

Banco de Dados Biorregional para o Semi-Árido no Estado da Bahia

Jocimara Souza Britto Lobão¹

José Antonio Lacerda Lobão²

Washington de Jesus Sant'anna da França-Rocha³

^{1 3} Universidade Estadual de Feira de Santana – UEFS

GP-SIG (Grupo de Pesquisas em Sistemas de Informações Geo-referenciadas) – UEFS

Br 116, Km 3 – Campus Universitário

wrocha@terra.com.br¹

juci.lobao@ig.com.br²

² Consultor Independente

Rua Alberto Campos Boaventura Nº 24, Bairro Brasília - FS-BA

Jose.lobao@uol.com.br

Abstract. This work describe the development of a database about the Semi-arid region on the Bahia State, with emphasis to the Caatinga Biome. Through the integration and combination of spatial data on an GIS environment with non-spatial biodiversity, social, environmental and economical data, varied nature informations (satellite images, vectorial files, aerial photos, photographic, bibliographical, textual, documental, historical data, researchers and Institutions cadastres, etc.) obtained on several Research and Teaching Institutions, public and private organs, NPOs, among other, were systematized in a friendly and accessible environment. The database was designed and implemented using MySQL DBMS integrated with a interface programming developed on Delphi language.

Palavras-chave: Geography Information, System, database, biodiversity, Delphi, MYSQL, Sistema de Informações Geográficas, banco de dados, biodiversidade.

1- Introdução

Bancos de dados espaciais vêm ganhando muita atenção pela sua capacidade em combinar dados não-espaciais relacionais com dados cartográficos georreferenciados, ampliando significativamente a possibilidade de agregar novas informações e, principalmente, de conectá-los (Lisboa Filho, 2002). A integração de dados em um Sistema de Informações Georreferenciadas (SIG) é de fundamental importância para estudos ambientais, devido à possibilidade de agregar dados de biodiversidade, dados sociais, econômicos, políticos e culturais, potencializando a capacidade de análise.

Os Bancos de Dados convencionais sobre biodiversidade agregam dados alfanuméricos, alguns deles com uma grande quantidade de campos para contemplar as especificidades que envolvem a catalogação de espécies, porém, estão desassociados do ambiente físico, e desconsideram os dados socioambientais, históricos, econômicos e culturais. Além disso, não incluem dados analíticos, mapas, e muito menos georreferenciamentos. Como exemplo pode-se citar o Sistema ALICE, usado pela Universidade Federal de Pernambuco – UFPE, Checklist das Plantas do Nordeste <<http://umbuzeiro.cnip.org.br/db/pnechk/geog.html>>; o Sistema de gerenciamento de herbário da Universidade Estadual de Feira de Santana – UEFS ou ainda uma estrutura mais incrementada como o SinBiota, (<http://sinbiota.cria.org.br>) que flexibiliza a recuperação de cada coleção por meio do software spLinker, incorporado num sistema online, mas também não agrega os dados ambientais, econômicos e sociais.

Estas e outras alternativas apesar de se constituírem em excelentes bancos de dados estão limitadas a objetivos específicos a que se destinam.

Dentro de uma perspectiva de sistematizar em um mesmo ambiente, dados de natureza bastante variada (espaciais, textuais, tabulares, fotográficos etc) é que se estruturou o Projeto

da Base Biorregional Caatinga Semi-Árido, executado pela Universidade Estadual de Feira de Santana - UEFS e o Centro de Recursos Ambientais - CRA, como parte de um programa de Bases Biorregionais no Estado da Bahia. Este banco de dados integra dados ambientais, econômicos e sociais, por meio de SIG, incorporando também dados convencionais.

