UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

CENTRO DE CIÊNCIAS AGRÁRIAS

DEPARTAMENTO DE COMPUTAÇÃO

DANILO BOECHAT SEUFITELLI

DATA WAREHOUSE ESPACIAL PARA ANÁLISE HISTÓRICA DOS CASOS DE DENGUE NO ESTADO DO ESPÍRITO SANTO

ALEGRE - ES

2014

DANILO BOECHAT SEUFITELLI

DATA WAREHOUSE ESPACIAL PARA ANÁLISE HISTÓRICA DOS CASOS DE DENGUE NO ESTADO DO ESPÍRITO SANTO

Trabalho de conclusão de curso apresentado ao Departamento de Computação do Centro de Ciências Agrárias da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Bacharel em Sistemas de Informação.

Orientador: Prof. MSc. Antonio Almeida de Barros Junior.

ALEGRE - ES

2014

DANILO BOECHAT SEUFITELLI

DATA WAREHOUSE ESPACIAL PARA ANÁLISE HISTÓRICA DOS CASOS DE DENGUE NO ESTADO DO ESPÍRITO SANTO

Trabalho de conclusão de curso apresentado ao Departamento de Computação do Centro de Ciências Agrárias da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Bacharel em Sistemas de Informação.

Aprovado em 18 de fevereiro de 2014.

BANCA EXAMINADORA

Prof. MSc. Antonio Almeida de Barros Junior
Universidade Federal do Espírito Santo
Orientador

Prof. Dr. Bruno Vilela Oliveira
Universidade Federal do Espírito Santo

Prof. MSc. Valéria Alves da Silva
Universidade Federal do Espírito Santo

“Nenhum de nós chegou onde está, exclusivamente por meio do impulso de nossos próprios pés. Nós chegamos até aqui porque alguém se inclinou e nos ajudou.”

Thurgood Marshall

AGRADECIMENTOS

A Deus, pelo dom da vida e pela oportunidade de aprender a cada dia.

À minha mãe e a meu pai, pelo amor incondicional.

Aos meus irmãos, que mesmo de longe sempre estiveram comigo.

Aos meus familiares, amigos e namorada, pela presença, palavra e companheirismo.

RESUMO

As organizações gestoras da saúde pública sempre buscam promover o aumento da eficiência na administração de seus programas de gerenciamento da saúde. Necessita-se então de agilidade na produção de informação e consequentemente em sua análise. Por esta razão, existe uma grande importância em analisar dados referentes à Dengue (doença causada por um vírus), devido ao aumento de sua incidência nos últimos anos. Tais análises podem ser feitas por meio de modernas técnicas computacionais a fim de se obter o comportamento epidemiológico da doença na dimensão do tempo e espaço, o que, por sua vez, coopera com a melhoria de programas de prevenção e controle da doença. Dentre as técnicas computacionais utilizadas no processo de tomada de decisão, a técnica de *Data Warehouse* Geográfico (DWG) apresentou-se como uma importante ferramenta nesse processo. Portanto, o objetivo deste trabalho foi desenvolver um DWG destinado à análise dos casos de Dengue do estado do Espírito Santo, bem como a construção de uma ferramenta *Extract Transform and Load* (ETL), capaz de efetuar a carga do DWG a partir das bases de dados do Departamento de Informática do SUS – DATASUS.

Palavras-chave: Saúde Pública, Dengue*, Data Warehouse* Geográfico, *Extract Transform and Load*.

SPATIAL DATA WAREHOUSE FOR HISTORICAL ANALYSIS OF CASES OF DENGUE IN THE STATE OF ESPÍRITO SANTO

ABSTRACT

The management of public health organizations always seek to promote increased efficiency in managing their health management programs. Then there is need of agility in information production and consequently in their analysis. For this reason, there is a great importance in analyzing data relating to Dengue (disease caused by a virus), due to the increase in incidence over the last years. Such analysis can be done through modern computational techniques in order to obtain the epidemiological behavior of the disease in the dimension of time and space, which, in turn, cooperates with the improvement of prevention and control programs. Among the computational techniques used in the decision- making process, the technique of Geographic Data Warehouse (DWG) was presented as an importante tool in this process. Therefore, the aim of this work was to develop a DWG for the analysis of cases of dengue in the state of Espírito Santo, and the construction of a tool Extract Transform and Load (ETL), able to carry the load of the DWG from bases data from the Department of the SUS - DATASUS.

Keywords: Public, Dengue, Geographic Data Warehouse, Extract Transform and Load Health.

LISTA DE FIGURAS

FIGURA 1 - Fêmea do Aedes Aegypti – Transmissor da Dengue..... 15

FIGURA 2 - Características do *Data Warehouse*.....18

FIGURA 3 - Diagrama dos elementos do DW.....19

FIGURA 4 - Escopo Global do Projeto.....24

FIGURA 5 – Modelo Multidimensional.....25

FIGURA 6 - Tela Inicial da Ferramenta ETL.....27

FIGURA 7 - Tela Carga Intermediária.....28

FIGURA 8 - Tela Carga Cubo.....29

FIGURA 9 – Mapa de casos de Dengue por ano.....31

FIGURA 10 - Gráfico de Frequência de Incidência x Dia da semana.....32

FIGURA 11 - Gráfico de Casos de Dengue x Ano.....33

FIGURA 12 - Gráfico de Frequência de Casos de Dengue x Mês ao longo de 6 anos.....34

FIGURA 13 - Gráfico percentual de registros de Dengue de acordo com sua classificação.....35

FIGURA 14 - Gráfico de registro de Dengue por ano de acordo com o sexo...36

FIGURA 15 - Gráfico de frequência de registros por idade.....37

FIGURA 16 – Gráfico de séries temporais.....40

SUMÁRIO

1 Introdução 8
1.1. O Problema e sua Importância 10
1.2. Objetivos 13
1.2.1. Objetivo Geral 13
1.2.2. Objetivos Específicos 13
2 Revisão de Literatura 14
2.1. Sistema Único de Saúde 14
2.2. Dengue 15
2.3. Data Warehouse 17
2.3.1. Orientado a assunto 18
2.3.2. Integrado 18
2.3.3. Não volátil 18
2.3.4. Variável no tempo 18
2.3.5. Acessível 19
2.3.6. Processo Orientado 19
2.4. *Data Warehouse* Geográfico 21
3 Metodologia 24
4 Resultados 30
4.1. Análise Geográfica 30
4.2. Análises Descritivas 32
4.3. Análise de Regressão 38
5 Conclusão 41
6 Referências 42
ANEXO A – Diagrama uml das transformações 46

1 INTRODUÇÃO

As organizações gestoras da saúde pública buscam, a todo o tempo, promover o aumento da eficiência na administração de seus programas de gerenciamento da saúde. Como o governo produz uma enorme quantidade de dados provenientes dos registros de incidências de doenças, necessita-se de agilidade na produção de informação e consequentemente de sua análise.

Atualmente, devido aos inúmeros casos de eventos epidemiológicos emergentes e reemergentes no Brasil, enxerga-se a necessidade de uma ferramenta que integre as diversas bases de dados do governo, de forma a contribuir com agilidade nas análises destes dados, levando à informação.

Doenças emergentes são aquelas cuja incidência tem aumentado nas duas últimas décadas ou ameaçam aumentar no futuro. São consideradas doenças reemergentes, aquelas que aparentemente foram erradicadas ou tiveram registros de incidências consideravelmente reduzidos mas voltaram a apresentar ameaça para a saúde humana (BRUCE et al., 2009).

A Dengue por sua vez é uma doença reemergente, causada por um vírus, a qual é bastante conhecida devido ao aumento de sua incidência nos últimos anos (HUNG et al., 2013).

Existe uma grande importância em analisar dados referentes à Dengue por meio de modernas técnicas computacionais, como o *Data Warehouse* (DW), a fim de se obter o comportamento epidemiológico da doença na dimensão do tempo e espaço, o que pode cooperar com a melhoria de programas de prevenção e controle da doença (HUNG et al., 2013).

Souza et al. (2010) geraram um banco de dados com atributos climáticos. A partir desse conjunto de informações, averiguaram a distribuição dos casos notificados do ano de 2009 na cidade de Cáceres – MT, com o objetivo de contribuir com informações para o monitoramento e controle epidemiológico da Dengue.

Outro estudo que relaciona técnicas computacionais com casos de dengue, foi o de Lorbieski et al. (2010). Tais pesquisadores desenvolveram um sistema

computacional cujo o objetivo foi disponibilizar um Sistema de Informações Georreferenciados específico para a Dengue no município de Cascavel- PR.

Dentre as técnicas computacionais utilizadas na análise de dados para o processo de tomada de decisão, a técnica de *Data Warehouse* Geográfico (DWG) apresenta-se como uma importante ferramenta, por ser um facilitador da interpretação dos dados relacionando-os às áreas geográficas de estudo.

A partir dessas premissas, este trabalho visa apresentar um estudo e implementação de um DWG destinado à análise dos casos de Dengue do estado do Espírito Santo, bem como a construção de uma ferramenta *Extract Transform and Load* (ETL), que seja capaz de efetuar a carga do DWG a partir das bases de dados do Departamento de Informática do SUS (DATASUS), de forma a contribuir com o tema Cidades Inteligentes, que é o termo ao qual a Sociedade Brasileira de Computação – SBC, associa o papel da tecnologia da informação e comunicação na solução de problemas oriundos do progresso das cidades, como questões sobre saúde pública, tais como a Dengue (CSBC, 2013).

A seguir, será abordado mais detalhadamente o problema, para destacar a importância da criação da ferramenta. Em seguida, serão identificados o objetivo geral e os objetivos específicos do trabalho. No capítulo 2, será apresentada uma abordagem atual dos estudos na área. O Capítulo 3 descreve a metodologia utilizada para o desenvolvimento do trabalho. No Capítulo 4 são apresentados os resultados obtidos com a pesquisa e, no Capítulo 5, são enumeradas as referências.

1.1. O PROBLEMA E SUA IMPORTÂNCIA

Segundo Blane e Acheson (1999, 1998, apud PIRES, 2011) a Saúde Pública pode ser caracterizada como “a arte e a ciência de prevenir doenças, promover a saúde e prolongar a vida por meio de esforços organizados da sociedade”. Para Graham e Kelly (2004), a saúde pública é entendida como uma garantia de uma maior expectativa de vida para a população, onde deve-se buscar por diminuir a desigualdade, devido à posições sociais, e garantir a todos uma saúde pública de qualidade, por meio de tratamentos adequados e a prevenção de enfermidades.

Dentre as diversas definições para o termo, todas elas apresentam como cerne o controle, a prevenção e redução de doenças, bem como a manutenção e promoção da saúde de toda a população (BEAGLEHOLE, 2004).

De acordo com o SUS a Saúde Pública é garantida no contexto nacional conforme a Constituição Federal Brasileira. O DATASUS é o órgão máximo responsável por coletar, processar e armazenar todos os dados referentes à Saúde Pública no Brasil (DATASUS, 2013).

Ao se falar em Saúde Pública, termos como ‘controle’ e ‘prevenção’ sempre se fazem presentes em tais definições. Como matéria-prima para a realização destas ações, a informação é essencial, sendo impossível controlar e prevenir sem a disponibilidade e o uso adequado da informação. Para que seja possível produzir informações necessárias à gestão do SUS, o DATASUS possui sistemas administrativos como Sistema Estatísticas Vitais (SIM/SINAC), Cadastro Nacional de Estabelecimentos de Saúde (CNES), Sistemas de Informações Ambulatoriais (SIASUS), Sistema de Informações Hospitalares (SIHSUS) e Sistema de Informação de Agravo de Notificação (SINAN), conforme detalhado por Santos (2006).

Segundo Pires (2011), apesar dos sistemas do DATASUS conterem uma quantidade considerável de dados que produzem uma grande quantidade de informação, há o desafio e a necessidade de identificar e desenvolver aplicações

adequadas para manipular a informação disponível e proporcionar o conhecimento necessário aos objetivos da Saúde Pública.

Dentre os sistemas citados anteriormente, o SINAN se destaca por ser alimentado, principalmente, por notificações de casos de agravos (doenças) que se encontram na lista nacional de doenças de notificação compulsória (Portaria GM/MS Nº 104, de 25 de Janeiro de 2011) disponíveis no Portal da Saúde (2013). Dentre as doenças que constam na lista nacional, encontra-se a Dengue, que só na primeira semana de março de 2013 foram registrados 84.122 novos casos no Brasil.

A Dengue é uma doença infecciosa tendo como agente causador, um arbovírus, que se enquadra em quatro sorotipos diferentes conhecidos, que são identificadas pelas siglas DEN-1, DEN-2, DEN-3 e DEN-4 (DENGUE, 2013). Este arbovírus só pode ser transmitido ao homem por um transmissor, também chamado de vetor, que é popularmente conhecido como o mosquito da Dengue, denominado de *Aedes Aegypti* (OLIVEIRA et al., 2013).

Duarte et al. (2006) diz que deve-se ter agilidade suficiente para detectar precocemente as epidemias, bem como os casos de evolução grave, por parte da vigilância epidemiológica da Dengue, a fim de se reduzir a letalidade da doença. Para tanto, precisa-se ter disponível informação consistente e oportuna, diagnóstico laboratorial otimizado, critério de caso bem definido e profissionais de saúde com um bom conhecimento clínico da doença.

Nesse contexto, a Ciência da Computação oferece recursos (técnicas e ferramentas) que são capazes de integrar informações históricas e não voláteis, dentre as quais se incluem as técnicas de DW e *On-line Analytical Processing* (OLAP).

