

## **Procedimentos para Levantamento e Análise de Dados para Confecção de uma Banco de Dados Geográfico.**

**Eng. Mestrando Dirceu de Menezes Machado Júnior <sup>1</sup>  
Profa. Dra. Ruth Emília Nogueira Loch <sup>2</sup>**

**<sup>1</sup> UFSC - Depto. de Engenharia Civil  
Programa de Pós Graduação em Engenharia Civil PPGEC/UFSC  
88040-970 Florianópolis SC  
dirceu73@terra.com.br**

**<sup>2</sup> UFSC - Depto. de Engenharia Civil  
Programa de Pós Graduação em Engenharia Civil PPGEC/UFSC  
88040-970 Florianópolis SC  
renloch@ecv.ufsc.br**

**Resumo:** O levantamento e análise dos dados necessários para composição de um banco de dados geográfico é um processo que demanda uma maior atenção por parte dos administradores de banco de dados, pois a qualidade final de uma base de dados para aplicações em geoprocessamento é reflexo da qualidade dos dados componentes do mesmo. Este trabalho visa então estabelecer uma proposta para análise e controle de qualidade dos dados componentes de um banco de dados geográfico.

**Palavras chaves:** Geoprocessamento, banco de dados geográfico, qualidade de dados

**Abstract:** The collect and analysis of the necessary data for composition of a geographic data base is a process that demands a bigger attention on the part of the data base administrators, because the final quality of a database for geoprocessing applications is a reflection of the quality of the component data. Therefore this paper aims to establish a proposal for analysis and quality control of the component data for a geographic database.

**Keywords:** Geoprocessing, geographic database, data quality.

## 1 Introdução

A primeira etapa na confecção de um banco de dados geográfico para análise da região de influência de um empreendimento hidrelétrico diz respeito ao levantamento e análise dos dados pré-existentes sobre a área em questão. Nesta etapa devem ser pesquisados e adquiridos todos os dados relevantes sobre a área a ser estudada.

Segundo Burrough (1992), a aquisição de dados é a etapa mais longa e onerosa na estruturação de um banco de dados geográficos. A avaliação para obtenção de dados deve então ser feita de forma cuidadosa, evitando a inclusão de dados redundantes ou desnecessários à aplicação a que o banco de dados se destina.

É importante ressaltar que durante o período de aquisição dos dados deve-se levar em consideração a forma de armazenamento (analógico ou digital) em que se encontra a fonte original de um determinado dado. No caso da fonte do dado estar em meio analógico (papel, fotografia aérea), deve-se avaliar a melhor maneira de conversão do mesmo para o meio digital, de forma a manter a integridade da informação constante na mesma.

Durante a etapa de levantamento e aquisição dos dados, deve-se então, fazer uma análise prévia da qualidade da informação contida nos mesmos. A análise da qualidade dos dados deve ser efetuada com base em orientações bem definidas e parâmetros que possam ser mensuráveis, de forma a evitar a subjetividade (que por sua vez pode levar ao erro). Essa análise, então, possibilitará ao gestor do banco de dados, decidir sobre a inclusão e/ou a melhor maneira do ajuste das informações contidas na fonte de dados analisada. A análise da qualidade deve também ser feita durante a estruturação do banco de dados geográficos, principalmente no que diz respeito à atualidade da informação contida, muitas vezes uma informação pode perder o valor por se tornar um dado obsoleto (por exemplo, mudança de proprietário ou mudança de limite de propriedade) Através da análise prévia e posterior da qualidade dos dados se garante a qualidade não só dos dados, mas também do produto final, um banco de dados geográfico confiável e seguro.

As seções 2 e 3 desse trabalho argumentam sobre a natureza dos dados constantes em um banco de dados geográfico, bem como as fontes e a maneira de aquisição dos mesmos. A seção 4 discute acerca de fatores importantes para uma avaliação da qualidade dos dados e estimação de parâmetros para sua mensuração. Finalmente a seção 5 apresenta conclusões e recomendações.

