



Revista Brasileira de Cartografia (2012) N<sup>o</sup> 64/1: 113-133  
Sociedade Brasileira de Cartografia, Geodésia, Fotogrametria e Sensoriamento Remoto  
ISSN: 1808-0936

## FRAMEWORK COLABORATIVO PARA MINERAÇÃO DE DADOS ESPACIAIS: ENGENHARIA APLICADA AO CTM

*Collaborative Framework for Spatial Data Mining:  
Engineering Applied to Multipurpose Cadastre*

**André Fabiano de Moraes & Lia Caetano Bastos**

**Universidade Federal de Santa Catarina – UFSC**  
**Centro Tecnológico - Departamento de Engenharia Civil**  
Caixa Postal 476 - CEP 88040-900 – Florianópolis SC Brasil  
ecv3afm@ecv.ufsc.br, ecv1lcb@ecv.ufsc.br

*Recebido em 21 Abril, 2011/ Aceito em 05 Julho, 2011*  
*Received on April 21, 2011/ Accepted on July 05, 2011*

### RESUMO

Com a disseminação de diferentes estruturas de dados CAD/GIS produzidas através do rápido avanço das tecnologias de informação, são reais os desafios na implementação de uma infraestrutura interoperável de dados espaciais, sendo foco de várias discussões. Entretanto com o surgimento de tantas variações, a interoperabilidade entre os dados ainda é reduzida, tornando inevitável a produção de volumes expressivos de dados, que acabam enfrentando cedo ou tarde conflitos computacionais. A partir disso, a pesquisa apresenta uma investigação sobre os aspectos relevantes que influenciam na engenharia de projetos, originando o desenvolvimento do ambiente *Open Integration for Collaborative Geospatial Framework Web* (OpenICGFw), visando o reconhecimento de estruturas, integração, manipulação e colaboração, em sintonia com esforços da INDE, OGC e W3C. Durante a discussão, são apresentados os resultados através de aplicações webmapping visando a integração do cadastro técnico multifinalitário a outras bases distribuídas para uso das principais técnicas de mineração de dados espaciais. Ao final, o artigo discute a hipótese e a contribuição da pesquisa, questões prioritárias na implementação da mineração de dados espaciais, além do avanço tecnológico na engenharia territorial para colaboração e gestão do conhecimento.

**Palavras chaves:** Infraestrutura, Framework, Colaboração, Interoperabilidade, Mineração de Dados Espaciais.

### ABSTRACT

With the spread of different data structures, CAD/GIS produced by the rapid advancement of information technologies, are real challenges in the implementation of an interoperable spatial data infrastructure, the focus of many discussions. However with the emergence of many variations, the interoperability between the data is still low, making the inevitable production of large volumes of data, which sooner or later end up facing computational conflicts. From this, the study presents an investigation on the relevant aspects that influence the projects engineering, resulting in the development environment *Open Integration for Collaborative Geospatial Framework Web* (OpenICGFw), aimed at recognizing structures, integration, manipulation and collaboration, in line with INDE efforts, OGC and W3C. During the discussion, the results are presented through applications webmapping to integrate multipurpose cadastre the other distributed databases for use of the main techniques of spatial data mining. In the end, the article discusses the hypothesis and the contribution of research, priority issues in the implementation of spatial data mining, as well as technological advances in the

territorial engineering for collaboration and knowledge management.

**Keywords:** Infrastructure, Framework, Collaboration, Interoperability, Spatial Data Mining.

## 1. INSTRUÇÃO

Com a expansão do uso de tecnologias para lidar com os processos da engenharia, principalmente tratando-se do geoprocessamento como ferramenta para tomada de decisão, é observada uma forte adesão das instituições públicas nos processos de análise espacial. Porém, a ausência de ferramentas que facilitem a integração e a interoperabilidade entre diferentes fontes de dados geográficos tem dificultado esse processo. Aliado a isso é detectada a descontinuidade de projetos que demandam de expressivo conhecimento técnico, ocasionando na redução de recursos que, normalmente, afetam instituições públicas e privadas (STEUDLER, 2010). Também é observado o acesso aos dados geográficos via internet e todos os problemas referentes ao seu intercâmbio, devido à natureza particular de cada dado e sua definição. Para tal, observam-se alguns desafios citados por Câmara et al. (2000), como a complexidade e a inexistência de padrões de dados, a exemplo da interoperabilidade entre dados espaciais, e principalmente pela falta de modelos conceituais comuns, acarretando problemas na troca de dados entre sistemas de informações geográficas (SIG) distintos. Neste foco, detecta-se experiências para minimizar ou superar a problemática (MACÁRIO et al., 2009) e (KARABEGOVIĆ et al., 2010).

Em sistemas heterogêneos, a conversão de dados representa um custo entre 60% e 80% do custo total da implantação. Com relação ao Brasil, isso é agravado pela falta de padrões nacionalmente estabelecidos. Por outro lado, o surgimento da Infraestrutura Nacional de Dados Espaciais (INDE, 2009), normatiza as primeiras atividades concretas do plano de ação para promover a implantação de infraestrutura de dados espaciais no país, mas também sofre pela indisponibilidade de ferramentas interoperáveis de baixo custo.

Em vista de tal problemática, é apresentado neste artigo um método envolvendo aspectos ágeis e flexíveis para auxiliar nas atividades de elaboração, desenvolvimento, construção, implementação e também acompanhamento de geo-soluções através da internet.

Na definição do método são utilizados os conceitos do MCDA-C conforme (CARVALHO

et. al., 2007), possibilitando avaliar os recursos disponíveis nos principais softwares livres adotados em sistemas de informações geográficas, bem como a implementação de novas características. Para isso, um modelo flexível de desenvolvimento foi elaborado visando abordar questões relacionadas às constantes coletas para requisitos variáveis (JOHNSON et. al., 2008) e (SANTOS, 2009), além das especificações para padrões de dados e técnicas comumente utilizadas.

Este artigo apresenta duas contribuições visando superar essas deficiências. Na primeira, são investigados os aspectos fundamentais para uma infraestrutura de dados espaciais interoperáveis (OOSTEROM et. al., 2008), nos quais também são avaliados os requisitos necessários para elaboração do framework OpenICGFw através do MCDA-C (Seção 2 e 3). A segunda contribuição apresenta o estudo de caso utilizando o framework OpenICGFw para detectar padrões territoriais e urbanos (NEUNZERT, 2010), visando comprovar a hipótese verdadeira ao detectar conhecimento oculto e desconhecido juntamente com aplicabilidade do framework (Seção 4).

### 1.1 Estruturando Problemas para Avaliação Multicritério (MCDA-C)

A estruturação do processo de apoio à decisão visa construir uma estrutura aceita pelos atores modelando uma realidade preexistente. É a fase mais importante do MCDA-C, visto que proporciona aprendizado, clareza e representatividade através da definição e construção de um modelo que servirá de base comum, no qual os valores dos atores intervenientes possam ser validados. O propósito da estruturação é desenvolver um conjunto de instrumentos que permitam aos decisores entender bem o problema.

Inicialmente, para Bana e Costa (1993, p. 31), o processo de apoio à decisão é um sistema aberto, tendo componentes específicos a exemplo de atores, valores, objetivos, ações e características. A atividade de apoio à decisão pode, então, ser vista como um processo de interação com uma problemática mal-estruturada, em que os elementos e suas relações emergem de forma mais ou menos caótica.

Um problema é definido como uma situação prioritária na qual o decisor deseja mudar seu estado atual, mas não está seguro de como realizá-lo. Sendo assim, o conceito do MCDA estabelece quatro dimensões necessárias para a existência de um problema: a existência de uma insatisfação, um proponente, a importância para merecer o esforço de resolução e, por último, que haja possível solução (ENSSLIN, 2001). A abordagem MCDA, por sua vez, pode ser enquadrada como um processo para avaliar as abordagens de avaliação de desempenho, bem como um processo para desenvolver um modelo sistemático de avaliação de desempenho. No paradigma construtivista, o objetivo dos modelos é a geração de conhecimentos aos decisores. Esta fase proporcionará condições para que os decisores possam visualizar as consequências das alternativas com relação aos seus aspectos, sendo julgados como importantes no contexto.

A fim de delimitar as dimensões a serem trabalhadas na construção de um modelo estratégico, é apresentado inicialmente cada ponto de vista elementar (PVE), que representa um conjunto de especificações mínimas para a estruturação e funcionamento do método. A partir dos resultados obtidos com a estruturação do método, em sintonia com as propostas do Grupo Temático de Georreferenciamento – Governo Eletrônico (eGOV, 2010) – do INDE, e também com as especificações de padronização de dados espaciais propostas pelo (OGC, 2009), foi possível fundamentar a elaboração do Ponto de Vista Fundamental (PVF) e do Ponto de Vista Elementar (PVE), conforme detalhamento dos descritores, o que permitiu entender os requisitos e priorizar o desenvolvimento de um ambiente estratégico para integração e colaboração.

## **1.2 Definindo Padrões para o Método de Desenvolvimento**

Após a estruturação dos problemas, são aliados ao método de desenvolvimento critérios específicos para direcionar a sua implementação. Muitos desses critérios são oriundos de outras disciplinas, a exemplo da orientação a objetos e da engenharia web. Outros critérios são adaptados para o processo de desenvolvimento de softwares de forma ágil e flexível.