Em suma, como os sistemas DATASUS produzem uma enorme quantidade de dados e informações sobre registros de Dengue espalhados pelo Brasil, localizados em bases de dados distintas, como o SIM, SIASUS, SINAN entre outras, há a necessidade de formatar e integrar os dados dessas bases em um único repositório (DWG). De posse desse repositório, pode-se então extrair informações sumarizadas onde a partir delas consegue-se promover a

descoberta de novos conhecimentos sobre os casos encontrados de Dengue no estado do Espírito Santo, visando sempre contribuir com os objetivos da saúde pública: controle e prevenção.

1.2. OBJETIVOS

1.2.1. Objetivo Geral

Elaborar e construir uma ferramenta de Extração, Transformação e Carga de Dados (ETL) capaz de extrair dados heterogêneos das diversas bases de dados do SUS de todo estado do Espírito Santo e armazenar os dados sumarizados em um banco de dados multidimensional (DWG), que também será elaborado e construído a partir dos metadados das bases de dados do SUS. A partir disso, objetivou-se buscar por conhecimentos baseados nos fatos gerados por meio do DWG e consultados por uma ferramenta OLAP e Sistemas de Informações Geográficas (SIG).

1.2.2. Objetivos Específicos

- Analisar os metadados do DATASUS;
- Propor uma modelagem multidimensional espacial (DWG) para armazenar as informações sumarizadas;
- Criar o DWG;
- Desenvolver uma ferramenta que seja capaz de extrair os dados do Sistema Único de Saúde e inseri-los, de forma sumarizada, no DWG. Ou seja, criar uma ferramenta ETL;
- Analisar os resultados dos fatos gerados de acordo com a modelagem feita;

2 REVISÃO DE LITERATURA

2.1. SISTEMA ÚNICO DE SAÚDE

A Saúde pública é um assunto de interesse a toda a população, principalmente quando o assunto é o Sistema Único de Saúde (SUS). O Art. 196 da Constituição Federal (CF) diz que “a saúde é direito de todos e dever do Estado (...)”, o que idealiza que todos os brasileiros detém o direito da saúde, e também a partir desta, encontra-se clara, a universalidade da cobertura do SUS.

Antes da implantação do SUS, em 1988, o acesso à saúde era garantido aos cidadãos que podiam pagar por serviços de saúde privados, aos que detinham o direito à saúde pública, por serem assegurados pela previdência social e aos que não possuíam direito algum sobre a saúde (BRASIL, 2013). Essa definição leva a interpretação de que apenas um pequeno grupo da população tinha o acesso gratuito à saúde.

Após a criação do SUS, houve um aumento significativo de beneficiados pela saúde pública, pois o número de usuários era de 30 milhões e passou para 190 milhões com a implementação desse sistema. Cerca de 80% dessa população depende exclusivamente do SUS para ter acesso aos serviços de saúde (BRASIL, 2013). Segundo o Ministério da Saúde (2013), existem cerca de 261.605 hospitais credenciados ao SUS no Brasil, e 48% deles são privados. O Espírito Santo se encontra com 2,05% deste total, com 5.582hospitais integrados ao SUS (CNES, 2013).

Por intermédio da informatização das atividades do SUS, dentro de diretrizes tecnológicas adequadas, grande quantidade de dados é gerada e mantida pelo Departamento de Informática do SUS (DATASUS), o qual faz jus a uma de suas competências: “manter o acervo das bases de dados necessário ao sistema de informações em saúde e aos sistemas internos de gestão institucional” (DATASUS, 2013).

A CF dita pelo Art. 197, que as ações e serviços de saúde são de relevância pública, e o Poder Público, em termos da lei, deve fiscalizar e controlar a saúde,

podendo ser executada “diretamente ou por meio de terceiros e, também, por pessoa física ou jurídica de direito privado”.

Com isso, temos que a informação é fundamental para a democratização da Saúde e o aprimoramento de sua gestão, onde busca-se por atender às necessidades do SUS e aproximar as inovações científicas e o desenvolvimento tecnológico das ações de prevenção e controle da Dengue, um dos problemas de saúde que mais acometem a população brasileira.

2.2. DENGUE

A Dengue, é uma infecção viral causada por um arbovírus transmitida pela fêmea do mosquito *Aedes Aegypti*, ilustrado na Figura 1, que também se infecta com este vírus e o dissemina por meio de sua picada. É um problema de saúde que cresce rapidamente com uma estimativa de 2,5 bilhões de pessoas em situação de risco, principalmente nos países do sul e sudeste do globo terrestre com destaque na Ásia, Caribe, Américas Central e do Sul, e também na África. (VARELLA, 2013 e WHITEHORN et al., 2010).

Figura 1 - Fêmea do Aedes Aegypti – Transmissor da Dengue

Fonte: Enciclopédia Britânica, 2013.

A propagação da Dengue pode ser interpretada por uma combinação de diversos fatores como o aumento da urbanização, o crescimento populacional, a migração de viagens internacionais, bem como a dificuldade do controle eficaz do transmissor da doença. O clima também pode ser um fator determinístico para a propagação global da Dengue (WHITEHORN et al., 2010).

Ainda há poucos estudos que tragam contribuições para a eliminação da Dengue no Brasil. Em um estudo realizado por Barcellos et al. (2005), no estado do Rio Grande do Sul, foi proposto um SIG para o controle de Dengue, visando apontar as áreas de maior intensidade de transmissão da doença e identificar as áreas mais vulneráveis ao vetor.Com isso, foi possível planejar ações para o controle da Dengue.

No estado do Espírito Santo, em 2011, foram registrados mais de 51 mil casos de Dengue, com cerca de 25 óbitos. No ano de 2013, somente no primeiro mês foram registrados mais de 5 mil casos, com suspeita de 84 desses serem da forma mais grave da doença. O Espírito Santo é o terceiro estado com maior número de casos de Dengue do Brasil, ficando atrás apenas do Mato Grosso do Sul e Minas Gerais (G1, 2013a, 2013b).

No Mato Grosso, um estudo realizado por Souza et al. (2010), no qual foi proposto a geração de um banco de dados geográfico da cidade de Cáceres ao qual foram adicionadas informações climáticas (precipitação e temperatura) do período entre 1971 e 2009, além dos casos de Dengue notificados em 2009. Com os resultados, foram criados mapas temáticos mensais da distribuição dos casos de Dengue e suas respectivas análises em uma perspectiva que contribuiu para o monitoramento, controle e combate à doença na cidade.

Anualmente, o SUS registra em suas bases de dados, milhares de casos de Dengue espalhados por todo o estado do Espírito Santo. Com isso, ao inspirar-se em iniciativas como os estudos acima apresentados, e sabendo da importância da extração de informações nas bases de dados do governo, é de extrema importância para o auxílio do controle e prevenção desta doença no estado, que se realize um estudo dos registros de Dengue, por intermédio de uma ferramenta computacional voltada para a análise de dados relacionados a um mesmo assunto, denominado DWG.

2.3. DATA WAREHOUSE

Muitos sistemas de informações atualmente operam sobre banco de dados transacionais, que contém informações detalhadas que permitem às instituições acompanhar e controlar seus processos operacionais. Em contrapartida, existe uma crescente demanda, por parte dos gestores dessas instituições, por sistemas de informação que auxiliem no processo de tomada de decisão, o que resulta na necessidade de recursos computacionais que forneçam subsídios para apoiar o processo decisório, sobretudo a níveis táticos e estratégicos da instituição (PIRES, 2011).

Goldschmidt (2005) e Inmon (1997), definem um *Data Warehouse* (DW) como um conjunto de dados que se baseiam em assuntos, de forma integrada, não volátil, variável ao tempo e destinado a auxiliar no processo de tomada de decisão do negócio.

Elmasri e Navathe (2011) caracterizam o DW como provedor de acesso a dados para análises complexas, descoberta de conhecimento, e tomada de decisão. Ele dá suporte a demandas de alto desempenho sobre os dados e informações de uma organização.

O DW possui características conforme as apresentadas na Figura 2. Em seguida, são descritos detalhes dessas características conforme descreve Goldschimidt (2005).

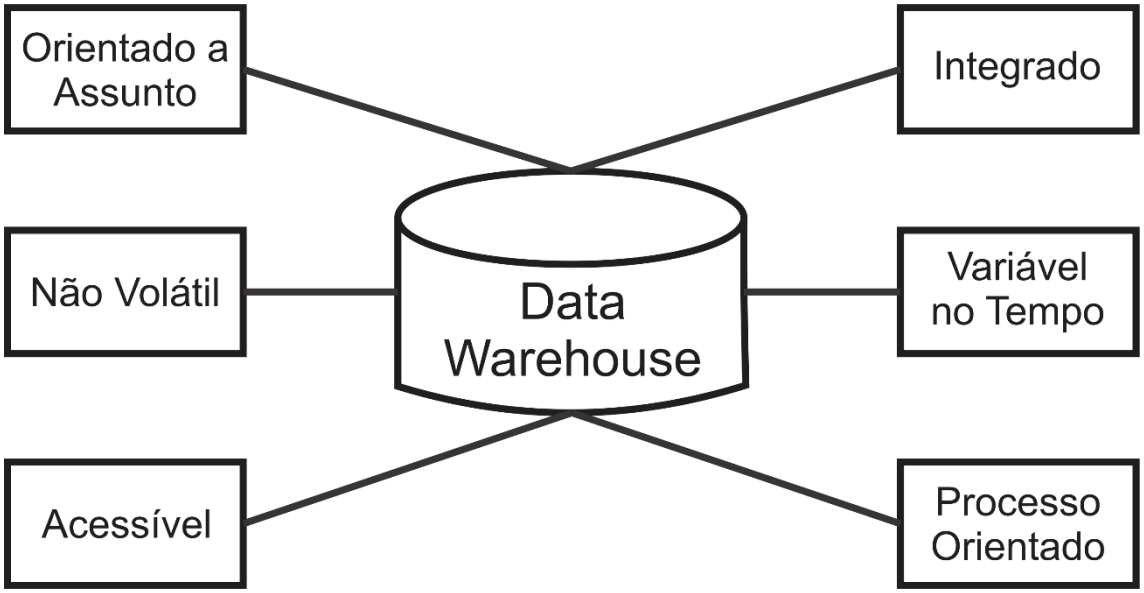


Figura 2 – Características do *Data Warehouse*.

Fonte: Saurabh et al., 2011.

2.3.1. Orientado a assunto

No DW as informações são armazenadas e apresentadas de acordo com um assunto específico, de modo a apresentar informações apenas sobre um determinado tema.

2.3.2. Integrado

Os dados operacionais, devem ser integrados e consolidados no DW, independente da fonte. Diferentes formatos em fontes diversas são integradas em um único formato no DW. A este processo de transformação dos dados, chama-se de integração dos dados.

2.3.3. Não volátil

Uma vez carregados no DW, estes dados não poderão mais sofrer alterações. A informação é consistente, independentemente de quando o repositório for acessado.

2.3.4. Variável no tempo

Representa a história dos dados bem como sua informação atual. A informação histórica é importante em um DW, pois com ela é possível analisar o passado; relacionar com informações para o presente e permite previsões para o futuro.

2.3.5. Acessível

O principal objetivo de um DW é fornecer informações de fácil acesso para os usuários finais.

2.3.6. Processo Orientado

É importante considerar que a forma como deve ser visto um DW é de um processo para a entrega de informação. Por isso, a manutenção de um DW é contínua e interativa.

Santos e Gutierrez (2008) apud Pires (2011) dividem o *Data Warehouse* em quatro elementos: dados operacionais; processos de carga (ferramentas ETL); informações analíticas (ferramentas OLAP); metadados. Na Figura 3 é representado o esquema dessa divisão para este trabalho. Kimball (2002), apresenta diferenças como: sistemas operacionais (fonte dos dados); *data staging area*; apresentação de dados e ferramentas de acesso aos dados.

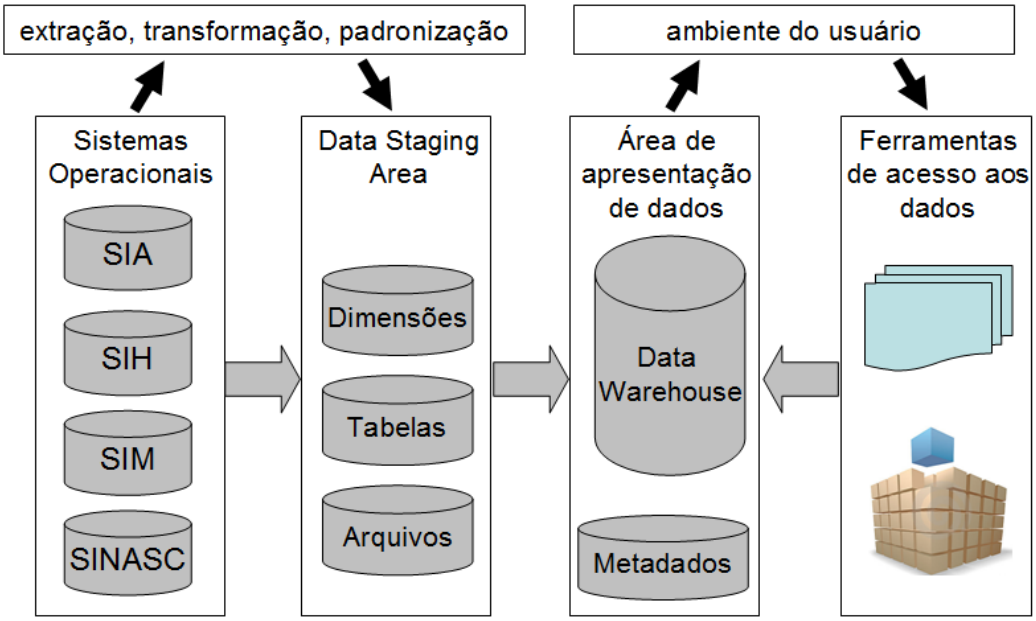


Figura 3 – Diagrama dos elementos do DW.

