

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

**Projeto Conceitual de Banco de Dados Geográficos
através da Reutilização de Esquemas, utilizando
Padrões de Análise e um *Framework* Conceitual**

por

Jugurta Lisboa Filho

Tese submetida à avaliação,
como requisito parcial para a obtenção do grau de
Doutor em Ciência da Computação

Prof. Dr. Cirano Iochpe
Orientador

Porto Alegre, janeiro de 2000.

CIP - Catalogação na Publicação

Lisboa Filho, Jugurta

Projeto Conceitual de Banco de Dados Geográficos através da Reutilização de Esquemas, utilizando Padrões de Análise e um *Framework* Conceitual / Jugurta Lisboa Filho. - Porto Alegre: PPGC da UFRGS, 2000.

212 f.: il.

Tese (doutorado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR-RS, 2000. Orientador: Iochpe, Cirano.

1. Sistemas de Banco de Dados 2. Modelos Conceituais de Dados 3. Sistemas de Informação Geográfica. 4. Reutilização. 5. Padrões de Análise. 6. Framework. I. Iochpe, Cirano. II. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitora: Prof^a Wrana Panizzi

Pró-Reitor de Pós-Graduação: Prof. Franz Rainer Semmelmann

Diretor do Instituto de Informática: Prof. Philippe O. A. Navaux

Coordenador do PPGC: Prof^a Carla Maria Dal Sasso Freitas

Bibliotecária-chefe do Instituto de Informática: Beatriz Haro

*Dedico esta tese aos meus pais
Jugurta e Diva*

*à minha amada
Janaina*

*e aos meus filhos
Pedro (14)
Tomas (10)
Maria (8)
Francisco (3)*

Agradecimentos

Um trabalho de quase cinco anos só é possível com a ajuda de muitas pessoas. Agradeço a todos aqueles que participaram, direta ou indiretamente, desta minha longa jornada. Citando alguns nomes, certamente estarei cometendo algumas injustiças, mas por outro lado, não poderia deixar de mencionar os seguintes agradecimentos.

Inicialmente, agradeço a meu orientador Prof. Cirano Iochpe por todo o apoio, orientação, oportunidades e, principalmente, pela sua motivação e coragem para enfrentarmos juntos os desafios de um novo e fascinante campo de pesquisa.

Devo um agradecimento especial a alguns colegas que me ajudaram a entender um pouco sobre a área de geoprocessamento e seus sistemas de informação geográfica (SIG). Gostaria de agradecer ao Prof. Heinrich Hasenack (Centro de Ecologia/UFRGS) pela valiosa e fundamental ajuda, tanto no esclarecimento das minhas inúmeras dúvidas sobre o funcionamento dos SIG, como fornecedor de vários exemplos de aplicações que serviram de base para as pesquisas desenvolvidas nesta tese. Também a Milton Ramires (UFRJ), Íria Garaffa (UFRGS e Pró-Guaíba) e Fátima Pires (Unicamp) pela valiosa ajuda nos primeiros momentos do meu doutorado. Ao Eliseu Weber (Centro de Ecologia/UFRGS) pela constante troca de idéias sobre SIG e pelo chimarrão, sempre disponível no Laboratório do Centro de Recursos Idrisi. À minha amiga Karla Borges (Prodabel-BH), por me ajudar a aprofundar os meus conhecimentos sobre modelagem conceitual de banco de dados geográficos.

Obrigado a toda a comunidade do Instituto de Informática da UFRGS. A todos os professores que de alguma forma contribuíram para a minha formação, seja através de aulas ou avaliações seja através de uma simples conversa ou ajuda mas, principalmente, pela agradável convivência durante todos esses anos. Agradeço à Prof^a. Carla Maria Dal Sasso pela presteza e simpatia com que coordena o PPGC. Aos funcionários, muito obrigado pelo apoio e companheirismo. Agradeço em especial aos funcionários Luis Otávio, Silvania, Maria do Carmo, Lourdinha, Eliane, Beatriz, Ida e à Tia Maria que tanto me ajudou a ficar acordado com seu cafezinho.

Durante esses anos, reparti com vários colegas, aqui da UFRGS, momentos de muita alegria, mas também alguns momentos de aperto. No entanto, em todos esses momentos nunca estive sozinho. Em especial, agradeço a ajuda e a força dos amigos Marcelo Ladeira, Lúcia Giraffa, Ricardo Bastos, Ricardo Silveira, Beatriz Franciosi, Ricardo Custódio, JB e Hércules.

Agradeço aos colegas que participaram comigo em vários projetos de pesquisa, inclusive sendo co-autores de algumas das publicações inseridas nesta tese. Em especial, gostaria de agradecer à Íria Garaffa, Andréia Costa e Roni Anzolch (Projeto SIGMODA) e, novamente, aos colegas Prof. Heinrich Hasenack e Eliseu Weber (Centro de Ecologia/UFRGS). Agradeço ao acadêmico Willian Wazlawik pela confecção dos caracteres especiais da fonte Sigmoda.

Tive a grande oportunidade de passar três meses como *Visiting Scholar* no NCGIA da Universidade do Maine, USA, onde decidi pelo meu tema de tese. Para lá, gostaria de enviar os meus agradecimentos à minha orientadora Prof^a. Kate Beard, ao *my friend* João Paiva (INPE) e a alguns colegas que tornaram mais agradável a minha estada naquele país, entre eles Panayotis (Panos), Sotirios, Thomas, Andrea, Andreas, Diana e Suelaine.

Agradeço aos colegas do Departamento de Informática da Universidade Federal de Viçosa (UFV), por carregarem o piano durante os cinco anos de minha ausência e por terem lutado, e conseguido, criar o nosso curso de mestrado, para o qual espero poder contribuir ativamente a partir de agora. Um agradecimento, em especial, aos colegas Luiz Raggi (meu procurador) e Luiz Albuquerque (Chefe do DPI), pelos “pepinos” resolvidos junto à UFV, e aos colegas Cacau, Carlos e Zeluís pelas mensagens freqüentes, o que me manteve eletronicamente próximo da UFV durante todo esse tempo.

Não existe uma forma escrita capaz de traduzir o meu agradecimento à minha esposa Janaina e aos meus filhos Pedro, Tomas, Maria e Francisco. Apenas gostaria de deixar registrado aqui, que todos os momentos em que estive, ora sem paciência ora ausente de casa, foram para o bem deles. Inclusive hoje, neste lindo domingo de sol e véspera de feriado (14/11/99), em que tive de abandoná-los em casa para trabalhar na conclusão desta tese.

Agradeço o carinho e apoio de todo os meus familiares, em especial aos meus pais Jugurta e Diva, e aos meus irmãos Magda, Gizelda, Mirna e Antônio.

Devo um agradecimento ao povo brasileiro que, mesmo sem saber, financiou todo o meu trabalho, seja através do meu salário, pago pela UFV, seja através das bolsas de estudo dos programas CAPES/PICDT e RHAEC/CNPq. Agradeço, também, o apoio financeiro do Instituto de Informática da UFRGS que custeou algumas das minhas viagens para apresentação de artigos em congressos.

Por fim, agradeço à Deus por tudo.

J.L.F.

Sumário

Lista de Figuras.....	9
Lista de Tabelas	13
Resumo	15
Abstract.....	17
1 Introdução.....	19
1.1 Motivação	19
1.2 Hipótese	22
1.3 Objetivos.....	22
1.4 Estrutura da tese.....	23
2 Revisão Bibliográfica.....	27
2.1 Introdução	27
2.2 Artigo I: Um Estudo sobre Modelos Conceituais de Dados para Projeto de Bancos de Dados Geográficos.....	28
2.3 Artigo II: Mecanismos de Reutilização em Sistemas de Informação.....	42
3 Modelagem Conceitual de Banco de Dados Geográficos	55
3.1 Introdução	55
3.2 Artigo III: Adaptando o Modelo de Objetos OMT para Modelagem Conceitual de Aplicações de SIG	57
3.3 Artigo IV: Metodologia para Modelagem Conceitual de Dados de um SIG Interinstitucional	64
3.4 Artigo V: Modelos Conceituais de Dados para Aplicações Geográficas: uma experiência com um SIG interinstitucional.....	69
3.5 Artigo VI: BCH/GIS - A Basic Class Hierarchy to Support GIS Conceptual Design.....	80
4 Instrumentos de Reutilização	89
4.1 Introdução	89
4.2 Artigo VII: Applying Analysis Patterns in the GIS Domain.....	91
4.3 Artigo VIII: Specifying Analysis Patterns for Geographic Databases on the basis of a Conceptual Framework.....	100
4.4 Artigo IX: Padrões de Análise para Banco de Dados Geográficos	113
4.5 Artigo X: Modelagem Conceitual de Banco de Dados Geográficos: o estudo de caso do Projeto PADCT/CIAMB	128
4.6 Artigo XI: Projeto de Banco de Dados Geográficos: mapeando esquemas GeoFrame para o SIG <i>Spring</i>	143
5 Conclusões.....	151
5.1 Resumo das Contribuições da Tese	155
5.2 Trabalhos Futuros	156

Anexo 1.....	159
Artigo XII: Geospatial Metadata in Brazil: an experience in data documentation of an environmental GIS application	159
Anexo 2.....	167
Artigo XIII: Um servidor de consultas à <i>clearinghouse</i> em português para distribuição de metadados geográficos.....	167
Anexo 3.....	191
Artigo XIV: Requisitos especiais e modelos conceituais de dados para aplicações de sistemas de informações geográficas	191
Anexo 4.....	195
Outros trabalhos publicados durante a tese.....	195
Anexo 5.....	197
Fonte de caracteres Sigmoda.....	197
Bibliografia	199

Lista de Figuras

Artigo I

FIGURA 1 - Processo de modelagem conceitual.....	31
FIGURA 2 - Evolução cronológica dos modelos conceituais de dados para SIG	31
FIGURA 3 - Diferenciando fenômenos geográficos e objetos convencionais	35
FIGURA 4 - Um exemplo de modelagem no modelo Geo-IFO	36
FIGURA 5 - Modelando temas segundo o modelo UML-GeoFrame.....	37
FIGURA 6 - Exemplo de modelagem de aspectos temporais no modelo GeoOOA	40

Artigo II

FIGURA 1 - Processo de reutilização de software	44
FIGURA 2 - Padrão de análise LocalizaçõesGeográficas	48
FIGURA 3 - Padrão de projeto Composite	49
FIGURA 4 - Exemplo de uso do padrão de projeto <i>Composite</i>	50

Artigo III

FIGURA 1 - Modelos de dados na visão de campo [LIS 96a]	59
FIGURA 2 - Hierarquia de classes básicas	60
FIGURA 3 - Substituição simbólica de esquemas	61
FIGURA 4 - Modelagem de dados na visão de campo.....	62
FIGURA 5 - Modelagem de dados na visão de objetos.....	63

Artigo IV

FIGURA 1 - Biblioteca de Classes Básicas	67
--	----

Artigo V

FIGURA 1 - Módulos do modelo Modul-R [CAR 93].....	72
FIGURA 2 - Fases do Projeto de Banco de Dados Geográfico	76
FIGURA 3 - Exemplo do esquema conceitual do SIGPROGB.....	78
FIGURA 4 - Exemplo do esquema de representação espacial do SIGPROGB.....	78

Artigo VI

FIGURE 1 - BCH/GIS-v.1 – Version One.....	83
FIGURE 2 - BCH/GIS-v.2 – Version Two.....	84
FIGURE 3 - BCH/GIS-v.3 – Current Version	85
FIGURE 4 - Example of the SIGPROGB spatial representation schema.....	86

Artigo VII

FIGURE 1 - Geographic Locations Analysis Pattern [HAY 95].....	95
FIGURE 2 - Environmental Quality Parameters Pattern	97
FIGURE 3 - Example of modeling water quality parameter	98

Artigo VIII

FIGURE 1 - GeoFrame classes diagram	103
FIGURE 2 - Example of <i>relateWith</i> relationships	104
FIGURE 3 - An example of theme diagram	108
FIGURE 4 - Stereotypes for generalization	108
FIGURE 5 - Stereotypes for <i>represent</i> association.....	109
FIGURE 6 - Example of class diagram using GeoFrame	109
FIGURE 7 - Urban Way System analysis pattern.....	110
FIGURE 8 - Example of use of the analysis pattern.....	111

Artigo IX

FIGURA 1 – Processo de captura e aplicação de padrões	115
FIGURA 2 – Notação gráfica do diagrama de classes UML (resumida).....	117
FIGURA 3 – Estereótipos do <i>framework</i> GeoFrame	117
FIGURA 4 – Diagrama de classes do padrão “Parâmetros de Controle Ambiental” ..	119
FIGURA 5 – Exemplo de uso do padrão “Parâmetros de Controle Ambiental”	119
FIGURA 6 – Diagrama de classes do padrão “Campo Geográfico Categórico”	120
FIGURA 7 – Exemplo de uso do padrão “Campo Geográfico Categórico”	120
FIGURA 8 – Diagrama de classes do padrão “Campo Geográfico Numérico”	121
FIGURA 9 – Exemplos de uso do padrão “Campo Geográfico Numérico”	121
FIGURA 10 – Diagrama de classes do padrão “Hidrografia”	122
FIGURA 11 – Exemplo de uso do padrão “Hidrografia”	122
FIGURA 12 – Diagrama de classes do padrão “Malha Viária Urbana”	123
FIGURA 13 – Diagrama de classes do padrão “Rede de Circulação Viária Urbana” ..	124
FIGURA 14 – Exemplo de uso dos padrões “Malha Viária Urbana” e “Rede de Circulação Viária Urbana”	125
FIGURA 15 – Diagrama de classes do padrão “Cadastro de Lotes Urbanos”	126
FIGURA 16 – Exemplos de uso do padrão “Cadastro de Lotes Urbanos”	126

Artigo X

FIGURA 1 - Processo de modelagem conceitual.....	129
FIGURA 2 - Notação gráfica do diagrama de classes UML (resumido)	131
FIGURA 3 - Diagrama de Classes do GeoFrame	132
FIGURA 4 - Estereótipos para generalização	135
FIGURA 5 - Estereótipos para associação	136
FIGURA 6 - Região Carbonífera do RS, Micro-região São Jerônimo	137
FIGURA 7 - Diagrama de temas.....	138
FIGURA 8 - Tema Meio Antrópico.....	139
FIGURA 9 - Tema Meio Biótico e Abiótico	141

Artigo XI

FIGURA 1 - Diagrama de Classes do GeoFrame	144
FIGURA 2 - Estereótipos para generalização	145
FIGURA 3 - Estereótipos para associação	146
FIGURA 4 - Nível conceitual do <i>Spring</i>	147
FIGURA 5 - Exemplo de mapeamento GeoFrame- <i>Spring</i>	149

Artigo XII

FIGURE 1 -The Caí watershed in the Rio Grande do Sul State and in the Pró-Guaíba program area.....	163
---	-----

Artigo XIII

FIGURA 2.1 – Topologia do Sistema.....	170
FIGURA 2.2 - Passos de uma consulta à Clearinghouse	171
FIGURA 2.3 – Exemplo de consulta Isearch.....	173
FIGURA 2.4 – Servidor Zserver ativo	174
FIGURA 2.5 – Resultado de uma consulta Zclient.....	175
FIGURA 2.6 – Página de consultas MetaStar Gateway do FGDC.....	176
FIGURA 2.7 – Visualização de um conjunto de metadados selecionado a partir de uma consulta	176
FIGURA 3.1 – Topologia do Sistema <i>Clearinghouse</i> com um tradutor para o português instalado na máquina do usuário	178
FIGURA 3.2 – Topologia do Sistema <i>Clearinghouse</i> com um pseudo servidor de consultas com um tradutor para o português.....	179

FIGURA 3.3 – Topologia do Sistema <i>Clearinghouse</i> com um servidor de consultas e um tradutor para o português	180
FIGURA 4.1 – Servidor de consultas e tradutor na máquina servidora.....	181
FIGURA 4.2 – Servidor de consultas e tradutor na máquina servidora vistos de um nível mais baixo de abstração.....	182
FIGURA 4.3 – Página de consultas do servidor em português no momento do preenchimento dos campos a consultar.....	185
FIGURA 4.4 – Resposta da consulta sobre palavras existentes no texto.....	185
FIGURA 4.5 – Resposta para a consulta sobre uma palavra inexistente nos arquivos indexados.....	186
FIGURA 4.7– Consulta sobre o Banco de teste 1, no Servidor UFRGS, reproduzindo uma consulta real.....	187

Anexo 5

FIGURA 1 – Conjunto de caracteres da fonte SIGMODA e seus respectivos estereótipos.....	197
---	-----

Lista de Tabelas

Artigo XI

TABELA 1 - Classes do GeoFrame x classes Spring 147

Artigo XIV

TABELA 1 - Requisitos e Modelos Conceituais para SIG 193

Resumo

Esta tese aborda problemas relacionados ao projeto de banco de dados geográficos, componente responsável pelo armazenamento dos dados nas aplicações de Sistemas de Informação Geográfica (SIG). Especificamente, a pesquisa enfoca problemas que ocorrem durante a fase do projeto conceitual de bancos de dados geográficos, identifica deficiências e vantagens nos métodos e modelos conceituais descritos na literatura e propõe alguns instrumentos que facilitam a reutilização de soluções já desenvolvidas e validadas, anteriormente, por outros projetistas mais experientes. Esta tese está centrada na hipótese de que a maior parte das falhas de projeto de banco de dados geográficos deve-se, principalmente, à ausência de um método adequado que auxilie o projetista a elaborar o esquema conceitual do mesmo. Um fator agravante é que, normalmente, os usuários, na maioria das vezes com formação fora da área de computação, projetam seus próprios sistemas.

O objetivo geral desta tese foi a investigação de modelos, metodologias e métodos que possam ser aplicados durante o projeto conceitual de banco de dados geográficos. Dentre os objetivos mais específicos está o desenvolvimento de instrumentos que, utilizados de forma conjunta, possibilitem o emprego da idéia de reutilização de projetos já existentes. Como resultado desta pesquisa, as principais contribuições obtidas foram: (1) a identificação e especificação de um conjunto de padrões de análise que documentam soluções para problemas de modelagem que são recorrentes em aplicações nas áreas de controle ambiental e de gestão urbana; (2) um *framework* conceitual, denominado GeoFrame, que serve como ponto de partida para a modelagem conceitual de bancos de dados geográficos, bem como para a especificação de padrões de análise em aplicações de SIG.

Palavras-chave: Banco de Dados, Projeto Conceitual, Padrões de Análise, *Framework*, Reutilização, Sistema de Informação Geográfica.

TITLE: “Conceptual Design Of Geographic Databases Through Schema Reuse: Applying Analysis Patterns And A Conceptual Framework”

Abstract

This thesis describes a research work in the area of geographic database design. Geographic databases are the data storage components of geographic information systems (GIS). Main focus of this thesis is the conceptual design of geographic databases. The work investigates strengths and weaknesses of existing conceptual geographic data models, and proposes forms of reuse of well known design solutions in this area. It relies on the idea that the occurrence of failures in geographic database design is mainly due to the absence of a well suited design method. This fact is yet more relevant when one observes that GIS database design is mainly carried out by professionals with no knowledge of database design techniques.

In this work we have investigated models, methodologies and methods that can be applied to support conceptual design of geographic databases. A set of concepts as well as constructs are introduced that enable the reuse of existing solutions to some recurrent conceptual design problems in the GIS domain. Main contribution of this thesis is, on the one hand, the introduction as well as specification of a set of analysis patterns for geographic database design. A conceptual framework called GeoFrame complements the main results. It can be used as a starting point of the design process as well as a basis for the description of new analysis patterns for geographic applications.

Keywords: Database, Conceptual Design, Analysis Patterns, Framework, Reuse, Geographic Information System.

1 Introdução

Este documento constitui a tese submetida à avaliação, como requisito parcial para obtenção do grau de Doutor em Ciência da Computação, junto ao Programa de Pós-Graduação em Ciência da Computação (PPGC) da Universidade Federal do Rio Grande do Sul (UFRGS). A tese aborda problemas na área de Banco de Dados aplicados a Sistemas de Informação Geográfica.

A pesquisa enfoca, especificamente, problemas que ocorrem durante a fase de projeto conceitual de bancos de dados geográficos, identifica deficiências e vantagens nos métodos e modelos conceituais descritos na literatura, e propõe alguns instrumentos que facilitam a reutilização de soluções já desenvolvidas e validadas anteriormente, por outros projetistas mais experientes. Neste capítulo, apresenta-se a motivação para o desenvolvimento da pesquisa através da descrição dos vários problemas identificados, a hipótese investigada, os objetivos e contribuições desta pesquisa e, por último, a organização do documento.

1.1 Motivação

Sistemas de Informação Geográfica (SIG) são sistemas computacionais que permitem a captura, armazenamento, manipulação, recuperação, análise e apresentação de dados referenciados geograficamente [WOR 95]. Dados referenciados geograficamente ou, simplesmente, dados georreferenciados são dados que descrevem fenômenos geográficos cuja localização está associada a uma posição sobre/sob a superfície terrestre.

Devido à necessidade de gerenciamento de grandes volumes de dados, a maioria dos SIG disponíveis comercialmente trabalham de forma integrada com sistemas gerenciadores de banco de dados (SGBD). Assim, as aplicações desenvolvidas a partir de um SIG apresentam complexidade semelhante ou maior aquela apresentada por aplicações que são desenvolvidas com base em SGBD. No entanto, técnicas de projeto de banco de dados, já consagradas no desenvolvimento de aplicações em diversas áreas (ex.: aplicações comerciais), não têm sido empregadas pela comunidade de projetistas de aplicações de SIG. Isto é decorrente de alguns fatores, dentre os quais pode-se citar:

- as aplicações de SIG são projetadas e desenvolvidas, muitas vezes, pelos próprios usuários, os quais, normalmente, possuem formação em diferentes áreas como, por exemplo, geografia, cartografia, biologia, agronomia e engenharia. Em geral, esses profissionais não se preocupam com a modelagem conceitual do banco de dados a ser utilizado pela aplicação de SIG [PIR 97];
- a inexperiência por parte desses projetistas, em relação ao uso de metodologias de desenvolvimento de sistemas de informação, tem resultado na construção de sistemas com sérias limitações de uso (ex.: ocorrência freqüente de dados redundantes e, muitas vezes, inconsistentes); e
- o elevado custo do processo de aquisição e preparação de dados georreferenciados, conseqüência de sua característica espaço-temporal, faz com que o esforço do projetista esteja voltado, principalmente, para este problema.

Atualmente, no Brasil, as aplicações de SIG estão sendo desenvolvidas, em sua grande maioria, por instituições públicas, prefeituras e concessionárias de serviços

públicos. Estas instituições enfrentam uma série de problemas comuns, dentre os quais estão as limitações quanto à disponibilidade de recursos financeiros e a falta de pessoal técnico especializado.

Um dos pontos mais críticos de uma aplicação de SIG é o problema da aquisição dos dados [ARO 89]. Como muitas instituições não dispõem de recursos financeiros suficientes para aquisição ou produção dos dados georreferenciados, torna-se fundamental o intercâmbio de dados entre as instituições, a fim de que o maior número possível de usuários possa beneficiar-se de dados pré existentes.

Durante o desenvolvimento da modelagem conceitual de diversas aplicações de SIG [LIS 96, LIS 97, LIS 97a, LIS 99b], através da execução de projetos de cooperação entre o Instituto de Informática da UFRGS e várias instituições usuárias, duas tendências importantes, com relação ao projeto de banco de dados geográficos, puderam ser observadas:

- o problema da falta de recursos humanos qualificados deve perdurar por vários anos, sendo necessário, portanto, buscar soluções que contribuam para a melhoria da qualidade dos sistemas desenvolvidos pelos usuários;
- o problema da aquisição de dados tende a ser gradativamente reduzido, à medida que conjuntos de dados espaciais venham a ser documentados por meio de metadados, divulgados, por exemplo, através de catálogos de metadados disponibilizados na Internet [WEB 98].

Dentre os projetos de cooperação desenvolvidos pelo Instituto de Informática da UFRGS, dois merecem destaque. O primeiro é o Sistema de Informação do Gerenciamento Costeiro (SIGERCO-RS), desenvolvido em parceria com a Fundação Estadual de Proteção Ambiental do Rio Grande do Sul (Fepam), no âmbito do projeto integrado de pesquisa SIGMODA, financiado pelo CNPq. O segundo grande projeto desenvolvido foi a modelagem conceitual do SIG do Programa Pró-Guaíba (SIGPROGB) [BIN 98], no qual estão envolvidas mais de dez instituições estaduais e municipais, além da UFRGS. O Programa Pró-Guaíba tem como objetivo o desenvolvimento de uma série de ações visando a melhoria da qualidade da água em toda a bacia hidrográfica do Rio Guaíba. Este programa é financiado pelo BID – Banco Interamericano de Desenvolvimento e pelo Governo do Estado do Rio Grande do Sul.

Além dos problemas citados anteriormente, foi identificado um outro, o qual está mais especificamente relacionado com o projeto conceitual de banco de dados geográficos para aplicações de SIG. Os bancos de dados geográficos impõem uma série de requisitos de modelagem que devem ser atendidos pelos modelos conceituais de dados para as aplicações de SIG. Este problema tem levado ao desenvolvimento de diversas propostas de extensão dos modelos conceituais tradicionais (ex.: Entidade-Relacionamentos e modelos orientados a objetos). Pode-se citar, por exemplo, os modelos Modul-R [HAD 96], Geo-ER [HAD 97], Geo-OMT [BOR 97], GMOD [PIR 97], GeoOOA [KOS 97], GISR [SHK 97] e MADS [PAR 98]. Um estudo comparativo desses modelos pode ser encontrado em [LIS 99d]. Nenhum desses modelos, no entanto, obteve grande aceitação por parte dos projetistas de SIG que continuam, em sua maioria, não realizando a etapa de modelagem conceitual. Uma prova disso é que novas extensões de modelos conceituais continuam sendo propostas [BED 99, PAR 99 e RAM 99].

Desde o segundo semestre de 1995, tem-se experimentado várias destas idéias propostas na literatura, mas a maioria destas soluções falharam ou foram insuficientes, principalmente quando de sua utilização em grandes sistemas não hipotéticos. Um exemplo real é o caso do SIGPROGB [LIS 97], no qual foram identificadas e modeladas mais de 250 classes apenas durante a primeira fase do projeto. Atualmente,

já são mais de 300 classes, pois o projeto encontra-se ainda em fase de desenvolvimento.

Paralelamente à aplicação das soluções encontradas na literatura, buscou-se uma solução própria, na qual, partindo-se de um modelo orientado a objetos (foram experimentados os modelos OOA [COA 91], OMT [RUM 91] e UML [BOO 98]), desenvolveu-se um diagrama de classes para servir de base à modelagem conceitual de banco de dados geográficos.

Através destas experiências reais, desenvolvidas no âmbito do projeto SIGMODA, e de várias experiências acadêmicas, realizadas como trabalho de disciplinas oferecidas pelo PPGC da UFRGS (inclusive em alguns cursos de mestrado remoto), pôde-se observar uma série de problemas e deficiências na estrutura do diagrama de classes. Estas constatações práticas levaram ao desenvolvimento de uma série de versões do diagrama, cuja documentação constitui uma das contribuições desta tese [LIS 96, LIS 97a, IOC 98, LIS 99].

A partir da experiência adquirida durante a modelagem conceitual dos dados destes sistemas, constatou-se que muitos fenômenos geográficos são modelados de forma semelhante quando as aplicações pertencem a um determinado domínio, neste caso, o domínio das aplicações de controle ambiental.

Normalmente, durante as fases de análise de requisitos e modelagem conceitual de dados das aplicações de SIG é identificado um grande número de entidades geográficas e também de relacionamentos entre as mesmas. Segundo Gordillo [GOR 97], um projetista experiente em modelagem de dados desenvolve seus modelos a partir do conhecimento prévio de um conjunto de entidades interrelacionadas, em vez de desenvolver cada novo projeto do início. Os projetistas reutilizam parte dos trabalhos desenvolvidos anteriormente para resolver novos problemas similares. No entanto, esta experiência não tem sido passada para outros projetistas devido à inexistência de instrumentos adequados que possibilitem o compartilhamento de projetos já testados e validados.

Por outro lado, o intercâmbio de dados georreferenciados vem aumentando entre diferentes organizações, principalmente, através do uso da *Web* e de outros serviços da Internet. Para facilitar o compartilhamento de dados, os conjuntos de dados georreferenciados existentes são documentados por meio de metadados, que são organizados em um formato padronizado e disponibilizados através de um sistema de catálogo de metadados [FED 97]. Embora já exista uma avançada infra-estrutura para auxiliar no compartilhamento de dados georreferenciados, projetos de banco de dados para aplicações de SIG continuam sendo feitos, quase sempre, a partir do zero.

A abordagem de padrões¹ tem sido utilizada na área de Engenharia de Software, principalmente em projetos de sistemas orientados a objetos, para possibilitar a reutilização de soluções bem sucedidas nas diversas etapas do desenvolvimento de sistemas (ex.: análise, projeto e implementação). Como consequência, desenvolveram-se várias categorias de padrões, as quais representam diferentes níveis de abstração e cujos padrões podem ser aplicados em cada uma dessas etapas. Padrões de análise, por exemplo, têm sido elaborados com o objetivo de auxiliar a reutilização de soluções existentes durante as fases de levantamento de requisitos e modelagem conceitual de dados [HAY 95, COA 97, FOW 97].

Segundo Gamma e outros [GAM 94], um padrão apresenta a essência de uma solução para um problema recorrente em um contexto específico. Padrões de análise descrevem soluções para problemas de modelagem de aplicações com base em um

1 Do inglês pattern

modelo semântico de alto nível, onde são aplicados termos e conceitos usados no próprio domínio da aplicação [RIE 96].

Devido à natureza dos dados geográficos, suas aplicações são promissoras candidatas à reutilização de projetos de banco de dados [LIS 98]. Por exemplo, aplicações na área de controle ambiental são sempre baseadas em um conjunto de temas comuns como geologia, hidrografia, hipsografia, geomorfologia, temperatura, etc. As principais variações dizem respeito à área em estudo, ao nível de detalhamento da informação (ex.: escala cartográfica utilizada) e aos objetivos da análise ambiental que se deseja realizar. Desta forma, existem alguns conjuntos de entidades ou fenômenos geográficos que aparecem relacionados entre si, quase sempre da mesma forma, em diferentes aplicações. A descrição destes conjuntos de entidades interrelacionadas, na forma de padrões de análise e a ampla divulgação destes padrões permitem que novas aplicações façam uso de soluções existentes.

1.2 Hipótese

Acredita-se que o uso de padrões de análise possa possibilitar um ganho significativo em termos de qualidade e produtividade no desenvolvimento de aplicações de SIG por permitir, por exemplo, que projetistas menos experientes possam reutilizar soluções já testadas e validadas por outros projetistas mais experientes. Além disso, padrões de análise podem ser úteis no treinamento de novos projetistas de bancos de dados geográficos.

Assim, tornar o processo de modelagem conceitual mais simples pode incentivar profissionais de outras áreas, que não a Informática, a procederem de forma metodológica no projeto do banco de dados de suas aplicações de SIG, ou seja, partindo de um esquema conceitual que represente, de forma não ambígua, os conceitos da realidade, até transformá-lo no esquema proprietário de um SIG comercial.

A idéia de reutilização em projeto de SIG foi explorada, inicialmente, por A. Marr, R. Pascoe e G. Benwell [MAR 97], que propuseram uma estrutura de modelagem de processos de análise espacial², que pode ser reutilizada a partir de sua disponibilização através da *Web*. Tal abordagem é complementar à que se está propondo neste trabalho. Pode-se, por exemplo, definir padrões de processos funcionais para documentar procedimentos de análise espacial em SIG.

O emprego da abordagem de padrões no desenvolvimento de aplicações de SIG é um assunto praticamente inexplorado hoje em dia. Alguns dos raros exemplos podem ser vistos em [BAL 97, GOR 98] que propõem o uso de padrões de projeto no desenvolvimento de software de SIG.

1.3 Objetivos

O objetivo geral desta pesquisa de tese é a investigação de modelos, metodologias e métodos que possam ser aplicados durante a realização da modelagem conceitual de banco de dados geográficos, desenvolvidos para o armazenamento de dados em aplicações de SIG.

² Na comunidade de usuários de geoprocessamento, o termo análise espacial corresponde à realização de consultas ao banco de dados geográficos, usando-se para isto, as funções de análise espacial disponíveis nos SIG (ex.: *overlay* de mapas). Nesta tese, o termo análise diz respeito à fase de análise de requisitos de um sistema computacional.

Entre os objetivos mais específicos está o desenvolvimento de uma série de instrumentos que, utilizados de forma conjunta, possibilitam o emprego da idéia de reutilização no projeto de banco de dados geográficos.

Uma das características fundamentais desta pesquisa foi a busca por uma solução que pode ser utilizada na prática, não só por especialistas em projetos computacionais mas, principalmente, pelos próprios usuários de SIG, uma vez que estes, na maioria dos casos, projetam e desenvolvem seus próprios sistemas.

Além das contribuições técnicas, espera-se que os resultados desta pesquisa contribuam diretamente para a melhoria da qualidade dos resultados obtidos através do uso da tecnologia de SIG no País e no exterior, bem como para o fortalecimento da cultura de cooperação intra e interinstitucional, através do compartilhamento tanto de dados como de projetos de banco de dados. Isto leva, como consequência direta, à redução dos investimentos necessários para o desenvolvimento de aplicações de SIG.

1.4 Estrutura da tese

Esta tese foi elaborada de acordo com o novo formato recomendado pela Comissão do Programa de Pós-Graduação em Computação da UFRGS. A tese está organizada como uma coletânea de artigos produzidos durante o curso. Ao todo são quatorze artigos, dentre os quais doze já foram publicados, um foi submetido para publicação e o outro será submetido a um evento (a ser definido) no início do próximo ano.

O trabalho está organizado da seguinte forma. O Capítulo 1 apresenta a introdução da tese. Os artigos estão agrupados por assunto nos capítulos 2, 3 e 4. Cada um desses capítulos inicia-se por uma subseção de introdução, na qual estão descritas as contribuições de cada artigo constante do capítulo. O Capítulo 5 apresenta as conclusões e trabalhos futuros. Três artigos estão descritos como anexo. Os dois primeiros documentam contribuições adicionais da tese e o terceiro é um resumo estendido.

O Capítulo 2 apresenta dois artigos, nos quais encontra-se documentada a revisão bibliográfica sobre os dois principais temas dessa pesquisa. O Artigo I (Seção 2.2), descreve o estudo realizado sobre os modelos conceituais de dados que foram propostos especificamente para projeto de banco de dados geográficos. Este artigo apresenta o estado-da-arte do assunto até o mês de agosto de 1999, quando foi submetido para publicação na revista *IP-Informática Pública*, editada pelo Centro de Desenvolvimento e Estudos da Prodabel, Empresa de Informática e Informação do Município de Belo Horizonte. O Artigo II (Seção 2.3) descreve as principais abordagens de reutilização usadas nas fases iniciais do desenvolvimento de sistemas de informação. O artigo foi submetido, em setembro de 1999, para publicação na revista *Acta Scientiae*, editada pelo Centro de Ciências Naturais e Exatas da Universidade Luterana do Brasil (ULBRA).

O Capítulo 3 apresenta a evolução do diagrama de classes BCH/GIS, que serve como ponto de partida para a modelagem de um novo banco de dados geográficos. O diagrama BCH/GIS é resultado da evolução ocorrida durante o desenvolvimento de diversos sistemas. O Artigo III (Seção 3.2) apresenta a primeira versão do diagrama, a qual foi utilizada na modelagem do sistema SIGERCO-RS como parte de um projeto de cooperação entre a Fepam - Fundação de Proteção Ambiental do RS e o Instituto de Informática da UFRGS. Os artigos IV e V (Seções 3.3 e 3.4) apresentam resultados parciais da modelagem do sistema SIGPROGB, o SIG do Programa Pró-Guaíba. O Artigo IV descreve a metodologia de modelagem do sistema SIGPROGB e apresenta a segunda versão do diagrama BCH/GIS, enquanto o Artigo V descreve a necessidade de

uma nova fase no ciclo de projeto de grandes bancos de dados geográficos, a fase do Projeto de Representação Espaço-Temporal. Por último, o Artigo VI (Seção 3.5) documenta os problemas identificados nas duas primeiras versões do diagrama BCH/GIS, propondo uma nova versão. Esta terceira versão foi avaliada como parte de uma dissertação de mestrado [GAR 98] e suas sugestões foram incorporadas quando da transformação do diagrama BCH/GIS no *framework* GeoFrame, documentado no Capítulo 4.

O Capítulo 4 apresenta os principais resultados desta tese. No Artigo VII (Seção 4.2), é descrita a adequação do uso de padrões de análise no domínio das aplicações de SIG. O Artigo VIII (Seção 4.3) introduz o *framework* conceitual GeoFrame, que é derivado do diagrama de classes BCH/GIS, servindo de base à modelagem de novos bancos de dados geográficos e, também, como suporte à especificação de padrões de análise em SIG. O Artigo IX (Seção 4.4) introduz um conjunto de padrões de análise que foram identificados a partir de um estudo detalhado de diversos esquemas conceituais, desenvolvidos no âmbito do projeto SIGMODA ou por terceiros. O Artigo X (Seção 4.5) descreve um exemplo de modelagem no qual foram aplicados alguns dos padrões descritos no Artigo IX. Por último, o Artigo XI (Seção 4.6) exemplifica a usabilidade do *framework* GeoFrame, apresentando um exercício de mapeamento de um esquema conceitual, construído com base no GeoFrame, para o modelo da interface do sistema Spring [INP 99].

No Capítulo 5 são apresentadas as conclusões, são comentados os resultados obtidos, as dificuldades encontradas e os avanços alcançados. O capítulo descreve, ainda, uma série de trabalhos que darão continuidade a esta pesquisa, através de projetos de cooperação entre o Instituto de Informática da UFRGS, o Departamento de Informática da Universidade Federal de Viçosa (UFV) e possíveis parceiros institucionais.

O Anexo 1 inclui o Artigo XII, o qual descreve alguns resultados do estudo sobre qualidade de dados georreferenciados e distribuição de metadados, apresentando uma experiência de documentação de dados em uma aplicação de SIG para a área ambiental. Este artigo, embora não trate diretamente do assunto principal desta tese, foi incluído neste volume devido a este tema de distribuição de dados ter sido alvo de pesquisa relativa aos modelos conceituais de dados para SIG. Além disso, este artigo é referenciado por vários dos artigos constantes dessa tese.

O Artigo XIII, descrito no Anexo 2, apresenta a experiência de implementação de um protótipo, desenvolvido no âmbito do Projeto SIGMODA, de um sistema para distribuição e busca a metadados através da Internet, para servir aos usuários de países de língua portuguesa.

O Anexo 3 inclui o Artigo XIV, que é um estudo sobre os modelos conceituais de dados, publicado apenas como resumo estendido.

O Anexo 4 apresenta a relação das demais publicações produzidas pelo autor durante seu curso de doutorado, mas que não foram incluídas neste volume de tese. São tutoriais, relatórios técnicos e monografias.

Por último, o Anexo 5 inclui a descrição de como utilizar o arquivo fonte de caracteres (sigmoda.ttf), que contém os caracteres especiais utilizados na elaboração dos diagramas de classes. Este arquivo está disponível no sítio do projeto SIGMODA (<http://www.inf.ufrgs.br/gpesquisa/sigmoda>).

A seguir estão relacionadas as referências completas dos artigos que compõem este volume de tese.

Capítulo 2

- LISBOA FILHO, J.; IOCHPE, C. Um estudo sobre modelos conceituais de dados para projeto de bancos de dados geográficos. **Revista IP-Informática Pública**, Belo Horizonte, v.1, n.2, 1999.
- LISBOA FILHO, J.; IOCHPE, C. Mecanismos de reutilização em sistemas de informação. Submetido à revista **Acta Scientiae**, Canoas-RS, 1999.

Capítulo 3

- LISBOA FILHO, J.; IOCHPE, C. Adaptando o modelo de objetos OMT para modelagem conceitual de aplicações de SIG. In: SEGEO-RJ, 1., 1996, Rio de Janeiro. **Anais...** Rio de Janeiro: Esc. Engenharia, 1996.
- LISBOA FILHO, J.; GARAFFA, I. M.; IOCHPE, C. Metodologia para modelagem conceitual de dados de um SIG interinstitucional. In: GISBRASIL CONGRESSO E FEIRA PARA USUÁRIOS DE GEOPROCESSAMENTO, 3., 1997, Curitiba. **Anais...** Curitiba: Fator GIS, 1997.
- LISBOA FILHO, J.; IOCHPE, C.; GARAFFA, I. M. Modelos conceituais de dados para aplicações geográficas: uma experiência com um SIG interinstitucional. In: SIMPÓSIO BRASILEIRO DE GEOPROCESSAMENTO, 4., São Paulo, 1997. **Anais...** São Paulo: UPUSP, 1997.
- IOCHPE, C.; LISBOA FILHO, J. BCH/GIS: a basic class hierarchy to support GIS conceptual design. In: INT. CONF. ON MODELING GEOGRAPHICAL AND ENVIRONMENTAL SYSTEMS WITH GEOGRAPHICAL INFORMATION SYSTEMS (GIS). 1998, Hong Kong, China. **Proceedings...** Hong Kong: The Chinese University of Hong Kong, 1998.

Capítulo 4

- LISBOA FILHO, J.; IOCHPE, C.; BEARD, K. Applying analysis patterns in the GIS domain. In: ANNUAL COLLOQUIUM OF THE SPATIAL INFORMATION RESEARCH CENTRE, 10., 1998, Dunedin, New Zealand. **Proceedings...** Dunedin: SIRC, University of Otago, 1998.
- LISBOA FILHO, J.; IOCHPE, C. Specifying analysis patterns for geographic databases on the basis of a conceptual framework. In: ACM SYMPOSIUM ON ADVANCES IN GEOGRAPHIC INFORMATION SYSTEMS, 7., 1999, Kansas City, USA. **Proceedings...** Kansas City: ACM Press, 1999.
- LISBOA FILHO, J.; IOCHPE, C. **Padrões de análise para banco de dados geográfico**. A ser submetido.
- LISBOA FILHO, J.; IOCHPE, C.; HASENACK, H.; WEBER, E. J. Modelagem conceitual de banco de dados geográficos: o estudo de caso do projeto PADCT/CIAMB. In: UFRGS/CENTRO DE ECOLOGIA. **Energia e Meio Ambiente: a questão do carvão no Rio Grande do Sul**. Porto Alegre: Editora da Universidade, 1999. No prelo.
- LISBOA FILHO, J.; COSTA, A. C.; IOCHPE, C. Projeto de banco de dados geográficos: mapeando esquemas *GeoFrame* para o SIG *Spring*. In: BRAZILIAN WORKSHOP ON GEOINFORMATICS - GEO-INFO, 1., 1999, Campinas. **Anais...** São José dos Campos: INPE, 1999.

Anexo 1

- WEBER, E. J.; LISBOA FILHO, J.; IOCHPE, C.; HASENACK, H. Geospatial metadata in Brazil: an experience in data documentation of an environmental GIS application. In: INT. CONFERENCE & EXHIBITION ON GEOGRAPHIC INFORMATION - GISPlanet, 1998, Lisbon, Portugal. **Proceedings...** Lisbon: USIG,1998.

Anexo 2

- VILLAVICENCIO, F.; IOCHPE, C; LISBOA F., J. Um servidor de consultas à clearinghouse em português para distribuição de metadados geográficos. In: GISBRASIL'99 - CONGRESSO E FEIRA PARA USUÁRIOS DE GEOPROCESSAMENTO DA AMÉRICA LATINA, 5, 1999, Salvador. **Anais...** Salvador: Fator GIS, 1999.

Anexo 3

- LISBOA FILHO, J.; IOCHPE, C. Requisitos especiais e modelos conceituais de dados para aplicações de sistemas de informações geográficas. In: IDEAS - WORKSHOP IBEROAMERICANO DE ENGENHARIA DE REQUISITOS E AMBIENTES DE SOFTWARE, 1998, Torres-RS. **Proceedings...** Torres: UFRGS/Instituto de Informática,1998. Resumo estendido.

2 Revisão Bibliográfica

2.1 Introdução

Este capítulo apresenta, inicialmente, o levantamento bibliográfico realizado sobre o estado-da-arte em modelos conceituais de dados específicos para projeto de banco de dados geográficos. Em seguida, é apresentado um estudo no qual são identificados os principais mecanismos de reutilização que podem ser aplicados nas etapas iniciais do processo de desenvolvimento de sistemas de informação (*early reuse*).

O primeiro artigo, intitulado “*Modelos conceituais de dados para projeto de bancos de dados geográficos*”, apresenta os últimos resultados sobre o estudo que vem sendo desenvolvido desde a publicação do primeiro relatório de pesquisa [LIS 96a]. Neste artigo é apresentado um conjunto de requisitos que foram identificados ao longo desta pesquisa que teve como base principal, a experiência obtida durante a modelagem conceitual de diversos sistemas. Posteriormente, pôde-se compará-lo com outros conjuntos de requisitos publicados por outros autores [EUR 96, HAD 97 e RUG 98]. Tais requisitos são utilizados na análise das características apresentadas pelos diversos modelos específicos para projeto de banco de dados em aplicações de SIG.

Algumas versões preliminares deste conjunto de requisitos, juntamente com algumas análises de modelos encontrados na literatura, também foram publicados em alguns de nossos artigos [LIS 96, LIS 97, LIS 98a] e na monografia do exame de qualificação [LIS 97b]. Os artigos estão documentados no Capítulo 3 com exceção da referência [LIS 98a], que foi publicada como resumo estendido e está incluída no Anexo 3.

No segundo artigo, intitulado “*Abordagens de reutilização em sistemas de informação*”, são descritos os principais mecanismos de reutilização aplicados nas etapas iniciais do processo de desenvolvimento de sistemas de informação. São descritas algumas pesquisas sobre reutilização de especificações de requisitos, sobre padrões (*patterns*) de projeto e de análise, e sobre *frameworks*. Padrões e *frameworks* são os novos mecanismos propostos, inicialmente, para auxiliar na reutilização de projetos orientados a objetos [GAM 94, JOH 92]. O estudo também descreve mecanismos atuais de suporte à reutilização de dados na área de SIG.

2.2 Artigo I: Um Estudo sobre Modelos Conceituais de Dados para Projeto de Bancos de Dados Geográficos

Jugurta Lisboa Filho e Cirano Iochpe

Revista IP – Informática Pública, Belo Horizonte, v.1, n.2, 1999.

RESUMO

Modelagem conceitual tem sido aplicada com sucesso no projeto de bancos de dados em geral. Independentes de plataformas de hardware e software, os modelos conceituais permitem representar, de maneira abstrata, formal e não ambígua, a realidade da aplicação, facilitando a comunicação entre projetistas e usuários. Aplicações geográficas, contudo, impõem alguns requisitos específicos de modelagem que não são satisfatoriamente atendidos pelos modelos de dados conceituais de propósito geral. Neste trabalho são apresentados os principais requisitos de modelagem que devem ser atendidos por modelos conceituais para Sistemas de Informação Geográfica (SIG). Com base nesses requisitos, o artigo faz uma análise comparativa de alguns dos mais conhecidos modelos conceituais de dados propostos especificamente para aplicações geográficas.

Palavras-chave: Sistema de Informação Geográfica (SIG), Modelo conceitual de dados, Sistema de banco de dados

ABSTRACT

Conceptual modeling has been applied successfully to database modeling in general. Independent of underlying hardware or software, conceptual models allow for a first, abstract representation of the application's reality helping designers to better communicate with one another and with the users. Geographic applications pose a set of specific requirements on conceptual data modeling that are either not at all or only partially supported by existing conceptual data models. This paper discusses a minimum set of requirements which should be supported by a data model for Geographic Information Systems (GIS) and compares some well known conceptual data models relying on these requirements.

Keywords: Geographic Information Systems (GIS), Conceptual data modeling, Database systems.

1 INTRODUÇÃO

O termo Sistema de Informação Geográfica (SIG) caracteriza os sistemas de informação que tornam possível a captura, modelagem, manipulação, recuperação, análise e apresentação de dados referenciados geograficamente (ou dados georreferenciados) [WOR 95]. De forma geral, um software de SIG é um sistema composto de quatro grandes componentes: componente de captura de dados,

componente de armazenamento, componente de análise e componente de apresentação dos dados.

O componente de armazenamento, denominado sistema de banco de dados geográficos, estrutura e armazena os dados de forma a possibilitar a realização das operações de análise e consulta. Devido à complexidade das aplicações que são desenvolvidas a partir de um SIG, projetar o banco de dados geográficos tem sido um dos grandes desafios para as organizações públicas e privadas.

O projeto de um banco de dados deve ser realizado com o apoio de um modelo de dados de alto nível, também conhecido como modelo conceitual. Durante vários anos, as pesquisas no campo dos modelos de dados para SIG centraram-se na busca por estruturas de dados para o armazenamento de dados georreferenciados, o que ficou conhecido como “debate *raster-vector*” [COU 92]. No final dos anos 80, iniciaram-se pesquisas sobre o uso dos conceitos de orientação a objetos no desenvolvimento de SIG [FRA 88, EGE 89, WOR 90].

A necessidade de novos modelos conceituais para o desenvolvimento de aplicações de SIG foi identificada somente no final da década de 80 [EGE 87]). O desenvolvimento de aplicações de SIG tem sido feito de forma não metodológica, tendo como resultado diversos problemas decorrentes de abordagens evolutivas desordenadas.

Segundo Williamson [WIL 91], inicialmente os tradicionais ciclos de vida de desenvolvimento de sistemas não eram aplicados em SIG porque cada sistema era visto como único e existiam, invariavelmente, muitas incertezas quanto aos requisitos do sistema. Desta forma, a implementação de projetos-piloto e de protótipos, sob os quais as modificações eram feitas durante o desenvolvimento, foi e tem sido a forma mais comum de desenvolvimento desses sistemas.

Para Williamson [WIL 91], as características dos dados geográficos impõem barreiras à implementação das aplicações de SIG, características essas que não aparecem na maioria dos outros sistemas de informação. Entre os problemas citados estão a grande variedade de métodos de aquisição de dados, a diferença de acurácia apresentada pelos conjuntos de dados, a falta de uma codificação adequada dos elementos sendo modelados, a ocorrência de coberturas de dados esporádicas, a necessidade de dados temporais e a incompatibilidade entre os conjuntos de dados.

Um dos fatores agravantes no desenvolvimento das aplicações de SIG diz respeito ao tamanho dessas aplicações. Por exemplo, diversas instituições podem estar envolvidas em um mesmo projeto tendo, no entanto, diferentes objetivos. Além disso, o período de implementação pode ser medido não em meses, mas sim em anos ou mesmo décadas. O desenvolvimento de modelos conceituais mais adequados para a modelagem das aplicações de SIG, incluindo seus bancos de dados, é um dos pontos mais importantes para possibilitar um melhor suporte ao projetista nos estágios iniciais do desenvolvimento dessas aplicações.

Diversos autores vêm propondo extensões dos modelos conceituais, tradicionalmente utilizados no desenvolvimento de sistemas de informação, para uso no projeto de banco de dados geográficos. Este artigo apresenta o estado-da-arte em modelos conceituais de dados propostos especificamente para SIG. A Seção 2 introduz alguns conceitos básicos sobre modelos de dados em banco de dados. A Seção 3 descreve o processo de modelagem conceitual de um banco de dados. A Seção 4 apresenta os principais requisitos de modelagem impostos pelas aplicações de SIG, através dos quais os diversos modelos conceituais de dados geográficos são analisados. Por último são apresentadas as conclusões do trabalho.

2 MODELOS DE DADOS

Um modelo de dados é uma coleção de conceitos que podem ser usados para descrever um conjunto de dados e as operações para manipular esses dados [BAT 92]. Segundo Navathe [NAV 92], os modelos de dados podem ser classificados, basicamente, em duas dimensões. Na primeira dimensão, os modelos de dados são classificados em função da etapa de desenvolvimento do projeto do banco de dados em que o modelo é utilizado (ex.: projeto conceitual, lógico e físico). A segunda dimensão classifica os modelos de dados quanto a sua flexibilidade e poder de expressão.

A tarefa de projetar um banco de dados está intimamente relacionada com o ciclo de vida de desenvolvimento de software onde, a cada etapa, novas informações e detalhes são acrescentados ao projeto do software [PRE 87]. No projeto de banco de dados, as informações que comporão o banco de dados são especificadas utilizando-se modelos de dados em diferentes níveis de abstração, iniciando por modelos de alto nível de abstração (ex.: modelo E-R [CHE 76]) e refinando-se o modelo até que sejam incorporados detalhes específicos, relacionados ao armazenamento dos dados (ex.: estruturas de dados).

Quanto à escala de flexibilidade e expressividade, o termo flexibilidade refere-se, neste contexto, à facilidade com a qual o modelo pode tratar com aplicações complexas, enquanto que a expressividade refere-se à habilidade de gerar diferentes abstrações em uma aplicação [NAV 92]. Segundo este aspecto, os modelos são classificados como: modelos primitivos (ou de arquivos); modelos de dados clássicos (ex.: relacional, hierárquico); modelos de dados semânticos (ex.: IFO [ABI 87]); modelos semânticos de propósito especial - desenvolvidos para atender as demandas em áreas de aplicações específicas (ex.: automação de escritório, VLSI, CAD/CAM e SIG) [BRO 84]. Mais recentemente esta lista foi acrescida dos modelos orientados a objetos.

Com base nessas duas classificações, este artigo aborda os modelos semânticos de propósito especial, ou seja, voltados para aplicações de SIG, com o nível de abstração empregado nos modelos conceituais, que são modelos independentes de software e/ou hardware a ser utilizado.

3 PROCESSO DE MODELAGEM CONCEITUAL DE DADOS

A característica básica de um modelo de dados, como o próprio termo explicita, é que ele é uma abstração da realidade. Um modelo conceitual de dados fornece uma base formal (notacional e semântica) para ferramentas e técnicas usadas para suportar a modelagem de dados. Modelagem de dados é o processo de abstração onde somente os elementos essenciais da realidade observada são enfatizados, descartando-se os elementos não essenciais. O processo de modelagem conceitual de banco de dados (Figura 1) compreende a descrição dos possíveis conteúdos dos dados, além de estruturas e de regras a eles aplicáveis.

A modelagem conceitual é sempre feita com base em algum formalismo conceitual (ex.: Entidade-Relacionamento, Orientação a Objetos) [EUR 96]. O resultado do processo de modelagem, denominado esquema conceitual, é apresentado através de uma linguagem formal de descrição que pode estar expressa através de uma sintaxe e/ou uma notação gráfica. Para cada formalismo conceitual podem existir diversas linguagens de descrição de esquema que são compatíveis com o formalismo.

O formalismo provê um conjunto de conceitos, elementos e regras que são usados no processo de modelagem da realidade, enquanto que a linguagem de descrição fornece uma gramática para a apresentação do esquema conceitual resultante da modelagem. A linguagem léxica possibilita o processamento computacional do

esquema, enquanto a notação gráfica é mais adequada para facilitar o entendimento e a comunicação entre seres humanos (ex.: usuários e projetistas). Portanto, um modelo conceitual de dados compreende um formalismo e uma linguagem de descrição, podendo ser uma linguagem léxica e/ou gráfica.

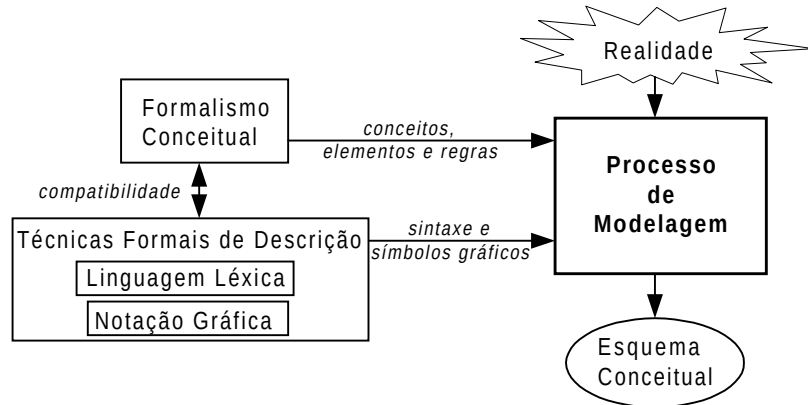


FIGURA 1 - Processo de modelagem conceitual

4 MODELOS CONCEITUAIS DE DADOS PARA SIG

4.1 Evolução Histórica

Modelos conceituais de dados têm sido adaptados para atender às necessidades impostas pelas aplicações de SIG desde o final dos anos 80 (Figura 2). A primeira extensão, da qual se tem conhecimento, foi publicada por Bédard e Paquette em 1989 [BED 89]. Os modelos conceituais de dados para SIG baseiam-se nos formalismos existentes. Além dos formalismos de Entidade-Relacionamento (ER) e de Orientação a Objetos (OO), algumas extensões de modelos têm como base o formalismo IFO (Modelo Semântico Formal de Banco de Dados) [ABI 87]. A seguir, são sintetizados alguns comentários sobre os modelos mostrados na Figura 2. Maiores detalhes sobre cada um desses modelos estão incluídos nas descrições apresentadas nas seções seguintes.

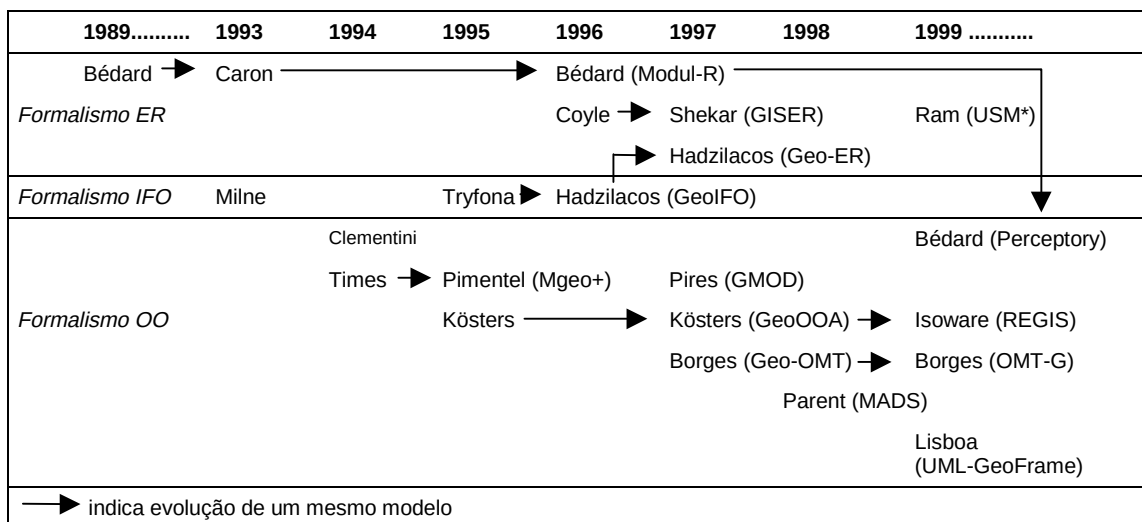


FIGURA 2 - Evolução cronológica dos modelos conceituais de dados para SIG

Bédard [BED 96] descreve uma extensão do modelo E-R [CHE 76], denominado Modul-R, para a representação de aplicações geográficas urbanas. Este modelo é resultado da evolução do trabalho de pesquisa desenvolvido na Universidade Laval, Canadá [BED 89, CAR 93]. O modelo Modul-R serve de base para a ferramenta CASE Perceptory [Bed 99], a qual emprega o formalismo OO e utiliza a sintaxe da linguagem UML [BOO 98].

A partir das pesquisas realizadas na Universidade de Minnesota [COY 96], Shekhar [SHK 97] propõe o modelo GISER (*Geographic Information System Entity Relational*), que é uma extensão para aplicações geográficas do modelo EER (E-R Estendido) [ELM 94].

Ram e outros [RAM 99] propõem o modelo USM* (*Unifying Semantic Model*), uma extensão do modelo E-R para solucionar o problema de acesso a diferentes bancos de dados geográficos. O modelo USM* foi desenvolvido na Universidade do Arizona.

Outro modelo derivado do formalismo ER é o modelo Geo-ER [HAD 97], que é uma adaptação do modelo GeoIFO [TRY 95, HAD 96], desenvolvido na Universidade de Patras, Grécia. O modelo GeoIFO, juntamente com o modelo proposto por Milne [MIL 93], são os únicos que têm como base o formalismo IFO [ABI 87, HUL 87].

O modelo de Clementini [CLE 94], desenvolvido na Universidade de L'Aquila, Itália, apresenta uma linguagem léxica para modelagem orientada a objetos de aplicações geográficas.

O modelo MGeo [TIM 94] foi desenvolvido na Universidade Federal de Pernambuco sendo, posteriormente, estendido no modelo MGeo+ [PIM 95]. Este modelo introduziu o uso de bibliotecas de classes que servem de base à modelagem de aplicações geográficas. O modelo MGeo+ é descrito com base no modelo de objetos OMT [RUM 91].

Pires [PIR 97] propõe uma extensão do modelo de objetos OMT [RUM 91], denominado GMOD, o qual serve de base para o ambiente computacional UAPE, que é específico para aplicações de SIG. O Modelo GMOD/UAPE é resultado de pesquisa desenvolvida na Unicamp, Campinas.

Kösters [KOS 95, KOS 97] descreve o modelo GeoOOA, uma extensão do modelo usado na metodologia de análise orientada a objetos OOA [COA 91], para análise de requisitos de aplicações em SIG. O modelo GeoOOA serviu de base para a implementação da ferramenta CASE REGIS [ISO 99], a qual emprega a notação gráfica da linguagem UML [BOO 98]. O modelo GeoOOA foi desenvolvido na Universidade de Hagen, Alemanha.

O modelo Geo-OMT [BOR 97], que é uma extensão do modelo de objetos OMT [RUM 91] para aplicações de SIG, foi desenvolvido na Fundação João Pinheiro, em Belo Horizonte. O modelo foi aprimorado para possibilitar a modelagem de restrições espaciais, passando a se chamar OMT-G [BOR 99]. Este modelo tem tido grande aceitação por parte de usuários/projetistas de SIG no Brasil.

O modelo conceitual MADS [PAR 98, PAR 99] também adota o formalismo de orientação a objetos através da inclusão de características do modelo padrão ODMG (*Object Database Management Group*) para sistemas orientados a objetos [CAT 97]. MADS está sendo desenvolvido na Universidade de Lausanne, Suíça.

Lisboa e Iochpe [LIS 99] propõem que a modelagem de banco de dados geográficos pode ser realizada a partir de um *framework* conceitual chamado GeoFrame, utilizando a linguagem de modelagem unificada UML [BOO 98], acrescida de alguns estereótipos. Estereótipo é o mecanismo de extensão fornecido pela própria UML. O *framework* GeoFrame foi desenvolvido na Universidade Federal do Rio Grande do Sul.

Existem ainda outras propostas de modelos conceituais para SIG, mas foram omitidas deste estudo devido à similaridade desses trabalhos com os aqui descritos. Algumas dessas extensões estão descritas em [EGE 92, BON 93, SUB 93, BAT 94, NAT 94, BON 95, CRO 96, ABR 98].

4.2 Requisitos de Modelagem Conceitual para Aplicações de SIG

Segundo Chrisman [CHR 97], um SIG provê um conjunto de atividades organizadas, através do qual as pessoas:

- medem os diversos aspectos de fenômenos e processos geográficos;
- representam estas medidas, normalmente na forma de um banco de dados digital, enfatizando temas espaciais, entidades e relacionamentos;
- operam sobre estas representações para produzir novas medidas e descobrir novos relacionamentos através da integração de fontes diversas; e
- transformam estas representações para ajustar a outras estruturas de entidades e relacionamentos.

Desta forma, um entendimento conceitual de um SIG inclui tanto os fenômenos estáticos quanto os eventos dinâmicos. Sendo que a localização e outras características dos fenômenos também são medidas sobre o tempo. O segundo item acima, enfatiza que um SIG inclui a representação das “coisas” sendo medidas, que são, normalmente, estruturadas na forma de temas, entidades e relacionamentos.

Com base na definição acima, Rugg e outros [RUG 98] descrevem um conjunto de requisitos para modelagem conceitual de SIG, onde alguns conceitos-chave são relacionados. São eles:

- fenômenos: habilidade para descrever e modelar a morfologia das coisas que existem no mundo real em um dado instante de tempo;
- processos: habilidade para descrever e modelar processos dinâmicos, incluindo atividades contínuas e eventos discretos;
- medidas: procedimentos e categorias definidas para registrar e representar as características dos fenômenos;
- localização: procedimentos específicos para registrar e representar as características de localização dos fenômenos;
- relacionamentos: habilidade para descrever e representar interações entre os fenômenos;
- temas: habilidade para agrupar fenômenos baseados em suas características, relacionamentos e outras propriedades;
- operações: habilidade para descrever e representar o comportamento dos fenômenos.

Com o objetivo de realizar um estudo comparativo entre diversas linguagens de esquema conceitual, no contexto de SIG, o Comitê Europeu para Padronização (CEN) também definiu um conjunto de requisitos que estas linguagens devem apresentar [EUR 96]. Entre os requisitos especificados pelo Comitê CEN, cabe ressaltar a determinação de inclusão, no modelo conceitual, dos aspectos geométricos e dos relacionamentos topológicos dos fenômenos geográficos.

Com base nestes conjuntos de requisitos e nos requisitos identificados em estudos anteriores dos autores [LIS 96a, LIS 97b, LIS 98a], o seguinte conjunto de requisitos de modelagem é usado para análise dos principais modelos conceituais descritos na literatura.

- Fenômenos geográficos e objetos convencionais: possibilidade de diferenciação entre fenômenos geográficos e objetos sem referência espacial.

- Visões de Campo e Objetos: possibilidade de modelagem dos fenômenos nas visões de campo e de objetos.
- Aspectos Temáticos: necessidade de organizar os fenômenos por tema.
- Aspectos Espaciais: possibilidade de modelagem das características espaciais dos dados.
- Múltiplas Representações: possibilidade de modelagem de que um fenômeno possa ter mais de uma representação espacial.
- Relacionamentos Espaciais: possibilidade de diferenciação dos relacionamentos espaciais, incluindo restrições de integridade espacial.
- Aspectos Temporais: possibilidade de modelagem das características temporais dos dados.
- Aspectos de Qualidade: possibilidade de modelagem de metadados de qualidade.

A seguir, cada um desses requisitos é apresentado em maiores detalhes, juntamente com a contribuição de cada modelo conceitual para SIG.

4.2.1 Fenômeno Geográfico e Objeto Convencional

Normalmente, em um banco de dados geográficos existem, além dos dados referentes a fenômenos georreferenciados (possui referência a sua localização em relação à superfície da Terra), objetos convencionais presentes em qualquer sistema de informação. Por exemplo, uma fazenda é um fenômeno geográfico se no banco de dados estão armazenadas suas informações espaciais como os limites da fazenda. Neste mesmo banco de dados pode-se ter dados sobre proprietários de fazendas, considerados objetos convencionais por não terem informações espaciais associadas. Em um esquema conceitual é importante que se possa diferenciar, facilmente, entre classes (ou entidades) descrevendo esses dois tipos de objetos.

O modelo GMOD [PIR 97] distingue, através de especialização, classes que representam entidades associadas a alguma localização geográfica (*Geo-Classe*) e classes que não possuem referência espacial (Classes Convencionais), mas que estão relacionadas com as primeiras. O modelo MGeo+ [PIM 95] utiliza uma abordagem semelhante.

Alguns modelos diferenciam os fenômenos geográficos dos objetos convencionais através da inclusão de pequenos símbolos gráficos, denominados pictogramas, nas classes que descrevem os fenômenos geográficos. A ausência de um pictograma significa que a classe modela um objeto convencional. Exemplos de modelos que seguem esta abordagem incluem GeoOOA [KOS 97], MADS [PAR 99], OMT-G [BOR 99] e Perceptory [BED 99].

Lisboa e Iochpe [LIS 99] definem um conjunto de estereótipos específicos para diferenciar objetos convencionais dos fenômenos geográficos, estes últimos ainda são especializados entre campos e objetos geográficos, como descrito na seção seguinte. A Figura 3 ilustra o uso de estereótipos definidos de acordo com o modelo UML-GeoFrame [LIS 99]. Alguns estereótipos são usados para modelar o tipo de representação espacial dos fenômenos geográficos e estão descritos na Seção 4.2.4.

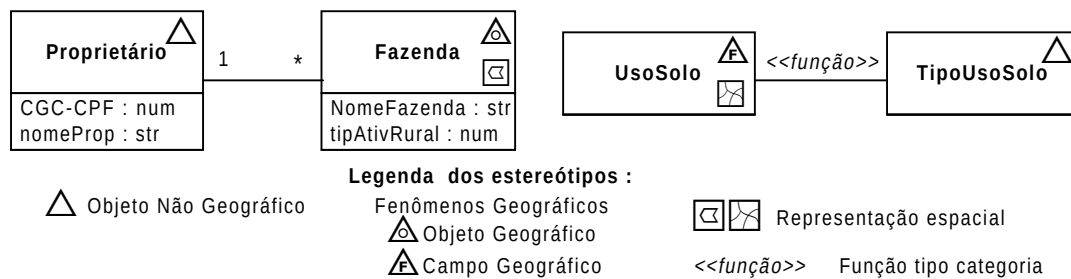


FIGURA 3 - Diferenciando fenômenos geográficos e objetos convencionais

4.2.2 Visões de Campo e de Objetos

Segundo Burrough [BUR 95], os modelos de dados geográficos devem refletir a maneira como as pessoas vêem o mundo. Um dos princípios filosóficos da percepção humana dos fenômenos geográficos é que a realidade é composta de entidades exatas e de superfícies contínuas.

Para Goodchild [GOO 92], a realidade geográfica pode ser observada segundo duas visões: de campo e de objetos. Na visão de campo, a realidade é modelada por variáveis que possuem uma distribuição contínua no espaço. Toda posição no espaço geográfico pode ser caracterizada através de um conjunto de atributos como, por exemplo, temperatura, tipo de solo e relevo, medidos para um conjunto de coordenadas geográficas. Já na visão de objetos, a realidade consiste de entidades individuais, bem definidas e identificáveis. Cada entidade tem suas propriedades e ocupa um determinado lugar no espaço. A realidade é modelada como um grande espaço onde entidades estão distribuídas sem que, necessariamente, todas as posições do espaço estejam ocupadas. Duas ou mais entidades podem estar situadas sobre uma mesma posição geográfica.

A classificação de fenômenos geográficos na visão de objetos é um processo natural e direto (ex.: rios são descritos pela classe *Rio*). No entanto, um campo geográfico como, por exemplo, *Altimetria*, não pode ser modelado diretamente como uma classe, pois *Altimetria* não é um objeto e, conseqüentemente, não pode ser descrito diretamente por uma classe.

Para Kemp [KEM 98], existem muitas questões filosóficas sobre o entendimento de campos geográficos. Por exemplo, os campos só existem quando são medidos? Os campos possuem características próprias? A atribuição de um determinado valor a uma região é real, ou é apenas um mecanismo conceitual direcionado a um objetivo específico? Apesar de estas questões não terem ainda uma resposta consensual, a definição de um campo como uma função em um domínio, o qual é um subconjunto do espaço-tempo, é suficientemente específico para os propósitos computacionais.

Desta forma, a modelagem conceitual dos fenômenos geográficos necessita de construtores especiais para modelar tanto os campos quanto os objetos geográficos. A maioria dos modelos existentes não suporta a modelagem adequada dos fenômenos geográficos que são percebidos na visão de campo. Por outro lado, todos os modelos estudados suportam a modelagem dos fenômenos na visão de objetos.

No modelo GMOD [PIR 97], a classe *Geo-Classe* (descrita na seção anterior) é especializada nas subclasses *Geo-Campo* e *Geo-Objeto*, ambas formando duas hierarquias de classes que representam fenômenos geográficos observados segundo as visões de campo e de objetos, respectivamente.

O modelo GeoIFO [TRY 95, HAD 96] permite a modelagem, de forma integrada, dos fenômenos nas visões de campo e de objetos, através dos conceitos de *posição de*

objetos e de *atributos variando no espaço*. Segundo Tryfona [TRY 95], objetos do mundo real estão ligados ao espaço através de sua posição. A *posição* de um objeto inclui sua localização, forma, tamanho e orientação. *Atributos variando no espaço* descrevem fenômenos do mundo real que estão relacionados diretamente ao espaço e não pertencem a nenhum objeto particular, sendo propriedades do espaço. Por exemplo, segundo o modelo Geo-IFO, em um sistema de transporte de combustível, o tipo de solo do terreno por onde passa um oleoduto pode ser modelado como um atributo da classe *Oleoduto*. A Figura 4 apresenta um exemplo de modelagem integrada de *posição de objetos* (ex.: posição de um oleoduto é representada por um objeto de duas dimensões, isto é, uma linha) e de *atributos variando no espaço* (ex.: tipo de solo é uma função do Espaço no domínio *Tipo-Solo*).