A partir de iniciativas de instituições de ensino, pesquisa, de órgãos governamentais e ONGs, diversos estudos foram realizados, passando-se a descobrir a importância e complexidade desse ecossistema. Entretanto, estas pesquisas encontram-se dispersas e fragmentadas nos diversos órgãos e instituições. Percebeu-se então a necessidade de organizar a grande variedade de dados de fontes e naturezas também diversas. Assim, os dados e informações coletadas sobre a região, nestes órgãos e instituições, vêm a suprir a necessidade, cada vez mais proeminente, sobre o conhecimento da diversidade da caatinga e fomentar novas pesquisas.

A área de estudo compreende a região do semi-árido na Bahia, uma das mais pobres e desconhecidas do estado em termos de biodiversidade, onde está localizado o Bioma Caatinga (**Figura 1**).

Fonte: SEI, 2003; IBGE, 2003; Elaborado por Lobão, 2004

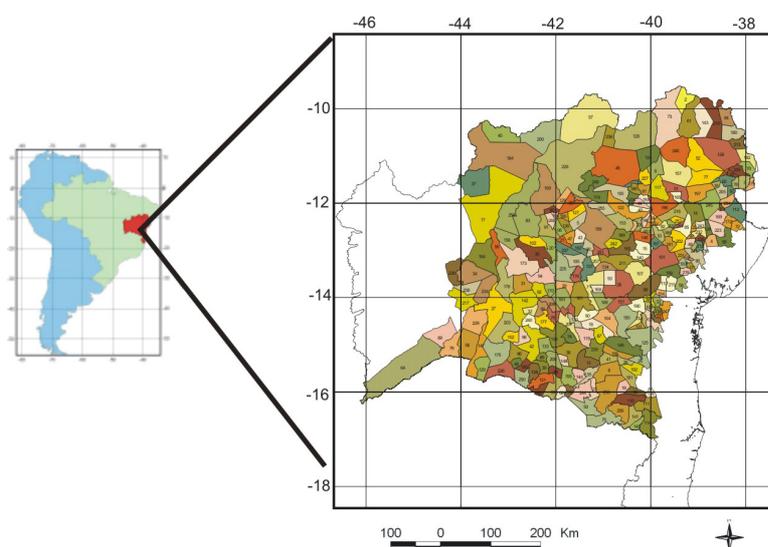


Figura 1 – Localização da Região Semi-Árida da Bahia e limites dos municípios do Bioma Caatinga

3- Projeto do Banco de Dados

O Banco de Dados construído representa aspectos do Mundo Real e é um repositório de onde pode-se armazenar e extrair uma vasta gama de informações derivadas, que possui um nível de interação com eventos que representará. Assim para sua concretização é necessário incorporar informações inter-relacionadas, dispostas numa ordem pré-determinada em função de um projeto de sistema, para o propósito definido.

Para determinar a área de abrangência dos dados a serem integrados, um amplo levantamento foi realizado a fim de discutir os limites físicos do Semi-Árido e do Bioma Caatinga (LOBÃO, 2004), com o objetivo de fundamentar teóricamente e conceitualmente as propostas de delimitação geográfica. Embora seja clara a diferença entre limites físicos e políticos (divisão municipal), a adoção do limite oficial decretado pela Superintendência de Desenvolvimento do Nordeste – SUDENE foi adotada por contemplar maior disponibilidade de dados sócio-econômico sem inviabilizar seus aspectos físicos e ambientais. Com base em

sua delimitação, iniciou-se a análise e organização dos dados existentes e possíveis de serem produzidos.

É comum planejar um banco de dados em três etapas distintas: *modelagem conceitual*, *modelagem lógica* e *modelagem física* (ELMASRI E NAVATHE, 1994). (**Figura 2**)

Figura 2 – Modelagem do Banco de Dados



3.1- Modelagem conceitual

A criação do Banco de Dados Caatinga/Semi-Árido foi realizada inicialmente com um amplo levantamento em várias instituições, bibliotecas e na Internet, catalogando os diferentes tipos de dados disponíveis. Para definição dos descritores dos dados de biodiversidade foram também consultados os bancos de dados globais de intercâmbio de dados de fauna e flora. Foram utilizados modelos entidade-associação para representação das relações entre as diferentes entidades e respectivos atributos descritores.