Fonte: PIRES, 2011.

Os sistemas operacionais destacados na Figura 3 são os chamados sistemas relacionais, que são responsáveis pela captura de todo o processo de transação das instituições. Santos e Gutierrez (2008) apud Pires (2011) os classificam como sendo os sistemas Processamento de Transações em tempo real (OLTP). Por se ter pouco, ou nenhum, controle sobre o conteúdo e o formato dos dados armazenados nesses sistemas operacionais, deve-se extraí-los, transformá-los e padroniza-los externamente ao DW.

Após os dados padronizados, temos o chamado *Data Staging Area*, que pode ser considerada como uma área de armazenamento com um conjunto de processos denominados como Extração, Transformação e Carga (ETL). O *data staging area* abrange tudo o que está entre os sistemas operacionais e a área de apresentação dos dados aos usuário do DW.

Na área de apresentação dos dados, temos o DW devidamente carregado, respeitando as características citadas acima. Os dados são armazenados de forma organizada e tornam-se disponíveis para serem consultados diretamente pelos usuários, geradores de relatórios ou por outras ferramentas de análise. Kimball (2002) refere-se à área de apresentação de dados como uma série de *data marts* integrados, sendo o *data mart* uma parte do todo que compõe a área de apresentação. O *data mart* pode ser definido ainda como uma representação dos dados de um único processo de negócio.

Por último, temos as ferramentas de acesso aos dados que, segundo Kimball (2002), pode ser simples como uma ferramenta de consulta ou tão complexa como uma ferramenta de exploração e modelagem de dados. Outra forma de se ter acesso aos dados, se dá pelo uso de mapas de calor, que a partir dos dados geográficos sobre os locais, gera-se informações plotadas em um mapa do local em questão, apresentando uma sumarização das informações disponibilizadas pelo DW.

É importante destacar algumas operações básicas que tais ferramentas de acesso aos dados devem disponibilizar aos usuários, conforme destacadp por Goldschmidt (2005):

- *Drill down/ drill up*: são utilizadas para aumentar ou diminuir o grão de detalhes da informação acessada. Por exemplo, a unidade federativa (UF) é um maior grão, quando comparada às cidades que pertencem àquela UF.
- *Slicing*: permite a seleção/escolha das dimensões a serem utilizadas na consulta. Por exemplo, visualizar a quantidade de registros de Dengue separados pelo ano e ocupação dos pacientes.
- *Dicing*: esse limita um conjunto de valores a serem exibidos, através de filtros nas dimensões. Exemplo: quantidade de registros de Dengue em 2012 na cidade de Vitória – ES.

Com isso, temos que as ferramentas de acesso aos dados, também chamadas de ferramentas OLAP e SIG, são responsáveis pela leitura dos dados do DW e consequentemente pela produção das informações.

2.4. *DATA WAREHOUSE* GEOGRÁFICO

Segundo Fonseca et al. (2007) *Data Warehouse* Geográfico (DWG) pode ser definido como uma extensão da abordagem tradicional de DW, acrescentando-se um componente geográfico. Basicamente, isso se resume à adição de atributos como coordenadas geográficas e também descrição do nome do local na dimensão correspondente no DW.

Um DWG deve manter as propriedades básicas de um DW, ou seja, orientado ao assunto, integrado, não volátil e variante no tempo, bem como oferecer suporte ao armazenamento, indexação, agregação e análises em mapas ou tabelas, provenientes de dados georeferenciados.

Ainda há poucos estudos que envolvam a utilização do DWG como parte de sua metodologia. Um estudo realizado pela Universidade Federal de Pernambuco conduzido por Fonseca et al. (2007), propôs o GeoDWM (Metamodelo para *Data Warehouse* Geográfico ou *Geographical Data Warehouse Metamodel)*, no qual

foi definido um conjunto de estereótipos e pictogramas e também especificou-se a forma como os diferentes tipos de dimensões e de medidas (convencionais e geográficas) podem ser associadas e organizadas para que se possa modelar um Armazém de Dados Geográficos.

Silva et al. (2004) propôs uma linguagem de consulta geográfica-multidimensional, a GeoMDQL, que buscou integrar sistemas SIG e OLAP, para que pudesse, em um único ambiente, dar suporte às decisões estratégicas baseadas em um *Data Warehouse* Geográfico.

Pereira et al. (2009) desenvolveu junto à Universidade Federal do Estado do Rio de Janeiro (UNIRIO) uma proposta de *benchmark* para *Spatial Online Analytical Processing* (SOLAPs), a qual inclui um modelo de dados, mecanismos para geração dos dados e exemplos de consultas multidimensionais geográficas. Nesse estudo foi possível identificar que o volume de dados de um DWG e a natureza complexa das operações geográficas requer uma estratégia de otimização para recuperação e agrupamento destas informações, que é possível através das ferramentas SOLAP.

Outro interessante estudo que envolve a utilização de DWG, foi a proposta do *SB-Index*, por Siqueira et al. (2009), que representa um eficiente índice para DWG, pois permite, entre outras coisas, consultas multidimensionais com predicados espaciais; provê suporte a hierarquias espaciais predefinidas; introduz o uso do índice Bitmap em DWG, herdando as suas vantagens (como as técnicas de codificação, *binning* e compressão); computa o predicado espacial e o transforma em um predicado convencional; e requer uma quantidade reduzida de espaço para armazenamento.

Outra iniciativa de sucesso que incentiva a adoção da tecnologia como um facilitador para o aumento da qualidade da saúde pública, foi o DW desenvolvido pela Universidade do Sul da Flórida para o gerenciamento da saúde comunitária, do que se extraíram relatórios completos e personalizados a partir dos dados relacionados à saúde, o que possibilitou análises rápidas e eficazes (BERNDT et al., 2003).

Este trabalho, aposta na utilização do DWG para auxiliar na identificação de novas informações sobre casos de Dengue no estado do Espírito Santo, que possam contribuir com a gestão da Saúde Pública não só no estado, mas também em todo o contexto nacional.

3 METODOLOGIA

O procedimento empregado durante o desenvolvimento do projeto pode ser dividido nas etapas que serão detalhadas abaixo. Para o auxiliar o entendimento delas, é ilustrado na Figura 4 todo o processo utilizado.

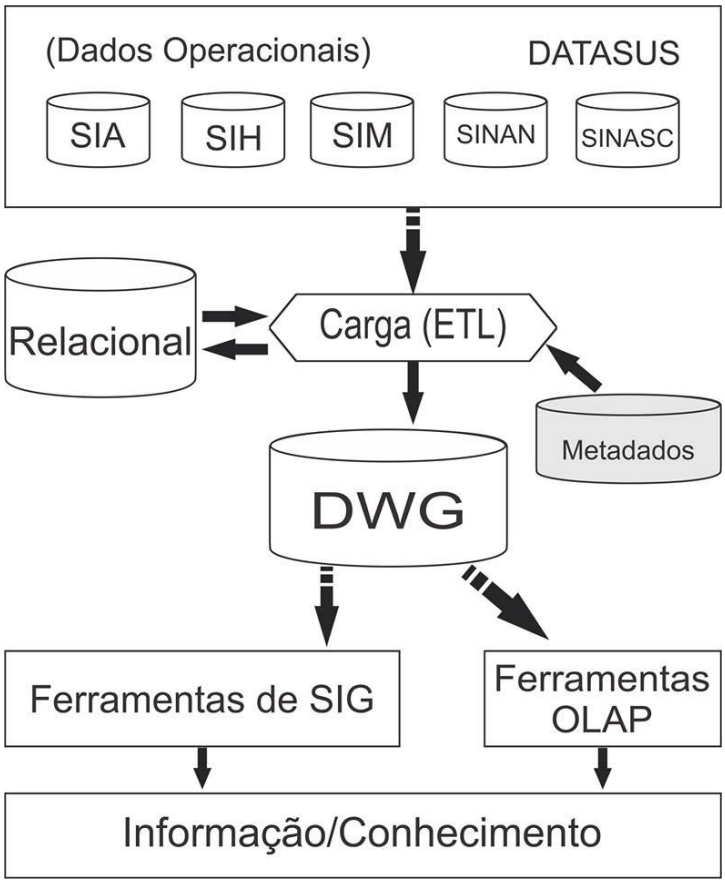


Figura 4 – Escopo Global do Projeto.

O passo inicial à execução do projeto se deu pela aquisição dos dados provenientes dos registros de incidências de Dengue, junto à Secretaria do Estado de Saúde do Espírito Santo (SESA). É importante destacar que do total de 215.468 registros do período de 2002 a 2012, foram identificados cerca de 349 que tiveram o registro da Dengue em cidades fora do Espírito Santo. Logo, esses dados foram descartados para a etapa de análise dos dados. Portanto, o estudo foi realizado com base nas informações dos 215.119 registros remanescentes.

Na segunda etapa do projeto, foi realizada uma análise dos metadados de Dengue, disponibilizados pelo DATASUS, para a identificação da modelagem de armazenamento dos dados utilizada nos sistemas do SUS. Como exemplo, o nome das colunas (atributos) e seu tipo (*integer, float, date, varchar)* bem como o seu significado.

Em paralelo a essa atividade, foi proposto um modelo multidimensional para armazenar as informações sumarizadas. Esse modelo (Figura 5) foi elaborado a partir das características de um DWG que, como explicado anteriormente, além das características tradicionais de um DW, tem a presença dos atributos geográficos na dimensão "Local". Com base nesse modelo, foram criados as tabelas no Sistema de Gerenciamento de Banco de Dados (SGBD) MySQL 5.6.

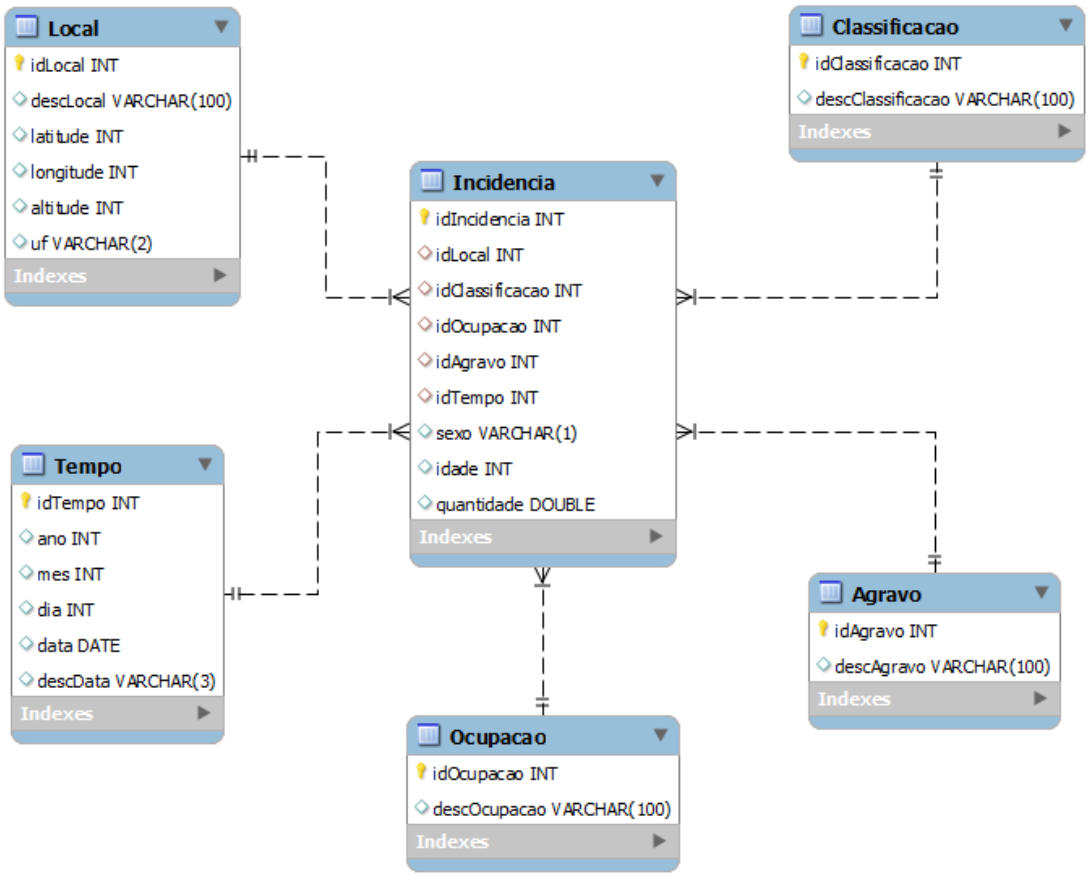


Figura 5 – Modelo Multidimensional.

No Quadro 1, destaca-se um resumo das informações do fato Incidência do modelo multidimensional proposto.

Descrição das dimensões do fato Incidência		
Dimensão	**Significado**	**Exemplos**
Local	Município onde ocorreu o registro da Dengue.	Alegre; Vitória.
Tempo	Data da ocorrência do registro da doença.	05/08/2009; 27/03/2010
Ocupação	Profissão do indivíduo que foi registrado com a doença.	Professor; Médico.
Agravo	Nome da doença.	Dengue.
Classificação	Classificação do grau de intensidade da doença.	Dengue Clássico; Dengue Hemorrágica.
Sexo	Sexo do indivíduo.	M; F.
Idade	Idade do indivíduo.	45 anos; 18 anos.