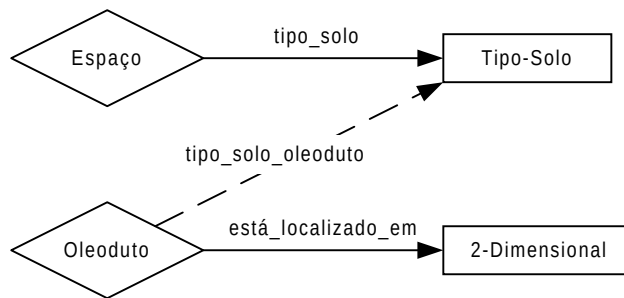


FIGURA 4 - Um exemplo de modelagem no modelo Geo-IFO

O modelo OMT-G [BOR 97] diferencia campos e objetos geográficos através do uso de dois conjuntos distintos de pictogramas, um para a representação espacial de campos e outro conjunto para a representação de objetos geográficos. O modelo GeoOOA [KOS 97] propõe uma solução na qual os conceitos de campo geográfico e modelo de representação *raster* são usados de forma ambígua. O modelo UML-GeoFrame [LIS 99] suporta esses dois primeiros requisitos de forma unificada, conforme mostrado na Figura 3.

4.2.3 Aspectos Temáticos

Os fenômenos geográficos possuem vários atributos, dentre os quais o que fornece sua localização geográfica. A localização e a forma dos fenômenos geográficos são representadas através de objetos espaciais, associados a um sistema de coordenadas. Uma entidade geográfica (ex.: um rio), deve estar associada a um ou mais objetos espaciais que representam sua localização e sua forma sobre a superfície terrestre. Em um SIG, as representações espaciais das entidades geográficas não são tratadas isoladamente, mas sim, em grupos de representações de entidades com características e relacionamentos em comum. Diferentes termos como *tema*, *camada* e *plano de informação* são encontrados na literatura, referindo-se a estes agrupamentos.

No nível interno de um SIG, a organização dos dados espaciais é feita em camadas físicas (ex.: conjunto de polígonos representando os limites dos lotes urbanos). No entanto, embora o projeto de camadas físicas seja um problema a ser tratado nas etapas de projeto lógico ou físico, diversos autores afirmam que é importante que camadas conceituais sejam definidas durante a fase de projeto conceitual [HAD 96, BED 96, EUR 96]. Segundo Hadzilacos [HAD 96], camadas conceituais não necessitam ter um relacionamento um-para-um com camadas físicas. Por exemplo, uma única camada conceitual, como *Hidrografia*, pode ser armazenada em diferentes camadas físicas, uma contendo somente os rios, outra contendo somente os lagos, etc. Neste

estudo adotou-se a nomenclatura utilizada em [EUR 96], onde uma camada conceitual recebe a denominação de tema.

Os modelos MGeo+ [PIM 95] e GMOD [PIR 97] permitem a modelagem de diferentes visões da área de estudo, correspondendo a temas, através do conceito de planos de informação.

De acordo com o GeoFrame [LIS 99], a área geográfica de interesse de uma aplicação de SIG pode ser retratada através de diversos temas, os quais agrupam conjuntos de classes fortemente relacionadas entre si. A Figura 5 mostra um conjunto de temas identificados e modelados formando uma hierarquia de temas. Temas são modelados através do construtor UML denominado pacote [BOO 98]. Ferramentas CASE para SIG estão adotando mecanismo semelhante. Exemplos incluem Perceptory [BED 99] e REGIS [ISO 99]. O modelo OMT-G [BOR 97] propõe uma notação própria para a modelagem hierárquica de temas.

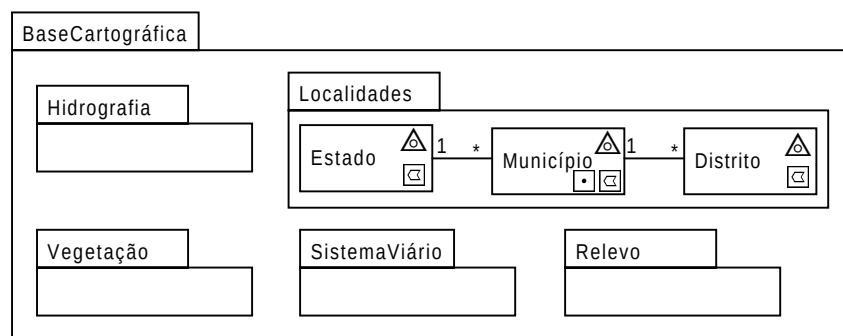


FIGURA 5 - Modelando temas segundo o modelo UML-GeoFrame

4.2.4 Aspectos Espaciais

Segundo Worboys [WOR 95], um fenômeno geográfico possui quatro dimensões ao longo das quais os atributos são medidos: *espacial*, *gráfica*, *temporal* e *textual/numérica*. Por exemplo, um objeto *município* pode conter os seguintes atributos: um polígono representando os limites municipais (*espacial*); um polígono e um símbolo gráfico (ex.: círculos concêntricos proporcionais à população) representando sua forma cartográfica em diferentes escalas (*gráfico*); data de emancipação e data em que os dados do município foram incluídos no sistema (*temporal*); e atributos descritivos como nome e população (*textual/numérico*). Portanto, os aspectos espaciais estão relacionados com a forma e localização dos fenômenos geográficos.

A localização geográfica e a abstração da forma espacial dos fenômenos geográficos são representadas, no SIG, através de objetos espaciais. Todo objeto espacial possui uma geometria, que representa a forma espacial do fenômeno, sendo que suas coordenadas devem estar registradas com base em um determinado sistema de coordenadas (ex.: latitude e longitude) e uma projeção (ex.: projeção UTM).

A experiência dos autores mostra que, embora os aspectos espaciais dos fenômenos geográficos possam parecer detalhes de mais para serem tratados durante a modelagem conceitual, a sua inclusão no esquema conceitual tem sido fator fundamental na comunicação com o usuário.

Na visão de objetos, os fenômenos geográficos são representados por objetos espaciais do tipo ponto, linha, polígono ou combinações destes, enquanto que na visão de campo uma superfície contínua pode ser representada, por exemplo, através de modelos numéricos, conjuntos de isolinhas, polígonos adjacentes e grade de células. Um SIG fornece diferentes tipos de modelos de representação, os quais permitem ao usuário

representar os diversos tipos de fenômenos geográficos. O modelo de representação a ser usado depende da finalidade da aplicação e das características do fenômeno.

Diferentes abordagens têm sido propostas para modelagem conceitual dos aspectos espaciais dos fenômenos geográficos. A abordagem mais comum é a que define uma associação entre a classe que descreve o fenômeno e a classe do tipo de objeto espacial correspondente a sua representação espacial. Exemplos de modelos que seguem esta abordagem são GISER [SHK 97], GeoIFO [TRY 95], Geo-ER [HAD 97] e GMOD [PIR 97]. Um dos problemas que ocorre nesta abordagem é a sobrecarga do esquema de dados resultante, devido ao grande número de novas associações.

Para obter esquemas de dados mais legíveis, alguns modelos utilizam pictogramas para substituir esta associação. Entre os modelos que utilizam pictogramas pode-se citar o Modul-R [BED 96], que foi o pioneiro, GeoOOA [KOS 97], OMT-G [BOR 99], MADS [PAR 99], UML-GeoFrame [LIS 99], além das ferramentas CASE Perceptory [BED 99] e REGIS [ISO 99]. Na Figura 3 duas classes (*Fazenda* e *UsoSolo*) apresentam pictogramas de representação espacial, definidos como estereótipos no modelo UML-GeoFrame [LIS 99].

Alguns modelos pioneiros, como o modelo de Clementini [CLE 94], induziam o projetista a cometer um equívoco de modelagem dos aspectos espaciais, que é o de modelar classes de fenômenos geográficos (ex.: classe *Rio*) como subclasse de classes de objetos espaciais (ex.: classe *Linha*).

4.2.5 Múltiplas Representações

Uma das características das aplicações geográficas é a possibilidade de existência de múltiplas representações para um mesmo fenômeno geográfico. Esta necessidade surge em resposta à complexidade da realidade a ser representada e às diferentes visões que os usuários têm de um mesmo fenômeno. Um fenômeno geográfico pode ser representado em diferentes escalas ou projeções, inclusive por diferentes objetos espaciais.

Múltiplas representações são modeladas através da inclusão de várias associações entre o fenômeno geográfico e os tipos de objetos espaciais correspondentes. Por exemplo, na Figura 5 a classe *Município*, contém dois tipos de representação espacial: pontual e poligonal.

Uma classificação de diferentes tipos de combinações possíveis de pictogramas é apresentada por Bédard e incluída no modelo Perceptory [BED 99]. Segundo Bédard, um fenômeno geográfico pode apresentar representação espacial com forma complexa, alternativa, múltipla, derivada ou complicada (explicada textualmente em um dicionário). Em todos os modelos citados na seção anterior é possível modelar múltiplas representações, embora muitos não descrevam isso de forma explícita.

4.2.6 Relacionamentos Espaciais

Uma das tarefas mais importantes quando se está modelando os dados de uma aplicação é a identificação de quais os relacionamentos que deverão ser mantidos no banco de dados, dentre os possíveis relacionamentos observáveis na realidade. No domínio das aplicações geográficas este problema é bem complexo, uma vez que o número de relacionamentos possíveis de serem mantidos no banco de dados é ainda maior, devido à existência dos relacionamentos espaciais entre os fenômenos geográficos.

Diversos tipos de relacionamentos espaciais são citados na literatura [EGE 89, LAU 92, WOR 95]. São eles:

- relacionamentos métricos: incluem os possíveis relacionamentos entre os atributos espaciais (geométricos) das entidades. Um exemplo é o relacionamento de distância entre coordenadas geográficas, que permite a execução de operações com base no conceito de proximidade;
- relacionamentos topológicos: propriedades topológicas são aquelas que se mantêm inalteradas após distorções (ex.: mudança de projeção). Assim, os relacionamentos topológicos referem-se aos relacionamentos que não dependem exclusivamente das coordenadas dos objetos (ex.: adjacência);
- relacionamentos de composição: ocorrem quando o componente espacial de um objeto é composto de outros objetos espaciais (objeto complexo). Por exemplo, um arquipélago é composto de um conjunto de ilhas.

A maioria dos SIG fornece estruturas especiais para o armazenamento explícito de alguns tipos de relacionamentos espaciais (normalmente são mantidos somente os relacionamentos de adjacência ou de conectividade), deixando os demais para ser calculados a partir das coordenadas espaciais dos objetos, durante a execução das operações de consulta. Muitas aplicações não necessitam manter os relacionamentos topológicos, embora eles sempre existam. Por outro lado, existem aplicações nas quais esses relacionamentos possuem significado semântico relevante, tal como vizinhança e cruzamento de ruas. Nestes casos o projetista necessita especificar estas informações no modelo de dados.

As cardinalidades associadas aos relacionamentos formam um conjunto de restrições de integridade que devem ser mantidas entre as instâncias dos objetos no banco de dados. Para os relacionamentos espaciais, novos tipos de cardinalidade podem ser definidos, tais como, associação espacial, pertinência, etc. Um modelo conceitual de dados para SIG deve fornecer meios para que o projetista represente os relacionamentos a serem mantidos no banco de dados geográficos. Isto inclui tanto os relacionamentos convencionais como os relacionamentos espaciais.

A maioria dos modelos estudados utiliza os próprios construtores de modelagem de relacionamentos já disponíveis no formalismo de origem. No entanto, alguns modelos definem novos construtores, como descrito a seguir.

O modelo GMOD [PIR 97] define dois novos tipos especiais de relacionamentos: *de versão* e *causal*. Relacionamentos de versão são usados para associar diferentes versões de uma mesma entidade, enquanto um relacionamento causal estabelece uma ligação de causa-efeito entre dois fenômenos modelados.

O modelo OMT-G [BOR 97] diferencia os relacionamentos espaciais dos relacionamentos convencionais através de novos símbolos gráficos. São definidos construtores abstratos para modelar alguns tipos de relacionamentos topológicos como, por exemplo, estruturas “todo-parte” e de rede. Um conjunto de 29 regras de restrições de integridade é apresentado em [BOR 99].

As novas primitivas do modelo GeoOOA [KOS 97] suportam abstrações de estruturas todo-parte espacial, diferenciando três tipos de estruturas: *cobertura*, *pertinência* e *partição*. Construtores específicos para modelagem de elementos de uma rede também são fornecidos pelo modelo GeoOOA.

De acordo com o modelo MADS [PAR 99], os relacionamentos espaciais, assim como a existência de atributos espaciais, são modelados através de linhas tracejadas.

4.2.7 Aspectos Temporais

A maioria dos SIG disponíveis atualmente considera as entidades como se o mundo existisse somente no presente. Informações geográficas são incluídas e alteradas ao longo do tempo, mas o histórico dessas transformações não é mantido no banco de

dados. Segundo Peuquet [PEU 95], estas limitações dos SIG vêm recebendo atenção crescente, devido à necessidade de um melhor entendimento dos processos geográficos e dos inter-relacionamentos de causa e efeito entre as atividades humanas e o meio ambiente.

Para Hadzilacos [HAD 96], a necessidade de os dados geográficos estarem qualificados com base no tempo, não se deve ao fato de os dados serem freqüentemente modificados, mas sim, à necessidade de se registrar estados passados de forma a possibilitar o estudo da evolução dos fenômenos geográficos. Para possibilitar uma análise de dados com base na evolução dos fenômenos geográficos é necessário adicionar aos SIG as potencialidades dos sistemas de bancos de dados temporais. Isto implica na inclusão de informações temporais relacionadas aos fenômenos geográficos e da extensão das linguagens de consulta disponíveis para que estas suportem cláusulas de condição associadas a aspectos temporais [EDE 94, WOR 95].

Alguns modelos estendem a notação gráfica para ressaltar a existência dos aspectos temporais em determinadas classes e/ou atributos. No modelo GMOD [PIR 97], Geo-Classes e classes Convencionais podem ser definidas como sendo temporais, através da definição de uma associação com a classe *Tempo*. A classe *Tempo* é raiz de uma outra hierarquia de classes, onde suas subclasses distinguem os tipos de características do tempo (ex.: discreto ou contínuo, variação linear ou intervalar).

O modelo GeoOOA utiliza um pictograma especial, que é um símbolo de um relógio, para diferenciar classes temporais (Figura 6). Além disso, a notação gráfica do modelo OOA [COA 91] é estendida para representar dois tipos especiais de relacionamentos temporais: *conexão ancestral* e *conexão de atributo temporal*. Um relacionamento de conexão ancestral permite a associação entre versões de um mesmo objeto. Por exemplo, um lote pode ser dividido, em um determinado momento, dando origem a dois novos lotes. A conexão de atributo temporal é usada para destacar uma associação decorrente da necessidade de modelar o aspecto temporal de um atributo. Por exemplo, na Figura 6, a classe *Escritura* é resultado da modelagem dos dados históricos do lote. Em uma modelagem não temporal, o usual é que cada lote tenha uma única associação com o proprietário atual.

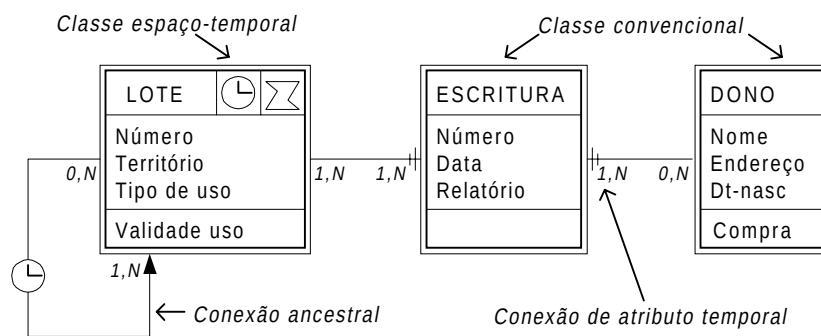


FIGURA 6 - Exemplo de modelagem de aspectos temporais no modelo GeoOOA

O modelo Modul-R também utiliza pictogramas para modelagem dos aspectos temporais. Provavelmente, devido à complexidade da solução apresentada por Bédard em [BED 96], esta solução foi abandonada na implementação da ferramenta CASE Perceptory [BED 99].

4.2.8 Aspectos de Qualidade

Dados com erros podem surgir nos SIG, mas precisam ser identificados e tratados. Os erros podem ser introduzidos no banco de dados de diversas formas: serem

decorrentes de erros já existentes nas fontes originais dos dados; serem adicionados durante os processos de obtenção (captura) e armazenamento dos dados; serem gerados durante a exibição ou impressão dos dados; ou surgirem a partir de resultados errôneos de operações de análise de dados [BUR 95].

Uma aplicação de SIG se diferencia das demais aplicações de banco de dados por fazer uso de grande volume de dados importados de outros sistemas, muitas vezes de outras organizações. Este fenômeno gera, freqüentemente, problemas de confiabilidade dos dados.

O padrão americano para troca de dados espaciais (SDTS) requer um relatório de qualidade, o qual fornece uma base para que o usuário faça seu próprio julgamento sobre a qualidade do dado importado [ASP 96]. Os aspectos de qualidade fazem parte de um conjunto maior de dados que são elaborados para possibilitar o compartilhamento de dados em aplicações de SIG. Diversas tentativas de padronização do formato de descrição de dados, denominados metadados, são encontradas na literatura [WEB 98].

No tocante a projeto conceitual, porém, pouco avanço tem sido observado. Os modelos Modul-R [BED 96] e GISER [SHK 97] apenas citam a necessidade de se manter dados de qualidade, mas não descrevem como isso é feito. Ram e outros [RAM 99] apresentam uma proposta de uso de metadados com o objetivo de acesso aos dados através de um modelo semântico (USM*), ao invés de comandos disponíveis em SIG.

5 CONSIDERAÇÕES FINAIS

A partir dos diversos modelos analisados observa-se que todos seguiram um mesmo caminho, onde se buscou acrescentar novas primitivas de abstração a um dos formalismos conceituais existentes com variações apresentadas, principalmente, na sintaxe da linguagem e na notação gráfica proposta.

No entanto, a maioria desses modelos, ao invés de realmente introduzir novas primitivas de modelagem, se restringe a descrever como e o que modelar, da realidade geográfica, utilizando-se quase sempre de primitivas de modelagem do próprio formalismo de origem.

Observa-se que alguns modelos já atendem, em maior ou menor grau, aos requisitos de modelagem identificados. Coloca-se, então, a pergunta: Por que os projetistas de SIG relutam, ainda, em realizar a fase de modelagem conceitual em seus projetos? Acredita-se que isto ocorra, não mais pela ausência de um modelo adequado para as aplicações de SIG, mas sim, devido a deficiências nas metodologias de desenvolvimento em que esses modelos são empregados. Uma das possíveis soluções para este problema é a utilização de metodologias de desenvolvimento de sistemas baseadas em abordagens de reutilização [LIS 98].

O estudo mostra que alguns dos requisitos analisados ainda necessitam ser melhor pesquisados como, por exemplo, os aspectos temporais e os aspectos de qualidade. Os modelos analisados também permitem concluir que, nos casos onde se buscou a definição de um modelo muito rico semanticamente, aumentou a dificuldade de se utilizar o modelo, devido ao grande número de primitivas (ex.: Modelo MADS).

2.3 Artigo II: Mecanismos de Reutilização em Sistemas de Informação

Jugurta Lisboa Filho e Cirano Iochpe

Submetido à **Revista Acta Scientiae**, Canoas-RS, 1999.

RESUMO

Reutilizar, quando bem feito, pode implicar no aumento de produtividade e qualidade em todas as etapas do desenvolvimento de software. Este artigo tem como objetivo identificar os principais mecanismos de reutilização aplicados nas etapas iniciais do processo de desenvolvimento de sistemas de informação (*early reuse*). Inicialmente são descritas algumas pesquisas sobre reutilização de especificação de requisitos desenvolvidas, principalmente, no início da década de 90. Em seguida são descritos novos mecanismos surgidos para auxiliar na reutilização de projetos orientados a objetos como padrões (*patterns*), com ênfase nos padrões de análise e *frameworks*. O estudo também descreve os mecanismos atuais de apoio à reutilização na área de Sistemas de Informação Geográfica (SIG).

Palavras-chave: Reutilização, Padrões de análise, *Framework*, SIG.

ABSTRACT

When reuse is well done it can results in productivity and quality increase in all the software development phases. This paper identifies and presents the mainly reuse mechanisms that may be applied in the early stage of the information system development. Researches about requirement analysis that were developed in the 90's are described. After that, new mechanisms that support reuse in object-oriented software design are investigated. It includes design and analysis patterns as well as frameworks. The paper also describes current mechanisms that support the reuse in Geographic Information Systems (GIS).

Keywords: Reuse, Analysis Patterns, *Framework*, GIS.

1 INTRODUÇÃO

A idéia de reutilização é inerente ao desenvolvimento de sistemas computacionais. Durante as décadas de 60 e 70, a idéia de reutilização centrava-se, basicamente, na reutilização de códigos de programas (ou subprogramas). Bibliotecas de funções, escritas em diversas linguagens de programação, foram criadas e disponibilizadas para que pudessem ser reutilizadas. A busca por mecanismos voltados a facilitar a reutilização de software teve como resultado uma série de eventos científicos dedicados ao tema. Em 1983 ocorreu o primeiro grande congresso na área, o *Workshop on Reusability in Programming* (Perlis, 1983 Apud [NEI 94]).

Como um dos principais instrumentos de aumento da produtividade dos programadores e da qualidade dos sistemas desenvolvidos, a idéia de reutilização

sempre esteve associada à criação de mecanismos que possibilitem a administração da complexidade dos sistemas. Por exemplo, o conceito de Tipo Abstrato de Dados (TAD) foi desenvolvido para possibilitar que estruturas de dados pudessem ser reutilizadas [EMB 87]. Para usar um TAD, o programador necessita conhecer apenas a interface e a funcionalidade de cada um dos subprogramas disponíveis. Outro exemplo de instrumento voltado à reutilização é o paradigma da orientação a objetos que tem, como um de seus princípios fundamentais, possibilitar a reutilização de classes existentes através de mecanismos como especialização e polimorfismo. Atualmente, na área de Engenharia de Software, pesquisas sobre reutilização exploram conceitos como *framework* e padrões [COL 96].

Em cada uma das fases do ciclo de vida de um sistema, existe a possibilidade de se reutilizar resultados obtidos anteriormente. Um exemplo que pode-se citar é o emprego de bibliotecas de funções de interface homem-máquina, durante a fase de codificação dos programas. O objetivo deste artigo é revisar alguns dos principais instrumentos que podem ser aplicados nas tarefas desenvolvidas nas fases iniciais do ciclo de vida.

O restante do artigo está organizado como segue. A Seção 2 apresenta os principais avanços obtidos pela pesquisa em reutilização de especificação de requisitos. Na Seção 3 é descrito o conceito de padrões e relacionados diversos tipos de padrões existentes. A Seção 4 descreve a técnica de reutilização baseada em *frameworks*. A Seção 5 descreve as abordagens de reutilização existentes na área de Sistemas de Informação Geográfica (SIG). Por último, a Seção 6 apresenta comentários, deficiências e vantagens das técnicas apresentadas.

2 REUTILIZAÇÃO DE ESPECIFICAÇÃO DE REQUISITOS

Nesta seção são apresentados alguns resultados de pesquisas sobre reutilização no campo da Engenharia de Software, pesquisas estas desenvolvidas até meados da década de 90.

Segundo Neighbors [NEI 94], no processo de desenvolvimento de software, quanto mais cedo forem identificados e criados recursos reutilizáveis, maior será o impacto da reutilização nas fases posteriores. Portanto, a reutilização de modelos durante a especificação de requisitos é o primeiro momento em que o projetista pode empregar algum tipo de recurso reutilizável, isto é, qualquer recurso existente que possa ser útil na construção de um sistema computacional. Para Edelweiss [EDE 94], o uso de abordagens de reutilização durante a fase de especificação de requisitos apresenta as seguintes vantagens:

- redução do custo do desenvolvimento da especificação;
- redução do custo de verificação e validação da especificação;
- aumento da produtividade no desenvolvimento de especificações;
- aumento da qualidade das especificações;
- padronização de especificações;
- facilidade de comunicação entre equipes que utilizam a mesma biblioteca.

Zirbes [ZIR 95] demonstrou, através de experimentos realizados com diferentes técnicas de análise de requisitos, que a modelagem baseada na reutilização de especificações de sistemas análogos, produz resultados mais completos e corretos do que a modelagem sem o apoio da reutilização sistemática. Os resultados da experiência desenvolvida por Zirbes mostram que os ganhos obtidos com o emprego da reutilização não estão apenas em ganhos de qualidade, mas também em produtividade onde, embora o tempo de desenvolvimento medido durante o experimento não tenha apresentado um

ganho muito significativo, o tempo total dedicado ao sistema é reduzido por diversos fatores indiretos, dentre eles, a redução no tempo de conhecimento do domínio e do tempo de manutenção, decorrentes de projetos de melhor qualidade.

Os objetivos da pesquisa sobre reutilização são “reunir, organizar e facilitar o acesso a todo o conhecimento sobre o desenvolvimento de software em um determinado domínio” [ZIR 95]. Desta forma, um fator fundamental em qualquer abordagem baseada em reutilização é a existência de um conjunto de recursos que possam ser reutilizados. Com isto, pressupõe-se a existência de um repositório (biblioteca) para disponibilizar esses recursos. Dois dos principais problemas que surgem na abordagem de reutilização, são: (1) a escolha de um método para classificação e organização dos recursos disponíveis; (2) a existência de mecanismos que auxiliem o projetista a localizar os possíveis recursos reutilizáveis em seu projeto.

Uma forma simples de se organizar uma biblioteca de recursos reutilizáveis é por meio de domínios. Um domínio, no contexto da Engenharia de Software, engloba um conjunto de aplicações caracterizadas por um conjunto comum de dados e objetivos relacionados [PET 91]. A descrição do recurso reutilizável, adicionada às informações auxiliares para localizá-lo e manipulá-lo, constituem a infra-estrutura básica para o processo de reutilização. Portanto, o processo de desenvolvimento de software com base em reutilização deve considerar algumas etapas adicionais, capazes de manter e consultar a infra-estrutura de recursos disponíveis. Segundo Zirbes [ZIR 95], o processo de reutilização constitui-se das seguintes etapas (Figura 1):

- construção da biblioteca de recursos: esta atividade inclui a identificação, escolha, classificação e armazenamento dos recursos reutilizáveis;
- reutilização dos recursos da biblioteca: esta atividade inclui a busca, seleção e adaptação dos recursos utilizáveis, de acordo com as necessidades do software sendo desenvolvido.

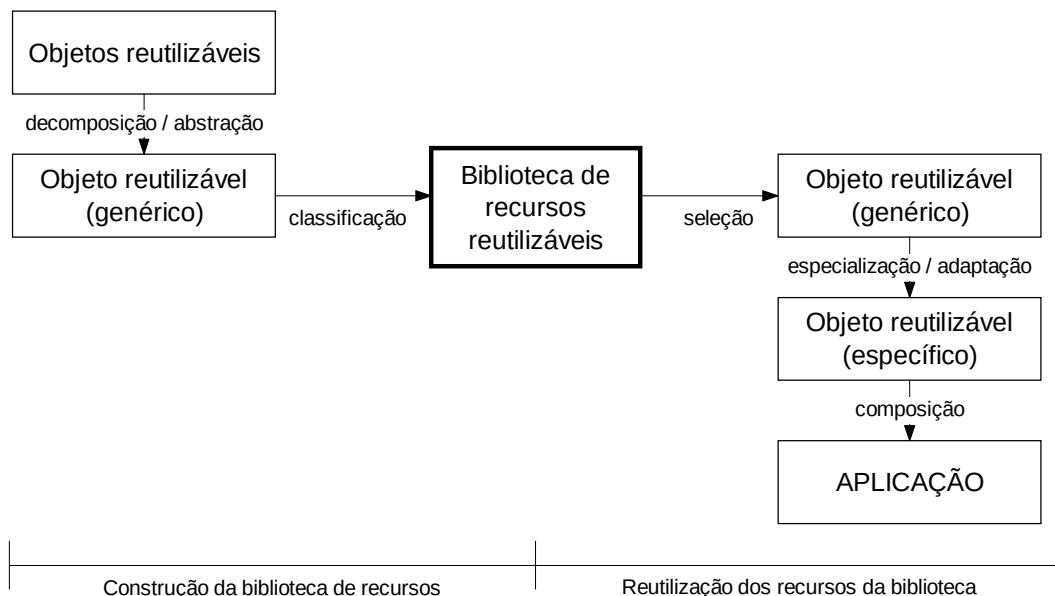


FIGURA 1 - Processo de reutilização de software

O processo de seleção de recursos reutilizáveis torna-se cada vez mais complexo, à medida em que a biblioteca aumenta de tamanho. Assim, é necessário o uso de métodos que auxiliem a localização de recursos reutilizáveis. Um deles, o método de seleção por analogia, foi desenvolvido por Maiden em sua tese de doutorado [MAI 92].

O processo de reutilização usando analogia possui duas etapas. Na primeira etapa é identificado um conjunto de domínios abstratos representando famílias de sistemas com características afins. Na segunda etapa é aplicado o raciocínio por analogia para reconhecer a semelhança entre o problema a ser definido e os domínios existentes, compreender esta semelhança e então, transferir o conhecimento sobre o domínio.

Além do método de seleção por analogia, uma série de outros métodos que auxiliam a localização de recursos reutilizáveis são descritos na literatura [GIR 90]. Alguns destes métodos utilizam técnicas de inteligência artificial para auxiliar o projetista a identificar e selecionar recursos reutilizáveis nos sistemas sendo desenvolvidos (Wood, 1988 Apud [PIM 98]). Ferramentas automatizadas são utilizadas para dar apoio ao uso desses métodos. Entre elas pode-se citar: o sistema Catalog [FRA 87] que permite classificar recursos reutilizáveis através do método de classificação de software baseado em vocabulário livre; o sistema ROSA - *Retrieval Of Software Artifacts* [GIR 96] propõe a classificação e recuperação automática baseada em vocabulário extraído de textos escritos em linguagem natural; a ferramenta CART - *Computer-Aided Reuse Tool* [LIA 93] utiliza um método de classificação baseado em facetas, isto é, grupos de classes elementares formando um esquema, para organizar uma biblioteca de software orientado a objetos.

Devido ao uso crescente do paradigma de orientação a objetos no desenvolvimento de software, novos instrumentos de reutilização têm sido propostos na literatura. As seções seguintes apresentam as tendências atuais em termos de abordagens de reutilização.

3 PADRÕES

A abordagem de padrões³ é ainda uma idéia muito recente no campo da Ciência da Computação. Desta forma, não existe um consenso sobre uma definição para o termo “padrão” neste contexto. Segundo Gamma [GAM 94], o uso de padrões está diretamente relacionado com a idéia de reuso em Engenharia de Software e teve origem no trabalho do arquiteto Christopher Alexander que, em [ALE 77], definiu:

“Cada padrão descreve um problema que ocorre diversas vezes em nosso ambiente e, então, descreve a solução desse problema de tal forma que você a reutilize milhares de vezes sem ter que fazer a mesma coisa duas vezes” (pag. x).

As idéias de Alexander deram origem ao desenvolvimento de pesquisas sobre padrões de projeto que, posteriormente, foram estendidas para diversos outros tipos de padrões.

A abordagem de padrões começou a ser conhecida após a publicação do livro *Design patterns: elements of reusable object-oriented software* [GAM 94], que apresenta um catálogo de soluções para problemas recorrentes em projeto de sistemas orientados a objetos. Neste livro, Erich Gamma, Richard Helm, Ralph Johnson e John Vlissides, que se tornaram conhecidos como a *gang* dos quatro (“*gang of four*” ou GoF), definem:

“Um padrão apresenta a essência de uma solução para um problema recorrente, em um contexto específico” (pag. 2-3).

³ Do inglês *Pattern*.

Esta definição compreende as idéias fundamentais de um padrão. A expressão “uma solução para um problema” significa que cada padrão identifica um problema e apresenta uma solução para ele. O termo “essência de uma solução” significa que somente os elementos essenciais são descritos, deixando os aspectos específicos para serem detalhados pelo projetista, dado que aspectos específicos normalmente não são reutilizados. O termo “problema recorrente” significa que os padrões devem ser descritos para problemas que já ocorreram diversas vezes e irão ocorrer novamente. Por último, “em um contexto específico” significa que a solução completa é válida para um contexto particular.

Buschmann [BUS 96], classifica os padrões em três categorias: padrões de arquitetura; padrões de projeto; e idiomas (também conhecido por padrões de implementação [PRE 95]). As duas primeiras categorias incluem os padrões relacionados com a fase de projeto do sistema, mas são aplicadas em problemas de diferentes escalas. Os padrões de projeto são mais abstratos (e menores) do que os padrões de arquitetura, enquanto que idiomas são padrões escritos em alguma linguagem de programação.

Uma quarta categoria de padrões, introduzida por Fowler [FOW 96], inclui os padrões de análise. Estes padrões são usados para descrever soluções empregadas durante as fases de análise de requisitos e modelagem conceitual dos dados. Os padrões de análise refletem estruturas conceituais representativas do domínio da aplicação e não soluções computacionais.

Fowler [FOW 97] define um padrão de análise como:

“Uma idéia que se provou útil em um contexto prático e que, provavelmente, será útil em outras situações similares” (pag xv).

A maioria dos padrões de análise publicados até o momento foram projetados, principalmente, para solucionar problemas de aplicações comerciais [COA97, FOW97, HAY95]. No entanto, a idéia de padrões de análise pode ser usada para aumentar a qualidade e a produtividade no desenvolvimento de aplicações não-convencionais como, por exemplo, as aplicações de SIG. Por exemplo, os padrões de análise podem ser usados para documentar como os projetistas de um determinado domínio de aplicação modelam fenômenos geográficos e seus relacionamentos. Padrões de projeto podem ser usados para documentar como os dados sobre o relevo de uma região são representados através de um Modelo Numérico de Terreno (MNT). Por último, idiomas podem definir como um MNT deve ser implementado em cada software de SIG específico. As subseções seguintes descrevem, em maiores detalhes, as características dos padrões de análise e de projeto.

3.1 Padrões de Análise

Devido à abordagem de padrões ter sido introduzida na área de Engenharia de Software, por projetistas de sistemas orientados a objetos, a categoria de padrões mais conhecida atualmente é a dos padrões de projeto. No entanto, diversas pesquisas vêm sendo realizadas buscando-se padrões que auxiliem projetistas de banco de dados a reutilizar soluções para problemas de modelagem de dados [COA97, HAY95, FOW97].

Segundo Coad [COA 92], as técnicas tradicionais de reutilização em análise orientada a objetos baseiam-se nos mecanismos de herança. Assim, para reutilizar classes pré existentes, o projetista necessita saber como localizar as classes adequadas a seu problema e derivar novas subclasses a partir dessas classes. Dessa forma, durante a

modelagem de uma nova aplicação, cada classe funciona como um simples bloco reutilizável.

Um padrão de análise (ou padrão conceitual, na terminologia de Riehle [RIE 96]) descreve um conjunto de classes, possivelmente pertencentes a diferentes hierarquias de classes, e as associações existentes entre elas. Padrões de análise podem ser vistos, portanto, como uma forma de descrever subesquemas de projetos mais complexos, os quais ocorrem com frequência durante o processo de modelagem de muitas aplicações. O uso de padrões melhora, de forma significativa, o tempo de desenvolvimento de novas aplicações, uma vez que a reutilização ocorre através de subesquemas e não através de classes isoladas [HAY 96].

Geralmente, um padrão de análise apresenta a solução do problema de uma forma mais sugestiva do que prescritiva, fornecendo um modelo e a discussão do por que a solução é proposta desta forma, suas vantagens e desvantagens. Segundo Fowler [FOW 96], a contribuição realmente importante de um padrão não é o modelo fornecido como solução, mas sim, o raciocínio que está por trás desta solução.

Hay [HAY 95] descreve um conjunto de padrões de análise, que ele batizou de padrões de modelos de dados, que são soluções de modelagem conceitual de dados para aplicações comerciais. Para ilustrar um padrão de análise, a seguir é mostrado o padrão *Localizações Geográficas*, extraído de [HAY 95], com algumas adaptações. O padrão está descrito segundo a forma proposta por Coplien [COP 95], sendo que os diagramas foram convertidos para a notação UML [BOO 98].

3.1.1 Exemplo de um Padrão de Análise

Nome: Localizações Geográficas

Problema: Descreve uma solução alternativa para o problema da modelagem de endereços, sendo útil tanto para pessoas, como para organizações.

Contexto: O padrão descreve uma alternativa de solução para o problema da modelagem de endereços, servindo tanto para pessoas físicas como para pessoas jurídicas. É comum, em muitas situações da vida real, um mesmo endereço estar associado a mais de uma pessoa (ou organização) e vice-versa. Nestes casos, quando o endereço é modelado simplesmente na forma de atributos, uma série de problemas conhecidos podem ocorrer como, por exemplo, informação incompleta ou a inconsistência do banco de dados.

Forças:

- Frequentemente é necessário ordenar/selecionar endereços por bairros, cidades ou por outro tipo de divisão onde o endereço está localizado. Assim, cada local tem de estar em uma localização geográfica, o que significa que cada localização geográfica pode ser a localização de um ou mais locais.
- Para emissão de etiquetas de endereçamento, o endereço pode ser tratado como um único objeto de informação. Se o propósito do endereço é apenas para emissão de etiquetas, a escolha da abordagem de atributos pode ser considerada.

Solução: A Figura 2 mostra o diagrama de classes que compõe o padrão. A classe PESSOA é uma generalização das classes PESSOA FÍSICA e PESSOA JURÍDICA. A classe PESSOA possui uma associação *LOCALIZA-SE* (de cardinalidade m:n) com a classe

ENDEREÇO. Esta associação pode incluir atributos como período e tipo de locação. A classe ENDEREÇO é que deve conter os atributos que caracterizam o endereço de uma instância de PESSOA. A classe LOCALIZAÇÃOGEOGRÁFICA é usada para criar uma estrutura hierárquica de localizações (ex.: País, Estado, Município, Bairro, etc), sendo que cada instância de ENDEREÇO deve estar associada a uma instância de LOCALIZAÇÃOGEOGRÁFICA.

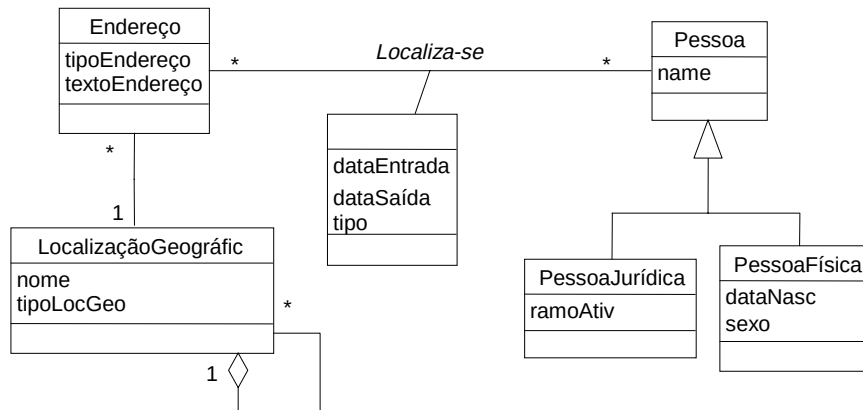


FIGURA 2 - Padrão de análise LocalizaçõesGeográficas

De acordo com Gerth [GER 96], a notação gráfica dos modelos de dados não são suficientes para a descrição dos padrões. A notação gráfica somente captura o resultado final da modelagem, representando-a na forma de um grafo de classes e associações. A descrição de um padrão, no entanto, deve também incluir a documentação das decisões tomadas, escolhas feitas, alternativas identificadas e restrições ao uso do padrão.

Segundo Hay [HAY 95], um padrão de análise pode e deve ser customizado para situações específicas. Por exemplo, o padrão *LocalizaçõesGeográficas* mostra uma situação (existem alternativas diferentes para a modelagem de informações sobre endereço) na qual pode-se observar que não existe o modelo totalmente correto, mas sim, o mais adequado para um problema em um determinado contexto. Outro ponto a ser observado é que a definição de um padrão pode estar associada ou incluir outros padrões. Por exemplo, a generalização da classe PESSOA é uma solução definida em outro padrão (*PARTY*), descrito tanto em [HAY 95], como em [FOW 97]. Por sua vez, o padrão *LocalizaçõesGeográficas* poderá ser reutilizado no projeto de outros padrões mais abrangentes.

3.2 Padrões de Projeto

Da mesma forma como não existe uma fronteira distinta, entre as atividades desenvolvidas durante as fases de análise e de projeto do processo de desenvolvimento de software, torna-se difícil definir, exatamente, as diferenças entre os padrões de análise e os padrões de projeto. O objetivo desta Seção não é apresentar em detalhes as características dos padrões de projeto, mas sim, o de evidenciar as principais diferenças entre um padrão de projeto e um padrão de análise. São elas:

- Um padrão de análise é um padrão cuja forma é descrita através de termos e conceitos pertencentes ao domínio da aplicação. Já, a forma de um padrão de projeto é descrita através de construtores genéricos de projeto de software (ex.: objeto, agregação) [RIE 96].
- Padrões de análise não são escritos para servirem a um propósito geral. É importante encontrar o equilíbrio entre obter um padrão muito abstrato e um padrão muito especializado. Por exemplo, apresentar como solução para o

problema de rede de transportes, que seus componentes podem ser modelados através de elementos do tipo conexão e ligação, em uma estrutura de rede é apresentar a solução de forma altamente abstrata, não auxiliando muito o projetista. No entanto, se o padrão de análise identifica os principais tipos de componentes de uma rede de transporte e como estes componentes podem (ou não) se relacionar, este padrão torna-se realmente útil para ser reutilizado. Um padrão de projeto descreve a estrutura e as dinâmicas de seus componentes, esclarecendo suas responsabilidades. Assim, os padrões de projeto podem ser vistos como complementares aos padrões de análise.

- Ao contrário dos padrões de análise, os padrões de projeto existentes são apresentados de forma bem estruturada e organizados separadamente em catálogos. Por exemplo, os padrões de Gamma [GAM 94] apresentam quatro elementos essenciais: nome do padrão, o problema, a solução e as conseqüências. No entanto, outros itens de descrição podem ser adicionados como motivação, participantes, sinônimos e outros padrões relacionados. A classificação dos padrões de Gamma é feita segundo dois critérios ortogonais: (1) *Propósito*, incluindo se o padrão está relacionado com o processo de criação de objetos, com a composição de classes ou com as interações comportamentais dos objetos; (2) *Escopo*, indicando se o padrão se aplica a classes ou a objetos. Os padrões de análise, por sua vez, são apresentados de forma mais explicativa, apresentando soluções alternativas, vantagens e desvantagens de cada um.

A Figura 3 ilustra o diagrama de classes do padrão de projeto *Composite* [GAM 94] e a Figura 4 mostra um exemplo de uso desse padrão no diagrama de classes do *framework* conceitual **GeoFrame** [LIS 99].

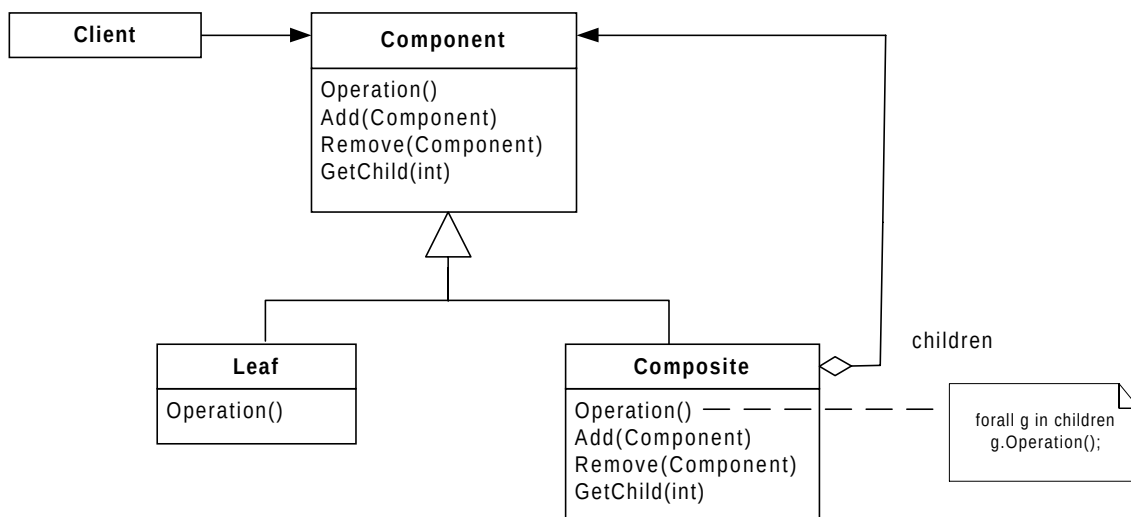
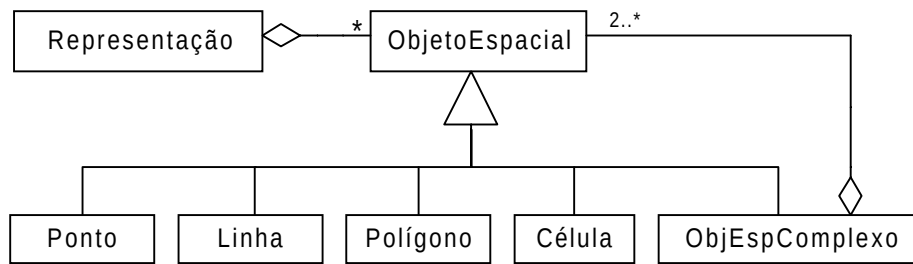


FIGURA 3 - Padrão de projeto Composite

FIGURA 4 - Exemplo de uso do padrão de projeto *Composite*

4 FRAMEWORKS

O uso de *frameworks*⁴ caracteriza uma técnica de reutilização de software orientado a objetos [JOH 97]. Segundo Pree [PRE 94], *framework* constitui um avanço real em termos de reutilização de software, uma vez que a reutilização ocorre não apenas a partir de pequenos blocos de programas, mas sim de todo um sistema (ou parte de um sistema), incluindo a reutilização do projeto.

Souza [SOU 98] define um *framework* como:

“Um projeto genérico em um domínio que pode ser adaptado a aplicações específicas, servindo como um molde para a construção de aplicações”.

Esta definição fornece uma visão bem mais abrangente sobre a potencialidade de um *framework* do que as definições apresentadas pelos autores mais ligados à programação orientada a objetos. Por exemplo, Wirfs-Brock e Johnson definem *framework* como sendo:

“Um esqueleto de implementação de uma aplicação ou de um subsistema de aplicação, em um domínio de problema particular” [WIR 90].

“Um projeto reutilizável de um programa ou parte de um programa, expresso como um conjunto de classes” [JOH 92].

Como um instrumento de reutilização, um *framework* não necessita estar implementado em uma linguagem de programação para fornecer a solução parcial a uma família de problemas. A vantagem de se ter o *framework* parcialmente implementado é que a solução final está bem mais próxima de ser atingida pelo desenvolvedor da nova aplicação. No entanto, o *framework* fica restrito a uma linguagem de programação específica ou necessita ser fornecido em diversas outras linguagens.

Uma das principais características de um *framework* é que o fluxo de controle entre o *framework* e sua aplicação cliente é invertido. Normalmente, quando uma biblioteca de classes é utilizada, o controle das chamadas às rotinas disponíveis é feito pelo programa que está utilizando a biblioteca. No caso de um *framework*, os métodos desenvolvidos para a aplicação é que são chamados pelos métodos do *framework*. Isto evidencia que, além de reutilizar todo um conjunto de classes, o conhecimento sobre o projeto do software também é reutilizado.

⁴ O termo *framework* não será traduzido, neste texto, devido a seu uso freqüente pela comunidade de Engenharia de Software no Brasil. Uma possível tradução para *framework* é arcabouço.

Padrões de projeto podem ser usados tanto na construção como na documentação de *frameworks* [JOH 92, MEU 97]. Um simples *framework* pode ser construído a partir de diversos padrões de projeto (também chamados de micro arquiteturas).

Outra forma de distinguir padrões de projeto e *frameworks* é que, enquanto um *framework* pode ser apresentado como um software inacabado, os padrões representam o conhecimento sobre como o software foi desenvolvido. Além disso, a complexidade dos problemas a serem resolvidos pelos *frameworks* são bem maiores do que os problemas resolvidos pelos padrões. Consequentemente, a construção de um *framework* é um processo bem mais complexo do que a definição de um padrão. Silva [SIL 96] descreve e compara uma série de métodos para construção de *frameworks*.

Um *framework* é composto por um conjunto de classes, muitas das quais abstratas, ou seja, não possuem instâncias. As classes de um *framework* podem ser classificadas como pontos fixos, isto é, não podem ser alteradas, ou como pontos adaptáveis, onde são feitas as mudanças e extensões necessárias à aplicação. Segundo Quadros [QUA 97], os pontos fixos determinam a arquitetura da aplicação, definindo sua estrutura geral, as responsabilidades e colaborações entre as classes e o fluxo de controle do *framework*, não podendo ser alterados pelo projetista da aplicação.

De acordo como as aplicações são derivadas a partir de um *framework*, este pode ser classificado em caixa-preta ou caixa-branca. Os pontos adaptáveis de um *framework* caixa-branca fornecem classes incompletas, as quais devem ser estendidas com as especificidades da aplicação. Um *framework* caixa-preta fornece um conjunto de classes alternativas para cada ponto adaptável. Neste caso o projetista deve escolher uma das alternativas disponíveis, mas não pode alterar suas propriedades. Um *framework* caixa-branca requer que o projetista tenha um maior conhecimento sobre o funcionamento do *framework*, enquanto que um *framework* caixa-preta requer do projetista um maior conhecimento sobre o domínio da aplicação.

As três abordagens de reutilização apresentadas até aqui (de especificação de requisitos, de padrões e *frameworks*) podem ser empregadas no desenvolvimento de aplicações em geral. A seção seguinte discute as características das abordagens de reutilização existentes atualmente na área de SIG.

5 ABORDAGENS DE REUTILIZAÇÃO EM SIG

Na área de Sistemas de Informação Geográfica (SIG) o conceito de reutilização tem sido aplicado, basicamente, no compartilhamento de dados. Observa-se, atualmente, um rápido crescimento da demanda por dados geoespaciais digitais [WEB 98]. Por sua vez, a todo momento, muitos dados geoespaciais estão sendo produzidos, tanto por instituições governamentais como por instituições privadas. Dados geoespaciais, neste caso, incluem os mais variados tipos de coleções de dados georreferenciados como, por exemplo, imagens de satélite, fotografias aéreas, amostragem geológica, observações biológicas, dados de redes de infra-estrutura e dados censitários.

A coleta de dados primários, ou seja, dados usados na produção de outros dados, tanto para aplicações ambientais como para aplicações urbanas, é uma tarefa de custo muito elevado. Com o objetivo de reduzir estes custos, as instituições necessitam de mecanismos eficientes para facilitar a reutilização de dados já existentes, visto que muitos conjuntos de dados georreferenciados podem atender a diversas aplicações de SIG, implementadas em instituições diversas.

Além da reutilização de dados, algumas pesquisas sobre mecanismos de reutilização de modelagem de processos espaciais têm sido realizadas [MAR 97, MAR 98]. Segundo Marr [MAR 97], modelagem do processo espacial é o procedimento

de delinear como os dados geográficos devem ser manipulados a fim de se obter resultados desejados.

Com relação a reutilização de projeto de banco de dados geográficos, porém, muito pouco foi feito até o momento [LIS 98]. Alguns padrões⁵ de metadados prevêm a documentação sobre informações semânticas associadas aos dados, mas com o intuito de facilitar o entendimento dos dados a serem reutilizados.

Metadados são dados sobre dados [ISO 97]. A existência de conjuntos de dados geoespaciais é documentada através de conjuntos de metadados. O termo metadados geoespaciais é usado para especificar metadados referentes a conjuntos de dados georreferenciados [SHE 95].

A importância dos metadados geoespaciais reside no fato de possibilitarem ao usuário avaliar o grau com que os dados satisfazem às necessidades de uma determinada aplicação. A informação prestada pelos metadados provê uma substancial economia de recursos na pesquisa e busca de dados existentes bem como no monitoramento, controle e aquisição dos mesmos. Metadados expressam informações sobre o conteúdo, a qualidade, a atualidade e outras características dos dados [WEB 98].

Para facilitar a busca e o acesso a conjuntos de dados existentes, alguns países estão especificando e institucionalizando padrões para a descrição uniforme dos metadados. Os padrões de metadados estabelecem categorias e elementos a serem contemplados para descrever adequadamente os dados. Entre os padrões de metadados mais conhecidos, pode-se citar: o padrão americano CSDGM [FED 97], o padrão canadense SAIF [BRI 95], o padrão CEN-TC287 da comunidade européia [CEN 96] e o padrão australiano ANZLIC [ANZ 97]. Estes padrões estabelecem uma terminologia comum aos fenômenos espaciais, classes de dados e seus modelos de informação, além de regras para produção de metadados.

A efetiva reutilização de dados geoespaciais digitais depende de diversos fatores, entre eles pode-se citar a interoperabilidade dos dados e a vontade de compartilhamento dos dados entre as instituições (de forma comercial ou não). Para que um maior número possível de usuários possa reutilizar dados geoespaciais existentes, é necessário que os catálogos de metadados estejam disponíveis para consulta. Alguns padrões já estão operacionais (ex.: CSDGM), dispondo inclusive de programas para cadastramento de metadados e esquemas para disponibilizar os metadados através da Internet [VIL 99].

Uma vez elaborado um conjunto de metadados, é necessário torná-lo acessível através de ferramentas de busca, o que vem suprir o primeiro objetivo dos metadados que é a localização dos dados. Conjuntos de dados geoespaciais são localizados através dos metadados. No entanto, estes dados não estão, necessariamente, disponíveis para serem obtidos automaticamente (ex.: *download* via ftp). Pode-se, por exemplo, localizar um conjunto de dados para uma determinada região geográfica de interesse. Através dos metadados de qualidade, avalia-se a adequabilidade desse conjunto para o uso planejado. Por último, pode ser necessário, caso o dado não esteja disponível, fazer o contato direto com o fornecedor do dado a fim de obtê-lo.

6 CONCLUSÕES

As abordagens de reutilização em sistemas de informação, embora apresentem diversas vantagens quando realmente aplicadas no desenvolvimento de sistemas, na prática não têm sido muito utilizadas. Alguns fatores podem justificar o pouco uso desses instrumentos:

⁵ Do inglês *standard*.

- ausência de uma metodologia de projeto de sistemas que induza o projetista a buscar por possíveis recursos reutilizáveis;
- baixa disponibilidade de recursos reutilizáveis;
- dificuldade de localização dos recursos reutilizáveis existentes. Inexistência de um mecanismo de apoio a busca dos recursos existentes;
- falta de cultura de trabalho em cooperação. É mais fácil para o projetista procurar recursos disponíveis para serem reutilizados do que disponibilizar seus próprios recursos para outros projetistas.

A abordagem de padrões (*pattern*) é ainda muito pouco conhecida. O projetista necessita, no entanto, estudar diversos padrões e exemplos de utilização para que possa identificar possíveis aplicações desses padrões. O treinamento dos projetistas sobre o uso de padrões pode ser compensado pelo ganho posterior em produtividade e qualidade dos sistemas desenvolvidos. Além disso, uma vez treinado, o projetista estará apto a aplicar padrões em diversos sistemas. Uma outra vantagem decorrente do uso de padrões é o da unificação de vocabulário entre os diversos membros de uma equipe de desenvolvimento.

Os problemas com o uso de *frameworks* são semelhantes aos encontrados no uso de padrões, porém, em escala diferente. Devido a sua maior abrangência, o uso de *frameworks* exige um conhecimento muito maior por parte do projetista, quando comparado ao uso de padrões.

O alto custo e a complexidade dos processos de aquisição de dados georreferenciados tornam a reutilização de dados geoespaciais um fator de grande importância. Toda aplicação de SIG utiliza-se de um conjunto de dados primários (ex.: hidrografia, vegetação ou divisas municipais). Desta forma, os conjuntos de dados digitais referentes a esses dados primários podem, na maioria das vezes, ser reutilizados por grupos de usuários trabalhando em aplicações diversas sobre a mesma região geográfica.

No Brasil, têm sido realizadas algumas iniciativas isoladas para disponibilizar dados georreferenciados. Por exemplo, o projeto *GeoMinas* (<http://www.mg.gov.br/geominas>), onde pode-se obter dados primários sobre o Estado de Minas Gerais. Um segundo exemplo é o projeto do grupo de geoprocessamento da Companhia de Desenvolvimento do Planalto Central – CODEPLAN (<http://www.gdf.gov.br/codeplan/geopro>). Estas iniciativas apenas fornecem os conjuntos de dados que podem ser reutilizados, mas não fornecem nenhum mecanismo que auxilie a busca e localização dos dados disponíveis. Um sistema experimental de busca a metadados geoespaciais é disponibilizado pelo Serviço Geológico do Brasil (CPRM) em (<http://www.cprm.gov.br>) [ALM 99].

Estudos no sentido da adoção de um padrão internacional para elaboração de metadados geoespaciais, no Brasil, têm sido realizados [RIB 96, WEB 98, ALM 99]. No entanto, para adoção de um padrão existente, ou mesmo para a definição de um padrão nacional, é necessário o envolvimento de toda a comunidade de usuários e produtores de dados, o que deve ser feito sob a coordenação de algum órgão responsável.

3 Modelagem Conceitual de Banco de Dados Geográficos

3.1 Introdução

Este capítulo descreve a evolução das abordagens de modelagem conceitual de dados empregadas durante o desenvolvimento de diversas aplicações de SIG, no âmbito de projetos de cooperação entre o Instituto de Informática da UFRGS e vários órgãos e instituições (públicas e privadas) que atuam no estado do Rio Grande do Sul.

O capítulo enfoca, principalmente, a evolução do diagrama de classes BCH/GIS, que serve como ponto de partida para o projeto de esquemas conceituais de banco de dados geográficos. Durante sua evolução, o diagrama de classes recebeu diferentes denominações como, por exemplo, Hierarquia de Classes Básicas e Biblioteca de Classes Básicas. Atualmente, o diagrama de classes BCH/GIS fornece a estrutura do *framework* GeoFrame [LIS 99], documentado no próximo capítulo.

Os quatro artigos que compõem o capítulo apresentam, ainda, outras características identificadas, principalmente, durante o desenvolvimento de um grande projeto de SIG, que foi a modelagem do sistema SIGPROGB, o SIG do Programa Pró-Guaíba [BIN 98]. Entre estas características, citar-se a identificação da necessidade de incluir uma etapa adicional à etapa de projeto conceitual, que se destina ao Projeto de Representação Espaço-Temporal (Seção 3.4).

Um segundo aspecto importante, identificado durante a definição da metodologia para a modelagem conceitual do sistema SIGPROGB (Seção 3.3), foi a necessidade da participação no desenvolvimento do sistema de um profissional da área de psicologia organizacional. Esta participação teve como meta aumentar a motivação dos usuários participantes do projeto e reduzir as resistências pessoais e políticas, uma vez que estavam envolvidos mais de dez órgãos e instituições em um único projeto.

O restante do capítulo está organizado da seguinte maneira. O Artigo III (Seção 3.2), intitulado “*Adaptando o modelo OMT para modelagem conceitual de aplicações de SIG*”, apresenta a primeira versão do diagrama de classes BCH/GIS, inclui uma descrição inicial dos requisitos de modelagem e de alguns modelos para SIG, já descritos no Artigo I (Capítulo 2), e apresenta um segmento da modelagem conceitual do banco de dados do sistema SIGERCO. Esta modelagem foi desenvolvida no âmbito do projeto integrado de pesquisa SIGMODA, financiado pelo CNPq, que foi desenvolvido em cooperação com a Fepam - Fundação Estadual de Proteção Ambiental do RS.

O Artigo IV (Seção 3.3), intitulado “*Metodologia para modelagem conceitual de dados de um SIG interinstitucional*”, descreve as três fases da metodologia de desenvolvimento da modelagem conceitual dos dados do sistema SIGPROGB, o SIG do Programa Pró-Guaíba. O artigo apresenta, ainda, a segunda versão do diagrama de classes BCH/GIS.

A Seção 3.4 apresenta o Artigo V, intitulado “*Modelos conceituais de dados para aplicações geográficas: uma experiência com um SIG interinstitucional*”, que inclui uma descrição resumida dos requisitos e modelos já descritos no Artigo I (Capítulo 2). A principal contribuição do artigo é a descrição da etapa especial do processo de modelagem, que é a Fase do Projeto de Representação Espaço-Temporal, uma etapa

fundamental para o desenvolvimento de grandes sistemas. O artigo apresenta, ainda, um exemplo de um segmento do esquema de dados do sistema SIGPROGB.

Por último, o Artigo VI (Seção 3.4), intitulado “*BCH/GIS - A basic class hierarchy to support GIS conceptual design*”, descreve os problemas identificados nas duas primeiras versões do diagrama de classes, as modificações efetuadas no diagrama para resolver estes problemas e propõe uma terceira versão. Na nova versão, o diagrama apresenta uma linha divisória entre as classes que devem ser consideradas em cada uma das fases iniciais, ou seja, Projeto Conceitual e Projeto de Representação Espaço-Temporal. Um exemplo de uso do diagrama de classes BCH/GIS é apresentado.

A terceira versão do diagrama de classes foi avaliada como parte do desenvolvimento de uma dissertação de mestrado [GAR 98]. Neste trabalho, novos problemas foram identificados, o que levou ao desenvolvimento de uma quarta versão, que corresponde ao diagrama de classes do *framework* GeoFrame [LIS 99]. As deficiências identificadas em [GAR 98] estão descritas na seção de introdução do próximo capítulo (Seção 4.1).

3.2 Artigo III: Adaptando o Modelo de Objetos OMT para Modelagem Conceitual de Aplicações de SIG

Jugurta Lisboa Filho e Cirano Iochpe

In: 1ª SEGEO-RJ – SEMANA ESTADUAL DE GEOPROCESSAMENTO, 1996, Rio de Janeiro. **Anais...** Rio de Janeiro: Escola de Engenharia, 1996.

RESUMO

Este trabalho apresenta uma adaptação do modelo de objetos OMT para utilização no projeto das aplicações de SIG, sendo sua aplicabilidade demonstrada através de um exemplo de SIG na área ambiental.

ABSTRACT

This paper presents an adaptation of the OMT object model for GIS application design, and shows its usage through an example of environmental GIS.

1 INTRODUÇÃO

Um Sistema de Informações Geográficas (SIG) é um sistema baseado em computador, que possibilita a captura, modelagem, manipulação, recuperação, análise e apresentação de dados referenciados geograficamente [WOR 95].

Uma das maiores dificuldades enfrentadas, atualmente, pelos usuários de SIG está na definição da base de dados geográfica, ou seja, na definição de quais entidades serão armazenadas no banco de dados geográfico. Neste artigo, chamamos de banco de dados geográfico (BDGEO) o componente do SIG capaz de armazenar ou gerenciar os dados espaciais e descritivos que serão manipulados pelas aplicações geográficas.

Segundo [ELM89], o projeto de banco de dados envolve três fases distintas: *projeto conceitual*, *projeto lógico* e *projeto físico*. Na fase do *projeto conceitual*, de que trata este artigo, o enfoque principal está na identificação de quais as entidades que serão representadas no BDGEO e como elas se relacionam, desconsiderando-se aspectos sobre a forma de armazenamento destas entidades no BDGEO e, até mesmo, qual o SIG que será utilizado.

Diversos trabalhos vêm sendo desenvolvidos com o objetivo de se encontrar um modelo de dados conceitual que seja, ao mesmo tempo, rico o bastante para representar a complexa realidade das aplicações geográficas e o mais simples possível, para que possa ser entendido e utilizado pelos usuários de SIG [WOR 94], [CLE 94], [CAR 93], [BON 93], [TRY 95], [KÖS 95], [TIM 94]. Uma forte tendência tem sido verificada no sentido de se utilizar os conceitos do paradigma de orientação a objetos para o projeto de aplicações em SIG [WOR94].

Este artigo apresenta os resultados da modelagem conceitual de dados de parte do sistema SIGERCO/RS, Sistema de Gerenciamento Costeiro, da Fundação Estadual de Proteção Ambiental do Estado do Rio Grande do Sul (FEPAM/RS). Tal modelagem,

baseada em uma biblioteca de classes de objetos OMT [RUM 91], foi desenvolvida no âmbito do Projeto Integrado de Pesquisa *SIGMODA*, financiado pelo CNPq.

O artigo está organizado da seguinte maneira. Na seção 2, são relacionados os requisitos especiais apresentados pelas aplicações de SIG. A seção 3 descreve as adaptações propostas ao modelo de objetos OMT, para a modelagem de aplicações geográficas. Na seção 4 são apresentados os resultados da modelagem de uma aplicação exemplo. Por último, são tiradas algumas conclusões sobre o trabalho e apresentadas perspectivas de trabalho futuro.

2 REQUISITOS DAS APLICAÇÕES DE SIG

As aplicações de SIG apresentam requisitos de modelagem específicos, que não são tratados adequadamente pelos modelos de dados empregados atualmente no desenvolvimento de aplicações de banco de dados convencionais [WOR 94]. Em [LIS 96a], relacionamos alguns destes requisitos. São eles:

- Dicotomia entre campos e objetos.
- Relacionamentos espaciais.
- Tratamento unificado dos dados espaciais e não-espaciais.
- Qualidade e disponibilidade dos dados.
- Temporalidade e representação de limites difusos.

Destes requisitos, o primeiro é o mais importante para a modelagem conceitual.

Para [GOO 92], a realidade geográfica pode ser observada segundo duas visões: de *campo* e de *objetos*. Na visão de *campo*, a realidade é modelada por variáveis que possuem distribuição contínua no espaço (ex.: relevo e temperatura). Todas as posições no espaço geográfico estão associadas a algum valor correspondente à variável representada. Na visão de *objetos*, entidades reais são observadas como estando distribuídas sobre um grande espaço vazio, ou seja, nem todas as posições estão necessariamente ocupadas.

[GOO 92] identifica seis tipos diferentes de modelos de representação, baseados na visão de campo, que são utilizados nos SIGs (Figura 1). São eles:

- *Amostragem Irregular de Pontos* : Dados são coletados em um conjunto finito de localizações irregularmente espaçadas (ex.: estações de medição de temperatura).
- *Amostragem Regular de Pontos* : Como no item anterior, porém com pontos distribuídos regularmente (ex.: Modelos Numéricos de Terreno).
- *Linhas de Contorno* : Dados são representados por um conjunto de linhas, cada uma com um valor associado (ex.: curvas de nível).
- *Polígonos* : A área de estudo é particionada em um conjunto de polígonos. Cada polígono tem um valor que é único em todas as posições dentro do polígono (ex.: tipos de solos).
- *Grade Regular de Células* : A área é dividida em uma grade regular de células, onde o valor da cada célula corresponde ao valor da variável para todas as posições dentro da célula (ex.: imagens de satélites).
- *Rede Triangular Irregular* : A área é dividida em triângulos irregulares, onde o valor da variável é definido em cada vértice e varia linearmente sobre o triângulo (ex.: TIN).

Esses modelos, muitas vezes são confundidos, equivocadamente, com as estruturas de armazenamento (*matricial* e *vetorial*) de dados espaciais. Cada um desses modelos pode ser mapeado em uma ou outra estrutura, sendo que alguns se adequam melhor à estrutura matricial e outros à estrutura vetorial. Por exemplo, os modelos

Amostragem Regular de Pontos e *Grade Regular de Células*, são mapeados naturalmente para a estrutura matricial, enquanto que os demais modelos são melhor organizados numa estrutura vetorial.

Na visão de objetos, as entidades da realidade são representadas como *pontos*, *linhas* ou *áreas*. Ao contrário da visão de campo, na visão de objetos dois objetos podem estar localizados na mesma posição geográfica, ou seja, podem possuir coordenadas idênticas.

Do ponto de vista dos usuários, esta divisão entre as visões de campo e objetos pode direcionar o projeto da base de dados geográfica. Na seção seguinte, apresentamos uma hierarquia de classes que serve de ponto de partida para o projeto de aplicações baseadas em SIG, permitindo a modelagem integrada dos dados nas visões de campo e objetos.

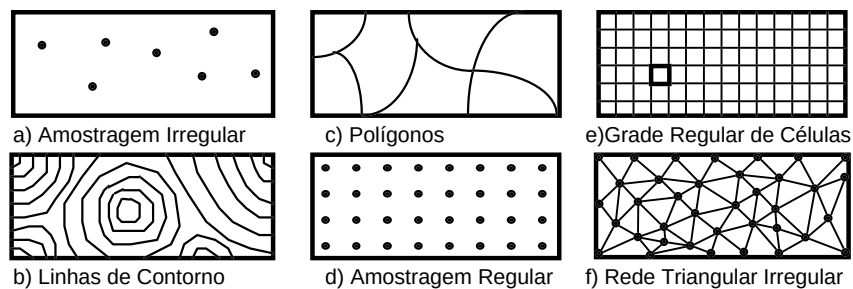


FIGURA 1 - Modelos de dados na visão de campo [LIS 96a]

3 MODELOS DE DADOS CONCEITUAIS

Modelos de dados são visões simplificadas de uma parte da realidade e são construídos de acordo com certas regras, para facilitar a implementação do banco de dados em um sistema de informações. Através de um formalismo gráfico, é possível abstrair os aspectos da realidade que se deseja modelar, e, devido à simplicidade dos construtores gráficos utilizados, torna-se simples a validação do modelo por parte dos usuários envolvidos no projeto.

A representação gráfica escolhida para ser utilizada em nosso projeto de SIG, foi a do modelo de objetos da metodologia OMT (Técnica de Modelagem de Objetos [RUM 91]), onde foram incluídos novos construtores para facilitar a modelagem das características espaciais da aplicação.

3.1 Modelo de Objetos OMT

Ao contrário dos modelos funcionais, onde a ênfase está nos procedimentos e funções executadas no sistema (ex.: modelo DFD [GAN 79]) e dos modelos estáticos, onde a ênfase está nos dados armazenados (ex.: modelo ER [CHE 76]), os modelos orientados a objetos utilizam uma filosofia que integra dados e funções em unidades denominadas *objetos* [GOL 82].

O modelo OMT é baseado em conceitos do paradigma da orientação a objetos, tais como, hierarquia de classes, herança de propriedades, atributos, métodos, entre outros, e permite a representação de conceitos de abstração semântica, tais como, classificação, generalização, agregação e associação.

3.2 Adicionando Semântica ao Modelo OMT

Nesta seção, primeiro é apresentada uma hierarquia de classes predefinidas, utilizada no projeto conceitual de aplicações geográficas. Depois é descrito uma estratégia de simplificação de esquemas por substituição simbólica.

3.2.1 Biblioteca de classes básicas

Muito antes da utilização dos SIGs, já era comum, na comunidade de Geoprocessamento, visualizar a realidade na forma de mapas sobrepostos, criando um conjunto de camadas, cada uma contendo as informações sobre um determinado tema (ex.: hidrografia, relevo, uso da terra, dados sócio-econômicos, etc). É natural, portanto, que um projeto de banco de dados para SIG inicie pela identificação dos temas a serem modelados.

Um BDGEO pode ser visto como sendo um repositório de dados sobre um conjunto de áreas geográficas. Cada área geográfica possui dados organizados em diversos temas, que podem ser modelados segundo as visões de campo e objetos. A Figura 2 mostra a hierarquia de classes básicas, seguindo o simbolismo do modelo OMT, que serve de ponto de partida para a modelagem conceitual de aplicações em SIG. Esta biblioteca permite a integração natural dos dados nas visões de campo e de objetos.

