CAPÍTULO 4

MODELAGEM DE DADOS EM GEOPROCESSAMENTO

"Na realidade, a ciência e a arte vêm aos homens por intermédio da experiência, porque a experiência criou a arte e a inexperiência, o acaso. E a arte aparece quando, de um conjunto de noções experimentadas, se exprime um único juízo universal de casos semelhantes".

Aristóteles (Metafísica).

4.1 APRESENTAÇÃO

Um *modelo de dados* é um conjunto de ferramentas conceituais utilizado para estruturar dados num sistema computacional. Aspecto fundamental no projeto de um SIG, o modelo descreve como a realidade geográfica será representada no computador. Nenhuma outra decisão limita tanto a abrangência e o crescimento futuro do sistema quanto a escolha do modelo de dados.

A necessidade de modelos gerais que representem avanços conceituais em Geoprocessamento foi assim expressa no programa de pesquisa do NCGIA (1989):

> *" Current GIS are either raster-based or vector-based and have limited abilities to combine the two. (...) While current systems are built using one model or the other, an ideal GIS would offer both and allow full interaction between them".*

Como indicado por Goodchild (1992a), a indústria de Geoprocessamento amadureceu a tal ponto que questões sobre estruturas de dados, algoritmos e funcionalidade contam com respostas padronizadas. Nesta situação, a modelagem de dados tem um papel crítico ao determinar a capacidade de uso e a rapidez de aprendizado do sistema.

Este capítulo apresenta um modelo de dados para Geoprocessamento, projetado com técnicas de orientação-por-objetos e que permite lidar com os diversos tipos de dados ambientais. As contribuições mais relevantes do modelo são:

- Apresentar uma abordagem unificada para a modelagem em Geoprocessamento, combinando as idéias de campos e objetos.
- Integrar imagens de Sensoriamento Remoto e Modelos Numéricos de Terreno com mapas temáticos, mapas cadastrais e redes.
- Prover suporte para representações geométricas múltiplas de uma mesma entidade do mundo real e permitir a coexistência de representações vetorial, matricial e grades num mesmo sistema.
- Engendrar uma interface de alto nível com conteúdo semântico.

Este modelo está sendo utilizado como base no projeto e implementação do sistema SPRING (Câmara et al., 1992) em sua versão 2.0. No âmbito do SPRING, o modelo serve como base para a definição de interface com o usuário, e de uma biblioteca de classes para SIGs em C++.

4.2 TRABALHOS ANTERIORES NA LITERATURA

4.2.1 VISÃO GERAL

O trabalho mais influente neste assunto é o de Goodchild (1992b). Nele, o autor apresenta as noções de *campos* e *objetos* como representações conceituais para dados geográficos. A maior parte dos trabalhos relativos a questão de modelos de dados em Geoprocessamento utiliza o paradigma de orientação por objetos, e serão descritos aqui em prioridade.

4.2.2 MODELAGEM ORIENTADA-A-OBJETOS EM GEOPROCESSAMENTO

O uso de conceitos de orientação por objetos em Geoprocessamento foi analisado pela primeira vez em Egenhofer e Frank (1987). Neste trabalho, os autores mostram como aplicar conceitos de *classificação, especialização, generalização e agregação* ao projeto de SIGs. Para uma revisão recente dos resultados teóricos em uso de orientação por objetos em projeto de SIG, veja-se Egenhofer (1995).

Uma das principais linhas de pesquisa deriva da comunidade de bancos de dados, ao procurar estender os conceitos de bancos de dados orientados a objetos ao domínio geográfico. Merecem destaque os trabalhos de Worboys (Worboys et al., 1990; Worboys, 1994), do grupo ligado ao SGBD O_2 (Scholl and Voisard, 1991; David and Voisard, 1993; Rigaux and Scholl, 1995) e de grupos brasileiros (Times e Salgado, 1995).

Os modelos propostos pela comunidade de banco de dados tendem a analisar um banco de dados geográfico como uma coleção de entidades individuais (objetos). Exemplo pode ser visto no trabalho de David e Voisard (1993):

> *"A map is a collection of geographic objects. A geographic object is usually derived from an entity of the real world. It has a spatial part as well as an alphanumeric description."*

Esta visão confunde o uso de modelos de dados orientados-a-objetos com a decomposição do espaço em objetos discretos. Veja-se este trecho do trabalho de Worboys (1994):

"There are two broad and opposing classes of models of geographical information. The class of field-based models treats such informations as collections of spatial distributions. (...) Patterns of topographic altitudes, rainfall and temperature fit this view. The class of entity-based models treats the information space as populated by discrete, identifiable entities (or objects), each with a geo-reference. (...) These two models result in two opposed GIS implementation approaches, namely vector and raster. Object-oriented models decompose the space into objects".

Particularmente interessante é o conceito de "realm", introduzido por Güting e Schneider (1993). Um *realm* descreve a geometria do espaço geográfico utilizando uma base discreta (conjuntos de coordenadas inteiras). Esta idéia é utilizada para desenvolver um álgebra formal (Güting et al., 1995) que define operações sobre o espaço discreto de um *realm*.

O modelo de dados orientado-a-objetos mais completo disponível é o proposto para o padrão SAIF (Spatial Archive and Interchange Format), utilizado no Canadá como formato de arquivo e intercâmbio de dados geográficos (Survey and Resources Mapping Branch, 1993; Sondheim, 1993). O padrão SAIF está sendo utilizado como base para implementar as extensões espaciais da linguagem SQL 3.

4.2.3 HISTÓRICO DOS TRABALHOS NO INPE

O modelo de dados proposto reflete a experiência do sistema SGI/INPE. Desenvolvido a partir de 1984, o SGI teve grande aceitação no mercado nacional para aplicações ambientais. O modelo do SGI foi desenvolvido com base nos conceitos de abstração de dados e contemplava a noção de *plano de informação* (Câmara, Velasco e Oliveira, 1982). Tratava-se essencialmente de modelo adequado a análise temática, que embutia a noção de geo-campos.

Uma noção ainda preliminar de *geo-objetos* foi introduzida em Erthal, Alves e Câmara (1988), que foi a base de implementação da primeira versão do SPRING. Posteriormente, Alves (1990) apresentou um modelo de dados orientado-a-objetos que apresentava avanços conceituais importantes.

As discussões que levaram à proposta deste modelo foram conduzidas num grupo aonde participaram Marco Casanova, Andrea Hemerly, Ubirajara Freitas, Ricardo Cartaxo, João Argemiro Paiva e o autor. As noções de *categoria* e de *visual* são devidas a Ricardo Cartaxo e a de *banco de dados geográficos*, a Ubirajara Freitas. Coube ao autor formalizar os conceitos, explicitar os diferentes níveis conceituais, introduzir a noção *de geo-objetos e objetos cadastrais* e elaborar a forma final do modelo.

O autor expressa ainda seu débito a Diógenes Alves pelas idéias expressas em Alves (1990) e por ter-lhe indicado o trabalho de Goodchild (1992b).

4.2.4 COMPARAÇÃO COM TRABALHOS DA LITERATURA

O modelo de dados aqui apresentado distingue-se dos trabalhos anteriores por sua visão abrangente das classes de dados presentes em SIGs e na separação entre os universo conceitual e o universo de representação. No nível conceitual, o modelo é uma evolução das idéias de Goodchild (1992b) e seu universo de representação é compatível com o padrão SAIF.

4.3 VISÃO GERAL DO PROCESSO DE MODELAGEM

O processo de Modelagem é a forma que dispomos para traduzir o mundo real em outros domínios. Um das abordagens mais úteis para este problema é o chamado "paradigma dos quatro universos" (Gomes e Velho, 1995), que distingue:

- O universo do *mundo real*, que inclui as entidades da realidade a ser modeladas no sistema.
- O universo *matemático (conceitual),* que inclui uma definição matemática (formal) das entidades a ser incluídas no modelo.
- O universo de *representação*, onde as diversas entidades formais são mapeadas para representações geométricas.
- O universo de *implementação*, aonde as estruturas de dados e algoritmos são escolhidos, baseados em considerações como desempenho, capacidade do equipamento e tamanho da massa de dados. É neste nível que acontece a codificação.

Esta visão do processo de modelagem está ilustrada na figura 4.1.