Quadro 1 – Descrição das dimensões do fato Incidência.

Em seguida, foi construído uma ferramenta *Extact Transform and Load* (ETL), na plataforma Java 6 no ambiente de desenvolvimento NetBeans IDE 7.4, que permite que seja modelado uma estrutura relacional (*Data Staging Area*) dinamicamente de um banco de dados, para que se possa efetuar uma carga intermediária dos dados, de forma a obter apenas os dados de interesse ao projeto, que seja compatível com o formato multidimensional produzido na etapa anterior.

Para se obter informações como a idade do indivíduo que registrou-se com Dengue, a aplicação ETL transforma os atributos DT_NASC e DT_NOTIFIC, que representam a data de nascimento e a data de notificação do paciente. A partir da subtração destes campos, obtêm-se o atributo idade. Todas as transformações realizadas na etapa do *Data Staging Area* estão representadas no Anexo A, pelo diagrama UML das transformações, conforme metodologia descrita por Colaço Júnior (2004).

Como pode se observar na Figura 6, na tela inicial do aplicativo ETL deve-se selecionar a base de dados com extensão DBF. Após selecionado, o aplicativo irá listar todas as colunas que o arquivo possui, o que possibilita uma flexibilidade na escolha dos atributos provenientes da base do DATASUS (DBF) que fará parte do *Data Staging Area* da aplicação ETL, ou seja, os elementos que integrarão a carga intermediária dos dados.

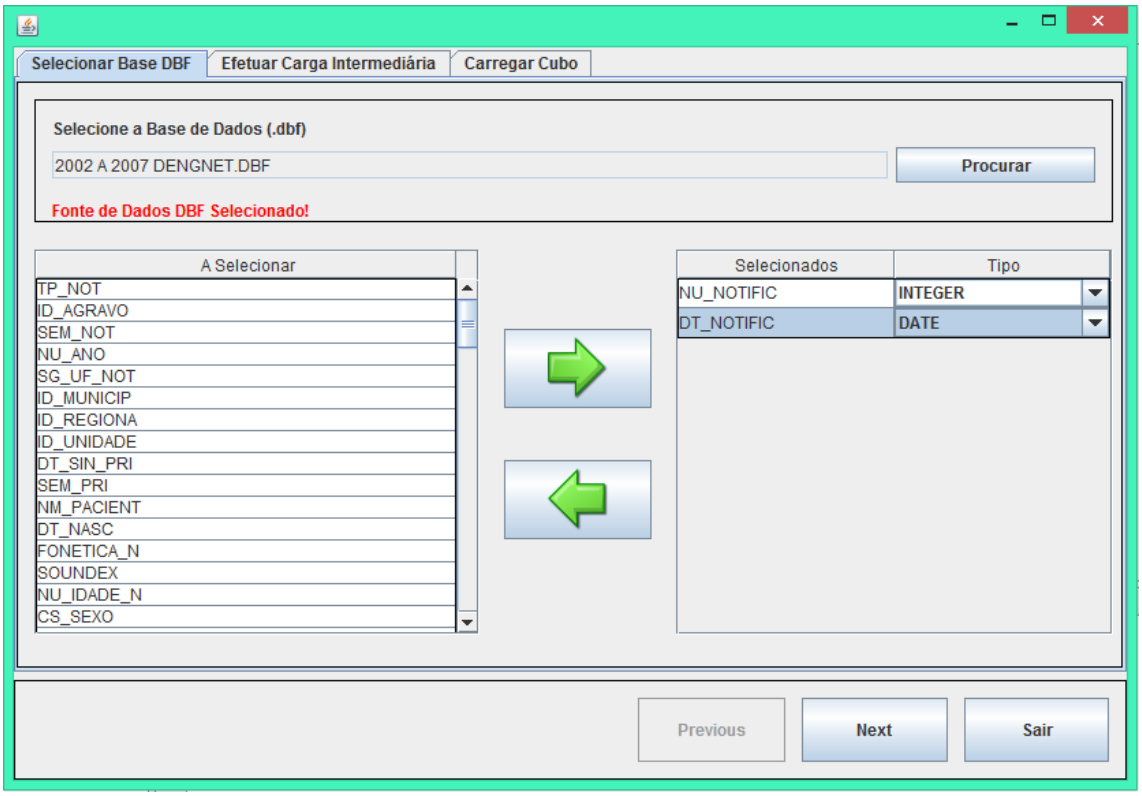


Figura 6 – Tela inicial da ferramenta ETL.

Devido ao fato de que a tabela utilizada para realizar a carga intermediária é construída dinamicamente, deve-se selecionar o tipo dos atributos que foram selecionados. Como por exemplo, na Figura 6, o atributo NU_NOTIFIC é do tipo "*INTEGER*" e o atributo DT_NOTIFIC é do tipo "*DATE*".

Após selecionado os atributos juntamente com os seus tipos, a próxima tela da aplicação (Figura 7) exibe uma pré-visualização da tabela de carga intermediária a ser criada, para fins de conferência. Ao verificar que a estrutura da tabela está correta, seleciona-se a opção "Criar" para que seja criada essa tabela no SGBD.

Com a tabela intermediária criada, é feito então a carga intermediária dos dados, por meio do botão “Carregar” da aplicação. Todo o processo de carga é exibido pela barra de progresso.

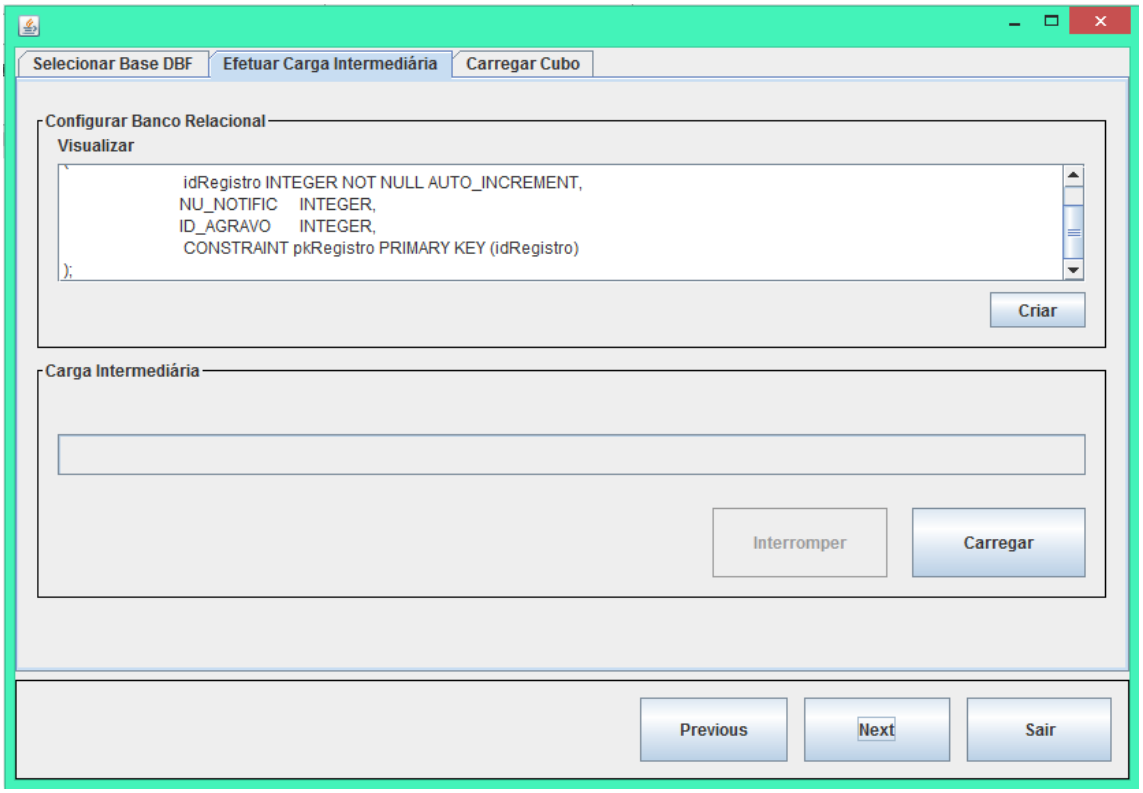


Figura 7 – Tela Carga Intermediária.

Com o banco de dados intermediário (*Staging Area)* carregado, finalmente é feito a carga de dados no DWG elaborado (Figura 8). Para isso, basta selecionar a opção “Carregar Cubo”. O processo de carga, pode ser acompanhado pela barra de progresso “Andamento da Carga”.

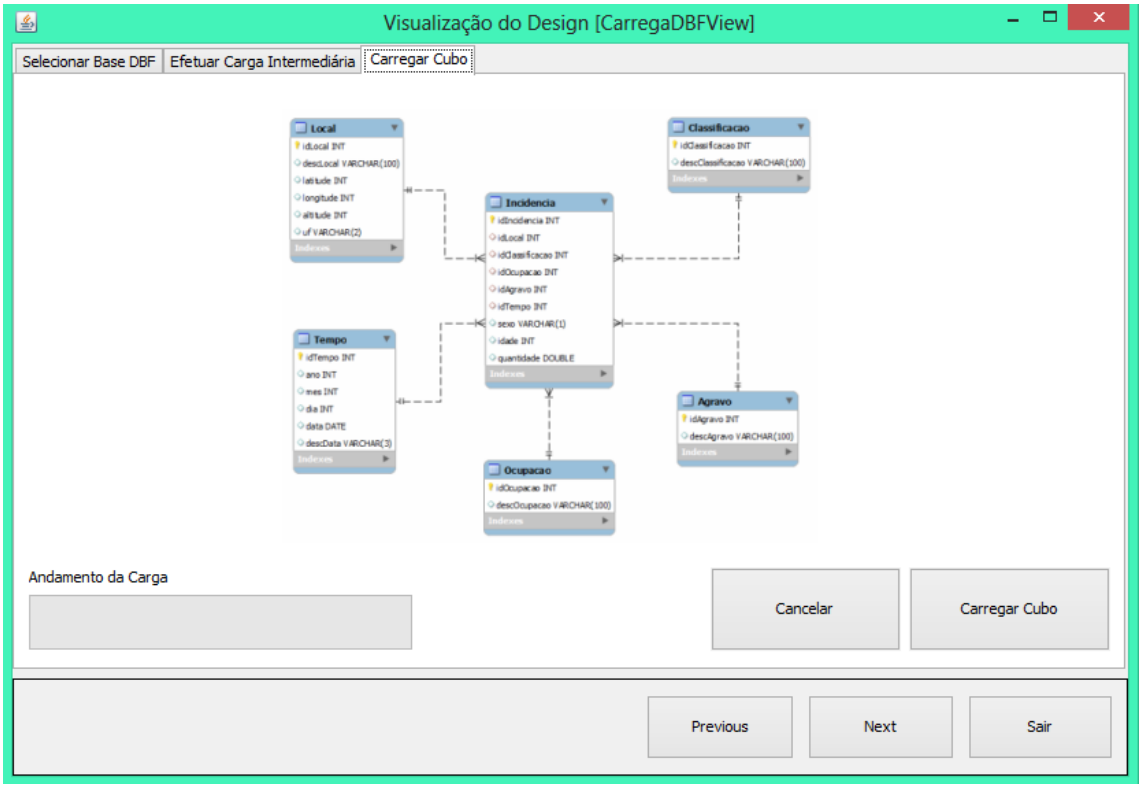


Figura 8 – Tela Carga Cubo.

As etapas finais consistem em utilizar ferramentas OLAP e SIG nos dados integrados do repositório geográfico, para assim realizar análises que trazem os objetivos de obter informações relevantes, através da aplicação dos conceitos de DW.

Os fatos gerados de acordo com a modelagem feita serão analisados no *Microsoft Excel*® 2003, por meio de gráficos descritivos e análise de regressão. Os mapas de calor foram gerados pelo aplicativo TerraView 4.2 (TERRAVIEW, 2010).

4 RESULTADOS

A partir do cubo gerado para o problema, podem-se realizar os mais variado tipos de análises. É possível combinar várias dimensões simultâneas para se extrair tabelas, gráficos e outros tipos de resultados estatísticos. Também é possível a realização análises descritivas e ou analíticas, utilizando as dimensões separadamente, o que permite ao gestor das informações, um conjunto de opções para analisar os casos da doença sob diversas perspectivas.

Neste trabalho, foram feitos alguns estudos utilizando as dimensões “Tempo”, “Local”, “Classificação”, “Sexo” e “Idade”, em que foram feitas consultas através do fato “Incidência”, e a partir dessas consultas, foram determinadas análises estatísticas descritivas e o estudo de regressão analítica foi aplicado com os dados gerados, através de gráficos e mapas.

4.1. ANÁLISE GEOGRÁFICA

Por intermédio das informações geográficas presentes na dimensão “Local”, foi feita uma análise geoespacial dos casos de Dengue por ano, no intervalo de 2008 à 2012 (Figura 9). Foram desconsiderados para análise os anos de 2002 a 2007 por não conterem dados em quantidades significativas.

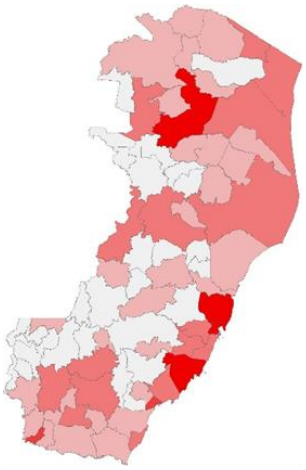


Figura 9 (a) - Casos de Dengue Ano 2008.