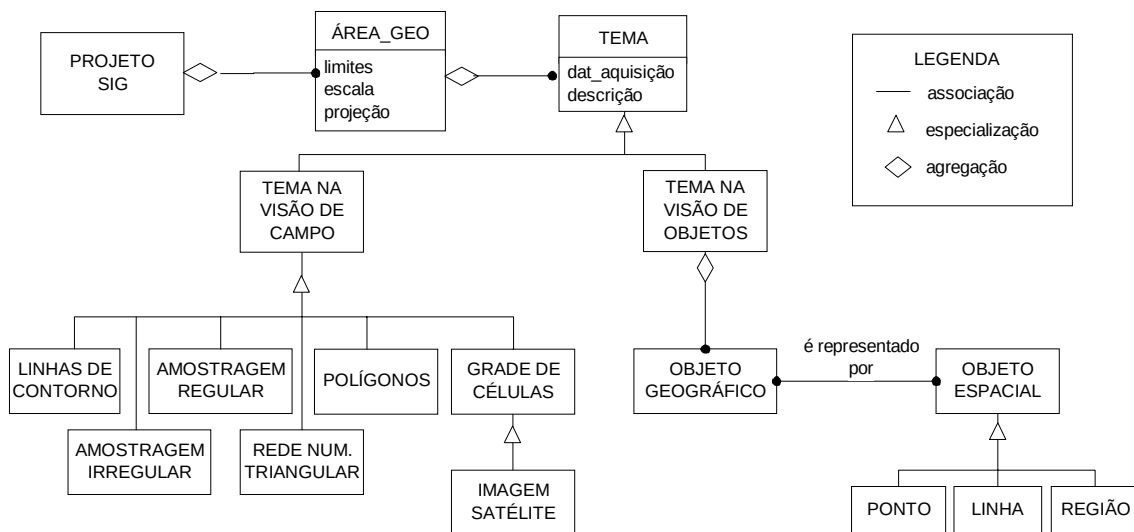


FIGURA 2 - Hierarquia de classes básicas

A biblioteca de classes apresentada aqui, foi baseada na biblioteca de classes descrita no modelo MGeo [TIM 94]. No modelo MGeo, a ênfase está na definição do conjunto de classes que descrevem os diversos tipos de objetos geométricos, que são empregados nas implementações de SIG.

De acordo com as características das variáveis presentes em cada tema, elas podem ser modeladas segundo uma ou outra visão. Por exemplo, temas como *relevo*, *tipo de solo* e *temperatura* possuem distribuição contínua no espaço e são modelados sob a visão de campo, enquanto que *hidrografia* e *dados sócio-econômicos* são temas que possuem entidades reais bem definidas e são modelados na visão de objetos.

3.2.2 Simplificando esquemas por substituição simbólica

Na visão de objetos, todas as classes de entidades geográficas, ou seja, aquelas que possuem atributos do tipo objeto espacial, estão associadas a, pelo menos, uma subclasse da classe *Objeto_Espacial*, isto é, *Ponto*, *Linha* ou *Região*. Isto causa a adição de um grande número de relacionamentos no diagrama final, tornando-o de difícil legibilidade. A inclusão de símbolos geométricos nas classes de entidades geográficas em substituição a esses relacionamentos, simplifica significativamente o diagrama final (Figura 3).

Esquemas semelhantes de simplificação de esquema por substituição simbólica, também foram empregados nos modelos MODUL-R [BÉD 89], [CAR 93] e no modelo GeoOOA [KÖS 95]. Estas substituições não se limitam às associações com objetos do tipo *Ponto*, *Linha* e *Região*. Substituições podem ser adaptadas a outros tipos de objetos espaciais, como por exemplo, *nodos* e *ligações*, na modelagem de objetos em aplicações de redes.

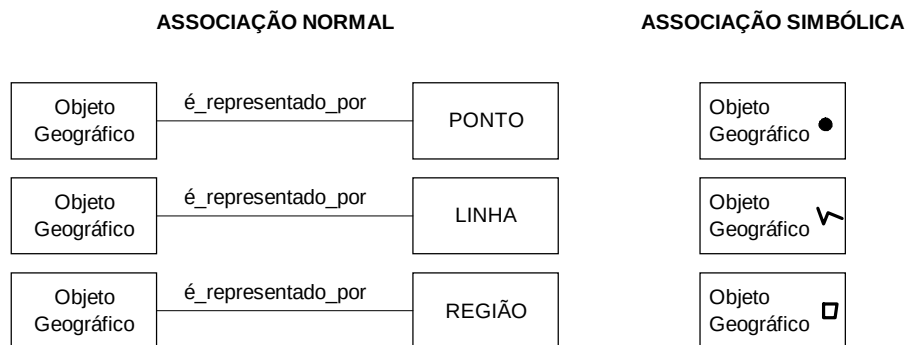


FIGURA 3 - Substituição simbólica de esquemas

4 MODELAGEM DE DADOS DO SIGERCO

O Projeto de Gerenciamento Costeiro (GERCO), da FEPAM-RS, tem como metas básicas, o planejamento e o gerenciamento, de forma integrada, descentralizada e participativa, das atividades sócio-econômicas na zona costeira do país, de forma a garantir a utilização, conservação e preservação dos recursos naturais e ecossistemas costeiros [IOC 94].

Esta seção apresenta, como exemplo de aplicação do modelo OMT adaptado para SIG, a modelagem conceitual de dados do SIGERCO/RS Sistema de Informações do Gerenciamento Costeiro.

O objetivo do SIGECO/RS é o de fornecer subsídios para a tomada de decisões nas diversas atividades do Projeto GERCO.

Uma das atividades do projeto GERCO, a do Macrozoneamento Costeiro, constitui-se na elaboração do *Mapa de Uso Projetado*, que é composto pelas cartas-síntese, representando as potencialidades e as restrições de uso em cada região costeira.

As subseções seguintes mostram os resultados da modelagem conceitual dos dados do SIGERCO/RS, relacionados com as etapas de Macrozoneamento Costeiro, Licenciamento e Monitoramento.

Por questões de simplificação do diagrama final da modelagem, as classes da aplicação foram separadas de acordo com a visão (de campo e de objeto) em que se enquadram.

4.1 Classes na Visão de Campo

As classes pertencentes ao domínio da aplicação, que se enquadram na visão de campo, foram definidas como especializações das subclasses de *Tema na Visão de Campo*, na hierarquia de classes básicas (Figura 4). Estas classes estão relacionadas, principalmente, com a execução da fase do Macrozoneamento Costeiro, no âmbito do Projeto GERCO e servirão de base às operações espaciais que serão realizadas para se obter o *Mapa de Uso Projetado*.

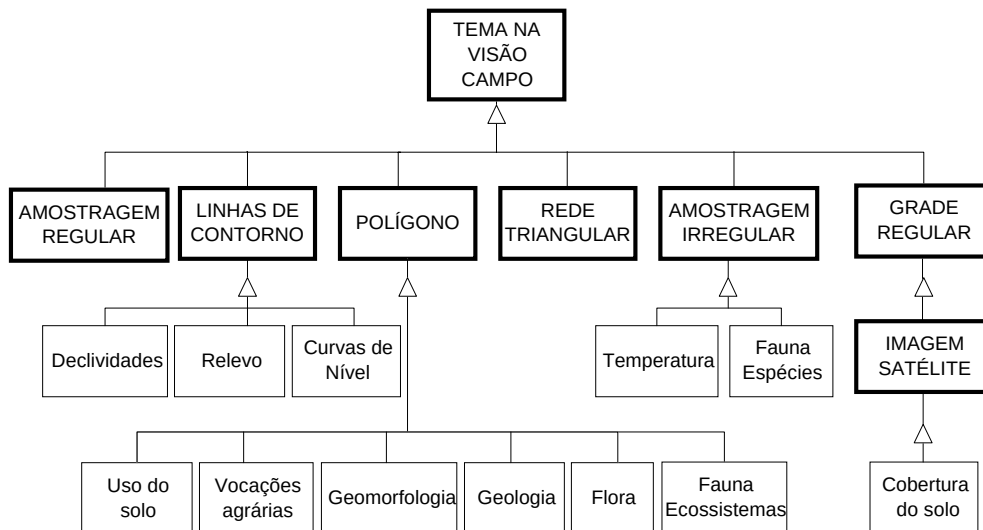


FIGURA 4 - Modelagem de dados na visão de campo

4.2 Classes na Visão de Objetos

As informações sobre os temas *Sócio-Economia* e *Uso da Água*, possibilitaram a identificação de diversas classes (com e sem representação espacial) que foram modeladas na visão de objetos, conforme mostra a Figura 5.

Todas as classes de entidades geográficas possuem um símbolo geométrico que indica o tipo de objeto espacial ao qual está associado. Esta associação é dependente da escala em uso e pode, inclusive, existir mais de um objeto espacial associado a uma mesma classe de objeto geográfico. Por exemplo, objetos da classe *Trecho de Rio* podem ser representados por uma linha ou uma região.

Na visão de objetos, os atributos espaciais e descritivos são descritos de forma integrada na definição das classes. Além disso, como pode ser observado no diagrama, os relacionamentos entre classes com e sem características espaciais, são modelados de forma integrada. Por exemplo, as classes *Ponto de Amostragem*, que é uma classe que possui representação espacial na forma de um ponto, e *Monitoramento dos Recursos Hídricos*, que não possui representação espacial, são representadas de forma integrada no modelo, através da associação (relacionamento) *realiza*. Isto permite que o projeto conceitual do banco de dados seja realizado sem que se leve em conta se os atributos das entidades modeladas são apenas descritivos ou se possuem representação espacial.

É importante observar que, somente nas fases subsequentes, é que serão considerados os aspectos do tipo: escolha da estrutura de armazenamento dos dados (matricial ou vetorial), ou quais camadas lógicas (layers) serão criadas. Por exemplo, um tema como *hidrografia*, pode dar origem a mais de uma camada lógica (layer), uma

contendo linhas representando os rios e outra contendo polígonos representando os lagos. O projeto lógico de um SIG depende do software que será utilizado.

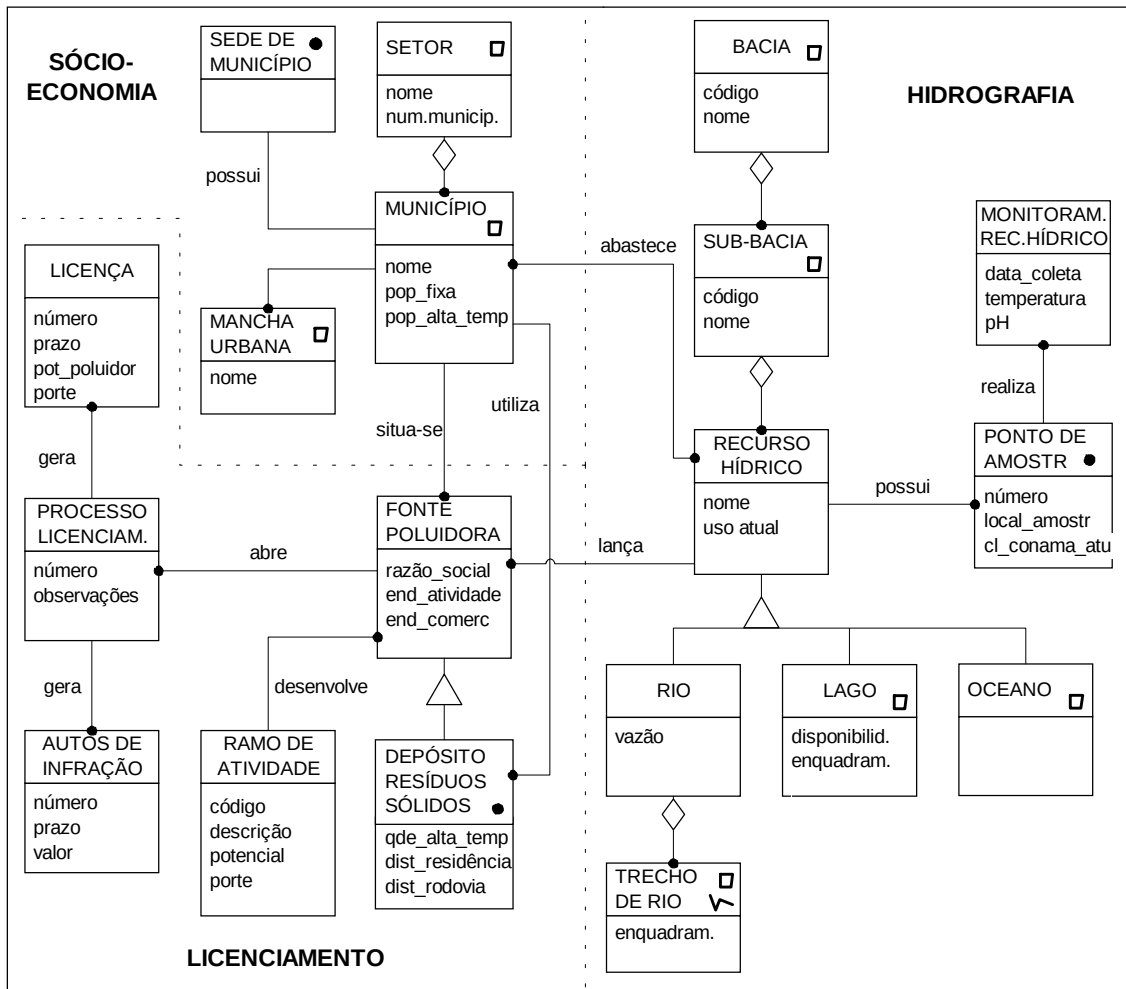


FIGURA 5 - Modelagem de dados na visão de objetos

5 CONSIDERAÇÕES FINAIS

Demonstrou-se a validade prática do modelo OMT, com algumas adaptações, na modelagem conceitual de um problema real de aplicação geográfica em SIG. Como trabalhos futuros, pretende-se analisar as técnicas de mapeamento do modelo lógico para um modelo de implementação em SIG.

Agradecimentos

Gostaríamos de agradecer à equipe do Projeto GERCO da FEPAM-RS, pela oportunidade de participação do processo de modelagem de uma aplicação real.

3.3 Artigo IV: Metodologia para Modelagem Conceitual de Dados de um SIG Interinstitucional

Jugurta Lisboa Filho, Íria Margarida Garaffa* e Cirano Iochpe

*UFRGS - Programa de Pós-Graduação em Computação

In: GISBRASIL - CONGRESSO E FEIRA PARA USUÁRIOS DE GEOPROCESSAMENTO, 3., 1997, Curitiba. **Anais...** Curitiba: Fator GIS, 1997.

RESUMO

Este trabalho apresenta a metodologia de desenvolvimento empregada na modelagem conceitual dos dados do *SIGPROGB* - Sistema de Informações Geográficas do *Pró-Guaíba*, um Programa para o Desenvolvimento Racional, Recuperação e Gerenciamento Ambiental da Bacia Hidrográfica do Guaíba. O processo de modelagem está sendo realizado em três fases. Na primeira fase foi realizado o treinamento em projeto de banco de dados e modelagem conceitual de aplicações geográficas, de técnicos das instituições co-executoras do Programa. A segunda fase consistiu na execução da modelagem conceitual dos dados descritivos dessas instituições e na terceira fase está sendo desenvolvido a modelagem conceitual dos dados espaciais do *SIGPROGB* e integração final de visões.

ABSTRACT

This paper presents the methodology of development employed on conceptual data modeling of *SIGPROGB* - Geographic Information System of *Pró-Guaíba*, a Program for the Rational Development, Recovery and Environmental Management of Guaíba's Hydrographic Basin. The process of modeling have tree phases. In the first phase was given a group training in database design and geographical application conceptual modeling, for the personnel of institutions partnership. The second phase consisted in the execution of descriptive data conceptual modeling, and in the third it has being developed the spatial data conceptual modeling and finish vision integration of *SIGPROGB*.

1 INTRODUÇÃO

O Estado do Rio Grande do Sul, através da Secretaria de Coordenação e Planejamento - SCP, elaborou o Programa para o Desenvolvimento Racional, Recuperação e Gerenciamento Ambiental da Bacia Hidrográfica do Guaíba (*Pró-Guaíba*) [SEC 92]. O objetivo do *Pró-Guaíba* é criar para a área da Bacia condições necessárias para a utilização racional de seus recursos naturais, recuperação da qualidade ambiental nas áreas urbanas e rurais, bem como o manejo ambiental sustentado da produção agrícola, pecuária, florestal e industrial da Bacia. As atividades de gerenciamento e monitoramento do *Pró-Guaíba* serão realizadas com base em informações mantidas pelo Sistema de Informações Geográficas do *Pró-Guaíba*

(SIGPROGB), que foi concebido através do Plano Diretor de Geoinformações (PDG) [SEC 92a].

A área geográfica de abrangência do Pró-Guaíba compreende a Bacia Hidrográfica do Guaíba, caracterizada como a mais importante do Estado do Rio Grande do Sul, com uma área de 85.950 Km², cerca de 30% da área do Estado. Abrange 222 municípios, entre eles a capital, Porto Alegre, e concentra aproximadamente 6 milhões de habitantes, correspondendo a 2/3 da população estadual.

Este trabalho tem como objetivo apresentar a metodologia empregada na modelagem conceitual de dados do SIGPROGB. Esta modelagem está sendo desenvolvida, de forma conjunta, pela SCP, ECOPLAN (Gerenciadora do Programa) e por pesquisadores do Instituto de Informática da Universidade Federal do Rio Grande do Sul - UFRGS.

O conjunto de informações que compõem o SIGPROGB é constituído pelas bases cartográficas e pelos bancos de dados pertencentes a onze instituições públicas (municipais e estaduais) que são co-executoras do Programa. Cada instituição participa do sistema como geradora ou usuária de determinado conjunto de informações, o que torna o SIGPROGB um sistema bastante complexo.

2 DESENVOLVIMENTO

O desenvolvimento da modelagem conceitual de dados do SIGPROGB teve como base o PDG, onde está definida a filosofia do sistema, as fontes de dados e os usuários responsáveis. A execução do processo de modelagem está sendo realizada em três fases.

Na primeira fase foi realizado o treinamento dos técnicos, envolvidos no SIGPROGB, de cada uma das instituições co-executoras. Este treinamento teve como objetivo primordial, o nivelamento do conhecimento dos técnicos em projeto de banco de dados e modelagem conceitual de aplicações geográficas, o que contribuiu sobremaneira no processo de elaboração da modelagem conceitual, permitindo a unificação de linguagem entre os diversos usuários e os projetistas. Foram ministrados pela UFRGS, cursos sobre projeto de sistemas de banco de dados, linguagem de consulta SQL e sobre uma metodologia de modelagem de dados baseada no paradigma de orientação a objetos (OMT - Técnica de Modelagem de Objetos), com ênfase em aplicações geográficas.

A segunda fase consistiu na execução da modelagem conceitual dos dados descritivos das onze instituições co-executoras. Inicialmente foram identificados grupos de instituições afins, ou seja, instituições que geram ou utilizam os mesmos tipos de dados (ex.: existem na área do Pró-Guaíba duas instituições que são concessionárias de distribuição de água). Esta divisão em grupos resultou na obtenção de diversos modelos conceituais. Posteriormente, através de uma etapa de integração de visões gerou-se um modelo conceitual integrado de todo o sistema. Para o desenvolvimento desta fase foi utilizado como software de apoio, a ferramenta CASE *Rational Rose*.

Na terceira fase, ainda em andamento, está sendo desenvolvido a modelagem conceitual dos dados espaciais do SIGPROGB. Numa primeira etapa estão sendo modelados os dados pertencentes à cartografia básica. Muitas das classes de entidades identificadas nesta etapa já haviam sido modeladas como dados descritivos, restando, portanto, a integração com os aspectos espaciais. Para a construção deste modelo estão sendo consideradas as bases cartográficas em suas respectivas escalas (1:1.000, 1:10.000, 1:50.000, 1:250.000 e 1:1000.000), as aplicações definidas no Plano Diretor de Geoinformações, que serão desenvolvidas por cada instituição co-executora e o modelo conceitual dos dados descritivos, elaborado na fase 2.

Como complemento desta terceira fase, está prevista uma nova etapa de integração de visões, o que resultará no modelo conceitual integrado (dados descritivos e espaciais) do banco de dados geográfico do SIGPROGB.

Em paralelo às fases de treinamento e execução da modelagem, foram realizadas reuniões sistemáticas com apoio de profissional da área de psicologia organizacional, com o objetivo de promover uma integração dos técnicos das diversas instituições co-executoras, a fim de diluir resistências pessoais e políticas, normalmente existentes em um ambiente interinstitucional.

3 METODOLOGIA

Modelos de dados são visões simplificadas de uma parte da realidade e são construídos de acordo com certas regras para facilitar a implementação do banco de dados em um sistema de informação [LIS 96b]. A utilização de modelos conceituais possibilita ao usuário abstrair os elementos da realidade que são importantes para a aplicação, sem se preocupar com os aspectos inerentes ao software de SIG que será utilizado para implementar a aplicação.

O projeto conceitual do SIGPROGB está sendo desenvolvido com base nos modelos orientados a objetos, cujo simbolismo gráfico utilizado é o da metodologia OMT (Técnica de Modelagem de Objetos) [RUM 91].

O modelo OMT é baseado nos conceitos do paradigma da orientação a objetos (objetos, classes, atributos e operações) e emprega conceitos de abstração semântica como classificação, generalização e agregação. Este modelo é definido pelos próprios autores, como uma tentativa de aperfeiçoamento do modelo E-R [CHE 76].

Devido às características espaço-temporais dos dados manipulados pelas aplicações geográficas, essas aplicações impõem alguns requisitos especiais aos SIGs, sendo que nem todos são suportados adequadamente pelos SIGs existentes atualmente.

Entre estes requisitos, o principal é o que diz respeito à característica espacial das entidades geográficas. Este requisito é resolvido em todos os SIGs, com maior ou menor potencialidade, a partir do armazenamento e recuperação da localização das entidades no espaço, suas formas geométricas e seus relacionamentos com outras entidades.

No entanto, a realidade geográfica é muito complexa e, em muitos casos, apenas o suporte aos requisitos espaciais não é suficiente, como por exemplo, em situações em que os aspectos temporais são relevantes, quando as entidades geográficas não possuem limites bem definidos, entre outras coisas. Em [LIS 96], relacionamos alguns destes requisitos especiais, apresentados pelas aplicações de SIG. São eles:

- Dicotomia entre campos e objetos
- Relacionamentos espaciais
- Tratamento unificado dos dados espaciais e não-espaciais
- Qualidade e disponibilidade dos dados
- Temporalidade e representação de limites difusos

Destes requisitos, o primeiro é o mais importante em relação à modelagem conceitual. Para Goodchild [GOO 92], a realidade geográfica pode ser observada segundo duas visões: de *campo* e de *objetos*. Na visão de campo, a realidade é modelada por variáveis que possuem uma distribuição contínua no espaço. Todas as posições no espaço geográfico podem ser caracterizadas através de um conjunto de atributos, como por exemplo, temperatura, tipo de solo e relevo, medidos para um conjunto de coordenadas geométricas no espaço Euclidiano.

Na visão de objetos a realidade consiste de entidades individuais, bem definidas e identificáveis. Cada entidade tem suas propriedades e ocupa um determinado lugar no

espaço. A realidade é modelada como um grande espaço onde entidades estão distribuídas sem que necessariamente todas as posições estejam ocupadas. Além disso, duas ou mais entidades podem estar situadas sobre uma mesma posição geográfica. Na visão de objetos, as entidades são representadas espacialmente por objetos geométricos (ou objetos espaciais) do tipo ponto, linha e polígono.

Um banco de dados geográfico pode ser visto como sendo um repositório de dados sobre um conjunto de áreas geográficas. Cada área geográfica possui dados sobre diversos temas, que podem ser modelados segundo as visões de campo e objetos. A Figura 1, mostra uma hierarquia de classes básicas, que é uma adaptação das hierarquias descritas em [TIM94] e [CAM 94], que serviu de ponto de partida para a modelagem conceitual do SIGPROGB. Esta biblioteca permite a integração natural dos dados nas visões de campo e objetos.

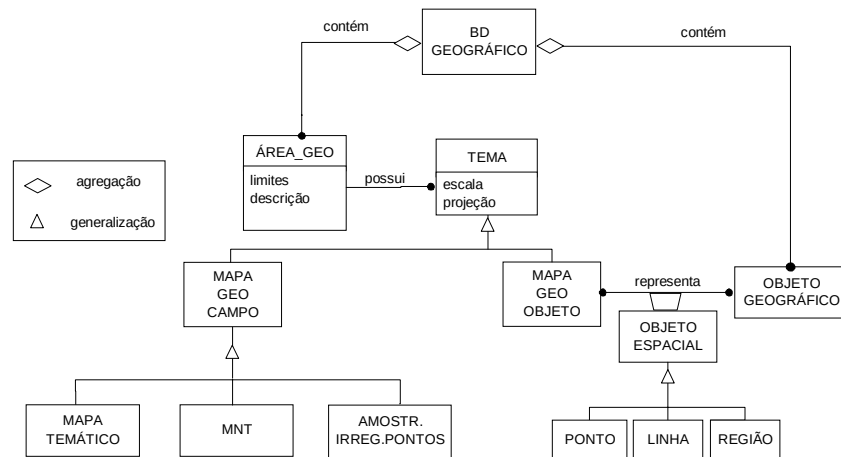


FIGURA 1 - Biblioteca de Classes Básicas

O esquema utilizado na modelagem do SIGPROGB está permitindo a construção de um modelo conceitual de dados com as seguintes características:

- As entidades estão totalmente integradas, independente se elas possuem, ou não, atributos espaciais.
- O tipo de objeto espacial associado às entidades geográficas varia de acordo com a escala utilizada. O modelo permite representar as entidades geográficas associadas com suas diferentes representações espaciais.
- O resultado do projeto conceitual constitui-se de um conjunto de diagramas de classes fortemente relacionadas, permitindo o gerenciamento da complexidade do modelo final.
- O modelo independe do software de SIG que será utilizado, podendo ser facilmente mapeado (transformado) em qualquer modelo de implementação existente.

4 CONCLUSÕES

O artigo descreve a metodologia de desenvolvimento utilizada na modelagem conceitual dos dados do SIGPROGB, uma aplicação geográfica que engloba informações de onze instituições ligadas ao Pró-Guaíba. São descritas as atividades executadas nas três fases do desenvolvimento da modelagem e apresentada a metodologia que está sendo utilizada na modelagem conceitual dos dados, a partir da utilização de uma biblioteca de classes básicas.

O modelo conceitual do SIGPROGB possui até o momento aproximadamente 250 classes. A utilização de ferramenta CASE foi imprescindível na modelagem, possibilitando a criação de esquemas gráficos de boa qualidade e geração do dicionário de dados.

3.4 Artigo V: Modelos Conceituais de Dados para Aplicações Geográficas: uma experiência com um SIG interinstitucional

Jugurta Lisboa Filho, Cirano Iochpe e Íria Margarida Garaffa*

*UFRGS - Programa de Pós-Graduação em Computação

In: SIMPÓSIO BRASILEIRO DE GEOPROCESSAMENTO, 4., 1997, São Paulo. **Anais...** São Paulo: UPUSP, 1997.

RESUMO

O artigo descreve os problemas encontrados na modelagem conceitual dos dados do *SIGPROGB* (*Sistema de Informações Geográficas do Programa Pró-Guaíba*), um SIG composto de dados georreferenciados compartilhados por onze instituições públicas (municipais e estaduais), co-executoras do Programa. As aplicações geográficas apresentam um conjunto de requisitos especiais que impõem novos desafios à modelagem dos dados no contexto de SIG. Entre estes requisitos, inclui-se: dicotomia entre as visões de campo e de objetos; relacionamentos espaciais; temporalidade; múltiplas representações; e qualidade dos dados espaciais. São apresentadas algumas propostas de solução existentes na literatura para a modelagem desses requisitos no projeto de aplicações de SIG. Uma nova fase de modelagem, denominada *Projeto de Representação Espaço-Temporal* foi identificada. Através de exemplos, o resultado da modelagem dos dados do *SIGPROGB* é apresentado, modelagem esta realizada, utilizando-se de uma abordagem orientada a objetos a partir de uma hierarquia de classes abstratas.

ABSTRACT

This paper describes the problems found during the conceptual data modeling of *SIGPROGB*, Pro-Guaiba Program's geographical information system, a GIS that shall maintain georeferenced data to be shared by eleven different public organizations. Geographical applications pose specific requirements to data modeling that traditional data models cannot cope with. Some of them are: dichotomy between field and object vision; spatial relationships; temporal dimension of data; multiple representations of data; and quality information about spatial data. Some proposals of data model extensions are presented and discussed upon. In addition, a new phase in the data modeling process, the spatial-temporal representation project, is introduced. Finally, the main results of *SIGPROGB*'s database project which relies on an object-oriented, abstract class hierarchy are presented and illustrated through examples.

1 INTRODUÇÃO

O termo *Sistemas de Informações Geográficas* (SIG) caracteriza os sistemas computacionais capazes de capturar, armazenar, analisar e exibir dados *georreferenciados*, isto é, dados referenciados espacialmente em relação à superfície da

Terra [WOR 94]. Os SIG manipulam dados espaciais, normalmente representados na forma de mapas, associados a dados não espaciais, denominados textuais ou descritivos. Devido à complexidade estrutural e operacional dos dados manipulados pelos SIG, suas aplicações têm sido classificadas como *não-convencionais*.

Segundo [ROM 90], um dos componentes mais importantes de qualquer especificação de requisitos é a modelagem dos dados. No entanto, as técnicas tradicionais de modelagem de dados e suas ferramentas não fornecem meios adequados para representar completamente a realidade das informações geográficas. A dificuldade maior está no fato de que muitas destas informações precisam ser qualificadas com respeito à localização onde são válidas, o tempo de observação e a sua acurácia.

Diversos trabalhos de pesquisa vêm sendo desenvolvidos com o objetivo de se encontrar um modelo conceitual que seja, ao mesmo tempo, rico o bastante para representar a complexa realidade das aplicações geográficas e o mais simples possível, para que possa ser entendido e utilizado pelos usuários de SIG [CAR 93, HAD 96, KOS 97, OLI 97, SHE 97]. Uma forte tendência tem sido verificada no sentido de se utilizar os conceitos da orientação a objetos para o projeto de aplicações em SIG [WOR 94].

No desenvolvimento do modelo conceitual do *SIGPROGB*, um SIG interinstitucional, foram identificados alguns tipos de requisitos especiais que não são representados adequadamente pelos modelos conceituais disponíveis para o projeto de aplicações convencionais. Além disso, uma nova fase de projeto de banco de dados geográfico (BDG)⁶ mostrou-se necessária.

O texto está organizado como segue. A Seção 2 relaciona um conjunto de requisitos básicos de modelagem, específicos das aplicações de SIG. Na Seção 3, são descritos alguns modelos conceituais propostos, exclusivamente, para aplicações de SIG. A Seção 4 descreve as fases a serem seguidas no projeto do BDG. A Seção 5 apresenta a metodologia empregada na modelagem conceitual do *SIGPROGB*. Finalmente, na Seção 6 são apresentadas as conclusões do estudo e as perspectivas de trabalho futuro.

2 REQUISITOS BÁSICOS DE MODELAGEM DE DADOS ESPECÍFICOS DAS APLICAÇÕES GEOGRÁFICAS

A realidade geográfica normalmente é bem mais complexa do que a realidade modelada em aplicações convencionais. A seguir, estão descritos alguns requisitos especiais a serem observados no projeto de aplicações geográficas, identificados a partir de um estudo bibliográfico e da experiência dos autores no projeto de aplicações para SIG [LIS 96, LIS 97].

Dicotomia entre Visão de Campo e Visão de Objetos - modelos de dados geográficos devem refletir a maneira como as pessoas vêem o mundo. Um dos princípios filosóficos da percepção humana dos fenômenos geográficos é que a realidade é composta de entidades exatas e de superfícies contínuas [BUR 95].

Para [GOO 92], a realidade geográfica pode ser observada segundo duas visões: de *campo* e de *objetos*. Na visão de campo, a realidade é modelada por variáveis que possuem uma distribuição contínua no espaço. Cada posição do espaço geográfico, em análise, pode ser caracterizada através de um conjunto de atributos como, por exemplo,

⁶ O termo banco de dados geográfico (BDG) será utilizado para identificar o banco de dados do SIG, independente se este será gerenciado por um sistema de gerência de banco de dados (SGBD) ou um gerente de arquivos.

temperatura, tipo de solo e relevo, medidos para um conjunto de coordenadas geográficas.

Na visão de objetos, a realidade consiste de entidades individuais, bem definidas e identificáveis. Cada entidade tem suas propriedades e ocupa um determinado lugar no espaço. A realidade é modelada como um grande espaço onde entidades estão distribuídas de forma discreta.

Na visão de objetos, as entidades geográficas são representadas espacialmente por objetos espaciais do tipo *ponto*, *linha* e *polígono*. Cada entidade geográfica possui uma localização definida por uma referência à superfície da terra. Dependendo da escala do mapa onde é apresentada, a entidade geográfica pode ser representada por um objeto espacial distinto. Exemplo, na escala 1:2.000, uma igreja pode ser representada por um polígono e na escala 1:10.000, por um ponto.

Alguns modelos propõem, de forma equivocada, a modelagem de objetos geográficos como especializações dos objetos espaciais, herdando suas propriedades [CLE 94]. Por exemplo, um objeto *ESTRADA* não deve herdar as propriedades de um objeto espacial *LINHA*, pois uma estrada não é uma linha, mas sim, um objeto complexo que possui diversos atributos e relacionamentos, entre os quais inclui-se relacionamentos com os objetos espaciais que o representam.

Relacionamentos Espaciais - os bancos de dados armazenam elementos de informação de dois tipos principais: dados que descrevem as *entidades* que são relevantes em um domínio de aplicação; e dados que descrevem os *relacionamentos* entre tais entidades. Uma das tarefas mais importantes quando se está modelando os dados de uma aplicação é a identificação de quais os relacionamentos presentes, na realidade, que deverão ser mantidos no banco de dados.

No domínio das aplicações geográficas, este problema é bem mais complexo, uma vez que o número de relacionamentos possíveis de serem mantidos no BDG é ainda maior, devido à existência dos relacionamentos espaciais entre as entidades. Segundo [CLE 94], em um contexto geográfico as entidades estão cercadas por outras entidades, o que torna o estudo dos relacionamentos espaciais entre entidades uma questão altamente relevante na modelagem de dados geográficos.

Temporalidade - a maioria dos SIG disponíveis atualmente consideram as entidades como se o mundo existisse somente no presente. Informações são incluídas e alteradas ao longo do tempo, mas o histórico dessas transformações não é mantido no BDG. Este problema vem recebendo uma atenção crescente devido à necessidade de um melhor entendimento dos processos geográficos e dos interrelacionamentos de causa e efeito entre as atividades humanas e o meio ambiente [PEU 95].

A necessidade dos dados geográficos estarem qualificados com base no tempo, não se deve ao fato dos dados serem freqüentemente modificados, mas sim, pela necessidade de se registrar estados passados, de forma a possibilitar o estudo da evolução dos fenômenos geográficos (ex.: estudos de séries históricas para entender a evolução de um fenômeno qualquer) [HAD 96].

A utilização de um modelo espaço-temporal aumenta a potencialidade das operações de análise através de um SIG. Segundo [LAU 92], o fator tempo necessita ser medido ao longo de, pelo menos, duas dimensões: *tempo de transação* - momento da entrada do dado no BDG; e *tempo do evento* - momento em que o evento ocorre na realidade. Diversos modelos integrando os aspectos espaço-temporais em BDG têm sido propostos na literatura [PEU 95, REN 97].

Múltiplas Representações - a existência de múltiplas representações para uma mesma entidade está associada à complexidade da realidade a ser representada. Uma

entidade pode ser mantida, ao mesmo tempo, no BDG, em diferentes escalas ou projeções, e representada por diferentes objetos espaciais.

Segundo [CAM 96], o gerenciamento de múltiplas representações em um SIG deve ser considerado nos níveis de modelagem e de estruturas de dados. Problemas decorrentes da existência de múltiplas representações em um SIG são: redundância de dados; inconsistências do banco de dados; e multiplicidade de comportamento de um mesmo fenômeno geográfico.

Qualidade dos Dados - dados com erros sempre existem nos SIG, mas precisam ser identificados e tratados. Os erros podem ser introduzidos no BDG de diversas formas: serem decorrentes de erros nas fontes originais; serem adicionados durante os processos de obtenção (captura) e armazenamento; serem gerados durante a exibição ou impressão dos dados; ou surgirem a partir de resultados de operações de análise dos dados [BUR 95].

Acrescente-se aos possíveis problemas acima o *envelhecimento* ou a *obsolescência* a que estão sujeitos os dados geográficos. Por exemplo, devido aos elevados custos da produção cartográfica, muitas cidades brasileiras dispõem somente de cartas defasadas no tempo. Sendo assim, análises baseadas em informações ultrapassadas geram resultados com baixo grau de confiabilidade.

Aplicações geográficas também se diferenciam das demais aplicações de banco de dados por fazerem uso de grande volume de dados importados de outros sistemas, muitas vezes de outras organizações. Este fenômeno gera, freqüentemente, problemas de confiabilidade dos dados.

3 MODELOS DE DADOS PARA APLICAÇÕES EM SIG

Nesta seção são descritas as principais características presentes em alguns modelos conceituais propostos para a modelagem de aplicações de SIG.

3.1 Modelo Modul-R

Em [CAR 93] é descrita uma extensão do modelo E-R para aplicações geográficas urbanas, denominado *Formalismo Modul-R*. O modelo Modul-R tem como objetivo estender o modelo E-R para tratar os seguintes requisitos: *referência espacial*, *referência temporal*, *complexidade de banco de dados* e *dicotomia entre dados e processos*.

Modul-R é baseado em uma estrutura modular em forma de pirâmide (Figura 1), onde o projetista escolhe, de acordo com o contexto de cada aplicação, um conjunto de componentes semânticos que comporão uma linguagem integrada.

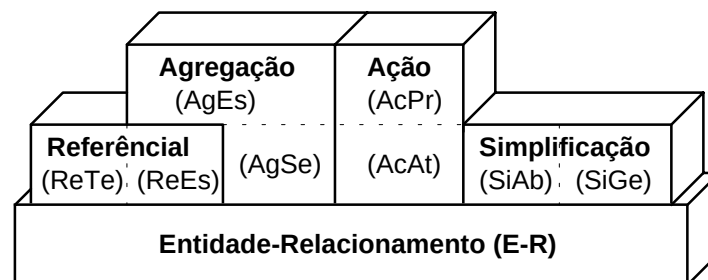


FIGURA 1 - Módulos do modelo Modul-R [CAR 93]

O modelo E-R forma a base da pirâmide e quatro novos módulos são adicionados. São eles:

- *Módulo Referencial (Re)* - permite que entidades e relacionamentos sejam referenciados no espaço e no tempo. Possui dois submódulos: Referência Espacial (ReEs) e Referência Temporal (ReTe).
- *Módulo Agregação (Ag)* - permite enriquecer o modelo facilitando sua expressão com base na tecnologia de orientação a objetos. Inclui os submódulos Agregação Semântica (AgSe) e Agregação Espacial (AgEs).
- *Módulo Ação (Ac)* - permite criar uma conexão entre a modelagem dos dados e dos processos, através de uma ligação entre elementos do modelo de dados conceitual e elementos de um diagrama de fluxo de dados (DFD). Compreende os submódulos Atividades (AcAt) e Processos (AcPr).
- *Módulo Simplificação (Si)* - permite a simplificação de esquemas complexos através dos submódulos Abstração (SiAb) e Generalização (SiGe).

O módulo E-R fornece os componentes básicos do modelo, ou seja, entidades, relacionamentos, atributos e cardinalidades, mas utiliza uma notação gráfica própria.

3.2 Modelo GMOD

[OLI 97] descreve a arquitetura do UAPE, um ambiente de desenvolvimento de aplicações geográficas que integra modelagem de dados e processos, com base em tecnologias de engenharia de software e de banco de dados. O objetivo do ambiente UAPE é fornecer suporte para dois tipos de atividade do usuário de SIG: projeto de aplicações e de banco de dados; e análise e manipulação de dados geográficos. O ambiente combina uma metodologia de projeto de aplicações com um modelo de dados semântico de alto nível, denominado GMOD, voltado para aplicações geográficas.

No modelo GMOD, uma aplicação geográfica possui dois tipos de classes: classes que representam entidades associadas a alguma localização geográfica (*Geo-Classe*); e classes que não possuem referência espacial, mas estão relacionadas com as primeiras (*Classes Convencionais*).

Toda instância de *Geo-Classe* possui um atributo espacial que é instância da classe *Geo-Região*, que descreve regiões da superfície terrestre segundo uma escala e uma projeção. As *Geo-Classes* são especializadas nas classes *Geo-Campo* e *Geo-Objeto*, as quais descrevem fenômenos geográficos observados segundo as visões de campo e objetos, respectivamente.

O modelo GMOD possui duas hierarquias de classes de representação, cujas raízes são as classes *RepObjeto* e *RepCampo*. A classe *RepObjeto* é especializada em classes cujas instâncias possuem atributos onde os valores são objetos espaciais. A classe *RepCampo* é especializada em subclasses correspondentes aos modelos de representação descritos em [GOO 92], recebendo denominações como *Grid*, *TriNet*, *Contour*, *PlanarSubdivision* e *PointSample*.

Geo-Classes e classes *Convencionais* podem ser definidas como sendo *temporais* (permite o gerenciamento da evolução dos fenômenos geográficos), através da definição de uma associação com a classe *Tempo*. Uma classe temporal de geo-objetos complexos permite a variação no tempo, tanto do geo-objeto complexo como de seus componentes.

Dois novos tipos de relacionamentos são definidos: de *versão* e *causal*. Relacionamentos de *versão* são usados para associar diferentes versões de uma mesma entidade. Podem ser usados, por exemplo, para relacionar diferentes representações espaciais de uma mesma entidade geográfica. Um relacionamento *causal* estabelece uma ligação de causa-efeito entre dois fenômenos modelados. Por exemplo, áreas de erosão estão diretamente relacionadas ao tipo de solo e à declividade do terreno.

3.3 Modelo GeoOOA

O modelo GeoOOA estende o modelo usado na metodologia OOA [COA 91] para análise de requisitos de aplicações em SIG [KOS 97]. As extensões incluídas pelo modelo GeoOOA permitem a representação dos seguintes contextos semânticos:

- Distingue entre classes de entidades com e sem representação espacial;
- Utiliza pictogramas gráficos para simbolizar o relacionamento entre uma classe de entidade geográfica e a classe de objeto espacial que a representa;
- Distingue os relacionamentos espaciais (ex.: topológicos) dos relacionamentos básicos (ex.: proprietário possui lote).
- As primitivas do modelo GeoOOA suportam abstrações do tipo classe espacial, estrutura todo-parte espacial, estruturas de rede e classe temporal.

3.4 Modelo GeoIFO

[TRY 95] propõe uma extensão do modelo IFO, denominada GeoIFO, onde um conjunto mínimo de novos construtores semânticos é introduzido para representar esquemas conceituais de BDG. Os conceitos de *posição de objetos* e *atributos variando no espaço*, que se adequam naturalmente aos modelos semânticos e orientados a objetos, possibilitam integrar as abordagens de campo e objetos.

No modelo GeoIFO, objetos do mundo real estão ligados ao espaço através de sua posição. A *posição de um objeto* inclui sua localização, forma, tamanho e orientação. Em modelos semânticos, a *posição de um objeto* pode ser modelada como um tipo especial de atributo, assim, *posição* seria uma função que retornaria para cada objeto geográfico uma parte do espaço.

Alguns fenômenos do mundo real estão relacionados diretamente ao espaço e não pertencem a nenhum objeto particular, como temperatura, relevo e índice pluviométrico. *Atributos variando no espaço* são propriedades do espaço. Por exemplo, em um sistema cadastral, o grau de declividade do terreno pode ser modelado como um atributo do lote. Porém, as dimensões do lote podem ser modificadas, influenciando no grau de declividade.

O modelo GeoIFO fornece uma representação semântica para o projeto de esquemas a nível conceitual. Uma linguagem com sintaxe definida formalmente é descrita em [HAD 96]. Esta linguagem equivale a uma DDL (*Data Definition Language*) e foi elaborada para permitir o mapeamento de um esquema semântico usando GeoIFO em um esquema lógico genérico para BDG.

Para [HAD 96], uma camada lógica é um conceito que pode ser usado para representar conjuntos de entidades geográficas. Camadas lógicas não necessariamente possuem um relacionamento um-para-um com camadas físicas, as quais são determinadas levando-se em consideração aspectos de eficiência do sistema. Por exemplo, uma única camada lógica denominada *RECURSOS_HÍDRICOS* pode incluir *LAGOS*, *RIOS* e *NASCENTES* e ser mapeada em três camadas físicas distintas.

3.5 Modelo GISER

[SHK 97] propõe um novo modelo chamado GISER (*Geographic Information System Entity Relational*), que é uma extensão ao modelo EER [ELM 94] para aplicações geográficas.

O modelo GISER é baseado em quatro conceitos básicos: *espaço-tempo*, *feições*, *coberturas* e *objeto espacial*. As *feições* representam as entidades geográficas que são

discretizadas em elementos denominados *coberturas*. Uma *cobertura* consiste de um conjunto de *objetos espaciais*, distribuídos no *tempo* e no *espaço*.

Uma aplicação é modelada a partir da criação de um esquema conceitual, onde entidades geográficas são relacionadas com entidades predefinidas que descrevem os principais elementos de um BDG como *polígono*, *espaço* e *objeto espacial*. Não são adicionadas novas primitivas ao modelo EER.

3.6 Avaliação dos Modelos

Dos modelos analisados, nenhum atende a totalidade dos requisitos descritos na seção 2. Os modelos GMOD, GeoIFO e GISER fornecem mecanismos para a modelagem integrada nas visões de campo e objetos. Relacionamentos espaciais são bem suportados pelos modelos GeoOOA e GISER. Apesar dos modelos Modul-R, GMOD e GeoOOA abordarem parcialmente os aspectos temporais, pesquisas específicas sobre BDG-Temporais estão sendo desenvolvidas [PEU 95, REN 97]. O requisito de múltiplas representações é abordado pelo modelo GMOD através do uso de relacionamentos de versões, mas de forma muito superficial. Por último, nenhum dos modelos analisados suporta a modelagem de dados sobre a qualidade dos dados utilizados (ex.: metadados de qualidade).

4 FASES DO PROJETO DE BANCO DE DADOS GEOGRÁFICO

Segundo [ELM 94], o projeto de banco de dados compreende três fases: *conceitual*, *lógico* e *físico*. O esquema conceitual de dados, gerado na fase de projeto conceitual, tem como objetivo representar **quais** as entidades (descritas por seus atributos) que serão armazenadas no banco de dados e os relacionamentos existentes entre essas entidades. O esquema conceitual é descrito em uma linguagem de alto nível (ex.: modelo E-R) e não considera os aspectos de implementação, o que o torna totalmente independente do software a ser utilizado. O esquema lógico, gerado a partir do esquema conceitual, define **como** as entidades serão armazenadas através dos tipos de dados disponíveis no SGBD (ex.: tabelas em SGBD relacional). Por último, no esquema físico são definidos detalhes de implementação dos dados (ex.: meios de armazenamento e índices de acesso).

Para [LAU 92], a utilização de diferentes níveis de modelagem é essencial para o projeto de BDG. No entanto, a maioria das aplicações geográficas têm sido projetadas com a atenção voltada, principalmente, para a escolha do formato de representação física dos dados espaciais (formato matricial ou vetorial) e pouca importância tem sido dada à integração com a modelagem das entidades não espaciais [LIS 96].

A partir da análise dos modelos descritos anteriormente e do desenvolvimento da modelagem do *SIGPROGB* (seção 5), concluiu-se que o projeto de BDG requer uma etapa adicional (Figura 2), que permita a especificação de questões relacionadas com a representação espaço-temporal das entidades geográficas. A seguir estão descritas cada uma das etapas propostas para o projeto do BDG.

Projeto Conceitual - nesta fase, a modelagem da aplicação geográfica deve ser realizada de forma tradicional, onde entidades importantes ao domínio da aplicação são identificadas e relacionadas entre si. Detalhes sobre o tipo de representação espacial não devem ser considerados. Segundo [REN 97], a modelagem dos aspectos temporais pode ser postergada. O esquema de dados resultante deve estar escrito em uma linguagem de alto nível, baseada em primitivas de abstração de dados como: *classificação*, *generalização*, *agregação* e *associação*.

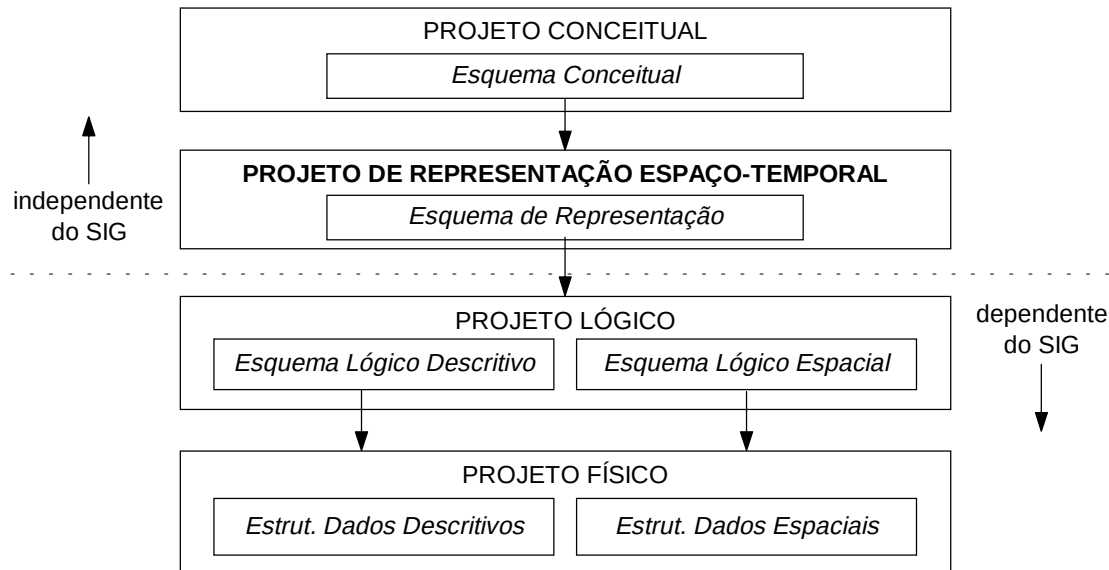


FIGURA 2 - Fases do Projeto de Banco de Dados Geográfico

Projeto de Representação Espaço-Temporal - como mencionado anteriormente, em um SIG as entidades geográficas são representadas, graficamente, por meio de objetos espaciais. Além disso, elas devem estar relacionadas com o período (ou instante) para o qual a informação é válida.

Na fase do projeto de representação espaço-temporal novos detalhes são adicionados ao esquema de dados. São eles:

- Diferenciação entre as classes com e sem representação espacial. Exemplo, a classe *PROPRIEDADE-RURAL* possui representação espacial (ex.: polígono), enquanto que a classe *PROPRIETÁRIO* não possui representação espacial.
- Para as classes que se enquadram na visão de objetos (ver seção 2), deve-se definir os tipos de objetos espaciais que serão usados para representá-la, sendo que cada tipo estará associado a uma escala de representação.
- Na visão de campo, classes representando os fenômenos geográficos devem ser modeladas como especialização em uma hierarquia de classes abstratas, conforme proposto no modelo GMOD [OLI 97].
- Por último, são adicionados os aspectos temporais, ou seja, é feita uma associação entre as classes que representam entidades geográficas e as classes que descrevem o tempo [KOS 97].

Projeto Lógico - o esquema de representação espaço-temporal é transformado em um esquema lógico de dados. Como a maioria dos SIG disponíveis atualmente utilizam uma arquitetura dual, ou seja, possuem dois componentes distintos, um para o gerenciamento dos dados espaciais e outro para os descritivos, a fase do projeto lógico pode ser dividida em duas subfases⁷: *Projeto Lógico Descritivo* (PLD) e *Projeto Lógico Espacial* (PLE). No PLD, técnicas convencionais de projeto de banco de dados podem ser empregadas. No PLE, torna-se necessário um mapeamento do esquema de representação espacial em um esquema específico proprietário do software de SIG a ser empregado. Por exemplo, o sistema *MGE* da *Intergraph* utiliza os conceitos de *PROJETO*, *CATEGORIA* e *FEIÇÃO* para organizar hierarquicamente os dados representados em uma determinada escala.

⁷ Um SIG baseado em uma arquitetura integrada (ex.: arquitetura de objetos) teria uma única subfase de projeto lógico de dados e um único conjunto de estruturas de dados.

Projeto Físico - nesta etapa, além dos detalhes de implementação tradicionais, são definidas as estruturas de dados espaciais que serão utilizadas como, por exemplo, se um mapa de propriedades rurais será armazenado em uma estrutura de dados vetorial (com/sem topologia), matricial ou ambas.

A formalização de uma linguagem para a modelagem dos dados na fase de projeto de representação espaço-temporal é tema atual de pesquisa dos autores. [CAM 96] descreve um Nível de Representação, onde as classes de representação propostas estão relacionadas a um SIG específico, o que equivale à fase do *Projeto Lógico Espacial* da Figura 2.

5 O EXEMPLO DO SIGPROGB

Esta seção descreve alguns resultados da modelagem conceitual do SIGPROGB, o SIG que dará suporte às atividades de gerenciamento e monitoramento do Pró-Guaíba (Programa para o Desenvolvimento Racional, Recuperação e Gerenciamento Ambiental da Bacia Hidrográfica do Guaíba). O conjunto de informações que compõem o SIGPROGB é constituído pelas bases cartográficas e respectivos bancos de dados pertencentes a onze instituições públicas (municipais e estaduais) que são co-executoras do Programa. Esta característica interinstitucional torna o SIGPROGB um sistema altamente complexo.

O SIGPROGB está sendo desenvolvido conforme as etapas descritas na seção anterior. Para o esquema conceitual adotou-se a notação da metodologia OOA [COA 91]. Devido ao grande número de entidades (~300), o esquema foi subdividido em sub-esquemas, onde as entidades foram agrupadas de acordo com grupos de instituições afins, que são geradoras dos dados.

Cada entidade geográfica está associada a uma classe, tendo suas características descritas através de atributos. O modelo enfatiza os aspectos estruturais em detrimento dos aspectos comportamentais (métodos). A Figura 3 mostra uma parte do esquema de dados resultante do projeto conceitual.

Na fase do *Projeto de Representação Espaço-Temporal*, as classes com e sem representação espacial foram diferenciadas através de pictogramas gráficos, conforme proposto nos modelos Modul-R [CAR 93] e GeoOOA [KOS 97]. O SIGPROGB prevê a elaboração de representações dos dados em diversas escalas (ex.: 1:2.000, 1:10:000, 1:50:000, 1:250.000 e 1:1.000.000). Uma vez que o conjunto de entidades geográficas (incluindo sua forma espacial) representadas em uma determinada escala varia de acordo com a escala utilizada, foi desenvolvido um sub-esquema para cada escala. A Figura 4 exemplifica parte do esquema de representação na escala 1:50.000.

Com base no modelo GMOD [OLI 97] e em [CAM 96], definiu-se uma hierarquia de classes abstratas (i.e. não possuem instâncias diretas) para a modelagem dos fenômenos geográficos observados na visão de campo. Um exemplo da aplicação dessa hierarquia pode ser visto em [LIS 96].

Analisando as Figuras 3 e 4, pode-se claramente identificar as classes que possuem representação espacial, a partir do pictograma gráfico que apresentam em seu canto superior. Nota-se também, que novas classes (ex.: as classes *TREstreito* e *TRLargo*) foram adicionadas ao esquema. Estas classes surgiram a partir do detalhamento do projeto e não haviam sido identificadas inicialmente.

O próximo passo do projeto será a conversão do esquema de representação para o modelo de dados do SIG que será utilizado (ainda não definido). Devido às limitações atuais, os aspectos temporais estão sendo tratados com base em *timestamps*, proposto em [PEU 95], ou seja, classes cuja evolução temporal necessita ser mantida no BDG estão sendo acrescidas de novos atributos temporais (ex.: data, período de tempo).

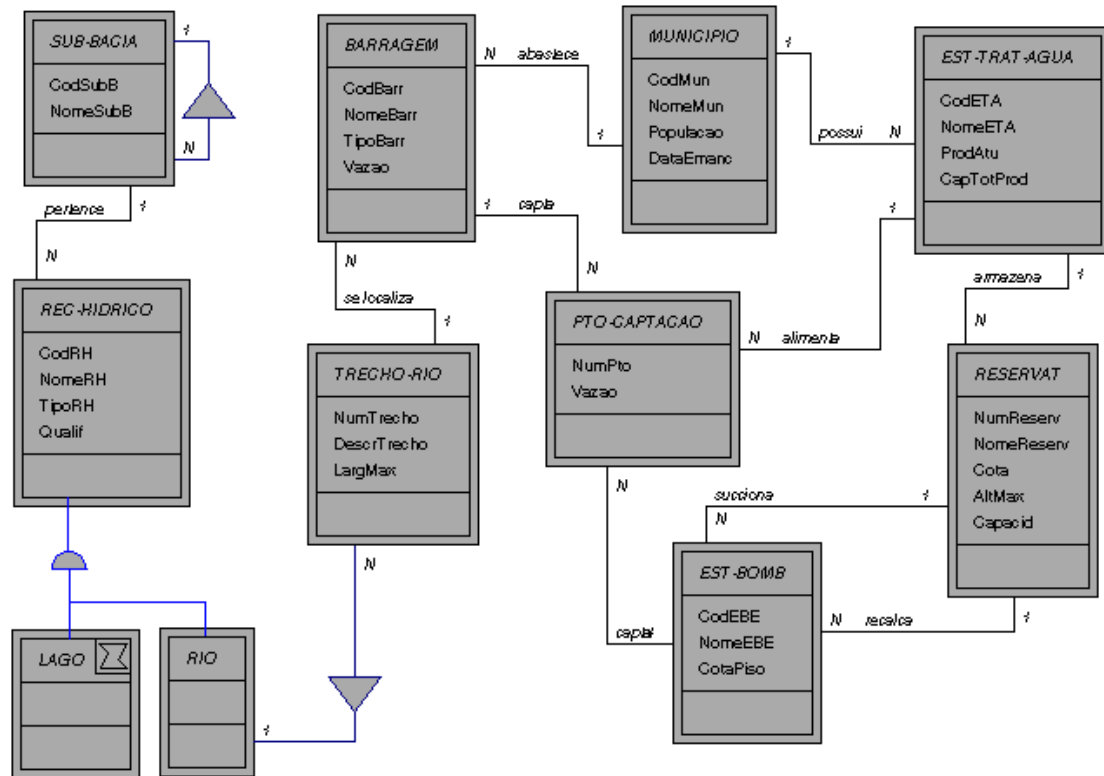


FIGURA 3 - Exemplo do esquema conceitual do SIGPROGB

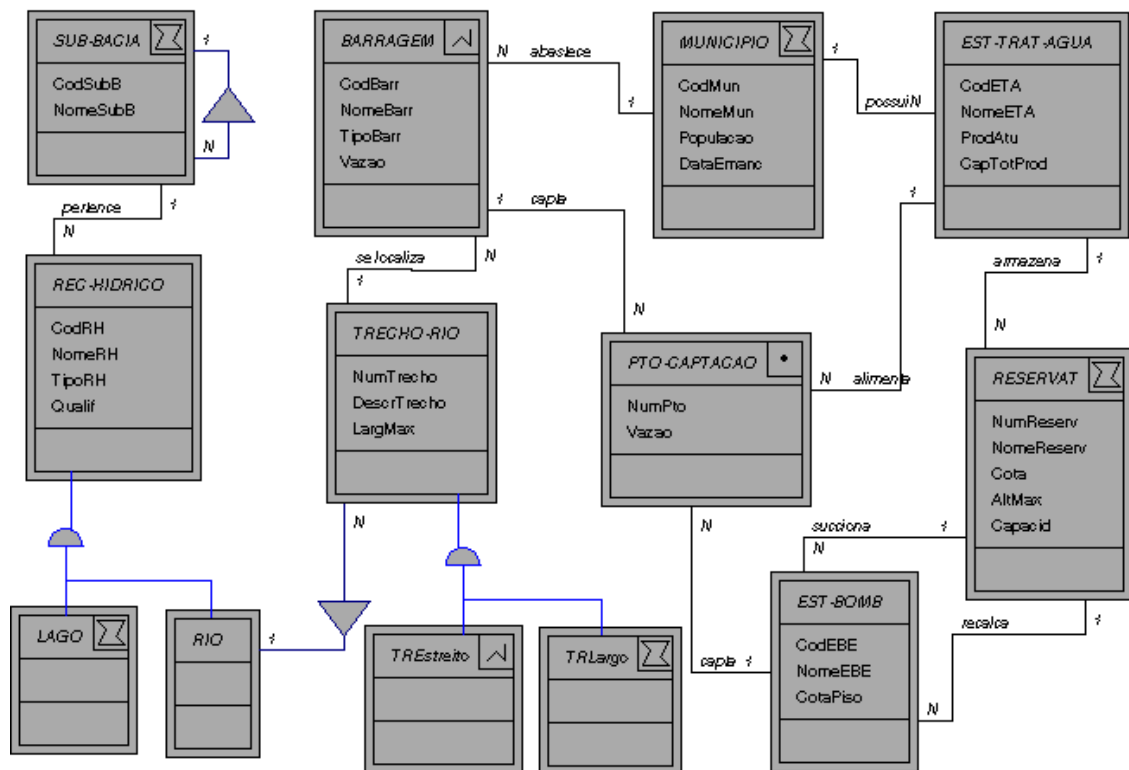


FIGURA 4 - Exemplo do esquema de representação espacial do SIGPROGB

6 CONCLUSÕES

O projeto do *SIGPROGB* permitiu a identificação de um conjunto de requisitos que impõe novos desafios aos modelos conceituais tradicionais. Estão em andamento diversas pesquisas que buscam definir um modelo adequado para suportar o projeto de aplicações de SIG.

Descreveu-se a fase do *Projeto de Representação Espaço-Temporal*, uma etapa adicional do projeto de banco de dados, onde, a partir de uma linguagem de alto nível (i.e. independente de software), pode-se modelar os aspectos de representação das entidades geográficas no espaço e no tempo. Esta nova etapa possibilita a simplificação da modelagem conceitual, uma vez que, projetistas e usuários podem concentrar suas atenções apenas na definição dos aspectos mais importantes da aplicação.

Demonstrou-se, através de exemplos, como o projeto do *SIGPROGB* está sendo desenvolvido apoiado em conceitos de diversos modelos da literatura e a partir de um conjunto de etapas adaptado ao projeto de BDG. Atualmente, os autores estão buscando uma formalização das primitivas de abstração necessárias para a fase do *Projeto de Representação Espaço-Temporal*. Está também sendo investigada a inclusão da definição de metadados de qualidade, um dos requisitos das aplicações de SIG, nesta mesma fase de projeto.

3.5 Artigo VI: BCH/GIS - A Basic Class Hierarchy to Support GIS Conceptual Design

Cirano Iochpe e Jugurta Lisboa Filho

In: INTERNATIONAL CONFERENCE ON MODELING GEOGRAPHICAL AND ENVIRONMENTAL SYSTEMS WITH GEOGRAPHICAL INFORMATION SYSTEMS (GIS). 1998, Hong Kong, China. **Proceedings...** Hong Kong: The Chinese University of Hong Kong, 1998.

ABSTRACT

Conceptual data models are used in the initial phases of a database project, i.e., during requirement analysis and the conceptual design. Although some general-purpose conceptual models have been successfully applied to model conventional information systems, they cannot cope well with some new, specific requirements posed on database modeling by geographical information systems. This paper presents a hierarchy of classes which has been used as a basis for modeling of large geographical information systems and discusses its evolution.

Keywords: Database Conceptual Modeling, GIS Application Modeling, Object Class Hierarchy.

1 INTRODUCTION

Conceptual data models are used in the initial phases of a database project, typically during requirement analysis and the conceptual design [NAV 92]. They are independent of hardware, operating system, or database management system (DBMS) and allow database designers to represent the application's data in an abstract, formal and unambiguous way. In the early phases of database design conceptual models help designers and users communicate about the reality being modeled. In the case of a geographical reality, both entities and the relationships among them can be abstracted by concepts such as classification, generalization-specialization, aggregation, and association.

Although some general purpose conceptual models (e.g., E-R, OMT) have been successfully applied to model conventional information systems, they cannot cope well with some new, specific requirements posed on database modeling by *geographical information systems* (GIS). Therefore, in the last few years many database researchers have proposed extensions to those models in order to support geographical database conceptual modeling [TRY 95, BED 96, KOS 97, OLI 97, SHE 97]. However, by modeling a large GIS application on the basis of some of those extensions we realized that some of the new requirements have not yet been satisfied [LIS 97].

This paper presents a hierarchy of object classes, which has been used as a basis for GIS conceptual database modeling and discusses its evolution. We applied this *Basic Class Hierarchy* (BCH/GIS) to integrate the conceptual modeling of both the descriptive and the geographical databases of the SIGPROGB, a large GIS project that

shall support the *Pro-Guaiba Environmental Program* [LIS 97]. The latter is a large program for environmental control in the context of the Guaiba River Basin in the state of Rio Grande do Sul, south of Brazil. The Inter-American Developing Bank supports this program, which involves more than twenty governmental and private institutions.

2 CONCEPTUAL DATA MODELS TO GIS APPLICATION

The GIS applications present some new modeling requirements that are not well supported by the general-purpose conceptual models. In [LIS 98a] five requirements which were identified from a bibliographical study and from the author's experience in the conceptual modeling of GIS environmental applications [LIS 96, LIS 97] are described. They are: *field* and *object* views; *spatial data quality*; *multiple representations*; *spatial relationship*; and *temporality*. The hierarchy described in this paper was designed to support the first three of the requirements above, which are shortly described below.

- *Fields and object view* – According to [GOO 92], the geographic reality can be observed in two perspectives: *field* and *object* views. In the field view, reality is modeled by variables with continuous distribution in space. Each space position is characterized by a set of attributes (e.g. temperature or relief). In the object view reality consists of discrete entities, each one with its own identity, being represented by spatial objects. These two visions use distinct modeling concepts.
- *Spatial Data quality* – Uncertain data always exist in GIS, but they need to be identified and treated. GIS applications use a large volume of data imported from other systems and this increase the data reliability problem. The GIS' users need to know the uncertainty degree of the data and how this uncertainty is propagated during spatial operations.
- *Multiple representations* – The same geographic entity can be represented, in a GIS many times and by several spatial objects (e.g. multiple scales). According to [CAM 96], in a GIS the multiple representation management should be considered both in the data structure level and in the conceptual level. The problems resulting from these characteristics are inconsistency, redundancy, and multiple behaviors for the same geographic phenomenon.

The last two requirements (*spatial relationship* and *temporality*) are not so complex in the applications we have worked, thus they did not receive much of our attention yet. We present below a brief comment about each one of the most recent data conceptual models found in the literature, based on which kind of requirements they support.

[OLI 97] describe the UAPE environment, where the GMOD data model support, mainly, the field and object view modeling. The GeoIFO model defines two new semantic constructors: *object position* and *space-varying attribute*, which supports the field and object views modeling as well [TRY 95]. The GISER model [SHK 97] is the only one to include some kind of mechanism to store metadata about quality, but in a superficial way. [BED 96] propose the MODUL-R formalism, an extension of the E-R model to GIS in the urban domain. This model supports multiple representations and temporal aspects. The GeoOOA model distinguishes between classes with and without spatial reference [KOS 97]. Moreover, it allows the user to represent the relationship between the geographic objects and spatial objects, support the definition of temporal classes, and define new special kinds of relationships to represent topological constraints.

When applied in a really large GIS application, as the SIGPROGB project, none of these models could appropriately support the conceptual modeling. The Modul-R is a good model to urban applications. The UAPE/GMOD was developed to environmental applications, but it is more appropriated to small problems, as shown by the examples in [PIR 97]. We are not proposing a new data conceptual model. Instead, we developed a basic class hierarchy, based in the object-oriented formalism where some good ideas from other models (e.g. graphic pictograms from MODUL-R and GeoOOA models) are used. Also, an extended design methodology is used, as described in the next section.

3 BCH/GIS - A BASIC CLASS HIERARCHY TO GIS

BCH/GIS relies upon the concepts of the OMT object-oriented conceptual data model [RUM 91] and serves as a starting point for both requirement analysis and conceptual database design during the early phases of GIS development. While leading database designers through the conceptual modeling process, BCH/GIS guarantees that the resulting schema meets the minimum set of modeling requirements, described below. When relying on the class hierarchy in order to both capture and represent the geographical reality of some application, designers are compelled to consider aspects such as: (1) data quality modeling; (2) the dichotomy between the object and the field visions of reality; and (3) multiple spatial representation of geographical phenomena that may be necessary either within the same map scale or for different scale representations.

During the modeling of the SIGPROGB database, we have tested and improved the class hierarchy. Relying on the experience of modeling more that two hundred object classes as well as building specific database views for various different map scales (e.g., 1:2000, 1:50000, 1:250000, and 1:1000000), we developed a series of evolving versions of the hierarchy. The following sections describe the BCH/GIS evolution, divided in three versions.

3.1 BCH/GIS Version-1

The first version of the BCH/GIS (Figure 1) was motivated by the MGeo model [TIM 94], which has a class hierarchy where object classes from different levels of abstraction are put together. Our contribution was to remove the classes concerned to the GIS implementation level (e.g. *OrientedLine*), to adapt the high level to the conceptual design phase, and to include the field representation classes, as defined in [GOO 92].

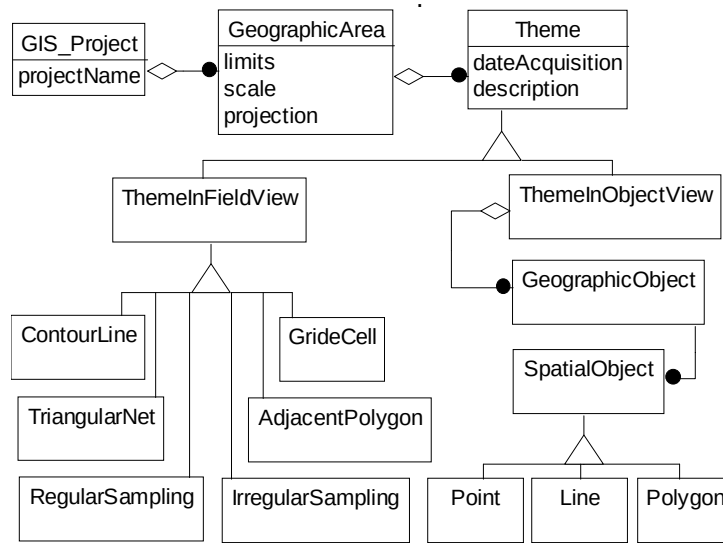


FIGURE 1 - BCH/GIS-v.1 – Version One

In the BCH/GIS-v.1, a GIS application is modeled as a subclass of *GIS_Project*, which aggregates several instances of the *GeographicArea* class, each one being represented by one or more *Theme* classes. The *Theme* Class is a generalization of the *ThemeInFieldView* and *ThemeInObjectView* subclasses, representing the dichotomy between fields and objets views. One *ThemeInObjectView* class is an aggregation of *GeographicObject* subclasses, which are related with one or more *SpatialObject*'s subclasses (i.e., *Point*, *Line* and *Polygon*) that represent the spatial location of the geographic entity.

The BCH/GIS-v.1 was used in the modeling of SIGERCO project, a GIS application in the environmental coastal domain [LIS 96]. The classes of the application's domain in the field view were defined as specializations of the *ThemeInFieldView*'s subclasses and each class representing a geographic entity of reality was related to one of *SpatialObject*'s subclasses. To avoid the unreadable data schemas generation, graphic pictograms were used, as proposed by [BED 96] and [KOS 97]. Graphic pictograms are small pictures put inside the class definition, whose semantic represents a relationship between the *GeographicObject* class and a *SpatialObject*'s subclass.

3.2 BCH/GIS Version-2

The second version of the BCH/GIS (Figure 2) borrowed ideas from the class library of SPRING GIS [CAM 96]. In the SPRING system, a geographic database (*GeoDB* class) is composed of maps (*GeoFieldMap*, *GeoObjectMap*) and of geographic objects (*GeoObject* class). Two significant changes were made: the substitution of the fields' representation classes by the classes *ThematicMap*, *DTM* and *Image*; and the kind of relationship between the *GeoObject* class and *SpatialObject* class.

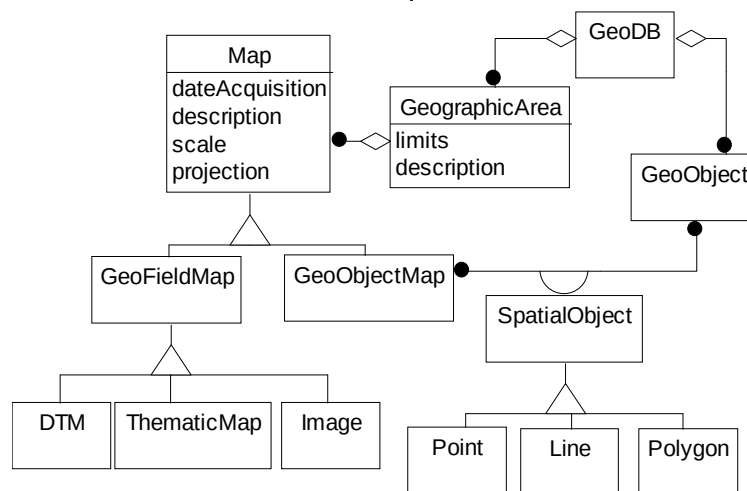


FIGURE 2 - BCH/GIS-v.2 – Version Two

The BCH/GIS-v.2 helped us to better understand several aspects in GIS modeling, which are listed below.