A partir disso, são iniciados os trabalhos de construção do software livre, denominado framework OpenICGFw. O seu desenvolvimento

foi realizado através do padrão Model View Controller (MVC) para funcionamento no ambiente web conforme (LISBOA, 2008) e (SOARES, 2009). Possibilitando a inclusão de novos conceitos para a produção de sistemas webmapping, dentre as quais o Model, View, Controller and Services (MVCS). Com essas especificações para o desenvolvimento, é possível produzir sistemas flexíveis que facilitem posteriormente a sua manutenção, visto que cada padrão adotado pode ser individualmente trabalhado sem interferir na comunicação global do sistema. Um framework agrega funcionalidades comuns em várias aplicações e as disponibiliza em estruturas que busquem ser facilmente manuseadas e compreendidas. Para isso, também foram realizados levantamentos sobre os frameworks de desenvolvimento comumente utilizados pela comunidade de software livre, visando proporcionar uma investigação dos pontos já consolidados, auxiliando na análise e na formulação de aspectos essenciais para a modelagem de um novo ambiente. Além disso, tais aspectos devem ser cuidadosamente investigados e pré-selecionados, visando ao reaproveitamento de funcionalidades existentes e disponibilizadas livremente para sua utilização, alteração e adaptação, variando de acordo com as necessidades individuais de cada projeto.

Para melhor compreensão, são apresentados os principais frameworks web utilizados no desenvolvimento de aplicações livres, principalmente na linguagem PHP. Dentre eles são citados: CakePHP, CodeIgniter, Drupal, Joomla, Miolo, Symfony, Zend Framework e Zoop Framework, além de frameworks específicos para o tratamento e disponibilização de dados espaciais, desenvolvidos em diferentes linguagens, sendo: Geodjango, Mapstraction, OpenGTS e OpenStreetMap. Com a identificação dos principais requisitos, é possível impulsionar a elaboração e a construção de um ambiente único, para e prover a interoperabilidade entre os dados distribuídos em diferentes órgãos públicos.

## **1.3 Provendo a Interoperabilidade para Dados Espaciais**

A interoperabilidade pode ser definida pelo intercâmbio e a manipulação de conjuntos de dados espaciais através de processos computacionais integrados, de tal forma que o resultado seja coerente

e novos valores possam ser agregados aos repositórios. A fim de resgatar as informações geradas pela cartografia e também por outras áreas que demandam maior integração, é necessário compreender claramente todas as fases de elaboração e construção das representações espaciais, tanto quantitativamente quanto com relação à qualidade descritiva dos dados. Certamente, o ponto chave neste processo são os metadados, que norteiam especificações para o desenvolvimento e rastreamento computacional.

Para dispor de dados interoperáveis é importante ter uma infraestrutura comum que descreva os dados para seu compartilhamento, em especial para o Brasil, possivelmente podendo ser adaptado para outros idiomas, ou seja, uma metodologia comum para auxiliar na definição de metadados, que será, por um lado, compatível com a regulamentação e sistematização de requisitos diretos e indiretos; por outro lado, que vise atender às necessidades da comunidade científica em função da diversidade de padrões e tipos de dados. Essa última questão é de particular interesse quando se trata de processos colaborativos. Para compreender as questões de interoperabilidade é necessário, contudo, investigar e observar as constantes mudanças tecnológicas, a fim de aperfeiçoar e fortalecer os processos existentes junto aos diferentes sistemas SIG amplamente distribuídos. A criação de metadados, no entanto, é essencial por três razões

principais (a,b,c), que estão ligadas a três grandes benefícios (d,e,f).

As razões para os metadados devem (a) refletir a necessidade de investimento para a organização manter o futuro da manutenção de suas estruturas de dados. Também com os metadados surge a (b) necessidade de acesso aos catálogos de dados armazenados, bem como a devida orientação para (c) auxiliar na transferência desses dados.

Por outro lado, os benefícios refletem o cuidado (d) com o conteúdo da descrição dos metadados e a precisão dos dados espaciais. Esses fatores incentivam a reutilização desses dados. Outro benefício visa (e) disponibilizar várias temáticas que nem sempre estão disponíveis devido ao seu intercâmbio. Por último, com os metadados é possível (f) identificar e acompanhar a transferência de grandes conjuntos de dados, facilitando sua estrutura para os requisitos organizacionais. Para isso, uma infraestrutura tecnológica é fundamental para estabelecer critérios técnicos que sejam viáveis de implementação, conforme ilustrado na Figura 1.

Ainda com relação à elaboração de uma infraestrutura de dados espacial padronizada, a INDE, a INSPIRE e a OGC, dentre várias outras organizações, buscam alavancar esforços para integrar ações na composição única entre diferentes tipologias e estruturas de dados espaciais. Embora este seja um tema de extrema relevância para a



Fig. 1 - Definição em camadas da plataforma tecnológica para estruturação do framework.

interoperabilidade de dados espaciais, tal aprofundamento não é objeto de investigação deste artigo, devido à sua prematuridade. A proposta de um modelo livre de dados (POSENATO, 2008), porém, deve ser detalhada através de conceitos, implementações e preocupações com relação à infraestrutura de dados espaciais, especialmente na gestão do sistema cadastral e da cartografia.

## **2. ARQUITETURA DO FRAMEWORK**

De acordo com Goodchild et al. (1999), durante muito tempo a descentralização e administração dos sistemas de informações geográficas impulsionaram a comunidade desenvolvedora desses sistemas para o fortalecimento da plataforma dos desktops (SHERMAN, 2008). Considerando a evolução desses sistemas, contudo, surgem as mudanças para os dias atuais, como a flexibilização dessas plataformas para as arquiteturas de serviços distribuídos pela internet. Diferentemente de outros tipos de serviços de informação, requer ambientes específicos de softwares que possibilitem aproximações entre as instituições através de mecanismos dinâmicos de comunicação.

Classificar as relações operacionais entre cliente, servidor, objetos geográficos e operações entre SIGs requer um detalhamento através dos metadados, possibilitando a comunicação entre diferentes serviços distribuídos (ISO/TS, 2008). Adotar uma estratégia de trabalho através de um ambiente único, proporcionando a colaboração, visa definir um conjunto de competências requeridas para a resolução de problemas comuns. Para isso, a implementação de ontologias (FONSECA et. al., 1999) e (MUN et. al., 2010) se faz uma alternativa interessante para leitura e identificação de estruturas conceituais, em especial para os padrões espaciais (PEACHAVANISHA et. al., 2006) e (GEHLERT et. al., 2007), elaboradas para auxiliar na organização e na integração entre sistemas distintos.

Com essa perspectiva, decorre a formulação de uma estratégia tecnológica, cujo ponto central é a definição de uma trajetória de aprendizado que visa desenvolver um conjunto de competências organizacionais em níveis. Observa-se que a estratégia responde a questões de longo prazo e que a tecnologia é tratada de forma ampla,

envolvendo aspectos relacionados aos produtos e aos processos.

O método foi desenvolvido em nove etapas, dentre as quais são contextualizadas neste artigo com maior detalhamento apenas as duas primeiras, visando auxiliar no planejamento e no desenvolvimento de novos frameworks. Sendo o método dividido para:

- 1º: Identificar o perfil de impacto das ferramentas.
- 2º: Identificar o perfil de impacto generalizado.
- 3º: Identificar os requerimentos do projeto.
- 4º: Codificar as especificações do INDE, OGC e W3C.
- 5º: Desenvolver o módulo de estatística.
- 6º: Desenvolver o módulo de mineração de dados.
- 7º: Desenvolver o módulo de parametrização.
- 8º: Desenvolver o módulo de visualização.
- 9º: Desenvolver o módulo de monitoramento.

### **2.1 Elaboração das Fases de Desenvolvimento**

Dentre as etapas, o método é dividido em três grandes fases. Na primeira, são abordadas as etapas fundamentais (1ª, 2ª, 3ª e 4ª), determinadas para o processo de elaboração. Já para a segunda fase, são tratadas as demais etapas, exploradas na Seção 3 deste artigo. A terceira e última fase não é tratada neste artigo, pois aborda aspectos operacionais de monitoramento e avaliação de desempenho do framework. Na primeira etapa, após a construção dos descritores com os respectivos valores, é realizado um comparativo que visa avaliar as principais características e funcionalidades de recursos e softwares existentes, a fim de identificar o perfil de cada objeto de investigação, proporcionando a geração do perfil de impacto, conforme discutido na Seção 2.2.

Para exemplificar, é realizada a avaliação de 5 softwares: SpringWeb (<http://www.inpe.br>), AlovMap (<http://alov.org>), CartoWeb (<http://cartoweb.org>), I3GEO (<http://mapas.mma.gov.br>) e Mapstraction (<http://mapstraction.com>), obtendo-se as escolhas através de pesquisas junto aos usuários.

### **2.2 Descritores –Primeira Etapa**

A construção dos descritores especialmente para auxiliar na identificação das necessidades e do perfil individualizado de tecnologias existentes (softwares) foi realizada em conjunto com os profissionais das prefeituras, engenheiros,

responsáveis pelo cadastro técnico e também com gestores denominados decisores. Visando atender exatamente cada ponto elementar, possibilitando a formulação da árvore com todos os pontos de vistas fundamentais. Com isso, foi possível selecionar 18 descritores, obtendo-se: D1 - Compatibilidade com Hardware; D2 - Compatibilidade com a Rede Internet, Intranet e Extranet; D3 – Portabilidade entre Sistemas Operacionais; D4 – Linguagem de Programação Adotada; D5 - Sistema Gerenciador de Banco de Dados; D6 – Interface de Navegação; D7 – Implementação do Padrão OGC; D8 – Implementação Padrão ISO; D9 – Incorporação de bibliotecas externas ao framework; D10 – Intercâmbio de dados geoespaciais; D11 – Catálogo de Estruturas Metainformação; D12 – Processamento de Ontologias; D13 – Integração e o relacionamento entre dados; D14 – Elaboração e o processamento de consultas; D15 – Construção de mecanismos para análises estatísticas; D16 – Implementação da Mineração de Dados; D17 – Colaboração; e D18 – Implementação de rotinas Organizacionais.

Ainda nesta etapa, para verificar o comportamento dos softwares e de futuras alterações nas taxas de substituição de um critério em específico, foi realizada análise de sensibilidade global, além das ações sensíveis a essa variação, aplicou-se análise de sensibilidade. Com isso, foi possível estabelecer um marco para diagnosticar o reflexo de mudanças evolutivas e tecnológicas sobre cada ponto de vista elementar para um framework integrador que atenda a especificações comuns para diferentes necessidades.