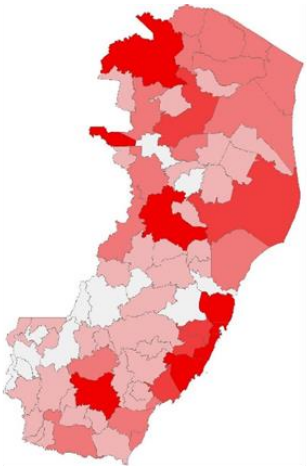


Figura 9 (b) - Casos de Dengue Ano 2009.

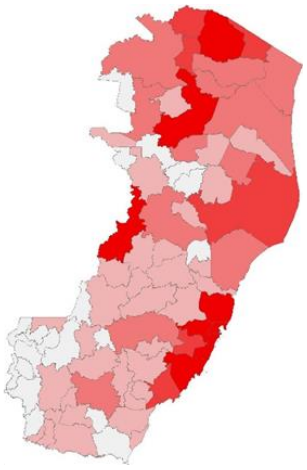


Figura 9 (c) - Casos de Dengue Ano 2010.

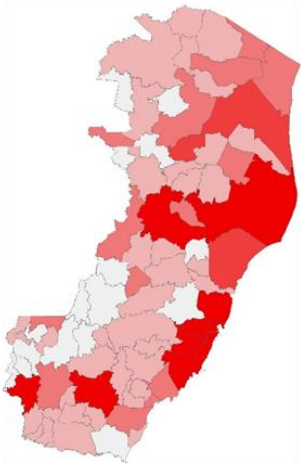


Figura 9 (d) - Casos de Dengue Ano 2011.

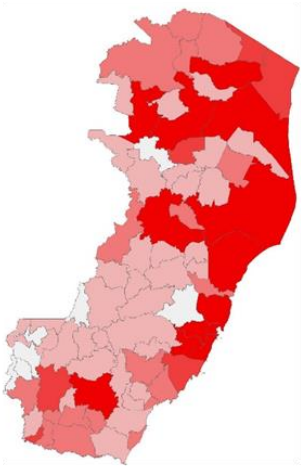
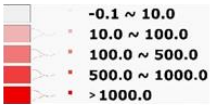


Figura 9 (e) - Casos de Dengue Ano 2012.

Legenda:



Além de uma crescente incidência de casos ao longo dos anos em análise, a região litorânea, em todos os anos, obteve mais registros de casos da doença em relação às demais regiões do estado, o que pode ser observado pelas tonalidades mais escuras dos mapas da Figura 9.

4.2. ANÁLISES DESCRITIVAS

A dimensão “Tempo”, foi modelada para que se possa obter a informação do dia da semana em que foram registradas incidência de Dengue (Figura 10).

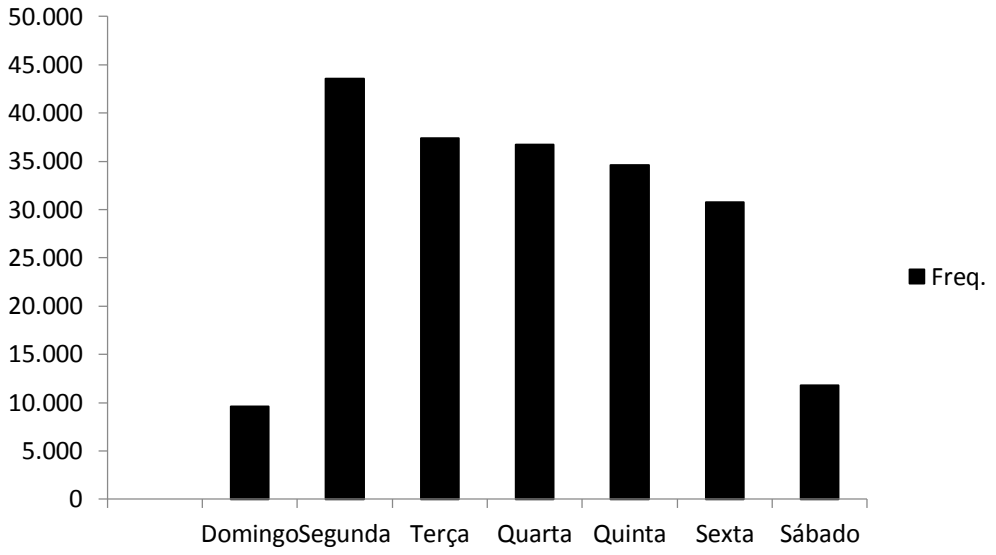


Figura 10 – Gráfico de Frequência de Incidência x Dia da semana.

Nota-se, que o ápice do registro de incidência ocorreu em uma segunda-feira, o que leva a interpretação de que, possivelmente, devido ao fato de que a maioria dos postos de saúde não funcionam aos fins de semana, apenas clínicas e hospitais, o número de registros são menores nestes dias.

Ainda com a dimensão “Tempo”, podemos analisar o registro de incidência por ano (Figura 11). Vale ressaltar que foi analisado o período de 2007 a 2012, excluindo os dados de 2002 a 2006 por não possuírem registros em quantidades significativas para análise.

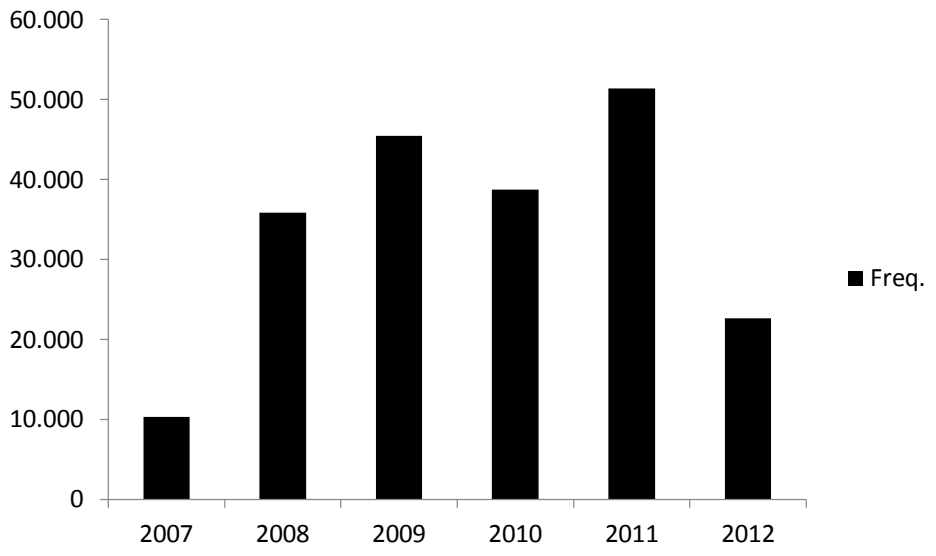


Figura 11 – Gráfico de Casos de Dengue x Ano.

Os anos de 2009 e 2011, foram os anos em que mais houve registro de casos de Dengue no Espírito Santo, em torno de 50 mil casos, em contrapartida, em um estudo realizado por Alves et al. (2011), na cidade de Aracajú – SE, o ano de 2008 foi o que mais teve casos confirmados de Dengue. Levando em consideração que são estados diferentes, os fatores climático dessas regiões, campanhas de combate à Dengue e/ou outros fatores, podem explicar essa diferença.

Ainda na dimensão “Tempo”, ao efetuar a operação de *Drill Down*, pode-se visualizar os registros de casos de Dengue por mês (Figura 12).

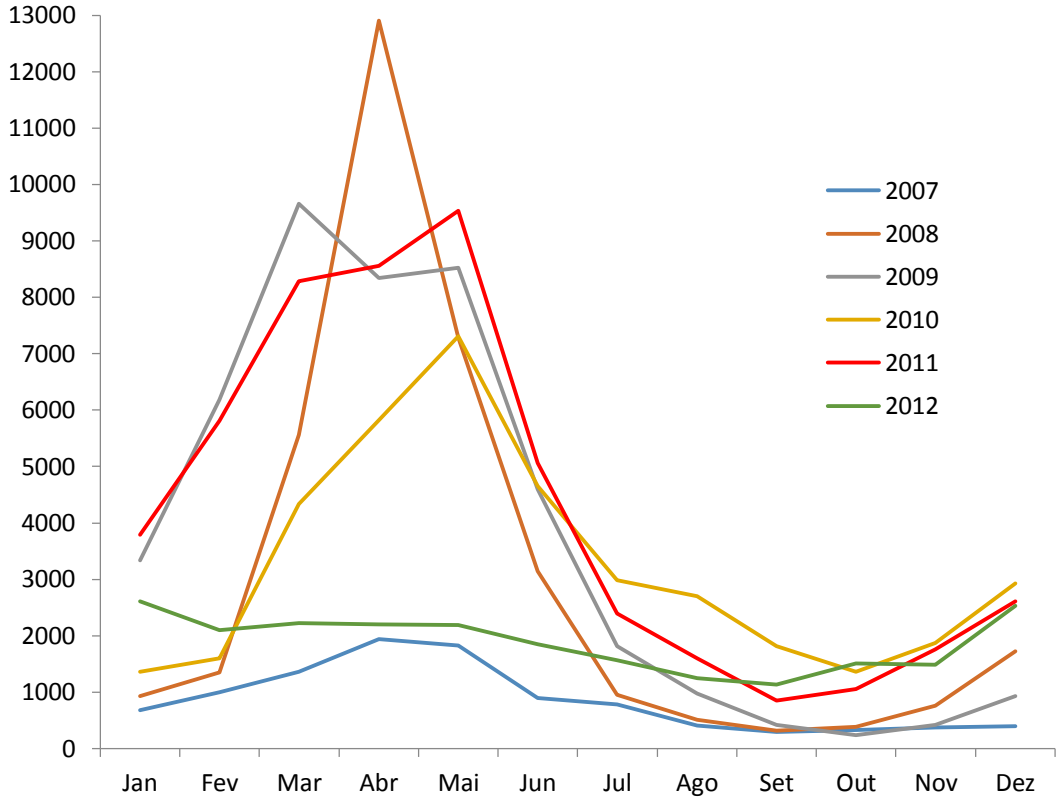


Figura 12 – Gráfico de Frequência de Casos de Dengue x Mês ao longo de 6 anos.

A partir desta figura, percebe-se que no primeiro semestre é que mais se concentram os registros de Dengue. Resultado similar foi encontrado por Alves et al (2011), em que o maior número de registros ocorreu entre os meses de janeiro e junho.

Com a dimensão “Classificação”, analisamos os casos mais comuns registrados nas bases de dados do SUS (Figura 13).

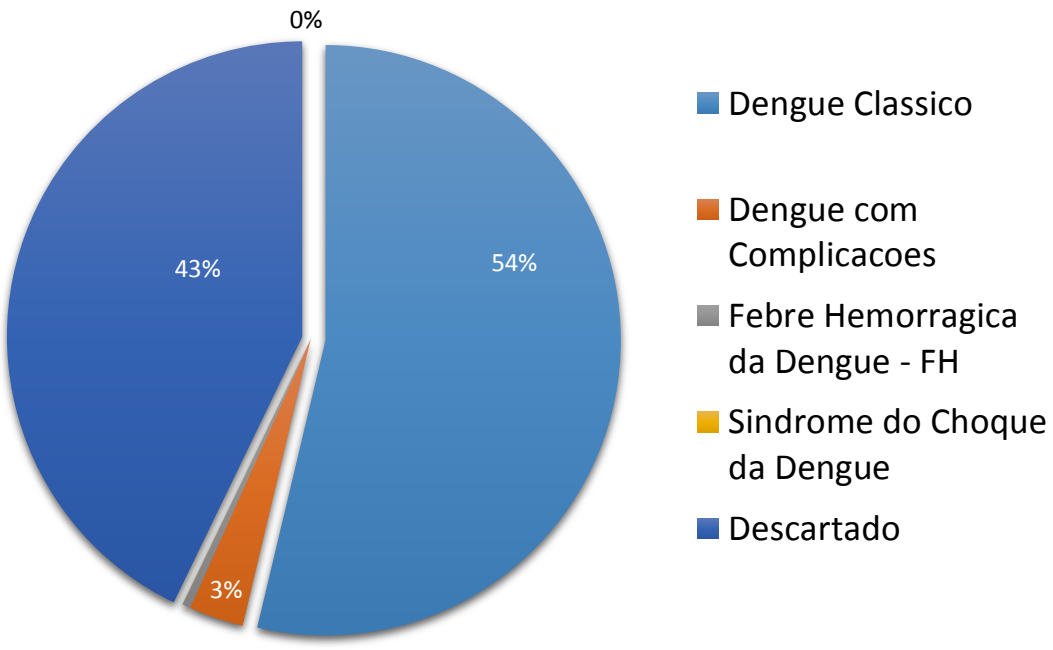


Figura 13 – Gráfico percentual de registros de Dengue de acordo com sua classificação.

A maioria dos casos são do tipo “Clássico”, com 54% dos casos. Os tipos mais graves da doença, como pode ser observado na Figura, não são tão comuns no Espírito Santo, pois somados, chegam cerca de 3% do total. Os demais casos foram classificados como “Descartados”, ou seja, casos não confirmados de Dengue.

Pela dimensão “Sexo”, pode-se visualizar os registros de Dengue por ano de acordo com o sexo (Figura 14).