- The majority of the conceptual data models proposed until now, give solutions for the designer to model maps and not the reality. As a map is a model of reality, when we model the components that appear in a map, we are not modeling reality, but a model of it (the map). This can be seen in the examples presented in [BED 96, KOS 97]. The result is that the conceptual data schema has entities both from the application domain and the computer system domain.
- The second problem identified from the BCH/GIS-v.2 concerns the application's classes modeling in the field view classes. As the object view, a field view class should not be modeled as specialization of the GeoFieldMap's subclasses, because it is the same mistake of modeling a Road class as subclass of the Line class, as proposed in [CLE 94]. The generalization/specialization abstraction is for relationships of the kind IS-A. A road is not a line as well as the relief is not a DTM (Digital Terrain Model).
- The third problem identified in the both BCH/GIS versions is the *GeographicArea* class should not be modeled as a class, but as an attribute of the representation models. Moreover, the *GeoDB* class (or *GIS_Project* class in the BCH/GIS-v.1) should not be in the data conceptual schema.

3.3 BCH/GIS Current Version

The BCH/GIS-v.3 includes some new characteristics (Figure 3). The first one is the partition line among the components should be present in the conceptual level and the components should be added during the *Spatial Representation Design Phase* [LIS 97]. This new database design phase becomes necessary during the SIGPROGB project development. In large-scale projects (like the SIGPROGB, which has more than 250 classes), aspects like spatial representation, temporal representation, and metadata cannot be treated during the requirement analysis phase. On the other hand, these aspects must be considered without taking the software restrictions (i.e., logical design) into account.

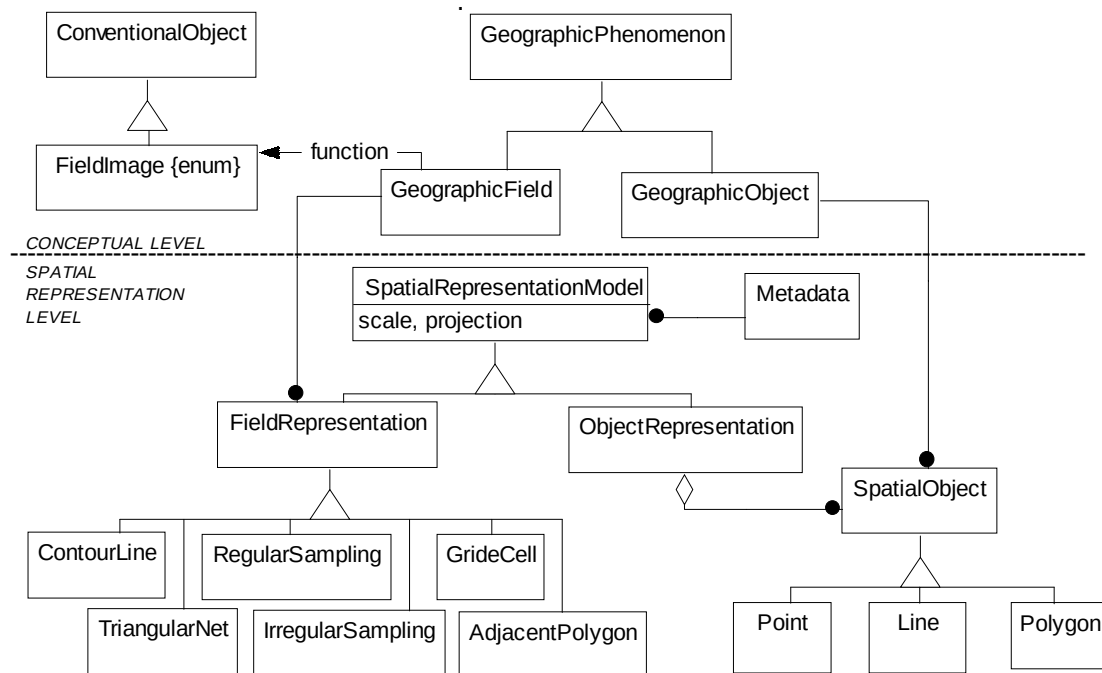


FIGURE 3 - BCH/GIS-v.3 – Current Version

In the conceptual design phase, the resulting schema (conceptual schema) shows only two kinds of classes: *ConventionalObject* class - the root of all classes without geographical reference; and *GeographicPhenomenon* - the root of all classes with geographical reference. It is useful to distinguish these two kinds of classes. We included one flag in the name of the class (e.g. adding the string “*GF*” in the class name. This means that the class is a *GeographicPhenomenon* subclass).

There are two kinds of *GeographicPhenomenon* subclasses: *GeographicField* and *GeographicObject*. Each *GeographicField* class has one special kind of association with one *FieldImage* class, which describes the possible values to the function represented by the field variable. For example, *LandUseType* is an enumeration with the following possible values {pasture, planting, naturalForest, and reforestation...}.

In the spatial representation design phase, the conceptual schema is transformed into a representation schema by the addition of new details, where the classes from the spatial representation level (Figure 3) are used. The *SpatialRepresentationModel* class is an abstract class where are defined some commons attributes, like *scale* and *projection*. The *SpatialRepresentationModel* class has an association with the *Metadata* class. The possible kinds of *Metadata* subclasses are still under development and are not included in the BCH/GIS-v.3 yet. We have concluded that the distinction between the *ThematicMap*, *DTM*, and *Image* classes (in the BCH/GIS-v.2) is not necessary and we cut these classes off the BCH/GIS-v.3.

The third innovation in the BCH/GIS-v.3 concerns the kind of representation of themes. One theme is an aggregation of several objects either from the same class or from different classes. For example, one could define a *Lakes* theme composed of instances of *Lake* class. On the other hand, one *Hydrological* theme could be defined including instances of *Lake*, *River* and *Basin*. There are two ways to modeling themes. The first one is to use the aggregation relationship as proposed in BCH/SIG-v.1. The second way is to use the concept of *Subject*, as proposed by Coad & Yourdon in the OOA method [COA 91]. To define a theme using the *Subject* concept, we draw a polygon (dotted line) around the classes in that theme. One *GeographicPhenomenon*

can take part in more than one theme. We have used different colors to identify the themes in the SIGPROGB project.

4 EXAMPLE OF THE BCH/GIS USE

In this section the BCH/GIS-v.3 use is shown and explained through a small piece of the SIGPROGB schema (Figure 4). Due to space restrictions, only the spatial representation schema is showed.

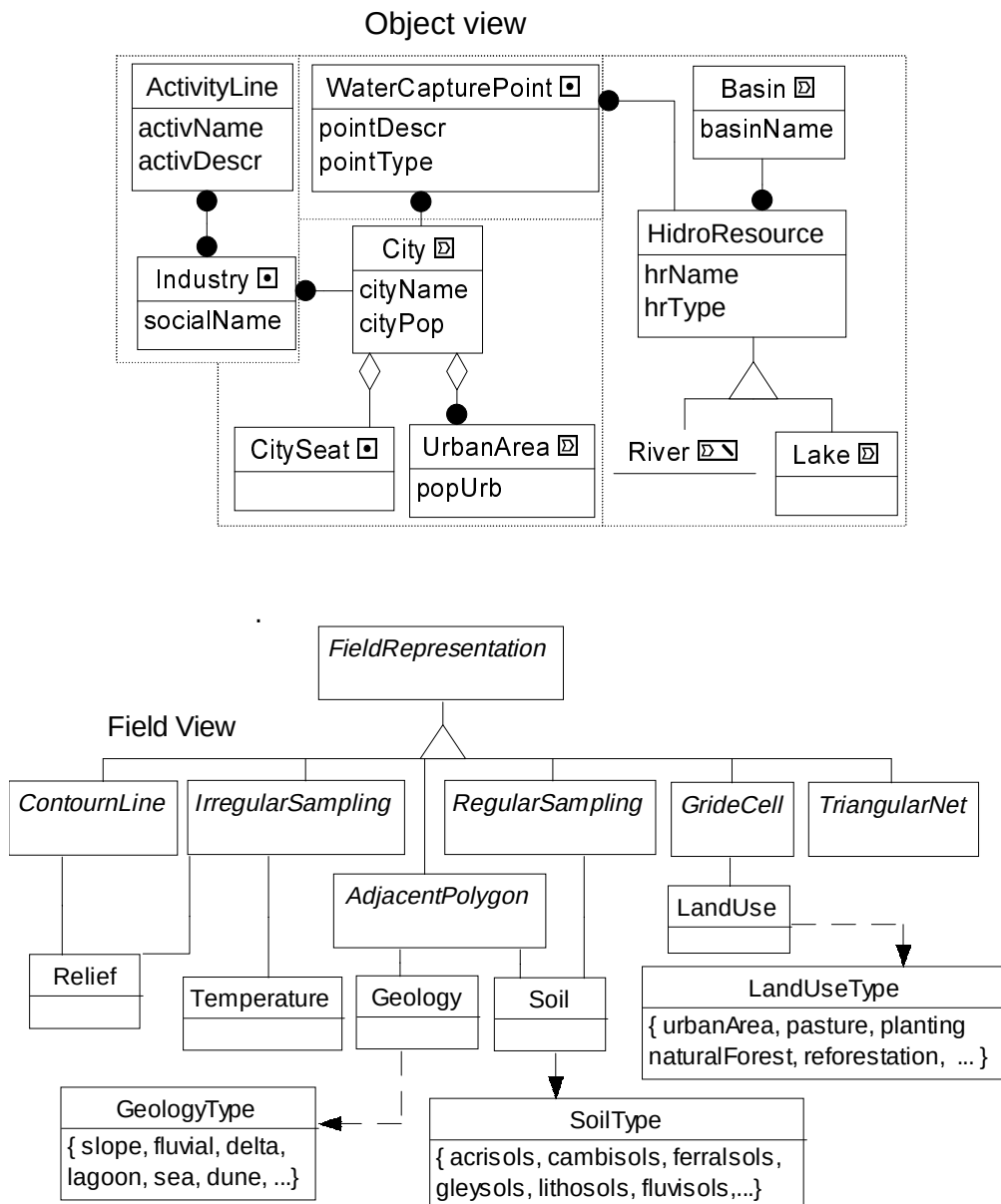


FIGURE 4 - Example of the SIGPROGB spatial representation schema

The following observations are presented:

- The dotted polygons show the theme specifications. The GIS CASE tool should permit the user to choose if s/he want to see the themes or not.
- Although not represented, each class in the object view is a subclass of *GeographicPhenomenon* class (all the classes with graphic pictograms) or a subclass of *ConventionalObject* class (without pictograms).
- The *FieldImage* subclasses (e.g. *SoilType*) must not be completely specified. A data dictionary can document all the possible values. Usually it is implemented as a table.
- A *GeographicObject* subclass can be related with more than one *SpatialObject* (e.g. the *River* class). This is represented by the use of multiple graphic pictograms.

5 CONCLUSIONS

This paper presented the evolution of the BCH/GIS development, a basic class hierarchy to be used as a starting point in GIS requirement analysis and data conceptual design. Based in a complementary database design phase, and by using graphic pictograms and the *Subject* concept to theme specification, we have shown that readable data schemas for large applications can be produced. The necessity of using a good CASE tool to support this work is essential. We used the Rational ROSE™ software and the graphic pictograms as a kind of Windows™ font, downloaded from Bedard's site. We are currently working with spatial metadata modeling and to extend the UML (*Unified Modeling Language*) to be used as an analysis patterns language to GIS applications.

4 Instrumentos de Reutilização

4.1 Introdução

Conforme descrito no Capítulo 1, o objetivo geral desta tese consiste na definição de instrumentos que permitem a reutilização de projetos de banco de dados geográficos, a fim de aumentar a qualidade dos projetos de banco de dados das aplicações de SIG, que são desenvolvidas, muitas vezes, pelos próprios usuários. A seguir, descreve-se a contribuição de cada um dos artigos que compõem este capítulo.

O Artigo VII (Seção 4.2), intitulado “*Applying analysis patterns in the GIS domain*”, apresenta a adequação do uso de padrões de análise na área de SIG. São descritas as diversas categorias de padrões encontradas na literatura, é feita uma descrição dos padrões de análise existentes, principalmente daqueles utilizados em aplicações comerciais. O artigo ainda introduz um primeiro padrão de análise na área de SIG.

A descrição de padrões de análise para banco de dados geográficos apresenta o mesmo problema dos projetos de banco de dados já descritos no Capítulo 2, ou seja, a necessidade de modelos conceituais específicos para descrever a solução de modelagem proposta pelo padrão, bem como para exemplificar o uso desses padrões. Uma vez que o objetivo de um padrão de análise é o de possibilitar a reutilização, pelo maior número de usuários possível, de uma solução para um problema conhecido, sua descrição deve ser de fácil entendimento por parte dos projetistas e usuários de SIG (que muitas vezes fazem o papel do projetista de banco de dados). Para isso, propõe-se o uso da notação do diagrama de classes da Linguagem de Modelagem Unificada – UML [BOO 98], com base no *framework* conceitual GeoFrame.

O GeoFrame originou-se a partir da transformação do diagrama de classes BCH/GIS (Seção 3.5), estando documentado no Artigo VIII, intitulado “*Specifying analysis patterns for geographic databases on the basis of a conceptual framework*” (Seção 4.3).

O objetivo de um *framework* conceitual é o de fornecer um diagrama de classes que pode ser usado como base para a modelagem das classes do domínio da aplicação. O termo *framework* conceitual significa que o produto gerado a partir do uso do *framework* não é um software executável, mas sim, um esquema conceitual de dados que, posteriormente, deverá ser traduzido para um esquema de dados específico do SIG no qual será desenvolvida a aplicação.

A adequação do uso do diagrama de classes BCH/GIS foi analisada através de sua utilização durante o desenvolvimento de uma sistema real (SIGPROGB [BIN 98]), que serviu como estudo de caso na dissertação de mestrado da aluna Íria Margarida Garaffa [GAR 98], desenvolvida no âmbito do Projeto SIGMODA [SIG 99], junto ao Curso de Pós-Graduação em Computação da UFRGS. As principais deficiências identificadas a partir desta avaliação prática e que serviram para orientar a especificação do GeoFrame estão relacionadas a seguir.

- Necessidade de mecanismos que possibilitem diferenciar, no esquema de dados, as classes que representam entidades geográficas, campos geográficos e classes convencionais, ou seja, não georreferenciadas.

- Tratamento uniforme para os tipos de representação espacial de campos e objetos como, por exemplo, através de pictogramas gráficos para ambos.
- Necessidade de diferenciar os relacionamentos espaciais dos convencionais. Por exemplo, duas classes podem estar associadas através de um relacionamento convencional, um relacionamento espacial ou ambos.
- Necessidade de aprimorar a forma de modelar fenômenos geográficos com múltiplas representações espaciais.

O Artigo VIII descreve como estas deficiências do diagrama BCH/GIS foram resolvidas no GeoFrame.

O Artigo IX (Seção 4.4), intitulado “*Padrões de análise para banco de dados geográficos*”, introduz alguns padrões de análise que foram identificados a partir da observação de projetos de aplicações de SIG nas áreas de controle ambiental e de gestão urbana. O artigo também apresenta uma introdução sobre padrões de análise, incluindo a descrição do processo de captura e documentação de um novo padrão, e propõe a adoção de uma linguagem para descrição de padrões de análise para bancos de dados geográficos.

Durante a especificação dos padrões de análise, descritos no Artigo IX, quando da utilização do GeoFrame foi identificado um problema relacionado à modelagem de campos geográficos categóricos. O detalhamento desse problema e a solução adotada está documentado no Artigo IX.

A Seção 4.5 inclui o Artigo X, intitulado “*Modelagem conceitual de banco de dados geográficos: o estudo de caso do projeto PADCT/CIAMB*”, que descreve uma experiência do uso de padrões de análise durante o desenvolvimento da modelagem conceitual dos dados do SIG do projeto PADCT/CIAMB. Durante o trabalho, pôde-se verificar a facilidade com que os projetistas desenvolveram a modelagem conceitual dos dados de um sistema no domínio de aplicações da área de controle ambiental, uma vez que os mesmos já haviam participado da modelagem de outros sistemas semelhantes [LIS 96, LIS 97]. Além disso, houve uma grande receptividade por parte dos usuários, com relação à notação gráfica utilizada na modelagem.

Por último, a fim de exemplificar como um esquema conceitual de dados, especificado em UML [BOO 98] e com base no GeoFrame, pode ser, facilmente, transformado para um software de SIG, realizou-se um exercício de mapeamento de um esquema conceitual de dados, gerado a partir do GeoFrame, para o software *Spring* [INP 99]. Este exercício de mapeamento está descrito no Artigo XI, intitulado “*Projeto de banco de dados geográficos: mapeando esquemas GeoFrame para o SIG Spring*” (Seção 4.6), no qual é apresentado um conjunto de transformações necessárias para o mapeamento entre um *Esquema GeoFrame* e um esquema de dados no *SIG Spring*. O artigo apresenta, como exemplo, o mapeamento de parte do esquema de dados do projeto PADCT/CIAMB (descrito na Seção 4.5) no *SIG Spring*. Este último artigo, embora não aborde, especificamente, o tema da reutilização, foi incluído neste capítulo por ser complementar à descrição do *framework* GeoFrame (Seção 4.3), exemplificando sua usabilidade como um instrumento para projeto de banco de dados geográficos.

4.2 Artigo VII: Applying Analysis Patterns in the GIS Domain

Jugurta Lisboa Filho, Cirano Iochpe e Kate Beard*

*NCGIA - Department of Spatial Information Science and Engineering, University of Maine, USA

In: ANNUAL COLLOQUIUM OF THE SPATIAL INFORMATION RESEARCH CENTRE - SIRC, 10., 1998, Dunedin, NZ. **Proceedings...** Dunedin: SIRC, University of Otago, 1998.

ABSTRACT

Patterns have been used in software engineering to enable the reuse of successful solutions for recurrent problems in various steps of the software development process. In the development of geographic information systems (GIS), usually only geographic data are reused. This paper shows the adequacy of deriving as well as applying analysis patterns in the conceptual data modeling process of geographic applications. By means of analysis patterns, already tested and validated solutions for recurrent problems can be reapplied to different GIS projects. The requirement of both a suitable pattern definition formalism and a structure for a GIS pattern catalogue are also discussed. Finally, an example of an analysis pattern for the GIS domain is presented.

Keywords and phrases: analysis patterns, conceptual data modeling, GIS applications.

1 INTRODUCTION

Geographic Information Systems are used to support database applications that both store and analyze georeferenced data (e.g., environmental control, utilities network). Conceptual data modeling in the domain of GIS applications typically identifies a large number of geographic entities as well as relationships between them. According to Gordillo [GOR 97], expert GIS designers do not solve every problem from scratch. On the contrary, they usually reuse previously worked out solutions to solve similar problems. Since the acquired experience remains with the designer him/herself and is usually not made public, other designers are not able to share or reuse that knowledge.

On the other hand, the exchange of geospatial data through the Internet among different users or organizations increases constantly. Standardized metadata describing sets of geospatial data are stored in clearinghouses [NEB 96]. In such a way, a GIS designer can find, assess, and obtain available data sets, avoiding having to produce them once again. However, the design of the GIS database must still be done from scratch, since nowadays no metadata for model description is available. Currently, the most well known system of clearinghouses is supported by the Federal Geographic Data Committee [FGD 97].

Patterns express a new concept which software engineers have been using to both document and share their experiences in object-oriented system analysis and design. There are different kinds of patterns that are used at different levels of abstraction as well as in different phases of the software development process. Though the use of patterns to enhance GIS database design has not yet been investigated in depth. Balaguer [BAL 97] describe one of the rare examples, where the use of design patterns is considered in the development of conventional applications with geospatial features. In the present paper, we discuss the suitability of both deriving and applying analysis patterns to respectively describe and reuse experiences of GIS conceptual data modeling.

The idea of reuse has already been explored in the GIS applications domain, but not in the data modeling process. Marr [MAR 97] proposed a structure of spatial process modeling that can be reused and was made available through the Web. We think that the reuse of spatial process modeling is complementary to our proposal, supposing that a spatial process could be defined as a functional pattern and could be integrated with the analysis patterns and the spatial metadata in a clearinghouse system.

We believe that the integration of both a GIS patterns catalogue (including all already existing pattern categories) and a set of spatial metadata into a clearinghouse will become important issues of research in the next years along with the increase of geographic data sharing through the Internet. Results of this investigation shall contribute to the development of higher quality GIS applications.

The remainder of this article is organized as follows. Section 2 briefly introduces today's existing pattern categories. Section 3 presents, in more details, the concept of analysis patterns. In Section 4, we emphasize the advantages of using analysis patterns in the process of modeling GIS applications. Section 5 describes expected features of a patterns catalogue for the GIS domain. Conclusions and suggestions of future work are presented in Section 6.

2 PATTERN CATEGORIES

The pattern approach is still very recent in computer science. Thus, there does not exist a consensus on a definition for the word "pattern" in this context. The use of patterns is directly related with the idea of reuse in software engineering and has origins in the work of Christopher Alexander, an architect that, in [ALE 77], defined:

"Each pattern describes a problem which occurs over and over again in our environment, and then describes the core of the solution to that problem, in such a way that you can use this solution a million times over, without ever doing it the same way twice. (Page x)"

The ideas of Alexander had given origin to the development of the research on design patterns and, later, several other pattern types. Fowler [FOW 97] defines a pattern as:

"An idea that has been useful in one practical context and will probably be useful in others. (Page xv)"

According to Gamma [GAM 94]:

"A pattern presents the essence of the solution for a recurring problem in a specific context. (Page 2-3)"

The pattern approach became more popular after the book *Design patterns: elements of reusable object-oriented software* [GAM 94] that presents a catalogue of solutions for recurrent problems in object-oriented design. Gamma's definition comprehends the fundamental ideas of a pattern. The expression "*the solution for a problem*" means that each pattern identifies one problem and presents one solution for

it. The term “*essence of the solution*” means that only the essential elements are described, leaving specific aspects to be detailed by the designer, since specific aspects of solutions are usually not reusable. “*recurring problem*” means that patterns should be created for problems that already occurred some times and will occur again. “*in a specific context*” means that the solution is valid for a particular context.

According to Buschmann [BUS 96], patterns can be divided in three categories: *architectural patterns*; *design patterns*; and *idioms*, the latter are also known as *implementation patterns* [PRE 95]. The first two categories include patterns related to the system's design process. Though, they differ from one another in the level of detail where each one is applied. Design patterns are more abstract (and of smaller granularity) than architectural patterns, while idioms are used in the level of programming languages.

A fourth category of patterns, introduced by Fowler [FOW 96], is the one of *analysis* or *conceptual patterns*, according Riehle and Züllichoven [RIE 96]. This type of patterns is used to describe solutions related to problems that arise during both the requirement analysis and the conceptual data modeling phases. Analysis patterns reflect rather conceptual structures of the application domain and are not concerned with computational solutions.

Most of the analysis patterns that were proposed until now have been designed mainly to solve problems that are recurrent in the conventional applications domain [HAY 95, COA 97, FOW 97]. However, the idea of analysis patterns can be applied to improve quality and productivity in the development of nonconventional applications (e.g. GIS applications) as well.

All pattern categories cited above can be used in a combined way. For example, analysis patterns can be used to register how designers from an application domain usually model the most important entities and its relationships in that domain. Design patterns can be used, for example, to document how data about the relief of a region are represented through a Digital Terrain Model (DTM). Finally, idioms can define how a DTM should be implemented in each commercial GIS product.

3 ANALYSIS PATTERNS

Since the pattern approach was introduced in the discipline of software engineering by the community of object-oriented system designers, the category of design patterns became more popular and well known. However, several researchers have been investigating patterns to help system analysts to reuse prior, already tested modeling strategies for recurring problems [HAY 95, GER 96, RIE 96, COA 97, FOW 97].

According to Coad [COA 92], traditional reuse techniques in the context of object-oriented analysis rely upon the inheritance mechanism. Thus, to reuse predefined classes, analysts need to know where to find the existing classes and then derive new subclasses from them. These existing classes are used as simple building blocks in a new application design.

An analysis pattern, from now on just pattern, describes a set of classes, possibly of different class hierarchies, and the associations among them⁸. Patterns can be seen, therefore, as a way of describing more complex design schemes, which are recurrent in the modeling process of many (similar) applications. Thus, they can speed up

⁸ The following nomenclature is used in this article. A *geographic class* is an abstraction of a *geographic entity*, which occurs in the reality (e.g. river, road), and an *association* is an abstraction of a *relationship* that occurs between entities of the reality (e.g. a river cross a road).

considerably the design of new applications and increase the quality of that design [HAY 96].

A pattern presents a solution in a more suggestive than prescriptive way, providing a model and a discussion of why the solution proposed has its actual form, and its strengths and weaknesses. According to Fowler [FOW 96], the really important thing about a pattern is not the model but the rationale behind it.

To illustrate how an analysis pattern can look, an example borrowed from [HAY 95] is showed. Hay presents a set of patterns that he called *data model patterns*. These patterns are recurrent in many conventional applications. Among them there is a pattern called *Geographic Locations*, which is shown below and represents a solution to the address-modeling problem. This pattern can be applied to modeling GIS applications as well.

Notice that the example was rewritten according to the object-oriented approach. The pattern description follows the form proposed by Coplien [COP 95], and the diagrams are in UML notation [RAT 97].

Data model graphic notations are not powerful enough to support patterns descriptions. The graphic notation only captures the final design result, representing it as a network structure of classes and associations between them. A pattern description, though, also includes the documentation of decisions taken, choices made, identified alternatives, and weaknesses of the model. In order to help designers to take a better pattern understand, examples should be included on the pattern description [GER 96].

Pattern name: Geographic Locations

Problem: The Geographic Locations pattern in such a way describes an alternative solution to the problem of address-modeling, being useful both for persons and organizations.

Context: It is common in many real life situations to find one street address related to more than one person (or organization) and vice-versa. The simplest way to model addresses is to include a set of address-related fields (e.g. street, number, ZIP code) as attributes of a class. In this case, a series of known problems can occur as, for example, incomplete information or database inconsistency.

Forces:

- Often it is of interest to collect addresses by neighborhood, city, state, or by other geographic location where the address is located. Thus, each site must be in one geographic location, which means that each geographic location may be the location of one or more sites.
- For mailing labels, an address is a single piece of information. If the purpose of the address is only for mailing labels, the class attributes choice could be considered.

Solution: Figure 1 shows the class diagram of the pattern. The class PARTY is a generalization of the classes PERSON and ORGANIZATION. The class PARTY has an association called PLACEMENT (of cardinality m:n) with the class SITE. This association can include attributes as period and type of the placement. The class SITE should contain the attributes that characterize the address of one PARTY instance. The GEOGRAPHICLOCATION class is defined to create a hierarchic structure of locations (e.g.: Country, State, City, Quarter). In this way, each instance of SITE must be associated with only one instance of GEOGRAPHICLOCATION.

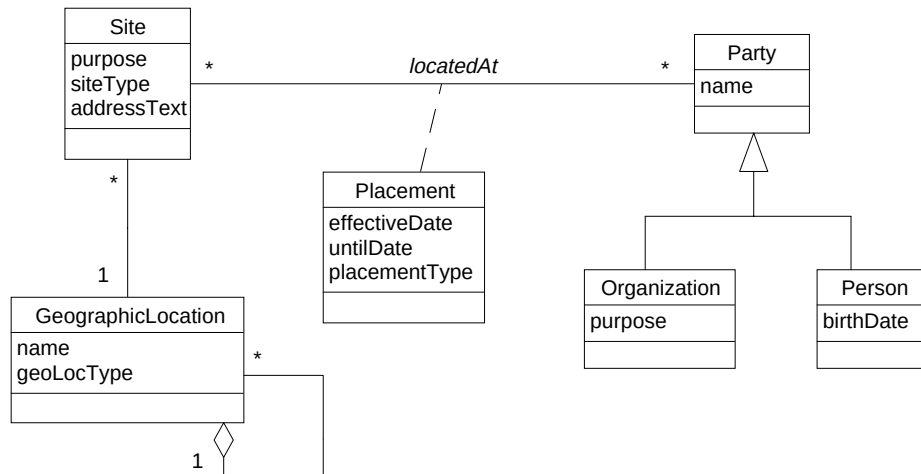


FIGURE 1 - Geographic Locations Analysis Pattern [HAY 95]

According to Hay [HAY 95], an analysis pattern should be customized for specific situations. The above described pattern provides an example of a case in which there is no final model. As can be seen in this example, a pattern definition may rely on other pattern definitions. The generalization of the PARTY class has actually been defined as a pattern itself in both [HAY 95] and [FOW 97]. On the other hand, the GEOGRAPHICLOCATION class can be reused in the design of several other GIS applications.

4 USAGE REQUIREMENTS OF ANALYSIS PATTERNS IN GIS DATA CONCEPTUAL MODELING

The domain of geographic information systems can benefit from the pattern approach, since several sets of geographic entities are recurring in many GIS applications. As an example, we can mention geographic entities that are usually represented in maps which form the basic layers of many GIS systems (e.g. hydrographic as well as topographical maps) concerning a common geographic area.

Usually, by starting the conceptual design of a new GIS application, the experienced analyst already knows a minimum set of classes, pertaining to the application domain, and the main associations between such classes [GOR 97]. Therefore, the data modeling process evolves from this initial set of classes. In fact, this development approach is being used for a long time already, but it has never been supported by any mechanism which could assist the analyst in this initial stage of the project.

The pattern approach provides an important contribution to this problem. Relying on it, provided that patterns can be defined and applied according to some formalism, it is possible to develop software mechanisms that are capable to both integrate pattern definitions into either existing or new geographic information models and also translate these definitions in definition languages accepted by commercial GIS products.

The use of analysis patterns implies some changes in the way conceptual data modeling is done today. Besides the usual requirements analysis (or engineering), the designer should search one or more pattern libraries trying to find already existing solutions for the modeling requirements he/she identified through the requirements analysis.

In fact, the study of existing patterns for some applications domain can give the designer helpful insight about the geographical reality being modeled. In this way, the requirements analysis process can be carried out with much more confidence as well as understanding.

According to Gerth [GER 96], besides allowing the reuse of knowledge and experience, patterns can improve the quality of software documentation as well as enhance the communication between designers.

As for the pattern approach to both be effective and evolve, every designer who makes use of pattern catalogues should try to optimize existing patterns relying on his/her own modeling experiences as well as to derive new patterns and insert them into a catalogue as soon as they have proven to be effective in the designer's own modeling process.

It should be clear by now that for the pattern approach to be helpful for designers pattern catalogues should be well organized and easy to both find and access. These aspects will be discussed in more detail in Section 5.

The use of patterns in GIS modeling depends on four important requirements:

- GIS designers should be interested in sharing their experiences;
- A standardized pattern description form should be adopted by the GIS designers community;
- Either a new or an extended data model should be adopted that is both suitable for GIS data modeling [IOC 98] and allows for the definition and integration of patterns as they can be described using the standardized pattern description form;
- Software mechanisms should be developed that support designers in selecting as well as locating, accessing, and reusing specific patterns, possibly on the basis of some query language.

The culture of data sharing through the Internet is widespread among GIS users all over the world. This is leading to a huge proliferation of sites that offer some kind of (interesting) data. The concepts of both metadata [WEB 98] that can describe available geographic data sets as well as their actual location in the network, and that of a clearinghouse system that keeps metadata bases and offers a query language interface on them are greatly contributing to organize this ever growing, worldwide federated geographic database system.

Similarly, we believe that GIS application designers should be supported by some patterns catalogue system distributed over a clearinghouse system where patterns are described and made available together with the metadata related to them. We also believe that GIS application designers will have interest in sharing their successful experiences in GIS data modeling as it already occurs in the community of object-oriented designers.

4.1 An Example of GIS Analysis Pattern

To illustrate the potential use of analysis pattern in the GIS domain, this subsection describes a pattern called *Environmental Quality Parameters*. We first identified this pattern during the data modeling process of a coastal management information system (SIGERCO-RS) that has been developed in partnership with FEPAM/RS - Foundation for Environmental Protection of the State of Rio Grande do Sul [LIS 96].

Some time later, the same pattern was reused during the data modeling process of SIGPROGB, the GIS of Pró-Guaíba, a program financed by the World Bank and the Government of the State of Rio Grande do Sul [LIS 97]. Both applications belong to the

environmental domain and, thus, present various similarities concerning their respective data models.

Pattern name: Environmental Quality Parameters

Problem: how to model data representing environmental quality parameters?

Context: the quality of some environmental elements (e.g. water, air) is assessed through periodical measurements of related quality parameters in different sampling points.

Forces:

- The number of parameters can vary for each sampling point during the monitoring time interval.
- The measurement periodicity can vary for each sampling point.
- Each parameter is always measured in the same unit.

Solution: Figure 2 shows a class diagram for the pattern. This pattern is based on the *Measurement Pattern* described in [FOW 96]. In each sampling point is taken, periodically, several measurements, corresponding to the analyzed parameters of quality control. We drew an association between the class UNIT and the class CONTROLPARAMETER because each parameter always is measured in the same unit of measurement. If this were not necessary, the association would be between the classes UNIT and MEASUREMENT, which would cause a problem of replication.

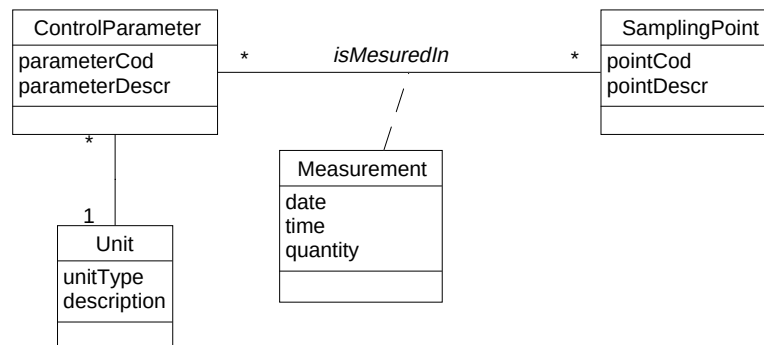


FIGURE 2 - Environmental Quality Parameters Pattern

Example: Figure 3 shows a segment of the conceptual data schema of the SIGPROGB Project, where the pattern was customized for the control of water quality parameters. In this case, each sampling point is located within a hydrological resource. Thus, the SAMPLINGPOINT class is associated with the HYDRORESOURCE class that, in turn, has other associations in the context of the project. The diagram also presents especial features of GIS conceptual data modeling as, for instance, the use of graphic pictograms, proposed by [CAR 93] and used in [IOC 98], for specifying the type of spatial object that represents each geographic entity.

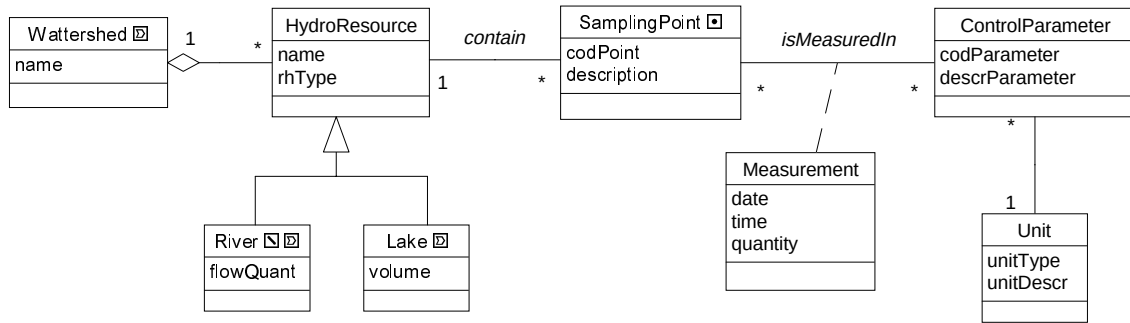


FIGURE 3 - Example of modeling water quality parameter

5 GIS PATTERNS CATALOGUE

The reuse of good data modeling solutions greatly depends on how hard are analysis patterns to be found. Therefore, patterns must be well organized in easy-to-find patterns catalogues.

On the other hand, since the number of registered patterns would tend to grow quickly, once designers really buy the patterns approach, the GIS designers community should count on either “de facto” or official standards of both an unified GIS conceptual data model which prescribes a pattern description form and a set of classification rules for GIS analysis patterns.

Also important in this context is the concept of a *Pattern Language*. It was first proposed by Alexander [ALE 77] and has been used by many design patterns authors. According to Coplien [COP 94], “a pattern language is a structured collection of patterns that build on each other to transform needs and constraints into an architecture”.

No analysis patterns language has been proposed until now. It would surely help GIS modeling designers to classify as well as describe sets of patterns on the basis of either common or related problems they can solve. This kind of classification can be very useful in the search of patterns by application domain. For instance, a pattern called *StreetNetwork* which describes object classes pertaining to a road network and a pattern called *Routing* which models street directions for each street segment in a network can be catalogued together into a pattern language intended to urban applications. This aggregation of patterns could be applied to help modeling a Bus Routes System.

A prototype that extends a clearinghouse system with a catalogue of GIS analysis patterns is under development at the Instituto de Informática of the Universidade Federal do Rio Grande do Sul. By accessing the clearinghouse’s query interface, it shall be possible for users to search for GIS analysis patterns much in the same way they would be searching for spatial data sets. We intend to create indices for patterns using the same mechanism which creates indices for spatial metadata. In addition, a patterns catalogue query subsystem shall be implemented.

An advantage that can be observed when metadata and analysis patterns are integrated in a clearinghouse system is that the metadata of the data produced for an application can make reference to the respective pattern which describes them. As a consequence, the pattern’s object schema can be used to update the database schema of the GIS in which the data will be incorporated.

6 CONCLUSIONS AND FUTURE WORK

This paper explores the potential of analysis patterns in what concerns conceptual data modeling reuse in GIS applications. As the pattern approach is a new research theme even for the software engineering area, there is no consensus about a description form for analysis patterns. According to [RIE 96], the best way to describe a pattern is by describing its foreseen usage. Currently, we are investigating proper approaches for patterns description as well as best structures for a GIS patterns catalogue.

The results of this research shall be extended, in a further step, to consider pattern types of different abstraction levels as well. With a mechanism based on reuse to support the development of GIS applications, the whole software development cycle can become more efficient. Analysis patterns can allow analysts and designers to reuse the knowledge of experienced professionals mainly during the initial stages of GIS design (e.g.: requirements analysis and conceptual modeling) which have lead to low quality projects due to the lack of experience of nonspecialist designers.

ACKNOWLEDGEMENTS

This research was developed in the scope of SIGMODA Project, financed by grants of RHAE-CNPq and CAPES-PICDT.

4.3 Artigo VIII: Specifying Analysis Patterns for Geographic Databases on the basis of a Conceptual Framework

Jugurta Lisboa Filho e Cirano Iochpe

In: ACM SYMPOSIUM ON ADVANCES IN GEOGRAPHIC INFORMATION SYSTEMS, 7., 1999, Kansas City, USA. **Proceedings...** Kansas City: ACM Press, 1999.

ABSTRACT

Frameworks and Patterns are important instruments that enable the reuse of successful software solutions in recurrent problems. Geographic information systems, on the other hand, are usually designed by people with only little knowledge of database modeling techniques. Therefore, we believe that analysis patterns and conceptual frameworks can both facilitate and improve geographic database design in many organizations.

In order to be used by designers with such a wide variety of backgrounds (e.g., cartographers, biologists, and architects), analysis patterns and conceptual frameworks for geographic database design must rely upon an easy to learn and to use, semantically rich conceptual data model.

This paper introduces the **GeoFrame**, a conceptual framework that serves as a starting point for the conceptual modeling of geographic databases. It relies on an extension of the Unified Modeling Language (UML) that can also be used to define new analysis patterns for geographic applications. In addition, we present an example of an analysis pattern that was extracted out of a database schema which was itself derived from **GeoFrame**. This pattern can be identified in various urban GIS applications.

KEYWORDS: Geographic databases; Conceptual design; Analysis pattern

1. INTRODUCTION

The nineties have witnessed a great evolution in the geo-processing technology as well as a significant increase in the number of installed Geographic Information Systems (GIS). Such systems have been adopted either in public or in private administration environments.

The increasing expansion of the GIS base installed is related, on the one hand, to the price reduction of data capture systems and of geo-processing software. On the other hand, such popularization is due to the wide spread of the geo-processing culture and the consequent expansion of the set of applications being implemented.

Along with the increase on the number of GIS users, the search for existing geo-spatial digital data to be reused in new applications and/or new types of spatial analysis is also growing.

Although there is a tendency to increase geo-spatial data sharing, mainly with the aid of systems running through the Internet as, for example, FGDC's Clearinghouse [FED 97], little has been done to facilitate the sharing of modeling solutions of

geographic databases. However, the unsystematic use of imported data from several (many times unknown) sources, can result in reliability problems in the system eventually leading to inconsistency and redundancy of data in the local geographic databases.

Another problem that can be observed is that when the data either do not exist, or are not appropriate for reusing, the respective existent geographic database design is in most cases also not reused, and new databases are developed from scratch [GOR 97]. For example, an information system for urban territorial tax collection, developed for a certain municipal district, could serve several other municipal districts (perhaps with minor adaptations), even though its related data can by no means be reused. In similar applications, many of the modeled entities tend to be repeated, alongside with the relationships among them. This leads us to research on new methods can support the reuse of geographic database design in an easy way.

While *design patterns* [GAM 94] can serve as reusability tools that provide for repeating design solutions for recurrent problems, *analysis patterns* [FOW 97, HAY 95] are defined to make the reuse of conceptual modeling solutions possible. *Analysis patterns* use terms and jargons from the application domain, while *design patterns* supply software implementation solutions for the modeled problems.

Due to the fact that GIS application designers have different backgrounds (e.g., cartographers, biologists, and architects), and usually have little knowledge of software development methods, geographic applications are strong candidates to the use of *analysis patterns* [LIS 98]. The work of these designers can be facilitated if database conceptual design, in several application domains, is documented and registered as *analysis patterns*.

An *analysis pattern* has to be described in such a way that any designer can easily understand the problem being presented and the solution being proposed. The existence of a common conceptual model for pattern specification is a fundamental requirement for the success of the adoption of *analysis patterns* by GIS application designers. Though, there is nowadays no consensus among GIS designers about a conceptual model for geographic applications.

Following the trend of research areas such as software engineering and databases, we chose the Unified Modeling Language (UML) [BOO 98] to use in geographic databases design. However it still needs some extensions as those proposed by several specific conceptual data models for GIS (e.g., GeoOOA [KOS 97], Geo-ER [HAD 97], Geo-OMT [BOR 97], MADS [PAR 98]).

Even some new CASE tools, like Perceptory [BED 99] and REGIS [ISO 99], that have been treated as belonging to a second generation of CASE tools for GIS conceptual modeling [BED 99], still present some limitations as, for instance, the lack of support to modeling geographic fields.

This paper presents **GeoFrame**, an object-oriented conceptual framework, that serves as a basis for class diagram construction to model geographic databases. A proposal of an UML extension for class diagram specification as well as *analysis patterns* description for geographic applications is also presented.

The remainder of the paper is organized as follows. Section 2 presents the structure of **GeoFrame**, describing each one of its classes. Section 3 explains how application class diagrams can be created starting from **GeoFrame** and using methods of schema simplification. In Section 4, an example of GIS *analysis pattern* is presented. The example was built based on **GeoFrame** and documented using the notation proposed in Section 3. Finally, Section 5 presents our conclusions.

2. GEOFRAME

GeoFrame is a conceptual framework that supplies a diagram of basic classes in order to assist the designer in the conceptual modeling of geographic phenomena as well as in specifying *analysis patterns* for geographic databases. **GeoFrame** class diagram is an evolution of a basic class library that has been developed to be used as a starting point for the conceptual modeling of geographic applications [LIS 96]. In [IOC 98], the reader can find a description of how this library has evolved over time. In [LIS 97], we discuss its use and evaluation within the development of a real GIS system called SIGPROGB.

Souza [SOU 98] defines a framework as being “a generic design in a domain that can be adapted to specific applications, being useful as a mold for constructing applications”. This definition provides a more comprehensive view of the potentiality of a framework than the definitions presented by object-oriented programmers. For instance, Ralph Johnson defines a framework as being “a reusable project of a program, or a piece of a program, expressed as a set of classes” [JOH 92].

As a reusability instrument, a framework does not need to be implemented in a programming language to supply a partial solution for a family of problems. In this work, the concept of framework is used in a more generic approach expressing the idea of an unfinished conceptual design schema for a family of applications.

GeoFrame relies on various interesting proposals of GIS specific, conceptual data models [CAM 96, HAD 97, KOS 97, OLI 97, SHE 97, TIM 94]. Each of the proposed models, though, presents limitations that can make the specification of *analysis patterns* more difficult. For example, the hierarchy of classes proposed in the MGeo model [TIM 94], later on extended to the MGeo+ model, mixes elements of different levels of abstraction (e.g. it includes classes like *GeographicDatabase*, *Symbol* and *OrientedLine*). A similar problem happens in the basic data schema proposed in the GISER model [SHK 97].

The GMOD model [OLI 97] supplies quite a complete class diagram. However, since the model is part of a computational environment for the development of geographic applications, its class diagram includes elements whose aim is to give support to the environment. For example, *Project* and *Document* are classes of the diagram related to the environment. Besides, processes, rules and relationships are all specified as classes, so that they can be kept as objects of the environment. These characteristics make the GMOD model unnecessarily complex for the definition of *analysis patterns* of geographic databases. Furthermore, the approach based on instances specification for user created class diagrams results in quite large final schema, thus reducing their readability. Despite these problems, several other ideas present in GMOD were incorporated to **GeoFrame**.

From the Modul-R [BED 89] and GeoOOA [KOS 97] models, we have taken (and extended) the idea of simplifying schemas through the use of graphical symbols (see Section 3).

2.1 GeoFrame Class Diagram

GeoFrame was defined in according to the rules of the Object Orientation formalism. This was done using the graphic notation of the UML class diagram [BOO 98]. The choice of UML was based on its wide acceptance by designers and by software industry in general, and on its tendency of to become a standard language for object modeling. One of the main requirement for *analysis patterns* specification is to use a well-known, easy to understand language. UML has demonstrated to be a good

answer to that requirement as can be noted through the facilities provided by new GIS CASE tools (e.g., Perceptory [BED 99] and REGIS [ISO 99]).

Figure 1 shows **GeoFrame** class diagram. There are four main classes: GEOGRAPHICREGION; THEME; NONGEOGRAPHICOBJECT; and GEOGRAPHICPHENOMENON. They generalize, in a high level of abstraction, the elements of a geographic data schema.

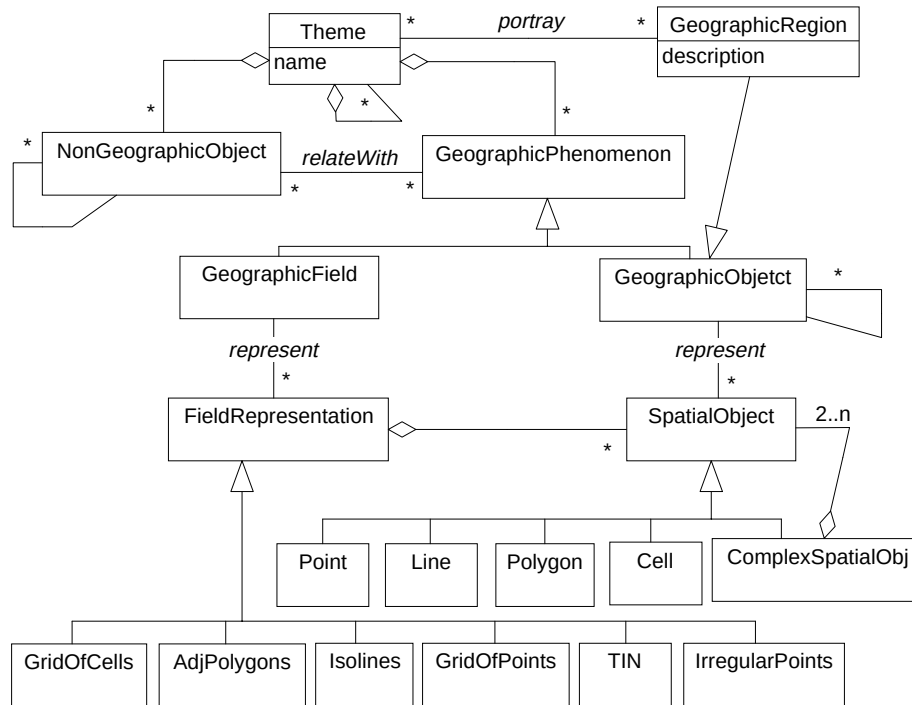


FIGURE 1 - GeoFrame classes diagram

The classes THEME and GEOGRAPHICREGION are the basis of any geographic application. All geographic applications have as their main objective the management and the manipulation of a data set for a certain region of interest, constituting a geographic database. For example, in an urban GIS application the city's area can be specified as being the geographic region of interest. For this geographic region, one can imagine that the following themes might be defined: limits of the urban zone; road network; neighborhoods; buildings (e.g., schools, hospitals); public transportation; and zones of garbage collection. It is possible that some additional themes need to be defined for some sub-areas. This would be the case, for instance, of special safety and emergency zoning for the downtown area.

Thus, a collection of themes can be specified for each geographic area. The organization of the geographic information in themes is directly related to the way the user uses it to both organize and process his/her geographic information. We have taken into account that before the advent of computerized GIS, the user of this kind of information (normally in the form of analog maps) already classified data according to some specific criteria in order to reduce the complexity of the system.

The conceptual specification of a theme does not necessarily entail the implementation of a data layer in a GIS. A theme defined during conceptual design can lead to the implementation of several layers. This issue depends on the software used as well as on the user's needs. Due to the limitations of some current GIS, a simple theme as *Rivers* can generate more than one data layer, for example, one containing only linear spatial objects and another containing polygonal spatial objects. These details should be

hidden from the user during conceptual design. Therefore they will not be discussed in this paper.

Another advantage of using the concept of theme, in the conceptual schema, is that it works as a complexity reduction mechanism for large database schemas. Geographic applications with hundreds of modeled entities (e.g., SIGPROGB [LIS 97]) are quite common. The use of themes allows the designer to divide the schema into cohesive sub-schemas, in which classes that are strongly related to each other are put together. In large projects, groups of related themes can be grouped in a more generic theme, thus forming a hierarchy of themes.

In a geographic database, as in any information system, there exist some objects with no geo-referencing, i.e., conventional objects. In **GeoFrame**, they are considered instances of subclasses of **NONGEOGRAPHICOBJECT**.

The abstract class **GEOGRAPHICPHENOMENON** generalizes any phenomenon whose location in relation to the earth's surface is being considered. For example, a land parcel could be considered as an instance of **GEOGRAPHICPHENOMENON** if its spatial attributes were to be represented in the database.

Geographic phenomena and non-geographic objects can be related to each other (Figure 1, *relateWith* association), as in the case that "every parcel belongs to some owner". The modeling of that kind of relationship allows the data stored in a GIS to be integrated with other information systems.

Geographic phenomena are perceived by users according to the view dichotomy of fields and objects [GOO 92]. These two views cause geographic phenomena to be modeled in different ways. The classes **GEOGRAPHICFIELD** and **GEOGRAPHICOBJECT** specialize the class **GEOGRAPHICPHENOMENON**, allowing to the designer to specify in a distinct (yet integrated) way, the geographic fields and objects, respectively.

The abstract class **GEOGRAPHICOBJECT** is a generalization of all classes of the domain that are perceived in the object view. Those classes representing geographic phenomena that can be individualized (possessing an identity and characteristics that can be described through a set of attributes) are included in this case. Classes such as *Parcel*, *River*, *Highway*, and *City* should be modeled as a specialization of the **GEOGRAPHICOBJECT** class.

It is important to notice the difference between the concept of geographic object, with is a geographic phenomena perceived in the object view (in opposition to the field view) and the concept of object, the basic constructor of object-oriented models.

The *relateWith* association (Figure 1), inherited from the **GEOGRAPHICPHENOMENON** class, can be specialized to represent the possible associations between geographic objects and non-geographic objects. Figure 2 shows a sub-schema of an urban GIS in which the *Parcel* class, a subclass of **GEOGRAPHICOBJECT**, is associated with both the *Owner* and the *TaxType* classes, subclasses of **NONGEOGRAPHICOBJECT**, as well as with the *Block* and *Street* classes, subclasses of **GEOGRAPHICOBJECT**.

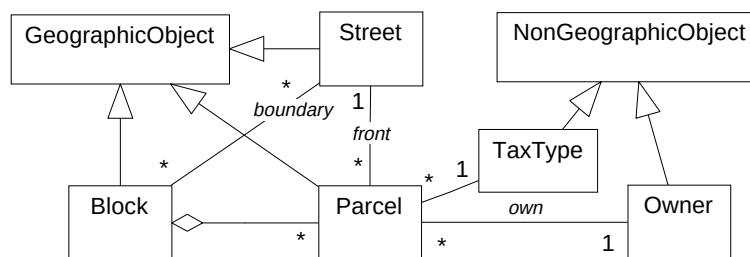


FIGURE 2 - Example of *relateWith* relationships

In its turn, the abstract class `GEOGRAPHICFIELD` generalizes phenomena that fit in the field view. Geographic fields are modeled as functions on a variable. Some fields refer to variables that are distributed on the surface, in a continuous way (e.g., *Relief*, *Temperature* and *Soil Covering*), and other fields refer to variables which are distributed in a discrete way (e.g., *Population* and *Epidemic Occurrences*) [OLI 97].

According to Câmara [CAM 96], a geographic field can be mathematically defined by a mapping that abstracts the space distribution of a variable V on an area R of the earth's surface. Thus, a geographic field can be modeled as a class of functions, with the following attributes: *domain*, whose value is a sub-area of R ; *image*, whose value belongs to the domain V ; and *mapping*, whose value is the mapping f of R in V . Thereby, the set of values assumed by V in the sub-areas of R determine the class to which the geographic space R belongs. For example, a geographic field *SoilCovering* can be characterized by the set of values $V = \{NativeForest, Culture, Reforestation, Pastures, \text{ and } UrbanArea\}$. For any point, inside the area R , there is a corresponding value element in the set V .

Thereby, a geographic field can be conceptually modeled as subclass of `GEOGRAPHICFIELD`. Some geographic field types, the so-called categorical fields [CHR 97], have an association with the class that represents the image of the field mapping function, a subclass of `NONGEOGRAPHICOBJECT`. This association is a specialization of the *relateWith* association for geographic fields. Fields whose the function *image* is composed by a set of numeric values do not specialize this association. For example, the *Relief* geographic field has the set of the real numbers as its image.

To Kemp [KEM 98], there are many philosophical questions in understanding of geographic fields. Some of these open questions include: “Do the fields only exist when they are measured? Do the fields possess their own characteristics? Is the attribution of a certain value to an area real, or is a conceptual mechanism addressed to a specific objective?”. Despite the fact that these questions still do not have a generally agreed answer, the definition of a field as a function in a domain which is a subset of the space-time domain is enough for computational purposes.

In a GIS, the implementation of the spatial representation and relationships among the geographic objects is based on spatial data structures. For example, the location of a pole can be stored through a pair of geographic coordinates (x,y) . Yet, the location of a street intersection should be stored in a network node data structure. Besides the coordinates (x,y) it is necessary to store topological information about the connectivity of this node with the arcs of the network that represent the streets.

The choice of the best data structure to implement the spatial representation of each geographic phenomenon is a task to be performed after the conceptual project, and is not to be considered at this level.

One of the fundamental principles of conceptual modeling is that a conceptual schema should only contain the elements of the domain, discarding implementation aspects [PAR 98]. The goal of including information on spatial objects related to the geographic phenomena in the conceptual schema is to allow the modeling of the spatial component of each phenomenon.

The information about geographic phenomena has three components: attribute, space and time [CHR 97]. These three components must be abstracted from the reality during the conceptual modeling process. Usually, in non geographic applications, only the component attribute (and in some cases the time component) is considered.

Therefore, when one considers points, polygons or grids of cells in **GeoFrame**, he/she is considering the forms of abstraction of the geographic phenomena spatial

component, but not the way in which they will be structured in the database. Nevertheless, the majority of the currently available GIS supplies very similar data structures for the storage of these abstract constructors.

For the purpose of modeling, the main concern is to determine, for example, if a *pole* will have punctual dimensions or if the *pipe* will have linear dimension. It is not necessary, however, to consider that the pipe will be stored by means of an arc directed in a vector structure with topology.

Some geographic phenomena can present complex spatial dimensions, that is, dimensions composed by other spatial objects (e.g., an archipelago). The **SPATIALOBJECT** class generalizes classes that are used for specification of the spatial component representation of the geographic phenomena in object view. They are **POINT**, **LINE**, **POLYGON**, **CELL** and **COMPLEXSPATIALOBJ**. The **SPATIALOBJECT** class was defined based on the *Composite* design pattern [GAM 94].

Notice, in Figure 1, that not even the possible associations among the classes **POINT**, **LINE** and **POLYGON** are represented in the **GeoFrame** class diagram, because any of these associations might not be true according to a specific GIS implementation.

The spatial aspects of a geographic field are abstracted in a different way from the spatial aspects of a geographic object. Chrisman [CHR 97] describes some geographic models through which the spatial component of the geographic information can be abstracted. The models presented by Chrisman, which are appropriate to modeling phenomena in the field view, can be summarized in the six spatial models described by Goodchild: grid of cells; adjacent polygons; isolines; grid of points; irregular triangular network; and irregularly sampled points. In **GeoFrame**, these models correspond to the subclasses of the **FIELDREPRESENTATION** class, namely **GRIDOFCELLS**, **ADJPOLYGON**, **ISOLINES**, **GRIDOFPOINTS**, **TIN** and **IRREGULARPOINTS**.

In a GIS, those models will be, at a later stage, implemented through the raster and vector representation models [GOO 92]. Each one of the six geographic field representation models can be implemented either in the raster or in the vector model, although some mappings are more natural. For example, the mapping of a field whose component has been abstracted by means of a grid of cells should be implemented in the raster model.

A geographic field can have its spatial component abstracted in different ways, i.e., through more than one of those field representation models. For example, the field *Temperature* can be abstracted by means of irregularly sampled points or by means of isolines. A similar situation happens with the geographic objects whose spatial components can be perceived, sometimes by alternative geometry, other times by double geometry, depending on aspects like the scale with which one intends to capture the spatial dimension of each phenomenon. The possibility of having multiple representations is shown in Figure 1, through the (1:n) *represent* association.

2.2 Relationships in GeoFrame

Figure 2 shows that geographic phenomena and non-geographic objects can be related to each other. The relationships among classes that are represented as associations in the diagram, allow for the specification of integrity constraints that should be hold in order for the database to remain consistent.

The relationships can be divided into three categories: *semantic*, *spatial* and *temporal*. Semantic relationships complement the description of knowledge regarding the descriptive aspects of the phenomena (e.g., the streets building blocks and a block can contain several parcels). Spatial relationships establish associations among the locations of the phenomena (e.g., the spatial limit of a block contains the limits of its

parcels). The modeling of the temporal aspects is under investigation, and will be treated in a separate paper.

Some GIS conceptual models propose new modeling constructors for spatial relationships. For example, the GeoOOA model [KOS 97] defines three new kinds of constructors for spatial aggregation: *covering*, *containment* and *partition*. The Geo-ER model [HAD 97] defines only one constructor for spatial aggregations. Even though these constructors allow the specification of spatial integrity constraints, they present two disadvantages. First, they make it more difficult for users/designers to understand the model and, second, they are incomplete, i.e., they do not allow several other relationships to be modeled through similar mechanisms. For example, in the GeoOOA model it is possible to specify that the location of an industry should be inside (*containment* structures) of an industrial district area, but at the same time there is not a constructor, as detailed as this, to specify that this district can be crossed by an avenue or that it can be adjacent to a river.

No matter how hard the designer tries to specify the restrictions of spatial integrity, he/she will always have a much bigger volume of relationships that will not have been specified. So, the relationship modeling approach in **GeoFrame**, that has been adopted in the development of our systems [LIS 96, LIS 97], is just modeling the semantic relationships, leaving the spatial relationships to be considered in the data acquisition and data quality assessment processes. Thus, we only use the relationship modeling constructors from the UML, i.e., association, specialization, aggregation and composition [BOO 98].

3. BUILDING CONCEPTUAL SCHEMA FROM THE GEOFRAME

GeoFrame provides a basic class diagram for the elaboration of conceptual schemas of geographic databases. In this section some modeling guidelines are described, as well as a proposal for simplification of schemas based on stereotypes, a UML mechanism of extensibility [BOO 98].

A conceptual schema is built starting from the specialization of **GeoFrame** classes. The geographic database conceptual modeling, using **GeoFrame**, is accomplished according to a top-down approach composed of three stages. Initially, for each geographic area considered, several themes (and sub-themes) are identified. In the second stage a class diagram is specified for each identified theme. The specification of the associations among classes of different themes is made still in that stage. Finally, the analysis and modeling of the spatial component of each geographic phenomenon is done. Each one of these phases is described in more detail in the following sections.

3.1 Theme Diagram

As described in Section 2, each geographic region is portrayed by zero or more themes. However, a theme can also be associated with more than one geographic region. A theme is specified as an aggregation of geographic phenomena classes, of non-geographic classes and of other more specific themes.

To increase the readability of the resulting schemas, themes are not modeled as subclasses of the THEME class, but as an UML constructor called *package*. A package is composed of a set of UML elements that could be of any type as, for example, classes, associations and other packages [BOO 98].

Figure 3 shows an example of theme hierarchical diagram defined for an application of environmental control (SIGERCO System) [LIS 96], developed for the coast of the Rio Grande do Sul state, Brazil. As it can be seen in Figure 3, themes are

associated with an object, an instance of `GEOGRAPHICREGION` class. For each region of a project the important themes should be specified. The theme diagram was based on the Geo-OMT model [BOR 97].

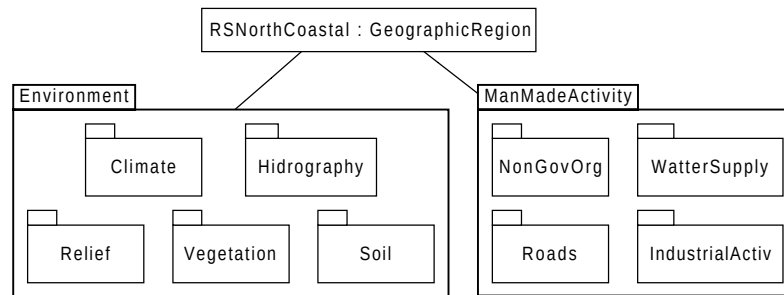


FIGURE 3 - An example of theme diagram

3.2 Specifying Geographic Phenomena and Non-Geographic Objects

After the theme identification stage, each theme should be refined. In the object-oriented approach, a geographic database schema is represented by the class diagram, which describes the geographic phenomena, the non-geographic objects and the possible relationships among them.

Each class identified in the domain should be modeled as a subclass of one of the following **GeoFrame** classes: `NONGEOGRAPHICOBJECT`, `GEOGRAPHICFIELD` or `GEOGRAPHICOBJECT`. To avoid visually overloading the diagram, stereotypes are used as a schema simplification mechanism. A stereotype extends the UML vocabulary (i.e., the metamodel) allowing the user to create new constructor types that can be used like any other element of the language [BOO 98].

To substitute the generalization relationships between the domain's classes and the **GeoFrame** classes, three stereotypes were defined (Figure 4). Figure 4 also shows examples of classes using these stereotypes, namely *NonGovOrg* (Non-governmental Organization), subclass of the `NONGEOGRAPHICOBJECT` class, *LandUse*, subclass of the `GEOGRAPHICFIELD` class and *River*, subclass of the `GEOGRAPHICOBJECT` class.

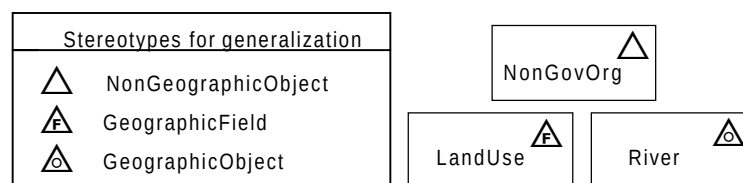


FIGURE 4 - Stereotypes⁹ for generalization

The specification of these stereotypes will make the stage of mapping between the conceptual and logical design simpler.

3.3 Modeling the Spatial Component of Geographic Phenomena

In **GeoFrame** every geographic field or object can be represented by multiple instances of the `FIELDREPRESENTATION` and `SPATIALOBJECT` classes, respectively (Figure 1). A geographic phenomenon can have multiple representations for several

⁹ These stereotypes are defined as characters of a special MS-Windows font, called SIGMODA, available for download at <http://www.inf.ufrgs.br/gpesquisa/sigmoda>. Some stereotypes were reused from the CASE tool Perceptory [2].

reasons. Among them, we can mention the need of multiple scales, users with different views of a same phenomenon and temporary versions.

The possibility of having multiple representations for a geographic phenomenon is modeled through different associations between geographic phenomena and possible forms of abstraction of their spatial components (*represent* association in Figure 1). Bédard [BED 99] proposes a set of possible variations for the modeling of the spatial component as, for example, optional geometry, simple and complex aggregates, alternative geometry and derived geometry. In **GeoFrame** these variations can be specified by the designer through the free combination of different stereotypes in the same class. However, this needs additional documentation in the data dictionary.

A second set of stereotypes (Figure 5) replace the great quantity of associations that results from modeling spatial components of the geographic phenomena. The semantic of this type of stereotype is the replacement of explicit associations between geographic phenomena and their spatial representations. Stereotypes also indicate geometric forms of spatial representations.

SpatialObject	FieldRepresentation
• Point □ Polygon	□ GridOfCells □ GridOfPoints
— Line ★ Complex	□ AdjPolygons □ TIN
	□ Isolines □ IrregularPoints

FIGURE 5 - Stereotypes for *represent* association

Figure 6 shows part of a conceptual database schema for an urban application developed from **GeoFrame**. In this example the mechanisms of schema simplification are showed. Some classes do not present stereotypes (e.g., *BuiltParcel*) because the respective relationships are inherited from its super-class.

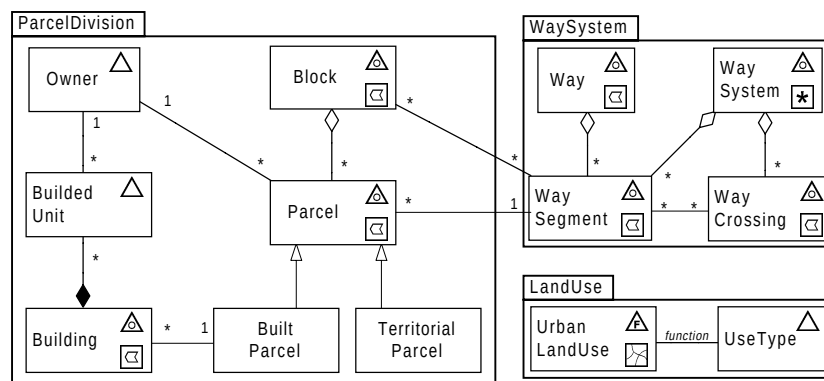


FIGURE 6 - Example of class diagram using GeoFrame

The use of graphical symbols as a schema simplification mechanism was presented, originally, by Bédard and Paquette [BED 89]. This technique has been used in several GIS conceptual models as, for example, GeoOOA [KOS 97], Geo-OMT [BOR 97] and MADS [PAR 98], as well as in the CASE tools Perceptory [BED 99] and REGIS [ISO 99]. However, these models present some limitations that make them inadequate for adoption as models for GIS *analysis patterns* description. Among these limitations, we can cite the following: (1) the models Modul-R and MADS do not support geographic fields modeling, while the GeoOOA model proposes a mistaken solution, in which the concepts of geographic field and raster structures are used in an ambiguous way; (2) the over use of graphical symbols reduces the readability of the resulting schemas in the models Geo-OMT and MADS; (3) the GeoOOA model

(including the CASE tool REGIS) does not support the specification of multiple representations of the spatial component.

GeoFrame approach allows the construction of easily readable schemas either by designers or by users, besides supporting the main requirements for the conceptual modeling of geographic information (e.g., geographic field modeling). An important contribution of this work is the stereotype use as a replacement for generalization-specialization relationships. This idea can be adopted as a simplification mechanism for schema containing complex class hierarchies.

4. EXTRACTING GIS ANALYSIS PATTERNS FROM GEOFRAME DERIVED SCHEMAS

From a careful analysis of existing geographic database designs, we can extract parts of class diagrams with potential to be transformed into *analysis patterns* to be reused in other geographic applications.

An *analysis pattern* presents, apart from the classes diagram proposed as the problem solution, a set of information (e.g., context, example) that helps the designer to understand the pattern and to evaluate its adequacy as a solution to a given problem. Thus, the description of a new pattern requires an additional effort, mainly from experienced designers, in creating a library of GIS conceptual modeling solutions.

An *analysis pattern* should just present the essence of the problem, discarding the specific particularities of each application. The following sub-section presents, as an example, a simple *analysis pattern* extracted from the schema presented in Figure 6.

4.1 An Example of Analysis Pattern

Pattern name: Urban Way System

Problem: How to model an urban way system?

Context: Almost all the cities in the world present a similar organization pattern. A city is subdivided in ways (e.g., streets, avenues, paths) forming a network of ways. Each way should have an identification code (e.g., ZIP code) and a name, apart from being divided into several segments. A way's segment corresponds to the space between two other way's segments that intercept or cross it. Each way's segment can be identified by a sequential number (associated to the way's identification) and has as attribute an interval of parcel numbers. The set of way segments and connections constitute a way system.

Solution: Figure 7 shows the pattern's class diagram.

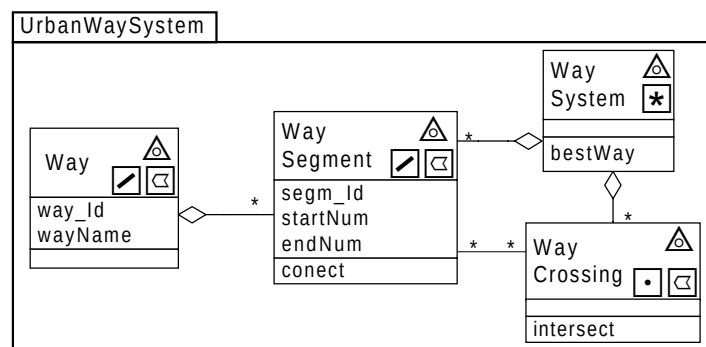


FIGURE 7 - Urban Way System analysis pattern

For each geographic phenomenon the pattern specifies only the more generic attributes and operations, which are extended in accordance with each specific

application. On the other hand, the possible abstractions of the spatial components are specified in the pattern. For example, the spatial component of the *Way* class is specified in Figure 7 as being linear or polygon, even though the same class is specified with distinct dimensions in the examples of Figure 6 and Figure 8. Due to space limitations, the specification of the attributes and of the operations were omitted in the examples.

Example: Figure 8 exemplifies the *analysis pattern* use in a public transportation application. The example was extracted from a conceptual schema presented in [BOR 97].

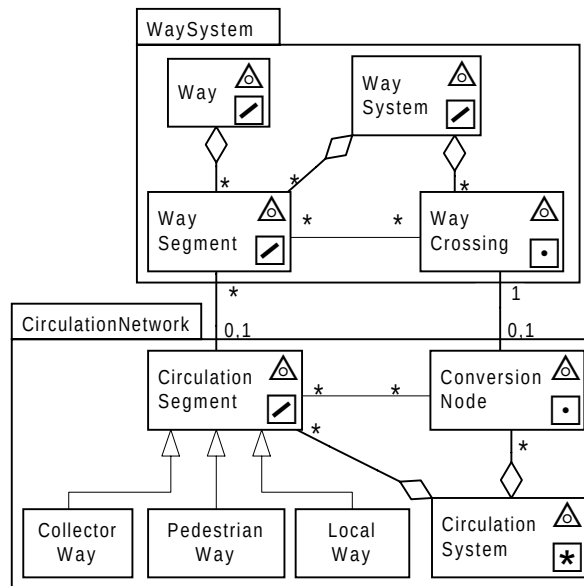


FIGURE 8 - Example of use of the analysis pattern

5. CONCLUSIONS

Frameworks and patterns have been used as tools for solutions reuse in recurrent problems, by the object-oriented system designers' community. The authors believe that the use of *analysis patterns* can facilitate and improve the geographic applications design, mainly due to the wide domain of GIS database and also to the great number of applications designed by users who are not familiar with modern techniques of modeling.