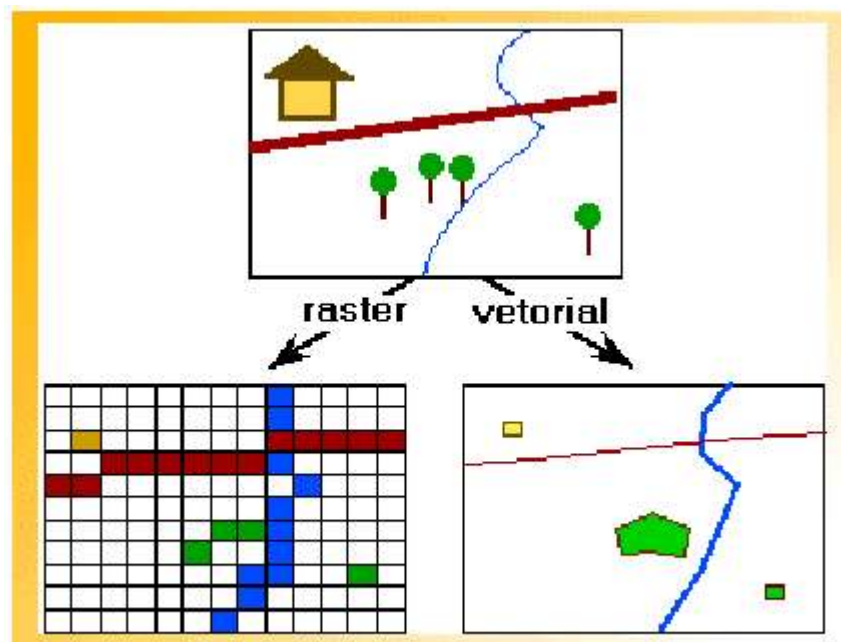
## 2 Natureza dos Dados Geográficos

Os dados componentes de um banco de dados geográfico podem ser classificados em três categorias: dados convencionais ou alfanuméricos - usados para descrever e caracterizar os objetos gráficos (por exemplo, nomes dos proprietários, quantidade de dependentes, número de usuários da terra, culturas cultivadas, tipo de técnica de cultivo utilizada, entre outros); dados espaciais – descrevem a geometria, a localização e os relacionamentos topológicos dos objetos geográficos; e dados pictóricos – atributos que armazenam imagens (por exemplo, fotografia da propriedade). (Lisboa Filho, 2001).

Por sua vez os dados espaciais podem ser representados por duas abordagens principais: o modelo matricial (ou raster) e o modelo vetorial.

Na estrutura matricial, a área em questão é dividida em uma grade regular de células de formato, normalmente, retangular. A posição da célula é definida pela linha e pela coluna onde está localizada na grade. Cada célula armazena um valor que corresponde ao tipo de entidade que é encontrada naquela posição. Uma área geográfica pode ser representada através de diversas camadas, onde as células de uma camada armazenam os valores associados a uma única variável (ex.: vegetação) (Chrisman, 1997). As camadas ficam totalmente preenchidas, uma vez que cada célula corresponde a uma porção do espaço sendo representado.

Na estrutura vetorial, cada fenômeno geográfico é representado, no banco de dados, por um objeto com identificação própria e representação espacial do tipo ponto, linha, polígono ou um objeto complexo. A posição de cada objeto é definida por sua localização no espaço, de acordo com um sistema de coordenadas. Objetos vetoriais não preenchem todo o espaço, ou seja, nem todas as posições do espaço necessitam estar referenciadas na base de dados. (Jensen, 1996)



**Figura 1** – Esquema ilustrando a diferença entre dados raster e vetoriais  
Fonte: Lisboa Filho, 2001

### 3 Fontes e Métodos para Aquisição de Dados

Os dados componentes de um banco de dados geográfico podem ser entidades ou fenômenos geográficos distribuídos sobre a superfície da terra, podendo pertencer a sistemas naturais ou criados pelo homem, tais como vegetação, tipo de solo, aspectos climáticos, dados hidrológicos, entre outros. Podem também ser objetos resultantes de projetos envolvendo entidades que ainda não existam, como o planejamento de uma estrada a ser implantada. (Ramirez, 1994).

As principais fontes de obtenção dos dados que compõem o banco de dados geográfico são obtidos a partir da fotogrametria, sensoriamento remoto e levantamento de campo. Sendo assim, a transferência dos dados do mundo real para o digital é apenas um passo no processo de aquisição dos dados. Muitas operações posteriores deverão ser realizadas como a associação entre os objetos gráficos e seus atributos não-gráficos, operações para corrigir e padronizar os dados, entre outros.

Outra forma de adquirir dados espaciais é via a digitalização manual (com mesa digitalizadora), a digitalização automática (ou escanização). Todos os dois permitem a transferência dos dados de mapas analógicos existentes para a base de dados geográfica. A digitalização é o processo de conversão de pontos e linhas de um mapa convencional para um formato compatível no uso em computador, podendo ser manual ou automática. (Tomlinson & Boyle, 1981).

