica), aquelas que drenam para bacias fechadas e aquelas que são tributárias dos dois primeiros casos (RUPERT, 2000, p. 19),.

A divisão das OTTO Bacias do Brasil pode ser adquirida diretamente do site do MMA (Ministério do Meio Ambiente) através de seu *web service*⁷ para informações geográficas. As informações disponibilizadas pelo MMA/ANA (Agência Nacional das Águas) permitem utilizar até cinco níveis para algumas bacias. Caso se tenham informações em uma escala mais precisa, pode-se aumentar essa classificação.

Além dessa divisão, existem ainda os comitês de bacias que podem ser entendidas como o fórum em que um grupo de pessoas se reúne para discutir sobre um interesse comum – o uso d'água na bacia (ANA, 2011). Esses comitês, através de seus planos de bacias que visam classificar e estudar uma bacia hidrográfica, definem novas sub-bacias.

As divisões oriundas até o nível cinco de Otto Bacias (*dim_fisica_hidrografica_regiao*), e a divisão por comitês (*dim_fisica_hidrografica_comite*) foram modeladas nas dimensões apresentadas na Figura 4.4.

⁶ Pfafstetter, Otto. Classificação das Bacias Hidrográficas, 1989 – Obra não publicada.

⁷<http://mapas.mma.gov.br/geonetwork/srv/br/metadata.show?id=252>

Figura 4.4 – Tabelas de dimensões hidrográficas

DIM_FISICA_HIDROGRAFICA_REGIAO	
ID_DIM_FISICA_HIDROGRAFIA_REGIAO	SERIAL (NN)
REGIAO_HIDROGRAFICA_NIVEL_1	VARCHAR
REGIAO_HIDROGRAFICA_NIVEL_2	VARCHAR
REGIAO_HIDROGRAFICA_NIVEL_3	VARCHAR
REGIAO_HIDROGRAFICA_NIVEL_4	VARCHAR
REGIAO_HIDROGRAFICA_NIVEL_5	VARCHAR
NIVEL_REGIAO_HIDROGRAFICA	INT

DIM_FISICA_HIDROGRAFICA_COMITE	
ID_DIM_FISICA_HIDROGRAFICA_COMITE	SERIAL (NN)
BACIA_COMITE	VARCHAR
BACIA_COMITE_NIVEL_1	VARCHAR
BACIA_COMITE_NIVEL_2	VARCHAR
BACIA_COMITE_NIVEL_3	VARCHAR
NIVEL_BACIA_COMITE	INT

4.2.2.2 Dimensão tipo de solo

A dimensão tipo de solo leva em consideração a classificação do tipo de solo existente no local do evento. A definição de solo pode ser entendida como:

O solo que classificamos é uma coleção de corpos naturais, constituídos por partes sólidas, líquidas e gasosas, tridimensionais, dinâmicos, formados por materiais minerais e orgânicos que ocupam a maior parte do manto superficial das extensões continentais do nosso planeta, contêm matéria viva e podem ser vegetados na natureza onde ocorrem e, eventualmente, terem sido modificados por interferências antrópicas (EMBRAPA, 2006, p. 27).

As classificações de solo para o Brasil seguem o sistema brasileiro de classificação dos solos (EMBRAPA, 2006). Os níveis categóricos aplicados para o Sistema Brasileiro de Classificação de Solos são seis:

- 1º nível categórico (ordens);
- 2º nível categórico (subordens);
- 3º nível categórico (grandes grupos);
- 4º nível categórico (subgrupos);
- 5º nível categórico (famílias);
- 6º nível categórico (séries).

Os níveis categóricos apresentados anteriormente podem ser vistos na modelagem da “dim_fisica_solo” apresentada na Figura 4.5.

Figura 4.5 – Dimensão física tipo de solo

DIM_FISICA_SOLO	
ID_DIM_FISICA_SOLO	SERIAL (NN)
ORDEM	VARCHAR
SUB_ORDEM	VARCHAR
GRANDE_GRUPO	VARCHAR
SUBGRUPO	VARCHAR
FAMILIA	VARCHAR
SERIE	VARCHAR
DESCRICAO	VARCHAR
FONTE_DESCRICAO	VARCHAR

4.2.2.3 Dimensão vegetação

A dimensão vegetação classifica o fato segundo as características da vegetação do local, que pode ser denominada fitogeografia. O Manual Técnico de Vegetação Brasileira (IBGE, 2012) traz um total de 24 classificações diferentes, divididas em três (3) grupos, conforme as áreas de abrangência: classificações universais; classificações continentais; e classificações brasileiras. Essa quantidade de classificações poderia ser modelada utilizando a técnica apresentada na Figura 2.2 para hierarquias implícitas. Contudo, nesta modelagem, optou-se por utilizar apenas a classificação proposta no Manual Técnico de Vegetação (IBGE, 2012).

A metodologia cartográfica usada no Sistema de Classificação da Vegetação Brasileira, apresentada no Manual Técnico de Vegetação, divide as classificações de acordo com os objetivos a serem alcançados. Essas classificações são:

- “regional” (1:10 000 000 a 1:2 500 000);
- “exploratória” (1:1 000 000 a 1:250 000);
- “semidetalhe” (1:100 000 a 1:25 000); e
- “detalhe” (maiores que 1:25 000).

As classificações foram modeladas até o nível de semidetalhe, porém os dados disponíveis publicamente o nível de detalhe por ser para estudos mais específicos, não foi contemplado neste trabalho.

Do nível da escala no Sistema Brasileiro de Classificação da Vegetação de regional até semidetalhe, a vegetação pode ser classificada em:

- Tipos de vegetação
 - Classes de formação

- Subclasses de formação
- Grupos de formação
- Subgrupos de formação
- Formações
 - Formações
 - Subformações

Essas classificações, assim como o nível de detalhamento, podem ser vistos na dimensão de vegetação da Figura 4.6.

Figura 4.6 – Dimensão vegetação e seus atributos

DIM_FISICA_VEGETACAO	
ID_DIM_FISICA_VEGETACAO	SERIAL (NN)
ESCALA_FAIXA_INICIAL	VARCHAR (NN)
ESCALA_FAIXA_FINAL	VARCHAR (NN)
ESCALA_NIVEL_DETALHAMENTO	VARCHAR (NN)
CLASSE_DE_FORMACAO	VARCHAR (NN)
SUBCLASSE_DE_FORMACAO	VARCHAR (NN)
GRUPO_FORMACAO	VARCHAR (NN)
FORMACAO	VARCHAR (NN)
SUBFORMACAO	VARCHAR (NN)

4.2.2.4 Dimensão clinografia

A clinografia é a declividade de um terreno. Para o monitoramento de desastres naturais, como deslizamentos e enchentes, a característica da clinografia é extremamente importante.

Declividade do terreno representa o desnível relativo entre dois pontos da superfície terrestre. Ele pode ser obtido através da determinação do ângulo de inclinação (α) ou através do valor da declividade em porcentagem ($d\%$). Para esse cálculo, utiliza-se a medida de elevação vertical (V) relacionada com a distância horizontal (H) correspondente. Assim, o declive é a relação entre V e H conforme as equações 1 e 2 (ANDERSON, 2002).

$$\alpha = \text{tangente } V/H \quad (1)$$

$$d\% = V/H * 100 \quad (2)$$

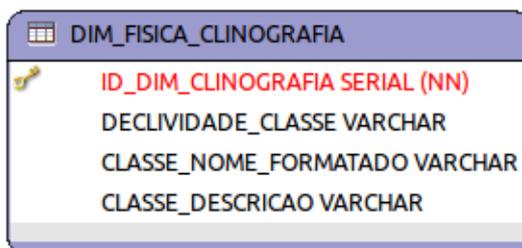
Para uma rápida classificação, a tabela de dimensão da clinografia foi populada com os intervalos rígidos, propostos em (LEPSCH; JÚNIOR, 1983), ilustrada na Tabela 4.1.

Tabela 4.1 – Intervalos rígidos de Lepsch

<i>CLASSES</i>	<i>INTERVALOS DE DECLIVIDADE</i>
A	$d\% < 2\%$
B	$2 \leq d\% < 5\%$
C	$5 \leq d\% < 10\%$
D	$10 \leq d\% < 15\%$
E	$15 \leq d\% < 45\%$
F	$45 \leq d\% < 70\%$
G	$70\% \leq d\%$

A dimensão clinografia e seus campos é apresentada na Figura 4.7.

Figura 4.7 – Dimensão de clinografia



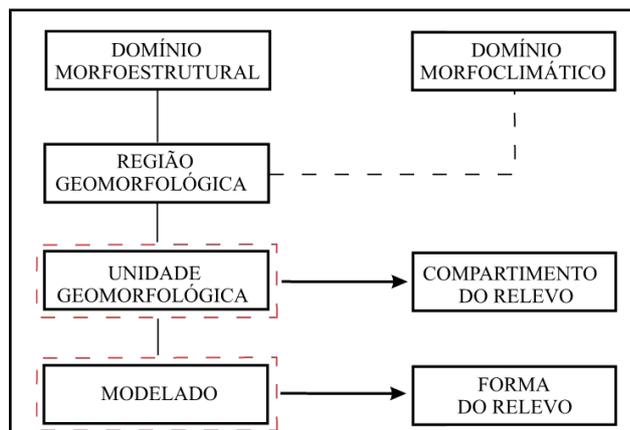
Como a declividade pode ser um valor escalar, esse valor, caso tenha importância para o projeto de implementação, pode ser modelado com uma dimensão descaracterizada.

4.2.2.5 Dimensão Geomorfologia

A geomorfologia estuda as formas da superfície terrestres assim como a relação de fatores que atuam na formação do relevo. O mapeamento do relevo pode ser agrupado de acordo com a taxonomia proposta no Manual Técnico de Geomorfologia. Para agrupamento, hierarquização e individualização dos conjuntos de feições, são considerados como parâmetros fatores causais, de natureza estrutural, litológica, pedológica, climática e morfodinâmica, responsáveis pela evolução das formas do relevo e pela composição da paisagem no decorrer do tempo geológico (IBGE, 2009).

No manual são identificados de acordo com a ordem decrescente de grandeza: Domínios Morfoestruturais, Regiões Geomorfológicas, Unidades Geomorfológicas, Modelados e Formas de Relevo Simbolizadas. Esses agrupamentos são representados na Figura 4.8, adaptada do Manual de Geomorfologia.

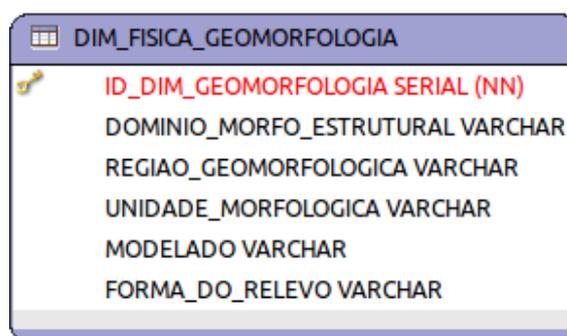
Figura 4.8 – Taxonomia de geomorfologia



Fonte: IBGE (2009)

Essa taxonomia foi modelada na dimensão apresentada na Figura 4.9.

Figura 4.9 – Dimensão geomorfologia



4.2.2.6 Dimensão clima

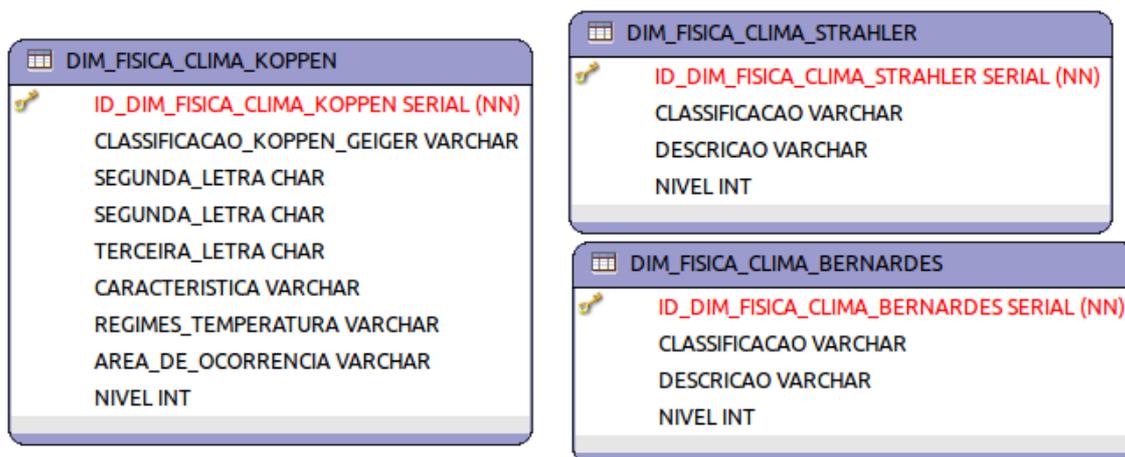
O clima pode ser entendido como as condições atmosféricas médias em determinada região. Os Sistemas de Classificações Climáticas (SCC) são de grande importância, pois analisam e definem os climas das diferentes regiões. Os SCC levam em consideração vários elementos climáticos ao mesmo tempo, o que facilita a troca de informações e análises posteriores para diferentes objetivos (ROLIM et al., 2007).

A classificação de Köppen-Geiger faz uso de três letras para apresentar as características do clima:

- A primeira letra, maiúscula, representa a característica geral do clima de uma região;
- A segunda letra, minúscula, representa as particularidades do regime de chuvas;
- A terceira letra, minúscula, representa a temperatura característica de uma região.

Existem ainda as classificações de Arthur Strahler e de Lysia Bernardes, porém elas não são hierárquicas. Para modelar essas 3 classificações, foram definidas 3 dimensões (Figura 4.10) e a dimensão que recebeu mais atenção foi a de classificação de Köppen, visto que a mesma é uma das mais abrangentes e possui uma hierarquia intrínseca (ROLIM et al., 2007).

Figura 4.10 – Dimensões modeladas para o clima



4.2.3 Dimensões socioeconômicas

Uma vez que o modelo propõe-se a ser genérico, as dimensões socioeconômicas são características fundamentais, especialmente em áreas humanas e da saúde, esta última, por sua vez, está intimamente ligada a fatores socioeconômicos, os quais são determinantes e condicionantes nos processos de saúde-doença e na implementação de políticas públicas de saúde (BATISTELLA, 2007).

As dimensões socioeconômicas podem contemplar faixa etária, densidade da população, assim como características econômicas, como por exemplo, se a região é tipicamente agrícola ou industrializada.

A complexidade desse dado é que ele pode não ter características geográficas e são valores escalares o que normalmente se enquadra em dimensões descaracterizadas (KIMBALL; ROSS, 2013).

É importante ressaltar que as dimensões socioeconômicas têm uma forte ligação com a dimensão política, pois os valores normalmente são obtidos a partir de um estado, município ou bairro. Sendo assim, além do nível hierárquico, é necessário saber até que escala podem ser utilizadas essas formas de agrupamento.

Devido aos fatores discutidos, as dimensões socioeconômicas não foram contempladas neste modelo.

4.2.4 Dimensões temporais

Um evento, quando adicionado ao mapa, possui características peculiares, como representar um evento que ocorreu na mesma posição mais de uma vez. As técnicas de grupo neste aspecto são interessantes, pois é possível facilmente agrupar uma informação pela sua dimensão temporal.

Em termos de hierarquia temporal, pode-se facilmente agregar a informação por dia, mês, ano, semestre, estação, entre outras granularidades temporais típicas. Para modelar as dimensões temporais e manter o formato estrela sem gerar a explosão combinatorial de para cada dia gerar as 24 horas e seus respectivos minutos, mantendo essas informações em dimensões que facilitam a criação das hierarquias, a dimensão temporal foi dividida em dia e hora. Essa divisão é comum em modelagem dimensional, principalmente para as questões temporais (KIMBALL; ROSS, 2013). A Figura 4.11 mostra as duas dimensões criadas para esse proposto.

Figura 4.11 – Dimensão temporal dia e hora

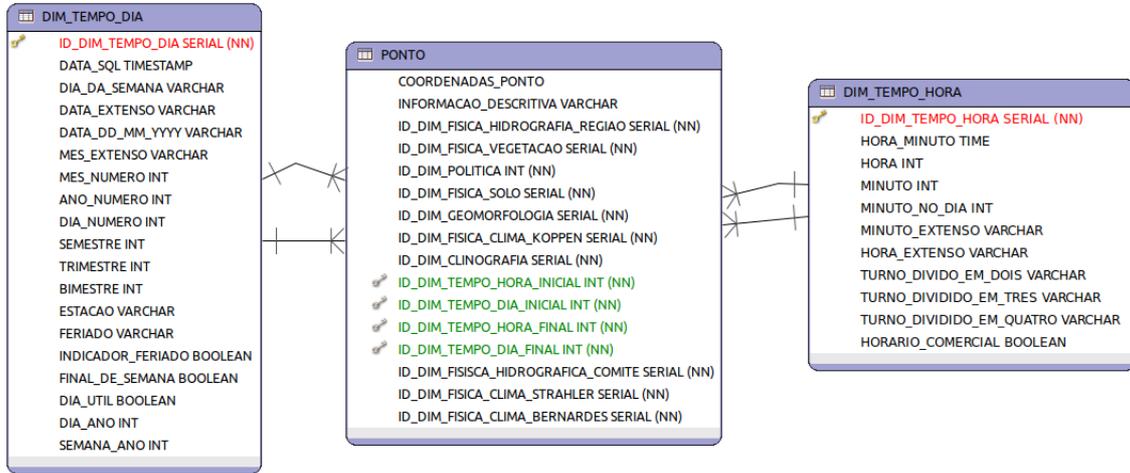
DIMENSAO_TEMPO_DIA	DIMENSAO_TEMPO_HORA
ID_DATASERIAL (NN)	ID_DIMENSAO_TEMPO_HORA SERIAL (NN)
DATA_SQL TIMESTAMP	HORA_MINUTO TIME
DIA_DA_SEMANA VARCHAR	HORA INT
DATA_EXTENSO VARCHAR	MINUTO INT
DATA_DD_MM_YYYY VARCHAR	MINUTO_NO_DIA INT
MES_EXTENSO VARCHAR	MINUTO_EXTENSO VARCHAR
MES_NUMERO INT	HORA_EXTENSO VARCHAR
ANO_NUMERO INT	TURNO_DIVIDIDO_EM_DOIS VARCHAR
DIA_NUMERO INT	TURNO_DIVIDIDO_EM_TRES VARCHAR
SEMESTRE INT	TURNO_DIVIDIDO_EM_QUATRO VARCHAR
TRIMESTRE INT	HORARIO_COMERCIAL BOOLEAN
BIMESTRE INT	
ESTACAO VARCHAR	
FERIADO VARCHAR	
INDICADOR_FERIADO BOOLEAN	
FINAL_DE_SEMANA BOOLEAN	
DIA_UTIL BOOLEAN	
DIA_ANO INT	
SEMANA_ANO INT	

A dimensão temporal para um objeto geográfico tem implicações diretas para a tomada de decisão e a característica de um dado evento ocorrer mais de uma vez em uma mesma posição necessita de uma forma de visualização mais complexa. O grupo de dados gerado permite uma navegação mais exploratória, gerando assim uma possibilidade de obter um resultado mais correto em relação ao tempo.

Para gerenciar o fato, é necessário elencar dois marcos em termos de tempo. O primeiro marco é a data em que o fato ocorreu, e o segundo marco de data a se

armazenar é a data de validade. A dupla relação das dimensões temporais com as tabelas de fato podem ser vistas na Figura 4.12.

Figura 4.12 – Dimensões temporais com dupla relação em relação ao ponto



Caso o dado ainda seja válido, ou seja, o dado esteja vigente, deve-se criar uma entrada temporal indefinida. Sendo assim, o registro com o id 1 na tabela de dimensão deve ser indefinido. A Figura 4.13 mostra a primeira linha da dimensão “dim_tempo_dia” e a “dim_tempo_hora” com essa representação. É melhor ter um apontador para uma data definida como indeterminada ou infinito do que deixar o campo da chave nula, visto que, além de deixar o banco mais conciso, essa linha pode ser enriquecida com descrições para o tomador de decisão.