One of the requirements for the success of the use of *analysis patterns* in geographic applications is that the patterns should be described in a language that is semantically rich but easy to understand. This paper presented **GeoFrame**, a conceptual framework that serves as a base for the conceptual modeling of geographic databases. Furthermore, an extension of the UML graphical notation for specification of geographic databases conceptual schemas, as well as for the description of GIS *analysis patterns* based on **GeoFrame** was presented.

Another contribution of this paper is the proposition of a method for the conceptual modeling of geographic phenomena that are perceived in the field view. Most of the models proposed in the literature deal only with phenomena in the object view. The method stresses the existence of different abstraction levels, in which the conceptual constructors should not be confused with implementation aspects. Thus, it must be clear that the inclusion of classes such as *Point*, *Polygon*, *Grid of Cells* in the class diagram, is related to the form of perception of the spatial component of each

geographic phenomenon and not to the data structures used to store those data in the GIS.

ACKNOWLEDGMENTS

The authors thank Willian Wazlawik for the confection of the graphical symbols for SIGMODA font and Karla Borges for supplying a series of conceptual design examples in urban GIS applications. This work was developed in part with SIGMODA project funds, a research project supported by CNPq.

4.4 Artigo IX: Padrões de Análise para Banco de Dados Geográficos

Jugurta Lisboa Filho e Cirano Iochpe

Artigo a ser submetido em 2000.

RESUMO

Um padrão de análise é qualquer parte de uma especificação de requisitos que se origina em um projeto e pode ser reutilizada em diversos outros projetos. O presente artigo introduz um conjunto de padrões de análise para bancos de dados geográficos, padrões estes identificados a partir de esquemas conceituais de aplicações de Sistemas de Informação Geográfica (SIG) nas áreas de controle ambiental e administração urbana. O artigo também apresenta uma introdução sobre padrões de análise e propõe a adoção de uma linguagem para descrição de padrões de análise para banco de dados geográficos com base no *framework* GeoFrame [LIS 99].

Palavras-chave: Projeto conceitual, Reutilização, Padrões de análise, SIG.

1 INTRODUÇÃO

O sucesso do desenvolvimento de grandes sistemas de informação tem como um de seus pontos chave a realização da análise de requisitos de forma metodológica e não ambígua. Experiências têm demonstrado que a etapa de análise de requisitos e projeto conceitual do banco de dados são atividades complexas e que demandam muito tempo. Segundo Johannesson [JOH 99], uma das razões para isso é que o conhecimento do domínio e o levantamento dos requisitos do sistema são feitos quase sempre desde o início, para cada novo sistema a ser desenvolvido.

Aplicações desenvolvidas com base em Sistemas de Informação Geográfica (SIG), embora apresentem alguns requisitos especiais (ex.: manipulação de dados referenciados geograficamente), devem ser desenvolvidas utilizando-se técnicas que são aplicadas com sucesso no desenvolvimento de qualquer sistema de informação. Uma das técnicas que vem recebendo atenção especial, principalmente, pela comunidade de projetistas de sistemas orientados a objetos, é o emprego de instrumentos que possibilitam a reutilização de componentes de software através da definição de padrões.

Um padrão é uma combinação recorrente de elementos de modelagem que ocorrem em algum contexto [FER 98]. Padrões podem ser aplicados nas diversas etapas do desenvolvimento de software, recebendo, conseqüentemente, diferentes denominações como padrões de análise, padrões de projeto, padrões de arquitetura, idiomas (padrões de implementação), etc.

A reutilização de componentes de software vem sendo feita, embora informalmente, desde a implementação dos primeiros programas de computador. O uso de padrões de projeto na área da Ciência da Computação é bem mais recente, tendo recebido um grande impulso após a publicação do livro de Gamma [GAM 94]. Segundo ele, “um padrão apresenta a essência de uma solução para um problema recorrente em um contexto específico”.

O uso de padrões na área de desenvolvimento de software está se tornando cada vez mais importante, devido a sua contribuição para reusabilidade e qualidade. O uso de padrões possibilita a disseminação do conhecimento e a troca de experiências entre os projetistas, bem como facilita a comunicação entre os membros envolvidos no projeto [ROB 93].

Embora ainda não tão famosos, uma outra categoria de padrões, a dos padrões de análise, vem sendo proposta como instrumento para a reutilização de soluções durante as fases de análise de requisitos e modelagem conceitual do banco de dados [ROB 93, HAY 95, FOW 97, FER 98, JOH 99].

Segundo Fernandez [FER 98], dentre os motivos que diferenciam os padrões de análise dos padrões de projeto, pode-se citar:

- padrões de projeto estão mais próximos da implementação por focar, principalmente, os aspectos típicos de projeto como, por exemplo, interfaces homem-máquina, criação de objetos, propriedades estruturais básicas;
- padrões de projeto podem ser aplicados a um número maior de aplicações. Por exemplo, a maioria das aplicações possuem interface homem-máquina;
- padrões de análise são dependentes da aplicação, pois sua semântica descreve aspectos específicos de algum domínio ou aplicação.

Uma comparação dos diversos tipos de padrões existentes pode ser encontrada em [BUS 95].

Uma particularidade das aplicações de SIG é que, normalmente, os dados manipulados por essas aplicações possuem um forte relacionamento entre si, devido a esses dados retratarem fenômenos geográficos que ocorrem sobre uma mesma região geográfica. Por exemplo, muitos dados espaciais (ex.: dados temáticos) são derivados de dados básicos (ex.: dados topográficos). O conjunto de dados que compõem a base cartográfica possui uma estrutura conceitual muito parecida na maioria das aplicações. Essa particularidade torna as aplicações de SIG fortes candidatas a se beneficiarem da reutilização de projetos de bancos de dados já existentes, como já vem ocorrendo com o compartilhamento de dados geo-espaciais em meio digital [LIS 98].

Este artigo tem como objetivo apresentar um conjunto de padrões aplicáveis as etapas de análise de requisitos e modelagem conceitual de bancos de dados geográficos. O restante do artigo está organizado como segue. A Seção 2 apresenta algumas características dos padrões de análise, detalhando o processo de criação e documentação dos mesmos. As seções 3 e 4 apresentam padrões de análise para aplicações de banco de dados geográficos nas áreas de controle ambiental e gestão urbana, respectivamente. A Seção 5 apresenta algumas considerações finais e perspectivas de trabalhos futuros.

2 PADRÕES DE ANÁLISE

Um padrão de análise é qualquer parte de uma especificação de requisitos que se origina em um projeto e pode ser reutilizada em diversos outros projetos [ROB 93]. Um padrão de análise descreve, em um nível arbitrário de abstração, um conjunto de objetos do mundo real, seus relacionamentos e as regras que definem seu comportamento e estado [JOH 99].

Padrões de análise têm sido desenvolvidos para possibilitar a reutilização de soluções de análise em diferentes sistemas. Alguns padrões são menos genéricos e podem ser reutilizados em diferentes aplicações dentro de um mesmo domínio, enquanto outros padrões podem ser aplicados em diferentes domínios. Como exemplos de padrões de análise, específicos para um domínio, pode-se citar os padrões para aplicações na área de seguradoras [ROB 93] e os padrões para aplicações na área

bancária [MAR 99]. Dentre os exemplos de padrões mais genéricos estão o padrão para reservas e locação de entidades reutilizáveis (ex.: reserva de quarto de hotel, reserva para aluguel de automóvel) [FER 99], os padrões *Observações e Medidas* [FOW 97] e os padrões *Contratos e Documentos* [HAY 95].

Johannesson [JOH 99] identifica dois requisitos de qualidade que um padrão de análise deve atender. O primeiro é que ele deve ser suficientemente genérico, isto é, um padrão de análise deve ser aplicável a diferentes sistemas. O segundo é que ele deve ser fácil de entender e mais fácil de usar do que modelar uma nova aplicação desde o início. Certamente, esses dois requisitos são conflitantes, uma vez que padrões altamente genéricos podem ser de difícil adaptação em novos projetos. Isso impõe um grande desafio aos projetistas que estão documentando padrões, que é a tarefa de encontrar o equilíbrio entre descrever os padrões de forma abstrata, para que possam ser realmente reutilizados, mas que também sejam de fácil entendimento por parte de seus possíveis usuários.

Embora a abordagem de reutilização tenha como uma de suas principais vantagens o aumento da produtividade dos projetistas, o uso de padrões requer a realização de algumas tarefas adicionais como, por exemplo, a descrição e divulgação dos padrões. Para isso, algumas etapas extras devem ser cumpridas, a primeira delas é a escolha de uma linguagem de consenso para a descrição dos padrões. Em seguida, vem a documentação dos próprios padrões e, por último, a necessidade de tornar desejável e efetiva a reutilização entre os projetistas. As subseções seguintes abordam estes aspectos em maior detalhe.

2.1 Reconhecendo novos padrões de análise

O processo de descoberta ou captura de um novo padrão de análise requer, do analista, habilidade e experiência suficientes para perceber analogias existentes entre partes de um projeto, de um sistema em estudo, que possam ser recorrentes em outros sistemas. Segundo Fernandez [FER 98a], o processo de criação de padrões é feito através da identificação de partes de um projeto que são possíveis candidatas à reutilização. Em seguida, esse sub-diagrama deve ser abstraído, suas especificidades devem ser eliminadas e o padrão deve ser documentado de acordo com uma linguagem de descrição de padrões.

Por sua vez, a aplicação de padrões requer a percepção de uma analogia entre um problema atual e os padrões disponíveis, com a conseqüente adaptação no projeto, do padrão identificado. A Figura 1 ilustra o processo de criação e uso de padrões.

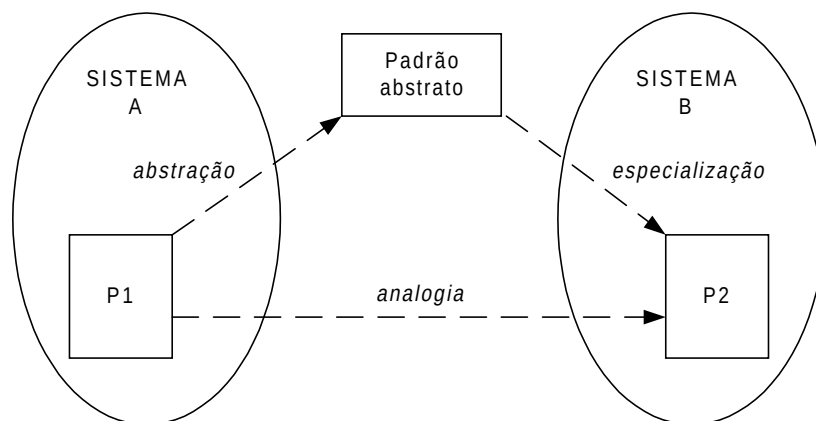


FIGURA 1 - Processo de captura e aplicação de padrões

Para Robertson [ROB 93], a característica mais importante que um analista necessita possuir para identificar novos padrões de análise é a habilidade de abstração. O processo de descoberta de padrões reutilizáveis é uma tarefa muito mais árdua do que sua utilização. Quando da definição de um novo padrão, detalhes possuindo características próprias de um sistema, ou da organização para a qual o sistema tenha sido desenvolvido, devem ser eliminados da descrição do padrão. O padrão deve incluir apenas os elementos que são genéricos o suficiente para que possam ser reutilizados. Durante a fase de análise, novos candidatos a padrão devem ser especificados e disponibilizados.

Rawsthorne [RAW 96] apresenta uma linguagem de padrões¹⁰ para análise de requisitos em um contexto de projeto de sistemas orientados a objetos. Um dos padrões propostos (*Pattern Scavenging*) fornece diretrizes para um analista que procura por novos candidatos a padrão. Isso é feito através do exame de modelos analíticos existentes, buscando-se ocorrências de similaridades na estrutura e comportamento, o que é feito a partir da busca por “formas” similares em diagramas existentes.

2.2 Descrevendo padrões de análise

Inicialmente, os padrões de análise eram descritos de forma narrativa sem nenhum cuidado em seguir uma estrutura padronizada, como em [HAY 95 e FOW 97]. Recentemente, a maioria dos padrões de análise está sendo descrita de forma semelhante aos padrões de projeto, os quais utilizam uma estrutura de tópicos pré estabelecida (*template*). Neste artigo, optou-se por adotar a estrutura mínima definida por Meszaros [MES 98], ou seja, *Problema-Contexto-Forças-Solução*.

- *Problema* – fornece uma declaração sucinta do problema que necessita ser resolvido. Normalmente colocado na forma de pergunta;
- *Contexto* – descreve o contexto no qual o problema foi identificado e para o qual a solução foi proposta;
- *Forças* – conjunto de restrições que foram consideradas quando da elaboração da solução do problema;
- *Solução* – fornece um diagrama de classes com a solução proposta.

Devido às especificidades dos bancos de dados geográficos, seus diagramas são elaborados com base no *framework* conceitual GeoFrame [LIS 99], descrito na seção seguinte.

2.3 Esquema conceitual de banco de dados geográficos

Uma das formas de documentar padrões de análise para projeto de banco de dados é através da elaboração de um esquema conceitual de dados, o que implica na escolha de um modelo conceitual de dados. Para atender aos requisitos de modelagem impostos pelas aplicações de SIG, diversos modelos conceituais têm sido propostos como, por exemplo, os modelos GeoOOA [KOS 97], Modul-R [BED 96], Geo-OMT [BOR 97] e MADS [PAR 99]. Um levantamento detalhado de modelos conceituais de dados para bancos de dados geográficos é apresentado em [LIS 99c].

¹⁰ Um padrão descreve uma solução para um único problema. Uma linguagem de padrão descreve um conjunto de padrões que são usados, de forma conjunta, para resolver um problema mais complexo. Diferente de um catálogo de padrões, uma linguagem de padrões inclui regras e orientações as quais explicam como e quando aplicar cada padrão para resolver um problema maior do que um único padrão possa resolver [APP 97].

Uma característica em comum desses modelos é que todos eles são, na verdade, extensões de um modelo conceitual já consagrado (ex.: E-R [CHE 76], OMT [RUM 91], OOA [COA 91]). Essas extensões introduzem novos construtores abstratos para possibilitar a captura e representação da semântica relacionada aos aspectos espaço-temporais da informação geográfica. Uma segunda característica comum desses modelos é que, se por um lado eles fornecem diversos novos construtores semânticos, tornando-os linguagens de alto poder expressivo, por outro lado, o número excessivo de novos construtores reduz a legibilidade dos diagramas resultantes e aumentam a dificuldade de entendimento, principalmente, por parte de usuários de SIG que muitas vezes projetam seus próprios sistemas.

Alguns dos requisitos mais importantes para a escolha de uma linguagem de descrição de padrões de análise são que ela seja de amplo conhecimento, por parte dos projetistas, que seja de fácil entendimento por parte de usuários leigos, que tenha alto poder expressivo e que seja extensível, a fim de atender a especificidades impostas por características próprias de um domínio de aplicação.

Em [LIS 99], nós mostramos que a Linguagem de Modelagem Unificada (UML) [BOO 98] atende a esses requisitos. Assim, para apresentar o padrão de modelagem proposto, utiliza-se a notação gráfica do diagrama de classes da linguagem UML (Figura 2).

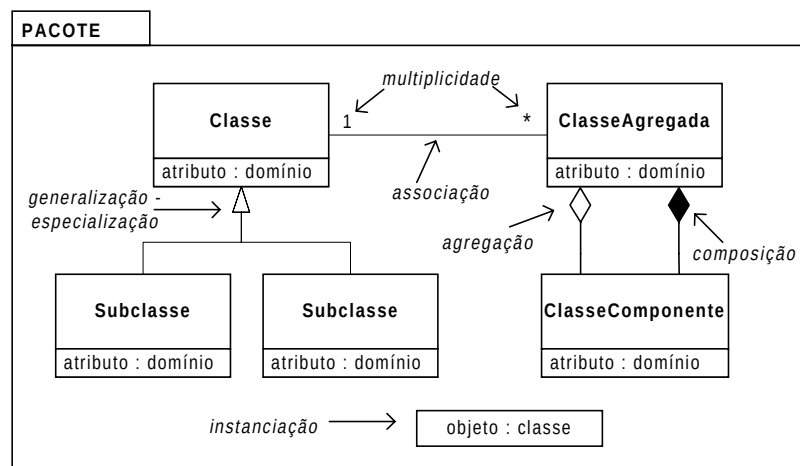


FIGURA 2 - Notação gráfica do diagrama de classes UML (resumida)

Além disso, a modelagem dos aspectos espaciais da informação geográfica é feita com base no *framework* conceitual denominado GeoFrame [LIS 99], juntamente com o conjunto de estereótipos mostrado na Figura 3.

<i>Fenômeno geográfico e Objeto convencional</i>	<i>Componente espacial de objetos geográficos</i>	<i>Componente espacial de campos geográficos</i>
Objeto geográfico Campo geográfico Objeto não geográfico	Ponto Linha Polígono Obj. espacial complexo	Pontos irregulares Grade de pontos Polígonos adjacentes Isolinhas Grade de células TIN
<<função>> <i>função categórica</i>		

FIGURA 3 - Estereótipos do *framework* GeoFrame

O primeiro conjunto de estereótipos (*Fenômeno geográfico e Objeto convencional*) é usado para diferenciar os dois principais tipos de objetos pertencentes a um banco de dados geográficos. *Fenômeno geográfico* é especializado em *Objeto geográfico* ($\triangle$) e *Campo geográfico* ($\triangle$), segundo as duas formas de percepção dos fenômenos geográficos, descritas por Goodchild [GOO 92]. *Objetos convencionais*, ou seja, objetos não referenciados geograficamente são modelados de forma tradicional e são identificados através do estereótipo ($\triangle$).

O segundo e o terceiro conjuntos de estereótipos (*Componente espacial de objetos geográficos e Componente espacial de campos geográficos*) são usados para a modelagem do componente espacial de fenômenos segundo as visões de objeto e de campo, respectivamente. A existência de múltiplas representações é modelada através da combinação de dois ou mais estereótipos. Por exemplo, uma classe *Município* pode ter duas formas de abstração de seu componente espacial, pontual e/ou poligonal, o que é especificado pelo par de estereótipos ($\square \square$).

Por último, o estereótipo <<função>> é usado para caracterizar um tipo especial de associação que ocorre quando da modelagem de funções categóricas. Segundo Chrisman [CHR 97], numa estrutura de cobertura categórica o espaço é classificado em categorias mutuamente exclusivas, ou seja, uma variável possui um valor do tipo categoria em todos os pontos dentro de uma região (ex.: tipos de solos).

3 PADRÕES DE ANÁLISE EM APLICAÇÕES DE CONTROLE AMBIENTAL

Durante o desenvolvimento da modelagem conceitual de diversas aplicações de SIG, na área de controle ambiental [LIS 96, LIS 97, GAR 98] e, também, analisando esquemas de banco de dados projetados por terceiros, percebeu-se que uma série de subconjuntos de classes (e seus inter-relacionamentos) são recorrentes. Esta seção apresenta alguns desses subconjuntos de classes descritos na forma de padrões de análise.

3.1 Padrão: Parâmetros de qualidade ambiental

Problema: Como modelar parâmetros medidos para fins de controle da qualidade ambiental?

Contexto: O acompanhamento da qualidade de fatores ambientais (ex.: água, ar, resíduos sólidos) é feito, tomando-se medições periódicas dos parâmetros de qualidade em diversos pontos de amostragem.

Forças:

- O número de parâmetros pode variar, durante o período de monitoração, para cada ponto de amostragem.
- Cada parâmetro é sempre medido em uma mesma unidade de medida.

Solução: A Figura 4 mostra o diagrama de classes que compõe o padrão. Em cada ponto de amostragem são realizadas, periodicamente, diversas medidas, correspondendo aos diversos parâmetros de controle analisados. Optou-se por associar a classe *Unidade* à classe *ParâmetroControle*, devido ao fato que cada parâmetro sempre é medido na mesma unidade de medida. Caso contrário esta associação seria entre as classes *Unidade* e *Medida*, provocando repetição de dados.

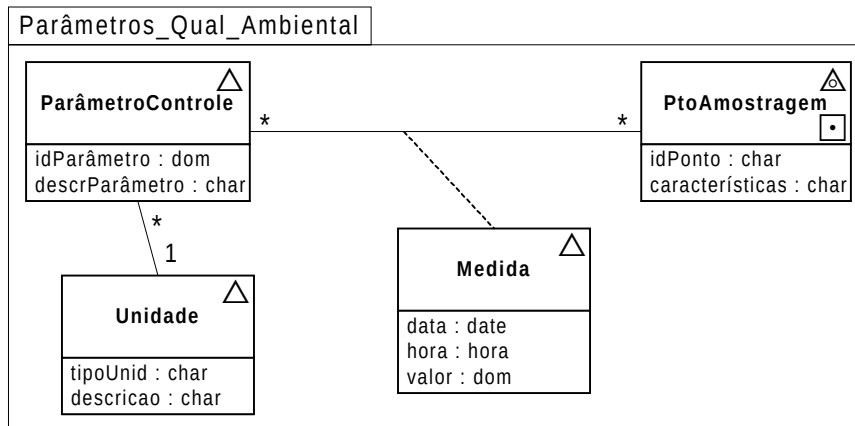


FIGURA 4 – Diagrama de classes do padrão “Parâmetros de Controle Ambiental”

Exemplo: Qualidade da água – Projeto SIGPROGB

No Projeto SIGPROGB [GAR 98], cada ponto de amostragem se localiza dentro de um recurso hídrico. Desta forma, a classe *PTOAMOSTRAGEM* foi associada à classe *RECURSOHÍDRICO* que, por sua vez, possui outras associações no contexto do projeto. O padrão *Parâmetros de Qualidade Ambiental* foi utilizado na modelagem dos parâmetros de qualidade da água, conforme ilustra a Figura 5. Também no projeto SIGPROGB, esse padrão foi reutilizado na modelagem dos parâmetros de controle ambiental em pontos de lançamento atmosférico (ex.: indústrias) e de resíduos sólidos (ex.: cortumes).

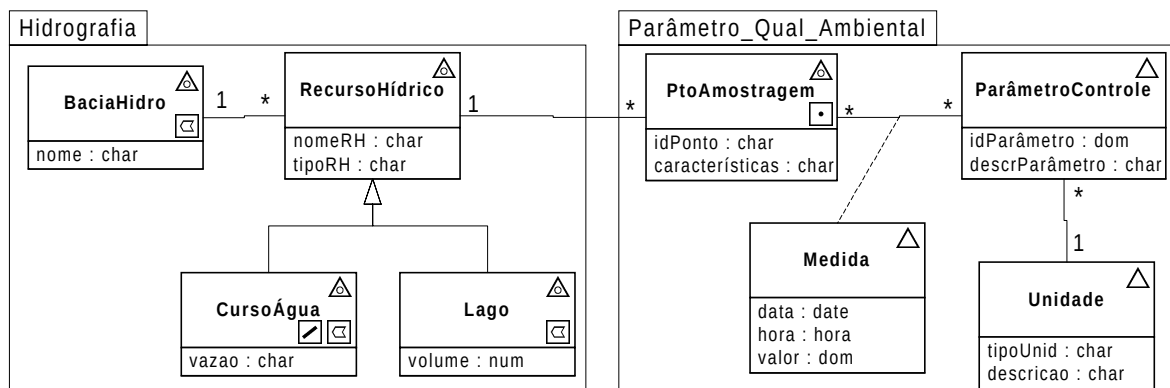


FIGURA 5 – Exemplo de uso do padrão “Parâmetros de Controle Ambiental”

3.2 Padrão: Campo Geográfico Categórico

Problema: Como modelar campos categóricos, isto é, campos geográficos que possuem para cada região, dentro da área de estudos, um valor associado a uma categoria?

Contexto: Campos geográficos são fenômenos geográficos percebidos na visão de campo, em contraposição a objetos geográficos que são fenômenos geográficos percebidos na visão de objetos.

Forças:

- Um campo geográfico é modelado como uma função, cujo domínio corresponde ao conjunto de pontos ou áreas dentro da região geográfica de interesse, e a imagem corresponde a um conjunto de valores representativos da presença do fenômeno geográfico naquela região.
- Um campo categórico só pode ter seus pontos associados a valores definidos em uma categoria (ex.: tipos de solos).

Solução: A Figura 6 mostra o diagrama de classes que compõe o padrão. A classe *CampoCategórico*, subclasse de *CampoGeográfico* ($\triangle$), tem seu componente espacial abstraído como polígonos adjacentes ou grade de células. Cada polígono/célula possui um único valor que corresponde ao tipo do campo categórico que ocorre em toda a região dentro do polígono/célula. Os atributos do campo categórico são modelados em uma subclasse de *ObjetoNãoGeográfico* ($\triangle$). Neste caso, ao invés de especificar cardinalidades para a associação, é utilizado o estereótipo $\llfunção\gg$, para enfatizar este tipo especial de associação. Na realidade, a associação que ocorre é entre uma instância de polígono/célula do campo categórico com uma instância de *TipoCampoCategórico*.

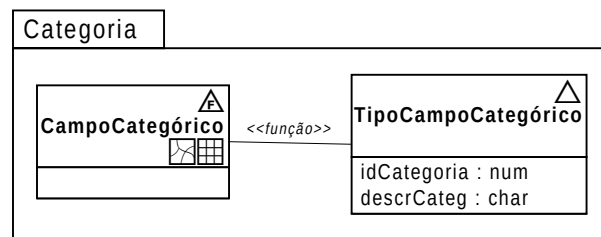


FIGURA 6 – Diagrama de classes do padrão “Campo Geográfico Categórico”

Exemplos: Uso do Solo e Geomorfologia

O uso do padrão de análise *Campo Geográfico Categórico* foi identificado nos sistemas SIGERCO-RS [LIS 96] e SIGPROGB [GAR 98], para os temas *Uso_do_Solo* e *Geomorfologia*. O tema *Geomorfologia* também foi identificado no Sistema ZEE/RJ - Zoneamento Econômico-Ecológico do Estado do Rio de Janeiro [SEM 99]. A Figura 7 ilustra o padrão de análise aplicado na modelagem do sistema SIGPROGB.

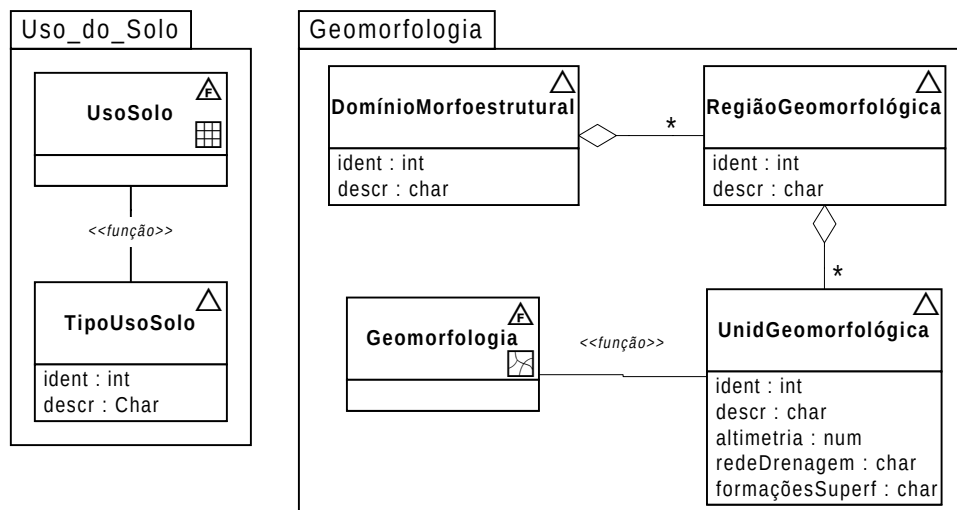


FIGURA 7 – Exemplo de uso do padrão “Campo Geográfico Categórico”

3.3 Padrão: Campo Geográfico Numérico

Problema: Como modelar campos numéricos, isto é, campos geográficos nos quais, cada região dentro da área de estudos está associada a um valor numérico não categórico?

Contexto: Campos geográficos são fenômenos geográficos percebidos na visão de campo, em contraposição a objetos geográficos, que são fenômenos geográficos percebidos na visão de objetos.

Forças:

- Um campo geográfico numérico é modelado como uma função, cujo domínio corresponde ao conjunto de pontos ou áreas dentro da região geográfica de interesse, e a imagem corresponde a um conjunto de valores numéricos.

Solução: A Figura 8 mostra o diagrama de classes que compõe o padrão. A classe *CampoNumérico*, subclasse da classe *CampoGeográfico*, representado pelo estereótipo Δ , tem seu componente espacial abstraído de diferentes formas alternativas como, por exemplo, Pontos Irregulares, Grade de Pontos, Isolinhas e TIN. Normalmente, um campo numérico é obtido a partir de levantamentos de campo (ex.: uso de GPS) com base em um ou mais desses modelos de representação e transformado nos demais. Operações de interpolação são freqüentemente utilizadas.

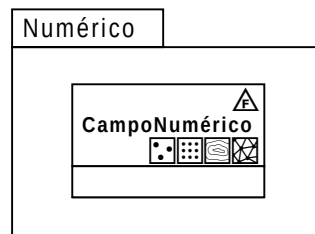


FIGURA 8 – Diagrama de classes do padrão “Campo Geográfico Numérico”

Exemplos: Clima e Relevo

O uso do padrão de análise *Campo Geográfico Numérico* foi identificado nos sistemas SIGERCO-RS [LIS 96] e SIGPROGB [GAR 98] nos temas referentes a clima e relevo (Figura 9).

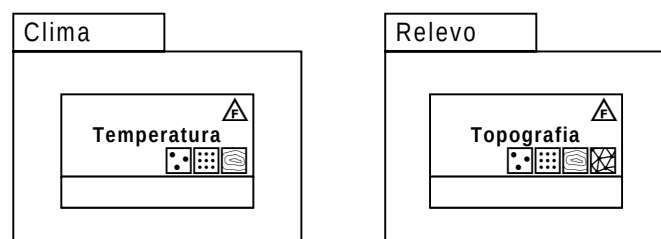


FIGURA 9 – Exemplos de uso do padrão “Campo Geográfico Numérico”

3.4 Padrão: Hidrografia

Problema: Quais os principais elementos presentes na modelagem do tema Hidrografia e como eles estão relacionados?

Contexto: Hidrografia é um tema básico, presente na maioria das aplicações da área de controle ambiental. No entanto, é possível obter-se diferentes visões dos elementos pertencentes ao tema como, por exemplo, hidrografia sob o enfoque dos diversos tipos de uso da água (ex.: balneários), hidrografia sob o enfoque de meio de transporte fluvial (ex.: rede hidrográfica), hidrografia como conjunto de recursos energéticos (ex.: usinas hidroelétricas e reservatórios).

Forças:

- Diferentes terminologias são usadas por usuários de diferentes áreas, inclusive com diferenças regionais (ex.: arroio, córrego, riacho, lago, lagoa).
- Embora não seja um recurso hídrico, o objeto *ilha* sempre está presente nos diagramas.

Solução: A Figura 10 apresenta um diagrama de classes, descrevendo os principais fenômenos geográficos relacionados com o tema. O projetista deve selecionar apenas as classes de seu interesse e acrescentar novas classes não incluídas no padrão. Também deve adequar a terminologia de acordo com a cultura local.

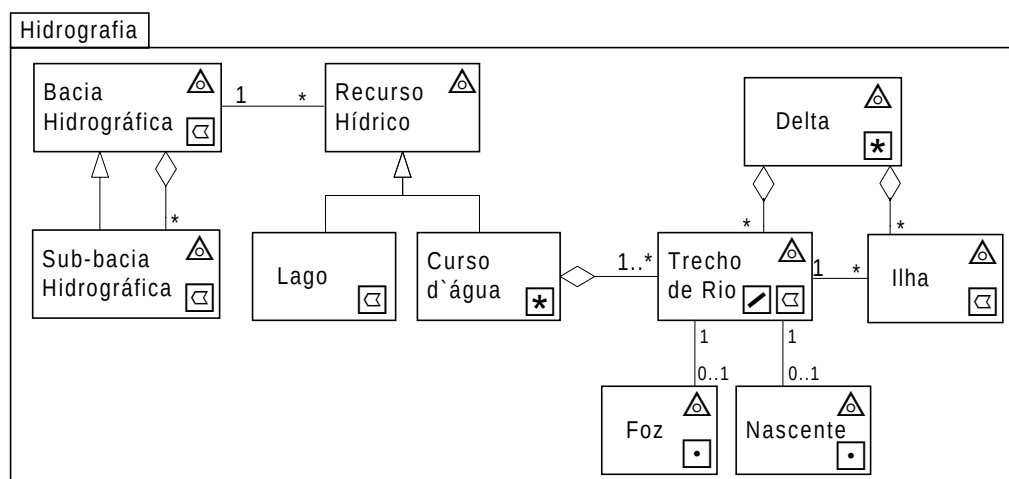


FIGURA 10 – Diagrama de classes do padrão “Hidrografia”

Exemplos: Bacia hidrográfica do rio Caí

Conforme citado anteriormente, *Hidrografia* é um tema base para diversas aplicações de SIG, principalmente, na área de controle ambiental. O uso do padrão de análise *Hidrográfica* pode ser visto no esquema de dados do sistema de informação sobre a qualidade da água da bacia hidrográfica do rio Caí [MAN 97], no sistema de informação do Programa de Limpeza dos Rios do Texas [LCR 99] e no exemplo de modelagem de dados geográficos do Sistema de Recursos Hídricos do Estado de Minas Gerais [FIG 98]. A Figura 11 apresenta o padrão de análise *Hidrografia* presente em parte do esquema de dados do sistema de informação sobre a qualidade da água da bacia hidrográfica do rio Caí.

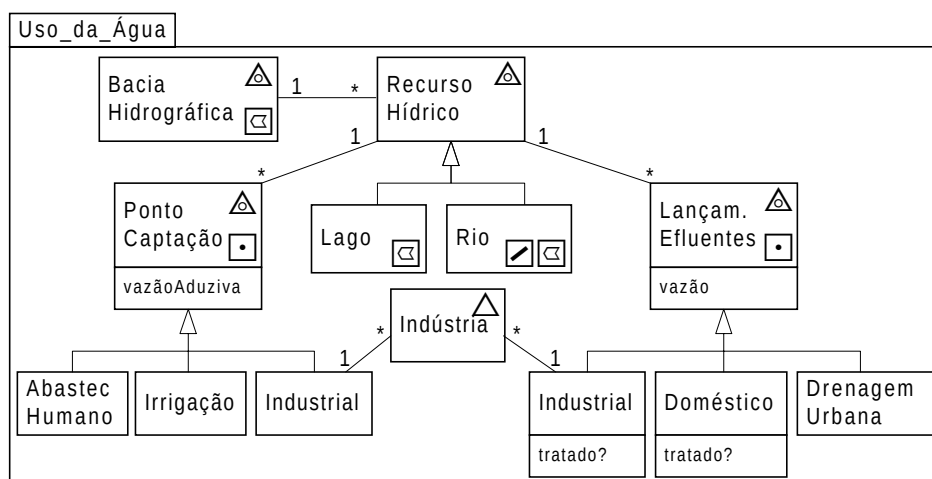


FIGURA 11 – Exemplo de uso do padrão “Hidrografia”

4 PADRÕES DE ANÁLISE EM APLICAÇÕES URBANAS

Nesta seção são descritos padrões de análise identificados através do estudo da modelagem de diversos bancos de dados para aplicações urbanas.

4.1 Padrão: Malha Viária Urbana

Problema: Como modelar uma malha viária urbana?

Contexto: No Brasil, praticamente todas as cidades apresentam um mesmo padrão de organização, no qual são estruturadas com base em suas vias de locomoção (ex.: ruas, avenidas, travessas). O conjunto de trechos de vias e seus cruzamentos formam uma rede viária urbana.

Forças:

- Cada via de locomoção, denominada logradouro, deve possuir um código de identificação e um nome, além de estar dividida em diversos trechos.
- Um trecho de logradouro corresponde ao segmento de via compreendido entre duas conexões, em seqüência, deste com outros logradouros que o cruzam ou interceptam.
- O conjunto formado pelas conexões (ou pontos terminais) e pelos trechos de logradouros constituem a malha viária urbana.

Solução: A Figura 12 mostra o diagrama de classes que compõe o padrão. Para cada fenômeno geográfico, o padrão especifica apenas os atributos e operações mais genéricos, os quais devem ser estendidos e especializados para cada aplicação específica. Consequentemente, são especificadas as possíveis abstrações de seus componentes espaciais. Por exemplo, o componente espacial da classe *TrechoLogradouro* é especificado como sendo linear ou poligonal, embora para um uso específico a mesma classe pode ser especificada com representações distintas.

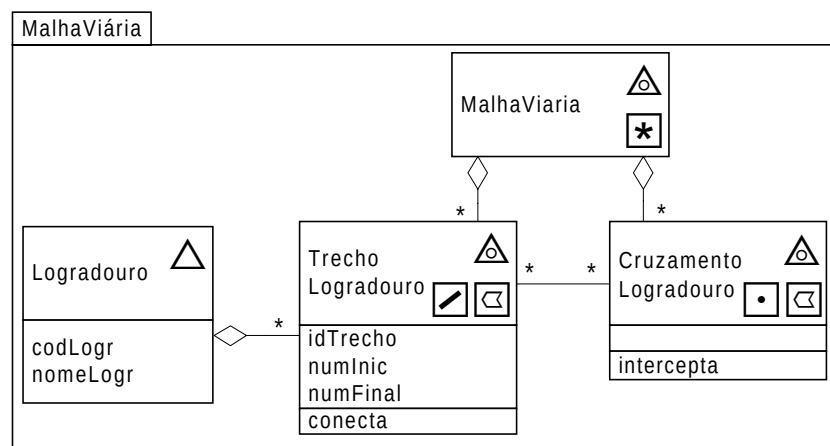


FIGURA 12 – Diagrama de classes do padrão “Malha Viária Urbana”

Padrões de análise podem ser utilizados em conjunto para solucionar um problema mais complexo. Além disso, dois ou mais padrões podem ser transformados em um novo padrão mais completo. Os padrões de análise *Malha Viária Urbana* e *Rede de Circulação Viária Urbana* (Seção 4.2) estão exemplificados, de forma conjunta, na Figura 14.

4.2 Padrão: Rede de Circulação Viária Urbana

Problema: Quais os elementos de uma rede de circulação viária urbana?

Contexto: A circulação de veículos em uma cidade é realizada sobre a malha viária urbana (ver padrão de análise *Malha Viária Urbana*). Circulação viária é uma rede que fornece o sentido do tráfego, enquanto a malha viária fornece a estrutura de vias. Algumas vias de locomoção possuem sentido único e outras sentido duplo. Cada trecho deve conter o sentido permitido de tráfego. Cada trecho é classificado de acordo com sua importância para o sistema viário como, por exemplo, se é uma via coletora, de ligação regional ou uma via local.

Forças:

- Cada trecho de circulação, que pode compreender vários trechos de logradouro, possui informações sobre o sentido permitido para tráfego de veículos.
- Alguns trechos de logradouros não permitem circulação de veículos. (Ex.: ruas de pedestres)

Solução: A Figura 13 ilustra o diagrama de classes que compõe o padrão.

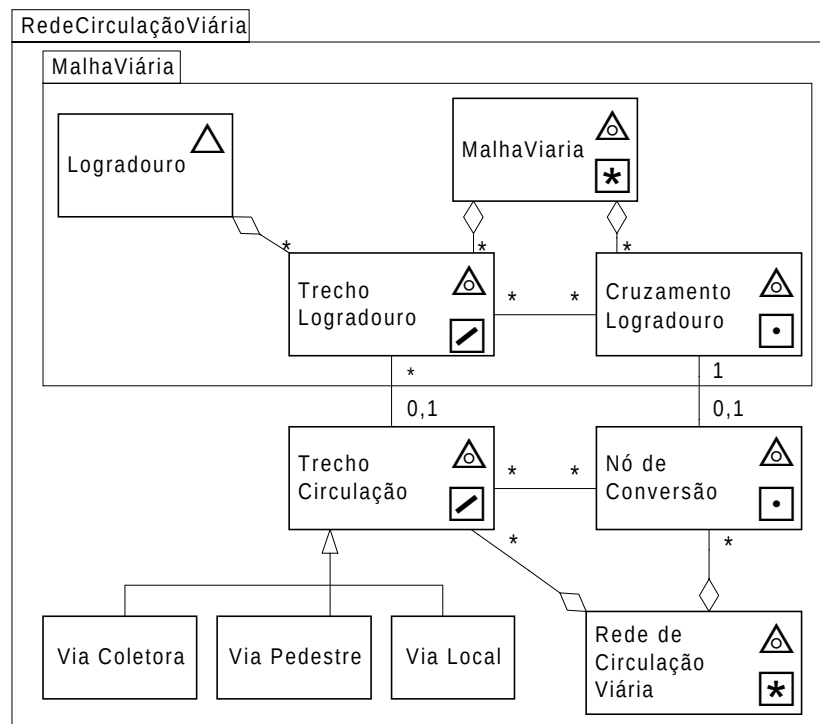


FIGURA 13 – Diagrama de classes do padrão “Rede de Circulação Viária Urbana”

Exemplo: Itinerário de Ônibus

O Sistema de itinerário de ônibus necessita da existência da circulação viária que, por sua vez, está sobre a malha viária, embora a malha viária possa também seja usada para outros fins. A Figura 14, extraída do esquema conceitual do banco de dados do sistema de transporte urbano da cidade de Belo Horizonte [BOR 97], ilustra o uso do padrão *Rede de Circulação Viária Urbana*. Observe que toda a modelagem referente aos temas *Malha Viária* e *Rede de Circulação Viária Urbana* foi reutilizada. O projetista necessita modelar somente a parte do sistema relativa à sua aplicação, ou seja, o sistema de transporte. Outros exemplos de possíveis uso deste padrão de análise incluem os

sistemas de roteamento de veículos para atendimento de ambulâncias, policiamento, corpo de bombeiros e roteamento de entrega de mercadorias.

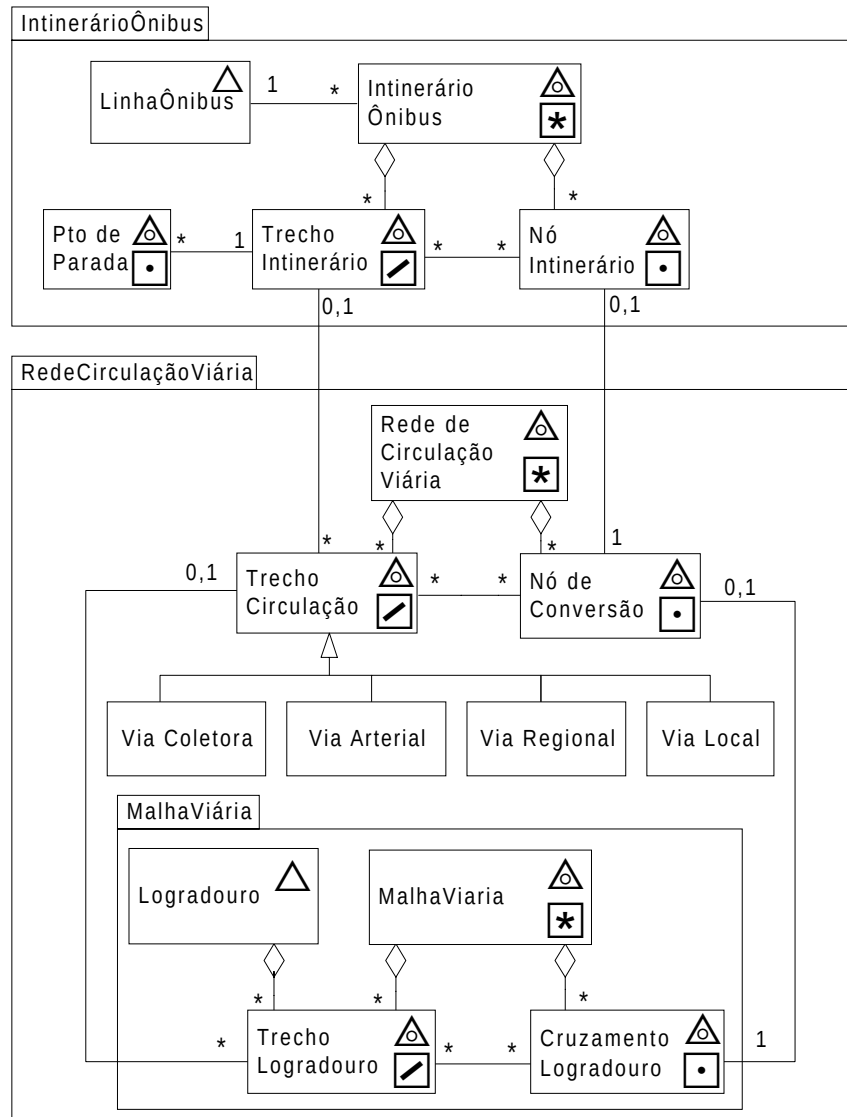


FIGURA 14 – Exemplo de uso dos padrões “Malha Viária Urbana” e “Rede de Circulação Viária Urbana”

4.3 Padrão: Cadastro de Lotes Urbanos

Problema: Como estruturar os dados de um sistema de cadastramento de lotes urbanos?

Contexto: Um dos problemas encontrados na maioria dos municípios brasileiros é a necessidade de realização do cadastramento de lotes, principalmente, para fins de tributação. Este é um dos melhores exemplos em que, embora não seja possível reutilizar dados georreferenciados de outro município, com exceção de dados regionais, é muito provável que o projeto do banco de dados possa ser reutilizado em grande parte. Isso ocorre devido à semelhança de legislação entre municípios de mesmo porte.

Forças:

- Todo lote possui pelo menos um dono, isto é, um mesmo lote pode estar registrado como pertencente a mais de um proprietário.
- Tanto pessoas físicas como pessoas jurídicas podem possuir lotes.

- Um lote pode estar edificado ou não.
- Um mesmo lote pode conter diversas edificações (ex.: condomínios com vários edifícios).
- Cada edificação pode conter diversas unidades (ex.: apartamentos, lojas), as quais estão registradas como pertencentes a diferentes proprietários.

Solução: A Figura 15 mostra o diagrama de classes que compõe o padrão.

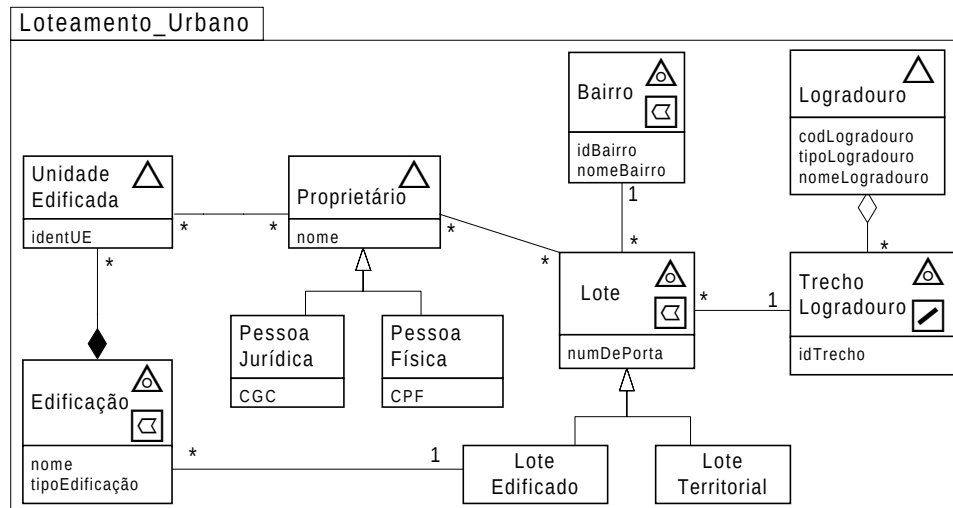


FIGURA 15 – Diagrama de classes do padrão “Cadastro de Lotes Urbanos”

Exemplo: Cadastro urbano para tributação

Sendo o loteamento urbano a base para qualquer cadastro do imposto predial e territorial urbano, o uso do padrão pode ser visto na Figura 16, a qual ilustra um sistema de cadastro urbano para fins de tributação do Imposto Territorial Urbano (IPTU). Embora se tenha conhecimento de modelagens semelhantes, este é um exemplo hipotético, no qual foram incluídos alguns atributos com base em suposições dos autores.

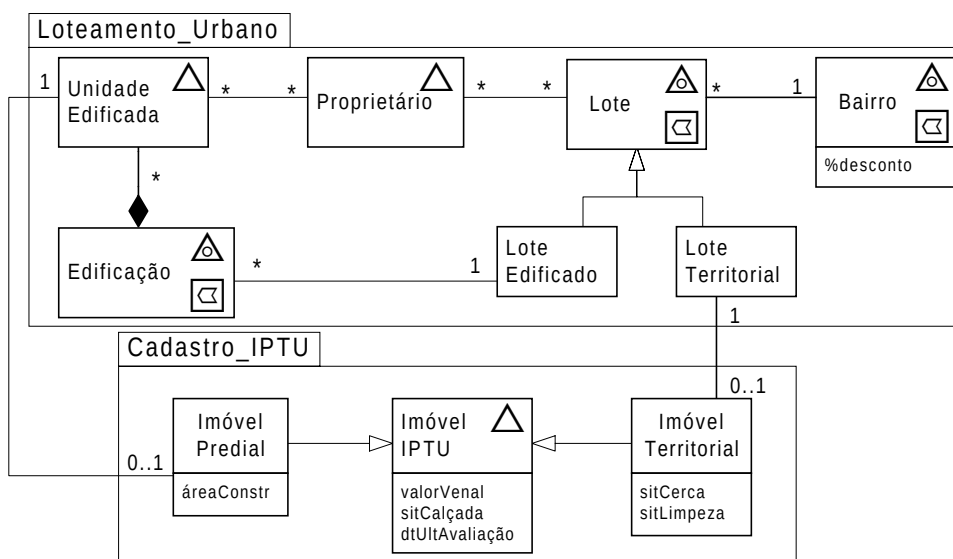


FIGURA 16 – Exemplos de uso do padrão “Cadastro de Lotes Urbanos”

5 CONCLUSÕES

Como se pode observar a partir dos diversos padrões de análise apresentados neste artigo, os padrões de análise não são soluções completas. Padrões descrevem orientações, projetos iniciais e propostas de solução para problemas recorrentes. Padrões de análise necessitam ser moldados em cada caso específico de reuso.

A abordagem de padrões de análise apresenta grande potencial para melhorar a qualidade das aplicações de SIG, bem como para reduzir o tempo e, consequentemente, os custos das etapas de análise de requisitos e modelagem conceitual do banco de dados.

Entretanto, para o sucesso desta abordagem é necessário se criar a cultura da cooperação entre os desenvolvedores de sistema. Por exemplo, muitos usuários reutilizam dados georreferenciados obtidos de terceiros, mas não disponibilizam seus próprios dados. Reutilizar uma boa solução documentada por outro projetista é uma idéia muito atraente, mas é necessário que todos contribuam com a abordagem de reutilização, documentando na forma de padrões de análise suas boas soluções.

Um padrão de análise não necessita apresentar uma solução original. Pelo contrário, padrões devem documentar soluções já testadas e validadas, pois são soluções para problemas recorrentes. Soluções para problemas únicos não necessitam ser documentadas, pois provavelmente não necessitarão ser reutilizadas. Assim, a evolução da documentação de um padrão, a partir da contribuição de outros projetistas que o tenham reutilizado, deveria ser decorrência natural de seu uso. Os padrões de análise descritos neste artigo podem e devem ser melhorados. Contribuições, comentários e críticas são bem vindos.

Como continuidade deste trabalho, está prevista a investigação de alternativas para disponibilização dos padrões de análise (para banco de dados geográficos) existentes e, também, o estudo de aplicações de SIG em diferentes domínios (ex.: redes de infra-estrutura, aplicações de marketing), com o objetivo de captura e documentação de novos padrões.

AGRADECIMENTOS

Os autores agradecem às pessoas que disponibilizaram exemplos de modelagem conceitual de bancos de dados geográficos, os quais foram utilizados na captura ou documentação dos padrões descritos neste artigo. Agradecimento especial segue para Karla Borges (Prodabel-MG), Maria Luíza Campos (NCE/UFRJ), Íria M. Garaffa (Ecoplan/Pró-Guaíba) e Eliseu Weber (Ecologia/UFRGS).

4.5 Artigo X: Modelagem Conceitual de Banco de Dados Geográficos: o estudo de caso do Projeto PADCT/CIAMB

Jugurta Lisboa Filho, Cirano Iochpe, Heinrich Hasenack* e Eliseu J. Weber*

*UFRGS - Centro de Ecologia/Centro de Recursos Idrisi

In: UFRGS/CENTRO DE ECOLOGIA. **Energia e Meio Ambiente: a questão do carvão no Rio Grande do Sul**. Porto Alegre: Editora da Universidade, 1999. No prelo.

1 INTRODUÇÃO

Projetar o banco de dados é uma das tarefas mais importantes no desenvolvimento de um sistema de informação. O projeto do banco de dados requer o uso de diferentes instrumentos, uma vez que as atividades necessárias a sua elaboração variam de acordo com a complexidade do sistema, com o tipo de pessoal envolvido, o sistema de gerenciamento de banco de dados (SGBD) utilizado, etc. Desta forma, o desenvolvimento de sistemas de banco de dados deve estar baseado em uma metodologia eficaz, a partir da qual são empregados instrumentos específicos de apoio às diferentes etapas do projeto.

A abordagem clássica de projeto de banco de dados consiste em dividir o processo em três etapas: *projeto conceitual*; *projeto lógico*; e *projeto físico* [ELM 94]. Na fase do projeto conceitual é elaborado o esquema conceitual do banco de dados, com base em modelos de dados que fornecem construtores de abstração de alto nível para descrever os requisitos de dados da aplicação. Para facilitar a comunicação entre usuários e projetistas são utilizadas linguagens bastante simples, como o modelo E-R [CHE 76]. No projeto conceitual não são considerados aspectos sobre o sistema de computação (software/hardware) utilizado.

Na fase de projeto lógico é elaborado o esquema lógico do banco de dados com base no modelo de SGBD que será utilizado. O esquema lógico é gerado, aplicando-se regras de transformação (mapeamento) dos construtores utilizados no esquema conceitual em elementos de representação de dados de um dos modelos de banco de dados implementados pelos SGBD disponíveis no mercado (ex.: relacional, hierárquico, objeto-relacional).

No projeto físico, define-se os aspectos de implementação física do banco de dados como, por exemplo, estruturas de armazenamento, caminhos de acesso, particionamento e agrupamento de dados. Estes fatores estão, diretamente, relacionados a um SGBD específico e permitem, ao projetista, planejar aspectos ligados à eficiência do sistema de banco de dados.

Este capítulo enfoca o processo de elaboração do esquema conceitual de dados do Projeto Energia e Meio Ambiente: a questão do carvão no Rio Grande do Sul (PADCT/CIAMB).

O sistema de informação geográfica do Projeto PADCT/CIAMB foi desenvolvido para servir de apoio aos diversos sub-projetos realizados. Estes sub-projetos foram executados por um grupo diversificado de profissionais e não houve a análise inicial dos

requisitos do sistema. A modelagem conceitual do Projeto PADCT/CIAMB foi realizada posteriormente ao desenvolvimento do mesmo em um processo conhecido como engenharia reversa. Desta forma, a elaboração do esquema conceitual teve como objetivo servir de instrumento para a integração e a documentação dos dados produzidos durante a execução dos diversos sub-projetos.

2 MODELAGEM CONCEITUAL DE BANCO DE DADOS GEOGRÁFICOS

Um modelo de dados fornece uma base formal (notacional e semântica) para ferramentas e técnicas usadas para suportar a modelagem de dados. Modelagem de dados é o processo de abstração onde somente os elementos essenciais da realidade observada são enfatizados, descartando-se os elementos não essenciais. O processo de modelagem conceitual de banco de dados compreende a descrição e definição dos possíveis conteúdos dos dados, além de estruturas e de regras a eles aplicáveis (Figura 1).

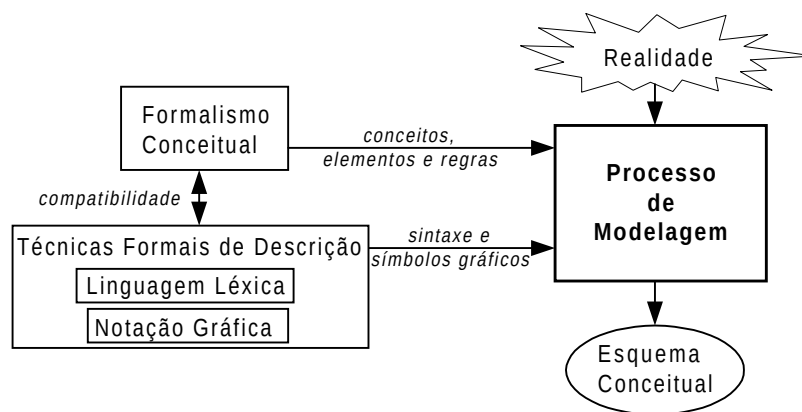


FIGURA 1 - Processo de modelagem conceitual

A modelagem conceitual é sempre feita com base em algum formalismo conceitual (ex.: Entidade-Relacionamento, Orientação a Objetos), independentemente do nível de abstração empregado [CEN 96]. O resultado do processo de modelagem, denominado esquema conceitual, é apresentado através de uma linguagem formal de descrição que possui uma sintaxe e uma notação gráfica. Para cada formalismo conceitual, existem diversas linguagens de descrição de esquema que são compatíveis com o formalismo.

O formalismo fornece um conjunto de conceitos, elementos e regras que são usados no processo de modelagem da realidade, enquanto que a linguagem de descrição fornece uma gramática para a apresentação do esquema conceitual resultante da modelagem. A linguagem léxica possibilita o processamento computacional do esquema, enquanto a notação gráfica é mais adequada para facilitar o entendimento e a comunicação entre seres humanos (ex.: usuários e projetistas).

Existem diversos modelos conceituais de dados propostos na literatura especificamente para aplicações de sistemas de informação geográfica (SIG) como, por exemplo, Modul-R [BED 96], GeoOOA [KOS 97], Geo-ER [HAD 97], GMOD [PIR 97], Geo-OMT [BOR 97] e MADS [PAR 98]. A maioria deles é baseada nos formalismos Entidade-Relacionamento e da Orientação a Objetos. No entanto, os modelos se diferem muito com relação à notação gráfica e quanto à linguagem léxica (quando definida).

A modelagem conceitual apresenta diversas vantagens para a modelagem de aplicações geográficas. Primeiro, por facilitar a execução do projeto lógico, o qual necessita atender as particularidades de um SIG específico. Os usuários podem expressar seus conhecimentos sobre a aplicação usando conceitos que estão mais próximos a eles sem a necessidade de utilizar jargões computacionais. Como a modelagem conceitual independe do software no qual o sistema é implementado, o projeto resultante se mantém válido caso ocorram mudanças de tecnologia. Neste caso, apenas a transformação entre os esquemas conceitual e lógico é afetada. No caso da tecnologia de SIG, isso se torna um fator muito importante, uma vez que grandes investimentos são preservados e há uma redução de custos e aumento das chances de sucesso em caso de mudança para tecnologias mais modernas. Por último, a modelagem conceitual facilita a troca de informações entre parceiros de diferentes organizações, uma vez que aumenta a capacidade de entendimento da semântica da informação, facilitando o uso correto da mesma.

2.1 Modelagem orientada a objetos

Um banco de dados pode ser visto como um modelo abstrato de uma porção da realidade, uma vez que seus dados representam um subconjunto de elementos pertencentes a esta realidade [BED 96]. Abstrair uma porção da realidade para projetar um banco de dados implica em selecionar os elementos (objetos) da realidade que são significativos (dentro do objetivo pretendido), identificar como eles podem ser estruturados e os relacionamentos entre eles.

O processo de modelagem conceitual é realizado, utilizando-se mecanismos de abstração, ou seja, construtores básicos definidos pelo formalismo. Uma descrição completa sobre modelagem de objetos pode ser obtida em [FUR 98]. Os principais mecanismos de abstração presentes no formalismo da orientação a objetos estão resumidos a seguir.

- *Classificação* - processo de abstração através do qual os objetos que representam elementos semelhantes têm suas propriedades descritas em uma única *classe*. Estas propriedades podem ser estáticas (estruturais) ou dinâmicas (comportamentais). Todo objeto é *instância* de uma classe. Todas as instâncias de uma classe possuem as mesmas propriedades estáticas, definidas como atributos da classe, e as mesmas propriedades dinâmicas, definidas como operações da classe.
- *Generalização e especialização* - classes que descrevem objetos semelhantes podem ser generalizadas em uma nova classe de mais alto nível. *Generalização* é o processo de definir classes mais genéricas a partir de classes com características semelhantes. *Especialização* é o processo inverso no qual classes mais específicas são detalhadas a partir de classes genéricas, adicionando-se novas propriedades na forma de atributos e/ou operações. Este tipo de abstração estabelece uma hierarquia na qual classes especializadas (denominadas subclasses) herdam as propriedades das classes genéricas (denominadas superclasses).
- *Associação* - tipo de abstração através do qual os relacionamentos entre objetos são especificados. A multiplicidade (ou cardinalidade) de uma associação indica quantos objetos podem estar relacionados através dessa associação. Por exemplo, se uma classe *Município* está associada com uma classe *Estado* com multiplicidade (N:1), significa que toda instância de *Município* está associada

a, no máximo, uma instância de *Estado* e cada instância de *Estado* pode estar associada a N instâncias de *Município*.

- **Agregação** - tipo especial de associação que descreve relacionamentos do tipo “é parte de”, onde um objeto complexo é definido como uma agregação de suas partes (ou objetos componentes). Uma variação deste tipo de abstração é a *composição*. Um objeto pertencente a um relacionamento do tipo composição só pode pertencer a um único objeto composto e tem sua existência dependente da existência do objeto composto.

Existem diversas linguagens para especificação de diagramas de classes segundo o formalismo da orientação a objetos. Entre as mais conhecidas pode-se citar OOA [COA 91], OMT [RUM 91] e UML [BOO 98]. Neste trabalho optou-se por utilizar a notação gráfica do diagrama de classes UML - *Unified Modeling Language* [BOO 98], seguindo a tendência das áreas de Engenharia de Software e Banco de Dados, reforçado pelo surgimento de ferramentas CASE para projeto de aplicações de SIG como, por exemplo, Perceptory [BED 99] e REGIS [ISO 99].

A Figura 2 apresenta os principais símbolos gráficos utilizados no diagrama de classes UML. Além dos construtores descritos anteriormente, a figura mostra o elemento *pacote*. Um pacote constitui-se de um conjunto de elementos do modelo UML podendo ser de qualquer tipo como, por exemplo, classes, associações e outros pacotes [BOO 98]. Pacotes são usados para dividir um esquema de classes em sub-esquemas de forma a tornar seu entendimento mais simples.

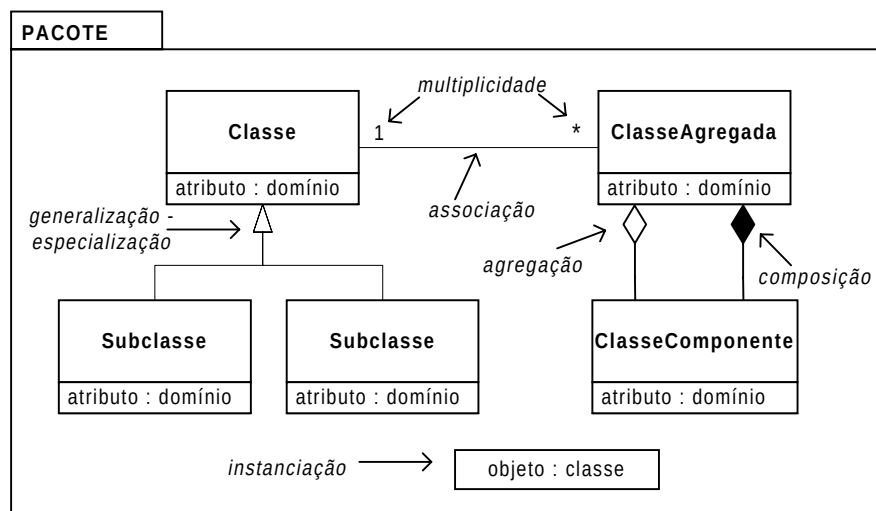


FIGURA 2 - Notação gráfica do diagrama de classes UML (resumido)

A seguir são descritos os instrumentos de reutilização utilizados no desenvolvimento da modelagem conceitual do Projeto PADCT/CIAMB.

2.2 O framework GeoFrame

GeoFrame é um *framework* conceitual que fornece um diagrama de classes básicas para auxiliar o projetista tanto na modelagem conceitual de dados geográficos como, também, na especificação de padrões de análise em bancos de dados geográficos [LIS 99].

Souza [SOU 98] define um *framework* como “um projeto genérico em um domínio que pode ser adaptado a aplicações específicas, servindo como um molde para a construção de aplicações”. Esta definição fornece uma visão bem mais abrangente sobre a potencialidade de um *framework* do que as definições apresentadas por autores

mais ligados à programação orientada a objetos. Por exemplo, Johnson [JOH 92] define um *framework* como sendo “um projeto reutilizável de um programa, ou parte de um programa, expresso como um conjunto de classes”. GeoFrame é um *framework* definido sob o enfoque mais genérico, onde o mesmo expressa a idéia de um projeto conceitual parcial para uma família de aplicações.

O GeoFrame foi definido de acordo com as regras do formalismo da orientação a objetos, utilizando a notação gráfica do diagrama de classes da linguagem UML [BOO 98]. A Figura 3 mostra o diagrama de classes do GeoFrame. A seguir, estas classes são descritas em maior detalhe.

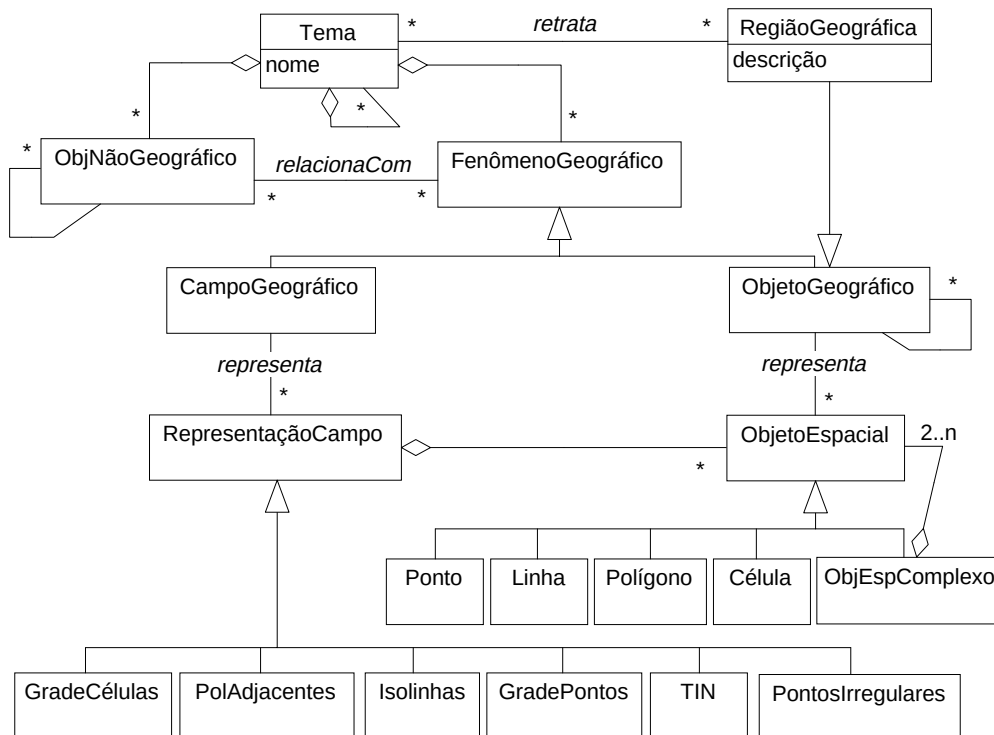


FIGURA 3 - Diagrama de Classes do GeoFrame

2.2.1 Tema e RegiãoGeográfica

As classes **TEMA** e **REGIÃO GEOGRÁFICA** formam a base de qualquer aplicação geográfica. Cada aplicação geográfica tem como objetivo o gerenciamento e a manipulação de um conjunto de dados para uma determinada região de interesse, constituindo o banco de dados geográfico.

Para cada região geográfica pode-se especificar uma coleção de temas. O agrupamento de classes que descrevem os fenômenos geográficos em temas funciona como um mecanismo para redução da complexidade em grandes esquemas de dados. O uso de temas permite, ao projetista, dividir o esquema de dados em subesquemas coesos, nos quais são agrupadas classes que estão fortemente relacionadas entre si. Conjuntos de temas afins podem ser agrupados em um tema mais genérico, formando uma hierarquia de temas.

2.2.2 ObjetoNãoGeográfico e FenômenoGeográfico

Em um banco de dados geográficos podem existir, além dos dados referentes a fenômenos georreferenciados, objetos convencionais presentes em qualquer sistema de

informação. Objetos que não possuem referência a uma posição geográfica dão origem a subclasses da classe OBJETO NÃO GEOGRÁFICO.

A classe FENÔMENO GEOGRÁFICO generaliza qualquer fenômeno cuja localização em relação à superfície terrestre seja considerada. Por exemplo, um distrito municipal é uma instância de FENÔMENO GEOGRÁFICO, se seus atributos espaciais estiverem representados no banco de dados. Caso contrário, o distrito é definido como um objeto não geográfico. Fenômenos geográficos e objetos não geográficos estão, muitas vezes, relacionados entre si (associação *relacionaCom* - Figura 3).

2.2.3 Campo Geográfico e Objeto Geográfico

Fenômenos geográficos são percebidos, na realidade geográfica, segundo as visões dicotômicas de campo e de objeto [GOO 92]. Essas duas visões acarretam diferentes maneiras de modelagem dos fenômenos geográficos. As classes CAMPO GEOGRÁFICO e OBJETO GEOGRÁFICO especializam a classe FENÔMENO GEOGRÁFICO, permitindo ao projetista especificar, de forma distinta porém integrada, os campos e os objetos geográficos respectivamente.

A classe OBJETO GEOGRÁFICO é uma generalização de todas as classes do domínio da aplicação que são percebidas na visão de objetos. Neste caso estão incluídas aquelas classes que representam fenômenos geográficos que podem ser individualizados, ou seja, que possuem identidade própria e suas características podem ser descritas através de atributos (ex.: *Mina, Rio, Rodovia, Município*).

A classe CAMPO GEOGRÁFICO generaliza os fenômenos que se enquadram na visão de campo. Campos geográficos são modelados como funções sobre variáveis. Alguns campos referem-se a variáveis distribuídas sobre a superfície, de forma contínua (ex.: *Altimetria, Temperatura e Cobertura do Solo*), enquanto outros referem-se a variáveis distribuídas de forma discreta (ex.: *População e Ocorrências Epidemiológicas*) [PIR 97].

2.2.4 Objeto Espacial

Em um SIG, a implementação da representação espacial e dos relacionamentos espaciais de um conjunto de objetos geográficos é feita com base em estruturas de dados espaciais. A escolha da melhor estrutura de dados para implementar a representação espacial de cada fenômeno geográfico é uma tarefa posterior ao projeto conceitual, não devendo ser considerada ao longo do mesmo.

Um dos princípios fundamentais da modelagem conceitual é que um esquema conceitual deve conter apenas os elementos do domínio, desconsiderando os aspectos de implementação [PAR 98]. O objetivo de se incluir, no esquema conceitual, informações sobre objetos espaciais relacionados a fenômenos geográficos é o de permitir a realização da modelagem (abstração) do componente espacial de cada fenômeno.

Portanto, quando se considera pontos, polígonos, isolinhas ou grade de células no GeoFrame, estão sendo tratadas as formas de abstração do componente espacial dos fenômenos geográficos, mas não as formas com que eles serão armazenados no banco de dados, embora na maioria dos SIG atuais existam estruturas de dados muito semelhantes para o armazenamento destes construtores abstratos.

Para efeito de modelagem, o que importa é determinar, por exemplo, se o componente espacial de uma estação meteorológica terá representação pontual ou se um rio terá representação linear. Não é necessário, no entanto, considerar que a representação do rio será armazenada por meio de um arco em uma estrutura vetorial com topologia.

Alguns fenômenos geográficos podem apresentar dimensão espacial complexa, ou seja, composta por outros objetos espaciais (ex.: um arquipélago). A classe `OBJETOESPACIAL` generaliza as classes necessárias para a especificação da representação do componente espacial dos fenômenos geográficos percebidos na visão de objetos. São elas: `PONTO`, `LINHA`, `POLÍGONO`, `CÉLULA` e `OBJESPACIALCOMPLEXO`.