A digitalização automática ou rasterização é a discretização do mapa ou imagem analógica em unidades retangulares homogêneas ou "pixels" através do uso do scanner. Este é um dispositivo ótico-eletrônico, composto por uma fonte de luz e um sensor ótico. Cada "pixel" detectado possui tamanho e cor, características que variam conforme a resolução espacial e radiométrica do sistema sensor usado. (Scarim et al, 1994). O arquivo gerado nesse processo é um arquivo matricial (raster).

A digitalização manual é o método no qual um mapa é colocado sobre uma mesa digitalizadora e, através de um dispositivo de apontamento um operador vai assinalando diversos pontos, que são calculados e interpretados como pares de coordenadas x e y. Pode ser feita de modo ponto a ponto, onde as coordenadas são lidas pela mesa e enviadas ao computador cada vez que o operador emita um sinal, ou de modo contínuo, onde as coordenadas são lidas continuamente, na medida em que o operador percorre uma feição e enviadas seqüencialmente ao computador. O modo contínuo é indicado para a digitalização de arcos extensos como curvas de nível, já para poligonais (como limites de propriedades) é mais indicado o modo ponto a ponto. Esse tipo de digitalização cria um arquivo vetorial.

Atualmente essa digitalização vem sendo feita em software que trabalham com o resultado da escanização do mapa em meio digital (após operações para georreferenciamento do mesmo em um sistema de coordenadas), criando dessa forma um arquivo vetorial composto por linhas, polígonos e pontos que representam as principais feições constantes no mapa utilizado, tais como, limites de propriedade, curvas de nível, edificações, entre outros. Essa tecnologia tem permitido a facilitação e a agilização do processo de digitalização que por muitas vezes era feito de forma cansativa e lenta, bem como a diminuição de erros cometidos pelo operador com o uso de mesas digitalizadoras uma vez que a digitalização feita de forma digital permite que a mesma seja feita na tela com o mapa original ao fundo, permitindo um maior controle do processo.

Existe ainda a digitação via teclado a qual é usada para a inserção dos atributos não gráficos (alfanuméricos ou tabulares). Informações provenientes de levantamento de campo normalmente são inseridas no banco de dados via teclado.

Outro método que tem sido bastante usado atualmente para obter dados espaciais é o GPS (Global Positioning Systems). Esse sistema possibilita a realização de levantamentos de campo, com alto grau de precisão e com o registro dos dados podendo ser realizado diretamente em meio digital.

#### 4 Análise da Qualidade dos Dados

Em virtude dos dados componentes de um banco de dados geográficos serem obtidos de variadas formas (digitalização, imageamento, incursões a campo, aplicação de questionários, entre outras), deve ser feita uma análise cuidadosa no que diz respeito à qualidade desses dados. É essa análise que permitirá selecionar os dados que comporão ou não o banco de dados geográfico a ser usado.

Segundo Chrisman (1984), qualidade representa o quanto um conjunto de dados se ajusta às necessidades de determinados usuários considerando-se suas aplicações geográficas.

Deve ser ressaltado que para a avaliação da qualidade de um conjunto de dados não basta apenas a análise dos dados em si, mas também a análise de todo o processo obtido para a aquisição do dado. (Bähr et al, 1999)

Também é muito importante saber a qualidade dos dados obtidos nos processos de aquisição, isto é, dos dados convertidos. Portanto, a garantia de qualidade é uma atividade de verificação final dos dados convertidos antes de serem carregados no banco de dados geográfico. Os dois principais objetivos da garantia de qualidade são: monitorar o nível de qualidade final dos dados convertidos e assegurar que todo o processo de controle de qualidade esteja sendo desenvolvido apropriadamente. (Medeiros & Alencar, 2000)

No gerenciamento da qualidade dos dados é necessário estimar os valores dos componentes de qualidade antes da montagem da base de dados, e é necessário haver uma redundância nos dados para uma estimativa posterior dos componentes da qualidade.

##### 4.1.Fatores a serem analisados para a qualidade dos dados.

Os fatores mais importantes para se ter um controle da qualidade dos dados em um banco de dados geográficos são (NCGIA, 1990):

- Acurácia posicional (qualidade do georreferenciamento);
- Dados de atributos (Acurácia dos atributos descritivos);
- Consistência das ligações entre dados geométricos e descritivos;
- Consistência das relações geométricas;
- Completude dos dados;
- Consistência lógica dos dados;
- Linhagem dos dados.