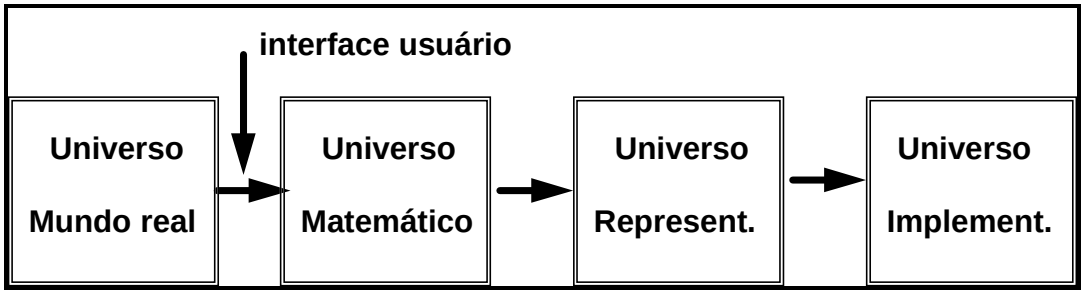

Figura 4.1 - Modelagem segundo o paradigma dos quatro universos.

A visão apresentada não se limita a sistemas de Geoprocessamento, mas representa uma perspectiva unificadora aos problemas de Computação Gráfica e Processamento de Imagens. Sua aplicação ao problema de Geoprocessamento é particularmente apropriada pois permite equacionar os problemas da área, como se pode constatar:

- no universo do *mundo real*, encontramos os fenômenos a ser representados (tipos de solo, cadastro urbano e rural, dados geofísicos e topográficos).

- no universo *conceitual (matemático)*, pode-se distinguir entre as grandes classes formais de dados geográficos (*dados contínuos* e *objetos individualizáveis*), e especializar estas classes nos tipos de dados geográficos utilizados comumente (dados temáticos e cadastrais, modelos numéricos de terreno, dados de sensoriamento remoto).

- no universo de *representação*, as entidades formais definidas no universo conceitual são associadas a diferentes representações geométricas, que podem variar conforme a escala e a projeção cartográfica escolhida e a época de aquisição do dado. Aqui se distingue entre as representações matricial e vetorial, que podem ainda ser especializadas.

- o universo de *implementação* é onde ocorre a realização do modelo de dados através de linguagens de programação. Neste universo, escolhemos estruturas de dados (tais como árvores quaternárias e árvores-R) para implementar as geometrias do universo de representação.

Com base nesta visão, as dicotomias tradicionais de Geoprocessamento (campos-objetos e matricial-vetorial) podem ser resolvidas, mostrando-se que elas encontram-se *em níveis distintos de abstração.*

Esta análise também indica que a interface de usuário de um SIG deve, tanto quanto possível, refletir o universo conceitual e esconder detalhes dos universos de representação e implementação. No nível conceitual, o usuário lida com conceitos mais próximos de sua realidade, e minimiza a complexidade envolvida nos diferentes tipos de representação geométrica[1].

[1]No entender do autor, este representa um dos principais diferenciais do software SPRING com relação a seus concorrentes.

4.4 O UNIVERSO DO MUNDO REAL

Pretende-se construir um modelo não limitado a uma particular área de aplicação de Geoprocessamento e que possa incorporar aplicações tão diversas como Estudos Ambientais, Agricultura, Geologia e Redes.

Para tanto, será preciso mapear as diversas áreas de aplicação de Geoprocessamento, em função dos tipos de dados utilizados, como ilustrado na Tabela 4.1. Como se verá a seguir, o modelo proposto é aplicável a todas as áreas listadas, ilustrativas da diversidade de problemas em Geoprocessamento.

Tabela 4.1

Aplicações típicas de Geoprocessamento

Aplic.	*Escalas típicas*	*Tipos de dados*	*Represent. Gráficas*	*Operações*
Floresta	1:10.000 a 1:500.000	cadastro rural (talhões)	matricial, vetorial	classif. imagens, consulta espacial
Agricult.	1:25.000 a 1:250.000	dados temáticos, S. Remoto,MNT	matricial,vetorial grades, TIN	análise espacial, declividade
Geologia	1:100.000 a 1:2,500,000	MNT, dados temáticos,imagens	matricial, grades	transf. IHS, visualiz. 3D
Redes	1:1.000 a 1:10.000	redes lineares (topologia)	vetorial	consulta espacial, cálculos dedicados
Estudos Urbanos	1:1.000 a 1:25.000	redes, cadastro urbano	vetorial	consulta espacial

4.5 O UNIVERSO CONCEITUAL

4.5.1 VISÃO GERAL

Da análise do universo do mundo real pode-se constatar que as aplicações de Geoprocessamento lidam com dois grandes tipos de dados (Goodchild, 1992b):

- *Geo-Campos*: correspondem a grandezas distribuídas espacialmente, como tipo de solo, topografia e teor de minerais.
- *Objetos geográficos (ou geo-objetos)*: individualizáveis e tem identificação com elementos do mundo real, como lotes num cadastro urbano e postes numa rede elétrica. Estes objetos tem atributos não-espaciais e podem estar associados a mais de uma representação geométrica.

O aspecto fundamental da distinção entre campos e geo-objetos é a questão da identidade. Existem milhares de áreas no Brasil classificáveis como "latossolo roxo", mas há apenas um "Jardim Botânico do Rio de Janeiro".

Os *geo-campos* obedecem ainda ao princípio do "*planar enforcement*" (restrição de preenchimento do plano): cada localização de um *geo-campo* está associada a um e somente um valor da variável espacial representada.

No caso de *geo-objetos*, o princípio do "planar enforcement" não precisa ser obedecido, pois os objetos podem ser disjuntos no espaço ou o mesmo local do espaço corresponder a mais de um objeto. Tome-se, a título de exemplo, os casos de solos e de rede elétrica. Num geo-campo de solos, cada local do espaço está associado a um tipo específico de solo (p.ex. Latossolo Roxo); numa de rede elétrica, num mesma localização podem conviver geo-objetos distintos (um poste com um transformador e uma chave).

4.5.2 UM MODELO PARA DADOS GEOGRÁFICOS

Nossa abordagem segue a filosofia do modelo de dados do SGBD orientado-a-objetos O_2 (Lécluse, Richard e Velez, 1991), que suporta os conceitos de *identidade de objetos*, *tipos de dados* e *herança*, e considera os seguintes conjuntos disjuntos:

- O conjunto *I* de identificadores de dados geográficos.
- O conjunto *C* de nomes de classes geográficas.
- O universo *U* de atributos *descritivos* $\{A_1,..., A_n\}$, definidos nos domínios $D(A_1),...,D(A_n)$. Denotaremos por a_i um valor possível do atributo A_i.

Seguiremos os seguintes passos:

1. Definir as classes básicas do modelo e estabelecer as relações entre elas, dentro dos princípios de especialização, generalização e agregação.
2. Estabelecer como, a partir do modelo, é possível definir um esquema conceitual para um banco de dados geográfico, por especialização das classes básicas.

4.5.3 CLASSES DO UNIVERSO CONCEITUAL

As classes básicas do modelo, definidas a seguir, são: *geo-campo, geo-objeto, objeto cadastral, objetos não-espaciais, plano de informação e banco de dados geográfico.*

Definições Preliminares

Inicialmente, será importante estabelecer a base geométrica na qual as classes do modelo são definidas. Seguindo a idéia de Güting e Schneider (1993), utilizamos um conjunto *discreto* de pontos como suporte para as entidades espaciais.

Definição 4.1. Projeção geográfica planar.

Dadas uma região da superfície terrestre definida por suas coordenadas de latitude e longitude $\varphi = (\varphi_0, \varphi_f)$, $\lambda = (\lambda_0, \lambda_f)$ e uma transformação cartográfica ρ: $(\vartheta, \Lambda) \rightarrow \Re^2$ que mapeia as coordenadas terrestres (φ, λ) para pontos no $\Re^2$, uma *projeção geográfica planar G* induzida por φ, λ e ρ é uma região do $\Re^2$ definida como:

$$(x_0, y_0) = \rho(\varphi_0, \lambda_0), \quad (y_f, x_f) = \rho(\varphi_f, \lambda_f)$$

$$G = \{ (x,y) \in \Re^2 \mid x_0 \leq x \leq x_f, \quad y_0 \leq y \leq y_f \}.$$

A partir de uma região contínua da superfície terrestre, podemos definir nosso conceito de *região geográfica* (ou *reticulado geográfico*).

Definição 4.2. Região Geográfica.