2.2.5 RepresentaçãoCampo

Os aspectos espaciais de um campo geográfico são abstraídos de forma diferente dos aspectos espaciais de um objeto geográfico. Chrisman [CHR 97] descreve diversos modelos geográficos através dos quais pode-se abstrair o componente espacial da informação geográfica. Os modelos relacionados por Chrisman, os quais são adequados à modelagem de fenômenos na visão de campo, podem ser resumidos nos seis modelos espaciais descritos por Goodchild [GOO 92]. São eles: grade de células; polígonos adjacentes; isolinhas; grade de pontos; rede triangular irregular; e pontos amostrados irregularmente. No `GeoFrame`, esses seis modelos correspondem às subclasses da classe `REPRESENTAÇÃOCAMPO`.

Em um SIG, esses modelos serão, posteriormente, implementados através dos modelos de representação matricial e vetorial. Cada um dos seis modelos de representação de campos geográficos pode ser implementado tanto no modelo matricial como no modelo vetorial, embora alguns mapeamentos sejam mais naturais. Por exemplo, um campo geográfico cujo componente tenha sido abstraído por meio de uma grade de pontos é mapeado para o modelo de representação matricial.

Um mesmo campo geográfico pode ter seu componente espacial abstraído de diferentes formas, ou seja, através de mais de um desses modelos. Por exemplo, o campo *Temperatura* pode ser abstraído por meio de um conjunto de pontos irregularmente distribuídos ou por meio de isolinhas. Situação semelhante ocorre com os objetos geográficos cujos componentes espaciais podem ser percebidos ora por formas alternativas (ex.: municípios podem ser representados por pontos ou polígonos), ora por formas duplas (ex.: um mesmo rio pode ter um trecho representado por uma linha e outro trecho representado por um polígono), dependendo de aspectos como a escala com a qual se pretende capturar a forma espacial de cada fenômeno.

2.3 Criando esquemas de dados a partir do GeoFrame

Um esquema conceitual de banco de dados geográficos pode ser elaborado a partir da especialização das classes do `GeoFrame`. A modelagem conceitual de banco de dados geográficos, usando o `GeoFrame`, é realizada segundo uma abordagem top-down composta de três etapas. Inicialmente são identificados, para cada área geográfica, os diversos temas (e sub-temas) a serem projetados. Na segunda etapa é definido um subesquema de classes para cada tema identificado. Ainda nessa etapa, é feita a especificação das associações entre classes de diferentes temas. Por último, é realizada a análise e modelagem do tipo de representação espacial de cada fenômeno geográfico identificado. A seguir, cada uma dessas fases é descrita. Ao longo desta descrição serão introduzidos alguns mecanismos de simplificação de esquemas do `GeoFrame`.

2.3.1 Diagrama de temas

Conforme mostrado na Figura 3, cada região geográfica é retratada por zero ou mais temas. No entanto, um mesmo tema também pode estar associado a mais de uma

região geográfica. Um tema é especificado como uma agregação de classes de fenômenos geográficos, de objetos não geográficos e de outros temas mais específicos.

Para aumentar a legibilidade do esquema resultante, os temas definidos pelo projetista não são modelados como subclasses da classe TEMA, mas através do construtor *Pacote*, da linguagem UML.

2.3.2 Especificando fenômenos geográficos e objetos não geográficos

Na abordagem orientada a objetos, um esquema de banco de dados geográficos é representado pelo diagrama de classes que descrevem os fenômenos geográficos, os objetos não geográficos e os possíveis relacionamentos entre eles.

Cada classe identificada no domínio da aplicação deve ser modelada como subclasse de uma das seguintes classes do GeoFrame: OBJETO NÃO GEOGRÁFICO; CAMPO GEOGRÁFICO; ou OBJETO GEOGRÁFICO.

Para evitar a sobrecarga visual do diagrama devido ao grande número de ligações, é utilizado um mecanismo de simplificação de esquemas denominado *estereótipo*. Um estereótipo estende o vocabulário UML permitindo ao projetista criar novos tipos de construtores que podem ser empregados como qualquer outro elemento da linguagem [BOO 98]. Para substituir os relacionamentos de generalização entre as classes do domínio e as classes do GeoFrame, são utilizados três estereótipos (Figura 4).

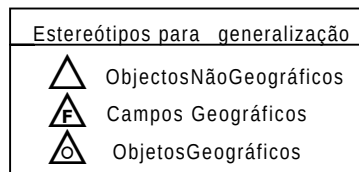


FIGURA 4 - Estereótipos para generalização

2.3.3 Especificando o componente espacial dos fenômenos geográficos

No GeoFrame todo campo e objeto geográfico pode ser representado por múltiplas instâncias das classes REPRESENTAÇÃO CAMPO e OBJETO ESPACIAL respectivamente. Um fenômeno geográfico pode ter múltiplas representações por vários motivos. Dentre eles, pode-se citar a necessidade de múltiplas escalas, usuários com diferentes visões de um mesmo fenômeno e versões temporais.

A possibilidade de se ter múltiplas representações para um mesmo fenômeno geográfico é modelada através de diferentes associações entre o fenômeno geográfico e as possíveis formas de abstração de seu componente espacial (associação *representa* na Figura 3). No GeoFrame estas variações podem ser especificadas através da combinação livre de diferentes estereótipos em uma mesma classe com documentação adicional no dicionário de dados.

Um segundo conjunto de estereótipos (Figura 5) é usado para substituir as associações que resultam da modelagem do componente espacial dos fenômenos geográficos. A semântica de cada estereótipo, neste caso, é a substituição de uma associação entre o fenômeno geográfico e sua representação espacial, além da indicação da forma geométrica de tal representação.

ObjetoEspacial	RepresentaçãoCampo	
 Ponto	 GradeCélulas	 GradePontos
 Linha	 PolAdjacentes	 TIN
 Polígono	 Isolinhas	 PontosIrregulares
 ObjComplexo		

FIGURA 5 - Estereótipos para associação

2.4 Padrões de análise

O segundo instrumento de reutilização empregado na modelagem do Projeto PADCT/CIAMB é o padrão de análise. Normalmente, durante a fase de modelagem conceitual de dados de aplicações geográficas, um grande número de fenômenos geográficos (e de relacionamentos entre eles) é identificado. Segundo Gordillo e Balaguer [GOR 98], um projetista experiente desenvolve seus modelos a partir do conhecimento prévio de um conjunto de entidades interrelacionadas, ao invés de sempre partir da estaca zero. Estes projetistas reutilizam parte de trabalhos desenvolvidos anteriormente para resolver novos problemas similares. No entanto, projetistas menos experientes não podem reutilizar soluções já validadas pois, normalmente, a documentação dos sistemas ou não existe, ou é insuficiente.

Conceitos como *frameworks* e padrões (do inglês *patterns*) estão se tornando, cada vez mais, importantes instrumentos no desenvolvimento de sistemas orientados a objetos. Segundo Buschmann [BUS 96], existem três categorias de padrões: *padrões de arquitetura*, *padrões de projeto*; e *idiomas*. As duas primeiras categorias incluem os padrões relacionados com a fase de projeto do sistema, mas são aplicadas em problemas de diferentes escalas. Os padrões de projeto são mais abstratos (e menores) do que os padrões de arquitetura, enquanto que idiomas são usados em nível de linguagens de programação.

Uma quarta categoria de padrões, introduzida por Fowler [FOW 97], inclui os *padrões de análise*. Estes padrões são usados para descrever soluções empregadas durante as fases de análise de requisitos e modelagem conceitual dos dados. Padrões de análise refletem estruturas conceituais do domínio da aplicação e não soluções computacionais. Fowler define um padrão de análise como “*uma idéia que se provou útil em um contexto prático e que, provavelmente, será útil em outras situações similares*”.

O emprego de padrões de análise no projeto de aplicações geográficas facilita a modelagem dessas aplicações devido à grande interseção entre conjuntos de fenômenos geográficos que interessam às aplicações de um mesmo domínio [LIS 98]. Por exemplo, aplicações na área de controle ambiental, como é o caso do Projeto PADCT/CIAMB, manipulam fenômenos geográficos pertencentes a temas comuns da área (ex.: hidrografia, clima, vegetação). Embora existam diferenças, principalmente em relação aos atributos modelados, pois dependem do enfoque definido pelo usuário, o conjunto de classes e seus relacionamentos quase sempre é o mesmo dentro de um tema específico.

A modelagem conceitual de dados do Projeto PADCT/CIAMB baseou-se em padrões de análise identificados durante o desenvolvimento de outras aplicações geográficas na área de controle ambiental [LIS 96, LIS 97]. Além disso, possibilitou a identificação de novos candidatos a padrões de análise.

3 MODELAGEM CONCEITUAL DE DADOS DO PROJETO PADCT/CIAMB

3.1 Descrição do Projeto

O projeto Energia e Meio Ambiente: a questão do carvão no Rio Grande do Sul (PADCT/CIAMB) é um trabalho que envolveu inúmeras equipes de pesquisadores e estudantes de diferentes unidades da Universidade Federal do Rio Grande do Sul [UFR 97]. O objetivo foi o exercício de um trabalho interdisciplinar tendo como tema o estudo de uma região tradicionalmente produtora e consumidora de carvão, produto de potencial econômico limitado e grande potencial poluidor. O carvão já vem sendo explorado na região do Baixo Jacuí (Figura 6) desde a segunda metade do século passado em minas subterrâneas e há algumas décadas também a céu aberto. O comprometimento ambiental da extração e uso é bastante conhecido, embora estudado sempre parcialmente.

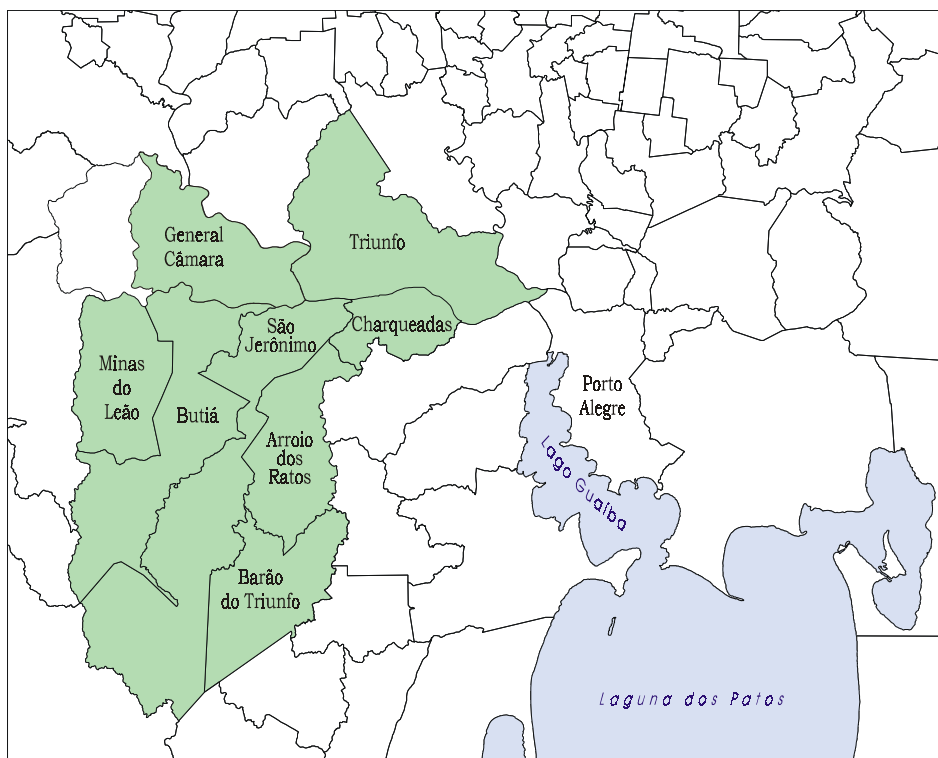


FIGURA 6 - Região Carbonífera do RS, Micro-região São Jerônimo

Para possibilitar a realização de uma análise multidisciplinar e integrativa, buscou-se investigar os aspectos sócio-econômicos e ambientais de modo a realizar um diagnóstico do potencial econômico regional e dos meios de recuperação ambiental de áreas degradadas em função da exploração ou da deposição dos rejeitos. Também a busca de alternativas tecnológicas para recuperação de áreas, para uma exploração mais racional bem como o encaminhamento político e jurídico de questões ligadas ao acompanhamento da exploração e a conscientização da população sobre as vantagens e

as limitações da exploração carbonífera (efeitos na economia e na saúde pública) foram alvo do projeto.

Esta gama variada de objetivos, ora de abrangência regional ora local, gerou um conjunto de dados amplo e complexo, devido à natureza e abrangência temporal e espacial dos dados gerados. Estes dados estão distribuídos em diversos meios digitais (SIG) e analógicos (relatórios técnicos). A integração dos diversos grupos de pesquisadores passa necessariamente pelo intercâmbio de dados, tanto para verificação quanto para a geração de novos dados derivados.

3.2 Esquema conceitual de dados do Projeto PADCT/CIAMB

A metodologia empregada na modelagem conceitual dos dados do Projeto PADCT/CIAMB seguiu uma abordagem inversa a que ocorre na maioria dos sistemas. Utilizando-se técnicas da engenharia reversa [HEU 98], partiu-se de uma análise dos dados existentes no SIG do projeto, bem como de análise em outras fontes como relatórios e entrevistas com executores do projeto, para gerar o esquema conceitual, tendo como base o GeoFrame e padrões de análise existentes.

Inicialmente, identificou-se os diversos temas para os quais existe algum tipo de dado. Tais temas, denominados de sub-temas, foram agrupados em dois temas mais genéricos: *Meio_Antrópico* e *Meio_Biótico_e_Abiótico* (Figura 7). De acordo com o GeoFrame, os temas retratam uma região geográfica. A região geográfica em questão é a região carbonífera do Baixo Jacuí, representada como uma instância da classe REGIÃO GEOGRÁFICA.

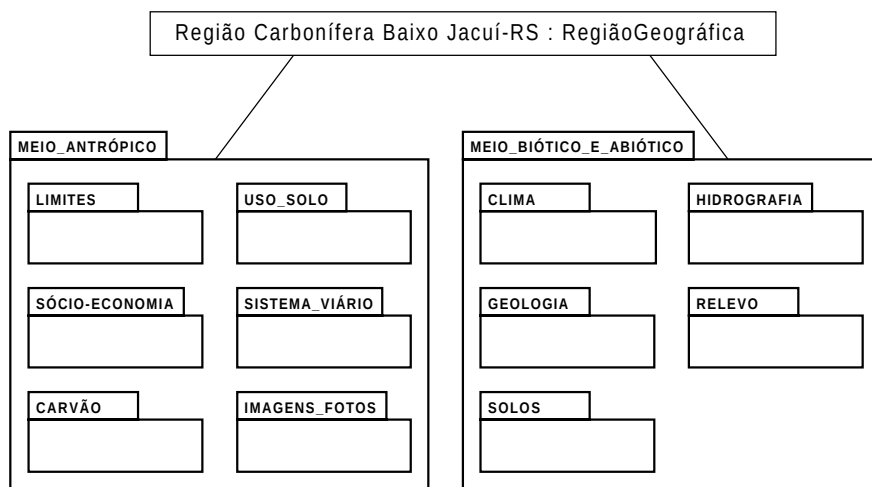


FIGURA 7 - Diagrama de temas

As Figuras 8 e 9 apresentam, respectivamente, o detalhamento (diagramas de classes) dos diversos sub-temas dos temas *Meio_Antrópico* e *Meio_Biótico_e_Abiótico*. A seguir é apresentada uma descrição sucinta das principais classes identificadas em cada tema.

3.2.1 Tema Meio_Antrópico

O tema *Meio_Antrópico* (Figura 8) agrega seis sub-temas: *Sócio-Economia*; *Uso_Solo*; *Limites*; *Ativ_Carbonífera*; *Sistema_Viário*; e *Imagens_Fotos*.

O sub-tema *Limites* engloba as classes daqueles fenômenos geográficos que são utilizados como unidades espaciais. A classe *Município*, subclasse de

OBJETOGEOGRÁFICO, é a principal unidade de mapeamento do projeto, cujos dados foram obtidos a partir de mapas analógicos em escala 1:50.000. Todos os municípios cadastrados fazem parte da única micro-região do projeto, a micro região de São Jerônimo. Um município possui diversos distritos (sem representação espacial), sendo que um deles, o distrito sede do município, possui representação espacial pontual. Dados de vários censos demográficos foram levantados junto ao IBGE - Instituto Brasileiro de Geografia e Estatística [NEV 97]. Para cada ano do censo demográfico existe uma divisão de setores censitários, modelados no classe *SetorCensitário*, subclasse de OBJGEOGRÁFICO com representação espacial do tipo polígono. A classe *Censo*, subclasse de OBJNÃOGEOGRÁFICO contém, para cada ano do censo, os dados sobre população rural e urbana.

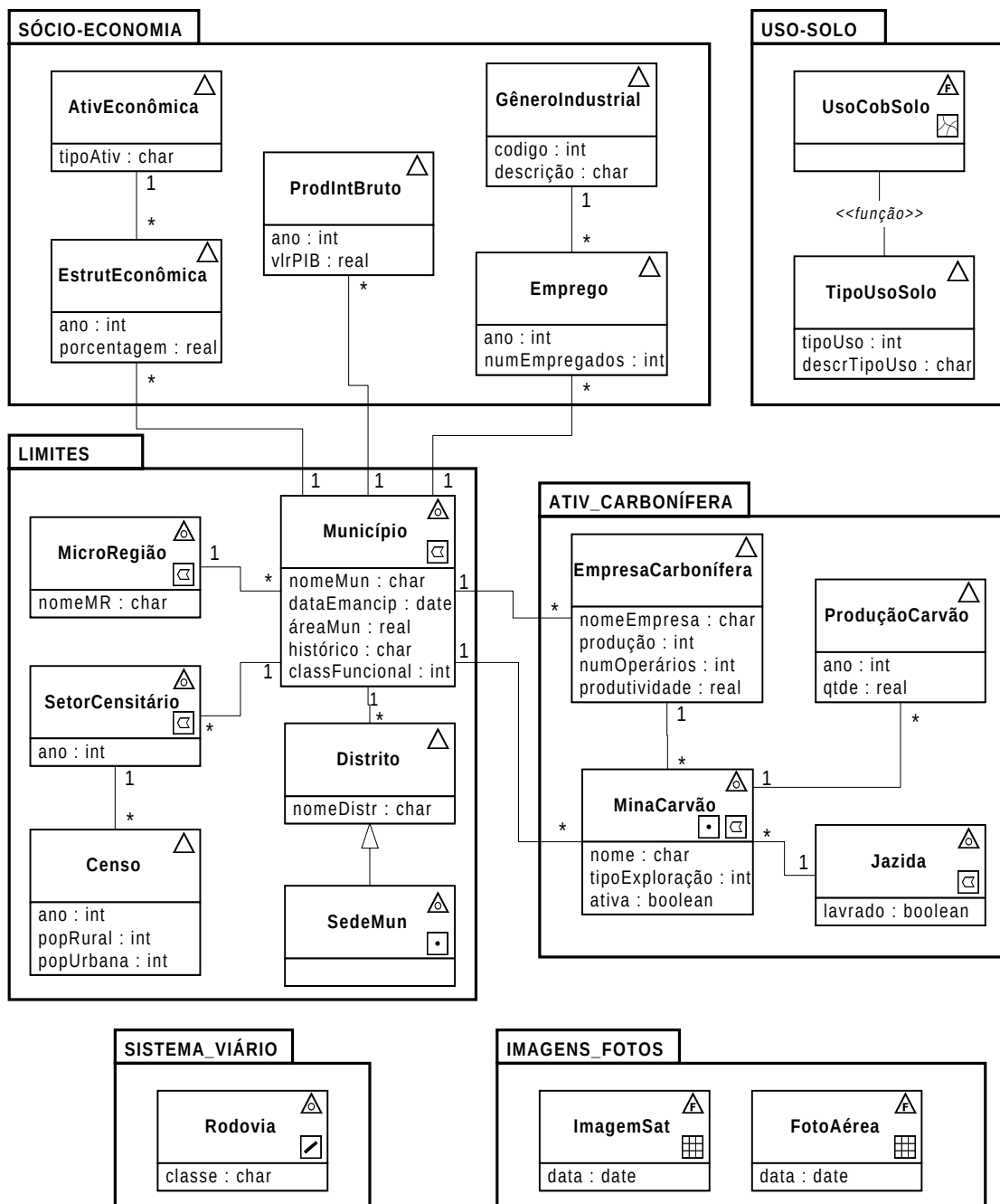


FIGURA 8 - Tema Meio Antrópico

O sub-tema *Sócio-Economia* agrupa, apenas, dados descritivos modelados como subclasses de OBJETO_NÃO_GEOGRÁFICO. Todos eles estão associados à classe *Município*. Exemplos de tipos de atividade econômica incluem indústria geral, agricultura, extração mineral e comércio. Informações sobre número de empregados estão separadas por gênero industrial como, por exemplo, extração de minerais, metalurgia, madeira, química e têxtil [SOU 97]. Para cada município, também é mantido o valor do produto interno bruto (PIB) anual.

O sub-tema *Uso-Solo* foi modelado com base no padrão de análise *Campo Geográfico Categórico* [LIS 9?]. A classe *UsoCobSolo*, subclasse de CAMPO_GEOGRÁFICO tem seu componente espacial abstraído como polígonos adjacentes. Os tipos de uso identificados são modelados na classe *TipoUsoSolo*, subclasse de OBJETO_NÃO_GEOGRÁFICO. A associação entre as classes *UsoCobSolo* e *TipoUsoSolo* é uma associação do tipo estereótipo, representada pela expressão <<função>>. O estereótipo, neste caso, foi usado para reforçar a idéia de que esta associação é uma função do espaço (região sobre a qual ocorre o campo Uso e cobertura do solo) em um domínio (Tipos de uso do solo), mas não uma simples associação entre duas instâncias de objetos.

Por último, os sub-temas *Sistema_Viário* e *Imagens_Fotos* são temas básicos, cujas informações são utilizadas para gerar outras informações. Por exemplo, imagens LandSat foram utilizadas para confecção do mapa de uso e cobertura do solo, enquanto fotos aéreas foram usadas na confecção do mapa de solos. As rodovias não estão individualizadas, estando apenas classificadas como principal, secundária e caminho.

3.2.2 Tema Meio_Biótico_e_Abiótico

O tema *Meio_Biótico_e_Abiótico* (Figura 9) agrega cinco sub-temas: *Hidrografia*; *Clima*; *Solos*; *Geologia*; e *Relevo*.

O sub-tema *Hidrografia* engloba classes como *RecursoHídrico*, que tem como principal instância o rio Jacuí em sua parte mais próxima da foz no lago do Guaíba. Os recursos hídricos estão classificados apenas como permanentes ou intermitentes. O rio Jacuí possui representação espacial poligonal, enquanto os demais rios da região estão representados de forma linear. Existem diversos pontos de coleta de amostragem de água, para os quais foram realizadas análises envolvendo diversos parâmetros [ALM 97]. Além das bacias hidrográficas dos principais afluentes do rio Jacuí, foram definidas bacias de drenagem para alguns pontos de amostragem. Devido à herança de propriedades na especialização, a classe *BaciaDrenagemPto* é uma subclasse de OBJETO_GEOGRÁFICO com representação espacial do tipo polígono. A existência de algumas pequenas ilhas originou a classe *Ilha*, para a qual existe apenas representação espacial. A modelagem do sub-tema *Hidrografia* foi realizada com base nos padrões de análise *Hidrografia* e *Parâmetros de Qualidade Ambiental*, definidos em [LIS 9?].

O sub-tema *Clima* foi especificado com base no padrão de análise *Campo Geográfico Numérico* [LIS 9?]. Com base em informações coletadas em uma série de estações meteorológicas, cuja classe é subclasse de *ObjetoGeográfico* (com representação pontual), várias variáveis contínuas foram medidas, as quais deram origem aos seguintes campos geográficos: *Precipitação*, *Temperatura*, *Umidade*, *Vento*, *Evaporação* e *Insolação* [FER 97].

O sub-tema *Relevo* engloba três classes: *Topografia*; *Declividade*; e *Exposição*. A classe *Topografia* foi modelada com três formas espaciais, ou seja, três formas distintas de abstração de seu componente espacial: *PontosIrregulares* (pontos contados); *Isolinhas* (curvas de nível); e *GradePontos* (modelo numérico de terreno - MNT). As classes *Declividade* e *Exposição* são derivadas do modelo numérico de terreno, mantendo o mesmo tipo de componente espacial, ou seja, *GradePontos*.

Como o sub-tema *Uso_Solo* (Figura 8), os sub-temas *Solos* e *Geologia* foram modelados com base no padrão de análise *Campo Geográfico Categórico* [LIS 9?]. O sub-tema *Solos* possui ainda a classe *PtoAmostrSolo*, cujas instâncias indicam a localização dos pontos de coleta para os quais existem uma série de dados referentes às análises de solos realizadas com as amostras de solos coletadas nestes pontos [SCH 97].

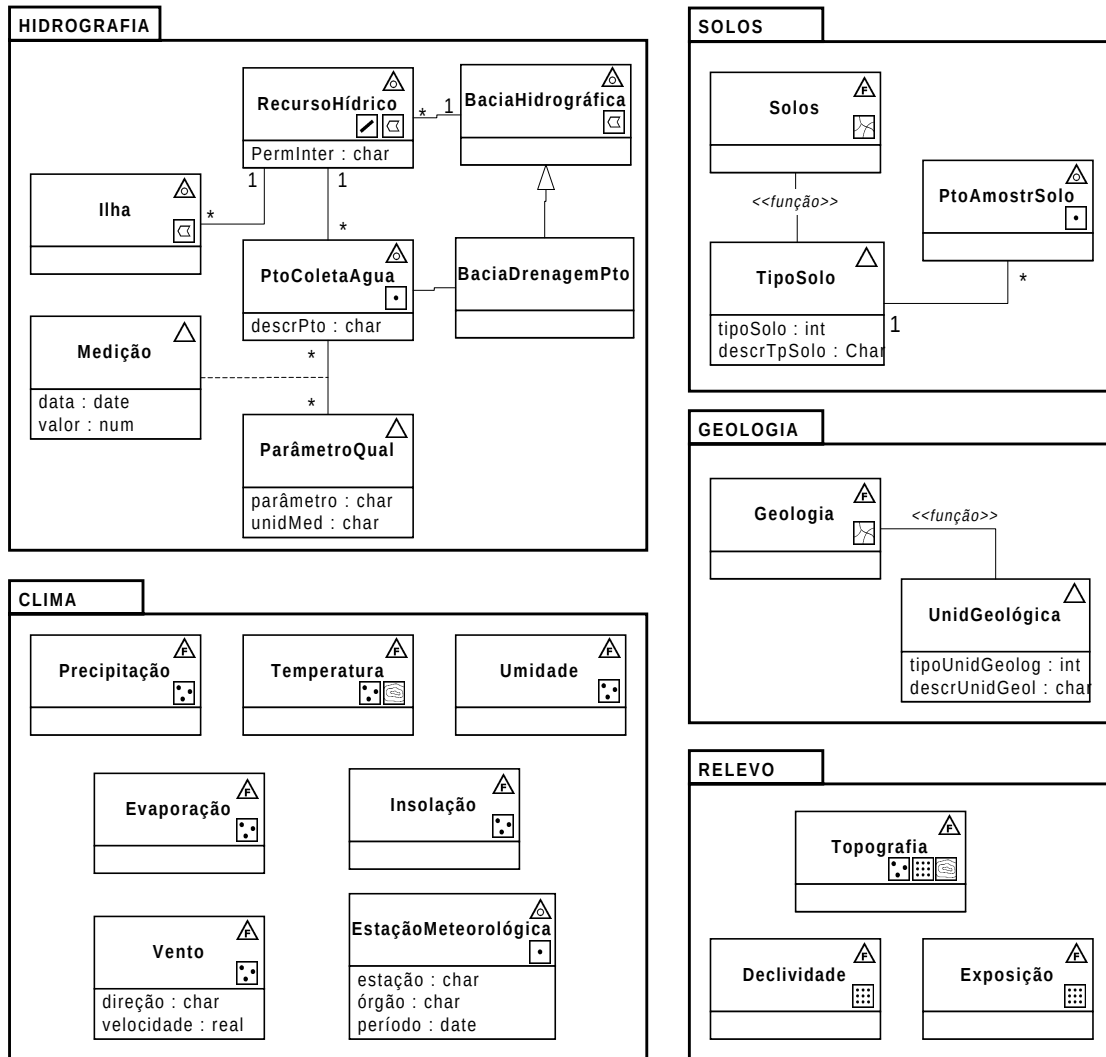


FIGURA 9 - Tema Meio Biótico e Abiótico

4 CONSIDERAÇÕES FINAIS

O projeto conceitual de banco de dados, embora seja uma etapa a ser realizada no início do ciclo de vida do desenvolvimento de sistemas, mostrou-se útil mesmo sendo realizado após a existência do sistema. Isto porque o esquema conceitual do banco de dados do Projeto PADCT/CIAMB possibilita ao usuário obter uma visão global dos dados produzidos durante o desenvolvimento do projeto.

Embora as metodologias de desenvolvimento de software incluam uma etapa inicial de projeto conceitual, o desenvolvimento de aplicações apoiadas em SIG tem sido realizado, muitas vezes, de forma incremental e diretamente no software de SIG. A consequência disso é que, com frequência, ocorrem problemas que poderiam ter sido

evitados através da modelagem conceitual (ex.: redundância de dados e ausência de relacionamentos importantes).

Durante a elaboração do esquema conceitual do Projeto PADCT/CIAMB, alguns conjuntos de dados redundantes foram identificados (ex.: dados sobre senso demográfico obtidos de diferentes fontes e estruturados de forma distinta e não relacionada). Também foram identificados alguns dados muito específicos a um determinado sub-projeto, que não foram incorporados ao esquema, uma vez que esses dados permanecerão apenas registrados nos relatórios dos sub-projetos e não farão parte do banco de dados digital (ex.: dados sobre índices de contaminação do sangue humano através de resíduos de carvão).

A elaboração do esquema conceitual com base no GeoFrame apresenta algumas vantagens importantes. Dentre as quais, pode-se citar:

- o esquema de dados final torna-se bastante claro, uma vez que apenas os elementos essenciais da aplicação são modelados;
- o uso de estereótipos permite, sem sobrecarregar visualmente o esquema, a fácil diferenciação entre os objetos não geográficos e os fenômenos geográficos (campos e objetos);
- a divisão do diagrama de classes em temas, especificado através de pacotes, torna o esquema fácil de ser lido, uma vez que a atenção do leitor pode se ater a apenas uma pequena parte do esquema por vez;
- devido ao alto nível de acoplamento entre as classes dentro de um mesmo tema, o número de associações entre classes de diferentes temas fica reduzido, contribuindo para a clareza do esquema.

A ocorrência de um mesmo tema em diversas aplicações de SIG possibilita ao projetista identificar padrões de análise a serem reutilizados durante a modelagem de uma nova aplicação. O uso dos padrões de análise *Parâmetros de Qualidade Ambiental*, *Campo Geográfico Categórico*, *Campo Geográfico Numérico* e *Hidrografia* reduziu o tempo gasto no projeto e a possibilidade de erros de modelagem, uma vez que esses padrões já foram testados anteriormente em outros sistemas [LIS 96, LIS 97]. Um novo sub-esquema candidato a padrão de análise diz respeito ao tema *Limites*, cujas classes (ex.: *Município*, *Distrito*) parecem ser recorrentes em outras aplicações de SIG.

4.6 Artigo XI: Projeto de Banco de Dados Geográficos: mapeando esquemas GeoFrame para o SIG Spring

Jugurta Lisboa Filho, Andreia Castro Costa* e Cirano Iochpe

*UFRGS - Programa de Pós-Graduação em Computação

In: BRAZILIAN WORKSHOP ON GEOINFORMATICS - GEO-INFO, 1., 1999, Campinas. **Anais...** São José dos Campos: INPE, 1999.

ABSTRACT

GeoFrame is a conceptual framework developed for geographic data conceptual modeling. This paper describes an experiment of mapping data conceptual schema developed on the basis of **GeoFrame**, into the Geographic Information System *Spring*.

RESUMO

Resumo. **GeoFrame** é um *framework* conceitual desenvolvido para modelagem conceitual de dados geográficos. Este artigo descreve um exercício de mapeamento de esquemas conceituais baseados no **GeoFrame**, no Sistema de Informação Geográfica *Spring*.

1 INTRODUÇÃO

Atualmente, cada sistema de informação geográfica (SIG) implementa seu próprio modelo de dados. Como consequência, não é possível definir um conjunto único de regras para o mapeamento de esquemas conceituais nos esquemas lógicos dos diversos SIG existentes. Com o avanço das pesquisas do consórcio OpenGIS [OPE 98], buscando a definição de um modelo de dados interoperável, será possível definir um conjunto de regras para o mapeamento no modelo de dados do OpenGIS.

O grupo de pesquisa em banco de dados geográficos do Instituto de Informática da UFRGS, tem investigado estratégias para projeto de bancos de dados geográficos (BDGeo), utilizando instrumentos de reutilização de esquemas, como *frameworks* e padrões de análise [FOW 97].

Um segundo tópico de investigação do grupo envolve a definição de regras para mapeamento de esquemas conceituais, elaborados com base no *framework* **GeoFrame** [LIS 99], em esquemas de um modelo de dados genérico. Entre outras, está sendo estudada, atualmente, a possibilidade de uso do padrão SAIF [BRI 95] como modelo genérico, já que este pode ser lido pelo software de conversão de dados FME [FEA 99].

O presente artigo descreve um exercício de mapeamento de esquemas **GeoFrame** no software *Spring* [INP 99].

2 O FRAMEWORK GEOFRAME

GeoFrame é um *framework* conceitual que fornece um diagrama de classes básicas para auxiliar o projetista na modelagem conceitual de dados geográficos e na especificação de padrões de análise em BDGeo [LIS 99].

Souza [SOU 98] define um *framework* como “um projeto genérico em um domínio que pode ser adaptado a aplicações específicas, servindo como um molde para a construção de aplicações”. O **GeoFrame** foi elaborado sob este enfoque, fornecendo a base do projeto conceitual para uma família de aplicações geográficas.

A Figura 1 mostra o diagrama de classes do **GeoFrame**, o qual está especificado segundo as regras do formalismo da orientação a objetos, utilizando a notação gráfica da UML [BOO 98].

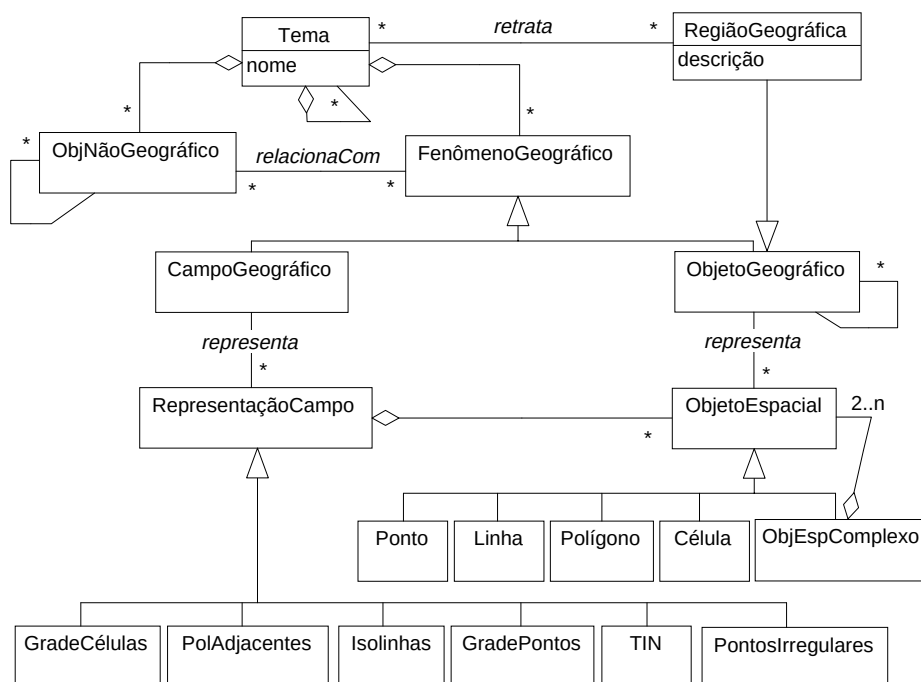


FIGURA 1 - Diagrama de Classes do GeoFrame

As classes TEMA e REGIÃO GEOGRÁFICA formam a base das aplicações geográficas, que têm como objetivo a manipulação de um conjunto de dados para uma determinada região de interesse, constituindo o BDGeo.

Para cada região geográfica, pode-se especificar uma coleção de temas. O agrupamento de classes que descrevem os fenômenos geográficos, em temas, funciona como um mecanismo para redução da complexidade em grandes esquemas de dados. O uso de temas permite, ao projetista, dividir o esquema de dados em subesquemas coesos, nos quais são agrupadas classes que estão fortemente relacionadas entre si.

Para aumentar a legibilidade do esquema resultante, os temas definidos pelo projetista não são modelados como subclasses da classe TEMA, mas sim, utilizando-se o construtor *Pacote* da UML [BOO 98].

Em um BDGeo, podem existir, além dos dados referentes a fenômenos georreferenciados, objetos convencionais presentes em qualquer sistema de informação. Objetos que não possuem referência a uma posição geográfica dão origem a subclasses da classe OBJNÃO GEOGRÁFICO.

A classe FENÔMENO GEOGRÁFICO generaliza os fenômenos cuja localização em relação à superfície terrestre seja considerada. Por exemplo, um distrito municipal é

uma instância de FENÔMENOGEOGRÁFICO, se seus atributos espaciais estiverem representados no BDGeo. Fenômenos geográficos e objetos não geográficos podem estar relacionados entre si.

Fenômenos geográficos são percebidos, na realidade geográfica, segundo as visões dicotômicas de campo e de objetos [GOO 92]. Essas duas visões implicam em diferentes formas de modelagem dos fenômenos geográficos. As classes CAMPOGEOGRÁFICO e OBJETOGEOGRÁFICO especializam a classe FENÔMENOGEOGRÁFICO, permitindo ao projetista especificar, de forma distinta, respectivamente, os campos e os objetos geográficos.

A classe OBJETOGEOGRÁFICO é uma generalização de todas as classes do domínio da aplicação que são percebidas na visão de objetos. Neste caso estão incluídas aquelas classes que representam fenômenos geográficos que podem ser individualizados, ou seja, que possuem identidade própria e suas características podem ser descritas através de atributos (ex.: *Estrada, Rio, Cidade*).

A classe CAMPOGEOGRÁFICO generaliza os fenômenos que se enquadram na visão de campo. Campos geográficos são modelados como funções sobre variáveis. Alguns campos referem-se a variáveis distribuídas sobre a superfície, de forma contínua (ex.: *Altimetria, Temperatura e Cobertura do Solo*), enquanto outros referem-se a variáveis distribuídas de forma discreta (ex.: *População e Ocorrências Epidemiológicas*) [PIR 97].

Com base no **GeoFrame**, cada classe do domínio da aplicação deve ser modelada como subclasse de OBJETONÃOGEOGRÁFICO, CAMPOGEOGRÁFICO ou OBJETOGEOGRÁFICO. Para evitar a sobrecarga visual do diagrama resultante, são utilizados estereótipos [BOO 98].

Para substituir os relacionamentos de generalização entre as classes do domínio e as classes do **GeoFrame**, são utilizados três estereótipos (Figura 2).

Um dos princípios fundamentais da modelagem conceitual é o de que os esquemas conceituais gerados devem conter, apenas, os elementos do domínio, desconsiderando os aspectos de implementação [PAR 98]. O objetivo de se incluir, no esquema conceitual, informações sobre objetos espaciais relacionados a fenômenos geográficos é o de permitir a abstração do componente espacial de cada fenômeno.

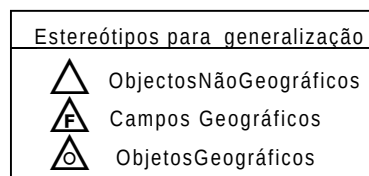


FIGURA 2 - Estereótipos¹¹ para generalização

Portanto, quando se considera pontos, polígonos, isolinhas ou grade de células, no **GeoFrame**, estão sendo tratadas as formas de abstração do componente espacial dos fenômenos geográficos e não as formas com que eles serão armazenados no BDGeo. Isso não exclui o fato de certos SIG implementarem estruturas de dados muito semelhantes para o armazenamento destes construtores.

Alguns fenômenos geográficos podem apresentar dimensão espacial complexa, ou seja, composta por outros objetos espaciais (ex.: uma rede hidrográfica). A classe

¹¹ Estes símbolos gráficos estão definidos em uma fonte especial de caracteres disponível em <http://www.inf.ufrgs.br/gpesquisa/sigmoda>.

OBJETOESPACIAL generaliza as classes necessárias para a especificação do componente espacial dos fenômenos geográficos percebidos na visão de objetos.

Os aspectos espaciais de um campo geográfico são abstraídos de forma diferente dos aspectos espaciais de um objeto geográfico. Chrisman [CHR 97] descreve diversos modelos geográficos através dos quais pode-se abstrair o componente espacial da informação geográfica. Os modelos relacionados por Chrisman, os quais são adequados à modelagem de fenômenos na visão de campo, podem ser resumidos nos seis modelos espaciais descritos por Goodchild [GOO 92]. São eles: grade de células; polígonos adjacentes; isolinhas; grade de pontos; rede triangular irregular; e pontos amostrados irregularmente. No **GeoFrame**, esses seis modelos correspondem às subclasses da classe REPRESENTAÇÃOCAMPO.

Um mesmo campo geográfico pode ter seu componente espacial abstraído de diferentes formas, ou seja, através de mais de um desses modelos. Situação semelhante ocorre com os objetos geográficos. Assim, todo campo ou objeto geográfico pode estar associado a múltiplas instâncias das classes REPRESENTAÇÃOCAMPO e OBJETOESPACIAL, respectivamente.

A possibilidade de se ter múltiplas representações para um mesmo fenômeno geográfico é modelada através de diferentes associações entre o fenômeno geográfico e as possíveis formas de abstração de seu componente espacial (associação *representa* na Figura 1). No **GeoFrame** estas variações podem ser especificadas através da combinação livre de diferentes estereótipos em uma mesma classe, com documentação adicional no dicionário de dados.

Um segundo conjunto de estereótipos (Figura 3) é usado para substituir as associações que resultam da modelagem do componente espacial dos fenômenos geográficos. A semântica de cada estereótipo é a substituição de uma associação entre o fenômeno geográfico e seu componente espacial, além da indicação da forma geométrica de sua representação.

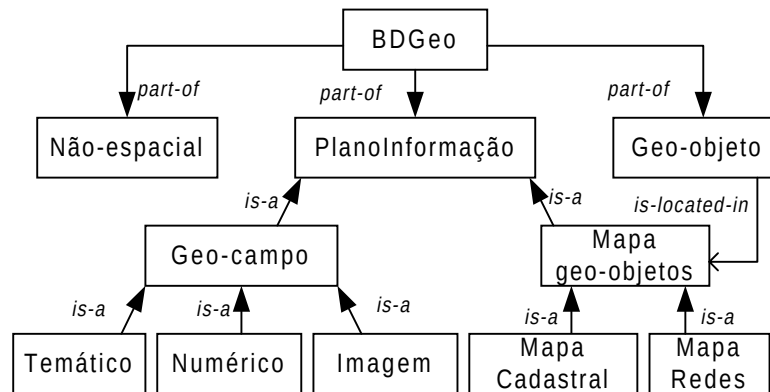
ObjetoEspacial	RepresentaçãoCampo	
 Ponto	 GradeCélulas	 GradePontos
 Linha	 PolAdjacentes	 TIN
 Polígono	 Isolinhas	 PontosIrregulares
 ObjComplexo		

FIGURA 3 - Estereótipos para associação

3 MAPEAMENTO GEOFRAME-SPRING

O SIG *Spring* [INP 99], desenvolvido pelo Instituto Nacional de Pesquisas Espaciais (INPE), foi escolhido para o exercício de mapeamento de um esquema conceitual elaborado a partir do **GeoFrame** (ou esquema **GeoFrame**), para um esquema proprietário de SIG.

O *Spring* foi escolhido por utilizar um modelo de dados orientado a objetos de alto nível, ou seja, mais próximo do usuário do que a maioria dos SIG disponíveis comercialmente. As classes do modelo conceitual do *Spring* (Figura 4) são semelhantes às classes do **GeoFrame**. Como exemplo, pode-se citar a diferenciação entre geo-campos e geo-objetos. Isto torna muito simples o mapeamento de um esquema **GeoFrame** para o *Spring*.

FIGURA 4 - Nível conceitual do *Spring*

Uma diferença básica entre o diagrama de classes do **GeoFrame** e o modelo *Spring* é que o **GeoFrame** foi definido para servir de base à modelagem conceitual de aplicações que podem ser implementadas em qualquer SIG, ou seja, fornece uma descrição do BDGeo de forma independente de aspectos de implementação. Por sua vez, o modelo *Spring* introduz elementos próprios de um software de SIG. Como exemplo, pode-se citar os conceitos de *plano de informação* e *mapa de geo-objetos*. Estes elementos do *Spring* fornecem as estruturas para a definição de mapas, o que no **GeoFrame** não tem correspondente.

No *Spring*, a especificação de um novo projeto inicia-se pela definição de um BDGeo, que corresponde à definição de um diretório, no qual são armazenados todos os dados e suas definições. Um esquema **GeoFrame** pode dar origem a um ou mais BDGeo, dependendo do tamanho e da complexidade do projeto modelado.

Cada região geográfica, especificada no esquema **GeoFrame**, dá origem a uma instância de *projeto*, cuja classe não consta no modelo *Spring* [INP 99]. No *Spring* um *projeto* define realmente a área física de trabalho. Na criação de um *projeto* são definidos um nome, a projeção cartográfica e o retângulo envolvente. Um único BDGeo pode conter diversos projetos.

Os temas que aparecem no esquema **GeoFrame** (Figura 5) não são considerados durante o mapeamento. Eles têm a função, apenas, de servir como instrumento de redução da complexidade em grandes esquemas.

A tabela abaixo apresenta um resumo das transformações a serem realizadas para mapear um esquema **GeoFrame** em uma especificação no *Spring*.

TABELA 1 - Classes do GeoFrame x classes Spring

	Obj Geo	GEOFRAME CampoGeo						Obj NãoGeo
SPRING								
Geo-objeto	x							
Temático	x*						x	
Numérico		x	x	x	x			
Imagem						x		
Cadastral	x							
Rede	x**							
Não-espacial								x

Conforme descrito na Seção 2, as classes do domínio são modeladas como subclasses de classes do **GeoFrame**, ou seja, OBJETO GEOGRÁFICO, CAMPO GEOGRÁFICO ou OBJNÃO GEOGRÁFICO. No *Spring*, estas subclasses são definidas como geo-objetos, planos de informação ou objetos não-espaciais.

As subclasses de OBJETO GEOGRÁFICO ($\triangle$) podem ser especificadas de diferentes maneiras. A forma mais natural (identificada por um $\star$) é a definição de um *geo-objeto*, para conter os atributos da classe, e um *mapa cadastral* para conter as representações espaciais dos geo-objetos.

Um caso especial ocorre quando os fenômenos geográficos, embora sejam percebidos na visão de objetos, não estão individualizados. Por exemplo, as rodovias podem estar classificadas apenas pela sua jurisdição (*federal*, *estadual* ou *municipal*). Assim, embora uma rodovia seja um fenômeno geográfico na visão de objetos, cada instância de rodovia não está individualizada. Neste caso, a classe Rodovia é especificada como um plano de informação *temático* (identificado por $\star^*$).

Por último, um objeto geográfico complexo ($\boxtimes$) pode ser especificado como um *mapa de rede* ($\star^{**}$).

Para as subclasses de CAMPO GEOGRÁFICO ($\triangle$), o *Spring* fornece três opções de planos de informação: *temático*, *numérico* ou *imagem*. Um campo geográfico, cujo componente espacial esteja modelado como polígonos adjacentes ($\boxplus$) é definido na categoria de geo-campo *temático*.

Se o campo geográfico está associado a uma subclasse de objeto não geográfico ($\triangle$), correspondente à função de mapeamento do campo, a subclasse ($\triangle$) poderá ser especificada ou como uma lista de classes temáticas, ou como uma categoria de *objeto não-espacial*. Na Figura 5, a classe *TipoUsoSolo* foi mapeada como uma lista de classes temáticas na categoria *UsoCobSolo*.

Um campo geográfico com componente espacial modelado como grade de células ($\boxplus$) é definido na categoria *imagem*. Os demais modelos espaciais de campos geográficos ($\boxdot$, $\boxtimes$, $\boxminus$, $\boxdiv$) levam à definição de elementos na categoria *numérico*.

O mapeamento de um ObjNãoGeográfico ($\triangle$) deveria ser feito em um elemento da categoria *objeto não-espacial*. Porém, na versão 3.3 do *Spring* esta categoria ainda não está disponível. Neste caso, a alternativa é especificá-lo como um geo-objeto.

A Figura 5 ilustra parte de um esquema **GeoFrame** referente à uma aplicação de controle de exploração de carvão mineral e o esquema correspondente no *Spring*.

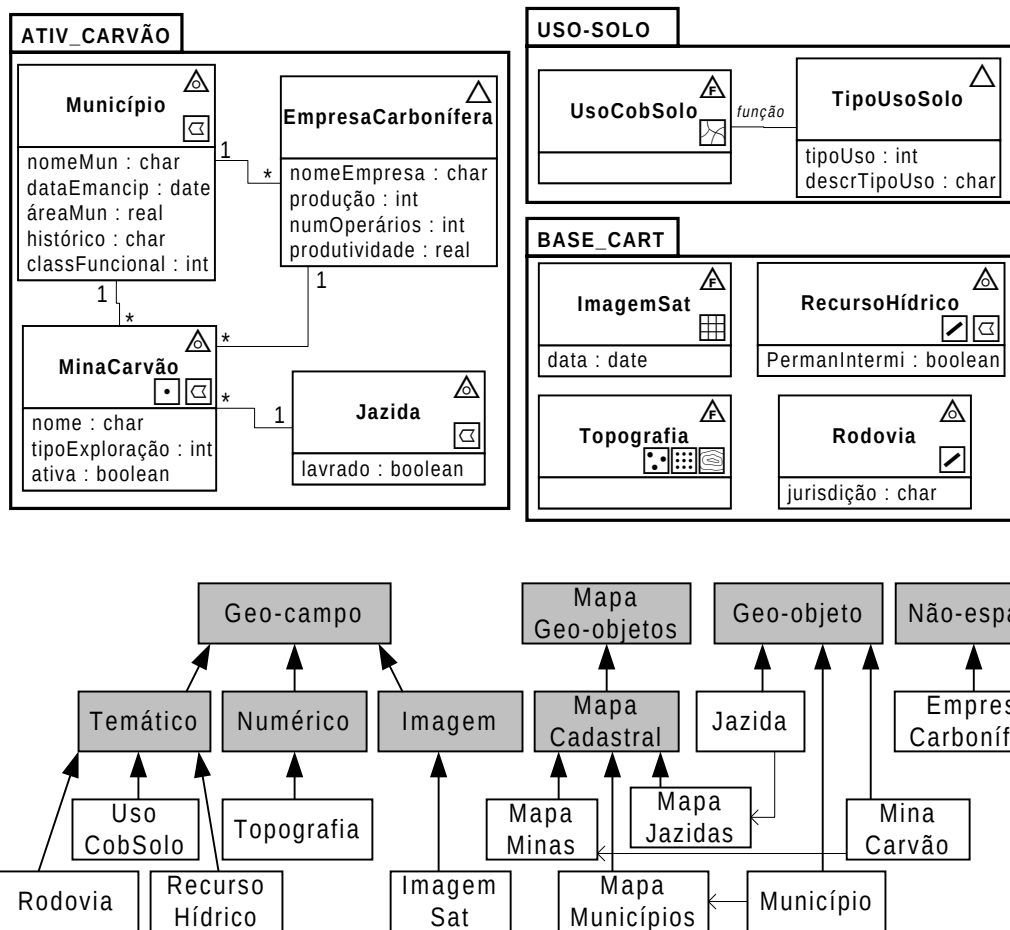


FIGURA 5 - Exemplo de mapeamento GeoFrame-Spring

4 CONCLUSÕES

O mapeamento de esquemas **GeoFrame** em projetos no *Spring* é facilitado devido a este SIG prover um modelo de dados, de alto nível, no qual são definidos os elementos do domínio da aplicação. O projetista não necessita, em um primeiro instante, se preocupar com as estruturas de armazenamento dos dados.

Os estereótipos referentes ao componente espacial dos fenômenos geográficos, além de indicar o tipo de categoria a ser definida no *Spring*, também orienta o usuário no processo de criação da base de dados propriamente dita, indicando a forma geométrica de suas representações.

5 Conclusões

Esta tese apresentou um conjunto de instrumentos que podem ser utilizados como apoio à modelagem conceitual de banco de dados geográficos, com base na abordagem de reutilização.

O objetivo geral da pesquisa de tese, que era a investigação de modelos, metodologias e métodos que possam ser aplicados durante a realização da modelagem conceitual de banco de dados geográficos, foi atingido. Primeiramente, através do estudo e análise de vários modelos conceituais, propostos na literatura, especificamente para aplicações de SIG [LIS 96a, LIS 97b, LIS 98a, LIS 99c].

Neste estudo, identificou-se uma série de requisitos de modelagem que devem ser satisfeitos pelos modelos conceituais para bancos de dados geográficos. Descreve-se abaixo as seguintes conclusões obtidas a partir da experiência de uso de alguns dos modelos propostos na literatura, durante o desenvolvimento da modelagem conceitual dos dados de sistemas reais.

- Todos os modelos seguem um mesmo caminho, no qual se buscou acrescentar novas primitivas de abstração a algum dos formalismos conceituais existentes, com variações ocorrendo, principalmente, na sintaxe da linguagem e/ou na notação gráfica proposta. No entanto, alguns modelos, em vez de realmente introduzirem novas primitivas de modelagem, restringem-se a descrever como e o que modelar, da realidade geográfica, utilizando-se primitivas de modelagem do próprio formalismo de origem. Ou seja, alguns modelos apresentam, na verdade, uma modelagem inicial dos aspectos da realidade geográfica, mas não um novo conjunto de construtores abstratos (ex.: GISER [SHK 97]).
- Nos casos em que se buscou a definição de um modelo conceitual mais rico semanticamente, aumentou-se, significativamente, a dificuldade de utilizar o modelo, devido ao grande número de primitivas introduzidas e da necessidade de existência de ferramentas de apoio (ex.: Modelo MADS [PAR 98]).
- Em conjunto, os modelos existentes já atendem, em maior ou menor grau, aos requisitos de modelagem identificados. No entanto, o estudo mostrou que alguns dos requisitos analisados ainda necessitam ser melhor pesquisados como, por exemplo, os aspectos temporais e os aspectos de qualidade.

Quanto aos aspectos temporais, os problemas de modelagem estão diretamente associados à falta de suporte para manipulação de dados temporais nos sistemas atuais. Alguns SIG provêm mecanismos específicos para o tratamento de aspectos temporais dos dados (ex.: Spring [INP 99]), mas apenas de forma superficial. Assim, mesmo sendo modelados no nível conceitual, a implementação dos aspectos temporais fica limitada pela falta, por exemplo, de uma linguagem de consulta contendo cláusulas temporais. Pesquisas estão sendo desenvolvidas com o objetivo de suprir essas deficiências [PEU 95, REN 97].

No caso dos aspectos de qualidade, o problema é mais crítico. Mais do que um problema de modelagem conceitual, qualidade de dados geo-espaciais é um problema que está sendo altamente pesquisado pela comunidade de usuários ligados à área das Geociências como, por exemplo, nas áreas de agrimensura, cartografia e sensoriamento remoto. A alternativa existente, mas ainda bastante fraca do ponto de vista das necessidades do usuário especialista,

é a inclusão de metadados de qualidade como documentação do conjunto de dados. Na forma apenas de documentação, os dados de qualidade não possibilitam que os usuários utilizem essa informação nas suas análises. Metadados de qualidade têm sido utilizados apenas como informação adicional para permitir que um usuário avalie a adequação de dados existentes, para possível reutilização [WEB 98].

Paralelamente ao estudo dos diversos modelos conceituais existentes foi desenvolvido, e também experimentado na prática, um diagrama de classes, denominado BCH/GIS, que serve como ponto de partida para a modelagem conceitual de novos bancos de dados geográficos.

A medida que novas aplicações foram sendo modeladas, com base no diagrama BCH/GIS, os problemas identificados foram originando novas versões do diagrama, conforme descrito no Capítulo 3. Os principais problemas identificados através dessas experiências práticas, e que foram solucionados até a terceira versão do diagrama BCH/GIS ou, posteriormente, quando da transformação desse diagrama no *framework* GeoFrame, estão listados a seguir.

- Alguns modelos (ex.: de Clementine [CLE 94]), induziam os projetistas a modelarem as classes da realidade como subclasses de um conjunto predefinido de classes. Desta forma, o projetista deveria especificar uma classe *Rio* como subclasse da classe *Linha*. Embora esse problema tenha sido identificado desde o início desta pesquisa, na modelagem de campos geográficos cometeu-se o mesmo equívoco, induzindo os projetistas a especificarem, por exemplo, uma classe *Relevo* como subclasse da classe de representação espacial *Isolinhas*. Este problema foi solucionado já na segunda versão do diagrama BCH/GIS.
- Necessidade de mecanismos que possibilitem diferenciar, no esquema de dados, as classes que representam fenômenos geográficos (campos e objetos geográficos) das classes de objetos convencionais, ou seja, não georreferenciados. Até a terceira versão do diagrama BCH/GIS, as classes que representavam objetos geográficos eram diferenciadas das demais, através da presença de um símbolo gráfico (ou pictograma) no interior do retângulo que representa a classe. No entanto, somente após a especificação do GeoFrame é que as classes de campos geográficos passaram a ser diferenciadas das classes de objetos convencionais, o que foi feito através do uso de estereótipos.
- Necessidade de tratamento uniforme para a modelagem dos componentes espaciais de campos e objetos geográficos. Até a terceira versão do diagrama BCH/GIS, a modelagem do componente espacial dos objetos geográficos era feita através do uso de pictogramas, enquanto que para os campos geográficos, seus componentes espaciais eram especificados através de associações com as possíveis formas de abstração de seu componente espacial, que eram definidas como classes do diagrama BCH/GIS. A partir do GeoFrame, foram definidos estereótipos para a modelagem do componente espacial de campos geográficos, unificando a forma de modelagem do componente espacial dos fenômenos geográficos.
- Necessidade de aprimorar a forma de modelar fenômenos geográficos com múltiplas representações espaciais. Este requisito, identificado no trabalho de dissertação de mestrado que avaliou a adequação do diagrama BCH/GIS através de sua utilização no Projeto SIGPROGB [GAR 98], foi facilmente contemplado quando da solução do problema anterior, ou seja, uma vez que se

definiu estereótipos para a modelagem de todos os fenômenos geográficos, a simples combinação de dois ou mais estereótipos resolve, em parte, este problema. Ao contrário do modelo Modul-R [BED 96], no qual foram definidas diferentes semânticas para diferentes formas de combinar os pictogramas, no GeoFrame optou-se por deixar para uma documentação adicional (ex.: dicionário de dados) as condições especiais que porventura ocorram. Esta decisão teve como base o requisito de simplicidade de uso e entendimento do modelo.

- Necessidade de diferenciar os relacionamentos espaciais dos convencionais. Por exemplo, duas classes podem estar associadas através de um relacionamento convencional, um relacionamento espacial ou ambos. Alguns modelos conceituais para SIG introduziram novos construtores para a modelagem de relacionamentos espaciais. Por exemplo, no modelo GeoOOA [KOS 97] são definidos três novos tipos de construtores de agregação espacial: *cobertura*, *pertinência* e *partição*. Estes construtores, embora permitam a especificação de restrições de integridade espaciais, apresentam duas desvantagens. Primeiro, tornam o modelo mais difícil de ser entendido por parte dos usuários/projetistas e, segundo, por serem incompletos, ou seja, não permitem que diversos outros tipos de relacionamentos possam ser especificados através de mecanismos semelhantes. Por exemplo, no modelo GeoOOA é possível especificar que a localização pontual de uma *indústria* deve estar dentro (“pertinência”) da região de um *distrito industrial* mas, ao mesmo tempo, não existe um construtor, tão detalhado quanto este, para especificar que um *distrito industrial* pode estar sendo “cruzado” por uma *avenida* ou ser “adjacente” a um *rio*.

Na nossa experiência, por mais que o projetista tente especificar as restrições de integridade espacial, sempre haverá um volume muito maior de relacionamentos que não terão sido especificados, embora tenham que ser verificados. Desta forma, a abordagem para modelagem de relacionamentos no GeoFrame, e que tem sido adotada no desenvolvimento de nossos sistemas [LIS 96, LIS 97 e GAR 98], é a de modelar apenas os relacionamentos semânticos, deixando os relacionamentos espaciais para serem observados quando dos processos de aquisição dos dados e verificação da qualidade dos dados adquiridos. Portanto, são usados apenas os construtores de modelagem de relacionamentos da própria UML, isto é, associação, especialização, agregação e composição [BOO 98]. Apesar disso, é possível, caso o projetista escolha pela especificação de relacionamentos espaciais, definir seus próprios estereótipos de associação, agregação ou composição, para diferenciar os relacionamentos espaciais dos demais relacionamentos.

Embora ainda apresentem algumas deficiências, os modelos conceituais, propostos na literatura, permitem, em maior ou menor grau, que os projetistas realizem a modelagem conceitual do banco de dados geográfico. A partir dessa constatação coloca-se, então, a pergunta: Por que os projetistas de SIG relutam, ainda, em realizar a fase de modelagem conceitual em seus projetos? Acredita-se que isto ocorra, não mais pela ausência de um modelo adequado para as aplicações de SIG, mas sim, devido a deficiências nas metodologias de desenvolvimento em que esses modelos são empregados. Além disso, credita-se tal comportamento, também, à falta de experiência dos projetistas que são, em sua maioria, profissionais de outras áreas diferentes da computação. Uma das possíveis soluções para este problema, conforme descrito nas

hipóteses levantadas no início dessa tese, é a utilização de metodologias de desenvolvimento de sistemas baseadas na abordagem de reutilização de projetos.

Para resolver essas limitações, definiu-se os objetivos mais específicos dessa tese, ou seja, o desenvolvimento de uma série de instrumentos que, utilizados de forma orquestrada, possibilitem o emprego da abordagem de reutilização no projeto conceitual de banco de dados geográficos.

Uma vez constatado que as aplicações de SIG são projetadas e desenvolvidas, em sua maioria, pelos próprios usuários, estabeleceu-se, como uma das condições fundamentais desta pesquisa, a busca por uma solução que possa ser verdadeiramente utilizada na prática, tanto por especialistas em projetos computacionais como pelos próprios usuários de SIG.

Inicialmente foram identificados e descritos alguns padrões de análise em aplicações nas áreas de controle ambiental e gestão municipal (Artigo IX – Seção 4.4). A identificação desses padrões de análise teve como base uma série de esquemas conceituais de banco de dados geográficos, tanto os desenvolvidos pelo grupo do Projeto SIGMODA, como esquemas desenvolvidos por terceiros, além de exemplos disponíveis na literatura.

Padrões de análise, embora fáceis de utilizar, não são simples de serem identificados e especificados. A identificação de um padrão de análise requer, do projetista, a habilidade de perceber sub-esquemas recorrentes, abstrair os elementos essenciais destes, de forma a torná-los independentes da aplicação e, por fim, descrevê-los utilizando uma linguagem de descrição de padrões de análise. Até o presente momento não foi encontrado, na literatura, nenhum padrão de análise específico para aplicações de SIG, embora alguns padrões de análise sejam tão genéricos que possam ser utilizados em diversos domínios de aplicação, inclusive em aplicações de SIG. As únicas publicações, nas quais os conceitos de padrões são aplicados à área de SIG, abordam o uso de padrões de projeto (não de análise), no projeto de software de SIG (não uma aplicação de SIG) [BAL 97, GOR 97, GOR 98].

Dos padrões de análise identificados e descritos nesta tese (Artigo IX – Seção 4.4), as seguintes conclusões adicionais podem ser apresentadas:

- alguns padrões são mais abstratos como, por exemplo, os padrões *Campo Geográfico Categórico* e *Campo Geográfico Numérico*. No entanto, tais padrões se parecem mais com os padrões de projeto, por não apresentarem uma solução para um domínio específico, podendo ser utilizados em diversos domínios de aplicação em que ocorra a presença de campos geográficos;
- outros padrões são mais específicos, por apresentarem soluções de modelagem para um único tema, como os padrões *Hidrografia*, *Malha Viária Urbana* ou *Cadastro de Lotes Urbanos*. Como esses são temas básicos que ocorrem em uma grande gama de aplicações geográficas, a reutilização desses padrões é praticamente direta, facilitando ainda mais sua reutilização, principalmente, por parte de projetistas não especialistas, uma vez que os padrões são descritos de forma a empregar termos e jargões com os quais os usuários já estão habituados. As diferenças de terminologia, como a que ocorre no padrão *Hidrografia*, não chegam a prejudicar o entendimento do padrão. Além disso, o uso de padrões de análise pode contribuir para a redução dessas diferenças de terminologia, pelo menos, dentro de uma mesma instituição ou entre instituições que trabalham de forma cooperativa.

Conforme apresentado nas conclusões do Artigo IX (Seção 4.4), para o sucesso da abordagem de reutilização é necessário desenvolver a cultura da cooperação entre os

desenvolvedores de sistema. Embora, no Brasil, muitos projetistas de SIG ainda tenham grandes limitações em compartilhar os seus dados, seja por qual for o motivo, a tendência que se observa é que, aos poucos, os usuários de SIG vão descobrindo as vantagens de compartilhar dados geo-espaciais e agora, também, projetos de banco de dados, o que não compromete, por exemplo, a concorrência entre empresas/instituições.

Outro ponto importante a ressaltar é que um padrão de análise não necessita apresentar uma solução original. Pelo contrário, os padrões devem documentar soluções já testadas e validadas, pois são soluções para problemas recorrentes. Assim, espera-se que outros projetistas venham a contribuir para a evolução da documentação de um padrão. Os padrões de análise descritos nesta tese não são diferentes, e espera-se que eles também sejam melhorados a partir de contribuições de outros projetistas.

5.1 Resumo das Contribuições da Tese

O instrumento básico da abordagem de reutilização proposta nesta tese é o *framework* conceitual GeoFrame [LIS 99]. O GeoFrame fornece um entendimento básico dos elementos presentes em qualquer banco de dados geográficos, servindo como ponto de partida para a modelagem conceitual de banco de dados geográficos e, também, para a descrição de padrões de análise para SIG. Embora o diagrama de classes do GeoFrame esteja documentado na linguagem UML (*Unified Modeling language*) [BOO 98], sua estrutura é bastante genérica e pode ser adaptada para utilização com base em outros modelos conceituais orientados a objetos (ex.: OMT [RUM 91], OOA [COA 91]) ou, até mesmo, em extensões do modelo Entidade-Relacionamentos (ex.: EER [ELM 94]).

A grande vantagem do uso da linguagem UML é que, através de sua escolha, chegou-se a uma solução bastante simples, na qual são utilizados apenas os construtores da própria linguagem juntamente com um pequeno conjunto de estereótipos. É importante ressaltar que o estereótipo é o mecanismo de extensão da própria UML. Assim, alcançou-se uma notação gráfica bastante compacta e, portanto, de fácil aprendizado e entendimento por parte de usuários e projetistas não especialistas em modelagem conceitual.

Esta solução UML-GeoFrame, que é independente de software, se enquadra nos modelos conceituais para SIG chamados de segunda geração [BED 99]. Outra vantagem dessa solução é a possibilidade do uso de qualquer ferramenta CASE para UML existente no mercado, juntamente com a fonte de caracteres Sigmoda (Anexo 5).

Uma contribuição adicional do GeoFrame para a pesquisa sobre modelos conceituais de dados para SIG é a forma através da qual deve ser modelado o componente espacial de campos geográficos, ou seja, de fenômenos geográficos caracterizados por estarem distribuídos continuamente sobre uma região. Este mecanismo, publicado inicialmente em [LIS 96] e [LIS 96a], já foi, inclusive, incorporado a outros modelos, como é o caso do modelo Geo-OMT [BOR 97].

O segundo instrumento proposto para suportar a reutilização de projetos de banco de dados geográficos é o uso de padrões de análise. Padrões de análise para reutilização de projetos de bancos de dados geográficos devem ser descritos com base em uma linguagem de descrição de padrões que inclui um modelo conceitual de dados. A integração UML-GeoFrame fornece uma linguagem bastante simples para a descrição das soluções apresentadas pelos padrões de análise para banco de dados geográficos.

Foi identificado e descrito um conjunto inicial de padrões de análise para aplicações nas áreas de controle ambiental e gestão municipal (Artigo IX – Seção 4.4). A aplicação de alguns desses padrões foi exemplificada através da modelagem de um

sistema real na área de controle ambiental (Projeto PADCT/CIAMB [LIS 99b] – Seção 4.5).

Além das contribuições obtidas como resultado da busca para atender os objetivos estabelecidos na tese, diversas contribuições indiretas foram obtidas. As mais importantes são:

- o resultado do processo de modelagem conceitual é um esquema conceitual de dados. Após validado e aprovado, o esquema conceitual tem de ser transformado em um esquema lógico, o qual é dependente do software a ser utilizado. Como não existe um modelo lógico de SIG que seja único, como ocorre com o modelo relacional, também não existe uma regra única para transformação de um esquema conceitual em um esquema lógico de SIG. Apresentou-se, através de um exercício de mapeamento, a transformação de um esquema conceitual, elaborado usando a notação UML-GeoFrame, em um esquema do SIG Spring [INP 99]. Essa experiência ([LIS 99a] – Seção 4.6), pode servir de base para o mapeamento de esquemas conceituais para outros SIG;
- atualmente, o compartilhamento de soluções de projeto de banco de dados geográficos é um tópico praticamente inexplorado. No entanto, para o compartilhamento de dados geo-espaciais já existe uma infra-estrutura bem mais avançada. Parte desta pesquisa de doutorado esteve voltada para a busca de soluções que facilitassem o compartilhamento de dados geo-espaciais entre diferentes usuários/instituições. Como resultado desse trabalho, desenvolveu-se, no âmbito do Projeto SIGMODA, um estudo detalhado sobre os padrões de metadados existentes [WEB 98, WEB 99] e foi implementado um protótipo de nó do sistema de *Clearinghouse* para os países de língua portuguesa, o qual foi desenvolvido como trabalho de conclusão de curso do acadêmico Fábio Villavicencio ([VIL 99] – Anexo 2).

5.2 Trabalhos Futuros

Devido às particularidades apresentadas pelas aplicações de SIG, existe uma forte possibilidade de se obter, cada vez mais, ganhos reais de produtividade e qualidade através da reutilização de projetos de banco de dados geográficos.

O uso de padrões de análise como instrumento de reutilização é um tema de pesquisa muito recente, mesmo quando se observa o seu uso em outros domínios da área de Engenharia de Software. Por ser, ainda, muito incipiente, existem muitos tópicos relacionados a padrões de análise que necessitam ser desenvolvidos.

