

Uma Hierarquia de Servidores de Dados Geopidemiológicos usando GML

Halisson Matos de Brito¹, Gilmar Luiz Vassoler², Saulo Bortolon³, Daniel de O. Arantes⁴

^{1,3}Programa de Pós-Graduação em Informática (PPGI)

²Programa de Pós-Graduação em Engenharia Elétrica (PPGEE)

⁴Graduando em Ciência da Computação

Universidade Federal do Espírito Santo (UFES), Brasil

Resumo - Este artigo apresenta um protótipo de sistema hierarquicamente distribuído para apresentação e atualização de dados geopidemiológicos, via Web. O sistema utiliza bases de dados Oracle para armazenar tanto os dados textuais quanto os dados geográficos, estes no formato Oracle Spatial. Possui integração com um protótipo de sistema clínico, construído a partir de sistemas de informação de saúde pública. Uma base nacional única é alimentada a partir de bases estaduais e municipais, e estas são alimentadas por Unidades de Saúde, consultórios ou equivalentes. As interfaces Web são implementadas utilizando-se o GeoTools, um pacote de ferramentas construídas em Java, de domínio público, e apresenta, além do mapa, gráficos estatísticos de diversos tipos. O sistema faz a exportação e importação de dados geopidemiológicos através do padrão OpenGis denominado GML (*Geography Markup Language*).

Palavras-chave: Base de dados geopidemiológicos, GML, Oracle Spatial, Internet.

Abstract - This paper presents a system prototype that shows and update geopidemiological data via Web. The system uses Oracle Databases to store both textual and geographic data, the last in the Oracle Spatial format. The prototype is connected to public health information systems, also customized into the Oracle database. Public health care units provide data to geopidemiological municipal databases. These provide data to state and national databases. The Web interfaces were built using GeoTools, a set of public domain Java tools, to show maps and many kinds of statistical charts. The communication among databases is done using GML-based protocols. GML (*Geography Markup Language*) is the OpenGis Consortium standard to geodata communication.

Key-words: Geopidemiologic database, GML, Oracle Spatial, Internet.

Introdução

O Ministério da Saúde, por meio do DATASUS, tem disponibilizado uma importante base de dados epidemiológicos para a comunidade científica e a população em geral, por meio da Internet [1]. Os dados coletados através do SIAB (Sistema de Informação de Atenção Básica) [2], SINAN (Sistema de Informação de Agravos de Notificação), SINASC (Sistema de Informações de Nascidos Vivos) [3], entre outros, são reunidos e disponibilizados a partir de Brasília. Utilizando o TabWin, é possível fazer a visualização destes dados em interfaces baseadas em mapas. Pesquisadores também têm desenvolvido interfaces baseadas na Web para facilitar a análise de tais dados, como em [4].

Contudo, como aponta [5], algumas características podem ser aprimoradas:

- falta integração entre sistemas, o que gera redundância e inconsistência de dados;

- mecanismo de transferência dos dados para atualização da base central no Datasus ainda é feito por disquete;
- dificuldade de georeferenciamento dos dados devido ao nível de agregação;
- dificuldade de utilização dos dados para vigilância epidemiológica;
- dificuldade de compartilhamento dos dados dos pacientes entre secretarias de saúde municipais/estaduais.

Por outro lado, com o Cartão Único de Saúde, cada pessoa passa a ter uma identificação única nos diversos sistemas de saúde em todo o país, o que reforça a necessidade de compartilhamento dos dados entre as diversas secretarias.

A integração de dados epidemiológicos com sistemas de informação geográfica (SIG) permite construir interfaces de visualização poderosas e efetuar diversos tipos de análise de correlação espacial, como: distribuição espacial de casos de dengue e ocorrência de focos de mosquito;

ocorrência de mortalidade infantil em áreas de pobreza; distribuição de casos de cólera em comparação com a localização de fontes de água para consumo, etc. Segundo [6], o estudo da ocorrência de doenças a partir de sua localização espacial é bastante difundido, especialmente na identificação de possíveis causas, sejam elas ambientais ou não.