Dada uma projeção planar *G*, uma escala χ associada a esta projeção e $\Delta x, \Delta y \in \Re$, uma *região geográfica* $R = [G, \chi, \Delta x, \Delta y]$ é um reticulado conectado contido no plano Z^2 tal que:

$$R = \{ (x_i, y_j) \in G \mid x_i = i.\Delta x, \quad y_j = j.\Delta y \quad \text{para } i, j \in Z \}$$

A região geográfica *R* serve de suporte geométrico para localizações geográficas, pois toda uma localização geográfica será representada por um ponto de R. O uso de um conjunto discreto de pontos facilita uma definição formal das classes de dados geográficas e das operações associadas. A definição de região geográfica proposta não restringe a escolha da representação geométrica (matricial ou vetorial) associada aos objetos geográficos.

Sabemos, da Cartografia, que uma área contínua da superfície da Terra será mapeada para uma superfície plana no $\Re^2$ através de transformações projetivas, para uma projeção cartográfica e escala definidas. A teoria cartográfica também nos ensina (Joly, 1990) que todo mapa possui uma precisão associada, que é usualmente definida em função de sua escala.

Com base nestas considerações, associaremos as dimensões Δx, Δy do reticulado *R* à precisão cartográfica necessária para representar univocamente cada elemento geográfico, dadas uma área da superfície terrestre, uma escala e uma projeção. Deste modo, mesmo se posteriormente (no universo de representação), escolhermos uma geometria vetorial para objetos definidos no universo conceitual, o reticulado R ainda poderá ser utilizado como base geométrica.

Geo-Campos

Definição 4.3. Geo-Campo.

Um geo-campo é um objeto $f = [R, V, \lambda]$, onde R é uma região geográfica, V um domínio de valores e $\lambda: R \rightarrow V$ um mapeamento entre pontos em R e valores em V.

Um geo-campo representa a distribuição espacial de uma variável que possui valores em todos os pontos de uma região geográfica. Os geo-campos podem ser especializados em:

- TEMÁTICO : a função λ caracteriza um *geo-campo temático* quando define um mapeamento $\lambda: R \rightarrow V$ tal que V é um conjunto finito enumerável. Os elementos de V definem os *temas* de um geo-campo (p.ex. um geo-campo de pedologia é caracterizado pelo conjunto de temas {latosolo roxo, litosolo, cambisolo, ...}).

- NUMÉRICO - uma instância desta classe, chamada de geo-campo numérico (ou MODELO NUMÉRICO DE TERRENO) é definida através de uma função $\lambda: R \rightarrow V$ tal que V é o conjunto dos reais (p. ex. um geo-campo de altimetria).

- DADO_SENSOR_REMOTO - esta classe é uma especialização de NUMÉRICO, definida a através de uma função $\lambda: R \rightarrow V$, tal que V é o conjunto dos naturais. Na prática, uma instância desta classe é obtida através de discretização da resposta recebida por um sensor (passivo ou ativo) de uma área da superfície terrestre.

Denotamos a classe de geo-campos por GEOCAMPO[2].

No caso de geo-campo temático, cada tema pode ser composto de um ou mais qualificadores[3] e pode ter atributos. Por exemplo, num geo-campo de solos,

[2]Na notação utilizada neste trabalho, o uso de *SMALLCAPS* denota uma classe.

cada tema de solo ({latosolo roxo, litosolo, cambisolo, ...}) pode estar associado a uma tabela de atributos (como teor de sódio, nitrogênio e potássio, pH do solo, umidade).

Note-se que um dado de sensor remoto é um caso particular de geo-campo numérico, onde o valor é proporcional à resposta do solo medida por um sensor numa faixa do espectro eletromagnético.

As figuras abaixo apresentam exemplos de geo-campos.

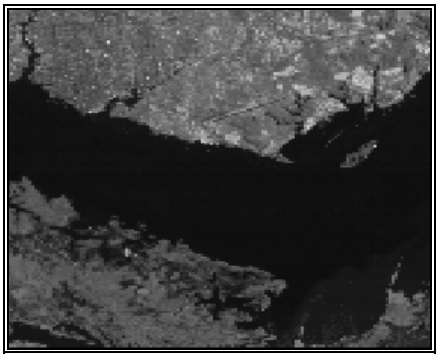

Figura 4.2 - Dado Sensor Remoto (Imagem LANDSAT de Manaus).

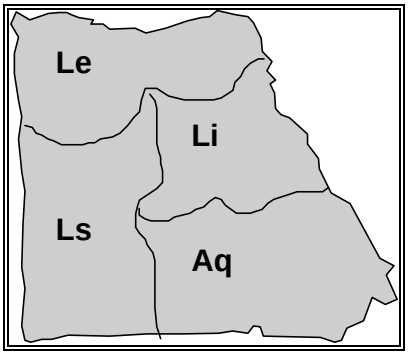

Figura 4.3 - Geo-campo temático (solos).

[3]O caso mais usual é um mapa temático onde cada localização está associada a um tema único. No entanto, é muito útil poder definir uma situação aonde mais de um conjunto de temas está associado a cada localização.

Especializações de Geo-Campos

As classes básicas TEMÁTICO, MNT e DADO_SENSOR_REMOTO podem ser utilizadas para derivar uma hierarquia de especialização, que reflita o domínio de aplicação. Por exemplo, para estudos de uso do solo, poderíamos criar uma classe TIPO_DE_SOLO (especialização de TEMÁTICO), cujas instâncias são os geo-campos de uso do solo da região de estudo, e também criar uma classe IMAGEM_LANDSAT (especialização de DADO_SENSOR_REMOTO) cujas instâncias seriam imagens do satélite LANDSAT sobre a região de estudo.

Geo-Objeto

Definição 4.4. Geo-Objeto

Dados um conjunto de regiões geográficas $R_1,...R_n$ e um conjunto de atributos $A_1,...A_n$ com domínios $D(A_1),...,D(A_n)$, um *geo-objeto* é um objeto $go = [a_1,...a_n, S_1,...,S_m]$, composto pelos valoes a_i (aonde $a_i \in D(A_i)$) e por localizações geográficas S_i (aonde $S_i \subseteq R_i$). Denotamos o valor do *i-ésimo* atributo descritivo de *go* por $go.A_i$ e a *i-ésima* localização geográfica de *go* por $go.R_i$.

Em outras palavras, um geo-objeto é um elemento único, que possui atributos descritivos e está associado a múltiplas localizações geográficas. Esta definição tem três grandes motivações adicionais:

1. *As projeções cartográficas:* A projeção planar da Terra, a partir de escalas macroregionais, é feita com o uso de quadrículas que estão particionados em sistemas de referência independentes, que definem recortes arbitrários no espaço e podem dividir a localização de um geo-objeto. Por exemplo, um particionamento cartográfico da Amazônia na projeção UTM, escala 1:250.000 faz com que os principais rios tenham representações geométricas descontínuas em vários mapas.
2. *Representações geométricas em diferentes escalas:* Na prática, num mesmo banco de dados geográfico, podem conviver representações da mesma realidade

geográfica em diferentes escalas geográficas. Por exemplo, consideremos um conjunto de mapas dos municípios do Estado de São Paulo, que inclui um mapa geral (na escala de 1:1.000.000) e mapas regionais (na escala de 1:500.000). Nesta situação, um mesmo geo-objeto (p.ex., o munícipio de São José dos Campos) teria duas representações geométricas: uma no mapa geral e outra no mapa regional do Vale do Paraíba.

3. *Múltiplas representações temporais*: As diferentes representações de um mesmo objeto podem corresponder a variações temporais do mesmo, como no caso de um lago que teve suas bordas alteradas[4].

Para ilustrar este conceito, consideremos a figura 4.4, que mostra um banco de dados da Amazônia, onde os quadrados pontilhados representam o recorte espacial do banco de dados. Entidades como os rios Amazonas e Xingú têm representações em diferentes particionamentos espaciais do banco de dados.