Além dos padrões de análise, outros assuntos abordados nesta tese também necessitam maiores pesquisas. A seguir, estão listados os principais tópicos de pesquisa que poderão dar continuidade a esta tese:

- investigação de novos padrões de análise para banco de dados geográficos em diferentes domínios de aplicação de SIG (ex.: aplicações na área de redes de infra-estrutura, agropecuária de precisão, pesquisas de marketing);
- evolução dos padrões de análise aqui propostos. A partir da identificação de novos usos para os padrões propostos nesta tese, deverão ocorrer atualizações e aprimoramentos desses padrões;
- transformação dos padrões de análise em componentes geográficos reutilizáveis. Da mesma forma com que um padrão de análise pode ser reutilizado, suas respectivas implementações, desenvolvidas através de software de SIG, poderão se tornar componentes reutilizáveis. O uso de

componentes reutilizáveis, derivados de padrões de análise, pode vir a ser uma forma ainda mais eficiente de desenvolvimento de aplicações de SIG. Por exemplo, para uma nova aplicação na área de controle ambiental, além de reutilizar um pré esquema do seu banco de dados, juntamente com seu comportamento definido através de operações/métodos, o projetista poderá obter toda a base de dados geo-espaciais, caso já tenha sido desenvolvida outra aplicação para a mesma área geográfica. No domínio das aplicações urbanas, por exemplo, essa possibilidade é ainda maior, devido ao grande número de aplicações que podem ser desenvolvidas para a região geográfica compreendendo os limites de uma cidade;

- desenvolvimento de catálogos de padrões de análise de forma a torná-los acessíveis ao maior número possível de usuários. Isto implica em pesquisas relacionadas à unificação de terminologias, tanto com o objetivo de facilitar o uso dos padrões como para facilitar a identificação de padrões existentes, mas que foram descritos usando termos diferentes daqueles que o usuário interessado está habituado a utilizar. Uma possibilidade é a integração dos mecanismos para disponibilizar padrões de análise para banco de dados geográficos juntamente com os mecanismos existentes que disponibilizam metadados através de um sistema de *Clearinghouse* [WEB 98, VIL 99];
- extensão do *framework* conceitual GeoFrame e conseqüente adaptação da linguagem UML para melhor representar os requisitos que ainda apresentam deficiências como, por exemplo, os aspectos temporais, de qualidade e relacionamentos espaciais. Além disso, é necessário o desenvolvimento de pesquisas que abordem a modelagem dos aspectos dinâmicos e funcionais das aplicações de SIG;
- desenvolvimento de uma metodologia para aplicação de engenharia reversa de aplicações de SIG, baseada no GeoFrame.
- desenvolvimento de ferramentas CASE para suporte à modelagem conceitual, com base no GeoFrame, usando a linguagem UML. Investigação e aprimoramento de regras para o mapeamento entre um esquema conceitual e uma implementação em um SIG;
- investigação no sentido de integrar esquemas conceituais elaborados com base no GeoFrame com seus respectivos modelos lógicos descritos, por exemplo, através do modelo OpenGIS [OPE 98].

Anexo 1

Artigo XII: Geospatial Metadata in Brazil: an experience in data documentation of an environmental GIS application

Eliseu J. Weber*, Jugurta Lisboa Filho, Cirano Iochpe e Heinrich Hasenack*

*UFRGS - Centro de Ecologia/Centro de Recursos Idrisi

In: INT. CONFERENCE & EXHIBITION ON GEOGRAPHIC INFORMATION - GISPlanet, 1998, Lisbon, Portugal. **Proceedings...** Lisbon: USIG, 1998.

ABSTRACT

This paper presents an experience in documentation of a spatial data set related to an environmental GIS application. The aim of this work is to analyze the advantages and problems of using existing standards in creating metadata to help Brazilian institutions sharing their geospatial data sets. Several standards were identified and analyzed with respect to their completeness, ease of use, available documentation, and availability of software tools. An example of the use of FGDC metadata standard is shown and advantages and disadvantages are discussed. Results show that the FGDC standard is a good alternative for supplying Brazilian documentation demands.

KEYWORDS: Metadata, Data Quality, Environmental Applications, Geospatial Data

1 INTRODUCTION

Mankind have used spatial information for a long time, having perceived its value in a series of important activities as navigation, land demarcation, military and commercial strategy. Initially, the spatial information was handled as manuscript or printed maps. Since the 60's the research on systems that work with geospatial information, later called Geographic Information Systems (GIS), has been done in parallel with the computer development [ANT 91]. In the 80's the development of GIS received a significant stimulus, giving rise to applications in a broad knowledge area, in public and private sectors. In the late 80's, GIS production achieved a strong commercial aspect.

From that date on an enormous amount of spatial data of different regions started to be stored in digital media. The high cost of structuring a database and the facility to reproduce digital data demanded the needs to avoid the duplication of efforts in the production of data. With this purpose, in many countries the production and distribution of digital spatial data were standardized and regulated [BUR 86].

Besides cost reduction, the standardized production brought as an additional advantage the fact that all the applications carried on one determined region use the same geographic database and can at any moment be adjoined or integrated. However, a new gap, related to the knowledge on the characteristics of the available digital data, could be identified.

Information that describes data is called metadata [SHE 97], and the information given by them represents a substantial economy in research resources and in the search for available geospatial data, as well as in the monitoring, control and acquisition. Metadata information allows the user to know where data is located, its accessibility and fitness to a specific application.

In the attempt to organize the metadata to facilitate the location and the access to the data, many proposals of metadata standards appeared, establishing categories and items to be filled. Currently, with the facility of data distribution through the Internet, the concept geospatial metadata has been revised under the concepts of interoperability, sharing, availability, search and transfer of data [TIM 96].

In Brazil, research with geospatial metadata is important because one strong demand for digital geospatial databases now exists. The contact with GIS technology began at the end of 80's and, passed one decade of the initial contact, many local and regional applications that need digital geospatial databases are in course. In every meeting of specialists and users in the field of geoprocessing, some attention is given to the problems related to the production and distribution of data. The advantage of creating national scope mechanism to assist institutions in the sharing of geospatial data has been discussed in the last years [RIB 96]. However, few concrete results have been achieved.

Looking for actions in this direction, a cooperation between Universidade Federal do Rio Grande do Sul (UFRGS) and the government of the State Rio Grande do Sul was signed. The cooperation involves many institutions, like the Fundação Estadual de Proteção Ambiental (FEPAM), in order to create a geospatial information system to support institutions which use GIS, schools and also to help individual users to find existing geospatial data.

This study describes the experience and the difficulties found in the documentation of a geospatial database, developed for a real application in the field of environmental management, in which different existing metadata standards have been analyzed.

2 ANALYSIS OF METADATA STANDARDS

Aiming to perform a practical experience with metadata production, a comparison of the following standards was done: SDTS (Spatial Data Transfer Standard) [DIG 88], CSDGM (Content Standards for Digital Geospatial Metadata) [FED 97], SAIF (Spatial Archive and Interchange Format) [BRI 97] and CEN (European Committee for Standardization) [CEN 96]. The study was based on the analysis of the documentation of the standard and on the available software tools to implement metadata in each of the standards.

The criteria used for the documentation analysis of each of the standards were:

- standard proposal: aim for which the standard was conceived;
- groups of metadata embraced by each of the standard and level of detail of the elements, especially of quality and spatial reference;
- if the standard includes data and metadata models;
- the possibility of including information about the metadata itself.

The criteria used for the analysis of the software tools available were:

- refinement;
- ease of use;
- operational system needed for running the software;
- availability of examples on produced metadata;
- formats for making metadata available.

For each standard the software systems below, downloaded from the Internet, were compared:

- STDS: SDTS; SDTS-EP Standard (Spatial Data Transfer Standard Encoding Program);
- CSDGM: Corpsmet 95 and NBII Metamaker 2.10;
- SAIF: SaifSuite1.1; FMEBC (Feature Manipulation Engine (BC)), Saif3.1 and Saif3.2;
- CEN: ST-Developer_1.5;

The chosen standard for the registration of metadata was the CSDGM, developed by the US Federal Geographic Data Committee. The selected software was the NBII Metamaker version 2.10, available through anonymous FTP at fgdc.er.usgs.gov, of the United States National Biological Service.

The CSDGM specifies the content of metadata information for a digital geospatial data set. The objective of the standard is to provide a common set of terminology and definitions for the documentation of these metadata. It establishes the name of the data elements and groups of data elements to be used to exchange data. Information on terms which are mandatory, mandatory under some conditions and optional is provided in the standard [FED 97].

The information included in the standard is based on four functions which the metadata generally should fulfill:

- *availability*: data needed to determine the data sets that exist for a geographic location;
- *fitness for use*: data needed to determine if a data set meets a specific need;
- *access*: data needed to acquire an identified data set;
- *transfer*: data needed to process and use a data set.

The CSDGM defines elements of metadata for the following general topics:

- Identification: basic information on the data set. Includes title, geographical extension, currentness and routines to be followed to acquire or to use data;
- Data quality: an evaluation of the data set quality. It includes positional and attribute accuracy, completeness, consistency, sources and methods information and procedures used to generate the data set;
- Organization of spatial data: the mechanism used to represent spatial information in the data set. It includes the methods used to directly represent spatial positions (like raster and vector), indirectly (like street addresses and zip codes) and the number of spatial objects in the data set);
- Spatial reference: description of the reference frame and ways to code coordinates in the data set. Includes projection name and parameters for map projections or coordinate grid systems, vertical and horizontal datum and resolution of the coordinate system;
- Entity and attribute: information on the data set content, including entity type and its attributes and domains of the attribute values. Includes the names and feature definitions, attributes and attribute values;

- Distribution: information on data acquisition. Includes contact with distributor, available formats, information on how to obtain data sets on-line, by mail and the cost of the data;
- Metadata reference: information on database currentness and responsibility on it.

In terms of use, the FGDC rules have been used as reference to the development of almost all of the other standards proposed up to now. There are bases of metadata in various regions of the world, mainly in the USA. In April 1994, the Executive Order (12096) *Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure* instructed all American federal agencies to use the CSDGM to document new geospatial data after 1995, and to make metadata available to the public through the *National Geospatial Data Clearinghouse*.

The clearinghouse is a distributed system for make spatial data available on the Internet through metadata [PET 98]. Using the data elements defined in the CSDGM, any institution worldwide can make their collections of spatial information searchable and accessible on the Internet using free reference implementation software developed by the FGDC. A Clearinghouse Node can be established wherever one has a collection of data and metadata to describe them. Material for technical support and development of the Metadata are provided through the Internet by the FGDC.

3 CASE STUDY

Identification of the Study Area

The study area is the Caí watershed, one of the five watersheds which form the Guaíba watershed (Figure 1). This watershed has a total area of 85,900 km² and around 6 million people living in 227 townships of the state Rio Grande do Sul. The Brazilian government, with aid of the World Bank is running a program called Pro-Guaíba, which aims at reducing water pollution [BIN 98, LIS 97b]. The Caí watershed measures 5,057.25 km² and estimated population of 383,929 inhabitants distributed in 41 township among latitude coordinates 29°06'S and 30°S and longitude coordinates 50°24'W and 51°40'W [AGR 71].

The GIS application in the Caí watershed consists of the implementation of an environmental information system with cartographic, thematic and descriptive data. Its aim was to survey the uses of water and soil all over the Caí, recollecting subsidies to a river segmentation proposal, as well as to identify conflicts about water and soil use in the watershed. It should contemplate the first step of the framing process of the water resources defined in the CONAMA Resolution 20/86 [CON 93]. The framing of the water resources is an important management instrument, since it establishes restrictions and potentials, defining objectives of water quality to be reached or maintained, based on the main uses of the water in each of the segments of the main river of the analyzed watershed [RGS 94].

FEPAM, in cooperation with GTZ (German Technical Cooperation Agency), in an unknown initiative, is now structuring the environmental information system of the Caí watershed to aid its framing and management [WEB 98b]. Like the main existing applications in Brazil, the structuring of the system was done by specialists in different application areas directly in a GIS software, IDRISI for Windows in this case, without neither a previous conceptual data modeling step nor a previous specification of which information would be recollecting about the used data.

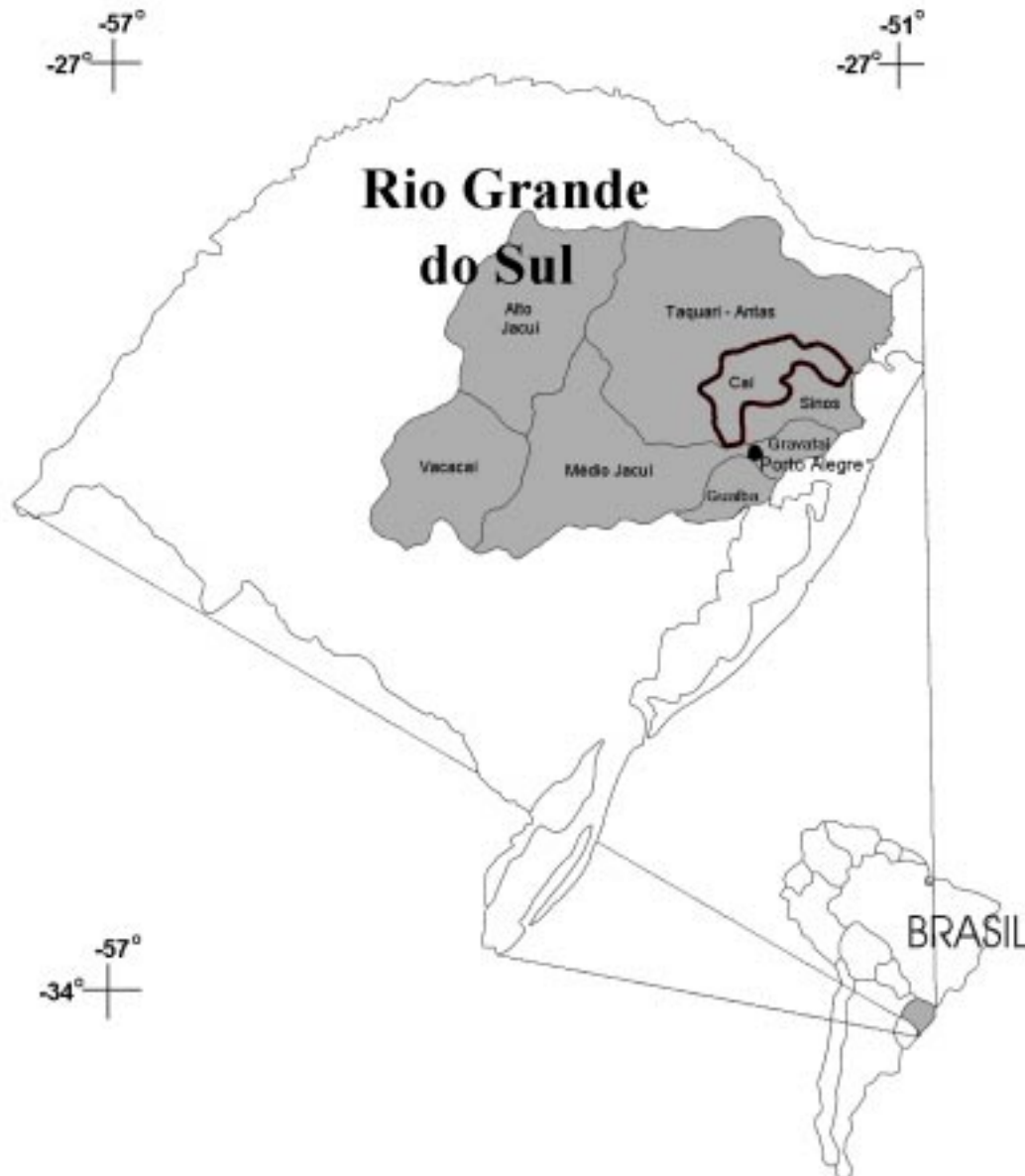


FIGURE 1 -The Caí watershed in the Rio Grande do Sul State and in the Pró-Guaíba program area.

Registering Metadata

Since the aim of the study is to evaluate the chosen standard, metadata for a data subset of the Caí watershed application were generated. The documented data were:

- Classification of satellite imagery with different soil cover classes in the watershed (raster);
- Hydrographic map (vector lines);
- Township boundaries in scale 1:250,000 of the IBGE (Brazilian Institute for Geography and Statistics, vector polygons);
- Map with the location of solid waste deposits (points);
- Attribute data associated to each deposit of solid waste (Access database table).

Metadata were surveyed based on the descriptive information given by the organizations which produce the respective data as well as by the procedure used to structure the data into the GIS.

After registering, the metadata were used to train the build up process of a clearinghouse node, using software available by the FGDC through the Internet. The steps developed included the exportation of the generated metadata by the software, its indexing in the clearinghouse node, specific consistency tests and the installation of a metadata server software.

4 RESULTS

The practical experience on recovering and registering metadata for a real GIS application allowed the identification of aspects of great value to the implementation of metadata in Brazil, especially related to difficulties concerning its generation.

The difficulties found in the test point out two basically distinct problems: (1) recovering of the information on the data (metadata) and (2) effective use of the standard.

As it generally happens, when the documentation of the data is made after its production, a considerable effort is necessary to recoup the needed information, and it will not always be possible to recoup it integrally. The registered data have very diverse origins and it does not have full information and cannot be trusted, reason for which the register of some items must be ignored. The greatest problem, in fact, was the difficulty to get the values to fill the elements of quantitative metadata, especially the quality metadata. Paradoxically, quality metadata are extremely important to the user who, after locating the existing data for the region of interest, needs to decide if they satisfy or not his/her needs.

Concerning the used standard (CSDGM), at a first moment, it was somewhat difficult to dominate its terminology, to know exactly what each item of the extensive stack of items represents and how to fill it in. This difficulty is in part due to the little experience people have with metadata in Brazil. However, this problem is easily overcome as more familiarity with the standard is achieved, and as more reference material is reviewed.

5 FINAL CONSIDERATIONS

The difficulties found in this case study related to the recovering of information on already produced data and to the use of the metadata standard deserve attention. They tend to be bigger in a systematic process of documentation, where the person making the register work probably had little contact with the geospatial data that being registered and is not always a specialist in the field of geospatial data or metadata.

The generation of metadata concomitantly to the production/digitalization of the data prevents a series of problems of the posterior documentation, with reduction of costs and greater allegiance of the information. In a country like Brazil, with very recent cartographic history, this aspect is much more critical. Therefore the importance of many referring details to the data not receiving the due attention during its production. Furthermore, the measurement methods for the various metadata elements are not well dominated by most of the people and norms are lacking. Consequently, it is important to deepen the discussion and research on metadata in Brazil in all these areas.

To recommend its adoption in Brazil, a metadata standard must be at the same time sufficiently inclusive, flexible to allow an adequate description of a data set and as simple as possible to allow its easy handling by producers and users of data. Although

presenting some limitations, the CSDMG presents some positive points, which makes it appropriate to be adopted as a metadata standard in Brazil. The CSDMG was the base for the development of other metadata standards, and is widely used in the US and has a growing acceptance in other countries (South Africa, for example). There is already a significant number of users, a large amount of metadata already registered with this standard and easy to use software available to implement metadata according to this standard.

Nevertheless, considering that the adoption of one standard implies the assignment of systematic tasks on the metadata generation to people with weak technical qualification, the existing software language represents an obstacle to take into consideration. There is a need to translate English menus to the desired language (Portuguese, in Brazil), in order to facilitate its adoption.

Related to setting up the clearinghouse, the experience with metadata allowed acquire the knowledge for the further establishment of a real clearinghouse node, to make accessible data coming from other environmental GIS-application, related to the extraction and use of coal in the state of Rio Grande do Sul, Brazil.

ACKNOWLEDGMENTS

To RHAEC/CNPq Research Program, for the fellowship; to GTZ-FEPAM Cooperation Program, on whose application data documentation was based on; to the fellowship holder Roni Anzolch for his contributions to the study; to the graduate student in Computer Science Andréia Castro Costa, for the research of software and metadata in the Internet.

Anexo 2

Artigo XIII: Um servidor de consultas à *clearinghouse* em português para distribuição de metadados geográficos

Fábio Villavicencio*, Cirano Iochpe e Jugurta Lisboa Filho

*UFRGS - Bacharelado em Ciência da Computação

In: GISBRASIL - CONGRESSO E FEIRA PARA USUÁRIOS DE GEOPROCESSAMENTO DA AMÉRICA LATINA, 5, 1999, Salvador. **Anais...** Salvador: Fator GIS, 1999.

RESUMO

Um dos principais problemas para o compartilhamento de dados geográficos, utilizados em Sistemas de Informação Geográfica, é a falta de divulgação sobre os dados existentes. Metadados, ou seja, dados sobre dados, têm sido propostos como um meio para permitir a localização de conjuntos de dados geográficos existentes. Para favorecer uma ampla utilização de metadados geográficos, vários padrões de metadados têm sido propostos em diversos países, sendo que o padrão americano do *Federal Geographic Data Committee* (FGDC) mostrou-se um forte candidato para ser adotado pela comunidade de geoprocessamento no Brasil [WEB98].

O FGDC disponibiliza um sistema de *Clearinghouse*, que é um sistema de armazenamento distribuído de metadados, através do qual os metadados são disponibilizados na Internet para busca, acesso e avaliação.

Considerando que, normalmente, o usuário brasileiro possui limitações em utilizar programas com interface em inglês, o principal obstáculo à real utilização do sistema de *Clearinghouse*, no Brasil, é a barreira idiomática.

Este trabalho descreve a implementação de um protótipo de servidor de consultas a metadados em português, desenvolvido no Instituto de Informática da UFRGS, mostrando as diferentes alternativas para sua implementação e as vantagens e desvantagens de cada uma. São descritos, também, os passos necessários à implantação de um nodo de *Clearinghouse*.

ABSTRACT

One of the main problems for sharing geographic data, which are used in Geographic Information Systems (GIS), is the spreading lack on the existing data. Metadata, i.e., data about data, have been considered as a way to allow the location of existing geographic data sets. To easy a wide use of geographic metadata, some metadata standards have been proposed in several countries. The American standard of

Federal Geographic Data Committee (FGDC) revealed be a strong candidate to adoption by the Brazilian GIS community [WEB98].

The FGDC developed a system of Clearinghouse, i.e., a system of distributed storage metadata, through which the metadata are put in the Internet for search, access and evaluation.

Considering that normally the Brazilian users have limitations in using programs with English interface, the main obstacle to the Brazilian Clearinghouse system, is the idiomatic barrier.

This paper describes the implementation of a prototype of metadata query server in Portuguese, developed in the Instituto de Informática/UFRGS (Computer Science Institute at Federal University of Rio Grande do Sul), presents the different alternatives of implementation and the advantages and disadvantages of each one. There are also described the necessary steps to the implantation of a Clearinghouse node.

1 INTRODUÇÃO

A necessidade crescente de informações cada vez mais precisas sobre dados geográficos, especialmente por órgãos de planejamento urbano e ambiental, tem impulsionado o desenvolvimento de Sistemas de Informação Geográfica (SIG). Segundo Lisboa [LIS97], estes sistemas manipulam dados obtidos de diversas formas como, por exemplo, mapeamento por *Global Positioning System*, GPS, digitalização de mapas e imagens de satélite por *scanner*, armazenando esses dados em arquivos de vários formatos.

A aquisição dos dados geográficos é, normalmente, bastante custosa, uma vez que os dados geográficos são, muitas vezes, coletados manualmente e necessitam tratamento pré-armazenagem. A coleta manual se faz através de digitalizações de mapas e mapeamentos de regiões com GPS, sendo o principal componente do custo de implantação de um SIG [ARO89].

Para evitar que os dados de uma determinada região sejam coletados mais de uma vez por instituições diferentes, gastando muitos recursos (tempo e pessoal qualificado), é desejável que exista uma forma de cessão dos dados, ou mesmo o comércio destes entre as instituições usuárias.

Para que possa existir a troca de dados geográficos entre instituições usuárias, deve-se conhecer informações sobre a qualidade dos dados, sua origem, a área coberta por eles, a sua data de aquisição e várias outras informações sobre o conjunto de dados geográficos disponíveis. Estas informações sobre os dados recebem o nome de metadados geográficos.

Através da consulta aos metadados de um conjunto de dados geográficos, a entidade usuária pode decidir se é conveniente a obtenção dos dados já prontos, se é melhor obter os dados através de coleta, ou mesmo saber se os dados que possui estão atualizados ou não.

Surge, então, o problema da disponibilização dos metadados a todos que tenham interesse em consultá-los. Para resolver este problema, o *Federal Geographic Data Committee*, FGDC [FED97], impulsionado por uma resolução do governo dos Estados Unidos que obriga os órgãos públicos a disponibilizarem informações a qualquer pessoa, sem ônus para o requerente, criou o sistema *Clearinghouse* [NEB 96], que consiste em uma rede de servidores contendo metadados. Essa rede conta com alguns servidores de consultas, que permitem que qualquer pessoa com acesso à Internet possa consultar os servidores com metadados através de formulários de consulta. O número de servidores de consulta é pequeno, sendo atualmente apenas três, para a demanda

existente de consultas, o que se traduz em um tempo de resposta da ordem de minutos em horários de maior congestionamento da rede.

Entre os principais usuários dos SIG estão os órgãos públicos e privados que trabalham com planejamento, meio-ambiente, limpeza urbana e estradas, entre outros, que necessitam de dados geográficos para o planejamento e execução das suas atividades. Por exemplo, o departamento de limpeza urbana necessita saber quais são e onde estão localizados os logradouros de uma cidade, para poder determinar quais serão os pontos de coleta de lixo. A mesma informação sobre os logradouros é necessária para que a secretaria dos transportes programe os itinerários de linhas de ônibus.

O intercâmbio dos dados geográficos no Brasil tem ocorrido de um modo muito precário, através da troca de dados digitais contendo apenas os dados geográficos, mas não os metadados que descrevessem a qualidade desses conjuntos de dados. A consequência imediata deste processo é a falta de controle sobre os dados utilizados, desconhecendo-se os objetivos para os quais foram coletados.

Com a necessidade de controlar melhor os dados que estavam sendo utilizados, algumas entidades usuárias começaram a desenvolver um método próprio de controle, que consistia em conjuntos de metadados com um número de informações bastante reduzido. Esses conjuntos continham apenas informações sobre a área de abrangência e a origem dos dados, não contemplando informações sobre autoria, disponibilidade dos dados e descrições mais detalhadas sobre o seu conteúdo.

Nos Estados Unidos, o FGDC definiu um padrão de metadados geográficos que abrange uma descrição bastante completa dos dados. Uma solução para resolver o problema da falta de informações sobre os dados geográficos no Brasil é a adoção do padrão de metadados do FGDC, que além de ter sido adotado pelo governo dos Estados Unidos, conta com grande aceitação em outros países como o Canadá, Austrália e África do Sul, todos eles oficialmente de língua inglesa. Para a disponibilização dos metadados, os órgãos públicos e privados destes países utilizam o sistema de *Clearinghouse*.

Apesar de o idioma inglês ser a língua padrão de fato da Internet, fora do meio acadêmico, são poucas as pessoas que dominam esse idioma. Como o usuário brasileiro típico do sistema de consulta à *Clearinghouse* será, normalmente, uma pessoa de nível de escolaridade médio, com nenhum ou pouco conhecimento desse idioma, um sistema de consulta em língua inglesa pode representar uma séria barreira para o usuário e, conseqüentemente, para o sucesso do sistema, em relação a seu público alvo.

Este artigo mostra as diversas alternativas de solução para o problema do idioma, discutindo suas vantagens e desvantagens. É descrita a implementação de um protótipo de servidor de consultas à *Clearinghouse*, desenvolvido no instituto de Informática da UFRGS, com uma interface em português, utilizando uma das soluções apresentadas. Dois requisitos importantes para o sucesso do sistema é que ele possua um tempo de resposta, no Brasil, inferior à média obtida em testes com um servidor de consulta do FGDC, e mantenha a compatibilidade com o sistema em inglês já existente.

Para o desenvolvimento deste trabalho, inicialmente, foi implantado um nodo de *Clearinghouse* cadastrado na rede do FGDC, contendo conjuntos de metadados reais obtidos em cooperação com o Centro de Ecologia da UFRGS. Em seguida, foi desenvolvido o servidor de consultas em português, baseado em documentos e ferramentas fornecidas pelo FGDC e outras fontes disponíveis na Internet.

2 O SISTEMA CLEARINGHOUSE

Metadados são dados sobre dados [WEB98]. Metadados geográficos são dados a respeito de conjuntos de dados geográficos disponíveis em meio digital.

O sistema *Clearinghouse* ou, simplesmente, *Clearinghouse*, constitui-se em um conjunto de servidores de metadados [NEB 96]. Atualmente constitui-se de cerca de 90 servidores, cujo acesso se faz através de um servidor de consulta, como, por exemplo, o MetaStar Gateway [MET98].

Cada servidor de metadados possui um ou mais conjuntos de metadados indexados segundo regras definidas no protocolo de nome Z39.50 [ACT98] e é completamente independente dos demais, o que significa que, no caso de uma falha em algum servidor, o sistema *Clearinghouse* continua funcionando normalmente, com exceção do servidor que falhou.

O servidor de consultas, também chamado *gateway*, atua como cliente dos servidores de metadados. O servidor recebe uma consulta de um cliente, que especifica qual o nome do servidor a ser consultado, qual é o banco de metadados e qual é o critério de busca, interroga os servidores de metadados envolvidos e recebe a resposta, devolvendo-a ao cliente que fez a consulta. A Figura 2.1 mostra a topologia do sistema.

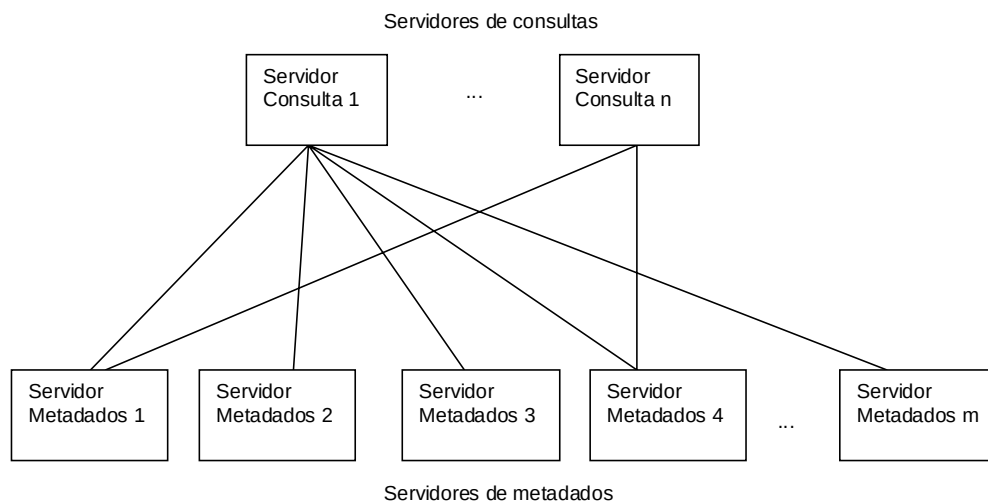


FIGURA 2.1 – Topologia do Sistema

2.1 O protocolo Z39.50

A comunicação entre o servidor de consultas (cliente dos servidores de metadados) e os servidores de metadados é feita através da Internet, utilizando-se um protocolo para buscas em bancos de dados chamado Z39.50 [ACT98].

O protocolo Z39.50 é bastante popular na comunidade científica, tendo sido concebido para realizar consultas a catálogos de bibliotecas. Como as bibliotecas possuem até hoje catálogos organizados de diferentes maneiras, o protocolo foi concebido para suportar dados armazenados de diferentes modos, porém, com algumas semelhanças. Por causa da sua robustez, com o passar do tempo, o Z39.50 foi se popularizando, tendo-se difundido a sua utilização para a indexação e consulta de textos estruturados quaisquer.

Através de um conjunto de parâmetros denominados perfil, o Z39.50 é informado da estrutura do texto a ser indexado e sobre quais campos pode-se realizar as consultas. Por exemplo, pode-se definir um perfil chamado *documento*, que especifica que todos os seus documentos deverão ter um título, um assunto e um corpo. Dentro deste *documento*, define-se que apenas os campos assunto e título poderão ser pesquisados. Então, o indexador criará apontadores apenas para as palavras do assunto e do título dos documentos que seguirem este perfil.

Para adequar o Z39.50 às consultas sobre metadados geográficos, o FGDC criou um perfil chamado GEO, onde a estrutura do conjunto de metadados é descrita.

Diversas implementações do protocolo Z39.50 estão disponíveis para diversas plataformas de hardware e sistemas operacionais, como Linux, Solaris, Windows NT, entre outros. O FGDC distribui, gratuitamente, algumas dessas implementações com o perfil GEO a todos os interessados em montar um nodo de *Clearinghouse*.

2.2 Consultas a metadados em um nodo de *Clearinghouse*

As consultas a um nodo de *Clearinghouse* são divididas em várias etapas. Na primeira etapa, o usuário que deseja fazer uma consulta comunica-se com um servidor de consultas. No sistema *Clearinghouse*, o modo de comunicação entre o usuário e o servidor de consultas é por meio de um navegador HTML. O usuário entra na página de consultas e, após informar quais os bancos de metadados e os critérios de busca, submete o seu pedido, passo 1 da Figura 2.2.

A partir deste momento, o servidor de consultas envia a pesquisa para todos os servidores de *Clearinghouse* envolvidos na busca e fica aguardando a resposta de cada um (passo 2). Na terceira etapa, cada servidor de *Clearinghouse* responde ao servidor de consultas, informando quais foram os conjuntos de metadados que satisfizeram os critérios de busca (passo 3). Em seguida, o servidor de consultas envia ao usuário a resposta de todos os conjuntos de metadados que satisfizeram a consulta por ele submetida (passo 4).

Neste ponto, o usuário pode dar-se por satisfeito e simplesmente terminar a conexão com o servidor, ou ainda, informar sobre qual dos conjuntos de metadados retornados ele deseja obter maiores informações. Caso a segunda opção seja a escolhida, um novo pedido é feito ao servidor de consultas (passo 5). Na etapa seis, o servidor de consultas envia um novo pedido ao servidor de *Clearinghouse* que contém o conjunto de metadados escolhido, solicitando o texto completo do conjunto de metadados. O servidor de *Clearinghouse*, escolhido, envia, então, ao servidor de consultas o texto solicitado (passo 7). Por último, o servidor de consultas envia a resposta ao usuário solicitante, contendo o texto do conjunto de metadados solicitados (passo 8).

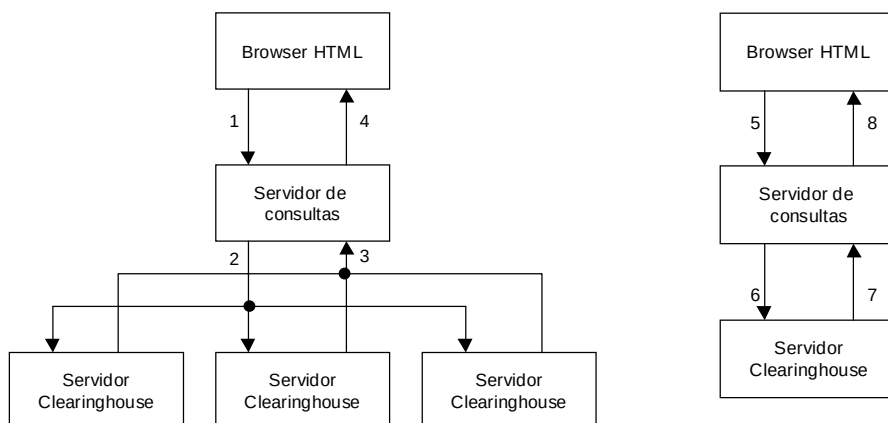


FIGURA 2.2 - Passos de uma consulta à Clearinghouse

2.3 Indexação e consultas dos conjuntos de metadados

A indexação dos metadados é necessária para que as consultas possam ser realizadas em um tempo razoável, uma vez que pesquisas exaustivas sobre os textos que um servidor de *Clearinghouse* pode armazenar são demasiadamente demoradas.

Para a indexação e consulta dos metadados e, de um modo geral, de todos os bancos de dados que utilizam o protocolo de buscas Z39.50, são utilizadas ferramentas de domínio público, disponíveis em vários locais na Internet.

Para este trabalho, foi escolhida a versão para Windows NT. Antes, porém, foram realizados alguns testes com a versão para Linux, que possuía alguns erros de programação, possibilitando apenas a indexação dos metadados, mas não a sua consulta. Uma vez que havia uma versão para Windows NT que funcionava adequadamente, optou-se por utilizá-la, já que vários testes com a ferramenta de consultas são necessários para a implantação de um nodo de *Clearinghouse*.

2.3.1 A ferramenta de indexação dos metadados

O indexador é chamado *lindex*, sendo possível a sua configuração com qualquer perfil desejado, bastando para isto, alterar-se seus arquivos de configuração. O *lindex* tem, como todas as outras ferramentas distribuídas pelo FGDC, uma interface de linha de comando, possuindo muitas maneiras, não compatíveis entre si, de indexar um texto, não sendo de utilização muito intuitiva.

Quando trabalha-se com metadados geográficos seguindo o padrão do FGDC é necessário configurá-lo com o perfil GEO. Este perfil vem junto com todas as versões das ferramentas para *Clearinghouse* distribuídas pelo FGDC, evitando-se, assim, o trabalho de definir os campos manualmente.

Ao ser utilizado para indexar algum documento, o indexador grava em diversos arquivos os índices necessários para o acesso ao texto segundo o perfil indicado. Por exemplo, o comando:

lindex -d db/testeall -t geo -m 8 -o fieldtype=geo.fields -o teste1 teste2

indica que o indexador deve criar o banco testeall, seguindo o padrão GEO, utilizando 8 Mb de memória e indexando os arquivos teste1 e teste2.

Como cada conjunto de metadados possui informações apenas sobre um único conjunto de dados geográficos, para tornar possível a consulta sobre vários conjuntos de metadados é necessário que eles sejam indexados ao mesmo tempo. No exemplo acima, esta característica foi levada em conta.

Uma outra característica importante da ferramenta de indexação é que ela não permite a atualização incremental, ou seja, toda vez que algum dos conjuntos de metadados que estiver indexado for alterado, mesmo que apenas uma palavra em um campo, é necessária uma nova indexação sobre todos eles. Uma nova indexação se faz necessária, também, para que algum conjunto de metadados seja incluído no banco indexado para consulta. O mesmo vale para a exclusão de um conjunto de metadados do banco indexado.

2.3.2 A ferramenta de consultas aos metadados

O programa para consultas sobre bancos de dados locais é chamado *Isearch*. Ele é relativamente simples de utilizar, se comparado ao *lindex*, tendo a capacidade de buscar palavras em posições determinadas do texto que foi indexado segundo o perfil indicado

no momento da indexação. Na consulta, não é necessário informar o Isearch sobre o perfil utilizado na indexação do texto, basta indicar a localização e o nome do banco de dados que se deseja consultar, além, é claro, dos critérios de busca.

Um exemplo de consulta válida para o Isearch é o seguinte:

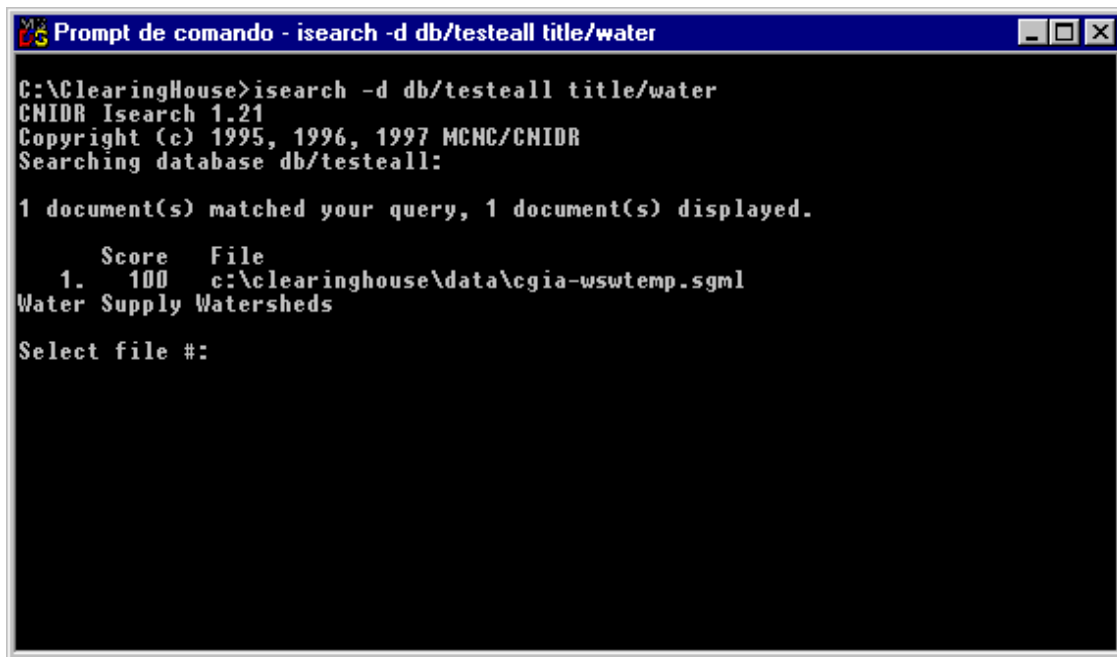
Isearch -d db/testeall title/water

onde **-d db/testeall** é o banco de dados a ser consultado, **title/water** significa procurar no título a palavra *water*. A execução da consulta é mostrada na Figura 2.3.

Outra consulta possível é:

Isearch -d db/testeall water

Nesta consulta, não está especificado em qual campo deverá ser procurada a palavra *water*, por isso, ela será procurada em todos os campos que foram indexados e em nenhum mais. Pode-se notar nas consultas acima, que não há nenhuma referência à localização do banco a ser consultado. Isto quer dizer que o banco deverá estar em um diretório correntemente mapeado pelo sistema operacional.



```
Prompt de comando - isearch -d db/testeall title/water

C:\ClearingHouse>isearch -d db/testeall title/water
CNIDR Isearch 1.21
Copyright (c) 1995, 1996, 1997 MCNC/CNIDR
Searching database db/testeall:

1 document(s) matched your query, 1 document(s) displayed.

      Score  File
    1.   100  c:\clearinghouse\data\cgia-wswtemp.sgml
Water Supply Watersheds

Select file #:
```

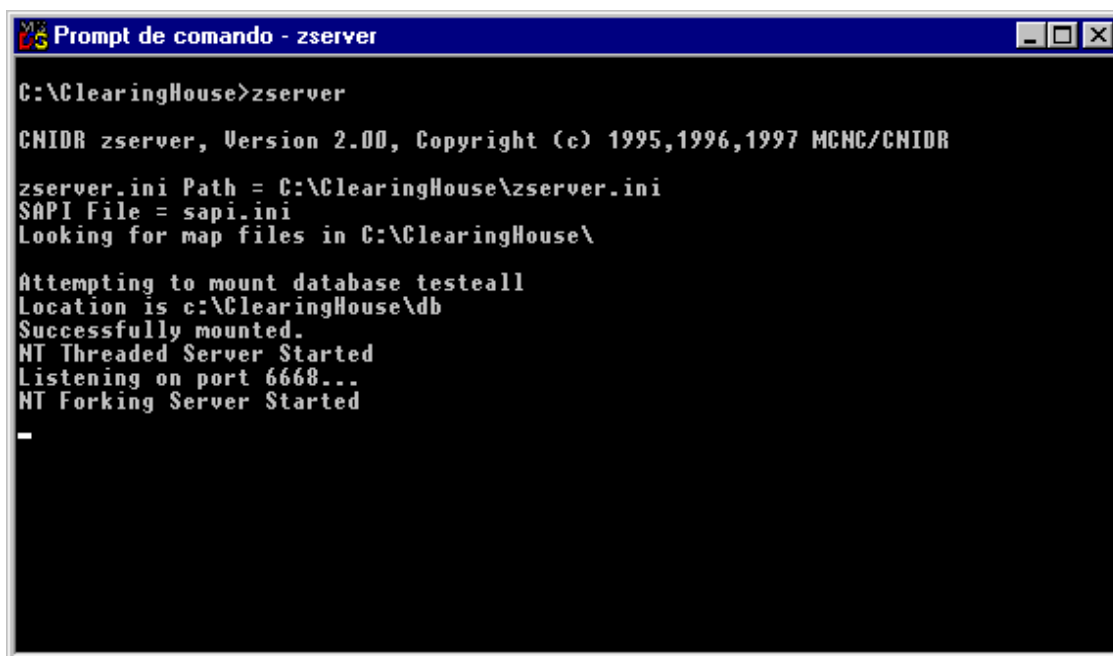
FIGURA 2.3 – Exemplo de consulta Isearch

2.3.3 Um servidor Z39.50

Juntamente com os programas de indexação e consulta local, é distribuído um servidor que segue o protocolo Z39.50. O servidor, chamado Zserver, no caso da *Clearinghouse*, é utilizado para responder às consultas submetidas pelos servidores de consultas.

Para que o Zserver funcione corretamente, é necessário que ele seja configurado com o perfil GEO e alguns outros parâmetros como, por exemplo, a porta de acesso e a localização do banco de metadados e índices. Para que as consultas possam ser efetuadas, o servidor deve ser executado na máquina que contém os conjuntos de metadados.

A execução do Zserver é mostrada na Figura 2.4, onde já aparece pronto para receber consultas sobre os bancos indexados. A inicialização do Zserver é bastante demorada, em comparação com o processo de indexação, sendo que na máquina em que foi testado, um PC 486 100MHz com 16Mb de memória RAM, levou cerca de oito minutos para ficar disponível para consultas. Depois que o servidor fica disponível, no entanto, ele consome poucos recursos da máquina em que está executando.



```
Prompt de comando - zserver

C:\ClearingHouse>zserver

CNIDR zserver, Version 2.00, Copyright (c) 1995,1996,1997 MCNC/CNIDR

zserver.ini Path = C:\ClearingHouse\zserver.ini
SAPI File = sapi.ini
Looking for map files in C:\ClearingHouse\

Attempting to mount database testeall
Location is c:\ClearingHouse\db
Successfully mounted.
NT Threaded Server Started
Listening on port 6668...
NT Forking Server Started
-
```

FIGURA 2.4 – Servidor Zserver ativo

2.3.4 Um cliente Z39.50

O cliente Z39.50, fornecido com as outras ferramentas no pacote para a construção de um nodo de *Clearinghouse*, serve, principalmente, para a realização de testes com o servidor. O programa é de utilização bastante simples, envolvendo poucos parâmetros e não necessitando de configuração em arquivos.

Para ser possível realizar testes com um conjunto de dados indexados, deve-se configurar o servidor e deixá-lo executando em uma máquina. Quando ele estiver disponível, esperando requisições de clientes, executa-se o programa Zclient, informando alguns parâmetros para a consulta como, por exemplo, o endereço de rede da máquina em que está o servidor, a porta a qual o servidor está monitorando, o banco de dados a ser consultado e o critério de consulta. Uma consulta válida é mostrada no exemplo a seguir:

Zclient 143.54.83.151 6668 testeall water

No exemplo, temos **143.54.83.151** como o endereço IP da máquina onde está o servidor e o banco com os conjuntos de metadados. Em seguida temos **6668**, que é o número da porta de acesso ao servidor Zserver, **testeall** é o nome do banco contendo os conjuntos de metadados que serão pesquisados e **water** é o critério de busca.

Na Figura 2.5 podemos visualizar a resposta para a consulta acima formulada.

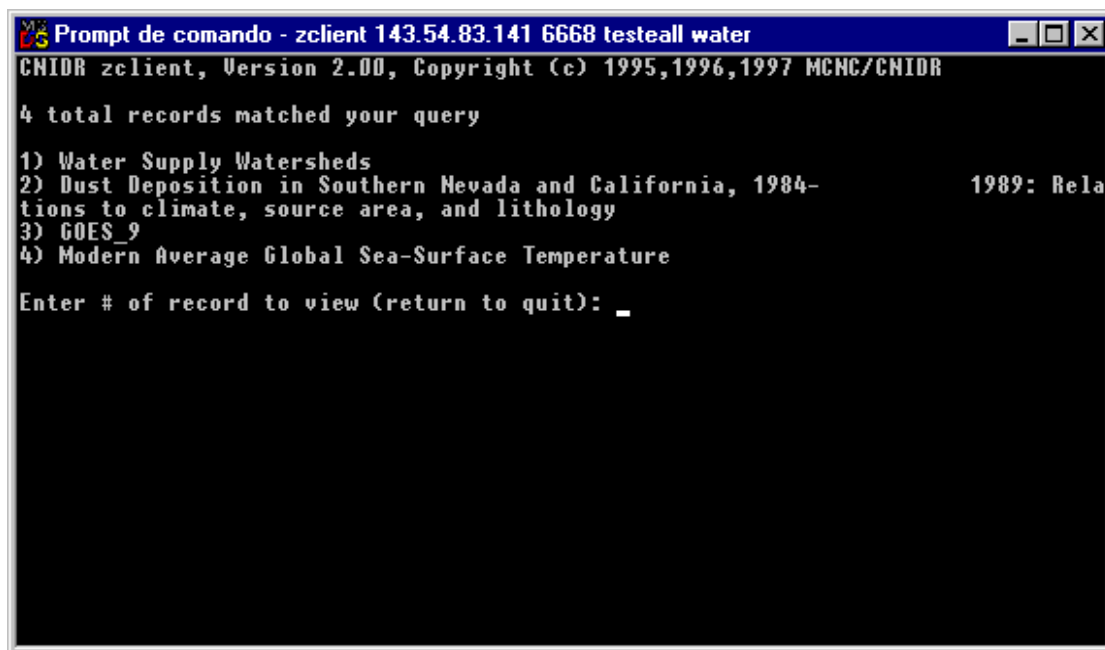


FIGURA 2.5 – Resultado de uma consulta Zclient

2.4 O servidor de consultas MetaStar

O servidor de consultas MetaStar [MET98] é o servidor de consultas oficial do FGDC. Através dele, cerca de 80% das consultas à *Clearinghouse* são realizadas. Por este motivo, um número muito elevado de mensagens se concentra em apenas um ponto da rede, o que significa um tempo de resposta relativamente elevado quando submete-se a ele uma consulta.

O MetaStar Gateway oferece um formulário para consultas bastante simples de utilizar, através de uma página web, conforme mostra a Figura 2.6. Basta selecionar alguns critérios de pesquisa, preencher alguns campos, submeter a consulta e ficar aguardando a resposta. Em alguns minutos a lista de bancos de dados que satisfazem os critérios de busca será mostrada.

As respostas para as consultas efetuadas através do MetaStar Gateway vêm na forma de uma lista com os bancos de dados que foram selecionados, informando em cada banco selecionado, quantos foram os resultados obtidos para a consulta efetuada.

A visualização dos títulos dos conjuntos de metadados, de um determinado banco de dados, que satisfizeram a consulta efetuada é possível quando o usuário seleciona de qual banco mostrado na tela ele quer os resultados.

Step 2 - Query Terms:

Field	Criteria	Boolean
Title	water	AND
Abstract	river	OR
Edition		AND
Originator		AND
Purpose		AND

Step 3 - Numeric Query Terms (Optional):

Field	Relation	Criteria	Boolean
Scale of Source	Equal To		AND
Cloud Cover	Equal To		

Step 4 - Select Databases:

- Africa Data Dissemination Service
- AGDC Alaska Geospatial Data Clearinghouse
- Australia - ERIN Environmental Data Directory
- Australia - ERIN Index

Document: Done

FIGURA 2.6 – Página de consultas MetaStar Gateway do FGDC

Somente após selecionar da lista de títulos de conjuntos de metadados qual é o desejado, é que o usuário pode ver na tela a listagem dos metadados daquele conjunto, conforme aparece na Figura 2.7.

Identification Information:

Citation:

Citation Information:

Originator: U.S. Geological Survey

Publication Date: 19970101

Title: Dekad Normalized Difference Vegetation Index (NDVI) for Africa - 1995

Edition: 1

Geospatial Data Presentation Form: remote-sensing image

Publication Information:

Publication Place: Sioux Falls, SD, USA

Publisher: EROS Data Center / International Program

Online Linkage: <URL: <http://edcent1.cr.usgs.gov/wadda/adds.html>>

Description:

Abstract:

The Normalized Difference Vegetation Index (NDVI) is a measure of the amount and vigor of vegetation at the surface. The magnitude of NDVI is related to the level of photosynthetic activity in the observed vegetation. In general, higher values of NDVI indicate greater vigor and amounts of vegetation. The NDVI is derived from data collected by the National Oceanic and Atmospheric Administration (NOAA) satellites, and processed by the Global Inventory Monitoring and Modeling Studies (GIMMS) at the National Aeronautics and Space Administration (NASA). The spatial resolution of the NDVI data is approximately 7 km. Vegetation indices derived from the (NOAA) Advanced Very High Resolution Radiometer (AVHRR) sensor have been employed for both qualitative and quantitative studies (e.g. Tucker et al. 1991 on the contraction and expansion of the Sahara desert, Sellers et al. 1995 on the calculation of biophysical

Document: Done

FIGURA 2.7 – Visualização de um conjunto de metadados selecionado a partir de uma consulta

2.5 Passos necessários para a implantação de um nodo de *Clearinghouse*

O processo de implantação de um nodo de *Clearinghouse* está bem definido pelo FGDC, existindo um tutorial específico para esta tarefa. Resumidamente, pode-se definir este processo como um seqüência de oito passos.

- Obter as ferramentas de indexação, consulta, o servidor Z39.50 e o cliente Zclient, fornecidos todos juntos em um pacote. Para obter-se os endereços de onde podem ser copiados os programas, deve-se acessar a página do FGDC [FED97].
- Preparar os arquivos de metadados, sendo necessários três formatos de arquivo diferentes para cada conjunto, quais sejam, HTML, SGML e texto simples. O HTML é utilizado para apresentação no *browser*, o texto simples para a apresentação na tela sem o uso do *browser* e o SGML para a indexação pelo *Index*.
- Indexar os arquivos com a ferramenta *Index* (Seção 2.3.1). Para ser realizada esta tarefa, o perfil deve ser informado ao indexador, assim como a localização e os nomes dos arquivos a indexar.
- Deve-se testar os arquivos de índice após a indexação para verificar se o passo anterior foi cumprido com êxito. Para isto, utiliza-se a ferramenta de consultas *Isearch* (Seção 2.3.2). Se tudo estiver funcionando corretamente será possível fazer pesquisas sobre os textos indexados, sendo possível visualizá-los na tela.
- Configurar o servidor *Zserver*, talvez a tarefa mais delicada do processo de implantação. Para a configuração do *Zserver*, existem arquivos de inicialização, que devem indicar, por exemplo, qual é o caminho para encontrar os arquivos de índice e qual o perfil a utilizar.
- Iniciar a execução do *Zserver* para possibilitar os testes com ele.
- Realizar testes com o servidor, utilizando-se para isto o cliente fornecido no pacote, o *Zclient*. O cliente deve ser informado do endereço do servidor de metadados, da porta a qual o servidor está escutando, o nome do banco de dados e o critério de pesquisa.
- Concluir a implantação do nodo de *Clearinghouse*, notificando o FGDC para que seja efetuado o cadastramento do servidor de consultas e do banco de metadados contido nele.

Tendo concluído todos os passos acima, o nodo de *Clearinghouse* está pronto para ser consultado.

3 ALTERNATIVAS PARA IMPLEMENTAÇÃO DE UM SERVIDOR DE CONSULTAS EM PORTUGUÊS

O público-alvo do servidor de consultas proposto é composto por pessoas que não dominam a língua inglesa, por isso deve ser construído com uma interface em português.

Para construir um servidor de consultas à *Clearinghouse*, há várias alternativas, tendo cada uma delas vantagens e desvantagens em relação às outras. Cada uma delas será descrita a seguir, com seus pontos favoráveis e desfavoráveis. No final, uma Tabela resumo é apresentada, indicando-se qual foi a solução escolhida.

3.1 Tradutor na máquina do usuário

A primeira alternativa consiste em alterarmos a topologia do sistema *Clearinghouse*, acrescentando um tradutor que fica entre o *browser* do usuário e o servidor de consultas em inglês, como mostrado na Figura 3.1.

O tradutor entraria em ação toda vez que o usuário dirigisse seu *browser* para a página do servidor de consultas em inglês, traduzindo campos de preenchimento pré-determinados da página de consultas e traduzindo, também, o nome de cada campo do conjunto de metadados retornado.

A principal vantagem de um sistema destes é a utilização de toda a estrutura já montada do sistema *Clearinghouse*, sem a necessidade de se cadastrar ou remover servidores em qualquer momento da operação do sistema. Outra vantagem, é que o usuário do sistema precisa lembrar-se de um só endereço para consultar qualquer página, tanto as em inglês quanto as em português.

Uma grande desvantagem, entretanto, é que o tradutor deveria ficar sempre executando na máquina do usuário, consumindo recursos desta máquina, necessitando ser copiado pela Internet e instalado, sem contar que o usuário que não possuir conhecimento suficiente pode encontrar problemas no momento da instalação.

Outra desvantagem significativa é que toda vez que a página de consultas mudasse, teríamos que atualizar o tradutor e distribuí-lo para todos novamente. Este problema é agravado pelo fato de os usuários de SIG utilizarem diferentes sistemas operacionais como, por exemplo, Windows e Unix, havendo a necessidade de manter atualizadas as versões para cada sistema operacional.

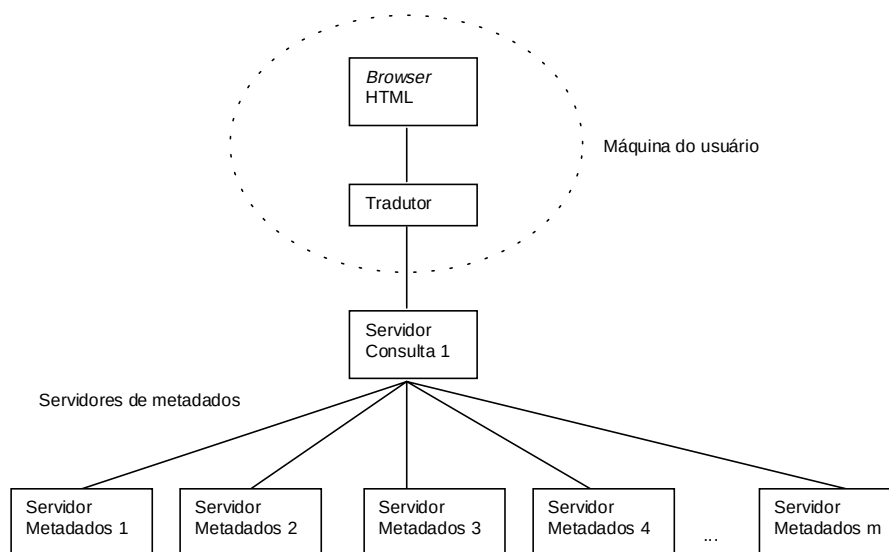


FIGURA 3.1 – Topologia do Sistema *Clearinghouse* com um tradutor para o português instalado na máquina do usuário

3.2 Pseudo servidor de consultas

Uma variação da solução anterior é a criação de um site acessível por *browser* contendo um tradutor. Este tradutor funcionaria como o do exemplo anterior, com uma diferença apenas: ficaria instalado em apenas uma máquina, podendo ser chamado de pseudo servidor de consultas, pois utiliza os serviços de outro servidor de consultas. A Figura 3.2 ilustra esta solução.

A utilização de um pseudo servidor é a mesma do que a utilização do servidor, do ponto de vista do usuário, já que ele somente digitará o endereço da página de consultas, embora tenha que lembrar de dois endereços distintos, um para fazer consultas em português e outro para as consultas em inglês.

A principal vantagem deste pseudo servidor é a melhora significativa da tarefa de manutenção, pois apesar de ser necessária a atualização do tradutor toda vez que a página de consultas mudar, não é mais necessário distribuir uma cópia do tradutor para cada usuário.

A grande desvantagem continua sendo o tempo de resposta muito elevado, o que poderia inviabilizar a utilização real desta solução, uma vez que todas as consultas serão repassadas para um servidor de consultas já sobrecarregado, concentrando ainda mais o tráfego da Internet.

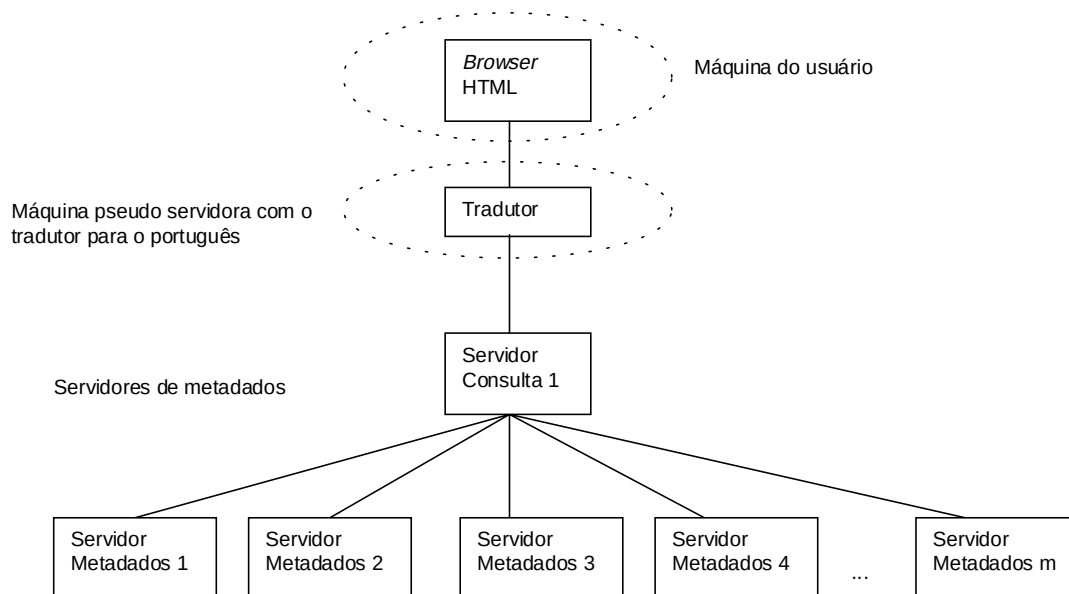


FIGURA 3.2 – Topologia do Sistema *Clearinghouse* com um pseudo servidor de consultas com um tradutor para o português

3.3 Servidor de consultas com tradutor

Uma terceira solução para o problema é deixar na máquina do usuário apenas o navegador WWW, construindo-se um servidor de consultas e um tradutor em outra máquina, conforme ilustra a Figura 3.3. Neste caso, o servidor de consultas construído comunicar-se-ia diretamente com os servidores de *Clearinghouse* selecionados na consulta, devolvendo para o usuário a resposta já traduzida.

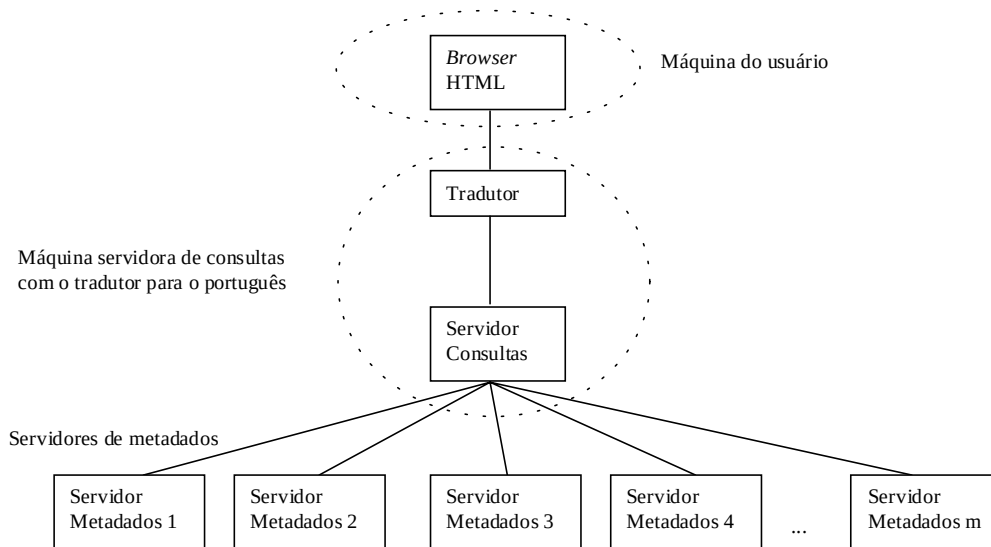


FIGURA 3.3 – Topologia do Sistema *Clearinghouse* com um servidor de consultas e um tradutor para o português

Esta abordagem tem algumas vantagens em relação às anteriores: a não necessidade de escrever um tradutor para cada sistema operacional; e a não necessidade de distribuir, a cada usuário, uma versão atualizada do tradutor. Entretanto, a maior vantagem desta solução em relação às outras é a diminuição considerável do tempo de resposta, fato conseguido graças ao redirecionamento do tráfego de mensagens na rede e à utilização de um servidor com menos carga de processamento.

Outro fato a considerar, enquanto o servidor for utilizado no Brasil, é a proximidade física do servidor em relação ao navegador do cliente, diminuindo o tempo de carga das páginas e reduzindo o tráfego intercontinental de mensagens por consulta.

Uma desvantagem que aparece no pseudo servidor e não no servidor de consultas com tradutor é a necessidade de alterar-se o tradutor a cada vez que a página de consultas em inglês for alterada. No caso do servidor de consultas com o tradutor, isto não ocorre, pois os servidores de *Clearinghouse* que serão consultados continuarão com o protocolo Z39.50, utilizando o perfil GEO.

As desvantagens desta solução, em relação às outras, também devem ser consideradas. A principal delas é a necessidade de manutenção da lista dos bancos disponíveis para consultas. Cada servidor de *Clearinghouse* novo ao qual desejar-se fazer consultas deverá ser acrescentado à lista, assim como os bancos que desistirem de fazer parte da *Clearinghouse*, deverão ser excluídos. Deve-se lembrar, contudo, que a desistência da *Clearinghouse* é um fato um tanto quanto raro.

4 A IMPLEMENTAÇÃO DE UM SERVIDOR DE CONSULTAS EM PORTUGUÊS

Para iniciar a implementação do servidor de consultas em português, foi necessário compreender a tecnologia envolvida no processo de distribuição dos metadados. Durante este processo, algumas decisões sobre a topologia de sistema a implementar e sobre as plataformas de hardware e sistema operacional tiveram que ser tomadas.

Após estudos sobre vantagens e desvantagens de cada alternativa de topologia possível para o servidor de consultas brasileiro, decidiu-se, principalmente por motivos de performance, utilizar a de servidor com tradutor, descrita na seção 3.2.

A adoção desta topologia permite flexibilidade na construção da página de consultas do servidor em português, que não necessita ser traduzida, uma vez que é escrita diretamente neste idioma. Não há, também, a desvantagem de precisar-se atualizar a página toda vez que um campo da página em inglês mudar.

Feita a escolha da topologia, havia a necessidade de definição da plataforma de hardware e sistema operacional. Como o grupo de pesquisa possui recursos escassos, a máquina a disposição para testes era um PC 486 DX4 100MHz com 16 megabytes de memória RAM e 800 megabytes de disco rígido. O sistema operacional inicialmente instalado foi o Linux, mas como os programas do pacote para este sistema estavam com problemas e o tempo era limitado, optou-se por utilizar o sistema operacional Windows NT 4.0 Workstation, para o qual os programas do pacote funcionavam a contento.

Simultaneamente à escolha do sistema operacional, optou-se por desenvolver o tradutor utilizando-se a linguagem Delphi 3 que é bastante fácil de aprender e, principalmente, conta com um suporte robusto para a programação utilizando Common Gateway Interface, CGI, e outros protocolos para a Internet.

Outra escolha foi a decisão de utilizar a ferramenta já pronta de cliente Z39.50, evitando programar novamente uma ferramenta. Esta escolha diminuiu sensivelmente o tempo de construção do servidor de consultas, algo que é muito importante na construção de qualquer sistema, embora isto signifique um tempo de resposta mais elevado em razão da comunicação entre os processos tradutor e servidor.

Tendo as principais decisões de projeto sido tomadas, a implementação do servidor em português passa pela obtenção de metadados, pela implantação do servidor de *Clearinghouse* para testes e pela programação do tradutor.

4.1 Detalhes da topologia do servidor de consultas em português

Antes de entrar na implementação do servidor em português é necessário ir um pouco mais a fundo na topologia do sistema. Na Figura 4.1, extraída da Figura 3.3, pode-se ver o servidor de consultas e o tradutor de um nível mais alto de abstração. Nesta macro visão, a entidade servidor de consultas é completamente independente do tradutor e comunica-se diretamente com o *browser* do usuário, o que, conforme será explicado a seguir, é um pouco mais complexo do que o mostrado aqui.

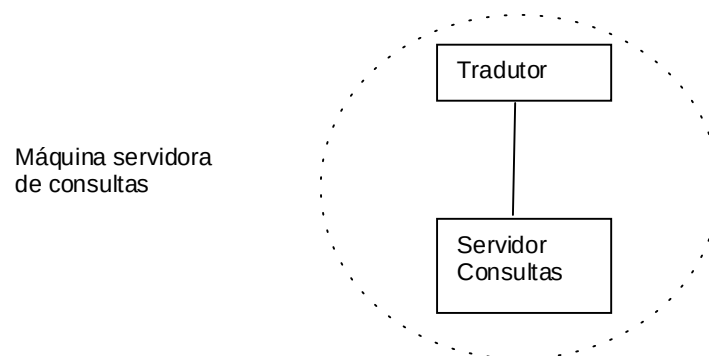


FIGURA 4.1 – Servidor de consultas e tradutor na máquina servidora

Na implementação, o servidor de consultas consiste em um servidor HTTP, um cliente Z39.50, e um programa que faz a ligação entre o formulário de consultas HTML e o cliente. Nesta implementação, tanto a ligação entre o servidor HTTP e o cliente quanto o tradutor inglês/português resultam em apenas um programa, programa este chamado, genericamente, de tradutor.

O servidor HTTP despacha as páginas HTML para o máquina do usuário e recebe os parâmetros da consulta efetuada pelo usuário. Os parâmetros da consulta são

transformados pelo tradutor em uma ou mais consultas no formato aceito pelo Zclient. Em seguida, o cliente Z39.50 submete a consulta para o servidor de *Clearinghouse* envolvido, aguardando a resposta durante um período de tempo determinado pelo usuário na consulta.

Uma vez respondida a consulta, o cliente Z39.50 repassa o resultado para o tradutor, que por sua vez, traduz os nomes dos campos de metadados para o português e, também, realiza a transformação do texto em uma página HTML, que é despachada pelo servidor HTTP até o *browser* do usuário. Esta topologia, vista de um nível mais baixo de abstração, é mostrada na Figura 4.2.

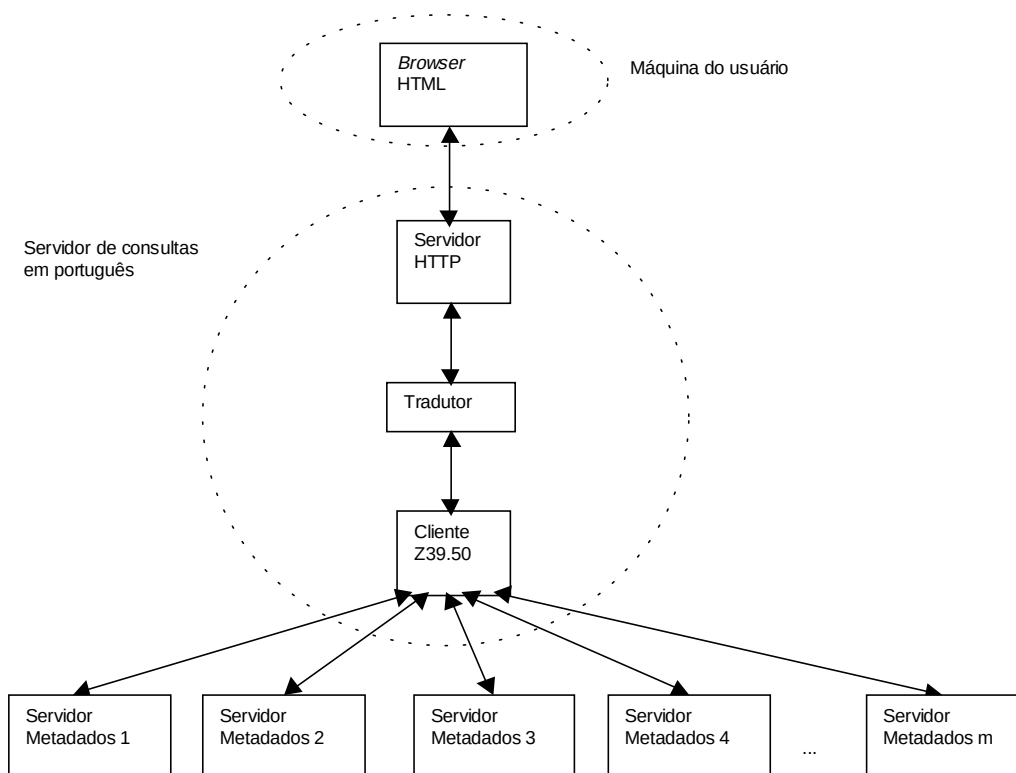


FIGURA 4.2 – Servidor de consultas e tradutor na máquina servidora vistos de um nível mais baixo de abstração

No caso de mais de um servidor de *Clearinghouse* ser selecionado para a consulta, uma instância do cliente Z39.50 é disparada para cada servidor envolvido, ficando o tradutor responsável pelo agrupamento das respostas e geração da página HTML correspondente.

4.2 O servidor HTTP

O servidor HTTP escolhido para ser utilizado na implementação do servidor de consultas proposto foi o mais compacto possível. É o Microsoft FrontPage Server, para Windows 95/NT, que é distribuído juntamente com o programa para confecção de páginas HTML, o FrontPage.