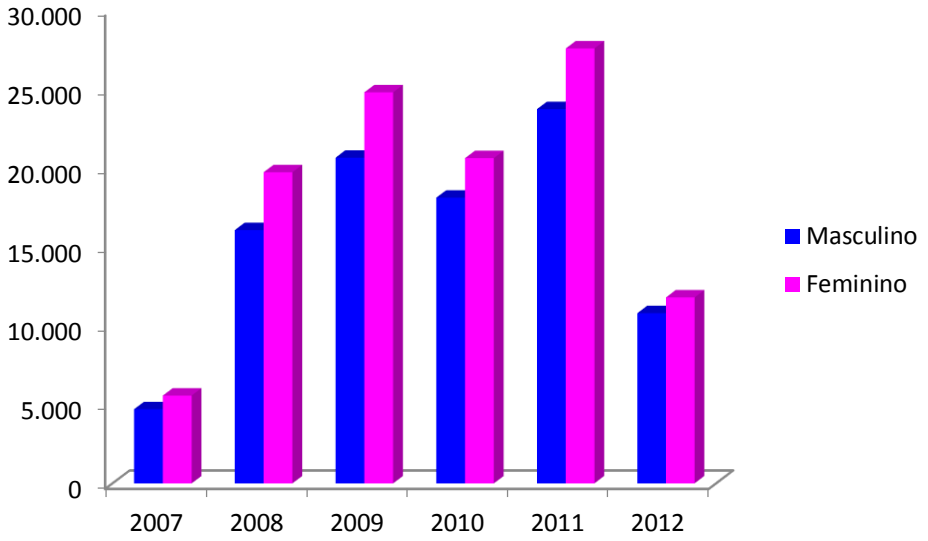


Figura 14 – Gráfico de registro de Dengue por ano de acordo com o sexo.

A partir deste gráfico percebe-se que há uma predominância do sexo feminino, durante todo o intervalo de tempo analisado. Em todos os anos analisados, a população feminina foi a que teve o maior número de registros. Podendo também perceber que com o maior número de registros foi o de 2011 com 27.500 registros do sexo feminino e 23.733 do sexo masculino. A média do número de registro no período de 2007 a 2012, foi de 15.696 (± 6.919) do sexo masculino e 18.361 (± 8.233) do sexo feminino.

Dados similares foram encontrados por Ferraz et al. (2013) e Alves et al. (2011), em que a maioria da população que contraiu Dengue é do sexo feminino. Assir, Masood e Ahmad (2014), obteve dados diferentes, onde o maior registro de casos de Dengue foi do público masculino, vale ressaltar que esse estudo foi realizado em outro país (Paquistão).

Na dimensão “Idade” pode-se visualizar a frequência de registros do agravo de acordo com as idades registradas na base de dados (Figura 15).

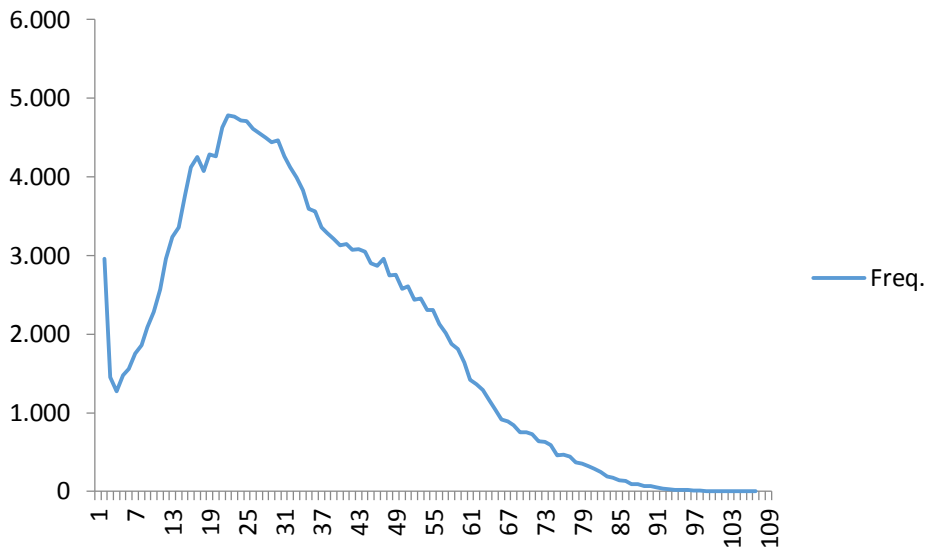


Figura 15 – Gráfico de frequência de registros por idade no período de 2007 a 2012.

A partir da Figura 15 pode-se observar que as idades entre 13 e 37 anos parecem ser mais suscetíveis à contrair a doença, com média de idade de 25 anos (± 7). Alves et al. (2011), encontrou um alto número de casos confirmados em pessoas com idades entre 15 e 49 anos. A média de idade com maior número de registros foi de 44 (±16) anos, no estudo realizado por Ferraz et al. (2013). A média de idade encontrada por Assir, Masood e Ahmad (2014), foi de 22 anos. Com base em todos os estudos é possível afirmar que o público mais propício a contrair a doença são os jovens e adultos.

4.3. ANÁLISE DE REGRESSÃO

Tendo em vista que temos disponível em nossa base de dados as informações de todos os indivíduos que realizaram o exame de diagnóstico de Dengue no estado do Espírito Santo, a análise estatística se dá em um Universo de pessoas e não em uma amostra aleatória.

A partir das variáveis disponibilizadas é possível se observar o efeito que algumas variáveis podem possuir sobre a probabilidade de que o resultado do exame seja positivo, ou seja, que o indivíduo em questão seja diagnosticado com Dengue.

Para isso, adotou-se a regressão Probit, na qual a variável dependente é a probabilidade de um evento ocorrer. Logo, os coeficientes estimados medem o efeito das variáveis explanatórias sobre esta probabilidade. Neste caso, as variáveis explanatórias são: 1) Idade do indivíduo, 2) Sexo (Mulher = 1, Homem = 0) e 3) Primeiros Meses (Primeiro Semestre =1, Segundo Semestre = 0).

Tabela 1 – Resultados da Regressão Probit.

Coeficientes	Estimado	Desvio Padrão	Z[1]	Valor	Pr(>|z|)
(Intercept.)	-0.307643	0.008011	-38.401	< 2e-16	***
Idade	0.002998	0.000157	19.094	< 2e-16	***
Mulher	0.040827	0.005659	7.215	5.39e-13	***
PrimeirosMeses	0.481617	0.006743	71.430	< 2e-16	***

Signif. Cod.	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 ' ' 1

[1] A probabilidade associada a estatística de teste Z, rejeita com elevado grau de confiança a hipótese nula de que o coeficiente é estatisticamente igual a zero, ou seja, que a variáveis explicativas não são relevantes.

Em primeiro lugar, a regressão mostra que todos os coeficientes são estatisticamente significantes. De acordo com os resultados, tanto a idade quanto o sexo apresentaram coeficientes pequenos e positivos, ou seja, a regressão indica que quanto mais velha a pessoa for, maior é a probabilidade de diagnóstico pelo exame. De forma análoga, podemos concluir que mulheres também possuem uma maior probabilidade de diagnóstico. Contudo, podemos esperar que o efeito destas duas variáveis seja pequeno na prática, dado o pequeno valor do coeficiente associado.

Já a variável “PrimeirosMeses” apresentou um coeficiente positivo elevado, o que indica que nos meses de mais calor, a probabilidade de diagnóstico pode aumentar consideravelmente. Contudo, os coeficientes estimados são de difícil interpretação, que necessita de uma complexa abordagem estatística que vai além do escopo deste trabalho.

Entretanto, é possível demonstrar o efeito dos Meses de Calor sobre a probabilidade de diagnóstico de uma maneira informal e intuitiva. O gráfico representado na Figura 16 apresenta três séries temporais construídas a partir dos dados fornecidos. A primeira série mostra a quantidade de exames realizados (obsmensais). Já a segunda fornece o número casos de dengue diagnosticados em uma base mensal. A terceira e última série apresenta a evolução da Razão de Descartes de Exames, isto é, aqueles em que a doença não foi detectada nos indivíduos em questão.

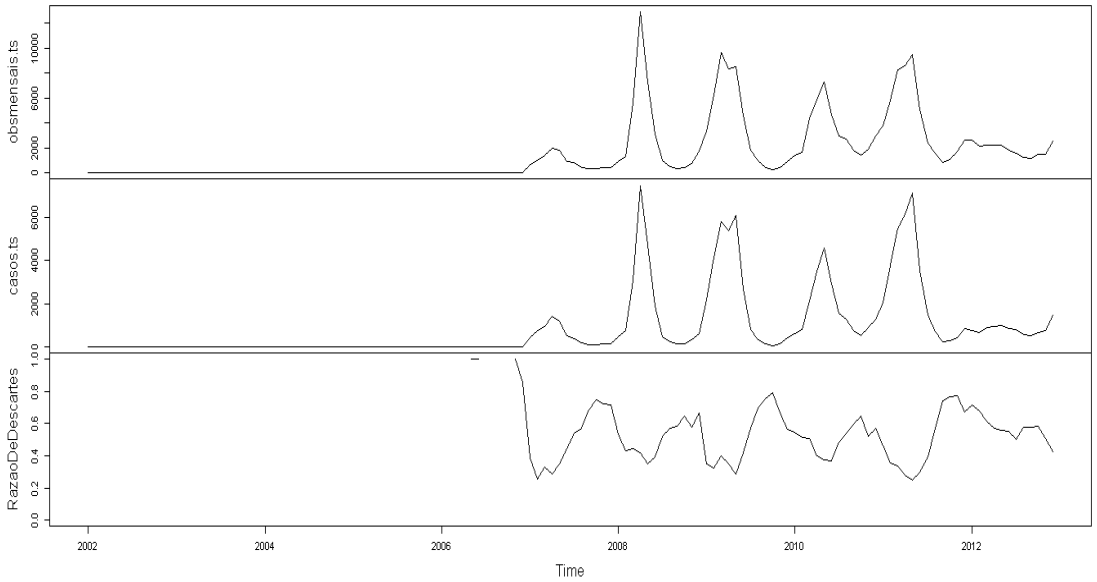


Figura 16 – Gráfico de séries temporais.

O gráfico apresentado na Figura 16, corrobora claramente com a regressão Probit apresentada. É fácil perceber que nos meses de maior incidência e exames de Dengue, a Razão de Descarte se reduz, ou seja, uma maior porcentagem de exames obtém resultado positivo. Já nos meses de poucos casos da doença, mesmo considerando a significante redução na quantidade de exames, a Razão de Descarte se eleva.

Dessa forma, podemos observar que o coeficiente associado à variável “Primeiros Meses” na regressão Probit, produz um forte impacto sobre a probabilidade de diagnóstico do exame, tendo em vista a elevada sazonalidade da Razão de Descarte entre os meses de maior e menor incidência da doença.

5 CONCLUSÃO

A partir da análise dos metadados do Departamento de Informática do SUS (DATASUS), foi possível modelar e criar um esquema multidimensional espacial para o armazenamento de dados sumarizados das bases de dados do SUS.

A ferramenta construída para fazer extração, transformação e carga dos dados dos bancos de dados do SUS para o DWG elaborado, se mostrou eficiente. Tal fato é observado de acordo com as análises estatísticas e geográficas realizadas sobre os dados sumarizados provenientes do DWG, apresentados no capítulo 4.

Com isso, os gestores públicos, de posse dessa ferramenta, terão condições de analisar os milhares de dados gerados do SUS com mais facilidade e flexibilidade, tornando a tomada de decisão para a Saúde Pública uma atividade mais rápida e eficaz.

6 REFERÊNCIAS

ACHESON R. Independent inquiry into inequalities in health: report. (Acheson Report). London: Department of Health, **The Stationery Office**, 1998.

ALVES, J. A. B. et al. Epidemiological aspects of dengue in Aracaju, state of Sergipe, Brazil. **Revista da Sociedade Brasileira de Medicina Tropical,** Aracajú, v. 6, n. 44, p.670-673, nov. 2011.

ASSIR, M. Z. K.; MASOOD, M. A.; AHMAD, H. I. Concurrent dengue and malaria infection in Lahore, Pakistan during the 2012 dengue outbreak. **International Journal Of Infectious Diseases,** Pakistan, v. 1, n. 18, p.41-46, jan. 2014.

BARCELLOS, C. et al. Identificação de locais com potencial de transmissão de dengue em Porto Alegre através de técnicas de geoprocessamento. **Revista da Sociedade Brasileira de Medicina Tropical**, Uberaba, p.246-250, jun. 2005.

BEAGLEHOLE, R. et al. Public health in the new era: improving health through collective action. **The Lancet**, San Francisco, v. 363, n. 9426, p.2084-2086, 19 jun. 2004.

BERNDT, D. J.; HEVNER, A. R.; STUDNICKI, J. The Catch data warehouse: support for community health care decision-making. **Decision Support Systems,** Tampa, p. 367-384, 2003.

BLANE, D. Health inequality and public policy: one year on from the Acheson report. **Journal of Epidemiology and Community Health**, v. 53, p. 748, 1999.

BRASIL (Org.). **Sistema Único de Saúde.** Disponível em: <http://www.brasil.gov.br/sobre/saude/atendimento>. Acesso em: 10 ago. 2013.

BRUCE, A. A.; CARRATÉ, J. A. Enfermedades reemergentes: factores causales y vigilancia. **Archivo Médico de Camagüey**, Cuba, v. 13, n. 2, 2009.

CNES (Org.). **Cadastro Nacional de Estabelecimentos de Saúde:** DATASUS. Disponível em: <http://cnes.datasus.gov.br/Lista_Tot_Es_Estado.asp>. Acesso em: 06 ago. 2013.

COLAÇO JUNIOR, M. **PROJETANDO SISTEMAS DE APOIO À DECISÃO BASEADOS EM DATA WAREHOUSE.** Rio de Janeiro: Axcel Books, 2004.