Figura 4.13 – Valores indefinidos para hora

Edit Data - Localhost (localhost:5432) - mestrado - dim_tempo_dia																			
	id_dim_tempo_dia [PK] serial	data_sql timestamp	dia_da_sen character	data_extenso character varying	data_dd_m character	mes_exteri integer	mes_nome integer	ano_nome integer	dia_numero integer	semestre integer	trimestre integer	bimestre integer	estacao character	feriado boolean	indicador boolean	final_de_si boolean	dia_util boolean	dia_ano integer	semana_an integer
1	1			Indeterminado	0	0	0	0	0	0	0	0	0	0	FALSE	FALSE	FALSE	0	0

Edit Data - Localhost (localhost:5432) - mestrado - dim_tempo_hora											
	id_dim_tempo_hora [PK] serial	hora_minu time	hora integer	minuto integer	minuto_no character	minuto_ex character varying	hora_extenso character varying	turno_divi character	turno_divi character	turno_divi character	horario_co boolean
1	1	00:00:00	0	0	0	Indeterminada	0	0	0	0	FALSE

4.3 Dimensões personalizadas

Por ser um modelo restrito a alguns domínios, o mesmo apresenta dimensões não abordadas ou então exploradas superficialmente. Para a implementação desse modelo para outras situações e/ou aplicações, torna-se necessária a implementação de outras dimensões.

Algumas características comuns foram elencadas entre as dimensões modeladas neste trabalho e permitem a criação de novas dimensões e carregamento de dados para o modelo. Dessa forma, o modelo pode se expandir e a ferramenta evoluir em tempo real.

Existem duas formas de criar dimensões: fazê-lo diretamente no banco de dados

ou utilizar o sistema de criação de dimensão apresentado no capítulo 5. No caso de fazê-lo de forma manual no banco de dados, é necessário construir uma dimensão com algumas características básicas:

- campos que a tabela de dimensão deve possuir;
 - *id_dimensao_personalizada* – chave de integridade referencial;
 - *nivel* – identificador numérico para controlar o nível de hierarquia;
 - *escala_geometria* – escala em que a informação foi gerada, e que possibilita o cálculo da acurácia da informação assim como um controle da confiabilidade dos dados em relação ao modelo;
 - *geometria* – campo que vai armazenar o polígono utilizado para o processo de ETL;
 - *fonte_dados* – informação de quem gerou os dados geográficos.
- preenchimento dos dados padrão apresentados na Tabela 4.2;
- marcar como obrigatório os campos que servirão para formar a hierarquia.

4.4 Limitações do modelo

Qualquer que seja a representação da realidade ou a representação da informação, corre-se o risco de representar apenas algumas facetas, pois a realidade é dinâmica e não pode ser representada na totalidade por um modelo estático. Outra característica relevante é o nível de informação de um modelo para a tomada de decisão, que pode ser genérico (necessitando simplificação de alguns aspectos, para poder ter dados) ou muito complexo, dificultando a sua interpretação.

Outra situação que o modelo não tem capacidade de prever é o fato de não se existir informação para uma ou outra dimensão em determinada região do mapa. Essa situação e a situação de uma dimensão não obrigatória são tratadas com duas tuplas pré-carregadas em cada dimensão do modelo. A Tabela 4.2 mostra os valores principais das tuplas que deverão ser inseridas para cada dimensão. Caso os dados tenham esse valor nos agrupamentos, os pontos são organizados no mapa por esses critérios.

Tabela 4.2 – Tuplas pré-carregadas no modelo

<i>Id</i>	<i>Significado</i>	<i>Resolve</i>
1	Não se aplica	- quando a dimensão não possui valor para ser agrupado; - não existe um valor possível no momento
2	Informação inexistente	O valor não é conhecido, mas pode existir – ao longo do tempo, essa informação tem chances de ser preenchida corretamente.

Essas características podem e devem ser controladas e medidas. Nesse sentido, é importante ressaltar alguns aspectos do modelo:

1. avaliação apenas para os padrões de classificação do Brasil – apesar de algumas classificações poderem ser aplicadas em nível global, a modelagem opta em contemplar apenas as características e classificações brasileiras;
2. incompleto – o modelo não é completo, não atende nem conseguirá atender o universo de dimensões que um objeto geográfico pode possuir;
3. imprecisão nos limites – a escala que a informação é gerada possui precisões geográficas diferentes;
4. ponto sobre o limite – ocorre quando um ponto fica exatamente no limite entre dois municípios;
5. contempla agregação apenas de pontos – apesar de muitas dimensões poderem ser aproveitadas caso a tabela de fato seja alterada, a utilização da agregação de outros objetos geográficos mais complexos, como linhas e polígonos, necessita de tratamentos diferentes. Um exemplo é um rio (representado por uma linha), que pode passar por diversas formações vegetais, e por isso é necessária uma abordagem diferente para o algoritmo de agregação.

Apesar dessas limitações terem sido detectadas, alguns ajustes foram feitos para não limitar a utilização. Para solucionar os problemas um e dois, foi desenvolvida a possibilidade de construção de dimensões personalizadas. O problema três é controlado através da propriedade escala. Para o problema quatro a solução foi controlar o processo de ETL.

4.5 Considerações sobre o modelo

O modelo proposto não visou ser completo, mas sim expansível. Os campos padronizados elencados de acordo com a necessidade de agrupamento de informações geográficas em *web maps* permitem criar um padrão que pode ser adotado para qualquer sistema. A modelagem aqui proposta elenca não só as características mais relevantes para os domínios de estudo, como também apresenta as limitações e formas de contornar e/ou controlar esses problemas.

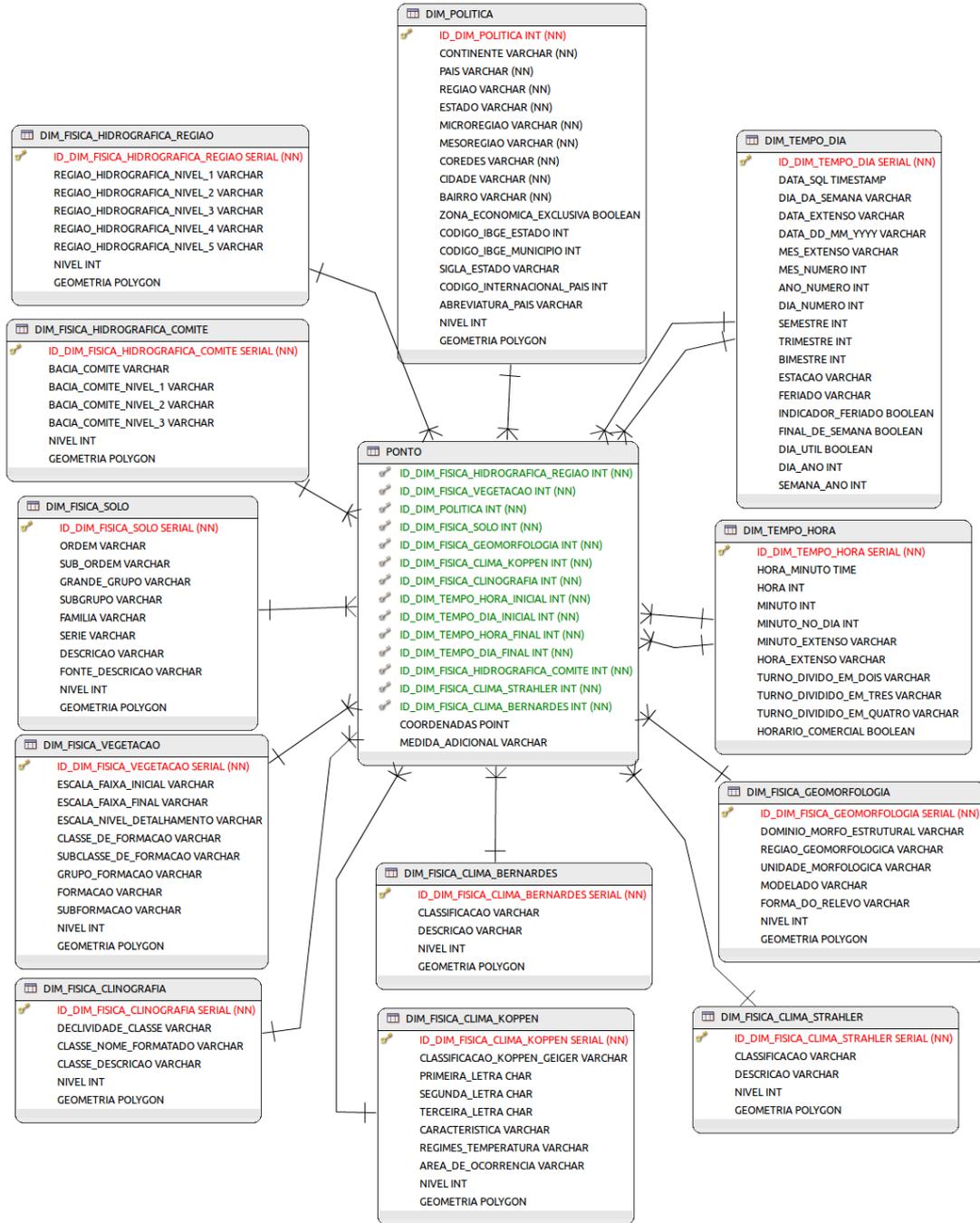
Cabe também destacar que o agrupamento de geometrias mais complexas como linhas e polígonos não foi abordado, uma vez que não é o objetivo deste trabalho. Para estender para outras geometrias, deve-se modelá-las como fatos. Entretanto, a forma de

representar no mapa, realizar as consultas e gerar os agrupamentos deve ser readequado. As linhas, por exemplo, necessitam técnicas de visualização em diferentes escalas como simplificação, generalização e não agrupamento.

Uma característica intrínseca a um modelo dessa magnitude é a mudança que ele pode sofrer. Por exemplo, caso a dimensão solo sofra uma nova classificação mais aceita ou mais precisa, o modelo precisa ser readequado. Dessa forma, além de o modelo não ser estático, ele tampouco pode ser tratado como definitivo. O modelo multidimensional, por ter hierarquia achatada e pouca dependência entre tabelas, permite se adaptar muito bem a mudanças.

O modelo completo desenvolvido neste trabalho, com todos os atributos para cada dimensão, é apresentado na Figura 4.14.

Figura 4.14 – Modelo multidimensional completo



5 A FERRAMENTA PARA O AGRUPAMENTO PERSONALIZADO DE PONTOS EM *WEB MAPS* (APPWM)

Devido à necessidade de avaliar o modelo multidimensional foi desenvolvida a ferramenta denominada APPWM – Agrupamento Personalizado de Pontos em *Web Maps*. A ferramenta tem como objetivo avaliar a aplicação do modelo multidimensional em um *web map* real. A APPWM é desenvolvida com vistas a facilitar o uso do modelo multidimensional montando as consultas e retornando as informações necessárias para a realização dos agrupamentos. Além das funcionalidades de consulta, a ferramenta apresenta diversas funcionalidades para manter e criar novas dimensões.

O desenvolvimento da ferramenta foi realizado com avaliações rápidas, de forma *ad-hoc* por possíveis usuários solicitando para verbalizarem o que pensavam sobre a ferramenta. Além da construção da ferramenta ter levado em consideração as necessidades dos usuários, foram usados componentes e métodos clássicos em *web maps* comerciais como Google Maps⁸ e Mapas Bing⁹.

Para facilitar o entendimento da ferramenta, este capítulo está dividido nas seguintes seções: a seção 5.1 apresenta a arquitetura da ferramenta; a seção 5.2 é voltada para as discussões sobre a implementação da visualização de dados; a seção 5.3 apresenta a forma de agrupamento de dados; a seção 5.4 apresenta o processo de carregamento de dados; a seção 5.5 apresenta como foram organizados os experimentos para utilizar o mesmo modelo; a seção 5.6 apresenta algumas telas e funcionalidades desenvolvidas para a interpretação de dados; e a seção 5.7 realiza uma discussão final sobre a ferramenta.

5.1 Arquitetura da APPWM

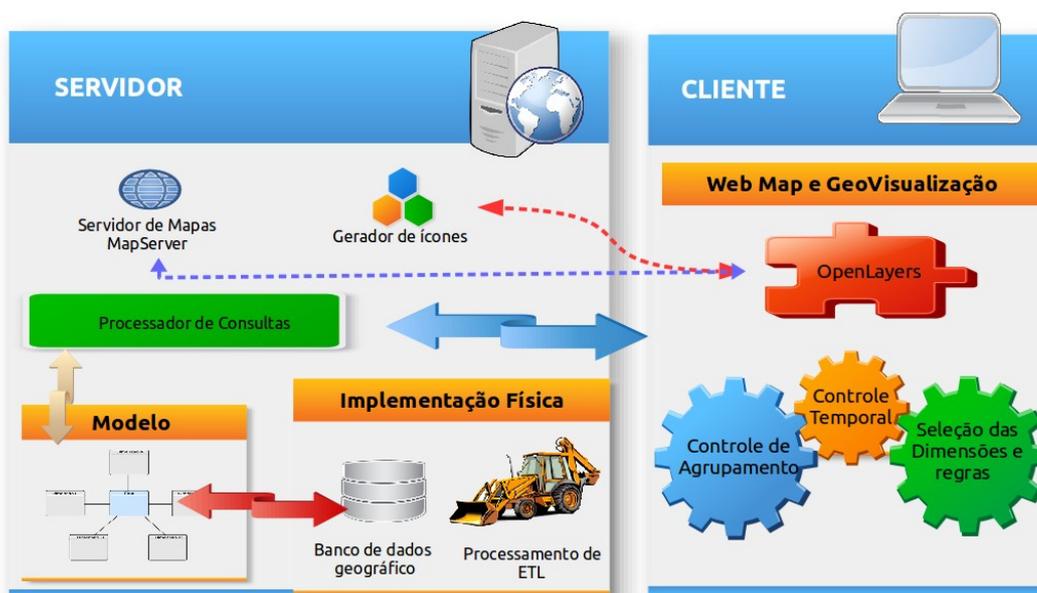
A ferramenta possui uma arquitetura dividida em camadas, e cada uma delas compõe uma das etapas de desenvolvimento da implementação. A Figura 5.1 apresenta essas 3 camadas: o modelo multidimensional que foi detalhado no capítulo 4; a camada de apresentação (bloco *Web map e geovisualização*), formada pelo *web map* e pelos recursos de visualização, é detalhada na seção 5.2; já a camada de implementação física

⁸ Google Maps – Disponível em <https://maps.google.com.br/>

⁹ Mapas Bing – Disponível em <http://www.bing.com/maps/>

dos dados (inferior), com o respectivo processo de carga dos dados, é detalhada na seção 5.4.

Figura 5.1 – Ferramenta APPWM



O *web map* utiliza a biblioteca OpenLayers que se conecta diretamente ao servidor de mapas MapServer (instalado e configurado no servidor) e aos servidores do Google, para poder gerar os mapas bases. O OpenLayers também utiliza o Gerador de Ícones (seção 5.2.4). A implementação do cliente ainda contempla três módulos compostos pela ferramenta de controle de agrupamento, controle temporal e seleção de dimensões e regras (ambas contempladas na seção 5.2.1).

Quanto ao servidor o mesmo possui um processador de consultas e sistemas de controle do banco de dados geográfico (seção 5.4.1) assim como o processamento dos dados de Extração e carregamento de dados (seção 5.4.2).

5.2 Camada de visualização.

A camada de visualização, devido à arquitetura intrínseca de sistemas *web*, é dividida em duas partes: a parte cliente (seção 5.2.1), na qual é gerado o mapa e executado o algoritmo de *cluster* e a parte servidor (seção 5.2.2), a qual fornece as informações para a geração do mapa, como a geração automática dos ícones.

5.2.1 Implementação do cliente

Esta seção visa apresentar o desenvolvimento do *web map* no que tange aos aspectos do cliente. A parcela de implementação do cliente são os códigos que devem ser baixados para o computador do usuário e essa parcela foi desenvolvida com a utilização da tríade clássica das páginas *web*: HTML (*HyperText Markup Language*), CSS (*Cascading Style Sheets*) e JavaScript.

O HTML é uma linguagem de marcação, e na aplicação serve para estruturar o *web map*. Para o *web map* desenvolvido, o HTML está de acordo com os novos padrões estipulados na versão 5 da linguagem. O HTML 5 contém recursos para aplicativos baseados na *Web*, suporte de vídeo, gráficos, mais efeitos de estilos, e um conjunto completo de APIs. HTML 5 adapta-se a qualquer dispositivo, seja *desktop*, celular, *tablet* ou televisão (W3C, 2012). O HTML 5 está presente nos navegadores mais atuais e uma lista das funcionalidades, assim como o suporte dos *browsers* pode ser obtido em (<http://html5readiness.com/>). Apesar dos navegadores mais antigos não suportarem HTML 5, são utilizados recursos que não comprometem a renderização correta da página, diminuindo a incompatibilidade do *web map*.

O CSS, que é responsável pela aparência de uma página “formata a informação entregue pelo HTML. Essa informação pode ser qualquer coisa: imagem, texto, vídeo, áudio ou qualquer outro elemento criado” (W3C, 2011 pg. 5). No desenvolvimento foi utilizada a versão 3, que assim como o HTML 5 possui limitações de funcionalidades em determinadas versões de *browsers*. Para conseguir uma compatibilidade maior com os *browsers* anteriores, além de facilitar a construção dos estilos e formatação, foram utilizados dois geradores de CSS *online*: CSS3generator¹⁰ e o CSS3Gen¹¹. Com a combinação de HTML e CSS foram desenvolvidos os menus e as legendas do mapa.

Para permitir a interação com o mapa foi utilizada a linguagem de programação JavaScript, e para a implementação seguiu-se o padrão ECMA-262 (ECMA INTERNATIONAL, 2011). Para otimizar o desenvolvimento e compatibilidade entre os *browsers* foi utilizado a biblioteca JQuery¹². Ainda, para verificar a performance e melhorar a padronização desse código, foi usado o analisador de otimização de código JSLint¹³. A biblioteca JQuery auxiliou a desenvolver os menus sensíveis ao contexto e a

¹⁰CSS3GENERATOR – Sistema gerador de CSS3 disponível em: <http://css3generator.com>

¹¹CSS3Gen – Sistema gerador de CSS3 disponível em: <http://css3gen.com>

¹²Jquery – Biblioteca JavaScript disponível em: <http://jqueryui.com>

¹³JSLint – Otimizador de códigos Javascript disponível em: <http://www.jshint.com>

montar o mapa, além das funcionalidades de expandir ou contrair menus de opções. Além da biblioteca JQuery, foi utilizado o seu *plugin* JQueryUI¹⁴ (*Unique Interfaces*), que possui componentes prontos como a barra deslizante do tamanho do agrupamento, a *FalseWindow* (janela falsa dentro do navegador, utilizado nas janelas de acesso aos formulários de avaliação do experimento e da legenda). A biblioteca JQuery foi escolhida devido à grande quantidade de *plugins* e documentação disponível.

A legenda é gerada a partir de um objeto JavaScript, que é configurável para fornecer mais ou menos funcionalidades. As propriedades implementadas desses itens da legenda são apresentados na Figura 5.2.