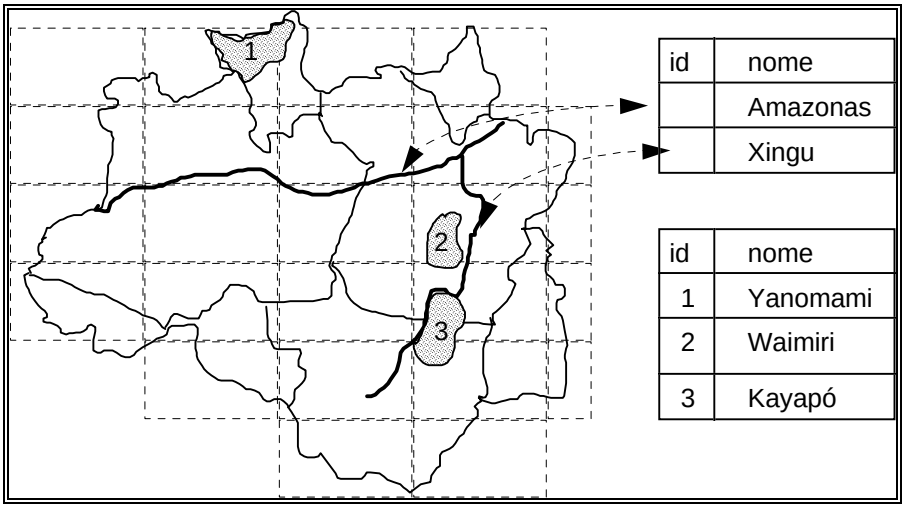

Figura 4.4 - Princípio de identidade em bancos de dados geográficos.

[4]Um caso particularmente dramático é o Mar de Aral, na ex-URSS.

Objeto não-espacial

Em muitas situações, é conveniente permitir a associação de informações não-espaciais a um banco de dados geo-referenciado. Por exemplo, considere-se uma aplicação de cadastro urbano em uma prefeitura, que já dispõe de um sistema para cálculo do IPTU baseado num cadastro alfanumérico de lotes. Neste caso, pode-se desejar associar o cadastro alfanumérico a dados georeferenciados contendo a localização geográfica destas propriedades. Para englobar estas entidades, introduzimos a noção de objeto não-espacial.

Definição 4.5. Objeto não-espacial

Dado um conjunto de atributos $A_1,...A_n$, um objeto não-espacial é um objeto $on = [a_1,...a_n]$ definido pelos atributos $a_i \in D(A_i)$, e que não possui localizações espaciais associadas.

Assim, a noção de objeto não-espacial engloba qualquer tipo de informação que não seja georeferenciada e que se queira agregar a um SIG. Denotamos a classe de objeto não-espaciais por NÃOESPACIAL.

O exemplo anexo mostra o caso de uma aplicação de cadastro rural. Neste caso, temos os geo-objetos da classe "fazendas" (que estão referenciados espacialmente) e desejamos estabelecer a ligação entre estes geo-objetos e a informação alfanumérica já existente, sob a forma de um cadastro de propriedades. Neste caso, as informações de cadastro são consideradas um *objeto não-espacial*.

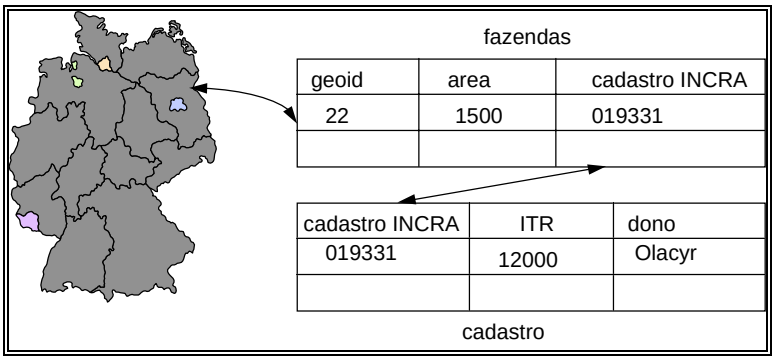

Figura 4.5 - Exemplo de ligação entre geo-objeto e objeto não-espacial.

Objeto Cadastral (Complexo)

Nossa definição de geo-objeto permite a associação de objetos geográficos a diferentes regiões no espaço. Como as aplicações de Geoprocessamento usualmente não armazenam ou manipulam elementos isolados, é conveniente armazenar a representação geométrica de um geo-objeto em conjunto com seus vizinhos, mantendo as relações de topologia. Por exemplo, num cadastro urbano, os lotes de um mesmo bairro são armazenados e apresentados em conjunto. Esta característica nos leva a introduzir a idéia de *objeto cadastral*, um *objeto complexo* que agrupa geo-objetos para uma dada projeção cartográfica e região geográfica, sendo instâncias da classe CADASTRAL.

Definição 4.6. Objeto Cadastral.

Um objeto cadastral é um objeto $oc=[R,GO,geo]$, onde R é uma região geográfica, GO é um conjunto de geo-objetos e $geo: GO \rightarrow 2^R$ é um mapeamento que atribui a cada geo-objeto $go \in GO$, um subconjunto não-vazio de 2^R (ou seja, uma localização no mapa).

A relação entre o objeto cadastral e os geo-objetos nele presentes é ESTÁ_MAPEADO_EM, definida pelo mapeamento *geo(go)*. A classe dos objetos cadastrais será denotada por CADASTRAL. A figura 4.6 ilustra um exemplo de objeto cadastral, para o caso de um cadastro urbano. Cada objetos cadastral deste classe contém os mapeamentos dos geo-objetos das classes LOTES, HOSPITAIS e ESCOLAS.

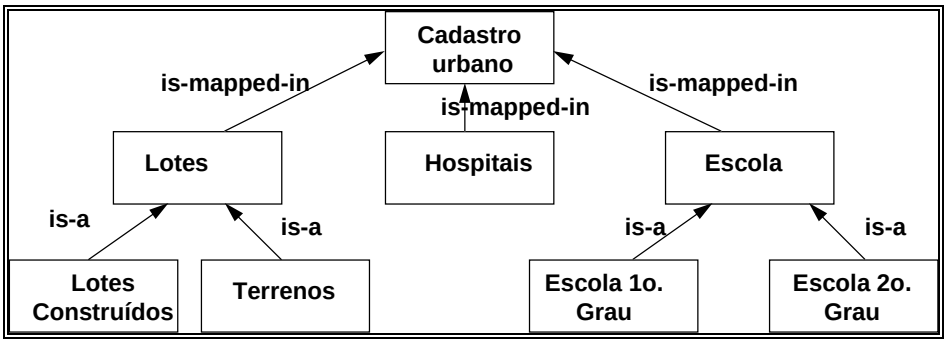

Figura 4.6 - Exemplo de objetos complexos (cadastro urbano).

Para ilustrar o conceito e o uso de objetos cadastrais, consideramos a ainda a figura 4.7, onde temos um caso de um objeto cadastral (cadastro urbano) contendo os lotes de uma parte do bairro "Jardim Esplanada" em São José dos Campos. Os

geo-objetos da classe "lotes" incluem todos os lotes existentes na cidades (na figura apresentados sob forma de tabela). Como o obejto cadastral está limitado espacialmente, contém apenas as representações geométricas de uma parte dos geo-objetos (na figura, indicados pelas setas).

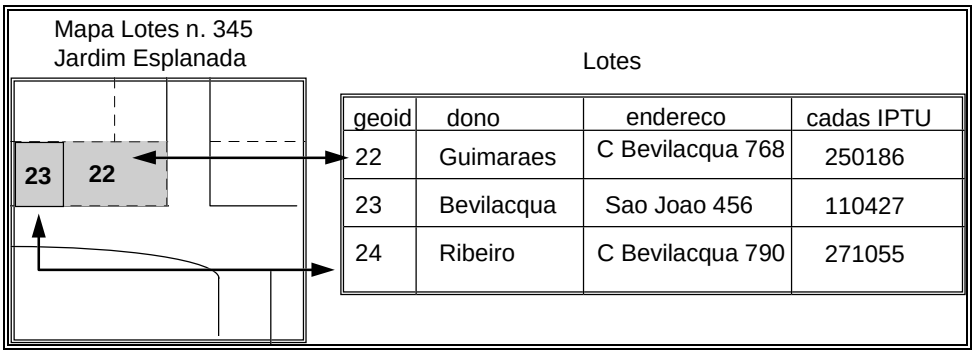

Figura 4.7 - Ilustração do conceito de objeto cadastral.

Consideramos ainda a especialização da classe CADASTRAL na classe REDE. Uma instância desta classe é um objeto $oc=[R,GO,geo]$, onde o mapeamento $geo:GO \rightarrow 2^R$ é tal que $geo(go)$ é um ponto ou uma linha em R. Como exemplo desta classe de objetos, considere-se a especilização da classe REDE na classe REDE ELÉTRICA. Uma instância desta classe conterá os mapeamentos de diferentes classes de geo-objetos, como mostrado na figura 4.8.