A **Acurácia posicional** (ou qualidade do georreferenciamento dos dados gráficos) diz respeito à medida de quanto o dado difere espacialmente (em termos de posição absoluta, relativa e forma) daquele tido como referência. É usualmente testada pela seleção de uma amostra específica de pontos predeterminados e comparação das coordenadas destes, com uma fonte independente, de qualidade

conhecida. A acurácia concerne à dispersão dos erros posicionais dos dados, sendo comumente estimada pelo cálculo do desvio padrão dos erros de posição dos pontos de teste selecionados. Outra forma de medição deste erro, usual em se tratando de coletas de campo e fotogrametria é o RMS (*Root Mean Square Error*). (Weber et al, 1999)

A **Acurácia dos atributos** retrata a fidelidade dos dados descritivos, com uma avaliação geral da identificação de entidades e atribuição de valores de atributo no conjunto de dados. A avaliação é feita basicamente através de funções estatísticas que medem a concordância/discordância dos atributos existentes em relação àqueles tidos como verdadeiros.

A **Completude** ou Completitude engloba informações sobre omissão (esquecimento) e comissão (criação do que não existe de fato), critérios de seleção, generalizações, definições utilizadas e outras regras de mapeamento.

A completude de um conjunto de dados é avaliada segundo três categorias: completude de cobertura, de classificação e de verificação. A completude de cobertura é a porcentagem dos dados disponíveis em uma área de interesse (em função de um total estimado). A completude de classificação e de verificação ajudam a determinar a conveniência de um conjunto de dados para uma dada aplicação. A completude de verificação refere-se à quantidade e distribuição das medidas de campo ou outras fontes independentes de informação utilizadas para validar os dados. A avaliação de completude é usualmente limitada à informação de cobertura. Normalmente informações de completude de classificação e verificação são ignoradas quando se considera a completude de dados geográficos (Medeiros & Alencar, 2000).

A **Consistência Lógica** informa sobre a manutenção de relações lógicas e topológicas consistentes. Os testes de consistência lógica incluem testes de valores válidos, testes gerais para dados gráficos (por exemplo, se os nós estão todos unidos, se os polígonos estão todos fechados, entre outros) e testes topológicos específicos (por exemplo, se o limite de polígonos vizinhos não estão se cruzando).

A **Linhagem** traz informação sobre os eventos, parâmetros e dados de origem que construíram o conjunto de dados documentados. As informações de linhagem incluem referências sobre quais dados originais foram usados para gerar o conjunto de dados, a escala e o meio de cada dado (papel, digital, dados de campo), descrição dos processos utilizados para obtenção dos dados, entre outros. A linhagem fornece informações com valores quantitativos e texto descritivo. (Weber et al, 1999).

Imagens, ilustrações, desenhos, tabelas e outras figuras podem ser criadas em qualquer padrão comum, com preferência em GIF (desenhos ou imagens com poucas cores diferentes) ou JPG (se tiver muitas tonalidades de cores).

#### 4.2. Parâmetros utilizados para mensuração da qualidade dos dados

Para que se possa fazer a avaliação da qualidade dos dados em uma base de dados geográfica é necessária a estimação de parâmetros que permitam a medição física da qualidade.

Existem vários parâmetros que podem ser usados para essa mensuração, dentre os quais pode-se destacar os seguintes (BÄHR, 1999):

- Acurácia
- Precisão
- Confiabilidade
- Grau de erro da classificação.

A **acurácia** é definida como o grau de proximidade das medições com seus valores reais. Normalmente a acurácia é caracterizada pelo desvio padrão ou desvio médio quadrático (RMS). Se o resultado é caracterizado pela variável aleatória  $x$ , o desvio padrão pode ser dado por:

$$s = \sqrt{E[(x - E(x))^2]}$$

onde :  $E$  = valor real conhecido.

O coeficiente  $E$  (valor real) pode ser obtido a partir de comparação com dados de outras fontes (mapas,

cartas topográficas), valores colhidos em campo (como pontos de controle para avaliação do georreferenciamento), análise dos dados de atributos (medições de limites em campo, medição de vazões reais para calibração do modelo), análise do erro conhecido pelos diferentes tipos de entrada (erro posicional do mapa de origem, acurácia na digitalização) entre outros.