Propostas têm sido feitas para integrar GIS ao sistema de saúde pública, inclusive de forma gratuita para as prefeituras, entre elas a do Spring e o SIG-EPI [7]. Algumas prefeituras, no entanto, já fizeram a opção por softwares GIS proprietários, como a prefeitura de Vitória, ES, que utiliza os produtos da ESRI [8]. Outras ainda podem optar por formatos de armazenamento específicos, como o Oracle Spatial [9] e o PostGis/PostGres (de código aberto) [10].

Se não é possível criar um padrão de armazenamento e manipulação de dados, é imperativo padronizar a comunicação de dados georeferenciados. Nesta direção, é obrigatório considerar o Consórcio OpenGis, que recomenda o uso da GML (*Geography Markup Language*) [11]. A GML é uma especificação baseada na XML (*eXtensible Markup Language*) [12], para definir as estruturas dos objetos geográficos a serem intercambiados entre sistemas diferentes.

Este trabalho apresenta um sistema para visualização de dados geoepidemiológicos, integrando mapas e gráficos estatísticos, onde os dados são armazenados no Oracle Spatial e visualizados utilizando o GeoTools [13]. O sistema propõe ainda uma hierarquia de bases de dados distribuídas, que vão desde bases municipais e estaduais, alimentadas por agentes comunitários, postos de saúde e consultórios, até uma base nacional. Esse esquema pode ser visto na Figura 1. O que diferenciaria essas bases seria o nível de detalhamento das informações que elas armazenariam. A base central reuniria dados detalhados por município (médias, índices e totais municipais), enquanto as bases estaduais e municipais, dados que poderiam ir até o nível de residências. Essas bases poderiam ser heterogêneas, ou seja, poderiam ter estruturas e softwares diferentes, desde que a troca de informações entre elas fosse padronizada.

O emprego da GML como padrão para troca de informações inter-bases permitiria que dados geográficos fossem compartilhados também “na horizontal”. Assim, um médico poderia verificar a densidade de focos de mosquitos em torno da residência de um paciente capixaba que estivesse sendo atendido por ele em São Paulo.

O artigo discute ainda, de forma preliminar, uma hierarquia de níveis de acesso aos dados, para garantir a privacidade do cidadão.

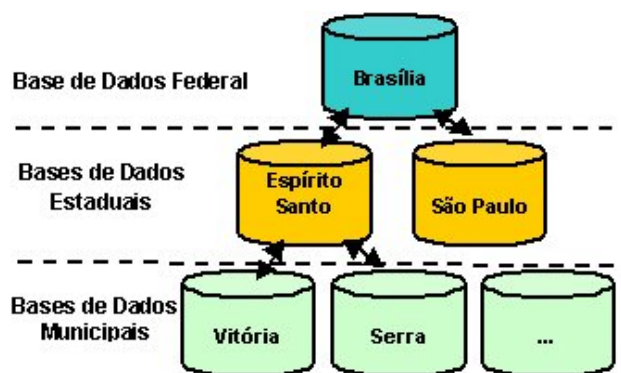


Figura 1 – Hierarquia de bases de dados geográficas.

Como estudo de caso, foi utilizado o cadastro de ocorrências de casos de dengue e focos de mosquito, além da apresentação de mapas interativos com gráficos de estatísticas sobre diversas doenças e dados sócio-econômicos.

Metodologia

Diversos sistemas de informação de atenção básica em saúde foram estudados, entre eles: SIAB, SIP, SINASC, SINAN, Cartão da Criança. Também se incluiu no estudo as especificações do PRC (Padronização de Registros Clínicos) do Datasus.

Estes sistemas de informação de atenção básica foram integrados em uma base de dados, após uma análise para eliminar casos de redundância, inconsistência e dificuldade de fluxo de dados entre os mesmos. A plataforma escolhida para a criação da base foi a do banco Oracle, devido à segurança, confiabilidade, performance, relativa facilidade em se encontrar mão-de-obra qualificada e grande capacidade de armazenamento de dados.