Figura 5.2 – Legenda do mapa com funcionalidades para controlar os agrupamentos



As funcionalidades adicionadas na legenda apresentada na Figura 5.2 foram numeradas em vermelho e são explicadas na sequência:

1. mais *cluster* ou menos *cluster* – essa funcionalidade permite ocultar ou apresentar o menu simplificado de *cluster* (segundo item de legenda dos produtores de frango);

¹⁴JqueryUI – Plugin da biblioteca jquery para componentes gráficos, disponível em: <http://api.jquery.com>

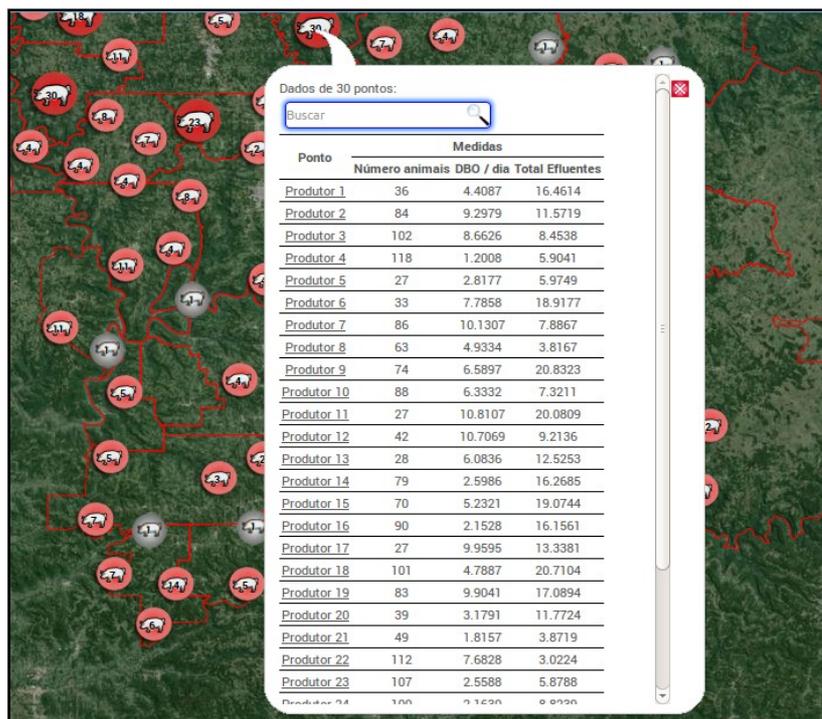
2. criar *cluster* – essa funcionalidade executa o algoritmo de agrupamento, do OpenLayers (seção 5.3);
3. selecionar critérios – essa funcionalidade permite ao usuário definir o contexto do agrupamento. Esse menu fornece uma entrada simples ao usuário para gerar os agrupamentos de acordo com o critério e o contexto que o mesmo necessita dos dados. A seleção do critério pode ser simples, ou seja, apenas uma dimensão e um nível hierárquico, ou pode ser múltipla com mais de uma dimensão. Vale destacar que os campos que aparecem para escolha são os campos com a propriedade *NOT NULL* no banco de dados;
4. tamanho agrupamento – o algoritmo de agrupamento funciona com uma grade. Esse parâmetro serve para aumentar ou diminuir o tamanho do grupo, ou seja, o tamanho da célula da grade. Essa funcionalidade permite melhorar a visualização dos grupos ou até fazer contagens;
5. duplicar – essa funcionalidade permite que se avalie o plano de informação por duas facetas diferentes, ou seja, duplica o plano de informação para ver outros aspectos da mesma informação;
6. reordenar – essa funcionalidade permite alterar as ordens dos planos de informação. Elas podem ser evidenciadas onde ocorre a sobreposição de informações.

Para o desenho no mapa, foi utilizada a biblioteca OpenLayers como alternativa à API do Google Maps, que foi a escolhida no início dos experimentos. Essa troca ocorreu pelo fato de a biblioteca OpenLayers ser *Open Source* e ser *standalone*, ou seja, não precisa de conexão com a internet para poder utilizar. Uma vantagem da OpenLayers é a possibilidade de utilizar o Google Maps API como servidor de imagens; assim, é possível usar o conjunto de imagens de satélites e ruas disponibilizados pelo Google na aplicação, além de camadas próprias sem limitações.

5.2.1.1 Exploração dos grupos

Para a navegação pelos agrupamentos foi criada uma pequena aplicação que permite explorar todos os elementos de um mesmo grupo, tornando a pesquisa pelo mapa mais exploratória. Dentre as principais funcionalidades da exploração dos grupos, está o filtro de elementos, que utiliza uma expressão chave qualquer, como pode ser visto na Figura 5.3.

Figura 5.3 – Visualização dos agrupamento de dados



5.2.2 Implementação do servidor

A implementação do servidor funciona como um *proxy* de consulta para o modelo multidimensional. Além disso, o servidor, com base nos dados inseridos no modelo, constrói a interface para que possa ser escolhido o contexto dos grupos de dados. A aplicação no servidor foi implementada na linguagem PHP 5.5 orientado a objetos.

A arquitetura da aplicação utiliza o padrão MVC (*Model View Controller*) desenvolvido inicialmente em 1979 por para a linguagem Smaltalk. Esse padrão possui 3 tipos de objetos (GUANGCHUN et al., 2003; COLLINS, 2012):

1. *Model* (modelo) – o objeto de dados, o controle dos dados armazenados – classes que acessam o banco de dados assim como os objetos de dados.
2. *View* (Visualização) – camada de visualização – no servidor utiliza-se o gerenciador de *templates* Smarty¹⁵ que facilita a separação da lógica da aplicação dos HTML que serão gerados para o cliente.

¹⁵ Smarty – disponível em - <http://www.smarty.net/>

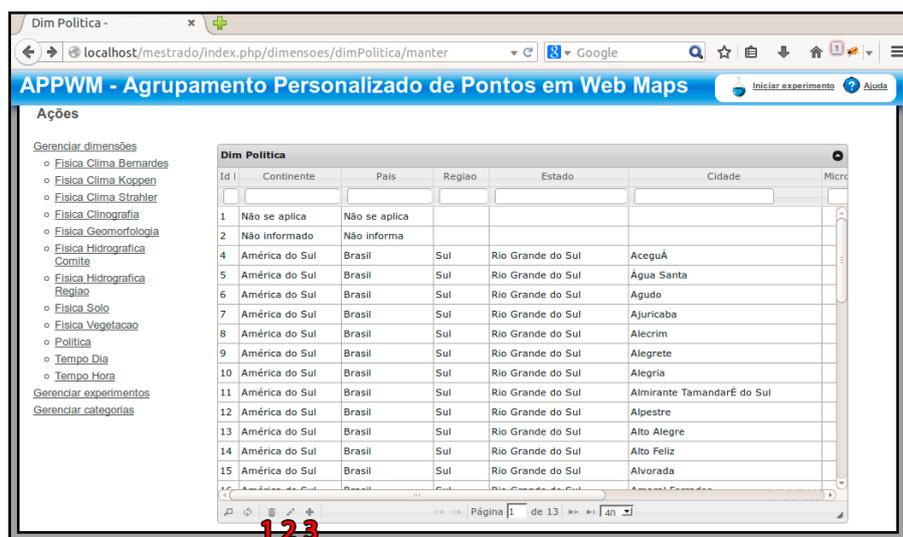
3. *Controler* (controlador) – a camada de controlador é responsável por controlar e receber os comandos dos usuários a fim de decidir e passar as instruções corretas para o modelo ou novamente para o visualizador.

São quatro as funcionalidades principais da aplicação em nível do servidor: o carregador simplificado de ETL (seção 5.4.2); a gerência dos dados das dimensões (seção 5.2.3); a geração de planos de informação para os mapas; e o gerador automático de estilos (seção 5.2.4).

5.2.3 Gerência dos dados das dimensões

Para a gerência dos dados das dimensões apresentadas no capítulo 4, foram criadas interfaces padronizadas. Essas interfaces fornecem algumas funcionalidades, conforme ilustrado na Figura 5.4.

Figura 5.4 – Manutenção dos dados das dimensões



Id	Continente	País	Região	Estado	Cidade	Micr
1	Não se aplica	Não se aplica				
2	Não informado	Não informa				
4	América do Sul	Brasil	Sul	Rio Grande do Sul	Aceguá	
5	América do Sul	Brasil	Sul	Rio Grande do Sul	Água Santa	
6	América do Sul	Brasil	Sul	Rio Grande do Sul	Agudo	
7	América do Sul	Brasil	Sul	Rio Grande do Sul	Ajuricaba	
8	América do Sul	Brasil	Sul	Rio Grande do Sul	Alecrim	
9	América do Sul	Brasil	Sul	Rio Grande do Sul	Alegrete	
10	América do Sul	Brasil	Sul	Rio Grande do Sul	Alegria	
11	América do Sul	Brasil	Sul	Rio Grande do Sul	Almirante Tamandaré do Sul	
12	América do Sul	Brasil	Sul	Rio Grande do Sul	Alpestre	
13	América do Sul	Brasil	Sul	Rio Grande do Sul	Alto Alegre	
14	América do Sul	Brasil	Sul	Rio Grande do Sul	Alto Feliz	
15	América do Sul	Brasil	Sul	Rio Grande do Sul	Alvorada	

1. Adicionar entrada na dimensão – criar uma nova linha na tabela de dimensão. Solicitam-se todos os dados da dimensão, assim como um arquivo do tipo *shape* ou KML para carregar os limites da entrada na dimensão.
2. Editar – permite a seleção e alteração de características de uma dimensão.
3. Deletar – permite excluir uma linha da dimensão. Ao se deletar uma linha que possui itens na tabela de fatos, a ferramenta tenta reenquadrar os pontos em outra dimensão, e caso não consiga, informa ao usuário e atribui o valor “não informado” na dimensão (id = 1).

O carregamento dos dados das dimensões é fundamental para o processo de ETL. Cada dimensão deve possuir um limite formado pelo seu polígono para que os

dados de pontos inseridos na tabela de fatos já sejam classificados de acordo com a dimensão. Portanto, o controle dos dados das dimensões é um requisito fundamental para o correto funcionamento do restante do sistema.

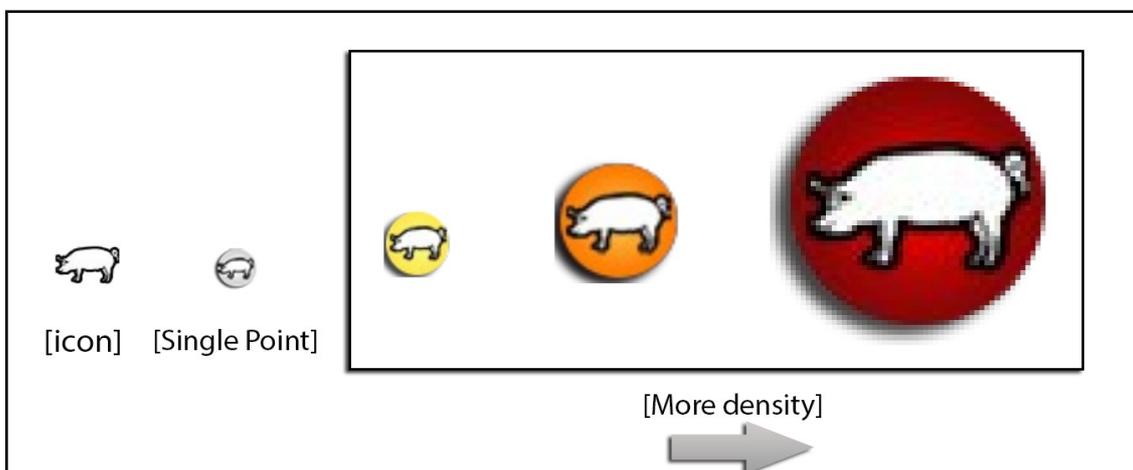
A ferramenta implementa essa funcionalidade devido à necessidade de permitir ao usuário um controle completo do modelo e de seus dados. Dessa forma, obtêm-se uma melhora significativa e uma confiança maior nos resultados das consultas. Além dessa confiança, o usuário pode explorar os dados do modelo e propor consultas diferentes das que a interface para o modelo prevê inicialmente. O conhecimento dos dados permite ao usuário a capacidade de explorar e descobrir novas informações ocultas inicialmente (DIAS et al., 2012).

5.2.4 Gerador de ícones e estilos

A funcionalidade de geração de estilos permite que, a partir de um ícone qualquer, seja possível gerar novos símbolos para representar as variações de quantidade em relação à escala. O objetivo principal de gerar automaticamente, porém controlado, os estilos, é tornar a leitura e interpretação de um mapa mais fácil e intuitiva. Essa geração automatizada de estilos torna a aplicação mais dinâmica e permite gerar inúmeras combinações do mesmo plano de informação utilizando diferentes critérios no mesmo *web map*.

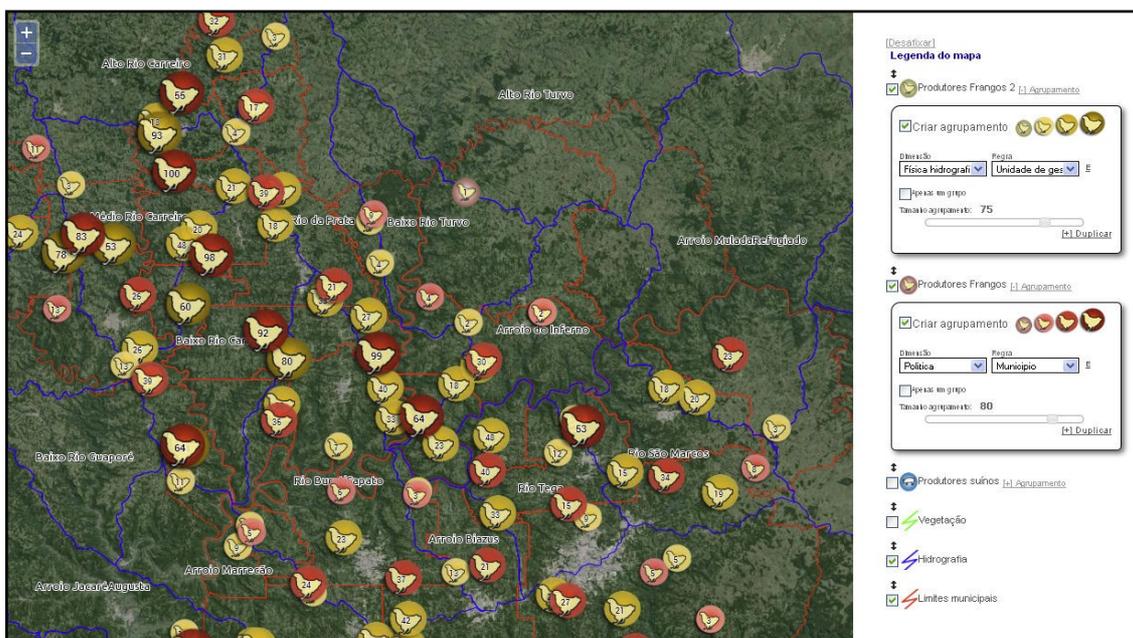
A Figura 5.5 mostra os ícones gerados a partir de um símbolo (imagem) de um porco. Esses ícones gerados servem para representar a densidade de pontos: o artifício do tamanho (quanto maior) e o artifício de cor (cor mais quente) representam maiores densidades e os opostos densidades menores.

Figura 5.5 – Geração automática de estilos



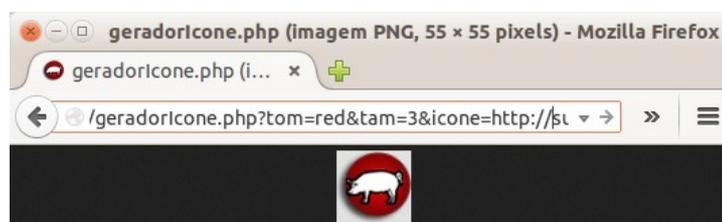
Outra vantagem de gerar os ícones através de um gerador controlado de estilos é atingir a possibilidade de visualizar um mesmo plano de informação através de aspectos e/ou visões diferentes. Essa funcionalidade pode ser vista com a Figura 5.6 onde é apresentado o mesmo plano de informação com agrupamento por município (ícones vermelhos) e agrupamento por bacia hidrográfica (ícones amarelos).

Figura 5.6 – Plano de informação duplicado



O gerador de estilos foi implementado usando uma aplicação *web service* no padrão REST. Para usá-lo, basta enviar a requisição via GET ou POST e o mesmo construirá o ícone. A documentação completa desse *web service* está demonstrado no apêndice B - Documentação web service de gerador de estilos. Um exemplo da utilização do *Web Service* pode ser visto na Figura 5.7. Observa-se ainda a requisição GET no endereço do navegador.

Figura 5.7 – Ícone gerado a partir de *web service*



5.2.5 Geração de planos de informação para os mapas

Em termos de implementação de servidor para facilitar a interpretação dos mapas, foi utilizado um servidor de mapas. Servidores de mapas permitem aos usuários

uma ampla interação com as informações espaciais disponibilizadas. Através do servidor de mapas, os usuários ou clientes podem acessar as informações no formato original e realizar consultas em diferentes níveis de complexidades (PIMENTA et al., 2012). O servidor de mapa escolhido nesta implementação foi o MapServer, devido a sua simplicidade e facilidade de instalação em ambientes Linux.

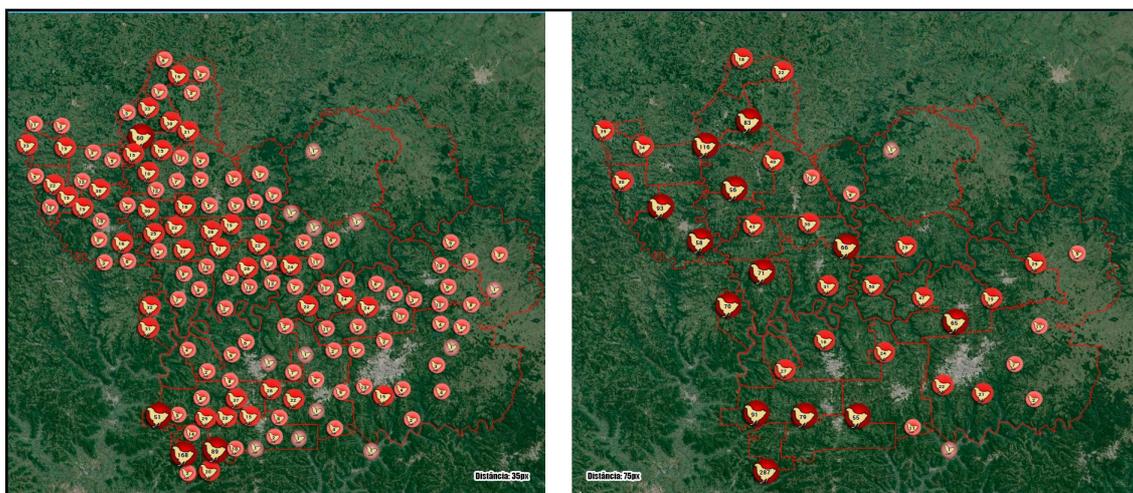
Nos experimentos, o MapServer foi utilizado para planos de informação como os limites municipais, e visualização de polígonos no mapa, como os limites base que foram utilizados para fazer o ETL.

5.3 Agrupamento de dados

A geração dos agrupamentos dos pontos utiliza o algoritmo disponibilizado na própria biblioteca OpenLayers. A escolha de utilizar o algoritmo implementado direto na biblioteca visou a possibilidade de focar os esforços no modelo e no fluxo de trabalho e não na performance e acurácia do algoritmo de agrupamento. A performance é abordada profundamente por outros trabalhos como (SILVA; SANTOS, 2011) .

A implementação do OpenLayers utiliza a técnica de grade. Um parâmetro que esse algoritmo pode receber é o tamanho das células da grade. Nesse sentido, foi permitido ao usuário configurar o tamanho dessa grade melhorando a visualização dos pontos no mapa. O recurso apresentado na Figura 5.8 mostra claramente a situação na qual o lado esquerdo evidencia o agrupamento com o tamanho da célula em 35px e o lado direito o tamanho da célula em 75px.

Figura 5.8 – Tamanho de agrupamento



5.4 Camada de Implementação dos dados

A implementação do modelo foi realizada no SGBD (Sistema Gerenciador de Banco de dados) PostgreSQL. Esse SGBD já é utilizado pelos grupos de pesquisa dos experimentos. Além disso, o mesmo é um banco de dados *open source* e possui uma extensão espacial o PostGIS, que permite a utilização de diversos tipos de dados espaciais adicionais além de funções para manipulação desses objetos. Já os processos de carga de dados foi utilizando *scripts* específicos para a implementação descrito na seção 5.4.2.

5.4.1 SGBD e extensão espacial

A seleção do banco de dados se baseou em um estudo da viabilidade e necessidade para a implementação do sistema. Optou-se por analisar, apenas tecnologias *open source* ou gratuita. A busca pelos SGBDs *open source* reconhecidos que foram encontradas foi: Firebird, MySQL, HSQL, SQLite e PostgreSQL.

Alguns critérios pessoais foram utilizados para a escolha do SGBD:

- quantidade de documentação;
- língua da documentação;
- experiência na utilização;
- tipo de licença; e
- suporte a dados geográficos.

Em consideração a estes critérios pessoais e testes realizados o SGBD PostgreSQL foi o escolhido. O principal fator foi a experiência de utilização e o suporte a dados geográficos. Referente à experiência de utilização, outro fator de escolha foi que ambos o GRID e o ISAM utilizam o SGBD PostgreSQL em suas aplicações.

Em se tratando de dados geográficos além do suporte nativo a dados geográficos o PostgreSQL permite instalar a extensão Postgis que adiciona outras funcionalidades. Atualmente, a empresa Refrations Research Inc¹⁶ mantém a equipe de desenvolvimento dessa extensão, que segue as especificações da SFSSQL(*Simple Features Structured Query Language*) (QUEIROZ; FERREIRA, 2005).