Durante o período de aquisição dos dados uma estimativa preliminar do valor da acurácia dos dados é muito útil, desde que a acurácia necessária para entrada dos dados seja dada (por exemplo, estabelecimento de padrões). Neste caso, um método para aquisição de dados pode ser usado desde que, o valor estimado preliminarmente da acurácia do dado seja menor que o valor padronizado (ou requerido).

No período que segue a aquisição dos dados, a medida da acurácia dos dados adquiridos pode ser feita utilizando-se do desvio padrão como mostrado na fórmula anterior. O valor real dos dados pode então ser observado através de consulta a dados de referência mais acurados.

A **precisão** é definida como o número de casas decimais, ou dígitos significativos, nas medidas dos dados. A precisão não é o mesmo que acurácia, ou seja, um grande número de dígitos significativos não necessariamente indica que uma medida é acurada. A precisão expressa a repetibilidade das medidas e é determinada pelos instrumentos, ou meios, usados para a tomada da medição dos dados. (Weber et al, 1999)

A **confiabilidade** fornece informações a respeito da facilidade para controlar e sobre erros grosseiros nos dados. O valor do menor erro bruto detectável nos dados é caracterizado pela confiabilidade. No caso de medidas normalmente distribuídas, o valor da confiabilidade de um conjunto de dados, durante o período de aquisição dos mesmos, pode ser determinado pela avaliação das estatísticas das medidas. Para a determinação da confiabilidade de dados isolados, após os mesmos já terem sido adquiridos, podem ser usados métodos de ajustamento de observações fotogramétricas.

A condição preliminar para a determinação do valor numérico da confiabilidade de um conjunto de dados é a existência da redundância. No período projetado para a aquisição de dados pode ser estimada a redundância do conjunto de dados. A redundância de um conjunto de dados pode ser dado pela seguinte equação (BÄHR et al, 1999):

$$h = N - n$$

onde,

N é o número total de dados existentes para determinada medida;  
n é o número de dados necessários para a solução única.

A razão  $t=h/n$  é uma medida da facilidade para controle (de erros brutos) de um conjunto de dados. Se

- $0 < t < 0,01$  o conjunto de dados não é controlável
- $0,01 < t < 0,1$  o conjunto de dados é fracamente controlável;
- $0,1 < t < 0,3$  o conjunto de dados é suficientemente controlável;
- $0,3 < t$  o conjunto de dados é bem controlável.

Por sua vez, a confiabilidade de cada dado individual, após a efetiva aquisição dos dados pode ser determinada usando os métodos de ajustamento. Para o caso de uma distribuição normal de dados não correlacionados pode-se usar a seguinte fórmula:

$$r_i = g s_i \sqrt{\frac{1}{m_i}}$$

onde:

$g = g(\alpha, \beta)$  ( $\alpha$  e  $\beta$  são níveis de significância)  
 $s_i$  é o desvio padrão prévio da medida  
 $m_i$  é a raiz do elemento diagonal da matriz abaixo:

$$[Q-A (A^t W A)^{-1} A^t] W$$

onde:

Q é a matriz de cofator da medida  
A é a matriz projetada das equações das observações  
W é a matriz diagonal de peso das medidas.

Este método de cálculo da confiabilidade pode ser usado no caso de um ajustamento médio das medidas após a aquisição efetiva dos dados.

O **grau de erro da classificação** é uma medida da qualidade de um dado descritivo específico. Pode ser determinado como função de uma matriz de erros de classificação. As funções mais utilizadas são a porcentagem ou o índice Cohen-Kappa.

## 5. Considerações Finais

A etapa de aquisição e análise dos dados pré-existentes se constitui em uma etapa primordial na estruturação e implantação de um banco de dados geográficos. É nessa etapa que deve ser feita uma análise cuidadosa da qualidade dos dados que comporão a base de dados, de acordo com fatores e parâmetros já mencionados anteriormente. A análise da qualidade deve também contemplar a forma de aquisição dos dados e todo o processo sofrido pelo mesmo, desde sua geração até a sua disponibilização, passando, se existente, pela forma de conversão do mesmo. A qualidade de um banco de dados geográfico é reflexo direto da qualidade da informação nele contida.

É necessário então, que, previamente à implantação de uma base de dados geográfica sejam estabelecidos padrões e formas de mensuração dos dados que a comporão, garantindo dessa forma, a integridade e confiabilidade do banco de dados geográfico a ser implementado.

O referencial teórico abordado nesse artigo faz parte de um projeto de mestrado que tem por finalidade a estruturação de um banco de dados geográfico dentro da área de influência do reservatório da Usina Hidrelétrica de Itaipu. Até o momento já se obteve alguns dados acerca da área em questão, e os mesmos estão ainda em fase de análise.