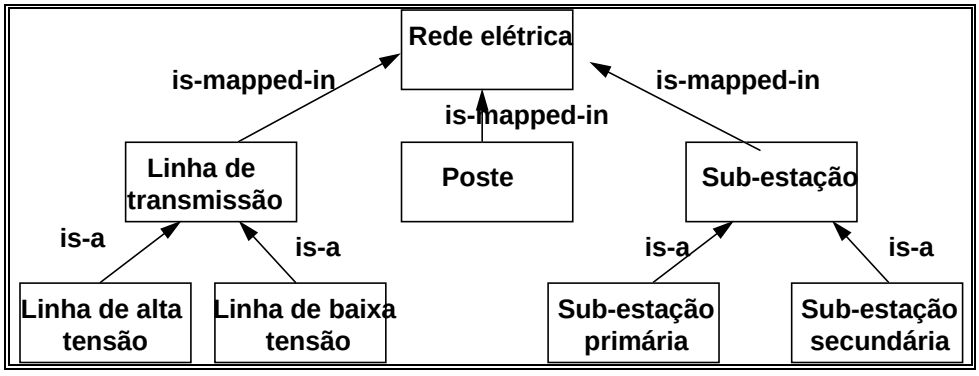

Figura 4.8 - Exemplo de objeto complexo (rede elétrica).

A figura 4.9 mostra o caso de uma rede elétrica, representada através de uma topologia arco-nó. Note-se que uma mesma localização geográfica está associada a dois geo-objetos (um poste e um transformador).

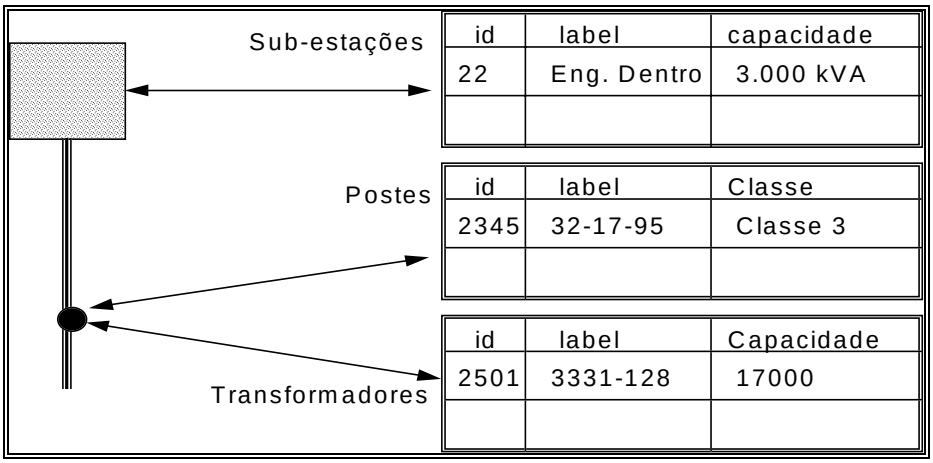

Figura 4.9 - Ilustração do conceito de rede.

Plano de Informação

Um conceito muito útil, para fins de definição de interface e de operações, é a idéia de plano de informação. Trata-se da generalização dos conceitos de geo-campos e de objeto cadastral.

Definição 4.7. Plano de informação

Um plano de informação é um objeto cadastral ou um geo-campo.

Uma instância da classe PLANO DE INFORMAÇÃO representa, para uma dada região geográfica, o lugar geométrico de um conjunto de dados geográficos (um geo-campo ou um objeto cadastral).

Banco de dados geográficos

Definição 4.8. Banco de dados geográficos

Um banco de dados geográficos é um objeto $bdg = [P, GO, ON]$, composto por um conjunto de planos de informação $P = \{pi_1, pi_2, ..., pi_n\}$, um de geo-objetos $GO = \{go_1, go_2, ..., go_n\}$ e um de objetos não-espaciais $ON = \{on_1, on_2, ..., on_n\}$.

Esta definição é particularmente interessante pois não faz restrição sobre escala dos dados e nem sobre a continuidade espacial dos planos de informação que compõem o banco. Tome-se, por exemplo, um banco de dados geográficos sobre a Amazônia Legal, com os seguintes dados:

- Vegetação (fito-fisionomia), especialização da classe TEMÁTICO, com 780 temas de vegetação definidas pelo IBGE (projeto RADAM): 26 cartas digitalizadas a partir de mapas na escala 1:1.000.000, projeção Lambert.
- Imagens de satélite, especialização da classe DADO_SENSOR_REMOTO, obtidas pelo satélite LANDSAT TM, em três anos (1986, 1990 e 1994): 228 imagens, projeção UTM.
- Geo-campos de uso do solo, especializações da classe TEMÁTICO, com as classes: floresta primária, floresta secundária, floresta de várzea, cerrado,

área de regeneração, rios e nuvem. Os mapas finais serão compatíveis com a escala 1:250.000, projeção UTM.

- Zoneamento econômico-ecológico do estado de Roraima, especializaçào das classe CADASTRAL, com 15 planos de informação gerados a partir do estudos na escala 1:100.000.

4.5.4 RESUMO DO UNIVERSO CONCEITUAL

Em resumo, o universo de representação do modelo tem como entidade básica um *banco de dados geográfico*. Este banco é composto por *planos de informação*, por *geo-objetos* e por *objetos não-espaciais*. Os planos de informação podem ser *geo-campos* ou *objetos cadastrais*. Estas classes podem ser ainda especializadas para construir o esquema conceitual do banco de dados geográfico. A figura 4.10 ilustra estas relações.

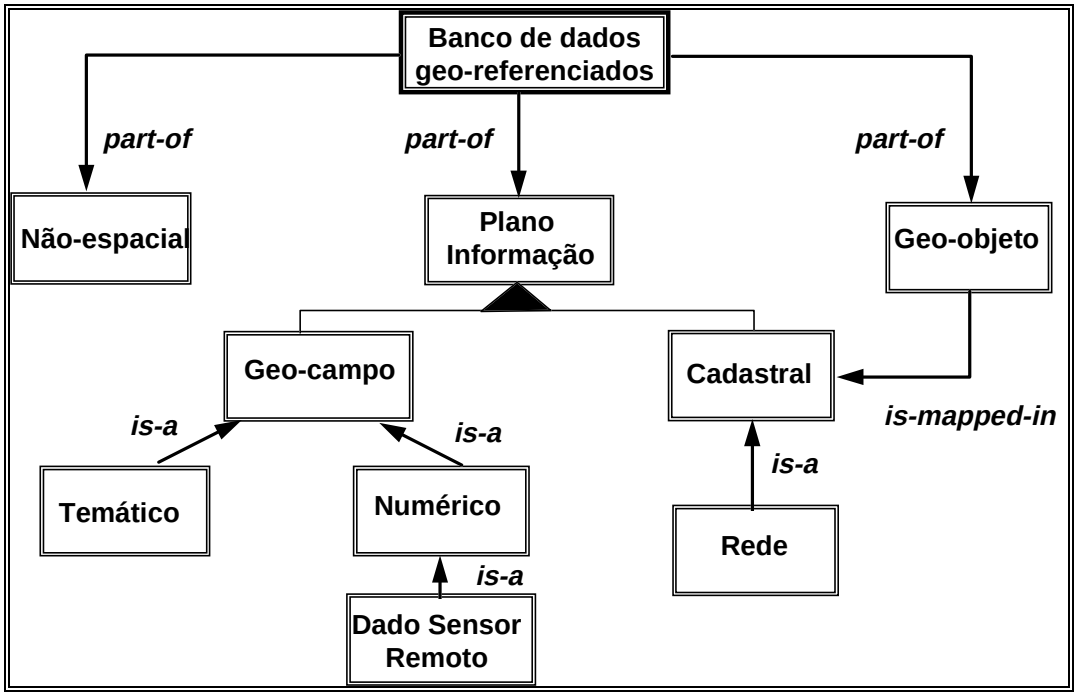

Figura 4.10 - Universo conceitual.

4.6 UNIVERSO DE REPRESENTAÇÃO

No universo de representação, estaremos definindo as possíveis representações geométricas que podem estar associadas às classes do universo conceitual. Inicialmente, devemos considerar as duas grandes classes de representações geométricas: REPRESENTAÇÃO VETORIAL e REPRESENTAÇÃO MATRICIAL.

Para detalhar o universo de representação, adotamos proposta semelhante à apresentada pelo padrão SAIF (Sondheim, 1993). A estrutura aqui proposta é uma simplificação do modelo do SAIF, e aproxima-se bastante da implementação utilizada no SPRING-2.0.