Os bancos de dados para os experimentos foram criados em um servidor com o sistema operacional Linux CentOS 5.9. A versão do PostgreSQL foi a versão 8.4 enquanto que a versão do Postgis utilizada foi a 1.5.

¹⁶ Site institucional: <http://postgis.refrations.net>

5.4.2 Extraction Transform Load (ETL)

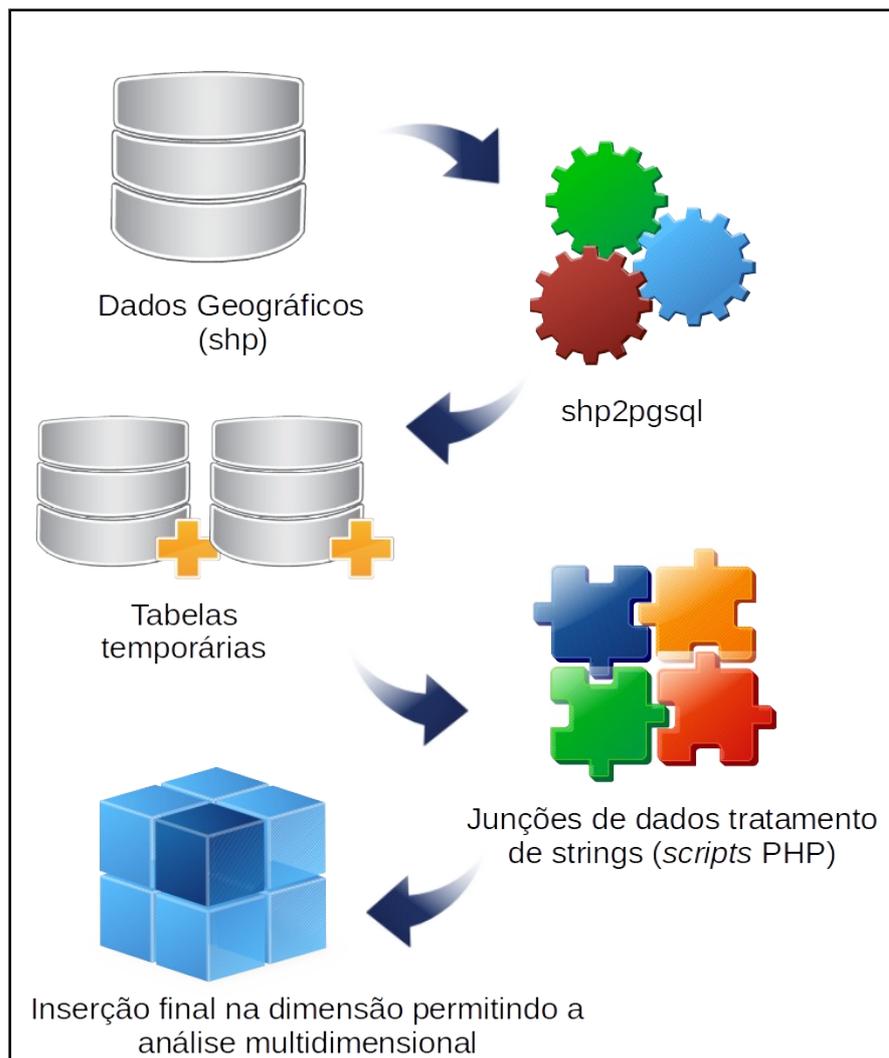
O processo de ETL é um passo fundamental na implementação de um *data warehouse*. Pela característica do modelo de dados não ser normalizado e pelo mesmo possuir redundância de dados, um controle mais rigoroso deve ser feito na hora de carregar os dados (Singh e Singh, 2010).

Para os experimentos e para a avaliação do modelo o processo de ETL foi realizado a partir de fontes distintas. Isso exigiu um critério maior na avaliação da carga dos dados. Apesar de utilizar um fluxo de dados simples o processo de carga precisou escolher dados cartográficos o mais homogêneos e fidedignos possível.

Os dados geográficos muitas vezes são disponibilizados através de *web services* ou por downloads em sistemas de arquivos com metadados ricos ou não. Um exemplo de sistema de dados espaciais organizados com metadados confiáveis é a Infraestrutura Nacional de Dados Espaciais (INDE) disponível em <http://www.inde.gov.br/>. A INDE foi instituída em 2008 e consiste segundo (BRASIL, 2008):

Conjunto integrado de tecnologias; políticas; mecanismos e procedimentos de coordenação e monitoramento; padrões e acordos, necessário para facilitar e ordenar a geração, o armazenamento, o acesso, o compartilhamento, a disseminação e o uso dos dados geoespaciais de origem federal, estadual, distrital e municipal.

A maioria dos dados geográficos disponibilizados em formato vetorial são disponibilizados em *shapefiles*. O formato *shapefile* da Esri® serve para armazenar a posição, formato e atributos de feições geográficas. O *shapefile* na realidade é armazenado como um conjunto de arquivos relacionados e contém uma classe de feição (ESRI, 2014). Os arquivos utilizados para o carregamento de dados são os de formato *.shp, *.shx e *.dbf. As etapas de carregamento dos dados podem ser visualizadas no esquema da Figura 5.9.

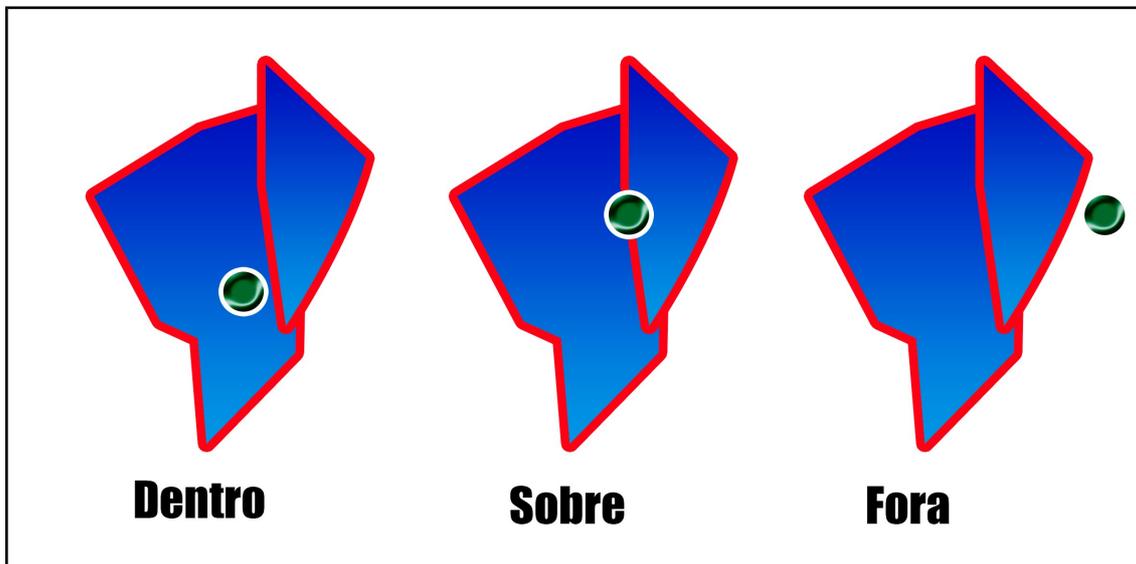
Figura 5.9 – Carregamento de dados do tipo *shape*

O PostGIS acrescenta recursos que permitem trabalhar com projeções e armazenar e converter geometrias para diversos formatos, além de possibilitar consultas geográficas. No processo de ETL a função `ST_CONTAINS` (que verifica se um objeto geográfico está contido em outro objeto geográfico) foi essencial para implementar a classificação dos pontos. Sendo assim, para popular a tabela de fatos, realiza-se o seguinte processo:

1. cadastra-se todos os pontos na tabela de fatos com os ids = 1. Deve-se inserir a medida geográfica coordenada também.
2. Para cada dimensão verificar se o ponto esta inserido na área delimitada por um polígono que vai representar um país um estado, ou uma cidade, ou outra área que represente uma forma de classificar o ponto.

Um problema na carga dos dados é a relação intrínseca que os objetos geográficos possuem (EGENHOFER et al., 1992). Um exemplo é verificar quando a coordenada está sobre uma linha por exemplo exatamente na divisa entre 2 municípios. Para solucionar essa exceção, o *script* armazena as possíveis escolhas para que o usuário no final possa determinar em qual valor de dimensão classificar. Caso o usuário não queira interferir no processo o *script* fornece uma opção para escolha automática. A escolha automática leva em consideração a ordem de avaliação dos polígonos. Essa exceção é conhecida como uma das relações entre as geometrias, conforme mostra a Figura 5.10.

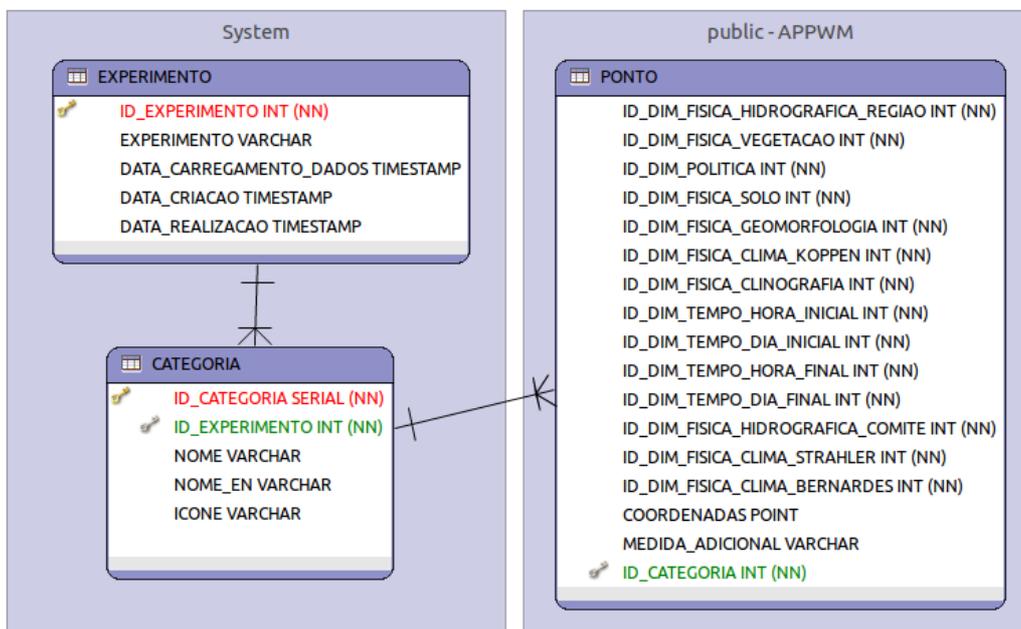
Figura 5.10 – Relações entre polígonos e pontos



5.5 Implementação dos experimentos

Com vistas a evitar a necessidade de replicação de dados, geração de aplicações personalizadas e reprocessamento de ETL foram adicionadas duas tabelas ao modelo, que ficaram armazenadas no *schema* “system”. A tabela *experimento* possui a descrição de cada experimento, enquanto que a tabela *categoria* apresenta cada camada do *web map* de cada experimento. A relação entre as tabelas juntamente com a tabela de fatos pode ser vista na Figura 5.11. Dessa forma os experimentos tornaram-se dinâmicos e de fácil replicação para demais domínios.

Figura 5.11 – Modelo de dados para Experimento



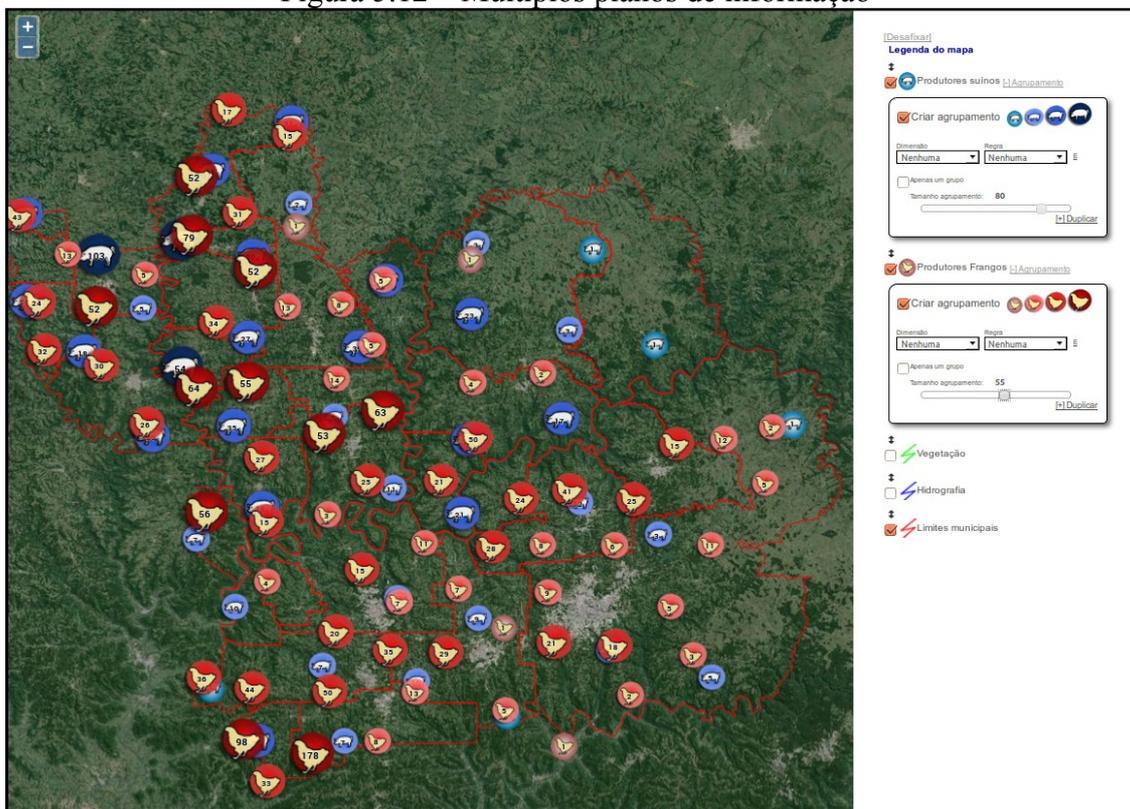
Com esse modelo a instância do experimento é informada para o servidor que seleciona as categorias de cada experimento formando assim as camadas no *web map*. Após o mapa formado, o sistema permite navegação livre ao usuário. Dessa forma o usuário pode explorar as funcionalidades oferecidas pela ferramenta assim como tomar decisões e gerar mapas.

5.6 Resultados da implementação

Nesta seção foram adicionados os resultados da implementação que não puderam ser evidenciados nas seções deste capítulo.

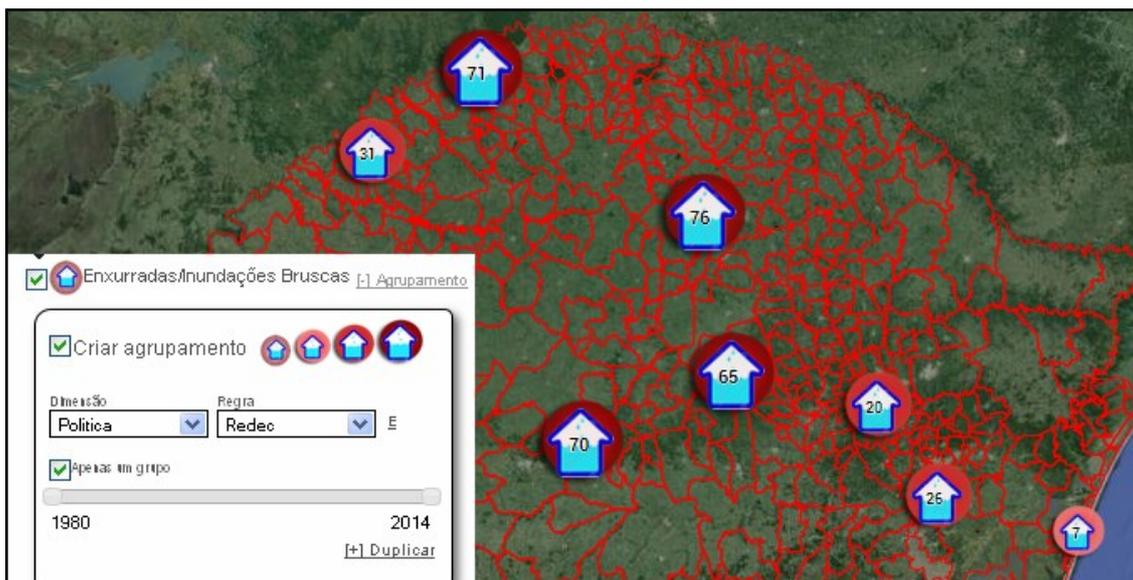
A primeira funcionalidade apresentada é a própria possibilidade de agrupamento a partir de um plano de informação com uma densidade de informação. Ao criar o agrupamento a ferramenta utiliza a proximidade e gera os grupos. A segunda funcionalidade destacada é a possibilidade de visualizar múltiplos planos de informação. Ambas funcionalidades podem ser observadas na Figura 5.12.

Figura 5.12 – Múltiplos planos de informação



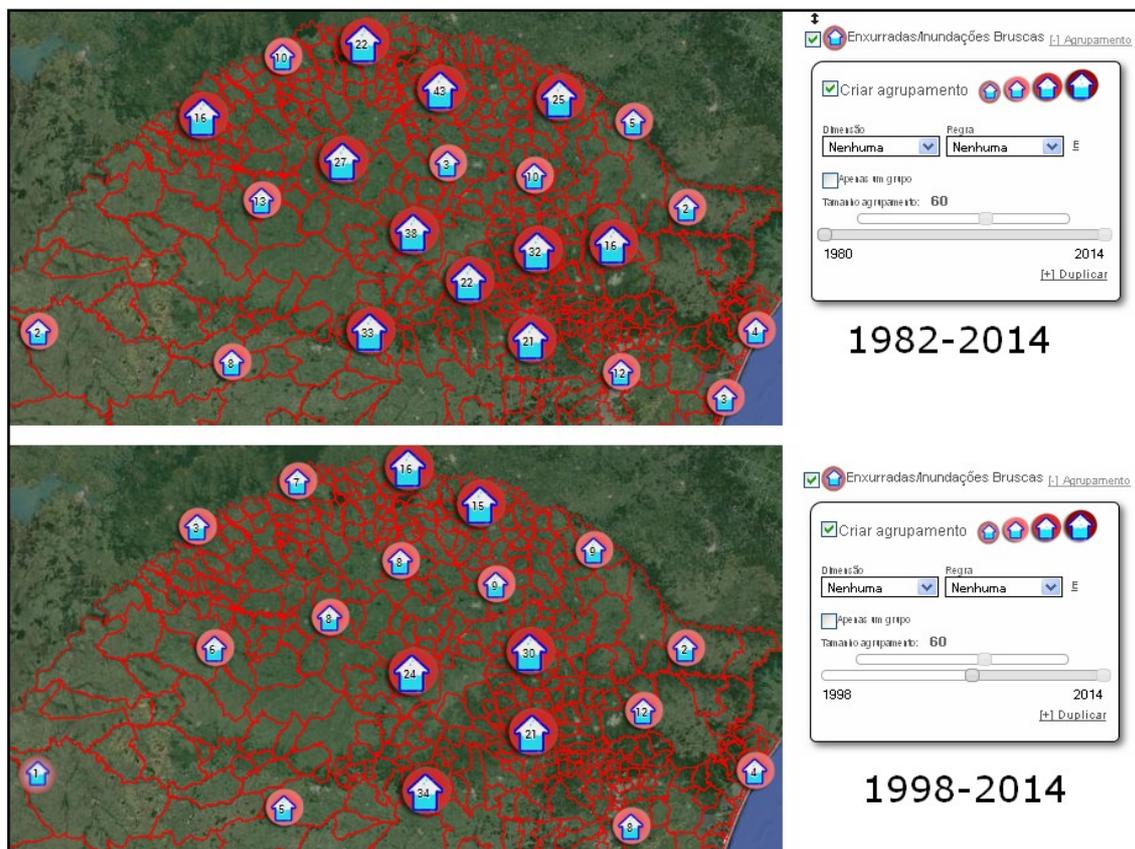
Para realizar contagens rápidas pode se seleccionar a opção “apenas um grupo” conforme apresenta a Figura 5.13. Essa funcionalidade é utilizada somente em situações onde se escolha uma dimensão e, respectivamente, uma regra para agrupamento.