Ao final da aplicação da primeira etapa e determinando o perfil de impacto dos softwares que disponibilizam dados pela internet, bem como as suas principais características, obtém-se através da presente etapa um valioso referencial para identificar pontos elementares que apresentem um nível de comprometimento, estabilidade ou de excelência. Tais pontos, por sua vez, proporcionam fortes indicadores para auxiliar no planejamento e na construção de recursos essenciais e inovadores em um novo framework. É realizada, assim, a distribuição dos respectivos valores para obtenção da pontuação de cada perfil através da seguinte fórmula:

$f_x$  = Frequência  $X_i = 1, 2, \dots, n$

$f_y$  = Frequência  $Y_i = 1, 2, \dots, n$

$f_z$  = Frequência  $Z_i = 1, 2, \dots, n$

Fator de Equivalência (a) = 5 “Excelência”

Fator de Equivalência (b) = 3 “Estável”

Fator de Equivalência (c) = 2 “Comprometedor”

$$\sum_{i=1}^n = \frac{(f_{x_i} \cdot a_n) + (f_{y_i} \cdot b_n) + (f_{z_i} \cdot c_n)}{10} \quad (1)$$

Após avaliação individual do i3Geo, entretanto, foi possível detectar quais pontos mais se destacam perante as demais funcionalidades apresentadas pelo software: o descritor D17 - Colaboração como sua característica mais positiva, e, por outro lado, o D11 - Metadados como a característica mais comprometedora perante os demais.

Também para a ferramenta CartoWeb, foi possível realizar o mesmo processo, apresentado o descritor D10 - Intercâmbio de Dados como sua característica mais positiva, e o descritor D9 - Incorporar bibliotecas externas ao framework como sua característica mais comprometedora.

Para a ferramenta AlovMap detectou-se o descritor D2 - Compatibilidade com a Rede Internet, Intranet e Extranet como sua característica mais positiva, e o descritor D9 - Incorporar bibliotecas externas ao framework como sua característica mais comprometedora. Ainda para a ferramenta MapStraction, detectou-se o descritor D2 - Compatibilidade com a Rede Internet, Intranet e Extranet como sua característica mais positiva, e o descritor D10 - Intercâmbio de Dados como sua característica mais comprometedora.

Para a ferramenta SpringWeb, por sua vez, detectou-se o descritor D2 - Compatibilidade com a Rede Internet, Intranet e Extranet como sua característica mais positiva, e o descritor D17 - Colaboração como a mais comprometedora. Também foram realizadas análises de sensibilidade e robustez dos descritores em comparação a todos os experimentos, ou seja, sobre cada software avaliado.

Para isso, são apresentados os gráficos que possibilitam a visualização de conflitos entre os descritores ou sua previsão para eventos futuros. A fim de validar o modelo e obter a confiança desejada sobre a construção do perfil de impacto individualizado, também foram avaliados os descritores através da análise de robustez e sensibilidade global, individualizando cada ponto elementar construído. A equação para análise de

sensibilidade trata da mudança das taxas de substituição em função de alteração da taxa de substituição de um critério. O somatório das taxas de substituição adotado pelo modelo multicritério discutido neste trabalho é igual a 1. Também todas as taxas de substituição devem ter um valor entre:

$$0 \leq w_i \leq 1$$

Para avaliação global são obtidos os resultados, após a investigação dos elementos  $VG_1(i3Geo)$ ,  $VG_2(CartoWeb)$ ,  $VG_3(AlovMap)$ ,  $VG_4(MapStraction)$  e  $VG_5(SpringWeb)$ :

$$\{VG_{(a)} = w_1 \cdot v_{1(a)} + w_2 \cdot v_{2(a)} + w_3 \cdot v_{3(a)} + \dots + w_n \cdot v_{n(a)}\}, \text{ ou seja:}$$

$VG_{(a)}$  - valor global da ação

$v_{i(a)}$  - valor parcial da ação dos critérios

$w_i$  - taxas de substituição dos critérios

$n$  - número de critérios do modelo

$$VG(a) = \sum_{i=1}^n w_i \cdot v_i(a) \quad (2)$$

Para cada elemento foram obtidos os respectivos valores na análise numérica, na qual foram escolhidas as taxas de cada critério modificando seu valor original:

$$\begin{aligned} VG_{(i3Geo)} &= 483,82; & VG_{(CartoWeb)} &= 318,69; \\ VG_{(AlovMap)} &= 224,15; & VG_{(MapStraction)} &= 142,11; \\ VG_{(SpringWeb)} &= 128,18; \end{aligned}$$

Esta alteração, entretanto, também afeta as demais taxas de substituição do modelo, haja vista a soma das taxas ser igual a um (1). Por consequência, as demais taxas do modelo foram recalculadas de tal forma que as proporções entre elas não se modifiquem. Para calcular as novas taxas de substituição do modelo, em função da modificação de uma delas, foi utilizada a seguinte equação:

$w_n'$  = taxa de substituição recalculada do critério  $n$ .  
 $w_n$  = taxa de substituição original do critério  $n$ .  
 $w_i'$  = taxa de substituição modificada do critério  $i$ .  
 $w_i$  = taxa de substituição original do critério  $i$ .

$$w_n' = \frac{w_n \cdot (1 - w_i')}{(1 - w_i)} \quad (3)$$

$$w_n' = (0,20 \cdot (1 - 0,65)) / (1 - 0,50) = 0,14$$

$$w_n' = (0,15 \cdot (1 - 0,65)) / (1 - 0,50) = 0,11$$

$$w_n' = (0,20 \cdot (1 - 0,35)) / (1 - 0,50) = 0,26$$

$$w_n' = (0,15 \cdot (1 - 0,35)) / (1 - 0,50) = 0,2$$

Para verificar a sensibilidade dos pontos de vistas fundamentais (PVF) do modelo, foram alteradas as Taxas de Substituição  $v_n$  (a) com aumento de 30% e, por outro lado, a redução do mesmo percentual. Contudo, a primeira alteração para 65% gerou:

$$VG_{(i3Geo)} = 620,68; \quad VG_{(CartoWeb)} = 408,44;$$

$$VG_{(AlovMap)} = 322,52; \quad VG_{(MapStraction)} = 228,07;$$

$$VG_{(SpringWeb)} = 180,90;$$

A segunda alteração das Taxas de Substituição  $v_n$  (a) para 35% gerou os seguintes valores:

$$VG_{(i3Geo)} = 350,20; \quad VG_{(CartoWeb)} = 228,05;$$

$$VG_{(AlovMap)} = 129,83; \quad VG_{(MapStraction)} = 62,53;$$

$$VG_{(SpringWeb)} = 73,93;$$

A sensibilidade do modelo fica evidenciada com a segunda alteração (35%), em que ocorre a troca de posições dos critérios devido à alteração nos valores globais para cada análise para avaliação global. Ainda nesta etapa, para verificar a sensibilidade de cada ponto elementar (PVE), são realizadas as alterações nas taxas individualizadas, além das ações sensíveis a essa variação, sendo aplicadas análises isoladas.

Para exemplificar a distribuição dos valores para D1 - Hardware, tem-se um grau de atratividade de 19.97%, em que são observados os comportamentos de cada ponto de vista elementar para o software em relação aos demais, possibilitando identificar a semelhança e também o choque de comportamento com relação ao critério avaliado. Destaca-se, após avaliação da sensibilidade dos descritores, o gráfico D9 - Biblioteca, que identifica o software Mapstraction com pequena diferença de atratividade comparado ao software SpringWeb, permitindo definir qual funcionalidade deve ser reutilizada.

Com relação ao D14 – Consultas, é detectada a proximidade entre funcionalidades existentes no software MapStraction em relação ao SpringWeb, tornando-os muito próximos até o grau de

atratividade 2,00. Com o aumento dessa pontuação de atratividade, o software SpringWeb apresenta fortes diferenças na reutilização de suas funcionalidades. Com isso, foi possível estabelecer um marco flexível para diagnosticar o reflexo de mudanças evolutivas e tecnológicas sobre cada ponto de vista elementar para reutilização de funcionalidades.

### 2.3 Avaliação do Perfil de Impacto – Segunda Etapa

Na segunda etapa, foi apontado um perfil de impacto generalizado, ou seja, o perfil que apresenta um equilíbrio entre todos os softwares submetidos ao processo de avaliação. Obteve-se também a união de elementos essenciais, como os indicadores, para estipular os principais objetivos e suas prioridades na formulação e no desenvolvimento das atividades de construção de um novo framework.

É realizada, contudo, a comparação entre o perfil médio de impacto (m) de todas as avaliações realizadas na primeira etapa, bem como a representação do melhor perfil com a maior pontuação obtida e também a representação do “*status quo*” a ser alcançado. A partir desses descritores, foram realizadas análises individuais da arquitetura e dos recursos disponibilizados em cada software. A partir do cruzamento de cada perfil, foram detectados quais pontos elementares deveriam ser priorizados na execução das atividades de construção do framework.

Conforme os valores extraídos no cruzamento de cada perfil, foi aplicada a mesma fórmula de cálculo utilizada para obter o ranking dos softwares analisados na primeira etapa, sendo possível visualizar os novos valores e quantificar os ganhos sob cada melhoria implementada na construção do framework, conforme ilustrado na Tabela 1.

Com base nos valores de cada nível (EX, ES e CO) são realizados os cálculos para obtenção da pontuação de cada perfil. Aplicando-se a mesma fórmula, é realizada a distribuição dos valores para obter a classificação final.