4.6.1 HIERARQUIA DE REPRESENTAÇÕES MATRICIAIS

A REPRESENTAÇÃO GEOMÉTRICA MATRICIAL pode ser especializada segundo a hierarquia de classes mostrada na figura 4.11.

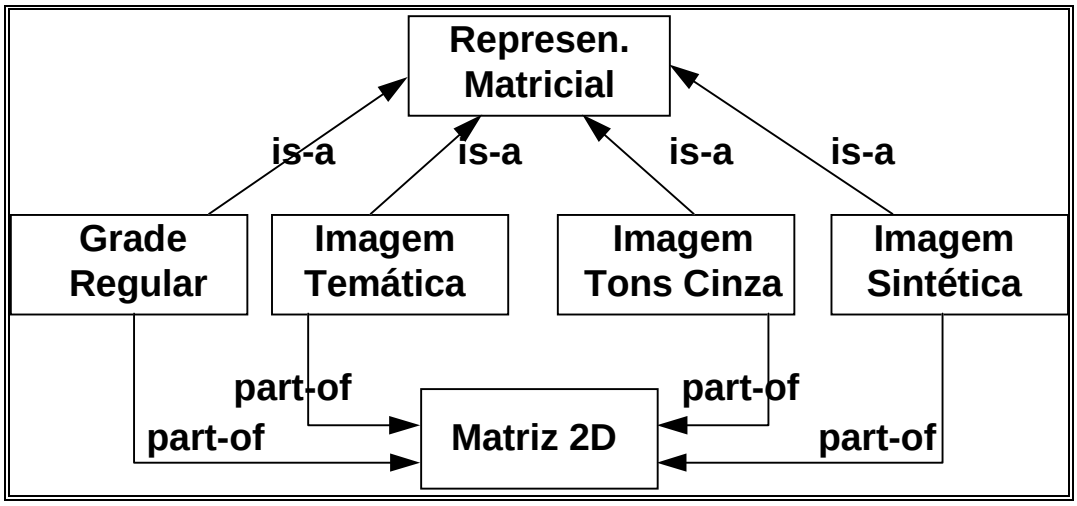

Figura 4.11 - Hierarquia de classes para REPRESENTAÇÃO MATRICIAL.

As classes derivadas de REPRESENTAÇÃO MATRICIAL são:

- GRADE REGULAR: uma grade regular é uma matriz de reais.
- IMAGEM EM TONS DE CINZA: imagem representada através de uma matriz sobre Q, onde $Q \subset \aleph$ denota o conjunto dos níveis de quantização da imagem.
- IMAGEM TEMÁTICA: representação matricial de um geo-campo TEMÁTICO, dada por uma tripla *(M, V, T)*, onde M é uma matriz sobre $Q \subset \aleph$, V são os valores possíveis do geo-campo, e T: Q→V uma função injetiva que associa valores em Q a temas em V. Por exemplo, numa imagem temática, um elemento da matriz de valor 2 pode estar associado ao tema "Floresta Ombrófila".
- IMAGEM SINTÉTICA: representação de uma imagem em cores, dada por um par *(M,T)*, onde *M* é uma matriz sobre $Q \subset \aleph$ e por T uma função que associa, para cada valor $q \in Q$, uma tripla de valores inteiros *(r, g, b)* que definem uma composição colorida do tipo RGB (vermelho, verde, azul). Utilizada para mostrar imagens em composição colorida em placas gráficas falsa-cor.

4.6.2 HIERARQUIA DE REPRESENTAÇÕES VETORIAIS

Para definir esta hierarquia, precisamos inicialmente precisar melhor o que entendemos pelas primitivas geométricas: *coordenadas2D, coordenadas 3D*, *nó2D, nó rede, arcos, arcos orientados, isolinhas e polígonos* (A definição de polígono foi obtidas em Puppo e Dettori, 1995).

Dada uma região geográfica R definida como

$$R = \{ (x_i, y_j) \subset [x_0, x_f] \times [y_0, y_f] \; ; \; x_i = i.\Delta x, \; y_j = j.\Delta y \qquad i, j \in Z, \; \Delta x, \Delta y \in \Re \},$$

podemos definir:

- COORDENADA2D: Uma *coordenada 2D* é um objeto composto por uma localização singular $(x_i, y_j) \in R$.
- COORDENADA3D: Uma *coordenada 3D* é um objeto composto por uma localização singular (x_i, y_j, z), onde $(x_i, y_j) \in R$.
- PONTO2D: Um *ponto 2D* é um objeto que possui atributos descritivos e uma *coordenada 2D*.
- LINHA2D: Uma *linha 2D* possui atributos e inclui um conjunto de *coordenadas 2D* $\{ (x_1,y_1), ..., (x_n,y_n) \; ; \; (x_i, y_i) \in R, \; x_{i+1} = x_i$ ou $y_{i+1} = y_i \}$. Denotaremos a primeira coordenada de uma linha l por $l(0)$, a última por $l(1)$ e o conjunto das demais por l^0.
- ISOLINHA: uma *isolinha* contém uma *linha 2D*, associada a um valor real (cota).
- ARCO ORIENTADO: um arco orientado contém uma *linha 2D*, associada a uma orientação de percorrimento.
- NÓ2D: um *nó 2D* inclui uma *coordenada2D* $(x_i, y_i) \in R$ e uma lista L de linhas 2D, tal que $\forall \; l_i, l_j \in L \; (l_i \neq l_j)$, $(x_i, y_i) = l_i(1) = l_j(0)$. Trata-se da

conexão entre duas ou mais linhas, utilizada para manter a topologia da estrutura;

- NÓ REDE: um *nó de rede* contém um *nó 2d e* uma lista de *arcos orientados*, onde a cada instância associamos uma impedância e um custo de percorrimento.

- NÓ 3D: uma instância desta classe contém uma *coordenada 3D* $(x_i, y_i, z))$ e um lista L de *linhas 2D*, tal que $\forall\ l_i, l_j\ \varepsilon\ L\ (l_i \neq l_j)$, $(x_i, y_i) = l_i(1) = l_j(0)$. Trata-se da conexão entre três ou mais linhas de uma grade triangular.

- POLÍGONO: Para caracterizar a classe dos polígonos, precisamos inicialmente de uam definição auxiliar, a de *área*. Uma *área* é um objeto a formado por um conjunto de linhas 2D, tal que $a = \{ l_1,..,l_n$ e $l_{i+1}(0) = l_i(1), i\ \varepsilon\ [1,n-1]$ e $l_1(0) = l_n(1)\}$. Uma área a divide o reticulado R em um conjunto *interior*, denotado por a^0, e um conjunto *exterior*, denotado por $\hat{a}$. Dado uma área a_e ("área externa") ε R, um conjunto de áreas internas ("buracos") $\{ a_1,...,a_n\}$ tais que $a_i \subseteq a_e$ $(i=1,..n)$, um *polígono* é um objeto definido pelo conjunto de pontos $pg = \{a_e^0 \cap \hat{a}_1 \cap \hat{a}_i ... \cap \hat{a}_n\ ,\ i=1,..,n\}$. Neste contexto, um polígono contém uma lista de *linhas 2D* e uma lista de *nós 2D*, que descrevem as coordenadas da área externa e das áreas internas que compõem o polígono.

Uma vez definidas as primitivas geométricas vetoriais, podemos estabelecer a hierarquia de representações geométricas vetoriais, como mostrado na figura 4.12.

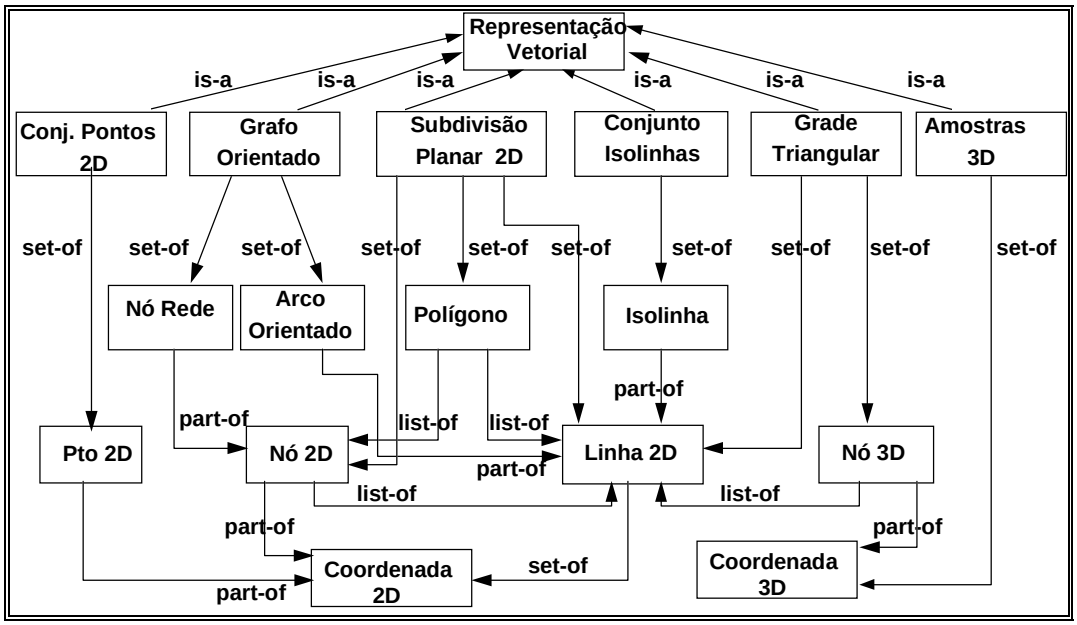

Figura 4.12 - Hierarquia de classes para REPRESENTAÇÃO VETORIAL.