Figura 5.13 – Funcionalidade apenas um grupo



Para a exploração de funcionalidades temporais foi desenvolvido e adicionado no APPWM o componente de tempo. Esse componente possui dois *sliders* possibilitando assim visualizar os dados em intervalos de tempo diferentes. Essa funcionalidade pode ser vista na Figura 5.14.

Figura 5.14 – Exploração temporal no APPWM



5.7 Considerações sobre a implementação

A implementação também se caracteriza como contribuição do trabalho no que tange a técnicas elencadas e apresentadas para montar *web maps* com a capacidade de agrupamento de pontos. As ferramentas e bibliotecas aqui apresentadas permitem, mais facilmente compreender a implementação para que o APPWM possa ser reproduzido e estendido de acordo com a necessidade de cada domínio.

A reprodução do trabalho em outras instâncias trazem contribuições que podem ser substanciais para o modelo base. A evolução do modelo permite que o sistema se adapte a novas situações e permita ao usuário novas consultas mais adequadas as suas necessidades. Além disso permite suprir a necessidade de diferentes usuários anteriormente não contemplados.

6 EXPERIMENTOS

Os experimentos foram elaborados visando dois aspectos principais, sendo o primeiro validar a aplicação do modelo multidimensional e se aplicação está de acordo com as necessidades dos usuários do modelo. O procedimento escolhido para avaliar o software foi por meio da realização de tarefas e consultas no modelo e após aplicação de questionários sobre as impressões dos usuários sobre aspectos que o modelo deve subsidiar.

A complexidade da validação de um modelo transcende os experimentos com os usuários. Um modelo é uma representação da realidade e esta representação pode possuir visões diferentes de acordo com o usuário. A validação da aplicabilidade do modelo é comprovada pela aplicação do mesmo em mais que um caso de uso em domínios diferentes. Para verificar se o modelo atendeu a necessidade de consultas e consumo de informação pelos usuários foram projetados algumas atividades para serem empregadas em cada um dos experimentos.

Cada um dos experimentos teve um objetivo em particular. O experimento da seção 6.2, realizado com a equipe do Instituto de Saneamento Ambiental (ISAM), visa definir as funcionalidades na ferramenta de consulta assim como uma validação do modelo. O experimento da seção 6.3, realizado em conjunto com o Grupo de Gestão de Riscos e Desastres Naturais (GRID), visou comprovar a aplicação do modelo em um outro domínio de aplicação comprovando a possibilidade de expansão e adaptabilidade do modelo.

O capítulo além dos experimentos possui duas outras seções: seção 6.1 - visando apresentar a metodologia para a construção e a avaliação dos experimentos; seção 6.4 - encerra o capítulo com as considerações sobre os resultados obtidos nos experimentos.

6.1 Configurações dos Experimentos

Nesta seção objetiva-se documentar o processo para o desenvolvimento dos experimentos. Os experimentos foram configurados com um mapa base para cada domínio fornecendo as camadas mais representativas para os usuários usarem o APPWM.

Para avaliar as funcionalidades foram usadas as normas ISO/IEC 9126 e ISO/IEC 14598. A metodologia é detalhada na seção 6.1.1, as métricas adotadas para avaliação são tratadas na seção 6.1.2.

6.1.1 Metodologia

A série de normas NBR ISO/IEC 14598 apresenta métodos para medição, julgamento e avaliação da qualidade de produtos de software. No processo de “Estabelecer os requisitos de Avaliação” os modelos de qualidade escolhidos com base na ISO 9126 foi a funcionalidade. A seleção das métricas de avaliação estão descritas na seção 6.1.2 que visa avaliar uma característica de qualidade interna da ferramenta. O plano de avaliação adotado está apresentado na seção 6.1.1.1, enquanto que cada uma das avaliações está definida em seus respectivos experimentos: experimento ISAM (seção 6.2), experimento GRID (seção 6.3).

6.1.1.1 Plano de avaliação

Para avaliar a capacidade do usuário em usar o modelo optou-se por usar uma técnica de questionários visto que poderia dessa forma, obter além do perfil do mesmo o *feedback* real da aplicação. A realização do experimento assim como a aplicação dos questionários utilizou a seguinte estratégia:

1. o usuário entra no sistema e é apresentado o treinamento (6.1.1.2);
2. ao iniciar o experimento o usuário preenche um formulário com dados pessoais, a fim de montar um perfil do usuário do sistema (6.1.1.3);
3. após o usuário deve realizar alguns procedimentos no sistema conforme o domínio do mesmo;
4. ao terminar a sequencia de perguntas o usuário clica em finalizar experimento e responde as perguntas do questionário de análise de funcionalidade (6.1.1.4).

Outra medida adotada para avaliar o usuário foi o tempo de realização do treinamento e do experimento. Sendo assim, o sistema gera *timestamps*: ao entrar no sistema; ao iniciar o experimento e; ao finalizar o experimento. O tempo para responder a cada questionário também é salvo.

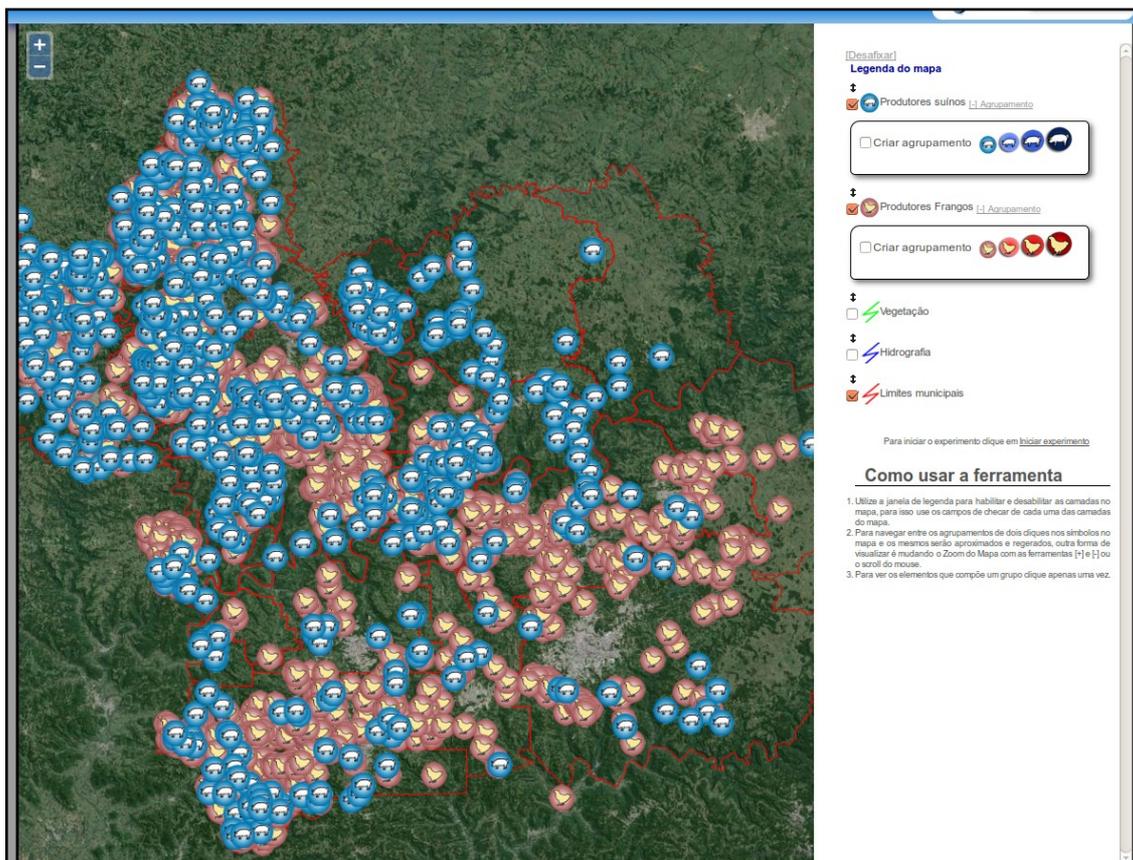
6.1.1.2 Treinamento

A fase de treinamento apresenta um menu auxiliar com algumas tarefas que podem ser feitas pelo usuário para entender o funcionamento da ferramenta. Nesta etapa

deixa-se a ferramenta livre para ser explorada pelo usuário. A única métrica que fica registrada é o tempo que o usuário permanece nessa fase. A fase de treinamento utiliza alguns artefatos para uma melhor compreensão do usuário do sistema.

O primeiro artefato para o treinamento é textual e fica disponível junto da legenda, como se pode observar na Figura 6.1.

Figura 6.1 – Ajuda e treinamento com a ferramenta



A transcrição completa do texto está apresentada a seguir:

A fase de treinamento objetiva familiarizá-lo com a técnica a ser avaliada. Você deve explorar livremente a ferramenta. Para ter mais controle sobre a ferramenta recomenda-se os seguintes passos:

1. Utilize a janela de legenda para habilitar e desabilitar as camadas no mapa. Para isso, use os campos de checkar de cada uma das camadas do mapa.
2. Para navegar entre os agrupamentos, dê dois cliques nos símbolos no mapa e os mesmos serão aproximados e regeados, outra

forma de visualizar é mudando o *Zoom* do Mapa com as ferramentas [+] e [-] ou o *scroll* do *mouse*.

3. Para ver os elementos que compõem um grupo, clique apenas uma vez.

Na fase de treinamento, se observar algum problema que afete a sua interação e seu entendimento, por favor, utilize o formulário de sugestão ao fim do experimento.

O segundo artefato produzido é um vídeo de demonstração adicionado na opção ajuda. Com esses artefatos acredita-se que seja possível permitir ao usuário explorar a ferramenta, que não possui recursos diferenciados do que os *web maps* normalmente oferecem.

6.1.1.3 Perfil do usuários

O questionário para definir o perfil do usuário foi baseado no documento do NISTIR 7742 (SCHUMACHER; LOWRY, 2010). Este documento apresenta um formato comum de avaliação para usabilidade tanto na avaliação como na interpretação de resultados. O documento prioriza que o mínimo necessário para uma boa avaliação de ferramenta é de 10 usuários.

Com base nesses itens o questionário para traçar o perfil do usuário foi criado conforme a Figura 6.2.

O SUS é composto por dez afirmações em escala de 5 pontos para que os participantes avaliem seu nível de concordância com o sistema. Metade das questões é de forma positiva e a outra de forma negativa. O SUS é extremamente útil para avaliar uma tecnologia nova em relação a outra existente, sendo realizado com perguntas voltadas para a usabilidade e neutro perante tecnologias. Desta forma pode ser reaplicado ao longo dos anos sem a necessidade de reinventar questionários (BROOKE, 2013).

Devido a essas características o questionário foi adaptado e pensado de acordo com a mesma filosofia, porém voltada para perguntas sobre a funcionalidade. A filosofia do SUS segundo Brooke (2013):

- forte correlação entre todos os itens selecionados para as perguntas;
- número total de perguntas limitadas devido as restrições de tempo para administrar questionários;
- perguntas que facilitem a reflexão sobre a questão permitindo que a resposta flua rapidamente evitando respostas trocadas.

Cada pergunta teve relação com uma funcionalidade proposta nas consultas do roteiro. O questionário com as perguntas adaptadas é apresentada na Tabela 6.1.

Tabela 6.1 – Questionário sobre as funcionalidades do sistema

		Discordo Totalmente				Concordo Totalmente
1.	Eu utilizei a escala de cores e o tamanho dos ícones para identificar a concentração de pontos no mapa.	<input type="checkbox"/>				
2.	Eu achei desnecessário o uso da escala de cores e o tamanho dos ícones para a representação da quantidade de pontos	<input type="checkbox"/>				
3.	Eu considero que os locais de menor/maior concentração da informação são facilmente encontrados.	<input type="checkbox"/>				
4.	Eu encontrei dificuldade de identificar os locais de maior/menor concentração utilizando o agrupamento.	<input type="checkbox"/>				
5.	Eu achei a troca de dimensão e regra do agrupamento adequada às tarefas de exploração de dados que costumo fazer	<input type="checkbox"/>				
6.	Eu achei as dimensões e regras de agrupamento das informações desnecessárias, faltando muitas regras para tornar os agrupamentos úteis.	<input type="checkbox"/>				
7.	Eu considero simples e adequado o agrupamento dos pontos utilizando essa ferramenta, comparativamente com as que utilizo normalmente.	<input type="checkbox"/>				
8.	Eu achei o sistema complexo demais para extrair as informações em comparação com outras ferramentas de meu cotidiano.	<input type="checkbox"/>				
9.	Eu me senti muito confiante/autônomo utilizando o sistema	<input type="checkbox"/>				
10.	Eu preciso aprender muito mais antes de continuar usando esse sistema.	<input type="checkbox"/>				

Além das questões sobre a funcionalidade foi solicitada mais 2 questões abertas: a primeira questão serve para enviar sugestões e dúvidas sobre a ferramenta; a segunda questão solicita ao usuário que informe duas consultas que gostaria de poder realizar sobre os dados.

A segunda questão visa obter o *feedback* do usuário no sentido de ver as necessidades que o mesmo tem sobre os dados e de que forma o modelo pode suprir ou não essas questões.

6.1.2 Métricas de avaliação

Foram montadas tarefas que são consideradas cotidianas para o usuário e que usando o modelo multidimensional podem ser realizadas com alguns cliques. Essas tarefas visam avaliar a funcionalidade da ferramenta e se essa atinge os requisitos propostos para solucionar o problema da visualização da densidade de informação.

6.1.2.1 Casos de sucesso e casos de falha

Para cada experimento foi apresentado os casos de sucesso e de falha. Isso teve como objetivo verificar se os objetivos definidos prioritariamente foram atingidos. Os casos de falha foram documentados a fim de permitir melhorias e refinamentos na proposta apresentada.

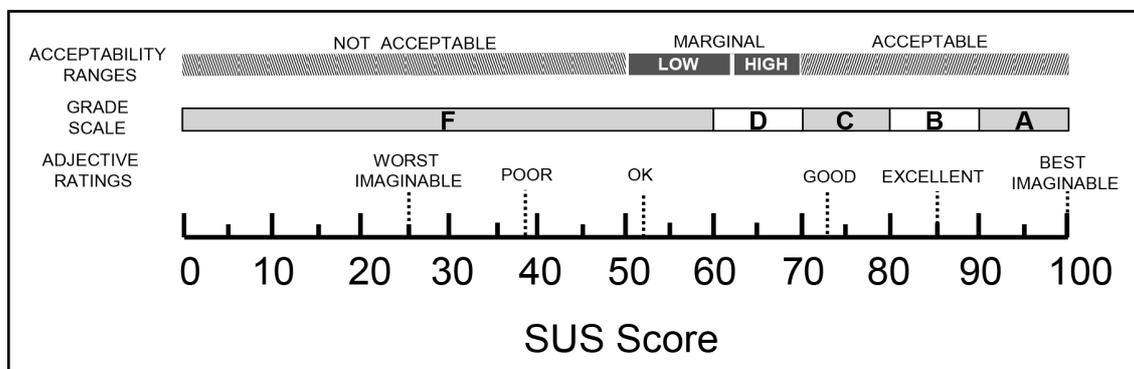
Quanto a realização da tarefa os estudos de sucesso e fracasso são contabilizados da seguinte forma: nota 2 (dois) para sucesso sem necessidade de auxílio no experimento; nota 1 (um) para realização das atividades com necessidade de auxílio; e 0 (zero) nos casos onde não conseguiram realizar os experimentos, descartando a avaliação.

Para avaliação do questionário objetivo usou o mesmo método do SUS. O cálculo do SUS segundo (BROOKE, 1996) tradução nossa é:

A contribuição de cada item de pontuação varia de 0 a 4. Para os itens 1, 3, 5, 7 e 9 (questões positivas), a contribuição da pontuação é a posição da escala marcada pelo participante menos um. Para os itens 2, 4, 6, 8 e 10 (questões negativas), a contribuição equivale a 5 subtraído da posição marcada. Multiplica-se a soma das pontuações por 2,5 para obter a pontuação geral do SUS.

O trabalho de Bangor et al. (2009) apresenta uma média para a escala do SUS sendo que esta classificação pode ser vista na Figura 6.3. Para fins de comparação usou-se os mesmos intervalos.

Figura 6.3 – Categorias para a escala do SUS



6.2 Instituto de Saneamento Ambiental - ISAM

Com o intuito de avaliar a ferramenta de consulta proposta inicialmente, assim como a aplicabilidade do modelo foi realizado o experimento denominado ISAM. Este experimento foi a base para a realização dos outros dois experimentos. Além de avaliar a ferramenta o experimento visou avaliar algumas das dimensões propostas.

Essa seção é dividida em quatro partes: a descrição do domínio (seção 6.2.1); o roteiro para o domínio (seção 6.2.2); o perfil dos usuários entrevistados nesse experimento (seção 6.2.3) e as discussões sobre os resultados do experimento seção 6.2.4.

6.2.1 Descrição do Domínio

O Instituto de Saneamento Ambiental (ISAM) é um instituto da Universidade de Caxias do Sul. Ele foi criado em 1999 e tem sua origem nas atividades de ensino, pesquisa e extensão desenvolvidas, desde 1987, com ênfase no gerenciamento de resíduos sólidos. Atualmente o ISAM tem sua atuação também voltada para o gerenciamento e planejamento ambiental, o que significa dizer que se trabalha dentro de uma visão de prevenção e minimização de impactos ambientais (ISAM, 2014).

Dentre os projetos de pesquisa se destaca o projeto (PDRS). O projeto visa manter informações dos principais produtores de animais da região do COREDE-Serra (FEE, 2011) que pertence ao nordeste do estado do Rio Grande do Sul. Os produtores

são organizados conforme o tipo de animal produzido e o principal objetivo desses dados é estimar a carga poluidora desses produtores (SCHNEIDER et al., 2007).

Para o experimento foram selecionados os produtores de frangos e os produtores de suínos, sendo que estes totalizavam o maior número de produtores. A área desse experimento é o COREDE-SERRA que compreende uma área de 8087km². Os produtores de frangos compreende um total de 1820 criadores e os produtores de suínos um número 720.

6.2.2 Roteiro do experimento

Para iniciar o experimento¹⁷ o usuário clica no botão experimento. A partir deste momento o tempo e algumas das ações no sistema são gravadas. O objetivo deste experimento é verificar a facilidade com que o usuário consegue utilizar a ferramenta e conseqüentemente explorar o modelo que da suporte a interação com os dados. As etapas do experimento aparecem na legenda do web mapa e são as seguintes:

1. A primeira tarefa é criar o agrupamento dos dois planos de informação (opção criar agrupamento).
2. Vamos explorar um pouco mais as informações: troque a forma de agrupamento do plano de informação de suínos para **dimensão** política → **regra** Município.
3. Troque o tamanho do *agrupamento* (utilize o *slider* -  - de controle: Tamanho do agrupamento).
4. Mude o zoom aproximando e afastando. Ao se aproximar de limites municipais a restrição se tornou útil?
5. Marque para essa camada a opção de grupo único.
6. Mude o *zoom* de forma que seja possível ver toda a área do COREDE-Serra
7. Mude o *zoom* de forma que seja possível ver toda a área do COREDE-Serra. O grupo único permite uma melhor visualização?
8. Clique em um grupo de dados dos suínos (qualquer grupo) e acesse as informações de um produtor específico. Foi possível identificar a que grupo o mesmo pertence?

¹⁷ Link para acesso ao experimento: <http://vbaco01.ucs.br/mestrado/?experimento=2>

9. Duplique o plano de informação (camada) ao qual está aplicado uma regra (opção duplicar no canto da camada);
10. Agrupe as informações pela **dimensão** → “física bacia hidrográfica” comitê e escolha em **regra** → “primeiro nível de unidade de gestão”. É possível visualizar e identificar os grupos de cada uma das regras de agrupamento?