### 2.4 Trabalhando com a Dinâmica dos Requisitos – Terceira Etapa

Após a identificação dos recursos mínimos para funcionamento do framework, nesta etapa é constituído o módulo de comunicação entre os usuários e desenvolvedores. É possível, a partir daí, promover a colaboração através de aplicações

TABELA 1- AVALIAÇÃO PARA RANKING FINAL.

| Perfil           | EX | ES | CO | Total | PO ** |
|------------------|----|----|----|-------|-------|
| Perfil Médio (m) | 4  | 7  | 7  | 18    | 5,5   |
| Melhor Perfil    | 9  | 3  | 6  | 18    | 6,6   |
| Perfil Desejado  | 17 | 1  | 0  | 18    | 8,8   |

EX (Excelência), ES (Estável), CO (Comprometedor) e PO\*\* (pontuação obtida após aplicação da fórmula para determinar o ranking final).

específicas que facilitam a comunicação entre todos os desenvolvedores e usuários no gerenciamento das atividades. Sendo assim, com o módulo desenvolvido, foram criados controles específicos para tratar aspectos relacionados aos procedimentos de elaboração e desenvolvimento flexível de sistemas com requisitos variáveis nos mais diferentes níveis, tendo o cadastro individual de cada requisito também denominado de tarefa, devidamente armazenado em um repositório.

Nesta etapa, alguns procedimentos são tratados internamente junto ao módulo para hierarquizar sua prioridade, com o propósito de delegar um desenvolvedor responsável para sua resolução. A partir disso são detectados padrões e necessidades comuns através de tarefas previamente inseridas por usuários, o que permite promover a colaboração organizacional e sua utilização como um mecanismo para auxiliar na coleta constante de requisitos para a melhoria permanente do framework. A principal discussão sobre a problemática do gerenciamento dos requisitos para um sistema de informação elaborado coletivamente não é restrita a uma dinâmica que aconteça uma única vez. Para a melhoria contínua dos softwares nessas áreas, detecta-se com frequência a ausência de mecanismos flexíveis e que facilitem o compartilhamento de necessidades indispensáveis para o trabalho colaborativo de um grupo.

### 2.5 Elaboração do Núcleo (Core) do Framework – Quarta Etapa

Nesta etapa é realizada a codificação do framework OpenICGFw, levando em consideração as discussões anteriores e obtidos os requisitos primários para o núcleo principal do respectivo framework.

A programação do software seguiu padrões de desenvolvimento para web 2.0, adotando-se o PHP a partir da versão 5.3, como a linguagem

padrão utilizada para a codificação da estrutura principal, incorporando novas bibliotecas e funcionalidades disponibilizadas pela comunidade de software livre. A escolha da linguagem foi realizada por dois motivos: o primeiro em observância às políticas, premissas e especificações técnicas definidas pelos Padrões de Interoperabilidade de Governo Eletrônico proposto pelo Governo Brasileiro através de leis e decretos; o outro motivo refere-se à expressiva comunidade de usuários distribuída pelo mundo que utilizam e auxiliam no suporte para manter a tecnologia estável, além da facilidade de desenvolvedores absorverem com maior rapidez a sintaxe.

Para o gerenciamento do framework, foi realizada a estruturação do repositório principal, caracterizando uma base fixa para o armazenamento dos dados relacionais, possibilitando a parametrização de funcionalidades para a personalização do ambiente. Inicialmente, o repositório foi projetado visando centralizar dados específicos a partir de processos executados diretamente no framework. Desenvolvido no SGBD-PostgreSQL foi necessário devido às várias funcionalidades já implantadas, principalmente, aos recursos automatizados para o tratamento de dados geográficos (CASANOVA et.al., 2005). A possibilidade de desenvolver aplicativos em três camadas (cliente, servidor e banco de dados) possibilitou o surgimento de vários sistemas webmapping (MITCHELL, 2005 pp. 12-35), apresentando componentes importantes para sua utilização junto às instituições públicas.

Para arquitetura do projeto, foram elaborados modelos através de artefatos contendo especificações da Unified Modeling Language UML 2.0, em conjunto com especificação da OMT-G para extensão geográfica (BORGES, 2001). Além de pacotes essenciais, também foram detectadas várias temáticas que auxiliaram no entendimento das necessidades para colaboração entre os usuários, proporcionando uma infraestrutura base de dados geográficos junto ao framework.

### **3. ESPECIFICANDO A MODULARIZAÇÃO APLICADA**

Quanto à modularização aplicada, é realizada brevemente uma descrição sobre a mesma, idealizada para inserção de recursos básicos e avançados através do framework, obtendo-se a

partir da quinta etapa a criação do módulo de estatística para a elaboração de consultas aos dados espaciais devidamente registrados.

Visando a detecção automática de tendências e associações “escondidas” nos dados, na sexta etapa é realizada a implementação do módulo de mineração de dados para extração de regras em busca de conhecimento até então não observado conforme (WITTEN et. al. 2005) e (TAN et. al., 2009). As aplicações de mineração de dados são direcionadas muitas vezes sobre dados coletados para outra finalidade ou para aplicações futuras sem especificação concreta. Por este motivo, a mineração de dados acaba não tendo os benefícios que possibilitam questionar a fonte dos dados. Em comparação, grande parte das estatísticas lida com o projeto de experimentos ou pesquisas, que podem sofrer interferência na coleta dos dados. O uso das técnicas, entretanto, possibilita obter a geração de regras e a composição de cartogramas através de métodos computacionais automáticos e semiautomáticos.

Inicialmente, foram realizados experimentos utilizando-se as técnicas de associação, classificação e agrupamento. Com a implementação da mineração de dados, buscou-se ampliar os recursos e auxiliar em formulações analíticas de maior complexidade em grandes bases de dados espaciais (BOGORNY, 2006). Através da técnica de classificação, é avaliado o algoritmo ID3 e sua evolução para o (C 4.5) proposto por (QUINLAN, 1993), além do algoritmo Apriori para a técnica de associação, e o K-means para a técnica de agrupamentos, adaptados para investigação de dados espaciais através da web.

Na sétima etapa, o módulo de desenvolvimento auxiliou na criação de novos recursos através da implementação de várias rotinas para automatização dos processos (BOUYER et. al., 2005), (BAUER, 2009), (BATCHELLER et. al., 2010) e (QIANG et. al., 2011). Na oitava etapa, refere-se ao módulo para visualização dos resultados através de textos, tabelas e cartogramas, pois muitos conjuntos de dados são de difícil compreensão humana, devido ao seu volume e complexidade, principalmente quando não há acesso aos metadados. Um exemplo é a árvore de decisão, que pode ser visualizada na íntegra, mas apresenta uma difícil interpretação, sendo que, para facilitar seu entendimento no framework OpenIGCFw, as

regras são transformadas em cartogramas através de consultas com a utilização de recursos visuais, contribuindo na implementação de aplicações webmapping.

Ao final, na nona e última etapa, é possível acompanhar a implementação do módulo para monitorar e gerenciar os processos executados com o framework juntamente com os processos do sistema operacional.

### 3.1 Módulo de Colaboração – Provendo a interoperabilidade entre dados geográficos

Para promover a integração de dados através do framework OpenICGFw, inicialmente são restringidas algumas situações, permitindo especificar neste artigo uma forma simples para integrar informações, gerando conhecimento atualizado para a colaboração, o planejamento e a coordenação das ações conjuntas entre instituições públicas e também privadas. Alguns desafios, porém, ainda são enfrentados, a exemplo de dados armazenados em estações de trabalho e não em um servidor específico; dados essenciais e produtos não catalogados; além da falta de estrutura física e tecnológica, limitando o armazenamento de grandes quantidades de dados (RAGHU et. al., 2007) e (YANG et. al., 2010).

Para isso é necessário o desenvolvimento de sistemas de integração e colaboração para tratar de problemas inerentes da arquitetura, engenharia, construção e gestão de instalações (SHEN, 2010) principalmente com relação a catalogação e atualização permanente que promova a identificação de todo o volume de dados local e distribuído,

produzindo vários metadados dos acervos digitais. Cita-se a experiência com a implementação do GeoNetwork (<http://geonetwork-openource.org>), como uma solução livre e de código aberto para o gerenciamento desses dados, que se soma a novos parâmetros e funcionalidades, possibilitando, a partir da catalogação, a manipulação desses dados para a descoberta de conhecimento, a exemplo do monitoramento de divisas territoriais, do meio ambiente, da inclusão social e também do desenvolvimento sustentável das regiões. Adotou-se, assim, uma arquitetura composta de vários procedimentos que possibilitam desde a autorização até a transformação de dados específicos para sua visualização final.

Após a autorização, ocorre a transferência para o servidor, mas, para identificação do processamento dos dados de entrada, são necessários metadados para especificar melhor as características de um ou mais conjuntos de dados, além do seu armazenamento, local ou remoto. A partir desse diagnóstico é possível proceder na execução dos demais procedimentos, além de novos módulos, conforme Figura 2.

Um dos importantes diferenciais deste trabalho é o módulo de intercâmbio de dados geográficos, que visa proporcionar através de uma interface web acesso a esses recursos de forma mais simplificada, adotando os seguintes processos:

- (1) Validar e prover a interoperabilidade;
- (2) Catalogar os dados;
- (3) Leitura, validação e limpeza dos dados;
- (4) Transformação e padronização;
- (5) Armazenamento dos dados;

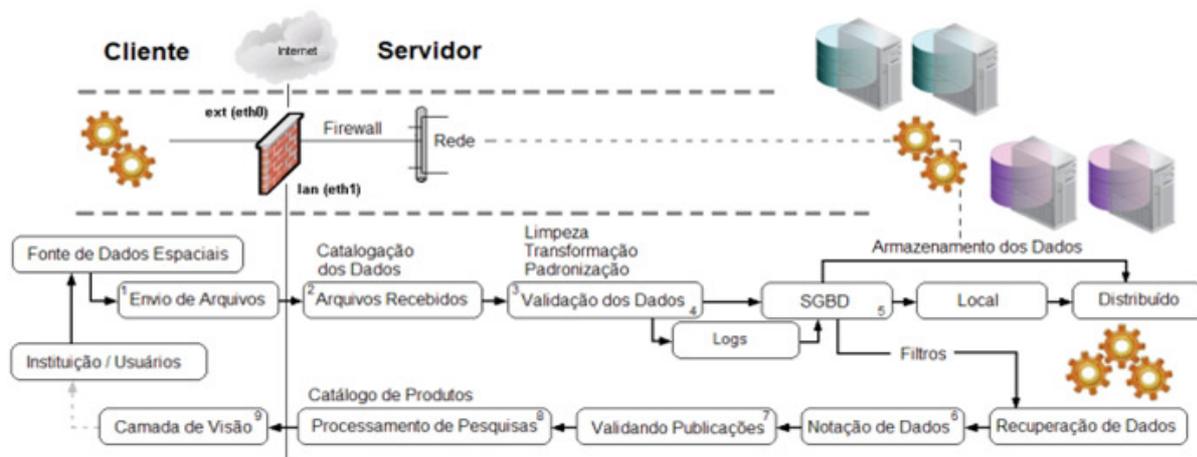


Fig. 2 - Aproximação dos processos para interoperabilidade dos dados geográficos através do framework.