Visando facilitar o entendimento entre usuários e construtores do Banco de Dados foram empregadas linguagens simples, com base no modelo ER(entidade relacionamento) (Chen, 1976).

O modelo conceitual foi construído de forma incremental, considerando-se que a modelagem é um processo de aprendizagem dos fatos referentes ao sistema que se quer implementar. A estratégia de construção do modelo partiu da necessidade de armazenamento, onde identificou-se que as principais fontes de informações eram os dados existentes e o conhecimento que as pessoas possuíam sobre o sistema.

Uma dessas fontes de informações - dados existentes, eram armazenados de diversas formas alternativas, inclusive em meio digital no formato de planilhas e documentos de texto, além de anotações físicas e fichários. A outra, foi o conhecimento que pessoas envolvidas no processo de pesquisas possuíam, necessitando serem sistematizados. Mas, o desafio maior era integrar tudo isso a um SIG, que estava sendo construído em paralelo.

Para o Banco de dados da Base Biorregional Caatinga Semi-Árido, existia a necessidade de controle das informações referentes aos diversos projetos existentes e que poderão vir a existir: pesquisadores e instituições a que estão vinculados, municípios que fazem parte dos projetos, bibliografia utilizada nas diversas ações implementadas e, principalmente, os dados de pesquisa produzidos. Entre os dados que compõem a Base Caatinga estão o cadastro de projetos de pesquisas, pesquisadores, check list das espécies vegetais e animais, dados bibliográficos, textuais, documentais, fotográficos, georreferenciados em ambiente de Sistema de Informações Geográficas- SIG e metadados.

Combinando estas duas fontes de informações, iniciou-se pela enumeração das entidades, em seguida os relacionamentos, a cardinalidade, a determinação dos identificadores e a verificação quanto ao aspecto temporal. No processo de detalhamento, completou-se o modelo e definiram-se as restrições de integridade. (LISBOA Filho, 2002; PARENT, 1998). Na validação, procurou-se identificar e eliminar as construções redundantes (HEUSER, 2004).

A **figura 3** exemplifica a modelagem conceitual, mostrando a arquitetura onde qualquer entidade (Pesquisador, município, projetos, fontes, dados, SIG e bibliografia) pode estar vinculada a uma outra. A vinculação mestre possui cardinalidade mínima igual a “0” e máxima igual a “1”; e, a detalhe, possui cardinalidade mínima igual a “1” e máxima igual a “n”. Esta estratégia de relacionamento, visou da flexibilidade em função da grande variedade de origem e natureza dos dados, com o cuidado de implementar mecanismos que evitam redundância.

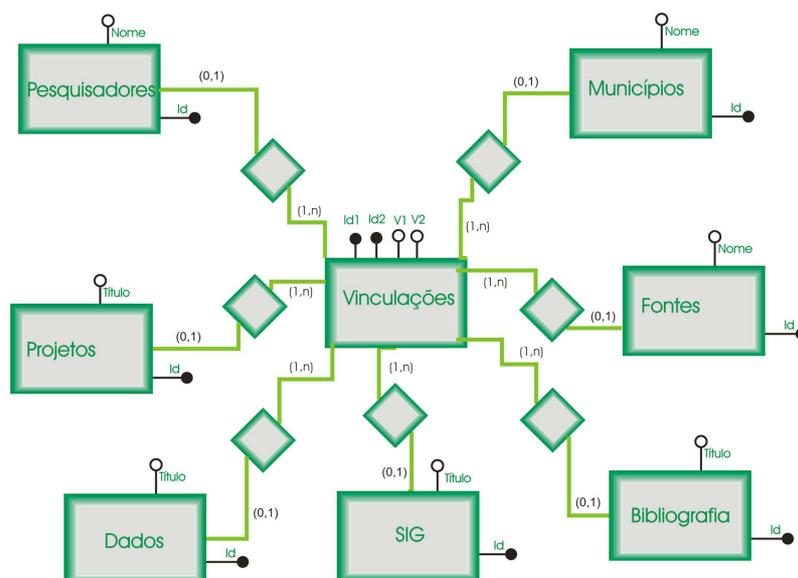


Figura 3 – Modelagem conceitual do Banco de Dados da Base Biorregional Caatinga Semi-Árido