O FrontPage Server é um servidor bastante compacto que funciona bem para despachar páginas e chamar programas externos, satisfazendo as necessidades do servidor de consultas proposto. É necessário, porém, configurá-lo através da alteração de parâmetros contidos em um arquivo de inicialização, informando em qual diretório estará o programa executável tradutor, quais serão os diretórios em que as páginas a

serem despachadas estarão e qual é o número máximo de requisições simultâneas que poderão ser aceitas por ele.

Uma vez em execução, pode-se acompanhar, na tela, o número de requisições sendo atendidas e, em caso de necessidade, pode-se verificar em um arquivo de registro quais foram as requisições feitas, em que horário e se houve algum erro em alguma delas.

4.3 O cliente Z39.50

Conforme descrito na Seção 2, o cliente utilizado para consultas com o protocolo Z39.50 foi o Zclient com interface de linha de comando, o qual faz as consultas especificadas em bancos de dados locais ou remotos, utilizando o conjunto de protocolos de comunicação TCP/IP. A estrutura de uma consulta, conforme mostrado anteriormente, é a seguinte:

Zclient endereço porta banco critério,

Onde, **endereço** é o endereço IP da máquina onde está o servidor de *Clearinghouse*, **porta** é o número da porta a qual o servidor está escutando, **banco** é o nome do banco de dados a ser pesquisado e **critério** é o conjunto de critérios de pesquisa. O conjunto de critérios de pesquisa é composto por palavras-chave e nomes de campos onde pesquisar, seguindo o formato:

critérios_pesquisa = nome_campo/palavra-chave operador | nome_campo/palavra-chave | palavra-chave

operador = AND | OR | NOT

palavra-chave = qualquer palavra para consulta

nome_campo = pubdate | title | descript | purpose | abstract | northbc | southbc | eastbc | westbc

No servidor em português proposto, o Zclient recebe uma consulta cujos parâmetros são fornecidos pela página HTML de consultas, devidamente traduzidos pelo tradutor. A resposta da consulta é manipulada pelo tradutor e repassada ao usuário, através do sevidor HTTP.

4.4 O tradutor

O tradutor desempenha o duplo papel de traduzir as consultas e as respostas, uma executada em momento completamente diferente da outra, conforme descrito nas subseções a seguir.

4.4.1 A tradução das consultas

As consultas precisam ser traduzidas de um conjunto de campos preenchidos pelo usuário, em uma página HTML, para parâmetros que o programa Zclient possa processar. Os campos que necessitam ser preenchidos ou selecionados de uma lista, pelo usuário, para a execução da consulta são os seguintes:

- nome do banco e servidor a consultar;
- data da última atualização dos conjuntos de metadados;
- campos onde se deseja pesquisar as palavras-chave digitadas;
- palavras-chave para consulta;
- latitudes e longitudes da área de cobertura da pesquisa.

Como exemplo, a consulta, já traduzida, para o caso em que se deseja saber quais são os conjuntos de metadados do banco testeall (supondo que o banco esteja localizado no endereço IP 143.54.83.131) cujos resumos possuam a palavra hidrografia é a seguinte:

Zclient 143.54.83.131 testeall abstract/hidrografia

A página em que é realizada a consulta acima é mostrada na Figura 4.3.

4.4.2 A tradução dos campos de metadados na resposta

A tradução dos campos de metadados na resposta ocorre após o usuário receber a lista de conjuntos de metadados que satisfazem sua consulta e requisitar a visualização dos metadados. Ela é feita pelo módulo tradutor que possui uma tabela com os nomes e respectivas traduções de cada campo.

Como as respostas seguem um formato definido, torna-se mais fácil a tarefa de tradução. O formato de cada linha de texto da resposta é **nome_do_campo: conteúdo** ou simplesmente **conteúdo**, onde **nome_do_campo** é o nome do campo de metadados onde as palavras compostas são separadas com sublinha ('_') e é seguido de dois pontos.

Inicialmente, um procedimento de busca de nomes de campos recebe uma linha do texto da resposta. Em seguida, procura por palavras seguidas de dois pontos que estejam no início da linha e as separa do resto, encaminhando-as ao tradutor.

O tradutor recebe como entrada as palavras que supostamente são nomes de campos. Em seguida, consulta em uma tabela a existência das palavras recebidas. Caso elas estejam presentes na tabela, as palavras correspondentes em português são recuperadas e são inseridas na linha de texto recebida, no lugar das palavras originais. Caso elas não sejam encontradas, as palavras são devolvidas como foram recebidas e são reintegradas à linha original. O processo de busca e tradução de palavras continua até que todo o texto da resposta seja examinado.

Em seguida, o texto traduzido passa por um processo de formatação, que adiciona quebras de linhas e destaca os nomes dos campos traduzidos para adequar-se ao formato HTML. Após a formatação, é despachado para o servidor HTTP, de onde é enviado para o usuário que realizou a consulta.

4.5 Testes realizados e resultados obtidos

Para verificar o funcionamento do sistema de consultas proposto, foram realizados vários testes com o protótipo em funcionamento no Instituto de Informática da UFRGS.

O primeiro tipo de teste consiste em verificar se o servidor é capaz de procurar por uma palavra indexada em qualquer ponto de um texto. Para isto, foi indexado o texto a seguir, que não possui nenhuma estrutura definida, sendo gravado com o nome pdtest1.txt:

Isto é um teste. As palavras serão indexadas pelo lindex e recuperadas pelo servidor de consultas em português.

As consultas sobre todas as palavras com mais de uma letra de extensão no texto acima foram realizadas utilizando a página de consultas do protótipo de servidor, conforme pode ser visto na Figura 4.3 abaixo:

Servidor de Consultas à Clearinghouse - Página de teste - Netscape

File Edit View Go Window Help

Back Forward Reload Home Search Guide Print Security Stop

Bem Vindo ao Servidor de Consultas à Clearinghouse em Português

Informe apenas os campos conhecidos e deixe os demais como estão

Servidor de Clearinghouse:

Banco de metadados:

Coordenadas geográficas limitantes:

Norte
90
Oeste -180 180 Leste
-90
Sul

Campo	Conteúdo	E/OU
Título:		OU
Descrição:		OU
Resumo:		OU
Qualquer campo:	palavras	

Document: Done

FIGURA 4.3 – Página de consultas do servidor em português no momento do preenchimento dos campos a consultar

Os resultados das consultas sobre palavras, feito somente com as palavras existentes no texto foram as seguintes, conforme ilustrado na Figura 4.4:

Teste de ClearingHouse - Netscape

File Edit View Go Window Help

Back Forward Reload Home Search Guide Print Security Stop

Resultado da Consulta

1 arquivo(s) satisfaz(ern) a sua consulta.

1) [pdtest1.txt](#)

Document: Done

FIGURA 4.4 – Resposta da consulta sobre palavras existentes no texto

Uma vez que o teste anterior obteve sucesso, o teste seguinte para comprovar o funcionamento correto do servidor é o de procura por palavras em um ponto específico do texto indexado. Para isto, o texto anterior foi alterado e novamente indexado, mas, desta vez, seguindo o padrão GEO do FGDC. O texto, em SGML, ficou o seguinte:

```
<!DOCTYPE metadata SYSTEM "metadata.dtd">
<metadata>
  <idinfo>
    <citation>
      <citeinfo>
        <origin>Teste
      </origin>
      <pubdate>19990115</pubdate>
      <title>Isto é um teste.</title>
    </citeinfo>
  </citation>
</idinfo>
<abstract>As palavras serão indexadas pelo Iindex e recuperadas pelo
servidor de consultas em português.</abstract>
</metadata>
```

Para este teste, submeteu-se algumas consultas ao servidor, perguntando-se sobre palavras em determinadas posições do texto como, por exemplo, título, origem e resumo. Todas as respostas foram corretas, conforme o esperado. Na Figura 4.5 pode-se visualizar a resposta para a consulta sobre uma palavra inexistente (montanha) no texto indexado. Na Figura 4.6 é mostrada a resposta para a consulta da palavra teste no campo de título do texto indexado.

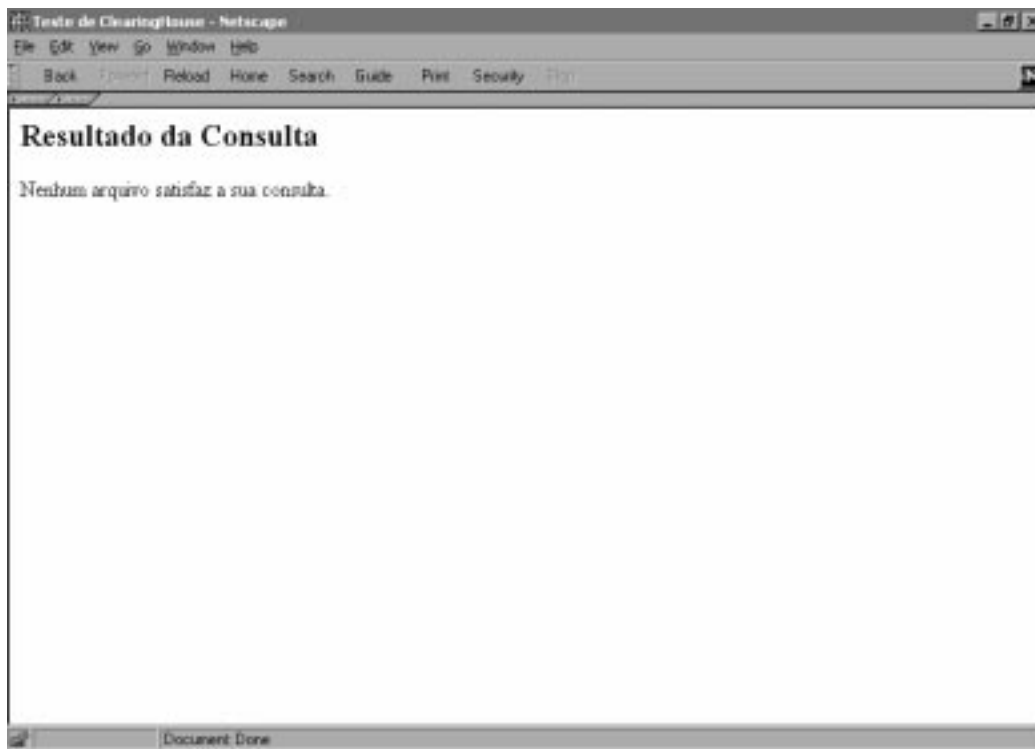


FIGURA 4.5 – Resposta para a consulta sobre uma palavra inexistente nos arquivos indexados.

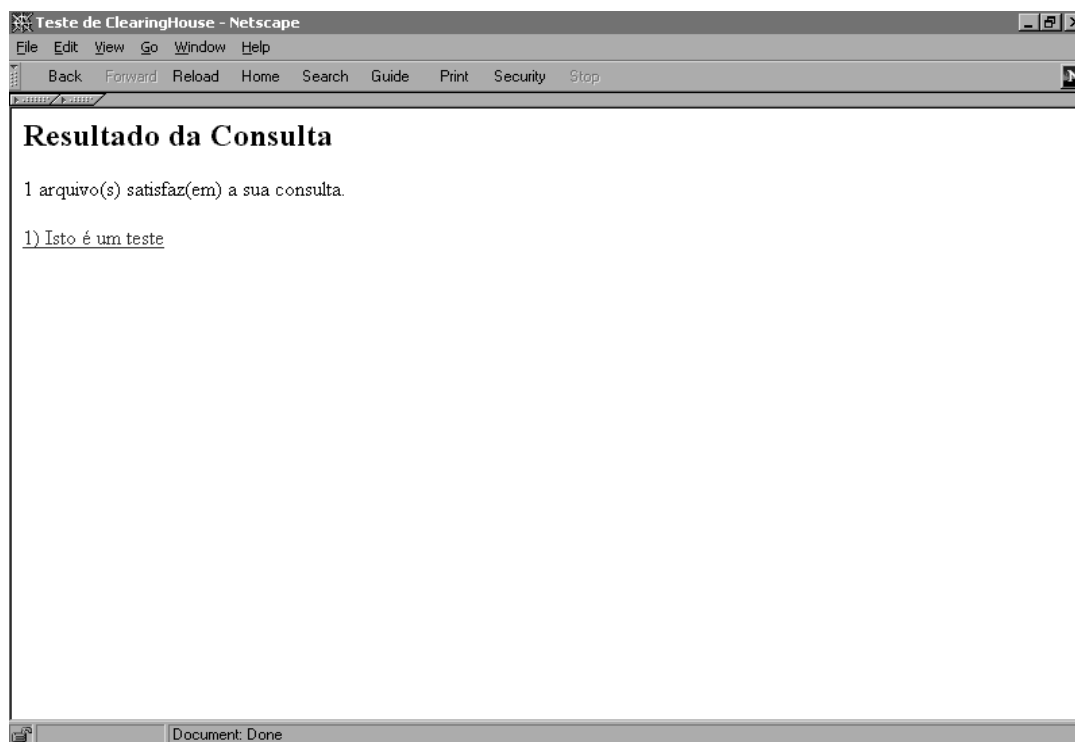


FIGURA 4.6– Resposta para a consulta da presença da palavra teste no campo título do arquivo indexado

Tendo-se comprovado o funcionamento adequado do mecanismo de consultas, um teste com um arquivo de metadados geográficos foi realizado, reproduzindo a utilização real do sistema. A Figura 4.7 mostra a tela de consultas, com critérios de busca preenchidos para uma consulta sobre o Banco de teste 1, no Servidor UFRGS.

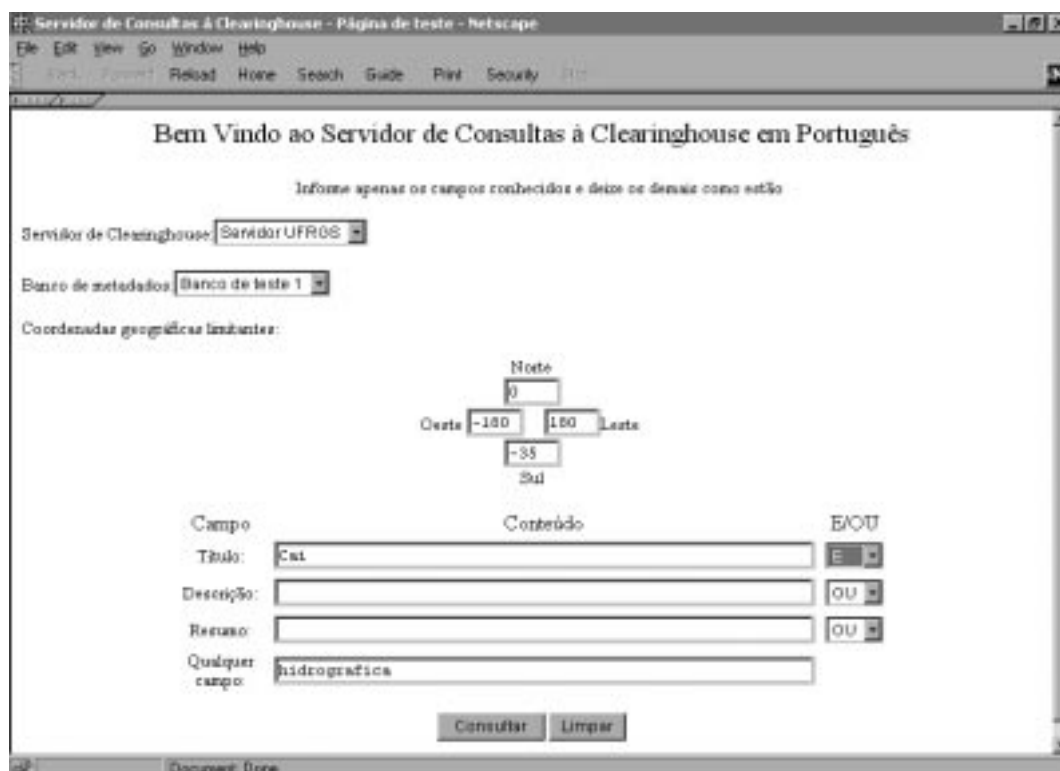


FIGURA 4.7– Consulta sobre o Banco de teste 1, no Servidor UFRGS, reproduzindo uma consulta real

5 CONCLUSÕES

Este trabalho abordou os aspectos necessários ao entendimento e resolução do problema do acesso, em português, a metadados geográficos, apresentando diferentes alternativas de solução para o problema.

Inicialmente, abordou-se o conceito de *Clearinghouse*, incluindo informações a respeito da topologia do sistema e quais são os passos necessários para a realização das consultas. As ferramentas distribuídas para a manutenção do servidor de *Clearinghouse* também são apresentadas, juntamente com as tarefas necessárias para sua criação. Em seguida, foram discutidas as diferentes alternativas possíveis para a solução do problema da tradução, abordando as vantagens e desvantagens de cada alternativa.

A solução escolhida foi implementada, utilizando-se as ferramentas disponíveis para a manutenção do sistema, sempre visando manter a compatibilidade com o sistema já existente em inglês. Sob os aspectos revisados acima, o trabalho cumpriu seus objetivos, possibilitando o acesso de usuários, de língua portuguesa, a um sistema distribuído de acesso a metadados, incentivando seu uso entre os usuários de SIG, principalmente no Brasil.

Durante o processo de desenvolvimento deste trabalho, surgiram várias dificuldades, todas elas de alguma forma contornadas. As principais dificuldades foram:

- A inexistência de referências detalhadas a respeito do funcionamento do sistema *Clearinghouse*. As informações sobre topologia tiveram que ser deduzidas, assim como o funcionamento das ferramentas de manutenção do servidor de *Clearinghouse*. O material disponível para consulta cobria o uso de poucas das funções das ferramentas utilizadas.
- A falta de colaboração do pessoal do FGDC, responsável pela divulgação do sistema, também foi um certo empecilho ao desenvolvimento do trabalho. As poucas informações divulgadas por eles foram, no entanto, valiosas.
- Tempo disponível para a realização deste trabalho foi, também, um fator determinante, principalmente na escolha do sistema operacional, visto que as ferramentas para o sistema operacional Linux não estavam funcionando bem no momento dos testes realizados.
- A dificuldade de depuração de programas que utilizam CGI também deve ser mencionada, sendo muito trabalhosa a tarefa de execução passo-a-passo deste tipo de programa, dentro do ambiente de desenvolvimento utilizado no decorrer da implementação deste projeto.

Existem diversas tarefas técnicas a serem realizadas no sentido de se obter um sistema de *Clearinghouse* para adoção no Brasil. Entre essas tarefas pode-se citar:

- Expandir a capacidade do servidor atual de consultas em português, de modo que ele permita a consulta a múltiplos servidores de *Clearinghouse* simultaneamente. Isto permitirá consultas sobre metadados mais abrangentes que as realizadas em apenas um servidor por vez. Essa tarefa inclui a transformação do servidor atualmente existente em um serviço do sistema operacional, de modo que não seja necessário o *login* de um usuário no sistema para executar o programa servidor.
- Construir uma programa para a conversão direta de arquivos gerados pelo MetaMaker (ferramenta para o cadastramento de metadados em inglês), para automatizar o trabalho de compatibilizar o formato adotado pelo MetaMaker com o adotado pelo Index. Atualmente é necessário converter manualmente os formatos utilizados por estas ferramentas.

- Facilitar a construção de arquivos de metadados por pessoas de língua portuguesa. Isto pode ser conseguido através do desenvolvimento de um programa de cadastramento de metadados em português, que tenha funções semelhantes ao MetaMaker, disponível na Internet.

Anexo 3

Artigo XIV: Requisitos especiais e modelos conceituais de dados para aplicações de sistemas de informações geográficas

Jugurta Lisboa Filho e Cirano Iochpe

In: IDEAS - WORKSHOP IBEROAMERICANO DE ENGENHARIA DE REQUISITOS E AMBIENTES DE SOFTWARE, 1998, Torres-RS. **Proceedings...** Torres: UFRGS/Instituto de Informática, 1998. Resumo estendido.

ABSTRACT

Conceptual modeling has been applied successfully to database modeling in general. Independent of underlying software, conceptual models allow for a first, abstract representation of the application's reality helping designers to better communicate with one another and with the users. Geographic applications pose a set of specific requirements on conceptual data modeling that are either not at all or only partially supported by existing conceptual data models. This paper discusses a minimum set of requirements which should be supported by a data model for Geographic Information Systems (GIS) and compares some well known conceptual data models relying on those requirements.

1 INTRODUÇÃO

Diversos trabalhos de pesquisa vêm sendo desenvolvidos com o objetivo de se encontrar um modelo conceitual de dados que seja, ao mesmo tempo, rico o bastante para representar a complexa realidade das aplicações geográficas e o mais simples possível, para que possa ser entendido e utilizado pelos usuários de *Sistemas de Informações Geográficas* (SIG) [BED 96, KOS 97, OLI 97, SHK 97, TRY 95]. Este trabalho faz uma análise comparativa dos principais modelos conceituais para SIG, com base em um conjunto de requisitos específicos apresentados pelas aplicações geográficas e aqui relacionados.

2 REQUISITOS DE MODELAGEM DE DADOS ESPECÍFICOS DAS APLICAÇÕES DE SIG

Nesta seção é descrito o conjunto de requisitos especiais de modelos de dados para projeto de aplicações de SIG, que foi identificado a partir de um levantamento bibliográfico e da experiência dos autores no projeto de banco de dados para SIG [LIS 97].

- *Visões de Campo e de Objetos* - Segundo [GOO 92], a realidade geográfica pode ser observada segundo duas visões: *de campo* e a *de objetos*. Na visão de campo, a realidade é modelada por variáveis que possuem uma distribuição contínua no espaço. Cada posição do espaço é caracterizada por um conjunto de atributos (ex.: temperatura ou relevo). Na visão de objetos a realidade geográfica consiste de entidades discretas e com identidade própria, que são representadas espacialmente por meio de objetos espaciais. Cada entidade geográfica pode ser modelada como uma classe, em um modelo de objetos. As duas visões requerem mecanismos distintos de modelagem.
- *Qualidade dos Dados* - Incerteza é inerente aos dados utilizados em SIG, mas precisa ser identificada e tratada. Os SIG também se diferenciam por fazer uso de grande volume de dados importados de outros sistemas. Este fenômeno gera, freqüentemente, problemas de confiabilidade dos dados. Além de saber o quão incerto são os dados armazenados no SIG, os usuários devem considerar que esta incerteza se propaga de diferentes formas, dependendo das operações de análise em que são utilizados.
- *Múltiplas Representações* - Uma mesma entidade geográfica pode ser representada, no SIG, em diferentes escalas e projeções, inclusive por diferentes objetos espaciais. Segundo [CAM 96], o gerenciamento de múltiplas representações em um SIG deve ser considerado nos níveis de modelagem e de estruturas de dados. Problemas decorrentes da existência de múltiplas representações incluem: redundância; inconsistência; e multiplicidade de comportamento de um mesmo fenômeno geográfico.
- *Relacionamentos Espaciais* – os SIG fornecem estruturas de dados especiais para o armazenamento explícito de alguns tipos de relacionamentos (ex.: adjacência e conectividade), deixando os demais (relacionamentos implícitos) para serem calculados durante as operações de consulta. A definição de quais relacionamentos serão mantidos de forma explícita e a especificação de restrições de integridade em relacionamentos topológicos são aspectos a serem considerados durante o projeto conceitual.
- *Temporalidade* – Atualmente, a maioria dos SIG comerciais consideram somente a realidade no presente. Informações geográficas são incluídas e alteradas ao longo do tempo, mas o histórico dessas transformações não é mantido no SIG. Para [HAD 96], a necessidade de qualificar os dados com base no tempo, não se deve apenas ao fato dos dados serem freqüentemente modificados, mas sim, pela necessidade de se registrar estados passados, de forma a possibilitar o estudo da evolução dos fenômenos geográficos.

3 MODELOS CONCEITUAIS DE DADOS PARA APLICAÇÕES EM SIG

Esta seção descreve e compara os principais modelos conceituais de dados propostos para atender aos requisitos acima. A Tabela 1 mostra a relação entre esses modelos e os requisitos descritos anteriormente.

[OLI 97] descreve o ambiente UAPE, cujo modelo de dados GMOD suporta, principalmente, a modelagem de campos e objetos. No modelo GeoIFO [TRY 95] são definidos dois novos construtores semânticos - *posição de objetos* e *atributos variando no espaço* – também para modelagem de campos e objetos. O modelo GISER [SHK 97] é o único a incluir algum tipo de mecanismo para armazenamento de metadados de qualidade, mas de forma muito superficial. [BED 96] propõe o modelo *Modul-R*, uma extensão do modelo E-R para SIG urbano, que suporta os requisitos múltiplas

representações e temporalidade. O modelo GeoOOA [KOS 97] distingue entre classes de entidades com e sem representação espacial; especifica um tipo de objeto espacial para cada classe de objeto geográfico; distingue os relacionamentos topológicos dos demais; e permite a definição de classe temporal. Uma versão completa do artigo pode ser obtida em www.inf.ufrgs.br/~jugurta.

TABELA 1 - Requisitos e Modelos Conceituais para SIG

Requisitos \ Modelos conceituais	GMOD	GeoIFO	Modul-R	GeoOOA	GISER
Visão de Campo e Objetos	*	*	-	-	-
Qualidade dos Dados	-	-	-	-	*
Múltiplas Representações	*	-	*	*	-
Relacionamentos Espaciais	-	-	-	*	-
Temporalidade	*	-	*	*	*

4 CONCLUSÕES E TRABALHOS FUTUROS

O problema da modelagem dos requisitos específicos das aplicações geográficas ainda necessita de maiores estudos. Uma integração das soluções propostas pelos diversos modelos poderá dar origem a um modelo que permita construir esquemas de dados mais fiéis à realidade em um contexto geográfico. O requisito qualidade dos dados é também o que menos atenção recebe e, devido a sua importância prática para as aplicações de SIG, que a cada dia utilizam mais dados compartilhados (ex.: via Internet), tal aspecto constitui nosso tema de pesquisa atual.

Anexo 4

Outros trabalhos publicados durante a tese

Tutoriais:

- LISBOA F., J. **Banco de Dados para GIS**. Salvador: Fator GIS, 1999. Apostila do curso apresentado no GIS Brasil'99.
- LISBOA F., J.; IOCHPE, C. **Introdução a Sistemas de Informações Geográficas com Ênfase em Banco de Dados**. Recife-PE: SBC, 1996. Apostila apresentada na Jornada de Atualização em Informática, JAI, 15.; Congresso da SBC, 16., 1996. Também editado por ocasião da Escuela de Ciencias Informáticas, 10., 1996, Buenos Aires.

Relatórios Técnicos e Monografias:

- WEBER, E. J.; ANZOLCH, R.; LISBOA F., J.; COSTA, A. C.; IOCHPE, C. **Qualidade de dados geoespaciais**. Porto Alegre: PPGC da UFRGS, 1999. (RP-293-99)
- LISBOA F., J. **Modelos Conceituais de Dados para Sistemas de Informações Geográficas**. Porto Alegre: CPGCC da UFRGS, 1997. 119p.- (EQ-12)
- LISBOA F., J.; IOCHPE, C. **Análise Comparativa dos Modelos de Dados Conceituais para Sistemas de Informações Geográficas**. Porto Alegre: CPGCC da UFRGS, 1996. (RP-266-96)
- LISBOA F., J. **Introdução a SIG - Sistemas de Informações Geográficas**. Porto Alegre: CPGCC da UFRGS, 1995. 69p.- (TI-491)

Anexo 5

Fonte de caracteres Sigmoda

O arquivo “sigmoda.ttf” é um arquivo de fontes de caracteres do sistema MS-Windows que contém o conjunto de caracteres para serem utilizados, como estereótipos, no projeto de banco de dados geográficos com base no GeoFrame [LIS 99].

O arquivo está disponível na sítio do Projeto SIGMODA no seguinte endereço:

<<http://www.inf.ufrgs.br/gpesquisa/sigmoda>>

Este arquivo deve ser instalado no diretório de arquivos de fontes de caracteres do sistema MS-Windows.

A Figura 1 mostra os caracteres a serem digitados e seus respectivos estereótipos.

Estereótipos para generalização	
Q – campo geográficos	
O – objeto geográfico	
P – objeto não geográfico	
Estereótipos para associação	
<i>Objetos Espaciais</i>	
B – ponto	
w – linha	
e – polígono	
u – obj. espacial complexo	
<i>Representação Campo</i>	
G – pontos irregulares	
H – grade de pontos	
I – TIN	
J – isolinhas	
K – grade de células	
L – polígonos adjacentes	

FIGURA 1 – Conjunto de caracteres da fonte SIGMODA e seus respectivos estereótipos.

Bibliografia

- [ABR 94] ABRANTES, G.; CARAPUCA, R. Explicit representation of data that depend on topological relationships and control over data consistency. In: EUROPEAN CONFERENCE AND EXHIBITION ON GEOGRAPHICAL INFORMATION SYSTEMS – EGIS/MARI, 1994. **Proceedings...** [S.l:s.n],1994.
- [ABR 98] ABRANTES, G.; GOMES, M. R. GEO-OMT: an object-oriented method supporting the development of facilities management systems. In: INT. CONFERENCE & EXHIBITION ON GEOGRAPHIC INFORMATION, 1998, Lisbon, Portugal. **Proceedings...** Lisbon: USIG,1998.
- [ABI 87] ABITEBOUL, S.; HULL, R. IFO: A formal semantic database model. **ACM Transactions on Database Systems**, New York, v.12, n.4, p.525-565, 1987.
- [ACT 98] ACTIVE THE Z39.50 SERVER PROVESS. **Activate the Z39.50 server process**. Disponível em <http://www.fgdc.gov/clearinghouse/tutorials/serve.html> (08/1998).
- [AGR 71] AGRAR UND HIDROTECHNIK/SOP-RS. **Planejamento hidrológico e estudo de desenvolvimento regional da bacia do rio Caí**. Porto Alegre: [s.n], 1971.
- [ALE 77] ALEXANDER, C. et al. **A Pattern Language**. New York: Oxford University, 1977.
- [ALM 97] ALMADA, C. M. W.; WÜRDIG, N. L. Avaliação da fauna bentônica em ambiente aquático da região carbonífera. In: UFRGS/CENTRO DE ECOLOGIA. **Energia e Meio Ambiente: a questão do carvão no Rio Grande do Sul**. Porto Alegre: UFRGS/Centro de Ecologia, 1997. v.5. Relatório final Projeto PADCT-CIAMB.
- [ALM 99] ALMEIDA, L. F. B. Padrão CPRMg para intercâmbio de dados gráficos digitais. In: GISBRASIL CONGRESSO E FEIRA PARA USUÁRIOS DE GEOPROCESSAMENTO DA AMÉRICA LATINA, 5., 1999, Salvador. **Anais...** Salvador: Fator GIS, 1999.
- [ANT 91] ANTENUCCI, J. C. et al. **Geographic Information Systems: a guide to the technology**. New York: Van Nostrand Reinhold, 1991.
- [ANZ 97] ANZLIC THE AUSTRALIA NEW ZEALAND LAND INFORMATION COUNCIL. **Core Metadata Elements for Land and Geographic Directories in Australia and New Zealand**. 1997. Disponível em <http://www.anzlic.org.au/metaelem.htm>.
- [APP 97] APPLETON, B. **Patterns and software: essential concepts and terminology**. Disponível em <http://www.enteract.com/~bradapp/> (11/20/97).
- [ARO 89] ARONOF, S. **Geographic Information Systems: a management perspective**. Canada: WDL Publications, 1989.

- [ASP 96] ASPINALL, R. J.; PEARSON, D. M. Data quality and spatial analysis: analytical use of GIS for ecological modeling. In: GOODCHILD, M.F. et al. (Eds.). **GIS and Environmental Modeling: Progress and Research Issues**. Fort Collins, CO: GIS World Books, 1996.
- [BAL 97] BALAGUER, F.; GORDILLO, S.; NEVES, F. Patterns for GIS applications design. In: PATTERN LANGUAGES OF PROGRAMMING, PLoP, 1997, Monticello, Illinois. **Proceedings...** Monticello: ACM, 1997.
- [BAT 94] BATTY, M.; XIE, Y. Research article: Modelling inside GIS: Part 1 Model structures, exploratory spatial data analysis and aggregation. **International Journal of Geographical Information Systems**, London, v. 8, n.3, p.291-307, 1994.
- [BCN 92] BATINI, C.; CERI, S.; NAVATHE, S. B. **Conceptual database design: an entity-relationship approach**. Redwood: Benjamin/Cummings, 1992.
- [BED 89] BÉDARD, Y.; PAQUETTE, F. Extending entity/relationship formalism for spatial information systems. In: AUTOCARTO, 9., 1989. **Proceedings...** [S.l.:s.n.], 1989. p.818-828.
- [BED 96] BÉDARD, Y. et al. Adapting data models for the design of spatio-temporal databases. **Computers, Environment, and Urban Systems**, Bristol-UK, v.20, n.1, 1996.
- [BED 99] BÉDARD, Y. Visual modeling of spatial databases towards spatial extensions and UML. **Geomatica**, [S.l.], v. 53, n. 2, p. 169-186, 1999.
- [BIN 98] BINS, C. A.; RUSCHEL, C.; MORAES, E. A. S. Pró-Guaíba: sistema de informações geográficas. In: GISBRASIL, 1998, Curitiba. **Anais...** Curitiba: Revista FatorGIS, 1998.
- [BON 93] BONFATTI, F. et al. Object-oriented support to the design of geographical information systems. In: EUROPEAN CONFERENCE AND EXHIBITION ON GEOGRAPHICAL INFORMATION SYSTEMS, 4., 1993, Genoa, Italy. **Proceedings...** Genoa, Italy: [s.n.], 1993.
- [BON 95] BONFATTI, F.; DALLARI, A.; MANARI, P.D. Capturing more knowledge for the design of geological information systems. In: INTERNATIONAL WORKSHOP ON ADVANCES IN GEOGRAPHIC INFORMATION SYSTEMS - ACM-GIS, 1995, Baltimore. **Proceedings...** Baltimore: ACM, 1995.
- [BOO 98] BOOCH, G.; JACOBSON, I.; RUMBAUGH, J. **The Unified Modeling Language User Guide**. Reading: Addison-Wesley, 1998.
- [BOR 97] BORGES, K. A. V. **Modelagem de Dados Geográficos: uma extensão do modelo OMT para aplicações geográficas**. Belo Horizonte: Fundação João Pinheiro, 1997. Dissertação de Mestrado.
- [BOR 99] BORGES, K.A.V.; LAENDER, A.H.F.; DAVIS JR, C.D. Spatial data integrity constraints in object oriented geographic data modeling. In: ACM SYMPOSIUM ON ADVANCES IN GEOGRAPHIC INFORMATION SYSTEMS, 7., 1999, Kansas City, USA. **Proceedings...** Kansas City: ACM, 1999.

- [BRI 95] BRITISH COLUMBIA SPECIFICATIONS AND GUIDELINES FOR GEOMATICS. **Spatial Archive and Interchange Format**: SAIF formal definition. Release 3.2. British Columbia, Canada: Ministry of Environment, Lands and Parks, Reference series. 1995. v.1. Disponível em <http://www.lii.crl.gov.bc.ca/saif/saifHome.html>.
- [BRO 84] BRODIE, M. L. On the development of data models. In: BRODIE, M. L.; MYLOPOULOS, J.; SCHMIDT, J.W. (Eds.). **On Conceptual Modeling**. New York: Springer-Verlag, 1984. p.19-48.
- [BUR 86] BURROUGH, P. A. **Principles of Geographical Information Systems for Land Resources Assessment**. Oxford: Clarendon Press, 1986.
- [BUR 95] BURROUGH, P.A.; FRANK, A.U. Concepts and paradigms in spatial information: are current geographical information systems truly generic? **International Journal of Geographical Information Systems**, London, v.9, n.2, p.101-116, 1995.
- [BUS 96] BUSCHMANN, F. et al. **Pattern-Oriented Software Architecture: A System of Patterns**. New York: John Wiley & Sons, 1996.
- [CAM 94] CÂMARA, G. et al. A general data model for integrating remote sensing and GIS data. In: SYMPOSIUM ON MAPPING AND GEOGRAPHIC INFORMATION SYSTEMS, 1994, Athen. **Proceedings...** Athen: [s.n.], 1994.
- [CAM 95] CÂMARA, G. **Modelos, Linguagens e Arquiteturas para Bancos de Dados Geográficos**. São José dos Campos: INPE, 1995. Tese de Doutorado.
- [CAM 96] CÂMARA, G. et al. **Anatomia de Sistemas de Informação Geográfica**. Campinas: Instituto de Computação, UNICAMP, 1996. (X Escola de Computação).
- [CAR 93] CARON, C.; BÉDARD, Y. Extending the individual formalism for a more complete modeling of urban spatially referenced data. **Computers, Environment and Urban Systems**, [S.l.], v.17, p.337-346, 1993.
- [CAT 97] CATTEL, R.G.G.; BARRY, D.K. (Eds.). **The Object Database Standard: ODMG 2.0**. San Francisco: Morgan Kaufmann, 1997.
- [CEN 96] CEN EUROPEAN COMMITTEE FOR STANDARDIZATION. **Geographic Information – Data Description Language – Conceptual Schema Language**. Brussels: CEN, 1996. (Report CR 287005).
- [CHE 76] CHEN, P.P.S. The entity-relationship model: Towards a unified view of data. **ACM Trans. Database System**, New York, n.1, 1976.
- [CHR 97] CHRISMAN, N. **Exploring Geographic Information Systems**. New York: John Wiley & Sons, 1997.
- [CLE 94] CLEMENTINI, E.; FELICE, P.D. Object-oriented modeling of geographic data. **Journal of the American Society for Information Science**, [S.l.], v.45, n.9, p.694-704, 1994.
- [COA 91] COAD, P.; YOURDON, E. **Object-Oriented Analysis**. 2.ed. New Jersey: Prentice-Hall, 1991.

- [COA 92] COAD, P. Object-oriented patterns. **Communications of the ACM**, New York, v.35, n.9, 1992.
- [COA 97] COAD, P. **Object Models: Strategies, Patterns, and Applications**. 2.ed. New Jersey: Yourdon Press, 1997.
- [COC 97] COCKCROFT, S. A taxonomy of spatial data integrity constraints. **Geoinformatica**, The Netherlands, v.1, n.4, 1997.
- [COL 96] COLEMAN, D. et al. **Object-Oriented Development: the fusion method**. New Jersey: Prentice-Hall, 1996.
- [CON 96] CONAMA. **Resoluções CONAMA entre 1984 e 1992**. Brasília: Ministério do Desenvolvimento Urbano e Meio Ambiente, 1993.
- [COP 94] COPLIEN, J. O. **Software design patterns: common questions and answers**. [S.l]: AT&T Bell Laboratories Technical Memorandum, 1994.
- [COP 95] COPLIEN, J. O. A generative development-process pattern language. In: COPLIEN J. O.; SCHMIDT D. (Eds.) **Pattern Languages of Program Design**. Reading: Addison-Wesley, 1995.
- [COU 92] COUCLELIS, H. People manipulate objects (but cultivate fields): beyond the raster-vector debate in GIS. In: **Theories and methods of spatial-temporal reasoning in geographic space**. Berlin: Springer-Verlag, 1992. p.65-77. (Lecture Notes in Computer Science 639).
- [COY 96] COYLE, M. et al. **Experiences with object data models in geographic information systems**. 1996. White paper. Disponível em <http://www.cs.umn.edu/research/shashi-group/>
- [CRO 96] CROSBIE, P. Object-oriented design of GIS: a new approach to environmental modeling. In: GOODCHILD, M.F. et al. (Eds.). **GIS and Environmental Modeling: Progress and Research Issues**. Fort Collins, CO: GIS World Books, 1996.
- [DAT 81] DATE, C. J. **An Introduction to Database Systems**. 3. ed. Reading: Addison-Wesley, 1981.
- [DIG 88] DIGITAL CARTOGRAPHIC DATA STANDARDS TASK FORCE. The proposed standard for digital cartographic data. **The American Cartographer**, [S.l], v.15, n.1, 1988.
- [EDE 94] EDELWEISS, N. **Sistemas de Informação de Escritórios: um Modelo para Especificações Temporais**. Porto Alegre: CPGCC da UFRGS, 1994. Tese de Doutorado.
- [EGE 87] EGENHOFER, M. J.; FRANK, A. U. Object-oriented databases: database requirements for GIS. In: INT. GEOGRAPHIC INFORMATION SYSTEMS SYMPOSIUM, 1987, Arlington. **Proceedings...** Virginia:[s.n.], 1987.
- [EGE 89] EGENHOFER, M. J.; FRANK, J.; JACKSON, J. A topological data model for spatial databases. In: SYMP. DESIGN AND IMPLEMENTATION OF LARGE SPATIAL DATABASES, 1989. **Proceedings...** [S.l.]:Spring-Verlag, 1989. p.271-286.

- [EGE 92] EGENHOFER, M. J.; FRANK, A. U. Object-oriented modeling for GIS. **Journal of the Urban and Regional Information Systems Association**, [S.l.], v.4, n.2, 1992.
- [ELM 89] ELMASRI, R.; NAVATHE, S. B. **Fundamentals of Database Systems**. Menlo Park: Addison-Wesley, 1989.
- [ELM 94] ELMASRI, R.; NAVATHE, S. B. **Fundamentals of Database Systems**. 2.ed. Menlo Park: Addison-Wesley, 1994.
- [EMB 87] EMBLEY, D. W.; WOODFIELD, S.N. A knowledge structure for reusing abstract data types. In: INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING, 9., 1987, Monterey. **Proceedings...** New York: IEEE Computer Society Press, 1987.
- [EUR 96] EUROPEAN COMMITTEE FOR STANDARDIZATION. **Geographic information – data description – conceptual schema language**. Brussels: CEN, 1996. (Report CR 287005).
- [FED 97] FEDERAL GEOGRAPHIC DATA COMMITTEE. **Content Standard for Digital Geospatial Metadata**. Washinton, D.C.: FGDC, 1997. Disponível em <http://www.fgdc.gov>.
- [FEA 99] FEATURE MANIPULATION ENGINE. **FME User Documentation Set**. 1999. Disponível em <http://www.safe.com/testdrv.htm>.
- [FER 97] FERRARO, L. W.; HASENACK, H.. Avaliação das variáveis climáticas de superfície do baixo Jacuí - RS. In: UFRGS/CENTRO DE ECOLOGIA. **Energia e Meio Ambiente: a questão do carvão no Rio Grande do Sul**. Porto Alegre: UFRGS/Centro de Ecologia, 1997. v.4. Relatório final Projeto PADCT-CIAMB.
- [FER 98] FERNANDEZ, E. B. Building systems using analysis patterns. In: INTERNATIONAL SOFTWARE ARCHITECTURE WORKSHOP (ISAW3), 3., 1998, Orlando-FL. **Proceedings...** Orlando: ACM Press, 1998.
- [FER 98a] FERNANDEZ, E. B. **Semantic Analysis Patterns**. In: INT. CONF. OF THE CHILEAN COMPUTER SOCIETY, 18., 1998, Antofagasta, Chile. Tutorial presented in 83 transparencies.
- [FER 99] FERNANDEZ, E. B.; YUAN, X. An analysis pattern for reservation and use of reusable entities. In: WORKSHOP IN THE CONFERENCE OF PATTERN LANGUAGE OF PROGRAMS - PLoP, 1999, Monticello-Illinois. **Proceedings...** [S.l:s.n], 1999. Disponível em <http://st-www.cs.uiuc.edu/~plop/plop99/proceedings/>.
- [FIG 98] FIGUEIREDO, L. C. C. **Modelagem de dados geográficos de sistema de informações de recursos hídricos**. Belo Horizonte: UFMG/Depto de Cartografia, 1998. Monografia do Curso de Especialização em Geoprocessamento.
- [FOW 96] FOWLER, M. Analysis patterns and business objects. **ACM SIGPLAN Notices**, New York, v.31, n.10, 1996. Trabalho apresentado na OOPSLA, 11., 1996, San Jose, USA.
- [FOW 97] FOWLER, M. **Analysis Patterns: Reusable Object Models**. Menlo Park, CA: Addison Wesley Longman, 1997.

- [FRA 88] FRANK, A. U. Requirements for a databases management system for a GIS. **Photogrammetric Engineering and Remote Sensing**, [S.l.], v.54, n.11, p.1557-1564, 1988.
- [FRA 89] FRAKES, W. B. et al. Information retrieval and software reuse. In: ANNUAL INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, 1989, Cambridge. **Proceedings...** New York: ACM, 1989.
- [FUR 98] FURLAN, J. D. **Modelagem de Objetos através da UML: the Unified Modeling Language**. São Paulo: Makron Books, 1998.
- [GAM 94] GAMMA, E. et al. **Design Patterns: Elements of Reusable Object-Oriented Software**. Reading, MA: Addison Wesley, 1994.
- [GAN 79] GANE, C.; SARSON, T. **Structured systems analysis: tools and techniques**. New York: Prentice-Hall, 1979.
- [GAR 98] GARAFFA, I. M. **Análise da Adequação de uma hierarquia de Classes Básicas para Modelagem Conceitual de SIG através de um Estudo de Caso**. Porto Alegre: CPGCC da UFRGS, 1998. Dissertação de Mestrado.
- [GER 96] GERTH, Th.; SCHACHTSCHABEL, R.; SCHÖNEFELD, R. Using patterns in design and documentation of software. In: THE WHITE OO NIGHTS -WOON, 1996, St. Petersburg, Russia. **Proceedings...** St. Petersburg: [s.n.], 1996.
- [GIR 90] GIRARDI, M. R. **Uma ferramenta de apoio à reutilização de software no desenvolvimento orientado a objetos**. Porto Alegre: CPGCC da UFRGS, 1990. Dissertação de Mestrado.
- [GIR 96] GIRARDI, M. R. **Classification and retrieval of software and their description in natural language**. Genève: Université de Genève, 1996. PhD Thesis.
- [GOL 82] GOLDBERG, A.; ROBSON, D. **Smalltalk-80: The language and its implementation**. Reading: Addison-Wesley, 1982.
- [GOO 91] GOODCHILD, M. F. Integrating GIS and environmental modeling at global scales. In: GIS/LIS, 1991. **Proceedings...** Atlanta, GE: [s.n.], 1991.
- [GOO 92] GOODCHILD, M. F., Geographical data modeling. **Computers & Geosciences**, London, v.18, n.4, p.401-408, 1992.
- [GOR 97] GORDILLO, S.; BALAGUER, F.; NEVES, F. Generating the architecture of GIS applications with design patterns. In: ACM WORKSHOP ON GEOGRAPHIC INFORMATION SYSTEMS, 5., 1997, Las Vegas, NE. **Proceedings...** Las Vegas: ACM, 1997.
- [GOR 98] GORDILLO, S.; BALAGUER, F. Refining an object-oriented GIS design model: topologies and field data. In: ACM SYMPOSIUM ON GEOGRAPHIC INFORMATION SYSTEMS, 6., 1998, Washington, D.C. **Proceedings...** Washington: ACM, 1998.
- [HAD 96] HADZILACOS, T.; TRYFONA, N. Logical data modelling for geographical applications. **International Journal of Geographical Information Science**, London, v.10, n.2, p.179-203, 1996.

- [HAD 97] HADZILACOS, T.; TRYFONA, N. An extended entity-relationship model for geographic applications. **SIGMOD Record**, New York, v.26, n.3, 1997.
- [HAY 95] HAY, D. C. **Data Model Patterns: Conventions of Thought**. New York: Dorset House Publishing, 1995.
- [HAY 96] HAY, D. C. Using data model patterns for rapid application development. In: INTERNATIONAL ORACLE USER WEEK, 1996. **Proceedings...** [S.l.]: Oracle, 1996.
- [HEU 98] HEUSER, C. A. **Projeto de Banco de Dados**. Porto Alegre: Sagra Luzzato, 1998. (Série livros didáticos, n.4)
- [HUL 87] HULL, R.; KING, R. Semantic database modeling: survey, applications, and research issues. **ACM Computing Surveys**, New York, v.19, n.3, p.201-260, 1987.
- [INP 99] INPE. **Manual do Spring - v.3.3**. São José dos Campos: Instituto Nacional de Pesquisas Espaciais, 1999. CD-ROM.
- [IOC 94] IOCHPE, C. **Requisitos de um Sistema de Banco de Dados Alfanumérico para o SIGERCO/RS**. Porto Alegre: FEPAM/GERCO-RS, 1994. Relatório Final de Consultoria.
- [IOC 98] IOCHPE, C.; LISBOA F., J. BCH/GIS: a basic class hierarchy to support GIS conceptual design. In: INT. CONFERENCE ON MODELING GEOGRAPHICAL AND ENVIRONMENTAL SYSTEMS WITH GEOGRAPHICAL INFORMATION SYSTEMS (GIS), 1998, Hong Kong, China. **Proceedings...** Hong Kong: The Chinese University of Hong Kong, 1998.
- [ISO 97] ISO/TC211 TERMINOLOGY PROJECT TEAM. **Collections of terms and definitions from ISO/TC211 geographic information/geomatics**. Disponível em <http://www.statkart.no/isotc211/terms/terms211.html>. (05/1997).
- [ISO 99] ISOWARE. **CASE-Toll REGIS**. Disponível em <http://www.isoware.de/>. (1999).
- [JOH 92] JOHNSON, R. E. Documenting frameworks using patterns. In: OBJECT-ORIENTED PROGRAMMING SYSTEMS, LANGUAGES AND APPLICATIONS CONFERENCE - OOPSLA, 1992, Vancouver. **Proceedings...** Vancouver: ACM, 1992.
- [JOH 97] JOHNSON, R. E. **Components, Frameworks, Patterns**. [S.l.:s.n], 1997. White paper.
- [JOH 99] JOHANNESSON, P.; WOHED, P. The deontic pattern – a framework for domain analysis in information systems design. **Data & Knowledge Engineering**, [S.l.], v.31, 1999.
- [KEM 98] KEMP, K. K.; VCKOVSKY, A. Towards an ontology of fields. In: INT. CONF. ON GEOCOMPUTATION, 3., 1998, Bristol-UK. **Proceedings...** Bristol: University of Bristol, 1998.

- [KOS 95] KÖSTERS, G.; PAGEL, B.; SIX, H. Object-oriented requirements engineering for GIS-applications. In: ACM-GIS INTERNATIONAL WORKSHOP ON ADVANCES IN GEOGRAPHIC INFORMATION SYSTEMS, 1995, Baltimore. **Proceedings...** Baltimore: [s.n.], 1995.
- [KOS 97] KÖSTERS, G. et al. GIS-Application Development with GeoOOA. **International Journal of Geographical Information Science**, London, v.11, n.4, 1997.
- [LAU 92] LAURINI, R.; THOMPSON, D. **Fundamentals of Spatial Information Systems**. San Diego: Academic Press, 1992.
- [LCR 99] LCRA LOWER COLORADO RIVER AUTHORITY. **The Texas Clean Rivers Program**: Colorado River Basin Data Management Plan Task 3D. Disponível em <http://www.lcra.org/lands/wsm/dmp1998.htm> (24/11/99).
- [LIA 93] LIAO, H. C. et al. Software reuse based on a large object oriented library. **Software Engineering Notes**, New York, v.18, n.1, p.74-80, 1993.
- [LIS 96] LISBOA FILHO, J.; IOCHPE, C. Adaptando o modelo de objetos OMT para modelagem conceitual de aplicações de SIG. In: SEGEO-RJ, 1., 1996, Rio de Janeiro. **Anais...** Rio de Janeiro: Esc. Engenharia, 1996.
- [LIS 96a] LISBOA FILHO, J.; IOCHPE, C. **Análise Comparativa dos Modelos de Dados Conceituais para Sistemas de Informações Geográficas**. Porto Alegre: CPGCC da UFRGS, 1996. (RP-266-96).
- [LIS 96b] LISBOA FILHO, J.; IOCHPE, C. **Introdução a Sistemas de Informações Geográficas com Ênfase em Banco de Dados**. Recife-PE: SBC, 1996. Apostila apresentada na Jornada de Atualização em Informática, JAI, 15.; Congresso da SBC, 16., 1996. Também editada por ocasião da Escuela de Ciencias Informáticas, 10., 1996, Buenos Aires.
- [LIS 97] LISBOA FILHO, J.; IOCHPE, C.; GARAFFA, I. M. Modelos conceituais de dados para aplicações geográficas: uma experiência com um SIG interinstitucional. In: SIMPÓSIO BRASILEIRO DE GEOPROCESSAMENTO, 4., 1997, São Paulo. **Anais...** São Paulo: UPUSP, 1997.
- [LIS 97a] LISBOA FILHO, J.; GARAFFA, I. M.; IOCHPE, C. Metodologia para modelagem conceitual de dados de um SIG interinstitucional. In: GISBRASIL - CONGRESSO E FEIRA PARA USUÁRIOS DE GEOPROCESSAMENTO, 3., 1997, Curitiba. **Anais...** Curitiba: Fator GIS, 1997.
- [LIS 97b] LISBOA FILHO, J. **Modelos Conceituais de Dados para Sistemas de Informações Geográficas**. Porto Alegre: CPGCC da UFRGS, 1997. 119p.- (EQ-12)
- [LIS 98] LISBOA FILHO, J.; IOCHPE, C.; BEARD, K. Applying Analysis Patterns in the GIS Domain. In: ANNUAL COLLOQUIUM OF THE SPATIAL INFORMATION RESEARCH CENTRE - SIRC, 10., 1998, Dunedin, NZ. **Proceedings...** Dunedin: SIRC, University of Otago, 1998.

- [LIS 98a] LISBOA FILHO, J.; IOCHPE, C. Requisitos especiais e modelos conceituais de dados para aplicações de sistemas de informações geográficas. In: IDEAS - WORKSHOP IBEROAMERICANO DE ENGENHARIA DE REQUISITOS E AMBIENTES DE SOFTWARE, 1998, Torres-RS. **Proceedings...** Torres: UFRGS/Instituto de Informática, 1998. Resumo estendido.
- [LIS 99] LISBOA FILHO, J.; IOCHPE, C. Specifying analysis patterns for geographic databases on the basis of a conceptual framework. In: ACM SYMPOSIUM ON ADVANCES IN GEOGRAPHIC INFORMATION SYSTEMS, 7., 1999, Kansas City, USA. **Proceedings...** Kansas City: ACM Press, 1999.
- [LIS 99a] LISBOA FILHO, J.; COSTA, A. C.; IOCHPE, C. Projeto de banco de dados geográficos: mapeando esquemas *GeoFrame* para o SIG *Spring*. In: BRAZILIAN WORKSHOP ON GEOINFORMATICS - GEO-INFO, 1., 1999, Campinas. **Anais...** São José dos Campos: INPE, 1999.
- [LIS 99b] LISBOA FILHO, J.; IOCHPE, C.; HASENACK, H.; WEBER, E. J. Modelagem conceitual de banco de dados geográficos: o estudo de caso do projeto PADCT/CIAMB. In: UFRGS/CENTRO DE ECOLOGIA. **Energia e Meio Ambiente: a questão do carvão no Rio Grande do Sul**. Porto Alegre: Editora da Universidade, 1999. No prelo.
- [LIS 99c] LISBOA FILHO, J.; IOCHPE, C. Um estudo sobre modelos conceituais de dados para projeto de bancos de dados geográficos. **Revista IP-Informática Pública**, Belo Horizonte, v.1, n.2, 1999.
- [LIS 99d] LISBOA FILHO, J.; IOCHPE, C. Mecanismos de Reutilização em Sistemas de Informação. Submetido à revista **Acta Scientiae**, Canoas-RS, 1999.
- [LIS 9?] LISBOA FILHO, J.; IOCHPE, C. **Padrões de análise para banco de dados geográficos**. A ser submetido.
- [MAI 92] MAIDEN, N. A. **Analogical specification reuse during requirements analysis**. London: Department of Business Computing, City University, School of Informatics, 1992. PhD Thesis.
- [MAN 97] MÂNICA, D. P.; BERNARDI, G.; SAGGIORATO, S. M. **Modelagem conceitual de dados da aplicação da Bacia do Rio Caí**. Porto Alegre: UFRGS/PPGC, 1997. Monografia da disciplina de Sistemas de Informação Avançados.
- [MAR 97] MARR, A. J.; PASCOE, R. T.; BENWELL, G. Interoperable GIS and spatial process modelling. In: GEOCOMPUTATION, 2., 1997, Dunedin, New Zealand. **Proceedings...** Dunedin: University of Otago, 1997.
- [MAR 98] MARR, A. J. et al. Towards the development of a documentation structure for modelling spatial process. In: GEOCOMPUTATION, 3., 1998, Bristol-UK. **Proceedings...** Bristol: University of Bristol, 1998.
- [MAR 99] MARSURA, P. **Banking patterns home page**. Disponível em <http://www.joeyoder.com/marsura/banking/> (03/09/1999).
- [MES 98] MESZAROS, G.; DOBLE, J. **A pattern language for pattern writing**. Disponível em http://hillside.net/patterns/Writing/pattern_index.html (01/12/98).

- [MET 98] METASTAR GATEWAY. **Metastar gateway**. Disponível em <http://www.fgdc.gov/clearinghouse/tutorials/index.html> (08/1998).
- [MEU 97] MEUSEL, M.; CZARNECKI, K.; KÖPF, W. A model for structuring user documentation of object-oriented framework using patterns and hypertext. In: EUROPEAN CONFERENCE ON OBJECT-ORIENTED PROGRAMMING – ECOOP, 1997, Finland. **Proceedings...** Finland: Springer-Verlag, 1997. (Lecture Notes in Computer Science, v.1241)
- [MIL 93] MILNE, P.; MILTON, S.; SMITH, J. L. Geographical object-oriented databases - a case study. **International Journal of Geographical Information Systems**, London, v.7, n.1, p.39-55, 1993.
- [MOE 88] MOELLEWING, H. The proposed standard for digital cartographic data: report of the digital cartographic data standards task force. **The American Cartographer**, [S.l.], v.15, n.1, 1988.
- [NAT 94] NATIVI, S. A conceptual modelling for the GIS developing. In: EUROPEAN CONFERENCE AND EXHIBITION ON GEOGRAPHICAL INFORMATION SYSTEMS – EGIS/MARI, 1994. **Proceedings...** [S.l:s.n],1994.
- [NAV 92] NAVATHE, S. B. Evolution of Data Modeling for Databases. **Communications of the ACM**, New York, v.35, n.9, p.112-123, 1992.
- [NEB 96] NEBERT, D. D. Information Architecture of a Clearinghouse. In: WWW ACCESS TO EARTH OBSERVATION / GEO-REFERENCED DATA WORKSHOP, 1996, World Web Conference. **Proceeding...** Disponível em <http://www.fgdc.gov/clearinghouse/>.
- [NEI 94] NEIGHBORS, J. M. An assessment of reuse technology after ten years. In: INTERNATIONAL CONFERENCE ON SOFTWARE REUSE, 3., 1994, Rio de Janeiro. **Proceedings...** Los Alamitos, California: IEEE Computer Society Press, 1994. 235p. p.6-13.
- [NEV 97] NEVES, G. R.; CHAVES, S. H. A. Notas para o estudo da região carbonífera tradicional do Rio Grande do Sul. In: UFRGS/CENTRO DE ECOLOGIA. **Energia e Meio Ambiente: a questão do carvão no Rio Grande do Sul**. Porto Alegre: UFRGS/Centro de Ecologia, 1997. v.2. Relatório final Projeto PADCT-CIAMB.
- [OLI 97] OLIVEIRA, J.L.; PIRES, F.; MEDEIROS, C. B. An environment for integrated modelling and analysis of geographic information. **GeoInformatica**, Boston, n.1, p.29-58, 1997.
- [OOS 89] OOSTEROM, P. V.; BOS, J. V. D. An Object-oriented approach to the design of geographic information systems. **Comput. & Graphics**, [S.l.], v.13, n.4, p.409-418, 1989.
- [OPE 98] OPEN GIS CONSORTIUM TECHNICAL COMMITTEE. **The OpenGIS Guide: introduction to interoperable geoprocessing and the OpenGIS specification**. Wayland-MA: Open GIS Consortium, 1998.
- [PAR 98] PARENT, C. et al. Modeling spatial data in the MADS conceptual model. In: INTERNATIONAL SYMPOSIUM ON SPATIAL DATA HANDLING. 1998, Vancouver, Canada. **Proceedings...** Vancouver: [s.n], 1998.

- [PAR 99] PARENT, C.; SPACCAPIETRA, S.; ZIMANYI, E. Spatial-temporal conceptual models: data structures + space + time. In: ACM SYMPOSIUM ON ADVANCES IN GEOGRAPHIC INFORMATION SYSTEMS, 7., 1999, Kansas City, USA. **Proceedings...** Kansas City: ACM Press, 1999.
- [PAR 99] PARENT, C. et al. Spatio-temporal conceptual models: data structures + space + time. In: ACM SYMPOSIUM ON ADVANCES IN GEOGRAPHIC INFORMATION SYSTEMS, 7., 1999, Kansas City, USA. **Proceedings...** Kansas City: ACM Press, 1999.
- [PET 91] PETTERSON, A. S. Model-based software reuse terminology. **Software Engineering Notes**, New York, v.16, n.2, p.45-51, 1991.
- [PET 98] PETERSON, D. L.; CHRISMAN, N. R.; SCHREINER, E. G. **A clearinghouse for geographic data on the Olympic Peninsula**. 1998. Disponível em <http://www.weber.u.washington.edu/~olympic/clearinghouse/>.
- [PEU 95] PEUQUET, D. J. An event-based spatiotemporal data model (ESTDM) for temporal analysis of geographical data. **International Journal of Geographical Information Systems**, London, v.9, n.1, p.7-24, 1995.
- [PIM 95] PIMENTEL, F. L. **Uma Proposta de Modelagem Conceitual para Dados Geográficos: o Modelo Mgeo++**. Recife: UFPE, 1995. Dissertação de Mestrado.
- [PIM 98] PIMENTA, A. **Especificação formal de uma ferramenta de reutilização de especificações de requisitos**. Porto Alegre: CPGCC da UFRGS, 1998. Dissertação de Mestrado.
- [PIR 97] PIRES, F. **Um ambiente computacional para modelagem de aplicações geográficas**. Campinas: Unicamp, Instituto de Computação, 1997. Tese de Doutorado.
- [PRE 87] PRESSMAN, R.S. **Software Engineering: a practitioner's approach**. 2.ed. New York: McGraw-Hill, 1987.
- [PRE 94] PREE, W. Meta patterns - a means for capturing the essentials of reusable object-oriented design. In: EUROPEAN CONFERENCE ON OBJECT-ORIENTED PROGRAMMING – ECOOP, 1994. **Proceedings...** Berlin: Springer-Verlag, 1994.
- [PRE 95] PREE, W. **Design Patterns for Object-Oriented Software Development**. Reading, MA: Addison-Wesley, 1995.
- [QUA 97] QUADROS, E. M.; RUBIRA, C. M. F. Construção de um framework para sistemas controladores de trens utilizando padrões de projeto e metapadrões. In: SIMPÓSIO BRASILEIRO DE ENGENHARIA DE SOFTWARE, 11., 1997, Fortaleza. **Anais...** Fortaleza: SBC, 1997.
- [RAM 94] RAMIREZ, M. R. **Sistemas Gerenciadores de Bancos de Dados para Geoprocessamento**. Rio de Janeiro: COPPE/UFRJ, 1994. Dissertação de Mestrado.
- [RAM 99] RAM, S.; PARK, J.; BALL, G. L. Semantic-model support for geographic information systems. **IEEE Computer**, Los Alanitos, v.32, n.5, 1999.

- [RAT 97] RATIONAL SOFTWARE CORPORATION. **UML - Unified Modeling Language**. Version 1.1. Santa Clara: Rational Software Corporation, 1997. Disponível em <http://www.rational.com/uml/html/notation/>.
- [RAW 96] RAWSTHORNE, D. A. A patterns language for requirements analysis. In: WORKSHOP IN THE CONFERENCE OF PATTERN LANGUAGE OF PROGRAMS - PLoP, 1996, Monticello-Illinois. **Proceedings...** [S.l:s.n], 1996.
- [REN 97] RENOLEN, A. Conceptual modelling and spatiotemporal information systems: how to model the real world. In: ScanGIS, 1997, Stocolm. **Proceedings...** Stocolm: [s.n.], 1997.
- [RGS 94] RIO GRANDE DO SUL. **Política Estadual de Recursos Hídricos**. Lei n. 10.350, de 30 de dezembro de 1994. Política para conservação e uso dos recursos hídricos.
- [RIB 96] RIBEIRO, G. P.; SOUZA, J. M.; FREITAS, A. L. B. Digital geospatial metadata: a Brazilian case federal databases. In: IEEE METADATA CONFERENCE, 1., 1996, Maryland. **Proceedings...** Maryland: IEEE, 1996.
- [RIE 96] RIEHLE, D.; ZÜLLIGHOVEN, H. Understanding and using patterns in software development. **Theory and Practice of Object Systems**, [S.l.], v.2, n.1, 1996.
- [ROB 93] ROBERTSON, S.; STRUNCH, K. Reusing the products of analysis. In: INTERNATIONAL WORKSHOP ON SOFTWARE REUSABILITY, 2., 1993, Lucca, Italy. **Proceedings...** Lucca: [s.n], 1993. Disponível em <http://www.atlsysguild.com/GuildSite/SQR/reusingAnalysis.html>.
- [ROM 90] ROMAN, G-C. Formal Specification of Geographic Data Processing Requirements. **IEEE Trans. on Knowledge and Data Engineering**, New York, v.2, n.4, p.370-380, 1990.
- [RUG 98] RUGG, R. D.; DOBSON, J. E.; EGENHOFER, M. J. **Formalisms and schema languages for modeling geographic information**. Santa Barbara, CA: Federal Geographic Data Committee, University Consortium for Geographic Information Science, 1998.
- [RUM 91] RUMBAUGH, J. et al. **Object-Oriented Modeling and Design**. New Jersey: Prentice-Hall, 1991.
- [SCH 97] SCHNEIDER, P.; KÄMPF, N.; GIASSON, E. Solos da bacia carbonífera do Baixo Jacuí, RS. In: UFRGS/CENTRO DE ECOLOGIA. **Energia e Meio Ambiente: a questão do carvão no Rio Grande do Sul**. Porto Alegre: UFRGS/Centro de Ecologia, 1997. v.3. Relatório final Projeto PADCT-CIAMB.
- [SEC 92] SECRETARIA DE COORDENAÇÃO E PLANEJAMENTO DO RS. **Programa para o Desenvolvimento Racional, Recuperação e Gerenciamento Ambiental da Bacia Hidrográfica do Guaíba**. Porto Alegre: Secretaria de Coordenação e Planejamento do RS, 1992.
- [SEC 92a] SECRETARIA DE COORDENAÇÃO E PLANEJAMENTO DO RS. **Plano Diretor de Geoinformações**. Porto Alegre: Secretaria de Coordenação e Planejamento do RS, 1992.

- [SEM 99] SEMADS-RJ. **Modelo Conceitual do Projeto de Zoneamento Ecológico Econômico – ZEE**. Rio de Janeiro: Secretaria de Estado de Meio Ambiente e Desenvolvimento Sustentável, 1999.
- [SHE 95] SHELLEY, E. P.; JOHNSON, B. D. Metadata: concepts and models. In: NATIONAL CONFERENCE ON THE MANAGEMENT OF GEOSCIENCE INFORMATION AND DATA, 3., 1995, Adelaide, Australia. **Proceedings...** Adelaide: [s.n.], 1995.
- [SHE 97] SHELLEY, E. P. et al. Spatial metadata in Australia – a learning experience. In: IEEE METADATA CONFERENCE, 2., 1997, Maryland. **Proceedings...** Maryland: IEEE, 1997.
- [SHK 97] SHEKHAR, S. et al. Data models in geographic information systems. **Communications of the ACM**, New York, v.40, n.4, 1997.
- [SIL 96] SILVA, R. P.; PRICE, R. T. Em direção a uma metodologia para o desenvolvimento de frameworks de aplicação orientados a objetos. In: SIMPÓSIO BRASILEIRO DE ENGENHARIA DE SOFTWARE, 10., 1996, São Carlos. **Anais...** São Carlos: SBC, 1996.
- [SOU 97] SOUZA, N. J.; BITTENCOURT, J. L. Aspectos globais da região carbonífera do Rio Grande do Sul. In: UFRGS/CENTRO DE ECOLOGIA. **Energia e Meio Ambiente: a questão do carvão no Rio Grande do Sul**. Porto Alegre: UFRGS/Centro de Ecologia, 1997. v.2. Relatório final Projeto PADCT-CIAMB.
- [SOU 98] SOUZA, C. R. B. **Um framework para editores de diagramas cooperativos baseados em anotações**. Campinas: Unicamp, 1998. Dissertação de Mestrado.
- [SUB 93] SUBRAMANIAN, R.; ADAM, N. R. The design and implementation of an expert object-oriented geographic information system. In: CIKM INFORMATION AND KNOWLEDGE, 2., 1993, Washington. **Proceeding...** Washington: [s.n.], 1993.
- [TIM 94] TIMES, V. C. **MGeo: Um Modelo Orientado a Objetos para Aplicações Geográficas**. Recife: UFPE, 1994. Dissertação de Mestrado.
- [TIM 94a] TIMES, V. C.; SALGADO, A. C. Uma Modelagem Orientada a Objetos para Aplicações Geográficas. In: SIMPÓSIO BRASILEIRO DE BANCO DE DADOS, 9., 1992, São Carlos-SP. **Anais...** São Carlos-SP: SBC, 1994.
- [TIM 96] TIMPF, S.; RAUBAL, M.; KUHN, W. Experiences with metadata. In: SPATIAL DATA HANDLING, 1996, Delft. **Proceedings...** Delft:[s.n.], 1996.
- [TRY 95] TRYFONA, N.; HADZILACOS, T. Geographic applications development: models and tools for the conceptual level. In: ACM-GIS INTERNATIONAL WORKSHOP ON ADVANCES IN GEOGRAPHIC INFORMATION SYSTEMS, 1995, Baltimore. **Proceedings...** Baltimore: ACM, 1995.
- [UFR 97] UFRGS/CENTRO DE ECOLOGIA. **Energia e Meio Ambiente: a questão do carvão no Rio Grande do Sul**. Porto Alegre: UFRGS/Centro de Ecologia, 1997. v.1. Relatório final Projeto PADCT-CIAMB.

- [ULL 82] ULLMAN, J. D. **Principles of Database Systems**. 2. ed. Rockville: Computer Science Press, 1982.
- [VIL 99] VILLAVICENCIO, F.; IOCHPE, C; LISBOA F., J. Um servidor de consultas à clearinghouse em português para distribuição de metadados geográficos. In: GISBRASIL CONGRESSO E FEIRA PARA USUÁRIOS DE GEOPROCESSAMENTO DA AMÉRICA LATINA, 5., 1999, Salvador. **Anais...** Salvador: Fator GIS, 1999.
- [WEB 98] WEBER, E. J.; LISBOA F., J.; IOCHPE, C.; HASENACK, H. Geospatial metadata in Brazil: an experience in data documentation of an environmental GIS application. In: INT. CONFERENCE & EXHIBITION ON GEOGRAPHIC INFORMATION - GISPlanet, 1998, Lisbon, Portugal. **Proceedings...** Lisbon: USIG, 1998.
- [WEB 98a] WEBER, E. J. et al. Estruturação de sistemas de informação ambiental em bacias hidrográficas: o caso da bacia hidrográfica do rio Caí - RS. In: GISBRASIL, 1998, Curitiba. **Anais...** Curitiba: Revista FatorGIS, 1998.
- [WEB 99] WEBER, E. J.; ANZOLCH, R.; LISBOA F., J.; COSTA, A. C.; IOCHPE, C. **Qualidade de dados geoespaciais**. Porto Alegre: PPGC da UFRGS, 1999. (RP-293-99)
- [WIL 91] WILLIAMSON, I., P.; HUNTER, G. J. The need for improved forms of conceptual models in geographical information system development. **Australian Surveyor**, Sidney, v.36, n.2, p.100-110, 1991.
- [WIR 90] WIRFS-BROCK, R.; JOHNSON, R. E. Surveying current research in object-oriented design. **Communication of the ACM**, New York, v.33, n.9, 1990.
- [WOR 90] WORBOYS, M. et al. Object-oriented Data Modeling for Spatial Databases. **International Journal of Geographical Information Systems**, London, v.4, n.4, p.369-384, 1990.
- [WOR 94] WORBOYS, M. F. Review Article: Object-oriented approaches to geo-referenced information. **International Journal of Geographical Information Systems**, London, v.8, n.4, p.385-399, 1994.
- [WOR 95] WORBOYS, M.F. **GIS: A Computing Perspective**. London: Taylor and Francis, 1995.
- [ZIR 95] ZIRBES, S. F. **A reutilização de modelos de requisitos de sistemas por analogia: experimentação e conclusões**. Porto Alegre: CPGCC da UFRGS, 1995. Tese de Doutorado.