- (6) Intercâmbio dos dados;
- (7) Validação para publicação;
- (8) Pesquisas;
- (9) Visualização dos resultados.

Com isso, para promover a integração de dados através do framework OpenICGFw, inicialmente são restringidas algumas situações, permitindo especificar neste trabalho uma forma simples para integrar informações e gerar conhecimento atualizado para a colaboração, o planejamento e a coordenação das ações conjuntas entre instituições públicas e privadas.

Uma alternativa para reduzir a produção de dados espaciais sem padrões definidos é o armazenamento centralizado e a catalogação estruturada, viabilizando seu gerenciamento distribuído. Mas, para este desafio em especial, identificar previamente a procedência dos dados é fundamental, tornando obrigatório o processo de geração dos metadados. A partir desses metadados, obtém-se a identificação para intercambiar estruturas previamente conhecidas, reduzindo significativamente a perda de qualquer característica original. Com isso, é obtido avanço significativo no processo de interoperabilidade entre os dados espaciais.

Ainda para o intercâmbio de dados foram avaliadas soluções de Extract, Transform, Load (ETL), Todas as soluções apresentaram fortes características positivas para sua adesão. Para o framework OpenICGFw, porém, surge a necessidade de recursos ainda inexistentes e não projetados para a parametrização desses processos.

### **3.1.1 Integração – Notação Adotada para Transformação de Dados no padrão XML/GML**

A integração de dados está diretamente ligada ao intercâmbio entre diferentes padrões dos dados conforme (FAN et. al, 2004), (RIGAUX et. al, 2007) e (SOIBELMAN et. al, 2008). Para a conversão dos dados, uma solução explorada refere-se à cópia dos dados originais para transformação, gerando um novo conjunto de dados, a fim de oportunizar a execução de processos computacionais complexos, apresentando uma notação específica em conformidade com OGC e W3C. Obtém-se tal notação pela implementação realizada devido à

necessidade de automação de processos paralelos, especificamente para finalidades distintas, tendo-se várias transformações e adequações a partir do mesmo conjunto de dados, para gerar novos datasets, submetidos posteriormente a diferentes algoritmos de mineração de dados através do framework proposto.

Vários aspectos já consolidados favorecem o armazenamento dos dados transmitidos para um repositório específico. Porém, inicialmente a integração no framework OpenICGFw parte da existência de um repositório no servidor, proporcionando a ligação entre as etapas citadas, através de um servidor de arquivos com a tecnologia freeNAS (<http://freenas.org/doku.php/>), melhor alternativa nos testes realizados por garantir acesso aos recursos de gerenciamento e, principalmente, pela conectividade através de vários protocolos de comunicação, a exemplo do Samba, FTP e NFS. Durante a execução dos testes, entretanto, detectou-se um crescimento surpreendente dos processos relativos à manipulação de grandes volumes de dados, levando rapidamente a uma capacidade limitada da estrutura física e lógica. A partir disso, foram restringidos o consumo e a manipulação de recursos durante o estudo de caso. Este fato desencadeou reflexões sobre novas alternativas para se trabalhar com dados distribuídos e também centralizados. Contudo, estas avaliações sobre os ambientes tecnológicos suportados são resultados a serem abordados em futuras discussões sobre a melhor implantação e operacionalização do framework.

Ainda com a realização dos processos de leitura, transformação e armazenamento, surgiu a necessidade e a padronização das conversões de dados para o framework, resultando na elaboração de uma notação específica para conversão de projetos na íntegra, conforme ilustrado na Figura 3.

Após identificação do projeto através de seu ID, são coletados os metadados necessários para catalogação junto à base de dados do framework.

Dentre seus principais atributos são citados: ID\_Project, Project\_Name, Format\_Origin e o Elaboration, que são fundamentais para as conversões e o devido rastreamento. Para realizar a implementação desta notação no framework, foi elaborado um programa na linguagem C++, contendo características específicas de um analisador sintático, denominado Parser,