3.2- Modelagem Lógica

A partir do modelo conceitual iniciou-se a modelagem lógica que foi projetada aplicando-se regras de transformações dos resultados utilizados no esquema conceitual ER, em elementos de representação de dados de um modelo SGDB relacional. Neste passo, cada entidade foi traduzida em uma tabela e cada atributo da entidade definiu uma coluna da tabela.

Visando o refinamento do modelo, uma vez que o projeto do banco de dados não pode ser visto como final, buscou-se algumas alternativas que possibilitaram uma melhor performance do sistema, mantendo o compromisso entre o ideal representado pelas regras de implementação e o alcançável frente às limitações de performance.

No modelo lógico, a transformação do relacionamento entre as entidades foi traduzido na tabela “Vinculações”, formada pelas colunas correspondentes aos identificadores das entidades relacionadas e pelas colunas correspondentes aos atributos identificadores do relacionamento, que foi determinada pelas cardinalidades mínima e máxima das entidades envolvidas.

A tabela vinculações é a única forma de relacionamento entre as demais, sendo que para cada entidade envolvida a sua participação é opcional (Cardinalidade (0,1) (Figura 3). Por isso, foi previsto um mecanismo de restrição pela verificação de relacionamento recursivo, para evitar a redundância. Tal mecanismo poderá ser implementado no SGDB com facilidade, utilizando-se como vantagem a característica da implementação em tabela única para as vinculações.

O esquema abaixo exemplifica a conversão ER para relacional da entidade participantes.

Participantes (Id, Nome, Endereço, Cidade, UF, telefones, e-mail, site, formação, atuação)

3.3- Modelagem Física

O projeto físico foi concebido a partir do modelo lógico, para o SGDB (*Sistema Gerenciador de Banco de Dados*) MySQL, por atender plenamente às necessidades concebidas.

O MySQL é um servidor robusto de banco de dados, sendo desenvolvido e aperfeiçoado desde o início dos anos 80. Este SGDB é muito rápido, possui multi-tarefa e é multi-usuário. Pode ser usado sob licença Open Souce/Free Software nos termos da GNU General Public License. Foi escrito em linguagem C++ para diversas plataformas, sendo as principais Windows e Linux. Possui suporte total a multi-threads usando threads diretamente no kernel, o que significa que pode utilizar facilmente múltiplas CPUs. Fornece mecanismo de armazenamento transacional e não transacional, com tamanho das tabelas podendo chegar a oito milhões de terabytes, praticamente limitado apenas pelo sistema operacional. Trabalha com senhas criptografadas e sistema de privilégios muito flexível e seguro, que permite a verificação baseada em estações/máquinas.

Os fabricantes do MySQL já testaram bancos com mais de 50.000.000 de registros, existindo usuários que mantêm servidor com 60.000 tabelas e aproximadamente 5.000.000.000 de linhas. Os clientes podem conectar-se ao servidor utilizando-se da rede, através de sockets TCP/IP, em qualquer das plataformas implementadas.

O MySQL possui extensões espaciais que permitem gerar, armazenar e analisar recursos geográficos que estão disponíveis através das tabelas MyISAM. Este modelo segue as especificações do OpenGIS (*Simple Feature Specifications For SQL*) Consortium (OGC) para estender suporte a dados espaciais. Assim, o MySQL implementa um subconjunto do ambiente SQL com Tipos Geométricos, onde cada objeto é associado com um Sistema de Referência Espacial e pertence a uma classe geométrica. As classes são implementadas numa hierarquia.