Na figura 4.12, distinguimos os relacionamentos de especialização ("*is-a*"), inclusão de uma instância ("*part-of*"), inclusão de um conjunto de instâncias ("*set-of*") e inclusão de uma lista de identificadores de instâncias ("*list-of*"). Este último relacionamento será utilizado para manter a topologia 2D. Consideramos aqui, as seguintes especializações de REPRESENTAÇÃO VETORIAL:

- CONJUNTO DE PONTOS 2D: uma instância desta classe é um conjunto de *pontos 2d*, utilizado para guardar localizações isolados no espaço (p.ex. no caso de poços de petróleo).

- CONJUNTO DE ISOLINHAS: uma instância desta classe é um conjunto I de *isolinhas* e $\forall\ i_i, i_j\ \varepsilon\ I\ (i_i \neq i_j)$, $i_i \cap i_j = \varnothing$. Num mapa de isolinhas, cada linha possui uma cota e as linhas não se interseptam.

- SUBDIVISÃO PLANAR: dada uma região geográfica R, uma instância desta classe contém um conjunto Pg de *polígonos,* L de *linhas 2D e* N de *nós 2D*.

- GRAFO ORIENTADO: uma instância desta classe é uma representação, composta de um conjunto de *nó rede* e de um conjunto de *arco orientado 2d.*

- GRADE TRIANGULAR: uma instância desta classe contém um conjunto de *nós 3D* e um conjunto L de *linhas 2D* tal que:

$\forall\ l_i \in L, \exists\ l_j, l_k,\ l_m, l_n \in L\ (i \neq j \neq k\ \neq m\ \neq n)$, onde

$l_i \cap l_j = l_i(0) = l_i \cap l_k$ e $l_i \cap l_m = l_i(1) = l_i \cap l_n$ e

$l_i^0 \cap\ l_p = \varnothing, \forall\ l_p \in L,\ i \neq p.$

Trata-se de uma grade onde todas as linhas se interseptam, mas apenas em seus pontos iniciais e finais.

- MAPA PONTOS3D: uma instância desta classe é um conjunto de *coordenadas 3d*. Trata-se de um conjunto de amostras 3D.

4.6.3 RELACIONAMENTO ENTRE PLANO DE INFORMAÇÃO E REPRESENTA"CÃO GEOMÉTRICA

No universo conceitual, definimos que uma generalização para as classes de objetos que possuíam suporte espacial, através da classe PLANO DE INFORMAÇÃO. Para relacionar esta classe com suas possíveis representações, introduzimos a classe REPRESENTAÇÃO GEOMÉTRICA como generalização de REPRESENTAÇÃO MATRICIAL e de REPRESENTAÇÃO VETORIAL, e o relacionamento está_representado_em para denotar a relação entre as classes PLANO DE INFORMAÇÃO e REPRESENTAÇÃO GEOMÉTRICA. A figura 4.13 indica a hierarquia básica do universo de representação.

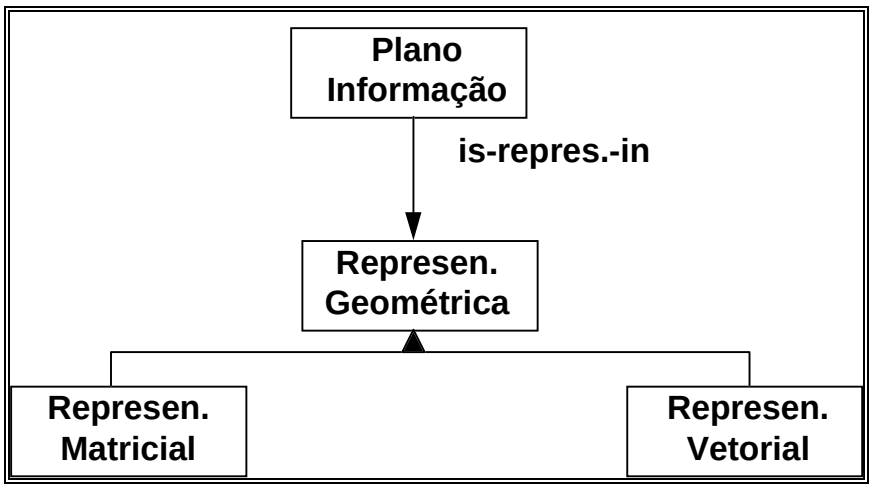

Figura 4.13 - Relacionamento entre universos conceitual e de representação.

Para maiores detalhes sobre as representações geométricas associadas a cada classe do universo conceitual, veja-se a seção 4.8 ("Relação entre os universos do modelo").

4.7 UNIVERSO DE IMPLEMENTAÇÃO

Ao discutir o universo de implementação, estaremos indicando quais as estruturas de dados a ser utilizadas para construir um sistema de Geoprocessamento. Aqui tratamos com decisões concretas de programação e que podem admitir número muito grande de variações. Estas decisões podem levar em conta as aplicações às quais o sistema é voltado, a disponibilidade de algoritmos para tratamento de dados geográficos e o desempenho do hardware. Para uma discussão sobre os problemas de implementação de operações geográficas, veja-se Güting et al. (1995).

Um dos aspectos principais a ser levados em conta no universo de implementação é o uso de estruturas de indexação espacial. Os métodos de acesso a dados espaciais compõem-se de *estruturas de dados* e *algoritmos de pesquisa e recuperação* e representam um componente determinante no desempenho total do sistema. Apanhados gerais da literatura são feitos em Cox Junior (1991) e Rezende (1992).

Estes métodos operam sobre chaves multidimensionais e dividem-se conforme a representação dos dados associados: pontos (ex: *árvores K-D*), linhas e polígonos (ex: *árvores R e R+*) e imagens (ex: *árvores quaternárias*). Estes e outros métodos têm possibilitado (principalmente no caso de pontos e linhas) grandes melhorias de desempenho no acesso a dados geográficos.

O grande fator limitante da maior parte dos métodos de acesso estudados é que foram projetados para operar em *memória principal*. Num SGBDG de grande porte, é preciso fazer acesso eficiente a dados em *memória secundária*. Isto vale tanto para dados vetoriais como dados matriciais. No caso de dados vetoriais, Mediano, Casanova e Dreux (1994) apresentam uma nova proposta de utilizar uma extensão das *árvores-B* para poder apresentar apenas os dados geográficos relevantes para uma determinada escala, sem incorrer em procedimentos desnecessários. Esta estrutura, chamada de *árvore-V,* permite ainda o acesso a dados multiresolução e é muito conveniente como suporte a métodos de folheamento de um grande banco de dados.

4.8 RELAÇÃO ENTRE OS UNIVERSOS DO MODELO

O paradigma dos "quatro universos de modelagem" (Gomes e Velho, 1995) parte do princípio que o mapeamento entre cada universo não é reversível e admite alternativas. Discutiremos a seguir estas relações.

4.8.1 DO MUNDO REAL PARA O UNIVERSO CONCEITUAL

A passagem do mundo real para o universo conceitual pode admitir algumas variações, conforme o domínio de aplicação. Em alguns casos, o mapeamento é direto: por exemplo, as imagens de satélite e grandezas topográficas e geofísicas são naturalmente mapeadas para instâncias de GEOCAMPO. No caso de mapas municipais e de divisão política, sua associação com as classes GEOOBJETO e MAPA DE GEOOBJETOS é também direta.