Como estudo de caso para o cadastro de informações e visualização de estatísticas, foi escolhido o acompanhamento de casos de dengue e focos do mosquito transmissor da mesma (*Aedes Aegypti*). A dengue é uma doença cujo controle ainda não foi alcançado em diversos municípios do país. A cada ano pode-se acompanhar novos surtos da doença e, nos últimos meses, pôde-se observar o avanço do tipo III do vírus, aumentando o risco de dengue hemorrágica.

Para o armazenamento/transferência de dados geográficos entre bases, foram analisados alguns formatos de dados: Oracle Spatial, GML e shapefile. Foi escolhido o Oracle Spatial para armazenamento, devido às vantagens do banco Oracle já citadas. Contudo, para a transferência inter-bases foi escolhida a GML por ser uma linguagem de padrão aberto e baseada no modelo de representação geográfica do OpenGis, que vem sendo adotado por várias empresas, entre elas a

Oracle. A GML, por ser baseada na XML, permite extração automática de informações (por ex. objetos geográficos) a partir de documentos texto. Sendo texto, permite utilizar algoritmos de compactação eficientes, reduzindo o problema quanto ao tráfego extra de dados que poderia causar.

Para visualização e atualização de dados geográficos via Internet, avaliaram-se várias ferramentas, entre elas o ArcIMS da ESRI, o GeoTools, JaGo, OpenMap e Jshape [14]. Deu-se preferência a ferramentas que fossem de código aberto e que possuíssem maior poder de visualização e edição de dados geográficos. Assim, foi escolhido o pacote GeoTools, devido ao conjunto de ferramentas que oferece, permitindo ainda sua utilização em conjunto com a API do Oracle Spatial, para apresentação e edição dos dados armazenados neste. Além disso, o GeoTools está em conformidade com o padrão OpenGis.

Resultados

Um protótipo do sistema está sendo desenvolvido para dar suporte às operações de inserção e apresentação de dados geográficos, bem como às operações de apresentação de estatísticas sobre os dados georeferenciados.

A interface para apresentação consiste de um mapa, na parte esquerda da tela e de dois painéis nas partes superior e inferior direitas. A Figura 2 apresenta a tela inicial do sistema. Nela pode-se ver:

- um mapa temático do Brasil com uma graduação de cores para representar a quantidade de casos de dengue em cada estado (cabe aqui observar que os dados apresentados no protótipo são fictícios);
- no painel superior direito tem-se uma legenda representando o intervalo (número de casos) representados por cada cor.
- à direita da legenda, botões permitem escolher qual informação (doença) será apresentada no mapa (dengue, tuberculose, ...)
- ao se passar o ponteiro do mouse sobre algum estado, por ex. o Espírito Santo, um gráfico de pizza aparece no painel inferior direito, mostrando a proporção/quantidade de casos das referidas doenças, no estado. Ao lado deste gráfico, tem-se a legenda para cada cor.

Cabe observar que o gráfico de pizza permite visualizar tanto a proporção de cada doença (através da largura do setor) quanto a quantidade total de casos de doença notificados (pelo raio da própria pizza).

As ferramentas de zoom e pan permitem ao usuário selecionar exatamente a parte do mapa que deseja ver, em um nível de ampliação adequado.



Figura 2 – Casos de dengue por estado, e número de casos de cada doença notificável no ES.

Ao se selecionar um estado, o mapa do mesmo é apresentado, subdividido em municípios, com a graduação de cores representando a quantidade de casos de dengue em cada um deles. A Figura 3 ilustra um outro uso das facilidades do GeoTools. Uma legenda semelhante ao mapa anterior representa a faixa de casos de dengue por município e a quantidade de municípios por faixa. À direita pode-se selecionar a visualização do índice pluviométrico por município. Ao se passar o mouse sobre algum município, o painel inferior direito passa a apresentar um histograma, representando a evolução dos casos de dengue, bem como a quantidade de chuva no último ano.