Os dados geoespaciais, por terem sido adquiridos em fontes variadas e em escalas têmporo-espaciais diferenciadas, necessitaram de tratamentos específicos: conversões e espacializações adequadas para que possa ser integrado a um SGBD.

Na modelagem física, foram definidos os aspectos de implementação do banco de dados (estruturas de armazenamento, caminhos de acesso, particionamento e agrupamento de dados). Esta modelagem está diretamente, relacionada ao SGBD MySQL e a linguagem de programação empregada para a interface do banco de dados. Foi escolhido o Delphi, por ser uma linguagem de alto nível, orientada a objeto, visando criar uma interface amigável e segura, facilitando sua manutenção e consulta (Figura 4).

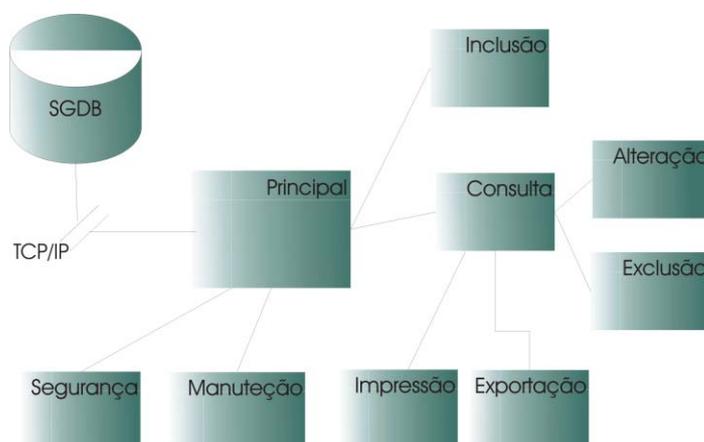


Figura 4 – Modelagem Física da Interface do Banco de Dados

A ferramenta de programação visual Delphi, versão 7.0, que serviu para criar uma interface amigável com o MySQL é baseada em ambiente integrado de desenvolvimento, orientada a objetos e amparada por uma vasta biblioteca de componentes, que agregam recursos das mais recentes tecnologias, oferecendo suporte às linguagens XML e HTML, possibilitando o uso da *Unified Modeling Language* – UML, que é uma linguagem de modelagem universal para sistemas orientados a objeto (MATOS, 2002).

Todos esses recursos favoreceram a abstração ao usuário em relação à camada de dados, poupando-o de detalhes de interesse apenas do administrador do sistema, sem abrir mão dos recursos oferecidos pelo MySQL (CANTÙ, 2003).

A modelagem física permitiu planejar aspectos ligados à eficiência do sistema, averiguando-se a consistência, segurança, assim como, a simplicidade no processo de administração, manutenção e saída de dados. O usuário interage com o sistema, por meio de funções de Inclusão, Alteração, Exclusão, Consulta e Exportação de dados espaciais e não-espaciais (**Figura 4**).

4. A implementação do Projeto

No projeto implementado o SIG, foi o foco principal sendo composto de dados gerais que englobam toda a região de estudo e dados específicos de determinadas áreas do semi-árido. Entre os dados gerais pode-se destacar os processamentos de imagens Landsat; Spot; Modelagem com Modelo Digital de Terreno (declividade, sombreamento, aspecto, hipsometria e convexidade), classificação de indicadores sociais; econômicos e de desenvolvimento, dentre outros, para toda a região (**Figura 5**).