CSBC. **Congresso da Sociedade Brasileira de Computação.** Disponível em: <http://www.ic.ufal.br/csbc2013/sobre-o-csbc>. Acesso em: 11 ago. 2013.

DATASUS (Org.). **Departamento de Informática do SUS.** Disponível em: <http://www2.datasus.gov.br/DATASUS/index.php?area=01>. Acesso em: 11 ago. 2013.

DENGUE. **Sobre a Dengue.** Disponível em: <http://www.dengue.org.br/dengue.html>. Acesso em: 17 jul. 2013.

DUARTE, H. H. P.; FRANÇA, E. B. Qualidade dos dados da vigilância epidemiológica da dengue em Belo Horizonte, MG. **Revista Saúde Pública**, São Paulo, p.134-142, 2006.

ELMASRI, R.; NAVATHE, S. B. **Sistemas de Banco de Dados.** 6ª Edição, São Paulo: Pearson Education. 2011.

FERRAZ, F. O. et al. Evaluation of laboratory tests for dengue diagnosis in clinical specimens from consecutive patients with suspected dengue in Belo Horizonte, Brazil. **Journal Of Clinical Virology,** Belo Horizonte, v. 1, n. 58, p.41-46, jun. 2013.

FONSECA, R.; FIDALGO, R. N.; SILVA, J.; TIMES, V. C. Um metamodelo para a especificação de Data Warehouses Geográficos. **XXII Simpósio Brasileiro de Banco de Dados**. Recife, p.193-207, 2007.

G1 (Espírito Santo). **Mais de 51 mil pessoas tiveram dengue no Espírito Santo, neste ano.** Disponível em: <http://g1.globo.com/espirito-santo/noticia/2011/11/mais-de-51-mil-pessoas-tiveram-dengue-no-espirito-santo-neste-ano.html>. Acesso em: 02 ago. 2013a.

G1 (Espírito Santo). **ES registra mais de 5 mil casos de dengue no primeiro mês de 2013.** Disponível em: <http://g1.globo.com/espirito-santo/noticia/2013/02/es-registra-mais-de-5-mil-casos-de-dengue-no-primeiro-mes-de-2013.html>. Acesso em: 02 ago. 2013b.

GOLDSCHMIDT, R.; PASSOS, E. **Data mining: um guia prático**. Rio de Janeiro: Elsevier, 2005 – 4ª impressão.

GRAHAM, H.; KELLY, M. P. Health inequalities: concepts, frameworks and policy. **Health Development Agency**: Briefing Paper, United Kingdom, 2004.

HUNG, E. L.; RODRÍGUEZ, L. C.; PUPO, M. M. G. Sistema de Procesamiento Transaccional Y Analítico Para La Vigilancia Epidemiológica Del Dengue. **Xv Convención Y Feria International**: IX Congresso Internacional de Informática en Salud 2013, Cuba, 2013.

INMON, W. H. **Como construir o Data Warehouse**. 2.ed. Rio de Janeiro: Campus, 1997.

KIMBALL, R.; ROSS, M. **The Data Warehouse Toolkit: o guia completo para modelagem multidimensional**. Rio de Janeiro: Campus, 2002.

LORBIESKI, R. et al. UMA CONTRIBUIÇÃO AO CONTROLE DA DENGUE EM CASCAVEL/PR ATRAVÉS DE UM SOFTWARE COM INFORMAÇÕES GEORREFERENCIADAS. **II Simpósio de Inovação Tecnológica,** Cascavel - PR, out. 2010.

MINISTÉRIO DA SAÚDE (Org.). **Portal da Saúde.** Disponível em:<http://portalsaude.saude.gov.br/portalsaude/index.cfm/?portal=pagina.visualizarArea&codArea=345>. Acesso em: 09 ago. 2013.

OLIVEIRA, E. et al. **Dengue:** ABC da Saúde. Disponível em: <http://www.abcdasaude.com.br/artigo.php?101>. Acesso em: 19 jul. 2013.

PEREIRA, D. N. S.; SANTOS, V. **Modelagem de Dados e Consultas em um DataWarehouse Geográfico.** Rio de Janeiro: Ufrj, 2009.

PIRES, F. A. **Ambiente para extração de informação epidemiológica a partir da mineração de dez anos de dados do Sistema Público de Saúde.** 2011. 206 f. Tese (Doutorado) - Faculdade de Medicina da Universidade de São Paulo, São Paulo, 2011.

PORTAL DA SAÚDE. **Dengue recua, mas prevenção deve ser mantida.** Disponível em:

<http://portalsaude.saude.gov.br/portalsaude/noticia/10651/785/dengue-recua-mas-prevencao-deve-ser-mantida.html>. Acesso em: 17 jul. 2013.

SANTOS, R.S., GUTIERREZ, M. A. MINERSUS – Ambiente computacional para extração de informações para a gestão da saúde pública por meio da mineração dos dados do SUS. **Revista Brasileira de Engenharia Biomédica**, v. 24, p. 77-94, 2008.

SANTOS, R. S. et al. Data Warehouse para a Saúde Pública: Estudo de Caso SES-SP. **X Congresso Brasileiro de Informática em Saúde**, Florianópolis, p.53-58, 14 out. 2006.

SAURABH, A. K.; NAGPAL, B. A SURVEY ON CURRENT SECURITY STRATEGIES IN DATA WAREHOUSES. **International Journal Of Engineering Science And Technology (IJEST),** New Delhi, p. 3484-3488. 04 abr. 2011.

SILVA, J. et al. PROPONDO UMA LINGUAGEM DE CONSULTA GEOGRÁFICA MULTIDIMENSIONAL. **VI Brazilian Symposium On Geoinformatics,** Campos do Jordão, p. 479-489. 01 jan. 2004.

SIQUEIRA, T. L. L. **SB-INDEX: um índice espacial baseado em bitmap para data warehouse geográfico.** 2009. 118 f. Dissertação (Mestrado) - Universidade Federal de São Carlos, São Carlos, 2009.

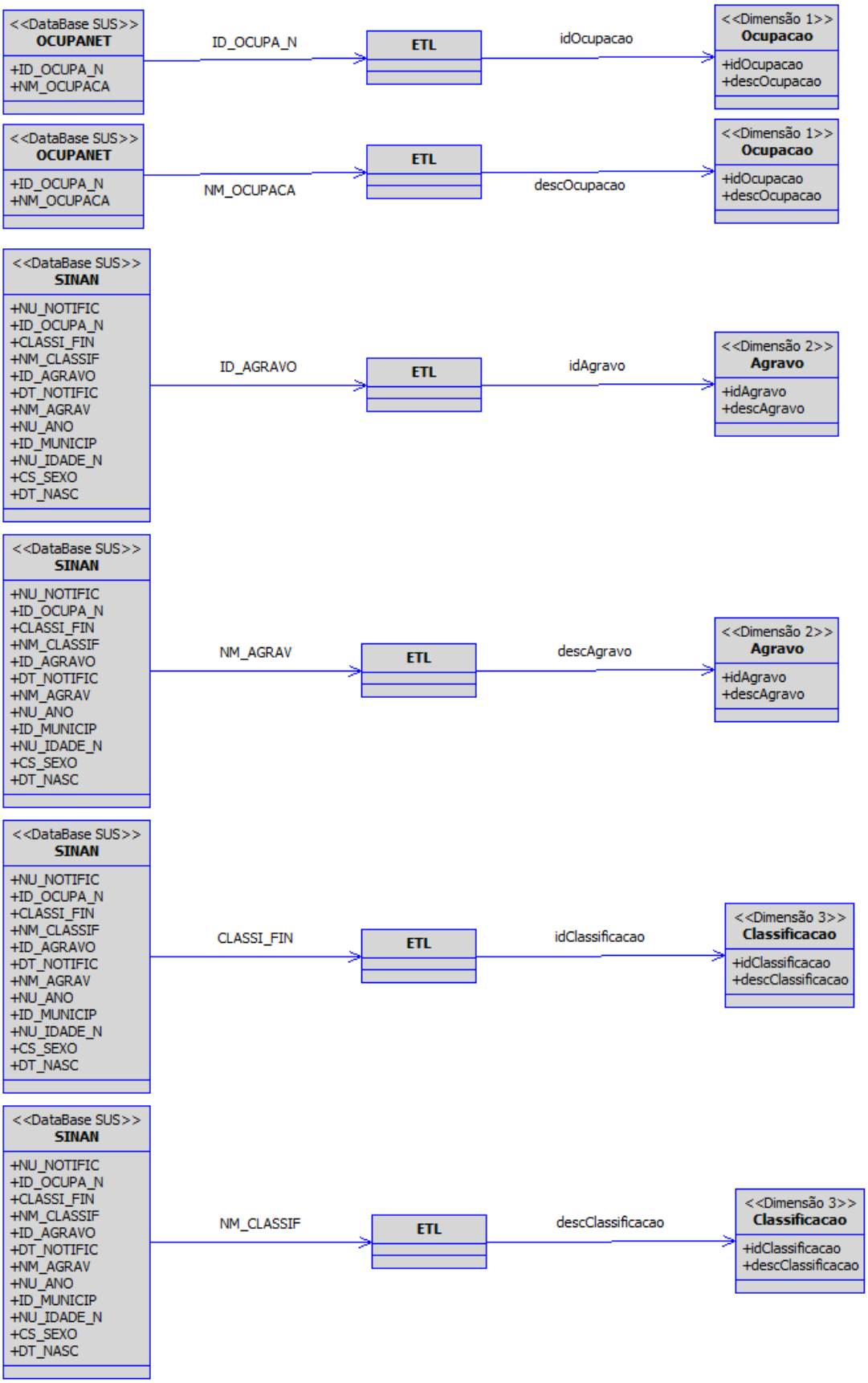
SOUZA, C. K. J.; NEVES, S. M. A. S.; COCHEV, J. S. Análise espacial da Dengue no município brasileiro de Cáceres, MT, no ano de 2009. **3º Simpósio de Geotecnologias No Pantanal**, Cáceres, p.999-1009, 20 out. 2010.

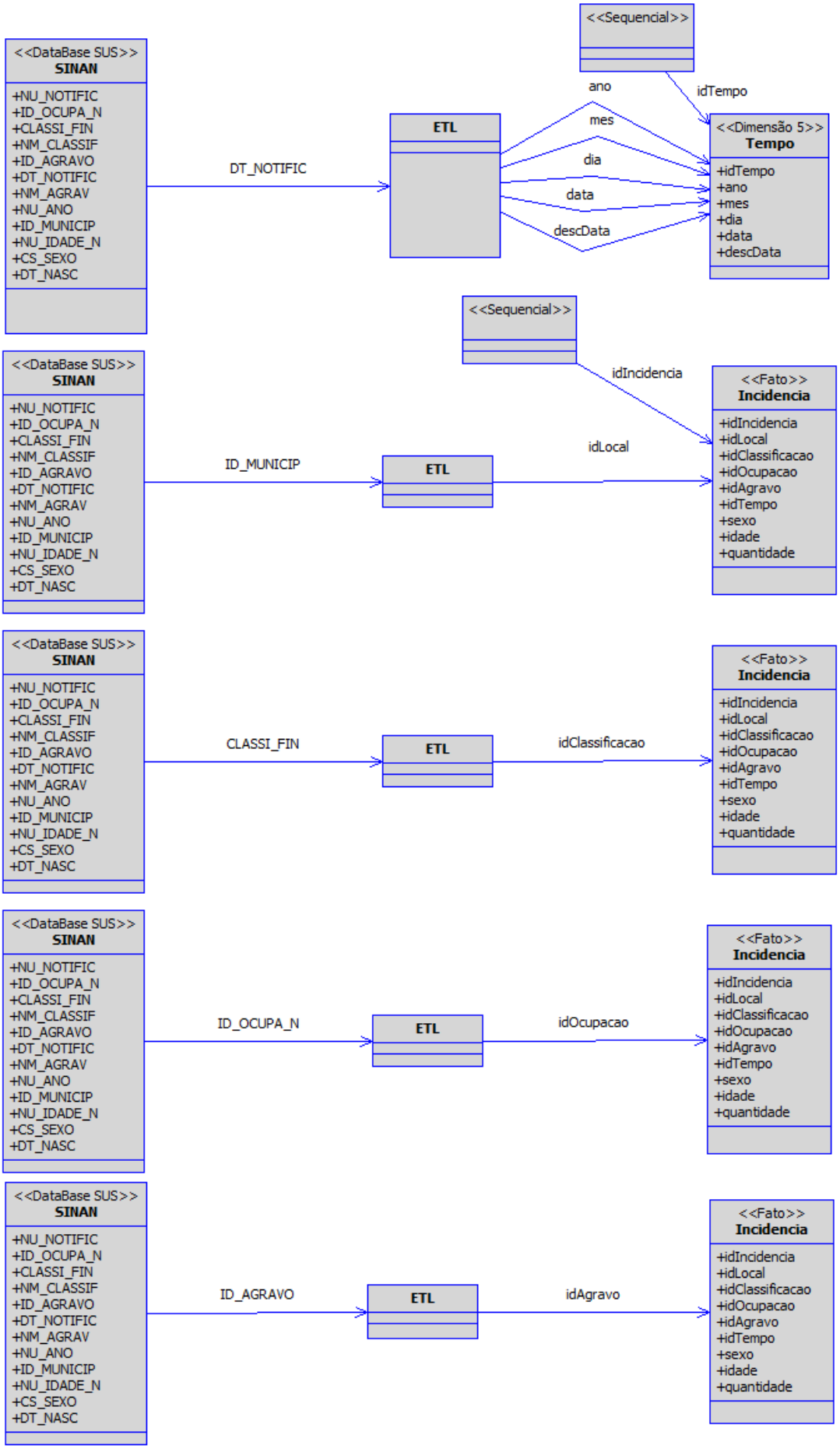
TerraView 4.1.0. São José dos Campos, SP: INPE, 2010. Disponível em: www.dpi.inpe.br/terraview. Acesso em: 13/01/2014.