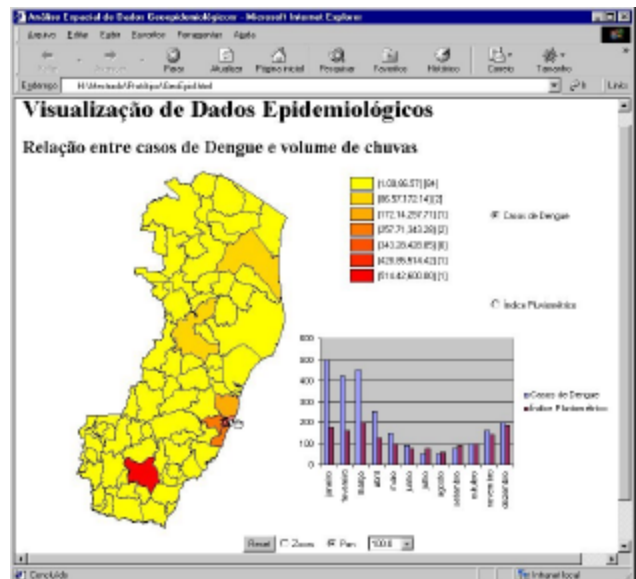


Figura 3 – Casos de dengue por município do ES, e histograma dengue/pluviosidade em Vitória.

Caso semelhante poderia ser implementado para outras divisões geográficas consideradas pelo Datasus, como macro e micro-regiões do estado, regionais de saúde, etc. Ou ainda, para o município, dividido por distritos e/ou cidades/vilas, conseguindo dessa forma outros níveis de informação. O protótipo, porém, não possui ainda estes casos implementados.

Como esses dados poderão estar em servidores web e/ou bases de dados diferentes, a busca aos mapas e aos dados epidemiológicos é direcionada dinamicamente ao servidor correto. Assim, é transparente para um usuário que busque informação o local físico onde ela está armazenada. O que importa é que ele pode visualizá-la a partir de uma mesma interface, o seu *web browser*.

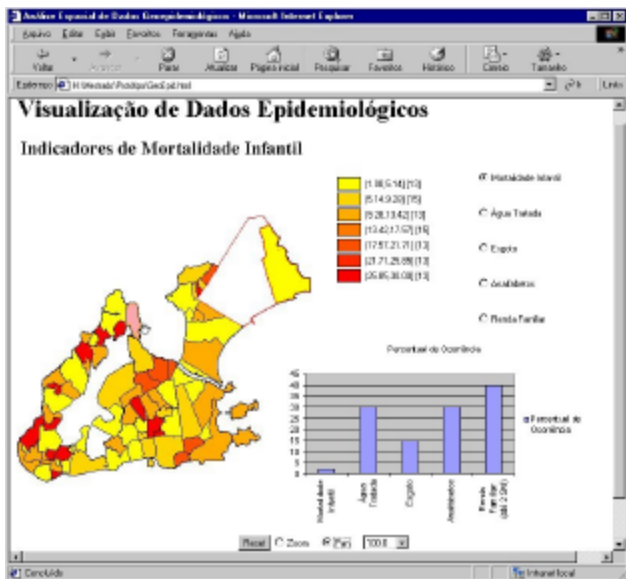


Figura 4 – Casos de mortalidade infantil por bairro e alguns indicadores do bairro Resistência.

Caso o usuário selecione uma cidade sobre o mapa (a sede de um município, por ex.), carrega-se o mapa da cidade com divisão por bairros, com graduação de cores indicando - por exemplo - casos de mortalidade infantil por bairro. A Figura 4 ilustra esta situação, ao se selecionar, no mapa, a cidade de Vitória (dados fictícios). O painel superior é semelhante aos outros, mas o painel inferior ilustra agora alguns indicadores sociais e de saneamento que indicam situações de risco social.

Ao se selecionar um bairro específico, são apresentadas as residências cadastradas no PSF/PACS local, com graduação de cores indicando casos de dengue notificados por residência, como pode ser visto na Figura 5. Os dois painéis são semelhantes aos anteriores. Ao se passar o mouse sobre alguma residência, o painel inferior direito apresenta um gráfico em forma de pizza, mostrando

ocorrência de casos de outras doenças na mesma residência. A apresentação dessas informações ao nível de residências, porém, deve possuir restrições quanto ao acesso, o que será discutido na próxima seção.



Figura 5 – Casos de Dengue por residência e número de casos de doenças em uma residência.