Os dados específicos são provenientes de projetos mais detalhados para algumas áreas do semi-árido, como a Região de Feira de Santana, Chapada Diamantina, Raso da Catarina e principalmente todas as áreas de Proteção ambiental e em particular as sete APAS, localizadas na área de estudo (Pedra do Cavalo, Gruta dos Brejões Vereda do Romão Gramacho, Dunas e Veredas do Baixo Médio São Francisco, Marimbus Iraquara, Lagoa de Itaparica, Serra do Barbado e Serra Branca) (**Figura 5**). Assim, uma vasta gama de dados espaciais de áreas específicas resultantes de outros trabalhos, como: Zoneamento ecológico econômico do Nordeste (CPATSA- EMBRAPA); Dados vetoriais do Projeto de Biodiversidade da Caatinga (TNC/APNE); Projeto de Desenhadores Rupestres (NOLASCO, 2004); dados do Projeto de Biodiversidade – PROBIO, 2004, entre outros, foram incorporados.

Todos esses dados espaciais e um significativo acervo de dados não espaciais podem ser acessados por telas do sistema que obedecem ao padrão do sistema operacional Windows da Microsoft, onde a interface gráfica foi projetada para tornar as operações auto-explicativas. (**Figura 5**).

A segurança de acesso aos dados, pode ser concedida em diversos níveis às funções e/ou tabelas específicas do banco, ou ainda, a todo o conjunto de dados. A permissão para incluir, alterar, excluir e consultar é liberada de forma diferenciada em nível de usuário. Os administradores, além das funções especificadas acima, podem também cadastrar usuários e conceder privilégios (**Figura 5**). Para todos os dados incluídos, alterados e/ou consultados é criado automaticamente um arquivo “.log” contendo informações do usuário, data e hora de acesso.

O backup dos dados pode ser realizado em três formatos: completo, incremental e programado.

Considerando a grande variabilidade de formatos, origens e natureza dos dados a serem incorporados foi necessário criar um software adicional que pudesse incorporar dados pré-existent sem a necessidade de uma digitação individualizada. Este software cria e importa

planilhas com campos do tipo: texto; memo; números; datas; lógicos (Sim/Não); foto. Os campos tipo foto podem visualizados na própria planilha ou em janelas diferenciadas (**Figura 5**).

A interface com o usuário ou programas clientes funcionam em rede por meio do protocolo TCP/IP, no modelo cliente/servidor. As consultas são enviadas ao servidor em SQL (*Structured Query Language*), que as processa e retorna apenas os dados solicitados, deixando a carga maior de processamento para o servidor. Mesmo quando as consultas resultam em grande quantidade de linhas, apenas o suficiente para ser apresentado em uma tela transitada pela rede e à medida que o usuário muda de tela, as demais linhas são requisitadas automaticamente, evitando transmissão de dados desnecessários. Quando são efetuadas alterações nos dados cadastrados o cliente recebe um registro completo, mas, apenas os campos que são efetivamente alterados é que retornam ao servidor.

As opções de consultas são baseadas principalmente em palavras chave, definidas no momento da entrada do dado específico, possibilitando a busca em diversos campos das tabelas. O usuário pode escolher entre várias opções de refinamento, para atender a sua necessidade de detalhamento. As respostas às consultas, dependendo da opção e do nível de refinamento escolhido, podem apresentar links para outros níveis de consulta. A forma de busca de dados pode ser feita nas próprias telas de manutenção dos dados. Além desse ambiente, o sistema gera código HTML de forma dinâmica por meio do protocolo HTTP propiciado pelo servidor embutido no programa, abrindo automaticamente o navegador internet padrão instalado (**Figura 5**).

Os dados podem ser visualizados de formas diferenciadas: por meio do software específico desenvolvido como um módulo dentro do próprio sistema; ou, por meio de softwares adicionais instalados na própria máquina do usuário. Esta filosofia empregada vai ao encontro do princípio de flexibilização que norteou a concepção do banco em todas as etapas. O resultado das consultas podem ser impressos ou exportados para os formatos (pdf Documento Adobe Acrobat; html, xls; csv e txt). Para os dados consultados podem ser gerados relatórios das buscas realizadas. Os dados também podem ser impressos diretamente por meio do sistema desenvolvido.