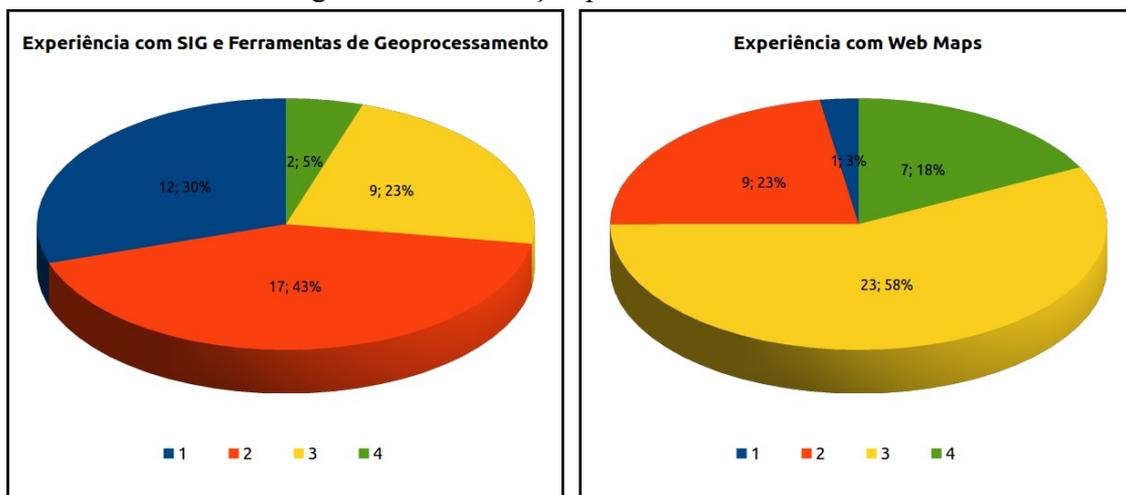
6.2.3 Perfil dos usuários

Um total de 54 usuários foram convidados (por e-mail) para esse experimento. Desses convidados 40 participaram da avaliação do sistema de agrupamento de dados. De acordo com o escore para sucesso e falha este experimento apresentou um escore de 77/108, sendo que 3 usuários necessitaram de apoio para terminar o experimento, 5 usuários desistiram de concluir o experimento e 9 usuários não responderam ao pedido nem solicitaram apoio.

Dentre os usuários, 19 usuários eram do sexo masculino e 21 do sexo feminino. Em relação à faixa etária, 37,5% dos usuários possuíam idade entre 15 a 25 anos, 47,5% entre 26 a 40 anos e 15% entre 41 a 55 anos. No tocante à profissão, os usuários são estagiários, bolsistas e professores, sendo que suas formações são predominantemente Engenharia Ambiental/química (15), Biologia(8), cursos nas áreas de tecnologia da informação (7), cursos na área da saúde (4), geografia(1) e artes plásticas (1) .

Em relação à experiência com ferramentas de GeoProcessamento e/ou SIGs, os avaliadores atribuíram-se um percentil, onde a escala varia de 1 (um) em que a experiência é inexistente a 4 (quatro), em que experiência com essa ferramenta é elevada. O gráfico da esquerda na Figura 6.4 apresenta as graduações indicadas pelos avaliadores. Já o gráfico da direita da Figura 6.4 apresenta à experiência com ferramentas *web maps* seguindo a mesma escala (1 a 4).

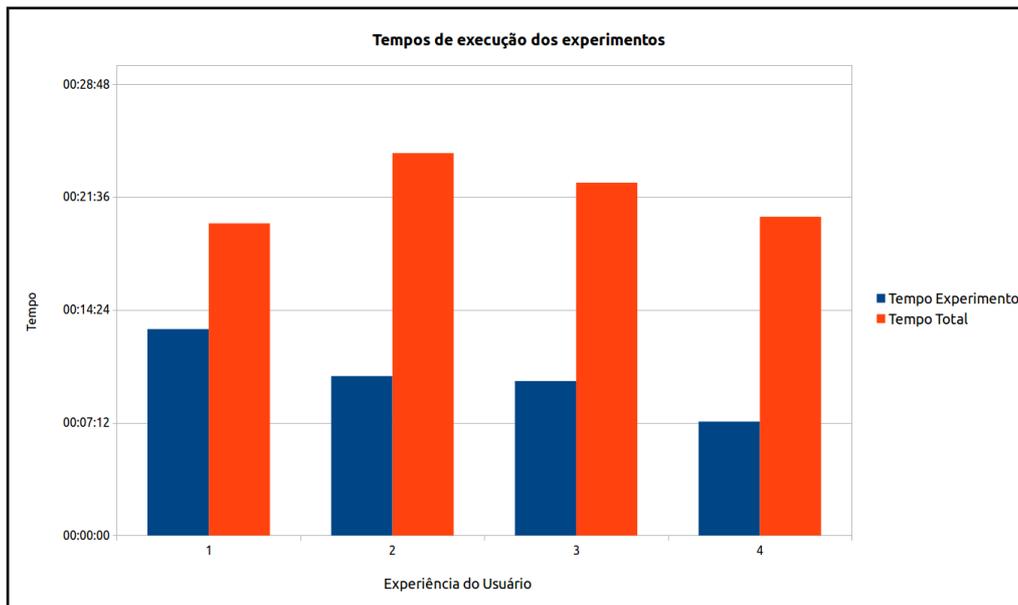
Figura 6.4 – Descrição perfil usuários ISAM



Referente ao tempo de execução dos experimentos, esses foram organizados da seguinte maneira: o tempo total de execução – diferença entre o *timestamp* do horário de entrada no APPWM e término de preenchimento do segundo formulário; e tempo de realização da atividade proposta – diferença entre o *timestamp* entre o término de preenchimento do formulário de identificação e o início do preenchimento do formulário de avaliação. Com essas duas medidas, o tempo médio de demora para a realização do experimento foi de 00:09:03 (nove minutos e trinta e três segundos), enquanto que o tempo médio em que o usuário ficou explorando a ferramenta foi de 00:22:29 (vinte e dois minutos e vinte e nove segundos).

O gráfico da Figura 6.5 apresenta as médias de tempo de acordo com o nível de conhecimento com softwares e informática apresentado pelos respondentes. Como pode ser visto nessa mesma figura as pessoas que possuíam mais conhecimento em informática executaram o experimento com menor tempo, mas dispensaram para a ferramenta um tempo mais significativo.

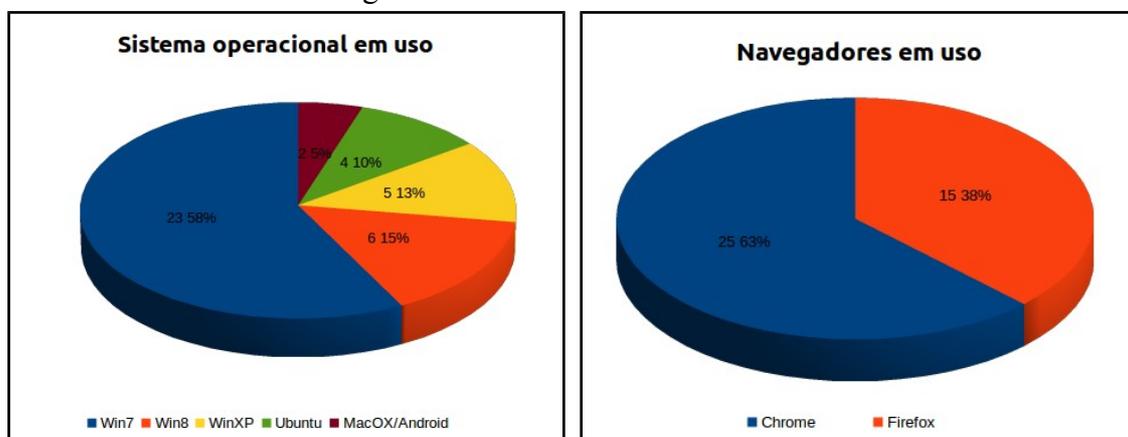
Figura 6.5 – Tempo médio no experimento e tempo médio na ferramenta



Além do tempo e do perfil dos usuários foi capturada a plataforma (sistema operacional e navegador, além da resolução em que o experimento foi executado) que os mesmos utilizavam. Essas informações são úteis para se ter um diagnóstico do perfil específico do usuário e em quais desses sistemas operacionais e navegadores o APPWM deve ser testado.

Os resultados referente à plataforma podem ser vistos nos gráficos da Figura 6.6: o gráfico da esquerda apresenta os sistemas operacionais – com predominância do Windows 7; e o gráfico da direita dos navegadores utilizados – Chrome (Predominante) e Firefox.

Figura 6.6 – Plataforma dos usuários



Quanto às resoluções estas variavam muito, sendo que a Tabela 6.2 apresenta uma contagem das resoluções mais comuns e a quantidade em que apareceram.

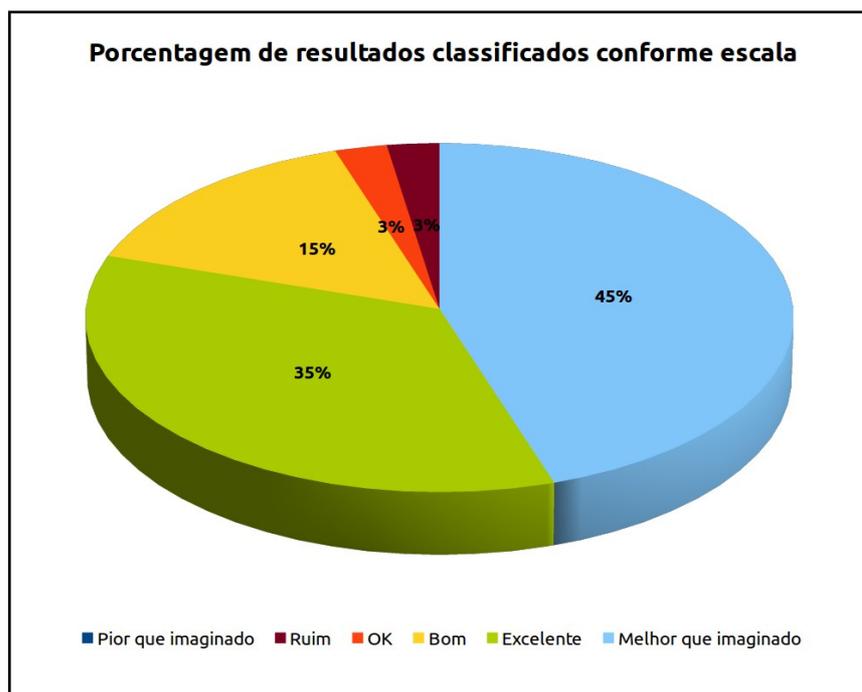
Tabela 6.2 – Resoluções de tela dos participantes do experimento no ISAM

Resolução	Quantidade
1366x768	15
1280x1024	5
1920x1080	5
1600x900	3
1440x900	3
1024x768	2
1280x800	2
1360x768	3
2400x1350	1
360x640	1
Total	40

6.2.4 Resultados obtidos

Dos 40 usuários que completaram todas as etapas do experimento obteve-se, através do cálculo a nota média de 79,26 comparando a escala proposta por Bangor et al. (2009) a aplicação se enquadrou em “Excelente”. A classificação em excelente também é visível na distribuição das notas, inclusive que apresentou mais resultados “melhor do que imaginado” do que “excelente”, que é apresentada no gráfico da Figura 6.7.

Figura 6.7 – Resultados de avaliação através do questionário



Dentre as consultas sugeridas pelos respondentes 4 apresentaram ser possível realizar com as restrições apresentadas, 7 usuários não sugeriram consultas, 10 usuários se focaram em totais que podiam ser gerados com o botão “apenas um grupo”. Algumas consultas que poderiam ser realizadas são apresentadas na sequência:

Uma possibilidade adicional pode ser a consulta a partir da seleção de uma área geográfica específica (através de uma ferramenta de desenho vetorial). Nesse caso, o sistema deve apresentar os resultados dos filtros selecionados pelo usuário dentro dos limites geográficos por ele desenhados. Esta funcionalidade é útil em casos de alta distribuição geográfica das informações de interesse, as quais podem pertencer a limites distintos, e portanto, necessitam ser agrupadas.

Mostrar produtores com quantidade de animais dentro de uma faixa informada.

Gostaria de saber o bairro ou província pertencente de cada ícone e quantidade aproximada de suínos ou frangos se tem em cada região.

Das 3 consultas que recebem destaque a primeira se apresenta como uma funcionalidade que a ferramenta pode apresentar em um trabalho futuro. Essa funcionalidade apresentaria uma grande vantagem em termos de performance visto que reduziria o espaço e depois as regras seriam aplicadas a esse espaço.

A segunda consulta que recebeu destaque apresenta uma medida (quantidade de animais) que deve ser adicionada na tabela de Fato. Portanto assim como as demais consultas sugeridas se esta se enquadra especificamente para o domínio. Devido o modelo ser expansível a funcionalidade para agrupar por fato pode ser desenvolvida na ferramenta. Apenas essa consulta específica do domínio recebeu destaque visto que seria necessário adicionar como medida e não como dimensão específica do domínio.

A terceira consulta se apresenta como um problema no carregamento de dados para o modelo. Como a região não possui polígonos em bases abertas para representar os bairros e províncias então não é possível liberar essa busca. Quanto à quantidade a mesma estaria disponível e caso o usuário quisesse realizar a contagem poderia selecionar a opção “Apenas um grupo”.

6.3 Grupo de Gestão de Riscos e Desastres Naturais – GRID

O experimento do GRID recebe destaque por explorar dados temporais e sobreposição temporal de informação. A organização dessa seção também obedece a seguinte estruturas: a descrição do domínio (seção 6.2.1); o roteiro para o domínio (seção 6.2.2); o perfil dos usuários entrevistados nesse experimento (seção 6.2.3); e as discussões sobre os resultados do experimento (seção 6.2.4).

6.3.1 Descrição do domínio

O grupo de Gestão de Riscos de Desastres – GRID é formado por uma equipe multidisciplinar, que hoje conta com pesquisadores das áreas de Engenharia Civil, Engenharia Ambiental, Geografia, Química, Arquitetura, Estatística, Biologia, Serviço Social e Comunicação Social, que estudam temas relacionados aos riscos de desastres. O grupo foi criado em 2009, dentro do programa de Pós-Graduação em Engenharia Civil da Universidade Federal do Rio Grande do Sul e entre as linhas de atuação estão o mapeamento de áreas de risco, análise de ocorrências (de desastres naturais, incêndios, acidentes com cargas perigosas), estudo de soluções para minimizar prejuízos humanos e materiais através do uso de medidas estruturais e não-estruturais (GRID, 2014).

Os dados que foram considerados no experimento são uma compilação dos desastres ocorridos no Rio Grande do Sul desde 1982. Os dados originalmente estavam em um banco de dados relacional.

Uma regra detectada neste experimento foi o REDEC (Coordenadorias Regionais de Defesa Civil). No caso do Rio Grande do Sul estas são divididas em 11. Para esse experimento o modelo foi modificado para conter mais essa regra na dimensão política.

6.3.2 Roteiro do experimento

O roteiro desse experimento apresentou, além do agrupamento de pontos de acordo com o contexto do usuário, a exploração temporal dos dados¹⁸.

1. Crie o agrupamento * da camada de informação das estiagens.
2. Crie o agrupamento * da camada de informação das inundações.
3. Utilize a ferramenta de *zoom* (*scroll* do *mouse*, ou [+] [-] no canto esquerdo superior) para aproximar e afastar os grupos.
4. Vamos explorar as informações! Troque a regra de agrupamento.
 - Para a camada de estiagens utilize a **Dimensão** ⇒ Política e a **Regra** ⇒ REDEC
 - Para a camada de inundações utilize a **Dimensão** ⇒ Física Solo e a **Regra** ⇒ Grande Grupo
5. Acesse as informações dos eventos que estejam em um grupo. Para isso, clique em um agrupamento.
6. Escolha a opção **apenas um grupo** para contabilizar de acordo com os REDECs as estiagens.
7. Vamos explorar um pouco mais os eventos temporais!
 - Limite os anos através do *slider* temporal no plano de informação das estiagens, para que apresente apenas a última década.

¹⁸ Link para acesso ao experimento: <http://vbaco01.ucs.br/mestrado/?experimento=2>

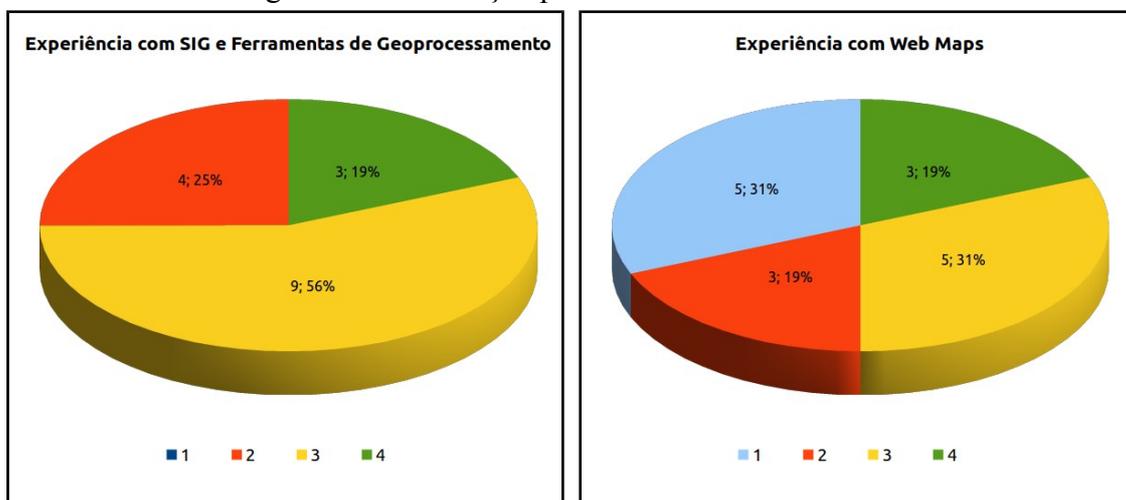
6.3.3 Perfil dos usuários

Para esse experimento foram convidados 31 usuários. Participaram da avaliação do sistema de agrupamento de dados 16 avaliadores. Este experimento teve uma menor adesão por parte dos usuários os quais precisaram ser lembrados muitas vezes da necessidade de realização do experimento. De acordo com o escore para sucesso e falha este experimento apresentou um escore de 31/62, sendo que 1 usuário necessitou de apoio para terminar o experimento, 7 usuários desistiram de concluir o experimento e 14 usuários não responderam ao pedido nem solicitaram apoio.

Os 16 usuários estavam divididos segundo o sexo em: 7 do sexo masculino e 9 do sexo feminino. Em relação à faixa etária, 46,7% dos usuários possuíam idade entre 15 a 25 anos, 33,3% entre 26 a 40 anos e 20% entre 41 a 55 anos. No tocante à profissão, os usuários são estagiários, bolsistas e professores, sendo que suas formações são predominantemente: Engenharia civil(9); Engenharia Ambiental/química (3); cursos nas áreas de tecnologia da informação (2); e arquitetura (2) .

Em relação à experiência com ferramentas de GeoProcessamento e/ou SIGs, os avaliadores atribuíram-se um percentil, onde a escala varia de 1 (um) em que a experiência é inexistente a 4 (quatro), em que experiência com essa ferramenta é elevada. O gráfico da esquerda na Figura 6.8 apresenta as graduações indicadas pelos avaliadores. Já o gráfico da direita da Figura 6.8 apresenta à experiência com ferramentas *web maps*, a apresenta a autoavaliação realizada pelos usuários à respeito de suas experiências, seguindo a mesma escala (1 a 4).

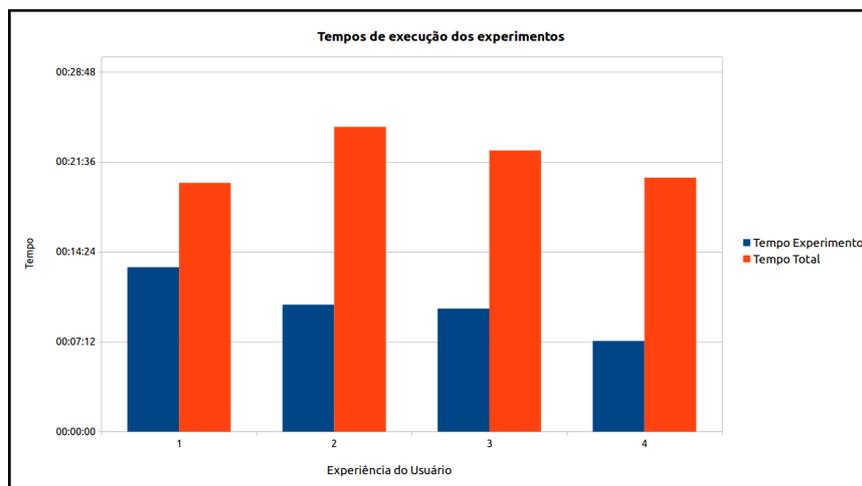
Figura 6.8 – Descrição perfil dos usuários do GRID



Para o experimento do GRID as medidas de tempo médio para a realização do experimento foram inferiores ao ISAM sendo de 00:07:46 (sete minutos e quarenta e seis segundos), enquanto que o tempo médio em que o usuário ficou explorando a ferramenta ficou em 00:18:37 (dezoito minutos e trinta e sete segundos).