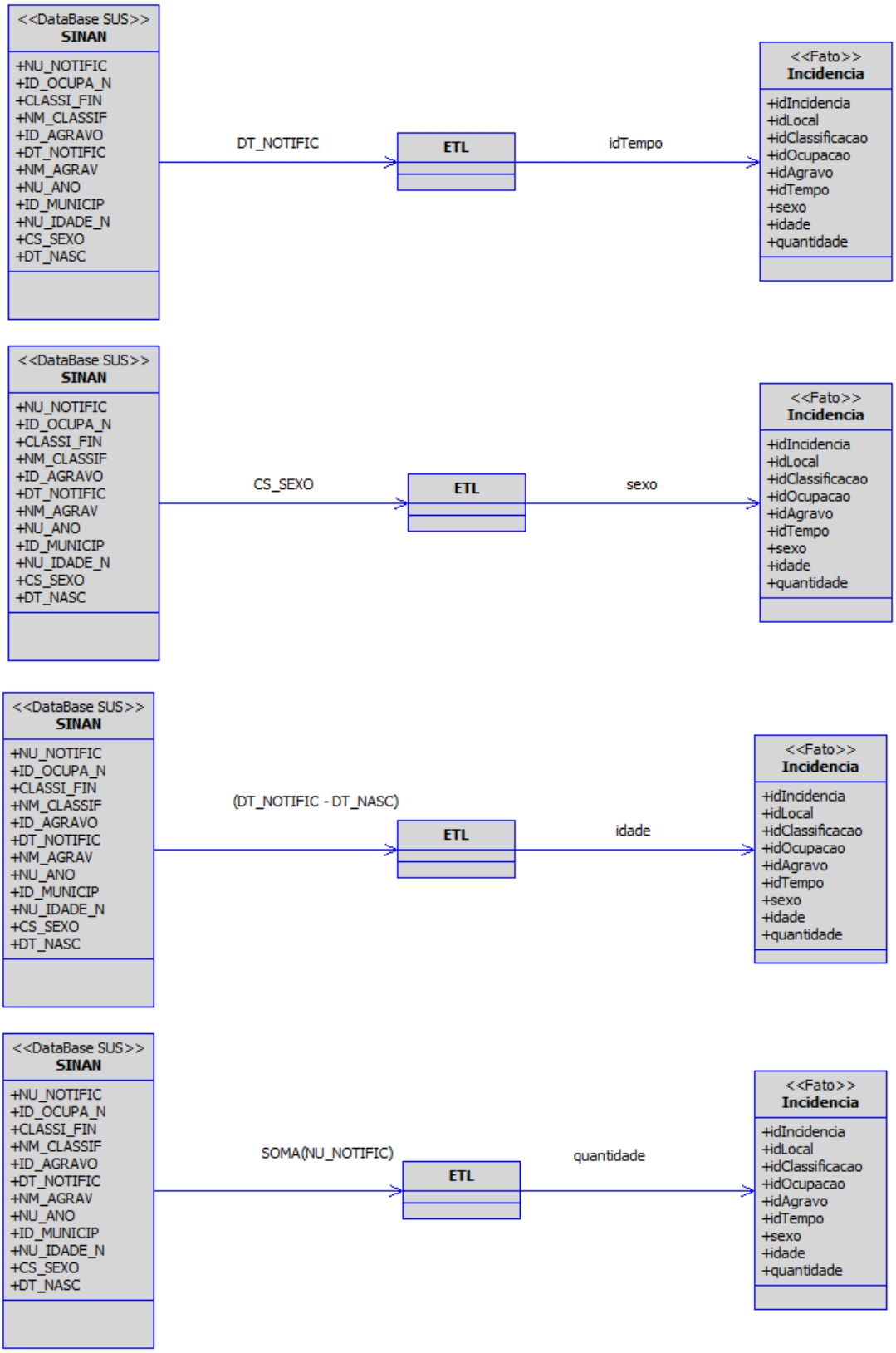
VARELLA, D. **Doenças e Sintomas - Dengue.** Disponível em: <http://drauziovarella.com.br/letras/d/dengue/>. Acesso em: 10 ago. 2013.

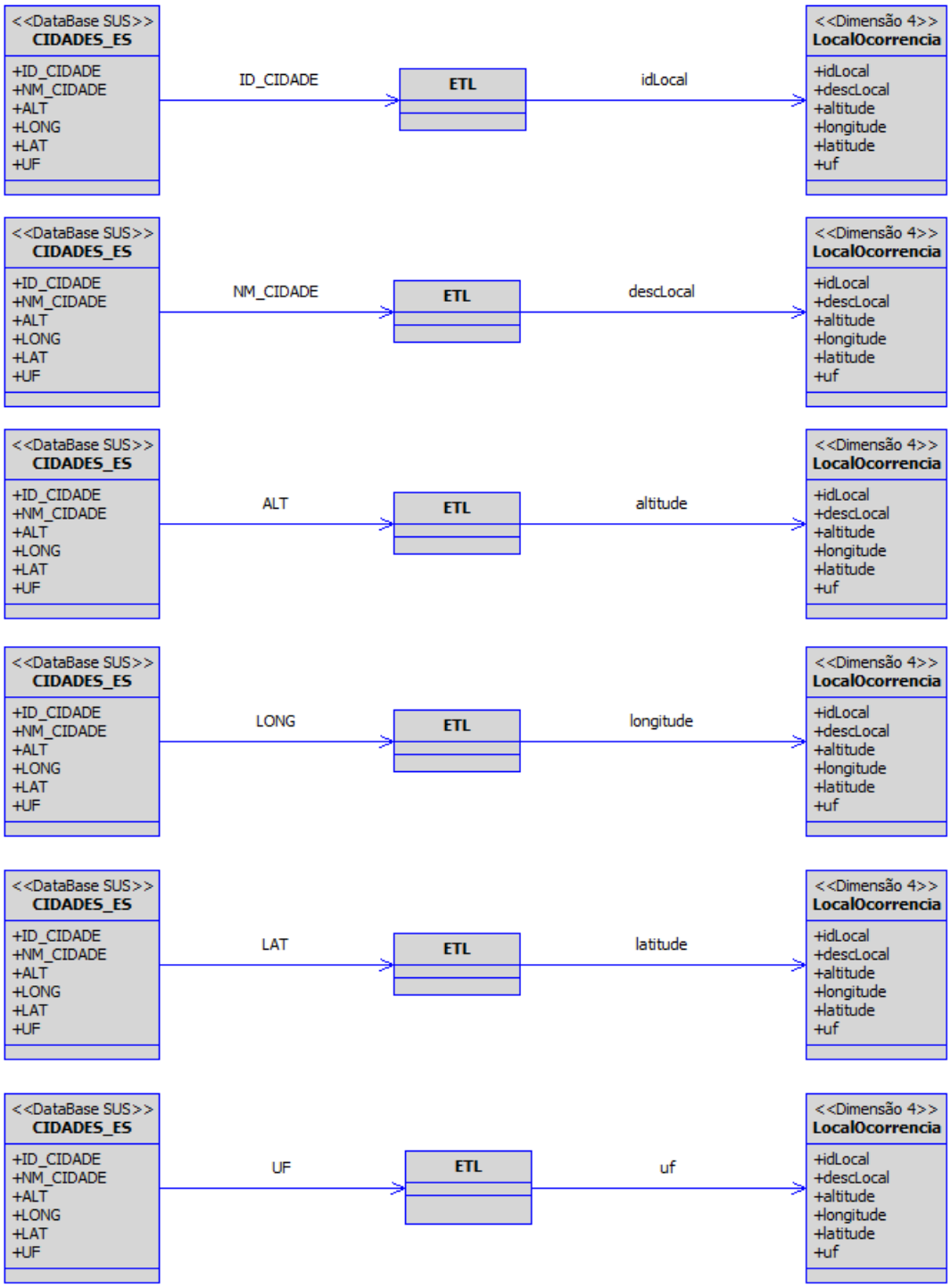
WHITEHORN, J.; FARRAR, J. Dengue. **British Medical Bulletin**, United Kingdom, p.161-173, 29 jul. 2010.

ANEXO A – DIAGRAMA UML DAS TRANSFORMAÇÕES




<<DataBase SUS>>
SINAN
+NU_NOTIFIC
+ID_OCUPA_N
+CLASSI_FIN
+NM_CLASSIF
+ID_AGRAVO
+DT_NOTIFIC
+NM_AGRAV
+NU_ANO
+ID_MUNICIP
+NU_IDADE_N
+CS_SEXO
+DT_NASC
DT_NOTIFIC
ETL
<<Sequencial>>
ano
idTempo
mes
dia
data
descData
<<Dimensão 5>>
Tempo
+idTempo
+ano
+mes
+dia
+data
+descData
<<Sequencial>>
idIncidencia
<<Fato>>
Incidencia
+idIncidencia
+idLocal
+idClassificacao
+idOcupacao
+idAgravo
+idTempo
+sexo
+idade
+quantidade
ID_MUNICIP
ETL
idLocal
CLASSI_FIN
ETL
idClassificacao
ID_OCUPA_N
ETL
idOcupacao
ID_AGRAVO
ETL
idAgravo


<<DataBase SUS>>
SINAN
+NU_NOTIFIC
+ID_OCUPA_N
+CLASSI_FIN
+NM_CLASSIF
+ID_AGRAVO
+DT_NOTIFIC
+NM_AGRAV
+NU_ANO
+ID_MUNICIP
+NU_IDADE_N
+CS_SEXO
+DT_NASC
DT_NOTIFIC
ETL
idTempo
<<Fato>>
Incidencia
+idIncidencia
+idLocal
+idClassificacao
+idOcupacao
+idAgravo
+idTempo
+sexo
+idade
+quantidade
<<DataBase SUS>>
SINAN
+NU_NOTIFIC
+ID_OCUPA_N
+CLASSI_FIN
+NM_CLASSIF
+ID_AGRAVO
+DT_NOTIFIC
+NM_AGRAV
+NU_ANO
+ID_MUNICIP
+NU_IDADE_N
+CS_SEXO
+DT_NASC
CS_SEXO
ETL
sexo
<<Fato>>
Incidencia
+idIncidencia
+idLocal
+idClassificacao
+idOcupacao
+idAgravo
+idTempo
+sexo
+idade
+quantidade
<<DataBase SUS>>
SINAN
+NU_NOTIFIC
+ID_OCUPA_N
+CLASSI_FIN
+NM_CLASSIF
+ID_AGRAVO
+DT_NOTIFIC
+NM_AGRAV
+NU_ANO
+ID_MUNICIP
+NU_IDADE_N
+CS_SEXO
+DT_NASC
(DT_NOTIFIC - DT_NASC)
ETL
idade
<<Fato>>
Incidencia
+idIncidencia
+idLocal
+idClassificacao
+idOcupacao
+idAgravo
+idTempo
+sexo
+idade
+quantidade
<<DataBase SUS>>
SINAN
+NU_NOTIFIC
+ID_OCUPA_N
+CLASSI_FIN
+NM_CLASSIF
+ID_AGRAVO
+DT_NOTIFIC
+NM_AGRAV
+NU_ANO
+ID_MUNICIP
+NU_IDADE_N
+CS_SEXO
+DT_NASC
SOMA(NU_NOTIFIC)
ETL
quantidade
<<Fato>>
Incidencia
+idIncidencia
+idLocal
+idClassificacao
+idOcupacao
+idAgravo
+idTempo
+sexo
+idade
+quantidade


<<DataBase SUS>>
CIDADES_ES
+ID_CIDADE
+NM_CIDADE
+ALT
+LONG
+LAT
+UF
ID_CIDADE
ETL
idLocal
<<Dimensão 4>>
LocalOcorrencia
+idLocal
+descLocal
+altitude
+longitude
+latitude
+uf
<<DataBase SUS>>
CIDADES_ES
+ID_CIDADE
+NM_CIDADE
+ALT
+LONG
+LAT
+UF
NM_CIDADE
ETL
descLocal
<<Dimensão 4>>
LocalOcorrencia
+idLocal
+descLocal
+altitude
+longitude
+latitude
+uf
<<DataBase SUS>>
CIDADES_ES
+ID_CIDADE
+NM_CIDADE
+ALT
+LONG
+LAT
+UF
ALT
ETL
altitude
<<Dimensão 4>>
LocalOcorrencia
+idLocal
+descLocal
+altitude
+longitude
+latitude
+uf
<<DataBase SUS>>
CIDADES_ES
+ID_CIDADE
+NM_CIDADE
+ALT
+LONG
+LAT
+UF
LONG
ETL
longitude
<<Dimensão 4>>
LocalOcorrencia
+idLocal
+descLocal
+altitude
+longitude
+latitude
+uf
<<DataBase SUS>>
CIDADES_ES
+ID_CIDADE
+NM_CIDADE
+ALT
+LONG
+LAT
+UF
LAT
ETL
latitude
<<Dimensão 4>>
LocalOcorrencia
+idLocal
+descLocal
+altitude
+longitude
+latitude
+uf
<<DataBase SUS>>
CIDADES_ES
+ID_CIDADE
+NM_CIDADE
+ALT
+LONG
+LAT
+UF
UF
ETL
uf
<<Dimensão 4>>
LocalOcorrencia
+idLocal
+descLocal
+altitude
+longitude
+latitude
+uf