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <DataFile xmlns:gml="http://www.opengis.net/gml"
   xmlns:xsi="http://www.w3.org/2000/10/XMLSchema-instance" >
3 <ID_Project> 000420820305-0001 </ID_Project>
4 <ProjectName> Setores Censitários do Município 420820305 - Itajai SC - Brasil </ProjectName>
5 <FormatOrigin> Shapefiles [shp] </FormatOrigin>
6 <Elaboration> IBGE – Instituto Brasileiro de Geografia e Estatística </Elaboration>
7 <GMLInputTemplate>
8 <CollectionElement>featureCollection</CollectionElement>
9 <FeatureElement>feature</FeatureElement>
10 <GeometryElement>geometry</GeometryElement>
11 <ColumnDefinitions>
12 <column>
13 <name>ID_</name>
14 <type>STRING</type>
15 <valueElement elementName="property" attributeName="name" attributeValue="ID_" />
16 <valueLocation position="body" />
17 </column>
18 <column>
19 <name>NAME1_</name>
20 <type>STRING</type>
21 <valueElement elementName="property" attributeName="name" attributeValue="NAME1_" />
22 <valueLocation position="body" />
23 </column>
24 </ColumnDefinitions>

```

\*\*\* Identificação dos projetos, formatação original e autoria.

Fig. 3 - Notação adotada para o XML/GML, gerado após a conversão de dados.

responsável pela leitura e escrita de arquivos com a codificação XML/GML, sendo possível, assim, transformar os dados para manipulação nos processos.

### 3.1.2 Intercâmbio para Dados Espaciais

Para promover o intercâmbio de dados, são incorporadas bibliotecas disponíveis no projeto INPE/TerraLib (2009), possibilitando integrar recursos específicos para interoperabilidade com produtos disponibilizados junto ao Instituto Nacional de Pesquisas Espaciais (INPE) do Brasil. Também foram incorporados softwares livres já consolidados, dentre eles o projeto *Geospatial Data Abstraction Library* GDAL (<http://www.gdal.org>), devido a vários recursos compatíveis para o framework.

Através do projeto Simple Feature Library (GDAL/OGR), foi possível realizar o intercâmbio de dados matriciais e vetoriais, por meio dos processos elaborados para execução no framework. Outra característica importante para a utilização da GDAL refere-se ao grande número de combinações que podem ser realizadas no processo de intercâmbio dos dados.

Especialmente para os dados matriciais foi possível alcançar um número maior de situações no processo de importação e transformação para formatos distintos, permitindo a conversão entre 50 diferentes formatos (50x50) possibilitando 2.500

combinações. Outra característica relevante são as 38 conversões possíveis para a criação e o georreferenciamento de dados (38x38), permitindo 1.444 combinações distintas.

### 3.2 Módulo de Mineração de Dados Espaciais

O módulo de mineração de dados consiste em um processo analítico projetado para explorar grandes quantidades de dados. Quando a mineração de dados é relacionada às pesquisas científicas para os dados geográficos, ocorre a busca de padrões consistentes e também de relacionamentos sistemáticos entre as variáveis. A partir disso, validar e aplicar os padrões detectados a novos subconjuntos de dados proporciona sua representação espacial.

Determinado dado espacial descreve uma localização ou uma forma, por exemplo: casa, hospital, estrada, rio, floresta, parque, entre outros. Um dado geográfico, por sua vez, descreve objetos ou fenômenos que acontecem na Terra e que têm associada uma posição geográfica. Três características principais descrevem um objeto geográfico: atributos não-espaciais (o quê), atributo espacial (onde) e as relações espaciais (como). Ainda os dados geográficos podem ser classificados em duas modalidades: objetos e fenômenos (KIM, 2009) e (ZHANGA et. al., 2010). Os principais relacionamentos espaciais ocorrem nos casos

topológicos quando se têm: a disjunção de objetos, um objeto está dentro de outro, objetos que se tocam, um objeto contém outro objeto, sobreposição de objetos, objetos se cruzam e quando os objetos são iguais.

Fenômenos geográficos possuem, na maioria das vezes, dependências espaciais complexas, tais como nos sistemas urbanos, bacias hidrográficas, rede viária, limites administrativos, dentre outros. Para lidar com essa complexidade, a caracterização espacial e a generalização são, portanto, tarefas significativas para a mineração de dados geográficos. A mineração de dados baseada em generalização pode seguir duas estratégias no framework OpenICGFw. A primeira refere-se à generalização espacial exclusiva e a segunda à generalização sem a representação espacial.

A primeira generalização agrega os dados com base em uma hierarquia de objetos fornecidos pelo usuário através de conceitos geográficos. O método de indução padrão é atribuído orientando a utilização em cada nível de agregação geográfica para determinar descrições compactas ou padrões de cada objeto (região). O resultado é uma descrição preexistente das regiões na hierarquia usando predicados de alto nível.

A segunda generalização gera unidades espaciais agregadas que compartilham da mesma descrição de alto nível. O método de indução é atribuído e usado para agregar atributos não espaciais em conceitos de alto nível. Em cada nível na hierarquia do conceito resultante, unidades geográficas vizinhas são fundidas, caso compartilhem da mesma descrição de alto nível. O resultado é uma hierarquia de agregação geográfica com base em várias dimensões.

Muitas das tarefas tradicionais de mineração de dados têm, contudo, funções análogas no domínio de mineração de dados geográficos. Além disso, o volume de dados geográficos combinado com a complexidade do acesso a eles para operações analíticas implica a redução da escalabilidade.

### **3.3 Módulo de Busca e Visualização da Informação**

A pesquisa em grandes acervos digitais de dados quase sempre detecta dados específicos sobre um fato ou uma ocorrência, tornando possível

fazer análises temporais sobre diferentes conjuntos de dados. O desafio está na operação desses conjuntos, a exemplo da união ( $A \cup B \rightarrow AB$ ), intersecção, diferença, produto cartesiano, junção, pois normalmente são armazenadas em estruturas de dados distintas pela falta de comunicação e padronização.

Para um fluxo de informação, a colaboração é detectada como uma das importantes etapas, a fim de minimizar a falta de comunicação, mas também é importante destacar que a cooperação acontece em diferentes situações, somente quando é possível obter todas as etapas realizadas, independentemente de qual sistema de informação seja utilizado. Para isso, não basta apenas que seja informada a existência de determinado conteúdo (dado espacial), se não for coordenada a sua distribuição (YANG et. al., 2010).

Somente após a distribuição e o livre acesso aos dados (conteúdos) é que a colaboração torna-se possível e, conseqüentemente, a cooperação dinâmica entre os usuários pode efetivamente acontecer. Com isso, busca-se com esse modelo de funcionamento conceber e facilitar a comunicação através da internet utilizando vários recursos, tais como: transferência de acervos, fórum de discussão, lista de discussão, blog de conteúdos específicos, calendário de atividades, enquetes para avaliar as questões de maior debate, que necessitam de decisões conjuntas, entre outros mecanismos. Ao intensificar a colaboração entre os diferentes usuários e, principalmente, entre as instituições públicas, a iniciativa acaba constituindo um importante meio de comunicação e aprendizagem, além de facilitar a integração e a catalogação de dados geográficos entre diferentes aplicações computacionais.

### **3.4 Módulo de Monitoramento do Framework**

O desenvolvimento dos mecanismos para avaliação do framework foi embasado na norma ISO/IEC 27002 (2005), que trata das técnicas de segurança – código e prática para a gestão da segurança da informação, contendo dois importantes aspectos. O primeiro refere-se ao monitoramento dos processos executados pelo próprio framework, gerando logs para avaliações, dentre as quais são citados os recursos de hardware mais utilizados, as funções mais executadas, o consumo do link da rede e o monitoramento da performance dos

algoritmos através de uma aplicação interativa na web. O segundo refere-se ao monitoramento de logs que são gerados automaticamente pelo sistema operacional Linux Debian (Squeeze), principalmente através de sensores instalados pelo software Syslog-ng, permitindo avaliar a demanda de serviços e a performance do servidor como um todo. Isto permite que as tarefas cotidianas no lado do servidor proporcionem a centralização dos processos com maior performance e baixo custo agregado. Com a utilização dos logs aliados aos parâmetros do projeto, é possível montar a árvore de avaliação do potencial, na qual são visualizados os elementos que necessitam de ações prioritárias, e, a partir dessa avaliação, planejar quais atividades serão realizadas para elevar o status de cada ponto elementar apresentado na seção 2.1.

#### 4. ESTUDO DE CASO – Detectando Padrões para Objetos Espaciais CAD/GIS

O estudo de caso apresenta a síntese de vários experimentos realizados para avaliar a implementação dos módulos junto ao framework OpenICGFW, visando principalmente a manipulação de dados espaciais para a área territorial do município de Itajaí, no Estado de Santa Catarina – Brasil.

Os dados foram obtidos através de fontes distintas submetidas a atualizações periódicas. No Instituto Brasileiro de Geografia e Estatística (IBGE,

2009), foram obtidos os dados censitários, disponibilizados livremente no formato SHP. Na Prefeitura Municipal de Itajaí (2010), foram obtidos através do Cadastro Técnico Urbano (CTU) dados específicos, sendo: (Conservação, Isento-iptu, Isento-taxas, Passeio, Situação-lote, Benfeitoria, Tipo-imóvel, Situação-construção, Espécie, Utilização, Acabamento, Patrimônio, Incra, Frente, Fundos, Área-construída, Topografia e Bairro). Já os dados parciais do mapeamento de áreas rurais, a exemplo da topografia, hidrografia, elevação foram obtidos da Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina (EPAGRI, 2010). Através do Instituto Nacional de Pesquisas Espaciais (INPE, 2010), foram obtidas imagens de satélite para acompanhamento de fenômenos naturais e sua temporalidade.