O gráfico da Figura 6.9 apresenta as médias de tempo de acordo com o nível de conhecimento com softwares e informática apresentado pelos respondentes.

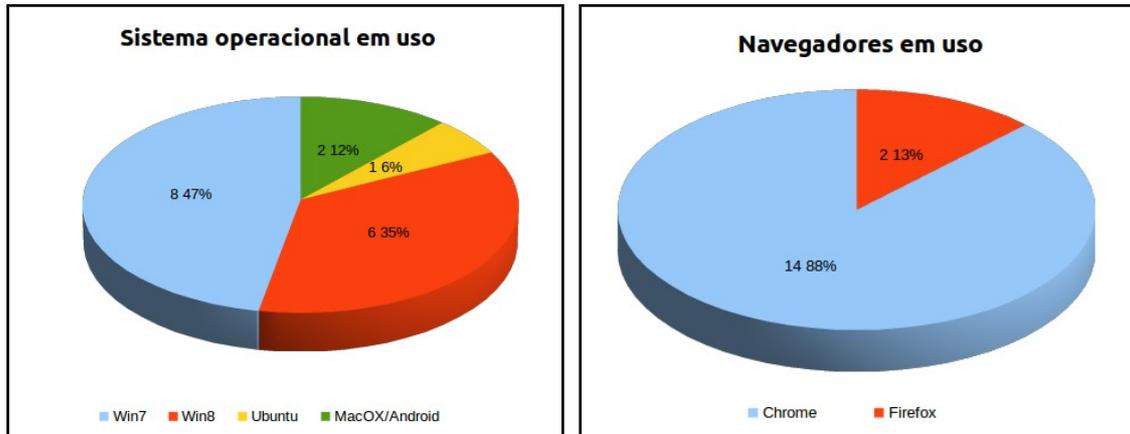
Figura 6.9 – Tempo médio no experimento e tempo médio na ferramenta



Além do tempo e do perfil dos usuários foi capturada a plataforma (sistema operacional e navegador, além da resolução em que o experimento foi executado) que os mesmos utilizavam. Essas informações são úteis para se ter um diagnóstico do perfil específico do usuário e em quais desses sistemas operacionais e navegadores o APPWM deve ser testado.

Os resultados referente a plataforma podem ser vistos nos gráficos da Figura 6.10: o gráfico da esquerda apresenta os sistemas operacionais – com predominância do Windows 7; e o gráfico da direita dos Navegadores utilizados – Chrome (Predominante) e Firefox.

Figura 6.10 – Plataforma dos usuários



Quanto as resoluções estas variaram muito, a Tabela 6.3 apresenta uma contagem das resoluções mais comuns e a quantidade em que apareceram.

Tabela 6.3 – Resoluções de tela dos participantes do experimento no GRID

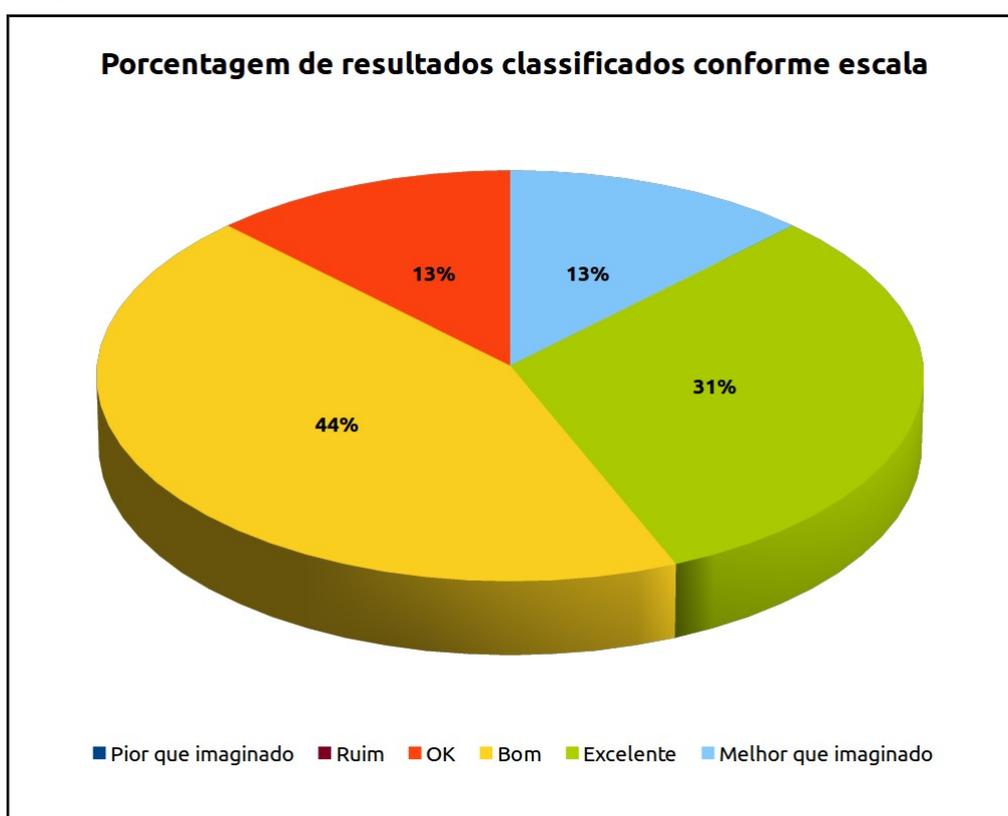
Resolução	Quantidade
1366x768	7
1920x1080	4
1152x864	2
1024x600	1
1280x800	1
1360x768	1
Total	16

6.3.4 Resultados Obtidos

Dos 16 usuários que completaram todas as etapas do experimento obteve-se através do cálculo a nota média comparando a escala proposta por Bangor *et al.* (2009). A aplicação se enquadrou em “bom”, sendo que a média ficou em 68,9.

A distribuição das notas conforme a escala de Bangor et al., (2009) é apresentada no gráfico da Figura 6.11.

Figura 6.11 – Distribuição das notas do experimento realizado com o GRID



Das consultas sugeridas, algumas que merecem destaque e poderiam ser realizadas, com alguns ajustes, são apresentadas na sequência:

- 1 – Seria interessante inserir o modelo digital de elevação. Também seria legal poder realizar consultas simples, como por exemplo: selecionar os eventos que ocorreram na REDEC 9 nos anos de 1990 a 2003.
- 2 – Quantidade de chuva acumulada num certo período Registro de ventos em estações próximo a eventos de ventanias, ciclones, etc Ou seja, dados climatológicos que me ajudassem a entender os eventos.
- 3 – Gostaria de poder baixar as informações em tabela ou txt. tabela de descrição do que é cada regra.

4 – Acho que poderia ter uma exibição de um polígono da regra para entender um pouco melhor o que foi agrupado.

5 – 1. combinar locais de estiagem e inundação com base no mesmo tipo de vegetação e relevo. 2. locais de danos por vendaval divididos por regra política e por divisão de renda per capita

Da primeira consulta pode se verificar que o modelo digital de elevação pode ser obtido como um WMS (*Web Map Service*), e a segunda consulta denominada “simples” pode ser obtida pelo modelo e inclusive pela ferramenta. Para realizar essa consulta pode se configurar os slides de tempo nesse intervalo, agrupar os eventos por REDEC e ativar o “botão apenas um grupo”.

Para a segunda consulta seria necessário a integração com outros sistemas para fornecer as médias e utilizar a visualização de múltiplos planos de informação (que a ferramenta já possui) para poder integrar ao sistema.

Para a terceira e quarta consultas apresentam funcionalidades que a ferramenta poderá disponibilizar em um trabalho futuro. Essas funcionalidades permitem ao usuário entender melhor o agrupamento realizado. Essa necessidade significa que as regras foram uteis para o cotidiano dos usuários, e necessitam ser melhor exploradas.

Para a quinta consulta observa-se duas situações distintas: a situação 1 apresenta uma mesclagem de dois tipos de eventos diferentes em um terceiro grupo, após fazer essa junção se poderia escolher as regras; a situação 2 é possível realizar estendendo o modelo com uma dimensão socioeconômica descaracterizada.

6.4 Considerações sobre os resultados obtidos

Para ambos os experimentos observou-se claramente a necessidade que os usuários possuíam de consumir dados geográficos e tomar decisão sobre os mesmos. Dentre as sugestões e consultas sugeridas tiveram informações específicas do domínio e dos dados da aplicação utilizada para o experimento.

Conclui-se que o modelo supre a necessidade de diversas consultas do cotidiano dos usuários sendo que para consultas mais complexas ainda é necessário a utilização de ferramentas específicas de SIG. Para usuários considerados não especialistas em SIG a ferramenta se apresentou de fácil aprendizado e os mesmos realizaram os experimentos com o tempo aproximado dos usuários avançados.

Para ambos os tipos de usuários as consultas que esses necessitavam praticamente foram supridas pelo *web map* do experimento. Isso se apresenta como um aspecto positivo importante, visto que uma das maiores vantagens de um *web map* é ser prático para tarefas que demandem pouco processamento geográfico.

7 CONCLUSÃO

A necessidade de melhorar a visualização de dados espaciais é um problema real e que precisa de atenção e desenvolvimento de novas técnicas nesta área. Este trabalho apresentou uma forma de qualificar a visualização de dados espaciais. Por meio de um modelo multidimensional, que possui em sua essência a simplicidade na consulta. O modelo, completamente extensível é voltado para uma síntese ou expansão da informação.

O modelo proposto viabiliza construção de regras que consideram as características geográficas intrínsecas ao ponto. A ferramenta em conjunto com o APPWM permite gerar grupos de pontos de forma automatizada e por características existentes nas dimensões criando assim regras de agrupamento.

As regras, organizadas por plano de informação, são voltadas à necessidade de tomador de decisão especialista, ou usuário esporádico que necessite consumir estas informações. Portanto, além de uma melhor visualização dos dados em uma área com informação densa, é possível rapidamente através do modelo realizar análises geográficas que necessitam de softwares específicos e complexos que rodam em ambiente *desktop*.

As análises propostas pela ferramenta com um modelo multidimensional permite uma rápida síntese em um ambiente de dados distribuído podendo facilmente ser acessado por qualquer usuário. Os usuários ao carregar novos dados poderão usufruir de mais informações relacionadas que o modelo provê aproveitando o relacionamento intrínseco dos dados.

Por se tratar de informação geográfica qualquer, informação pontual, ou seja, qualquer domínio que utilizar um ponto georreferenciado poderá usar esta técnica. O modelo multidimensional aqui proposto apresenta-se como método genérico para melhorar a visualização de grande número de informação, independente da escala do mapa. É um modelo que permite visualização de dados direcionada à necessidade de exibição de cada aplicação ou de cada usuário. Como o agrupamento tem a possibilidade de respeitar as relações geográficas, o modelo em conjunto com uma ferramenta é voltado ao contexto e não uma mera melhora na visualização dos pontos em um mapa.

7.1 Trabalhos futuros

Devido aos experimentos e a algumas tarefas solicitadas pelos usuários, foram surgindo algumas necessidades que precisam ser levadas em consideração. Para uma melhor organização essas propostas foram divididas em propostas para o modelo e as propostas para a ferramenta APPWM.

As principais propostas para a extensão e melhoria do modelo desenvolvido são:

- explorar as dimensões socioeconômicas com a sua relação perante a temporalidade;
- estudar dimensões que apresentam dados vagos, ou seja, dimensões que não tem o limite bem definido, como uma epidemia;
- Analisar outras formas geográficas como fatos, como por exemplo linhas e polígonos.

As principais propostas para a melhoria da ferramenta APPWM são :

- Apresentar qual o polígono ou linha representa a agregação dos dados;
- implementar um sistema de carregamento otimizado para dimensões descaracterizadas;
- separação da arquitetura da ferramenta, permitindo utilizar apenas o módulo de gestão do modelo ou o módulo de *web mapping*;
- criar uma função que permita agregar pontos de tipos diferentes com (chuvas e secas) em um mesmo plano de informação para que as regras possam ser aplicadas de forma uniforme.

REFERÊNCIAS

- ANA, AGÊNCIA NACIONAL DE ÁGUAS. **Glossario de Termos Hidrológicos**. Brasília: [s.n.], 2001.
- ANA, AGÊNCIA NACIONAL DE ÁGUAS. O Comitê de Bacia Hidrográfica: o que é e o que faz? In: ANA (Ed.); **Cadernos de capacitação em recursos hídricos**. v. 1, p.64, Brasília, 2011.
- ANDERSON, P. S. Perfil e outras representações do Relevo. **PRINCÍPIOS DE CARTOGRAFIA TOPOGRÁFICA**. p.1–14, [S.l.]: Illinois, 2002.
- ANDRIENKO, G.; ANDRIENKO, N.; KEIM, D.; MACEACHREN, A. M.; WROBEL, S. Challenging problems of geospatial visual analytics. **Journal of Visual Languages & Computing**, v. 22, n. 4, p. 251–256. [S.l.]: Elsevier, 2011.
- AUER, T.; MACEACHREN, A. M.; MCCABE, C.; PEZANOWSKI, S.; STRYKER, M. HerbariaViz: A web-based client–server interface for mapping and exploring flora observation data. **Ecological Informatics**, v. 6, n. 2, p. 93–110. [S.l.]: Elsevier, 2011.
- BANGOR, A.; STAFF, T.; KORTUM, P.; MILLER, J. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. **Journal of Usability Studies**, v. 4, n. 3, p. 114–123. [S.l.]:UXPA, 2009.
- BATISTELLA, C. Análise da Situação de Saúde: principais problemas de saúde da população brasileira. In: EPSJV/Fiocruz (Ed.); **O território e o processo saúde-doença**. p.35. Rio de Janeiro: EPSJV/Fiocruz, 2007.
- BEDARD, Y.; RIVEST, S.; PROULX, M.-J. Spatial On-Line Analytical Processing (SOLAP): Concepts, Architectures, and Solutions from a Geomatics Engineering Perspective. **Data Warehouses and OLAP: Concepts, Architectures and Solutions**. p.298–319. London, UK: IRM Press, 2007.
- BIMONTE, S.; TCHOUNIKINE, A.; MIQUEL, M. GeoCube, a Multidimensional Model and Navigation Operators Handling Complex Measures: Application in Spatial OLAP. In: T. Yakhno; E. Neuhold (Eds.); **Advances in Information Systems SE - 11**, Lecture Notes in Computer Science. v. 4243, p.100–109. Springer Berlin : Heidelberg, 2006. Disponível em: <http://dx.doi.org/10.1007/11890393_11>. Acesso em: 23/07/2013.
- BORGES, K. A. DE V. **Modelagem de Dados Geográficos**. [S.l.:s.n], 2002.
- BRASIL. DECRETO Nº 6.666, DE 27 DE NOVEMBRO DE 2008. , 2008. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2008/Decreto/D6666.htm>. Acesso em: 22/08/2014 .

BROOKE, J. SUS-A quick and dirty usability scale. **Usability evaluation in industry**, v. 189, p. 194. [S.l.:s.n], 1996.

BROOKE, J. SUS: A Retrospective. **Journal of Usability Studies**, v. 8, n. 2, p. 29–40. [S.l.]: UXPA, 2013.

BRUNET, R. La carte-modèle et les chorèmes. **Mappemonde**, v. 4, p. 4–6. [S.l.:s.n], 1986. Disponível em: <www.mgm.fr/PUB/Mappemonde/M486/p2-6.pdf>. Acesso em: 02/08/2014 .

CÂMARA, G. Representação computacional de dados geográficos. **Bancos de Dados Geográficos**. p.1–44. São José dos Campos: INPE - Instituto Nacional de Pesquisas Espaciais, 2005.

CÂMARA, G.; MEDEIROS, J. S. DE. **Geoprocessamento para projetos ambientais**. 2nd ed. São José dos Campos: INPE - Instituto Nacional de Pesquisas Espaciais, 1998.

CARDOSO, F. M. **Um Framework para Data Warehouse Espacial**. Universidade Federal de Campina Grande, 2009.

CHIARA, D. DE; FATTO, V. DEL; LAURINI, R.; SEBILLO, M.; VITIELLO, G. A chorem-based approach for visually analyzing spatial data. **Journal of Visual Languages & Computing**, v. 22, n. 3, p. 173–193, 2011. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S1045926X11000115>>. Acesso em: 14/4/2013.

CODD, E. F.; CODD, S. B.; SALLEY, C. T. **Providing OLAP to User-Analysts: An IT Mandate**..Sunnyvale, CA, 1993 .

COHEN, I. B. Florence Nightingale. **Scientific American**, v. 250, n. 3, p. 1–7. [S.l.:s.n], 1984.

COLLINS, M. J. MVC Web Applications. **Pro HTML5 with Visual Studio 2012**. p.41–73. New York, USA: Apress, 2012.

DAVIS, C.; NETO, G. C. Arquitetura de Sistemas de Informação Geográfica. In: INPE (Ed.); **Anatomia de SIG**. p.36. São José dos Campos, 2001. Disponível em: <<http://mtc-m12.sid.inpe.br/archive.cgi/sid.inpe.br/sergio/2004/04.19.14.10>>. Acesso em: 05/05/2014.

DIAS, M. M.; YAMAGUCHI, J. K.; RABELO, E.; FRANCO, C. Visualization Techniques: Which is the Most Appropriate in the Process of Knowledge Discovery in Data Base? In: A. Karahoca (Ed.); **Advances in Data Mining Knowledge Discovery and Applications**, 2012. InTech. Disponível em: <<http://www.intechopen.com/books/advances-in-data-mining-knowledge-discovery-and-applications>>. Acesso em: 30/11/2014.

ECMA INTERNATIONAL. **ECMAScript Language Specification**. 5.1 ed. Geneva, 2011.

EGENHOFER, M. J.; HERRING, J. R. **Categorizing Binary Topological Relations Between Regions, Lines, and Points in Geographic Databases**. [S.l.:s.n], 1992.

EMBRAPA. **Sistema Brasileiro de Classificação de Solos**. 2nd ed. Brasília, 2006.

ESRI. **The 50th Anniversary of GIS**. Disponível em: <<http://www.esri.com/news/arcnews/fall12/articles/the-fiftieth-anniversary-of-gis.html>>. Acesso em: 24/7/2014..

ESRI. **Shapefiles**. Disponível em: <<http://doc.arcgis.com/pt-br/arcgis-online/reference/shapefiles.htm>>. Acesso em: 24/7/2014.

FEE. **Unidades Geográficas - Coredes**. Disponível em: <http://www.fee.rs.gov.br/feedados/consulta/unidades_geo_coredes.asp>. Acesso em: 25/8/2013.

FOX, P.; LITTLE, G.; BRALA, B.; MITCHELL, T. **Google Maps Utility Library v3**. 2009. Disponível em: <<https://code.google.com/p/google-maps-utility-library-v3/>>. Acesso em: 03/04/2013.

GOLFARELLI, M.; MANTOVANI, M.; RAVALDI, F.; RIZZI, S. Lily: A Geo-Enhanced Library for Location Intelligence. In: L. Bellatreche; M. K. Mohania (Eds.); **Data Warehousing and Knowledge Discovery**, Lecture Notes in Computer Science. v. 8057, p.72–83, 2013. Berlin, Heidelberg: Springer Berlin Heidelberg. Disponível em: <<http://link.springer.com/10.1007/978-3-642-40131-2>>. Acesso em: 30/11/2014.

GOOVAERTS, P. Geostatistical analysis of health data with different levels of spatial aggregation. **Spatial and Spatio-temporal Epidemiology**, v. 3, n. 1, p. 1-10. [S.l.]:Elsevier, 2012.

GRID. **Grupo de Gestão de Risco de Desastres**, 2014. Disponível em: <<http://www.ufrgs.br/grid>>. Acesso em: 10/11/2014.

GUANGCHUN, L.; LU, W.; HANHONG, X. A novel web application frame developed by MVC. **ACM SIGSOFT Software Engineering Notes**, v. 28, n. 2, p. 7. [S.l.]:ACM, 2003. Disponível em: <<http://portal.acm.org/citation.cfm?doid=638750.638779>>. Acesso em: 10/1/2014.

HABIBULLAYEVICH, G. K.; CHEN, X.; SHIN, H. Efficient Filtering and Clustering Mechanism for Google Maps. **Journal of Advanced Management Science**, v. 1, n. 1, p. 107–111, Rowland Heights, CA: ETPUB, 2013. Disponível em: <<http://www.joams.com/index.php?m=content&c=index&a=show&catid=27&id=34>>. Acesso em: 8/8/2014.