Outro complicador é a existência de edifícios de múltiplos andares. Neste caso, a quantidade de residentes pode ser indicada pelo raio da pizza.

Apesar da seqüência de apresentação do sistema neste artigo partir da visão nacional para a visão ao nível de residência, o fluxo de dados se dá em sentido contrário. A Figura 6 ilustra este fluxo. Os dados clínicos do paciente são capturados pelos agentes do PSF/PACS nas residências ou por médicos em postos de saúde ou consultórios. Esses dados são passados para o sistema clínico, através da sua interface Web. O sistema clínico exporta periodicamente um sumário de dados para o sistema epidemiológico, no servidor de dados municipal. Um sumário com os dados de distrito (ou até mesmo bairro) é então passado para o servidor estadual, juntamente com as respectivas GMLs de ocorrências de doenças. Por sua vez, um sumário dos dados por município é passado para o servidor nacional, juntamente com as respectivas GMLs.

A transferência de dados geográficos entre bases via GML é ilustrada na Figura 7. Os objetos geográficos são extraídos, por exemplo, de uma base municipal (onde estão no formato Oracle Spatial), convertidos em um arquivo GML, enviado até o servidor estadual, convertido em seguida em objetos geográficos e armazenados na base estadual.

Um exemplo de um documento GML, contendo a descrição dos objetos geográficos de áreas de focos de mosquito no município de Vitória, além de outros dados, é ilustrado pela Figura 8.

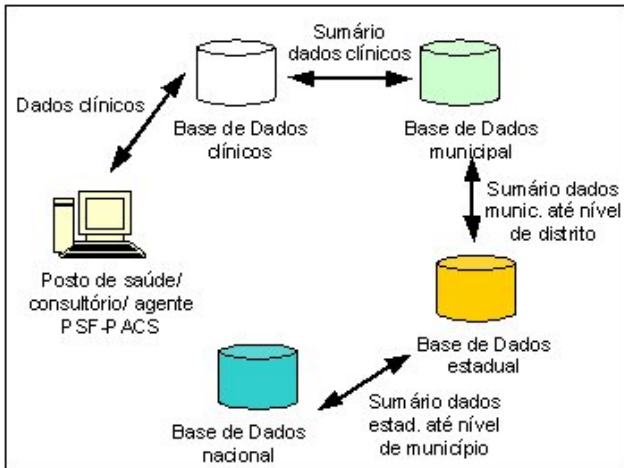


Figura 6 – Fluxo de dados da Unidade de Saúde até o servidor de dados nacional.