Os levantamentos temáticos podem se prestar a duas interpretações, conforme seu uso: quando se tratar de trabalhos de inventário (como o mapa de vegetação da Amazônia), devem ser modelados como instâncias de GEOCAMPO (ou, mais especificamente, da classe TEMÁTICO). No caso de estudos detalhados em médias e grandes escalas (como no zoneamento ecológico-econômico), onde cada região é caracterizada por qualificadores específicos, é conveniente que estes levantamentos sejam associados a instâncias de GEOOBJETOS e de MAPA DE GEO-OBJETOS.

4.8.2 DO UNIVERSO CONCEITUAL PARA A REPRESENTAÇÃO

Este mapeamento apresenta várias alternativas não-excludentes, a saber:

- Instâncias da classe DADO_SENSORE_REMOTO são usualmente armazenadas como IMAGENS EM TONS DE CINZA e IMAGENS SINTÉTICAS;
- Um geo-campo NUMÉRICO pode ser representados tanto como matrizes (GRADE REGULAR) como vetores (CONJUNTO DE ISOLINHAS, GRADE TRIANGULAR e CONJUNTO DE PONTOS 3D);

- um geo-campo TEMÁTICO pode ser representado tanto como vetores topologicamente estruturados (SUBDIVISÃO PLANAR), como por matrizes (IMAGEM TEMÁTICA).
- Instâncias de CADASTRAL são representados por instâncias de SUBDIVISÃO PLANAR ou CONJUNTOS DE PONTOS 2D.
- Uma REDE é representada por um GRAFO ORIENTADO.

A literatura tem consagrado a conclusão de que um SIG de propósito geral deve prover todas as alternativas de representação.

4.8.3 DAS REPRESENTAÇÕES À IMPLEMENTAÇÃO

Conforme dissemos anteriormente, a realização do universo de implementação é decisão concreta de programação. Faremos aqui algumas considerações de ordem prática:

- o armazenamento de pontos 3D em árvores K-D (Bentley, 1975) traz um ganho muito significativo para aplicações como a geração de grade regular a partir de um conjunto de amostras esparsas;
- o uso de árvores quaternárias (Samet, 1990) para armazenar imagens em tons de cinza não é eficaz. Para o caso de mapas temáticos, apesar de utilizado em pelo menos um sistema comercial (SPANS), os ganhos não são significativos;
- o uso de árvores-R (Gutman, 1984) ou árvores-V (Mediano et al., 1994) só se torna eficiente quando completado por algoritmos de busca e processamento que utilizam suas propriedades.

4.8.4 RESUMO

Para compreender melhor a relação entre os diferentes universos (níveis) do modelo, a tabela a seguir arrola vários exemplos de entidades do mundo real e suas correspondentes no modelo.

TABELA 4.2

CORRESPONDÊNCIA ENTRE UNIVERSOS DO MODELO

Universo do mundo real	*Universo conceitual*	*Universo de representação*	*Universo de implementação*
Mapa de vegetação	Geo-campo Temático	Imagem temática Subdivisão Planar	Matriz 2D Linhas 2D (com R-Tree)
Mapa altimétrico	Geo-campo Numérico	Grade regular Grade triangular Conjunto Pontos 3D Conjunto Isolinhas	Matriz 2D Linhas 2D e Nós 3D Pontos 3D (KD-tree) Linhas 2D
Lotes urbanos	Geo-objetos		
Mapa de lotes	Cadastral	Subdivisão Planar	Linhas 2D (com R-Tree)
Rede elétrica	Rede	Grafo Orientado	Linhas 2D (com R-Tree)

4.9 DEFINIÇÃO DO ESQUEMA DO BANCO DE DADOS GEOGRÁFICOS

O processo de definir o esquema conceitual de um banco de dados geográficos consiste de:

- estender a hierarquia de especialização definida pelo modelo, criando classes derivadas de GEOOBJETO, CADASTRAL, REDE, TEMÁTICO, MODELO NUMÉRICO DE TERRENO e DADO SENSOR REMOTO.
- definir restrições de integridade indicando, para cada nova classe definida como especialização de GEOOBJETO, quais especializações da classe CADASTRAL podem conter descrições de geo-objetos desta classe;
- definir restrições de integridade indicando, para cada especialização de PLANO DE INFORMAÇÃO (GEOCAMPO ou CADASTRAL), quais especializações de REPRESENTAÇÃO GEOMÉTRICA estão associadas.

A última operação só será realizada se o mapeamento entre o nível de representação e o de implementação não for fixo, o que seria uma forma desejável de extensibilidade.

Como exemplo, considere-se a seguinte definição de esquema conceitual para um banco de dados geográficos para cadastro rural (ilustrada na figura 4.14):

- uma classe FAZENDAS, especialização de GEOOBJETO, que pode ainda ser sub-especializada em LATIFÚNDIO e MINIFÚNDIO .
- uma classe MAPA DE PROPRIEDADES, especialização de CADASTRAL, que define um mapeamento para os objetos da classe FAZENDAS e suas especializações.
- uma classe MAPA DE SOLOS, especialização de TEMÁTICO, cujas instâncias armazenam o tipo de solos para as áreas de estudo;
- as classes ALTIMETRIA e DECLIVIDADE, especializações de MODELO NUMÉRICO DE TERRENO, cujas instâncias guardam (respectivamente) a topografia e a declividade da área de estudo;

- uma classe DADOS LANDSAT, especialização de DADO SENSOR REMOTO, cujas instâncias contêm as imagens do satélite LANDSAT sobre a região de estudo.

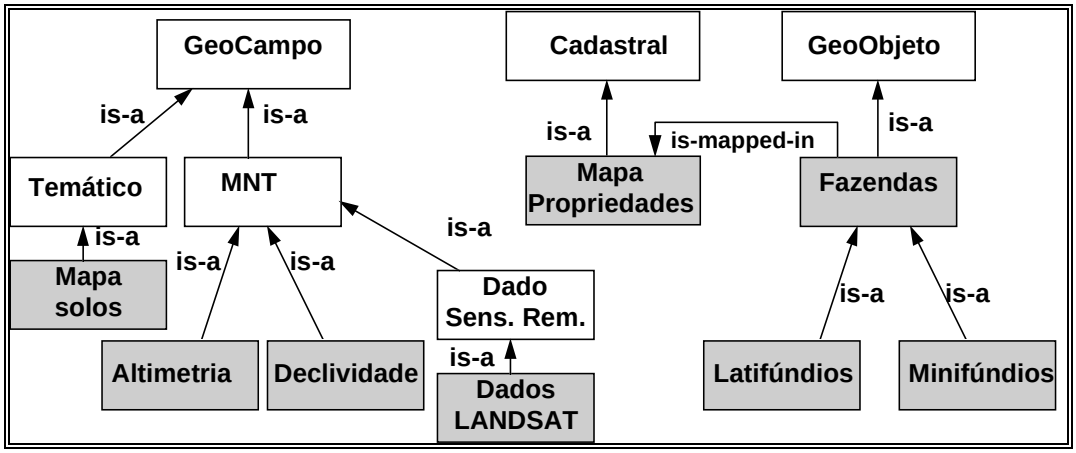

Figura 4.14 - Exemplo de definição de Esquema Conceitual.

4.10 ATRIBUTOS BÁSICOS DAS CLASSES DO UNIVERSO CONCEITUAL

Associamos a cada uma das classes do universo conceitual, atributos básicos que serão utilizados posteriormente pelos métodos definidos para cada classe. Estes atributos incluem (mudar isto.....)

- Para a classe GEOCAMPO, o identificador, nome, resolução horizontal, resolução vertical, escala e projeção cartográfica.
- Para a classe TEMÁTICO, a lista dos *temas* associados.
- Para a classe MNT, os valores máximo e mínimo da função.
- Para a classe GEOOBJETO, o identificador e o nome.

Para a classe CADASTRAL, o identificador, o nome, a escala e a projeção cartográfica.

4.11 INTERFACE COM O USUÁRIO

A nosso ver, um dos aspectos mais interessantes do modelo é permitir o projeto de interfaces amigáveis e eficientes, o que acelera em muito o aprendizado do sistema e aumenta a produtividade do usuário.

No SPRING-2.0, o uso do modelo permitiu o projeto de interface que permite a manipulação de dados geográficos num nível abstrato. Ao selecionar um *plano de informação*, apenas as operações disponíveis para a classe de dados geográficos associada ao plano estarão disponíveis ao usuário. Esta abordagem reduz muito a incerteza na escolha das funções válidas num SIG.