IBGE. **Manual técnico de Geomorfologia**. 2nd ed. Rio de Janeiro, 2009.

IBGE. **Manual Técnico da Vegetação Brasileira**. 2nd ed. Rio de Janeiro, 2012.

INMON, W. H. **Building the data warehouse**. 4th ed. Indianápolis: Wiley, 2005.

IOSIFESCU-ENESCU, I.; HUGENTOBLER, M.; HURNI, L. Web cartography with open standards – A solution to cartographic challenges of environmental management. **Environmental Modelling & Software**, v. 25, n. 9, p. 988–999. [S.l.]:Elsevier, 2010. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S1364815209002898>>. Acesso em: 1/5/2012.

ISAM. **Apresentação - Instituto de Saneamento Ambiental**. Disponível em: <<http://www.ucs.br/ucs/institutos/isam/apresentacao>>. Acesso em: 23/10/2014.

KIMBALL, R.; ROSS, M. **The data warehouse toolkit: guia completo para modelagem dimensional**. Rio de Janeiro: Campus, 2005.

KIMBALL, R.; ROSS, M. **The Data Warehouse Toolkit**. 3rd ed. Indianápolis, US: John Wiley & Sons, Inc, 2013.

KRYGIER, J.; WOOD, D. **Making maps - A visual guide to map design for GIS**. The [S.l.]:Guilford Press, 2011.

LAURINI, R.; SEBILLO, M.; VITIELLO, G.; SOL-MARTINEZ, D.; RAFFORT, F. Computer-generated Visual Summaries of Spatial Databases: Chorems or not Chorems? **S.A.P.I.EN.S [Online]**, v. 2, 2009. Disponível em: <<http://sapiens.revues.org/795>>. Acesso em: 22/10/2013.

LEPSCH, I. F.; JÚNIOR, R. B. **Manual para levantamento utilitário do meio físico e classificação de terras no sistema de capacidade de uso**. Sociedade Brasileira de Ciência do Solo, 1983.

MACEACHREN, A. M.; JAISWAL, A.; ROBINSON, A. C.; et al. SensePlace2: GeoTwitter analytics support for situational awareness. In: 2011 IEEE CONFERENCE ON VISUAL ANALYTICS SCIENCE AND TECHNOLOGY (VAST). **Proceedings...** Providence, US :IEEE. p.181–190, 2011.

MACIEJEWSKI, R.; HAFEN, R.; RUDOLPH, S.; et al. Forecasting Hotspots - A Predictive Analytics Approach. **IEEE transactions on visualization and computer graphics**, v. 17, n. 4, p. 440–453. [S.l.]:IEEE, 2011.

MACIEJEWSKI, R.; RUDOLPH, S.; TEBBETTS, G.; EBERT, D. S. Contextualizing Syndromic Hotspots - A Visual Analytics Approach. In: 5th INTERNATIONAL CONFERENCE ON GEOGRAPHIC INFORMATION SCIENCE. **Proceedings...** , [S.l.]:IEEE. p.1–4, 2008.

MAHE, L.; BROADFOOT, C. Too Many Markers! Disponível em: <<https://developers.google.com/maps/articles/toomanymarkers>>. Acesso em: 20/8/2014.

NARDI, A. R. Fundamentos e Modelagem de Bancos de Dados Multidimensionais. Disponível em: <<http://msdn.microsoft.com/pt-br/library/cc518031.aspx>>. Acesso em: 1/9/2013.

NEUMANN, A. **Encyclopedia of GIS**. Boston, MA: Springer US, 2008.

NIGHTINGALE, F. **A contribution to the sanitary history of the British army during the late war with Russia**. London, UK: Harvard University - Countway Library of Medicine, 1859.

OPENLAYERS. **Introduction**, 2013. Disponível em: <<http://openlayers.org/>>. Acesso em: 02/02/2014 .

PENTAHO. **Pentaho Analysis Services (Mondrian)**, 2013 Disponível em: <<http://mondrian.pentaho.com/>>. Acesso em: 08/02/2014.

PIMENTA, F. M.; LANDAU, E. C.; HIRSCH, A.; DANIEL PEREIRA GUIMARÃES. **Servidores de mapas: Programação para Disponibilizar Dados Geográficos Multidisciplinares Utilizando Tecnologias Livres**. Brasília, DF, 2012.

PROULX, M.-J.; BÉDARD, Y. Fundamental Characteristics of Spatial OLAP Technologies as Selection Criteria. In: LOCATION INTELLIGENCE. **Proceedings...** [S.l.]:Santa Clara, p. 1-15, 2008.

QUEIROZ, G. R. DE; FERREIRA, K. R. SGBD com extensões espaciais. **Bancos de Dados Geográficos**. 1st ed., p.33. Curitiba, PR: MundoGEO, 2005.

RIO DE JANEIRO. **Metodologia de Codificação de Bacias Hidrográficas**. Rio de Janeiro, 1986.

RIO GRANDE DO SUL, A. L. DO E. DO. **LEI Nº 10.283, DE 17 DE OUTUBRO DE 1994**. 1994.

RIVEST, S.; MARCHAND, P. Toward better support for spatial decision making: defining the characteristics of spatial on-line analytical processing (SOLAP). **GEOMATICA**, v. 55, n. 4, p. 539 to 555, 2001.

ROLIM, G. D. E. S.; BENTO, M.; CAMARGO, P. D. E.; et al. Classificação climática de köppen e de thornthwaite e sua aplicabilidade na determinação de zonas agroclimáticas para o estado de são paulo (1). **Bragantia**, v. 66, n. 4, p. 711–720, 2007.

ROSS, M. **The 10 Essential Rules of Dimensional Modeling**, 2009. Disponível em: <<http://www.kimballgroup.com/2009/05/29/the-10-essential-rules-of-dimensional-modeling/>>. Acesso em: 25/8/2013.

RUPERT, O. A. V. **Codificação Automática de Regiões Hidrográficas Utilizando Sistemas de Informações Geográficas**, 2000. Universidade Federal de Viçosa.

SCHNEIDER, V. E.; FINOTTI, A. R.; SILVA, M. D. A.; CEMIN, G.; PERESIN, D. Gerenciamento Integrado de Resíduos Rurais e sua Interface com a Gestão Ambiental: Panorama da situação na região de abrangência do COREDE Serra (RS). **Revista Limpeza Publica**, v. 1, n. 64, p. 39, 2007.

SCHUMACHER, R. M.; LOWRY, S. Z. **Customized Common Industry Format Template for Electronic Health Record Usability Testing**. 2010. Disponível em:

<http://www.nist.gov/itl/hit/upload/LowryNISTIR-7742Customized_CIF_Template_for_EHR_Usability_Testing_Publicationl_Version-doc.pdf>. Acesso em: 6/3/2013.

SILVA, J. DA; OLIVEIRA, A. G. DE; FIDALGO, R. N.; SALGADO, A. C.; TIMES, V. C. Modelling and querying geographical data warehouses. **Information Systems**, v. 35, n. 5, p. 592–614, 2010. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0306437909001033>>. Acesso em: 6/3/2013.

SILVA, R.; SANTOS, M. Y. Spatial Clustering to Uncluttering Map Visualization. In: ICCSA'11 - 2011 INTERNATIONAL CONFERENCE ON COMPUTATIONAL SCIENCE AND ITS APPLICATIONS. **Proceedings...** Espanha, p.253–268, 2011.

SILVA, R.; SANTOS, M. Y. Spatial Clustering in SOLAP Systems to Enhance Map Visualization. **International Journal of Data Warehousing and Mining (IJDWM)**, v. 8, n. June, p. 23–43, [S.l.]: IGI Global, 2012.

SINGH, R.; SINGH, K. A Descriptive Classification of Causes of Data Quality Problems in Data Warehousing. **IJCSI International Journal of Computer Science**, v. 7, n. 3, p. 41–50, [S.l.: s.n.], 2010.

TYNER, J. A. **Principles of Map Design**. 1st ed. New York, New York, USA: The Guilford Press, 2010.

UNITED NATIONS. **United Nations Convention on the Law of the Sea of 10 December 1982**. 1982. Jamaica. Disponível em: <http://www.un.org/depts/los/convention_agreements/texts/unclos/UNCLOS-TOC.htm>. Acesso em: 9/3/2013. .

VERDIN, K. L. A System for Topologically Coding Global Drainage Basins and Stream Networks. In: ESRI INTERNATIONAL GIS USER CONFERENCE. **Proceedings...** . p.1, [S.l.: s.n.], 1997. Disponível em: <<http://proceedings.esri.com/library/userconf/proc97/proc97/to350/pap311/p311.htm>>. Acesso em: 6/9/2013.

W3C. **CSS - Curso W3C escritório brasil**. São Paulo, 2011.

W3C. What is HTML5?, 2012 Disponível em: <http://www.w3.org/html/wiki/FAQs#What_is_HTML5.3F>. Acesso em: 6/1/2014.

YOUNG, S. G.; JENSEN, R. R. Statistical and visual analysis of human West Nile virus infection in the United States, 1999–2008. **Applied Geography**, v. 34, p. 425–431, [S.l.]:Elsevier, 2012.

ZAIANE, O. R.; FOSS, A.; LEE, C.; WANG, W. On Data Clustering Analysis: Scalability , Constraints , and Validation. **Advances in Knowledge Discovery and Data Mining**. p.28–39. London: Springer-Verlag, 2002.

APÊNDICE A - DESCRIÇÃO COMPLETA DAS TABELAS DE DIMENSÕES

DEFINIÇÕES

- O tipo de dado SERIAL, é um inteiro incrementável do PostgreSQL.

Dimensão política

dim_politica		
Nome coluna	Tipo	Descrição
id_dim_politica	SERIAL	ID com número inteiro que funciona como chave da integridade referencial.
continente	varchar	Informação textual para identificar o continente.
pais	varchar	Informação textual para identificar o país
regiao	varchar	Informação textual para identificar a região (exemplo: sul, sudeste).
estado	varchar	Informação textual para identificar o estado
microregiao	varchar	Informação textual para identificar a microregião
mesoregiao	varchar	Informação textual para identificar a mesoregião
coredes	varchar	Informação textual para identificar a Coredes
municipio	varchar	Informação textual para identificar a municipio
bairro	varchar	Informação textual para identificar a bairro
zona_economica_exclusiva	boolean	porção marinha entre 12 e 200 milhas náuticas da costa de um país, área sobre a qual este possui direitos exclusivos de exploração econômica, apesar de não possuir direito exclusivo de passagem.
codigo_ibge_estado	int	Código que o IBGE utiliza para referenciar um estado
codigo_ibge_municipio	int	Código que o IBGE utiliza para referenciar um município
sigla_estado	varchar	Sigla do estado
codigo_internacional_pais	int	Código do país segundo a ONU
abreviatura_pais	varchar	Abreviatura ou sigla do país
nivel	int	Nível da hierarquia
fonte_dado	varchar	Fonte de onde o dado geográfico foi gerado
escala_geometria	int	Escala do dado original
geometria	polygon	Geometria para o ETL e para possível visualização em mapa.

Dimensões físicas

dim_fisica_clima_bernardes		
Nome Coluna	Tipo	Descrição
id_dim_fisica_clima_bernardes	serial	Identificador único para manter a integridade referencial
classificacao	varchar	Identificador da classificação segundo Bernardes
descricao	varchar	Descrição da classificação
nivel	int	Nível da hierarquia
fonte_dado	varchar	Fonte de onde o dado geográfico foi gerado
escala_geometria	int	Escala do dado original
geometria	polygon	Geometria para o ETL e para possível visualização em mapa.

dim_fisica_clima_koppen		
Nome Coluna	Tipo	Descrição
id_dim_fisica_clima_koppen	serial	Identificador único para manter a integridade referencial
classificacao_koppen_geiger	varchar	Código de classificação
primeira_letra	char	Primeira letra de classificação
segunda_letra	char	Segunda letra de classificação
terceira_letra	char	Terceira letra de classificação
caracteristica	varchar	Descrição da classificação
regimes_temperatura	varchar	Qual o regime de temperatura
area_de_ocorrendia	varchar	Descrição das áreas de ocorrência
nivel	int	Nível da hierarquia
fonte_dado	varchar	Fonte de onde o dado geográfico foi gerado
escala_geometria	int	Escala do dado original
geometria	polygon	Geometria para o ETL e para possível visualização em mapa.

dim_fisica_clima_strahler		
Nome Coluna	Tipo	Descrição
id_dim_fisica_clima_strahler	serial	Identificador único para manter a integridade referencial
classificacao	varchar	Código de classificação
descricao	varchar	Descrição da classificação

dim_fisica_clima_strahler		
Nome Coluna	Tipo	Descrição
nivel	int	Nível da hierarquia
fonte_dado	varchar	Fonte de onde o dado geográfico foi gerado
escala_geometria	int	Escala do dado original
geometria	polygon	Geometria para o ETL e para possível visualização em mapa.

dim_fisica_clinografia		
Nome Coluna	Tipo	Descrição
id_dim_fisica_clinografia	serial	Identificador único para manter a integridade referencial
declividade_classe	varchar	Classe a qual a declividade pertence (abreviado)
classe_nome_formatado	varchar	Nome da classe por extenso
classe_descricao	varchar	Descrição da classe
nivel	int	Nível da hierarquia
fonte_dado	varchar	Fonte de onde o dado geográfico foi gerado
escala_geometria	int	Escala do dado original
geometria	polygon	Geometria para o ETL e para possível visualização em mapa.

dim_fisica_geomorfologia		
Nome Coluna	Tipo	Descrição
id_dim_fisica_geomorfologia	serial	Identificador único para manter a integridade referencial
dominio_morfo_estrutural	varchar	Domínio morfo estrutural
regiao_geomorfologica	varchar	Região geomorfológica
unidade_morfologica	varchar	Unidade morfologica
modelado	varchar	Modelado
forma_do_relevo	varchar	A forma de relevo predominante
nivel	int	Nível da hierarquia
fonte_dado	varchar	Fonte de onde o dado geográfico foi gerado
escala_geometria	int	Escala do dado original

dim_fisica_geomorfolgia		
Nome Coluna	Tipo	Descrição
geometria	polygon	Geometria para o ETL e para possível visualização em mapa.

dim_fisica_hidrografica_comite		
Nome Coluna	Tipo	Descrição
id_dim_fisica_hidrografica_co mite	serial	Identificador único para manter a integridade referencial
bacia_comite	varchar	Comite ao qual a bacia está regida
unidade_de_gestao_nivel_1	varchar	Primeira divisão de unidade de gestão
unidade_de_gestao_nivel_2	varchar	Segunda divisão de unidade de gestão
unidade_de_gestao_nivel_3	varchar	Terceira divisão de unidade de gestão
nivel	int	Nível da hierarquia
fonte_dado	varchar	Fonte de onde o dado geográfico foi gerado
escala_geometria	int	Escala do dado original
geometria	polygon	Geometria para o ETL e para possível visualização em mapa.

dim_fisica_hidrografica_regiao		
Nome Coluna	Tipo	Descrição
id_dim_fisica_hidrografica_re giao	serial	Identificador único para manter a integridade referencial
regiao_hidrografica_nivel_1	varchar	Região hidrográfica de primeiro nível
regiao_hidrografica_nivel_2	varchar	Região hidrográfica de segundo nível
regiao_hidrografica_nivel_3	varchar	Região hidrográfica de terceiro nível
regiao_hidrografica_nivel_4	varchar	Região hidrográfica de quarto nível
regiao_hidrografica_nivel_5	varchar	Região hidrográfica de quinto nível
nivel	int	Nível da hierarquia
fonte_dado	varchar	Fonte de onde o dado geográfico foi gerado
escala_geometria	int	Escala do dado original
geometria	polygon	Geometria para o ETL e para possível visualização em mapa.

dim_fisica_solo		
Nome Coluna	Tipo	Descrição
id_dim_fisica_solo	serial	Identificador único para manter a integridade referencial
ordem	varchar	Primeiro nível da hierarquia de classificação do tipo de solo
sub_ordem	varchar	Segundo nível da hierarquia de classificação do tipo de solo
grande_grupo	varchar	Terceiro nível da hierarquia de classificação do tipo de solo
subgrupo	varchar	Quarto nível da hierarquia de classificação do tipo de solo
familia	varchar	Quinto nível da hierarquia de classificação do tipo de solo
serie	varchar	Sexto nível da hierarquia de classificação do tipo de solo
descricao	varchar	Descrição textual do tipo de solo
fonte_descricao	varchar	Fonte da descrição
nivel	int	Nível da hierarquia
fonte_dado	varchar	Fonte de onde o dado geográfico foi gerado
escala_geometria	int	Escala do dado original
geometria	polygon	Geometria para o ETL e para possível visualização em mapa.

dim_fisica_vegetacao		
Nome Coluna	Tipo	Descrição
id_dim_fisica_vegetacao	serial	Identificador único para manter a integridade referencial
escala_faixa_inicial	varchar	Qual a escala inicial que permite avaliar o dado
escala_faixa_final	varchar	Qual a escala final que permite avaliar o dado
escala_nivel_detalhamento	varchar	Qual o nível de detalhamento da amostragem
classe_de_formacao	varchar	Classe de formação vegetal
subclasse_de_formacao	varchar	Subclasse de formação vegetal

dim_fisica_vegetacao		
Nome Coluna	Tipo	Descrição
grupo_formacao	varchar	Grupo de formação vegetal
formacao	varchar	Formação vegetal
subformacao	varchar	Subformação
nivel	int	Nível da hierarquia
fonte_dado	varchar	Fonte de onde o dado geográfico foi gerado
escala_geometria	int	Escala do dado original
geometria	polygon	Geometria para o ETL e para possível visualização em mapa.

Dimensões temporais

dim_tempo_dia		
Nome Coluna	Tipo	Descrição
id_dim_tempo_dia	serial	Identificador único para manter a integridade referencial
data_sql	timestamp	Data no formato de timestamp do sql YYYY-MM-DD HH:MM:SS
dia_da_semana	varchar	Dia da Semana por extenso
data_extenso	varchar	Data completa por extenso exemplo 23 de janeiro de 2014
data_dd_mm_yyyy	varchar	Data no formato dia/mes/ano exemplo 23/01/1988
mes_extenso	varchar	Mês por extenso
mes_numero	int	Mês em número
ano_numero	int	Ano em número
dia_numero	int	Dia em número
semestre	int	Semestre do ano
trimestre	int	Trimestre do ano
bimestre	int	Bimestre do ano
estacao	varchar	Estação climática hemisfério sul
feriado	varchar	Nome do feriado caso ser
indicador_feriado	boolean	Indicador se a data é feriado ou não
final_de_semana	boolean	Indicador se a data é em final de semana

dim_tempo_dia		
Nome Coluna	Tipo	Descrição
dia_util	boolean	Dia útil
dia_ano	int	Dia do ano por exemplo o dia 223
semana_ano	int	Semana do ano
mes_ano	varchar	Indicador mês/ano

dim_tempo_hora		
Nome Coluna	Tipo	Descrição
id_dim_tempo_hora	serial	Identificador único para manter a integridade referencial
hora_minuto	time	Hora: minuto
hora	int	Hora número
minuto	int	Minuto número
hora_12	int	Hora em formato 12 horas
am_pm	varchar(n)	Se é antes ou depois do meio dia
minuto_no_dia	int	Minuto no dia
minuto_extenso	varchar	Minuto escrito por extenso
hora_extenso	varchar	Hora escrito por extenso
turno_divido_em_dois	varchar	Turno em dia e noite
turno_dividido_em_tres	varchar	Turno em manhã tarde e noite
turno_dividido_em_quatro	varchar	Turno em manhã tarde, vespertino e noite
horario_comercial	boolean	Se é horário comercial

APÊNDICE B - DOCUMENTAÇÃO WEB SERVICE DE GERADOR DE ESTILOS

Gerado a partir do arquivo WADL (*Web Application Description Language*)
disponível em : <http://vbaco01.ucs.br/mestrado/>

Base	http://vbaco01.ucs.br/				
Path	/mestrado/geradorIcône.php				
Parâmetros					
Parâmetro	Tipo	Obrigatoriedade	Padrão	Valores possíveis	Descrição
Tom	Int/String	Não	1	RED=1, GREEN=2	Inteiro ou constante que representa o tom
tam	Int/String	Não	1	1,2,3,4	Tamanho do ícone 1 para o menor e 4 para o maior
icon	url	Sim	-	-	URI para um ícone que será utilizado para a geração