



UNIVERSIDADE DE SÃO PAULO  
ESCOLA DE ENGENHARIA DE SÃO CARLOS  
DEPARTAMENTO DE HIDRÁULICA E SANEAMENTO

# Introdução aos Sistemas de Informação Geográfica (SIG)

WOODROW NELSON LOPES ROMA  
MARCELO PEREIRA DE SOUZA

SÃO CARLOS  
2020

UNIVERSIDADE DE SÃO PAULO  
Escola de Engenharia de São Carlos  
Departamento de Hidráulica e Saneamento



Introdução aos Sistemas de Informação Geográfica

( SIG )

Woodrow Nelson Lopes Roma  
Marcelo Pereira de Souza

2001

# Capítulo 1

## Introdução ao SIG

### 1.1 - Um sistema de Informação Geográfica Manual

Vamos introduzir a linguagem dos sistemas de informação geográfica, com um simples exemplo inicial de uma aplicação de um SIG manual simples. Este SIG apareceu durante o início do desenvolvimento de um local para um campo de golfe. Assumimos para essa discussão que um local específico já está decidido. O projetista levantou e juntou um grupo de dados existentes para o local. Este grupo incluiu um mapa topográfico, um desenho dos limites do terreno e uma fotografia aérea do local. Referimo-nos a este conjunto de dados como **camada de dados** ou **plano de dados**.

O mapa topográfico revela informação sobre a elevação no local, representada por uma série de linhas de contorno. Estas linhas de contorno fornecem-nos uma quantidade limitada de informação sobre a forma do terreno. Certas coberturas do terreno são representadas por cores (o azul representa a água, e o verde a vegetação) e por textura ou figuras (tal como as figuras repetidas denotando terras úmidas). Dispositivos feitos pelo homem, incluindo estruturas e estradas, são indicadas por linhas e formas impressas em preto. Frequentemente as informações em mapas estão de 5 a 15 anos ultrapassadas, uma situação comum resultante de alterações na cobertura da área e do ciclo de atualização dos mapas. Cada uma dessas diferentes formas de informação, que podemos armazenar de diversas formas, é chamada de **tema**.

Uma prancha de desenho do órgão de planeamento municipal nos fornece um mapa com diferentes espécies de informação sobre a área. Este mapa é focado principalmente na infra-estrutura, e contém a descrição legal dos limites do campo de golfe pretendido, as estradas existentes e planejadas, as diferentes facilidades e a localização das

utilidades existentes e planejadas, tais como água potável, suprimento de energia elétrica e de gás e o sistema sanitário. O desenho geralmente é feito em escala maior que o mapa topográfico, e não são necessariamente baseados no mesmo mapa de projeção.

A fotografia aérea é uma fonte muito rica de informações particularmente para um analista treinado em interpretação de imagem. Um interpretador treinado pode detectar tipos de solo, topografia, drenagem e cobertura vegetal, baseado no conteúdo de uma foto. Infelizmente esta fotografia é provavelmente de escala diferente da dos dois mapas anteriores, e pode conter significantes distorções geométricas. Os dois mapas tendem a ser planimétricos, i.é., as relações espaciais entre objetos no solo são corretamente representadas nos mapas. A fotografia, por outro lado, provavelmente sofre de distorções de perspectiva inerentes a todas as fotografias e devido a um ponto de vista não vertical.

Um segundo passo no desenvolvimento dos planos para este local é tratar os três conjuntos de dados para que possam ser usados simultaneamente. A um cartógrafo ou desenhista é dada a tarefa de redesenhar a prancha do mapa de planejamento municipal e o mapa topográfico em filme plástico, de modo que os objetos nos novos mapas-filmes sobreponham-se às suas contrapartidas na fotografia aérea. Este processo, chamado de **registro**, causa aos objetos (prédios, estradas etc.) mover-se de suas posições originais, no mapa de planejamento, para novas posições nos mapas-filmes que se ajustam às posições em que são encontrados na fotografia. Alternativamente poderia ter sido tratada a fotografia para que os objetos se ajustassem ao mapa de planejamento. Em qualquer caso, este conjunto de planos de dados registrados espacialmente é agora uma base de dados geográficos úteis. Desde que os três conjuntos de informações foram convertidos para se sobreporem uns aos outros, os tratamentos futuros são muito mais fáceis. Note que, se a foto foi escolhida para ser a camada de base de dados, a base de dados resultante pode não ter nenhuma relação simples com um sistema de coordenadas geodésicas conhecido, tais como a latitude e a longitude. Entretanto para aplicações que cobrem uma pequena área isto não é um problema sério.