Por último, foram obtidos do Ministério do Meio Ambiente (MMA, 2010), dados da engenharia ambiental, ex. vegetação e matas nativas. Após a identificação da área observada no estudo de caso, foi realizado o levantamento sobre os metadados das principais temáticas conforme ilustrado na Figura 4. A partir do recebimento dos dados, foram realizados os procedimentos para catalogação junto ao framework.

Com a identificação da área a ser utilizada para implementação do estudo de caso, foi detectada a necessidade de obter os detalhes sobre a origem dos dados coletados, ou seja, os



Fig 4 - Contextualização das temáticas em níveis de granularidade.

metadados, extremamente importantes para as etapas seguintes do processamento.

Após o cadastramento das respectivas instituições e, também, a catalogação de suas bases, alguns dados foram submetidos ao processo de limpeza e transformação para uma notação específica, possibilitando a interpretação. Seguindo as atividades que tratam da aproximação com a interoperabilidade de dados geográficos entre padrões distintos, foram realizadas atividades de coleta, armazenamento e verificação, até a conversão final para integração. Durante esses processos de migração, ficou constatada a presença de objetos sem definição ou documentação sobre a estrutura, o que impediu o processamento devido à necessidade de tratamento diferenciado. Essas representações sem declaração receberam tratamento como objetos cognitivos, acarretando na intervenção de analista externo para adequar sua implementação.

Após adequação, os experimentos possibilitaram entender melhor tais problemáticas frequentemente encontradas na integração, colaboração e interoperabilidade entre os dados. A seguir são apresentadas duas avaliações que possibilitaram a codificação de rotinas específicas e otimizadas para esta face de integração junto ao framework.

Na primeira avaliação foram analisados aspectos entre diferentes fontes para a integração com o sistema de informações geográficas (S1) do IBGE/Estatcart, possibilitando o cruzamento de dados estatísticos desse sistema com as demais fontes junto ao framework (f) OpenICGFw. Na segunda avaliação, foram realizadas as atividades de recuperação, conversão e a importação dos conjuntos de dados do esquema referente ao boletim de informações cadastrais (S2) de Itajaí/BIC, tornando possível a intersecção dos dados espaciais, em que:  $(fn) \leftarrow (S1) \cap (S2)$ .

Ainda após a união dos dados (d1) da EPAGRI com os dados (d2) do MMA e com a união dos dados (d3) do INPE, foi possível a geração da função  $(fn) \leftarrow \{ (d1) \cup (d2) \cup (d3) \}$ .

#### 4.1 Implementando a Mineração de Dados Espaciais

Nesta fase de implementação do estudo de caso, são destacados os dados em formato vetorial para o experimento da hipótese, permitindo obter

informações preliminares da base de dados para aplicação de um modelo estatístico que possibilitou a visualização sintetizada da massa de dados através dos campos sumarizados. Esse procedimento é mais representativo para atributos com valores quantitativos, pois, além de fornecer informações sobre esses atributos, possibilita agir sobre os registros detectados com valores nulos ou incompletos. Por outro lado, para os registros que não apresentaram valores (falta de preenchimento), optou-se por não considerá-los para análise estatística, haja vista tal decisão não representar prejuízos. Na Figura 5 é apresentado o resumo dos atributos gerados após a intersecção dos dados.

Para o experimento  $(fn) \leftarrow (S1) \cap (S2)$ , foram selecionados todos os registros de (S1) sem perda ou descarte de dados. Para (S2), foram selecionados 49.413 registros de integrantes da base de dados original, acontecendo o descarte de 87 registros com valores nulos e que não atenderam às especificações, resultando num total de 49.325 registros.

O ponto forte deste experimento foi detectar conhecimento até então desconhecido, avaliando o estado de conservação dos imóveis urbanos e rurais no município (S2) através dos dados apresentados. Após o processo de implementação da mineração de dados sobre  $(S1) \cap (S2)$ , obteve-se a geração das regras.

Adotando-se a técnica de classificação, foram realizadas modificações no código fonte do algoritmo (C 4.5), sendo possível adequar sua execução no servidor e visualizá-lo diretamente no navegador web (browser). O experimento possibilitou criar uma derivação do C4.5 para a web, dando origem ao novo algoritmo inicialmente denominado OWC4.5-2010 (*open web classification*), visando à construção da árvore de decisão através de estruturas dinâmicas.

#### 4.2 Processando Estruturas

A vantagem de se utilizar um mecanismo para integrar diferentes fontes de dados sem dúvida alguma é otimizar significativamente o tempo destinado por seus usuários na elaboração e execução das pesquisas em busca de informações precisas. Para isso, no framework OpenICGFw foram elaborados recursos que permitem a passagem de parâmetros para o processamento em diferentes fontes de dados, a exemplo da origem



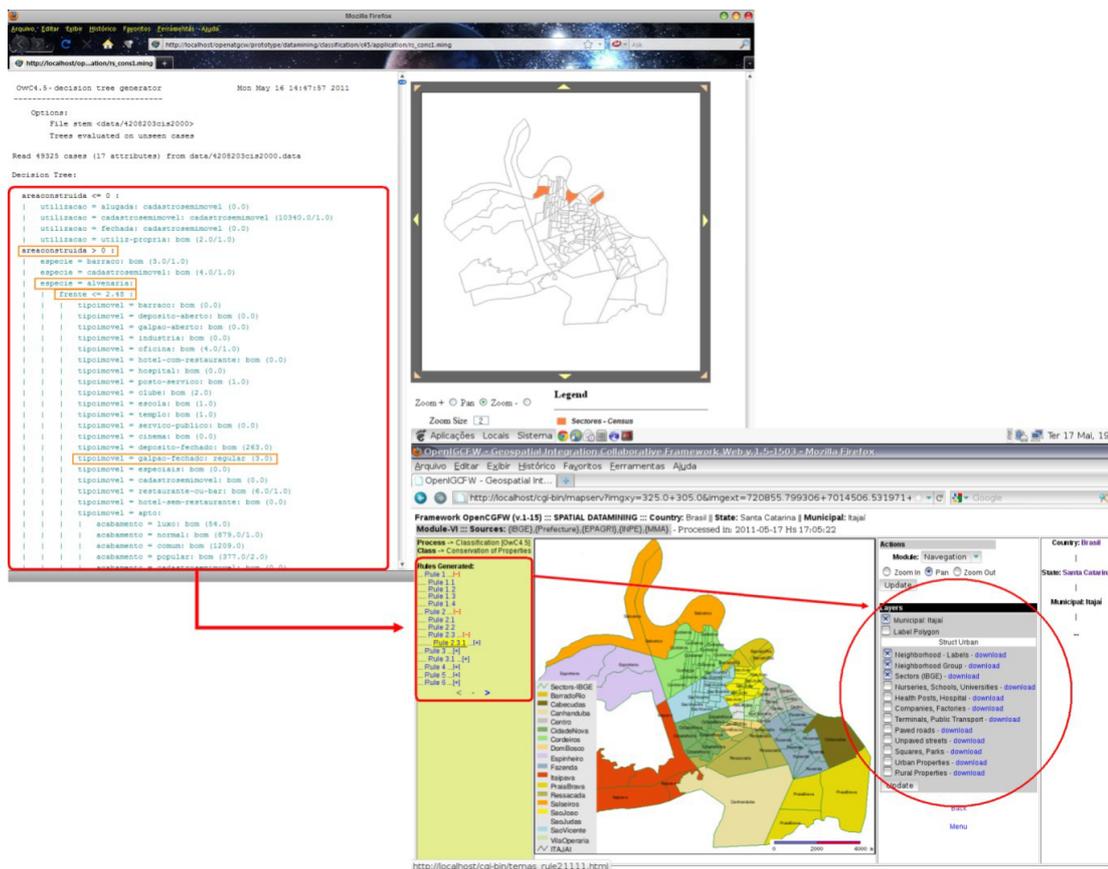


Fig 6 - Avaliação das regras individualizadas através das camadas para os mapas (cartogramas).

Isso também possibilitou acesso as consultas sem a presença de especialistas, trazendo aos decisores um mecanismo inovador para acompanhamento e auditoria de projetos. Finalmente, para as análises de alta complexidade, o módulo de visualização também proporcionou uma representação espacial (cartogramas).

## 5. CONCLUSÕES E TRABALHOS FUTUROS

Inicialmente, através da pesquisa foi possível detectar a existência de poucos estudos sobre a avaliação multicritério para o potencial de sistemas de informação geográficos colaborativos envolvendo principalmente decisores e usuários na construção de indicadores para o processo de desenvolvimento.

Tal fato impulsionou o desenvolvimento de um framework aberto em software livre, visando ao aumento da integração entre diferentes sistemas de informações, a maioria dos quais, ao serem elaborados, acabam deixando o ciclo de desenvolvimento fora das discussões junto às

comunidades. Diferentemente, neste trabalho, o desenvolvimento do framework para o modelo estratégico geoespacial de colaboração buscou compreender atividades distintas e relacionadas com o desenvolvimento de soluções webmapping (open source), além de compreender a estrutura de recursos existentes e possibilitar a reutilização de soluções existentes, principalmente nas etapas iniciais de coleta de requisitos essenciais para arquitetura do framework.

Através do trabalho, ficou comprovado o descarte da hipótese nula com relação à fraca sensibilidade do modelo de construção do framework apresentado através da análise de sensibilidade (seção 2.2). Ainda na arquitetura do framework, foram implementadas várias classes contendo atributos específicos, que permitiram estender funcionalidades para integração com outros sistemas de informações, visando principalmente ao tratamento e à integração de temáticas como meio ambiente, gestão territorial, administração pública, obras e construções, áreas construídas, entre outras. Ficando em conformidade com as diretrizes do

Ministério das Cidades (DOU, 2009, p.75 art.22), para integração de dados tendo como referência o CTM (Cadastro Territorial Multifinalitário).

A partir disso, no framework, foram otimizadas rotinas para melhorar a aproximação dos processos que tratam da interoperabilidade dos dados geográficos seção 3.1, possibilitando o aperfeiçoamento da comunicação, transmissão e colaboração entre usuários. Ainda com relação à interoperabilidade de dados geográficos, alguns trabalhos que servem de referência para este artigo apresentam apenas esquemas conceituais e padronizações para o intercâmbio. Originando adesão e padronização da notação em XML/GML, tornando-se item obrigatório de infraestrutura e implantação do framework junto às instituições.

Através do estudo de caso, foi possível constatar a performance dos processos responsáveis pelo intercâmbio dos dados espaciais, sendo que durante os experimentos ocorreram novas necessidades e, conseqüentemente, melhorias no código fonte do núcleo do framework foram implementadas. Por outro lado, ao possibilitar a colaboração e análise dos dados de forma dinâmica, foram implementados com sucesso os conceitos de computação distribuída, possibilitando análises em diferentes conjuntos de dados das temáticas apresentadas através da mineração de dados espaciais.

Avanços significativos foram alcançados através de experimentos exaustivos, utilizando-se principalmente algoritmos de classificação, associação e agrupamento. Porém, resultados mais expressivos foram extraídos com o método de classificação, alcançando-se os objetivos inicialmente estabelecidos para o artigo.

Destaca-se como a maior contribuição deste trabalho o levantamento das necessidades para identificação e adaptação de algoritmos, viabilizando a criação de regras através de um novo algoritmo para análise de objetos cognitivos voltados para a representação espacial. Contudo, através da pesquisa, tornou-se possível comprovar a aplicabilidade dos recursos de mineração de dados espaciais na web, provendo principalmente para os departamentos de engenharia um processo inovador de colaboração com o framework OpenICGfw, auxiliando principalmente na investigação de grandes volumes de dados, de forma que muitos consigam reutilizar de fato estruturas preexistentes e o

conhecimento adquirido. Ainda existem fatores, porém, que impedem a colaboração e a disseminação de dados espaciais junto às instituições. A partir deste trabalho, surgiram novas iniciativas e experimentos, sendo destacada a necessidade das melhorias para inclusão de novas bibliotecas, aumentando a compatibilidade para integração, devido à extensa lista de ferramentas comumente adotadas nos laboratórios de engenharia das instituições públicas.