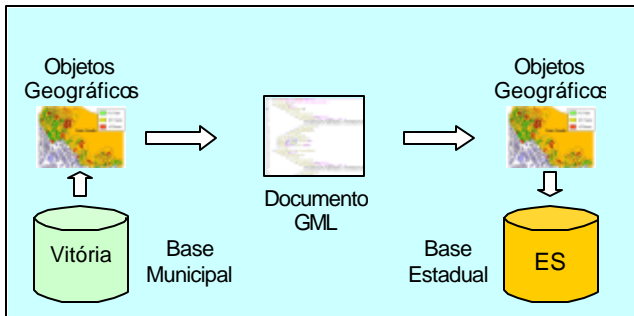


Figura 7 – Transferência de dados geográficos entre uma base municipal e uma estadual usando GML.

```

1 <?xml version="1.0" encoding="UTF-8" ?>
2 <collection xmlns:gml="http://www.opengis.net/gml" ?>
3   <gml:featureMember ?>
4     <feature gml:id="F1" ?>
5       <gml:polygons ?>
6         <gml:Polygon ?>
7           <gml:outerBoundaryIs ?>
8             <gml:LinearRing srsName="srs11" ?>
9               <gml:coordinates cs="s" decimal="." s="ss" ?>
10                6.6031174659729,48.621158539855516 6.636706829871045,48.4623031616231
11                6.571619387497793,49.490453627441485 6.515995893209797,49.49562454022
12                6.627163882591357,49.524078369149625
13              </gml:coordinates>
14            </gml:LinearRing>
15          </gml:outerBoundaryIs>
16        </gml:Polygon>
17      </gml:polygons ?>
18      <gml:CIPIID ?>Pltclclac /</gml:CIPIID ?>
19      <gml:ID ?>300000 /</gml:ID ?>
20      <gml:ID ?>300000 /</gml:ID ?>
21      <gml:ID ?>300000 /</gml:ID ?>
22      <gml:ID ?>300000 /</gml:ID ?>
23      <gml:ID ?>300000 /</gml:ID ?>
24      <gml:ID ?>300000 /</gml:ID ?>
25      <gml:ID ?>300000 /</gml:ID ?>
26      <gml:ID ?>300000 /</gml:ID ?>
27      <gml:ID ?>300000 /</gml:ID ?>
28      <gml:ID ?>300000 /</gml:ID ?>
29      <gml:ID ?>300000 /</gml:ID ?>
30      <gml:ID ?>300000 /</gml:ID ?>
31      <gml:ID ?>300000 /</gml:ID ?>
32      <gml:ID ?>300000 /</gml:ID ?>
33      <gml:ID ?>300000 /</gml:ID ?>
34      <gml:ID ?>300000 /</gml:ID ?>

```

Figura 8 – Exemplo de um documento GML para transferência de dados entre bases.

Discussão e Conclusões

A escolha da plataforma Oracle/Oracle Spatial para armazenamento dos dados é justificada por suas características apresentadas na seção de metodologia, e ainda por possuir bom suporte ao usuário. Além disso, os dados espaciais são armazenados junto aos dados alfanuméricos, o que permite o aproveitamento das características do SGBD, além de toda a parte de gerenciamento de backup ser feita em conjunto. Essa manipulação conjunta dos dados espaciais e não espaciais elimina os problemas de se trabalhar com dados geográficos separadamente em arquivos do sistema operacional ou a necessidade de se ter um SGBD geográfico em separado, como o *Spatial Database Engine (SDE)*, da ESRI.

A escolha de pacotes de domínio público, para o desenvolvimento da interface gráfica, foi feita visando reduzir o custo de implantação do sistema, principalmente para prefeituras de pequeno porte, além da relativa facilidade de se encontrar este tipo de ferramenta na Internet, com as funcionalidades necessárias. Ferramentas como Geotools, JaGo, JShape, OpenMap, etc, oferecem praticamente todas as funcionalidades para manipulação e apresentação de dados geográficos. Algumas delas oferecem ainda suporte a acesso a banco de dados e intercâmbio de dados geográficos usando GML.

A escolha do GeoTools foi motivada pela facilidade de implementação que este pacote oferece, além de permitir que suas funcionalidades sejam facilmente estendidas para atender às necessidades de cada sistema, como a necessidade de extensão feita para acessar a base de dados Oracle Spatial. Além disso, o GeoTools, assim como o Oracle, está em conformidade com o padrão OpenGIS, o que facilita o intercâmbio de informações entre sistemas semelhantes pelo uso da GML. Este intercâmbio é importante em situações, em que, por exemplo, uma pessoa reside numa cidade mas, em viagem, é atendida num Posto de Saúde em outra cidade, de outro estado, com suspeita de malária. O médico pode desejar saber como está a situação da malária na região em que o paciente reside. As informações, geográficas e textuais podem ser recuperadas pelo médico através da Internet, via GML, não importando se as bases de dados da cidade de origem e a de destino estejam no mesmo formato ou não. Isso permite que estados e municípios usem diferentes padrões de armazenamento ou visualização de dados espaciais, mantendo comunicação entre eles via GML.

Apesar das vantagens do Oracle, é sabido que seu custo é muito alto, o que inviabilizaria a implantação do sistema, com esta configuração, em prefeituras de pequeno/médio porte. Uma solução

para este problema seria a utilização de um banco de dados gratuito, como o PostGres, que também possui uma extensão para GIS, chamada PostGis.

Outra solução seria a instalação do Oracle no servidor estadual, que, além de manter a base de dados estadual, manteria também bases municipais. Essa alternativa possui ainda a vantagem de a prefeitura não necessitar contratar mão-de-obra especializada para a manutenção do sistema, o que para muitas prefeituras de pequeno porte é inviável.