Com base nos dados disponíveis sugere-se na tabela abaixo alguns itens que devem ser avaliados, considerando dados gráficos, cartográficos e alfanuméricos que possam vir a ser inseridos em um banco de dados geográfico da área estudada. É necessário se ter em mente durante a avaliação dos dados, fatores que atestam a qualidade dos dados, conforme descrito nas seções anteriores, para se certificar acerca da qualidade dos dados a serem inseridos no banco de dados.

**Tabela 1** – Parâmetros a serem avaliados na base de dados

Tipo de Dado	Parâmetros Avaliados
Imagens Aéreas	Forma de armazenamento: analógico ou digital; Ano do vôo; Escala do vôo; Câmera fotogramétrica e certificado de calibração; Georreferenciamento: sistema de projeção, datum, fuso, acurácia; Escala da restituição (se existente);
Imagens de Satélite	Formato de armazenamento do arquivo: jpg, tiff, geotiff; Ano da imagem; Resolução Espacial; Resolução Espectral; Resolução radiométrica; Composição das Bandas; Georreferenciamento: sistema de projeção, datum, fuso, acurácia;
Mapas	Forma de armazenamento: analógico ou digital; Escala do mapa; Entidade produtora; Ano de produção; Acurácia/precisão das informações contidas no mapa; sistema de projeção, datum, Qualidade das informações contidas;

Arquivos Vetoriais	Formato de armazenamento: shapefile, dwg, dgn, dxf, entre outros; Georreferenciamento: sistema de projeção, datum, fuso, precisão; Base cartográfica utilizada; Fonte da vetorização: mapas, levantamento de campo, retificação, entre outros; Confiabilidade das informações contidas;
Dados socioeconômicos	Forma de obtenção dos dados; Avaliação do questionário utilizado; Ano de obtenção e do cadastro; Formato de armazenamento: analógico, digital; banco de dados utilizado; Confiabilidade do cadastro; Necessidade de complementação de dados;
Outros dados descritivos	Tipo dos dados: hidrológicos, sedimentologia, qualidade da água, entre outros; Qualidade e consistência dos dados; Ano de obtenção; Necessidade de inclusão dos mesmos no sistema.

## 6 Referências Bibliográficas

- Aronof, S.** *Geographic Information Systems: a management perspective*. Canada: WDL Publications, 1989.
- Bähr, H.P.; Vögtle, T.;** *GIS for Environmental Monitoring*, Stuttgart:Scweizerbart, 1999.
- Burrough, P.A.;** *Principles of Geographical Information Systems for Land Resources Assessment*. Oxford: Oxford University Press, 1992.
- Chrisman, N.** *Exploring Geographic Information Systems*. New York: John Wiley & Sons, 1997.
- Jensen, J.R.** *Introductory Digital Image Processing 2º edition*, New Jersey, Prentice Hall Inc., 1996.
- Lisboa Filho, J.; lochpe, C.;** *Introdução a Sistemas de Informações Geográficas com Ênfase em Banco de Dados*; Outubro, 2001.
- Medeiros, C.B.; Alencar, A.C.;** *Qualidade dos Dados e Interoperabilidade em SIG*, UNICAMP, 2000.
- NCGIA;** *National Center for Geographic Information and Analysis: Core Curriculum – Introduction*, 1990.
- Ramirez, M. R.** *Sistemas Gerenciadores de Bancos de Dados para Geoprocessamento*. Rio de Janeiro: COPPE/UFRJ, 1994. Dissertação de Mestrado.
- Rech, J.V.;** *Base cartográfica Digital Comum para Concessionárias de Serviços Públicos e Prefeituras Municipais, Utilizando-se SIG Florianópolis*, 1997, 151p., Disertação de Mestrado, PPGEC/UFSC.
- Scarim, J.L.; Teixeira, L.A.** *Digitalização e Conversão Raster/Vector de Mapas*, Revista Fator GIS, nº06, Sagres Editora, 1994.
- Tomlinson, R.F.; Boyle, R.;** *The State of Development of Systems for Handling Natural Resources Inventory Data*. Cartographica, 1981.
- Weber, E.; Anzolch, R; Lisboa Filho, J.; Costa A.C.; lochpe, C.;** *Qualidade dos Dados Geoespaciais*, UFRGS, Instituto de Informática, 1999.