Na elaboração dos algoritmos para extração das regras, foram detectados aspectos que serão discutidos em trabalhos futuros, dentre os quais cita-se os processos de investigação sobre ontologias espaciais para extração de regras e identificação de padrões. Também com o presente trabalho tornou-se possível o monitoramento contínuo para acompanhamento dos processos, das perspectivas evolutivas e tecnológicas com a modernização dos recursos adotados no tratamento dos dados espaciais.

## AGRADECIMENTO

Ao Programa de Pós-Graduação em Engenharia Civil da Universidade Federal de Santa Catarina (UFSC), pela oportunidade, estrutura e compartilhamento das pesquisas. Ao Instituto Federal Catarinense – campus Camboriú, em particular ao Setor de Tecnologia da Informação, pelo apoio na execução das atividades.

## REFERÊNCIAS BIBLIOGRÁFICAS

BANA E COSTA, C.A. **Processo de apoio à decisão: actores e acções; estruturação e avaliação**. CESUR, v. 618, p. 31, 1993.

BATCHELLER, J.K.; REITSMA, F. Implementing feature level semantics for spatial data discovery: Supporting the reuse of legacy data using open source components. **Computers, Environment and Urban Systems**, v. 34 n. 4 p. 333–344, 2010.

BAUER, K.W.; **City Planning for Civil Engineers, Environmental Engineers, and Surveyors**. CRC Press, 2009.

BOGORNY, V. **Enhancing Spatial Association Rule Mining in Geographic Databases**. Universidade Federal do Rio Grande do Sul. Tese do Programa de Pós-Graduação em Computação. Brasil, 2006.

- BORGES, K.A.V., DAVIS Jr., C.A., LAENDER, A.H.F. OMT-G: An Object-Oriented Data Model for Geographic Applications. **Geoinformática**, v. 5, n. 3, p. 221-260, 2001.
- BOUYER, C.G.; SZNELWAR, I.L.; Análise Cognitiva do processo de trabalho em Sistemas Complexos de Operações. **Ciência & Cognição**, v. 4 p. 02-24, 2005. Disponível em: <<http://www.cienciasecognicao.org/pdf/v04/m31529.pdf>>. Acesso em: 15 nov. 2010.
- CÂMARA, G.; MONTEIRO, M.A.; PAIVA, A.J.; SOUZA, C.R.; LOPES, C.J.; BARBOSA C.C.; FREITAS, M.U. **Intercâmbio de Dados Geográficos no Brasil: um formato aberto**. INPE. São José dos Campos-SP, Brasil, 2000. Disponível em: <<http://www.dpi.inpe.br/geobr/>>. Acesso em: 10 abr. 2009.
- CARVALHO, N.F.; ENSSLIN, R.S.; GALLON, V.A.; ENSSLIN, L. Uma Metodologia Multicritério (MCDA-C) para Apoiar o Gerenciamento do Capital Intelectual Organizacional: Proposta de Construção de um Modelo. ABEPRO/ENEGEP - XXVII Encontro Nacional de Engenharia de Produção. **Anais...**, 2007. Disponível em: <[http://www.abepro.org.br/biblioteca/ENEGEP2007\\_TR640478\\_9625.pdf](http://www.abepro.org.br/biblioteca/ENEGEP2007_TR640478_9625.pdf)> Acesso em: 04 ago. 2011.
- CASANOVA, M.; CÂMARA, G.; DAVIS JR, C. **Banco de Dados Geográficos**. Ed. MundoGEO. Curitiba-PR, Brasil, p. 10-85, 2005.
- DOU - Diário Oficial da União – Ministério das Cidades. **Portaria Nº. 511 de 07 de dezembro de 2009. Diretrizes para criação, instituição e atualização do Cadastro Territorial Multifinalitário - CTM nos municípios brasileiros**, v. 234 n. 1 p. 75, 2009. Disponível em: <<http://www.in.gov.br/imprensa/visualiza/index.jsp?jornal=1&pagina=75&data=08/12/2009>> Acesso em 09 ago. 2011.
- eGOV - **Governo Eletrônico**, 2010. Disponível em: <<http://www.governoeletronico.gov.br>> Acesso: 05 mai. 2010.
- ENSSLIN, L.; MONTIBELLER, G.N.; NORONHA, M. **Apoio à decisão – Metodologia para Estruturação de Problemas e Avaliação Multicritério de Alternativas**. Insular, 2001.
- EPAGRI - **Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina – Mapas Digitais**, 2010. Disponível em: <<http://circam.epagri.sc.gov.br/mapoteca/>>. Acesso em: 12 dez. 2010.
- FAN, W.; GAROFALAKIS, M.; XIONG, M.; JIA, X. **Composable XML integration grammars**. ACM CIKM '04 Proceedings of the thirteenth international conference on Information and knowledge management. NY USA. p. 2-11, 2004.
- FONSECA, T.F.; EGENHOFER, J.M. Sistemas de Informação Geográficos Baseados em Ontologias. **Journals Cognition and Computation**, v. 1 n. 2 p.155-180, 1999.
- GEHLERT, A.; WERNER, E. Toward a formal research framework for ontological analyses, **Advanced Engineering Informatics**, v. 21 n.2 p. 119–131, 2007.
- GOODCHILD, M; EGENHOFER, J. Max; FEGEAS, R.; KOTTMAN, C. **Interoperating Geographic Information Systems**. Publishers Springer, v. 495 p. 30-110, 1999.
- IBGE. **Instituto Brasileiro de Geografia e Estatística. Geociências**, 2009. Disponível em: <[http://www.ibge.gov.br/servidor\\_arquivos\\_geo/](http://www.ibge.gov.br/servidor_arquivos_geo/)>. Acesso em: 12 dez. 2010.
- INDE – **Infraestrutura Nacional de Dados Espaciais – Perfil de Metadados Geoespaciais do Brasil em Conformidade com a Norma ISO19115:2003**, 2009. Disponível em: <[http://www.concar.ibge.gov.br/arquivo/Perfil\\_MGB\\_Final\\_v1\\_homologado.pdf](http://www.concar.ibge.gov.br/arquivo/Perfil_MGB_Final_v1_homologado.pdf)>. Acesso: 10 abr. 2011.
- INPE – **Instituto Nacional de Pesquisas Espaciais - Imagens de Satélite**, 2010. Disponível em: <<http://www.dgi.inpe.br/CDSR/>>. Acesso em: 12 abr. 2011.
- INPE – **Instituto Nacional de Pesquisas Espaciais. Softwares Livres - Tutorial de Programação TerraLib**, 2009. Disponível em: <<http://www.terralib.org/docs/v313/>>. Acesso em: 12 abr. 2011.
- ISO/IEC 27002:2005, **Information technology – Security technical – Code of practice for information security management**. ABNT.

- Disponível em: <<http://www.abntcatalogo.com.br/norma.aspx?ID=11549>>. Acesso em: 12 abr. 2011.
- ISO/TS 10303-1129:2008. **Industrial automation systems and integration — Product data representation and exchange— Part 1129: Application module: External properties**. ISO - International Organization for Standardization. Disponível em: <<http://www.abntcatalogo.com.br/norma.aspx?ID=59038>>. Acesso em: 11 fev. 2011.
- JOHNSON, B.; WOOLFOLK, W.; MILLER, R.; JOHNSON, C. **Projeto de software flexível: desenvolvimento de sistemas para requisitos variáveis**. LTC. Brasil, 2008.
- KARABEGOVIĆ, A.; PONJAVIĆ, M.; **Integration and interoperability of spatial data in spatial decision support system environment**. IEEE MIPRO - Proceedings of the 33rd International Convention, p. 1266-1271, 2010.
- KIM, C. **Spatial Data Mining, Geovisualization**. International Encyclopedia of Human Geography, p. 332-336, 2009.
- LISBOA, G.S.F. **Zend Framework: Desenvolvimento PHP5 orientado a objetos com MVC**. Novatec. Brasil, 2008.
- MACÁRIO, C.G.N.; MEDEIROS, C.B. Specification of a Framework for Semantic Annotation of Geospatial Data on the Web. **SIGSPATIAL Special**, v. 1 n.1 p.27-32, 2009.
- MITCHELL, T. **Web Mapping Illustrated. Using Open Source GIS Toolkits**. O'Reilly, p. 12-35: 59-209, 2005.
- MMA - **Ministério do Meio Ambiente do Brasil. Mapas Interativos – Geoprocessamento**, 2010. Disponível em: <<http://www.mma.gov.br>>. Acesso em: 12 dez. 2010.
- MUN, D.; RAMANI, K. **Knowledge-based part similarity measurement utilizing ontology and multi-criteria decision making technique**. *Advanced Engineering Informatics*, 2010.
- NEUNZERT, G. M. **Subdividing the Land: Metes and Bounds and Rectangular Survey Systems**. CRC Press, 2010.
- OGC - **Open Geospatial Consortium**. Disponível em: <<http://www.opengeospatial.org>>. Acesso em: 04 abr. 2009.
- OOSTEROM, P.; ZLATANOVA, S. **Creating Spatial Information Infrastructures: Towards the Spatial Semantic Web**. CRC Press, 2008.
- PEACHAVANISHA, R.; KARIMI, Hassan A.; AKINCI, Burcu; BOUKAMP, Frank; An ontological engineering approach for integrating CAD and GIS in support of infrastructure management. **Advanced Engineering Informatics**, v. 20 p.71-88, 2006.
- POSENATO, D.; LANATA, F.; INAUDI, D.; SMITH, I.F.C.; Model-free data interpretation for continuous monitoring of complex structures. **Advanced Engineering Informatics**, v. 22 n.1 p.135-144, 2008.
- PREFEITURA Municipal de Itajaí – Mapas**, 2010 Disponível em: <<http://www.itajai.sc.gov.br/mapas.php>>. Acesso em: 12 dez. 2010.
- QIANG, W.; KONSTANTINOV, A.; ZOU D.; YANG, L.T.; A Standards-based Interoperable Single Sign-on Framework in ARC Grid Middleware. **Journal of Network and Computer Applications**. v. 35, n 3, p. 892-904, 2011.
- QUINLAN, J.R. **C4.5 Programs For Machine Learning**. Morgan Kaufmann. California, p. 17-25; 45-55, 1993.
- RAGHU, T.S.; CHEN, H.; Cyberinfrastructure for homeland security: Advances in information sharing, data mining, and collaboration systems. **Decision Support Systems**, v. 43 n. 4 p. 1321-1323, 2007.
- RIGAUX, P.; SCHOLL, M.; VOISARD, A. **Logical Models and Query Languages**. Spatial Databases With Application to GIS, p. 69-111, 2007.
- SANTOS, D Carlos. **Open Source Software Projects' Attractiveness, Activeness, and Efficiency as a Path to Software Quality: An Empirical Evaluation of Their Relationships And Causes**. Southern Illinois University Carbondale (Thesis), p. 9-65, 2009.
- SHEN, W.; HAO, Q.; MAK, H.; NEELAMKAVIL, J.; XIE, H.; DICKINSON, J.; THOMAS, R.; PARDASANI, A.; XUE, H. Systems integration and collaboration in architecture, engineering, construction, and facilities management: A review. **Advanced Engineering Informatics**, v. 24 p. 196-207, 2010.

- SHERMAN, E.G. **Desktop GIS Mapping the Planet with Open Source Tools**. Publisher Pragmatic Bookshelf – Shelving Programming, 2008.
- SOARES, W. **Crie um Framework para sistemas web com PHP5 e Ajax**. Editora Érica. Brasil, 2009.
- SOIBELMAN, L.; WU, J.; CALDAS, C.; BRILAKIS, I.; LIN, K.Y.; **Management and analysis of unstructured construction data types**. *Advanced Engineering Informatics*, v. 22 p. 15–27, 2008.
- STEUDLER, D.; TÖRHÖNEN, M.P. **FLOSS in Cadastre and Land Registration**. Edition by Food and Agriculture Organization of the United Nations. p. 4-49, 2010.
- TAN, P.N.; STEINBACH, M.; KUMAR, V. **Introduction to DATAMINING**. Person Education, 2009.
- WITTEN, I.; FRANK, E. **Data mining: practical machine learning tools and techniques**. (Second edition). Morgan Kaufmann, v. 2 p. 37-76, 2005.
- YANG, C.; RASKIN, R.; GOODCHILD, M.; GAHEGAN, M. Geospatial Cyberinfrastructure: Past, present and future. **Computers, Environment and Urban Systems**, v. 34 n. 4 p. 264–277, 2010.
- ZHANGA, C., ZHAOB, T.; LI, W. Automatic search of geospatial features for disaster and emergency management. **International Journal of Applied Earth Observation and Geoinformation**, v. 12 n. 6 p. 409–418, 2010.