Sobre a disponibilização de dados epidemiológicos na Internet, há algumas considerações a serem feitas. Para níveis de apresentação de dados geográficos como os das figuras 2, 3 e 4, não há grandes problemas. Entretanto, para um nível como o da figura 5, a disponibilização não pode ser feita de forma aberta. Por exemplo, dados até o nível de município poderiam ser visualizados por qualquer tipo de usuário. Dados no nível de microárea ou residência só poderiam ser acessados pelo agente de saúde responsável pela microárea e pelo chefe da equipe PSF/PACS. Suponha, por exemplo, que esteja sendo apresentado o número de casos de HIV+ por microárea. Caso esse dado fosse disponibilizado na Internet, sem restrição, poderia haver especulação por parte da vizinhança sobre "quem está com AIDS". Mesmo a disponibilização dos dados de áreas maiores, como bairro ou área do PSF/PACS, também é discutível.

Uma solução seria a adoção de identificação de usuários, com níveis diferenciados de privilégio para a utilização do sistema, após uma discussão profunda a respeito das implicações éticas e do direito à privacidade do cidadão.

Referências

- [1] Datasus. "Informações de Saúde". <http://tabnet.datasus.gov.br/tabnet/tabnet.htm>.
- [2] Datasus. "Sistema de Informação de Atenção Básica". <http://www.datasus.gov.br/catalogo/pacs.htm>.
- [3] Funasa. "Sistemas de Informação em Saúde". <http://www.funasa.gov.br/sis/sis00.htm>.
- [4] Escola Paulista de Medicina. "On-line Searchable Database of Mortality Data from Death Certificate". <http://gamba.epm.br/pub/datasus/mortality.html>.
- [5] Fiocruz. "Proposta de implantação de um Sistema de Informações Geográficas para apoio ao PACS/PSF". <http://www.cpqam.fiocruz.br/nicc/ofrecife.htm>. Julho 1999.
- [6] Carvalho, Marília S. (1997). *Aplicação de métodos de análise espacial na caracterização de áreas de risco à saúde*. Tese de Doutorado, Programa de Engenharia Biomédica, COOPE/UFRJ, Rio de Janeiro, 179p.
- [7] Datasus. "CTI-GEO". <http://www.datasus.gov.br/catalogo/ctigeo/relatorio2.htm>.
- [8] ESRI. "Products". <http://www.esri.com/software/index.html>.
- [9] ORACLE. "Oracle Spatial". <http://otn.oracle.com/products/spatial/content.html>.
- [10] Refrations Research Inc. "PostGIS, Geographic Objects for PostgreSQL". <http://postgis.refrations.net>.
- [11] Open Gis Consortium. "Geography Markup Language (GML) 2.0". <http://www.opengis.net/gml/01-029/GML2.html>. Fevereiro 2001.
- [12] W3C. "Extensible Markup Language (XML) 1.0". <http://www.w3.org/TR/REC-xml>. Outubro 2000.
- [13] SourceForge. "GeoTools Open Source Mapping Toolkit". <http://geotools.sourceforge.net/>.
- [14] FREEGIS. "The FreeGIS Project". <http://freegis.org/index.en.html>.

Contato

Halisson Matos de Brito, Mestrando do Programa de Pós-Graduação em Informática - UFES. Prof. Substituto do DI – UFES. Tel: 0-xx-27-3335-2134. E-mail: hmbrito@inf.ufes.br.

Gilmar Luiz Vassoler, Mestrando do Programa de Pós-Graduação em Engenharia Elétrica (PPGEE) - UFES. Prof. Adj. Depto. Ciência da Computação da Faculdade de Ciências Humanas de Aracruz – ES. E-mails: gilmar@inf.ufes.br, gilmar@fsjb.edu.br.

Saulo Bortolon, Dsc. Eng. Elétrica, UNICAMP, Prof. Programa de Pós-Graduação em Informática da UFES. E-mail: bortolon@inf.ufes.br.

Daniel de Oliveira Arantes, graduando em Ciência da Computação - UFES. E-mail: doa_daniel@hotmail.com.