Figura 5 – SIG e Telas do Sistema desenvolvido

INTERFACE GRÁFICA



5. Conclusões

O banco de dados da Base Biorregional Caatinga Semi-Árido foi projetado com a meta de integrar dados e informações espaciais geométricos e convencionais e culminou na construção de um produto robusto, com uma interface de fácil acesso, contendo informações cartográficas georreferenciadas, compondo uma base física, social, econômica e biótica da Região Semi-Árida da Bahia. Sua implementação possibilitou consultas rápidas e seguras, facilitando novas pesquisas e aplicações. Isto foi cumprido com a utilização do SGDB MySQL e da linguagem Delphi, que propiciaram versatilidade e segurança. Apesar da grande quantidade de recursos implementados, a acessibilidade, ou melhor, a interface com o usuário final é muito amigável.

Ao incorporar uma imensa variedade de dados de origem e naturezas diversas, (dados georreferenciados, textos, documentos históricos, fotos aéreas, imagens de satélite, fotografias, tabelas, etc), com formatos variados, foi de fundamental importância que se restringisse o acesso indiscriminado a todos os dados preservando assim a integridade, confiabilidade e a propriedade intelectual dos pesquisadores. Logo, é a flexibilidade para incorporação de dados e a possibilidade de sua posterior divulgação num ambiente virtual, bem como sua segurança em nível de usuário que se constitui no grande diferencial desta implementação. Outro fato importante na concepção deste banco é a escalabilidade, ou seja, a possibilidade de desenvolvimento de novas funções, vinculações e novas demandas.

A visão e estrutura deste Banco de Dados vêm ao encontro da tendência de criação de bases de dados globais de biodiversidade, associados a dados físicos, sociais, econômicos, em ambiente SIG, numa visão holística com acesso remoto, possibilitando o estabelecimento de uma rede de especialistas em várias áreas do conhecimento.

6. Agradecimentos

Os autores agradecem ao CRA – Centro de Recursos Ambientais do Estado da Bahia pelo suporte financeiro para o desenvolvimento deste sistema.

7. Referências

- Cantú, M. **Dominando o Delphi 7 “A Bíblia”**. São Paulo: Makron Books, 2003
- Chen, P. P. S. (1976) **The entity-relationship model: Towards a unified view of data**. ACM Trans. Database System, New York, n.1.
- Dias, A. de S. **Delphi & Mysql**. Rio de Janeiro: Ciência Moderna Ltda, 2000
- Elmasri, R.; Navathe, S. B. (1994) **Fundamentals of Database Systems**. 2.ed. Menlo Park, CA: Addison-Wesley.
- Heuser, C. A. **Projeto de Banco de Dados**. Série Livros Didáticos n 4. Rio Grande do Sul: Sagra Luzzatto, 5ª ed., 2004
- Lisboa Filho, Jugurta et al. - **Modelagem Conceitual de Banco de Dados Geográficos: o estudo de caso do Projeto PADCT/CIAMB**. Porto Alegre: UFRGS, 2002 In: <http://www.ecologia.ufrgs.br/labgeo/artigos/jugurta.pdf> Acessado em: 02/03/2004
- Lobão, J. S. B. et al. **Semi-Árido da Bahia: A Delimitação da SUDENE**. In: II Feira do Semi-Árido. Universidade Estadual de Feira de Santana, 2004, Feira de Santana BA: UEFS, 2004 Seção de Comunicação.
- Matos, A. V. **Unified Modeling Language UML: Prático e Descomplicado**. São Paulo: Érica, 2002
- Parent, C.; Spaccapietra, S.; Zimanyi, E.; Domini, P.; Plazanet, C.; Vangenot, C. (1998) **Modeling spatial data in the MADS conceptual model**. In: **Proceedings of the International Symposium on Spatial Data Handling**. Vancouver, Canada.
- MySQL AB **“MySQL Reference Manual”**..(1999-2003)