Com as camadas de dados ajustadas a uma mesma vista da superfície da Terra, há algumas operações analíticas que podem ser feitas com este sistema manual de informações geográficas. O analista começa desenhando algum objeto novo em uma nova folha plástica que se sobrepõe as outras camadas. Por exemplo, nós podemos gerar

corredores de 25 m de largura nos limites da propriedade, corredores de 10m de largura ao redor das estradas existentes e a locação das facilidades propostas. Estas regiões recentemente derivadas podem sugerir alguns locais que são inviáveis para o desenvolvimento de novas facilidades, e outras que são particularmente desejáveis devido à proximidade de utilidades necessárias. Com isso estamos agora em condições de tomar decisões na locação da sede do clube, dos campos, das ruas de acesso e das áreas de estacionamento.

Com base nas decisões preliminares nós preparamos uma nova camada de dados que é um rascunho do layout do pretendido campo de golfe. Combinando as tentativas de orientações dos campos com o mapa topográfico original podemos passar aos cálculos para estimar volumes de terraplanagem. Estes cálculos são considerados no domínio da engenharia Civil. Definidos os campos, com os buracos, e dificuldades de cada quadra, é possível calcular as necessidades de sementes de grama e de fertilizantes.

De forma global, este processo envolve um certo número de passos chaves. Várias espécies de dados espaciais foram localizados, e tratados para que as características importantes de cada um fossem encontradas para a mesma posição. Uma vez que estes dados são levados para um sistema geográfico comum, é possível utilizá-los em conjunto para desenvolver informações derivadas. Como se vê este processo é um fluxo típico de dados e informação através de processamento de dados espaciais para a análise de um problema.

## 1.2 - Aplicações

O número de camada de dados que se necessita considerar varia grandemente de uma aplicação para outra. Considere um problema mais complexo: decidir a localização de um aeroporto. Algumas das camadas de dados ou temas que um planejador pode requerer incluem:

Administrativo

Proprietários das terras

Jurisdição governamental

Infra-estrutura

Rede de transporte

Corredores de utilidades

Direitos	Restrições de zoneamento
Mineração	
Uso atual da terra	Biota
	Espécies em extinção
	Cobertura vegetal
Abiótica	
Geologia da superfície	
Geologia subsuperficial	Clima
Águas superficiais	Temperatura
Águas subterrâneas	Precipitação
Planície de inundação	Neblina
Sítios arqueológicos	Vento
Elevação	Fotoperíodo

Sistemas de Informação Geográfica são amplamente utilizados. Os arquitetos adotaram os conceitos de SIG para tratamento paisagístico, analisando a adequabilidade do local e desenvolvendo capacidades de planejamento para um uso especificado (McHarg, 1969). Engenheiros civis e arquitetos envolvidos com o desenvolvimento de grandes locais tem interesses semelhantes incluindo considerações de impacto ambiental. Profissionais florestais usam esta tecnologia para mapeamento e gerenciamento das áreas, e no controle de pestes e doenças. Planejadores urbanos estão usando SIG para ajudar na automatização das taxas de assentamentos, roteamentos de emergência, manutenção dos meios de transportes e das terras públicas.

Gerentes ambientais e cientistas usam este sistema para aplicações tais como manutenção de inventário de espécies raras e em extinção e seus habitats, e para monitorização de locais de disposição de resíduos perigosos. Em adição a estes tipos de uso os planejadores militares adicionaram vários estudos estratégicos.

### 1.5 - Conceitos geográficos

**Objetos espaciais** são áreas geográficas delimitadas, com um número de diferentes espécies de atributos ou características associadas. Um **ponto** é um objeto espacial sem área. Um dos atributos chaves de um ponto é sua localização geodésica, usualmente representada por um par de números (por exemplo latitude e longitude). Pode haver um campo de dados associado com o ponto dependendo da aplicação.

Uma **linha** é um objeto espacial composto de uma seqüência de pontos conectados. Linhas não tem largura, e por isso uma localização estará de um lado ou do outro da linha, nunca sobre ela. **Nós** são pontos especiais geralmente indicando o cruzamento de linhas ou a extremidade de segmentos de linha.

Um **polígono** é uma área fechada. Polígonos simples contém uma só área, e complexos são divididos em mais de uma área com características diversas. **Cadeias** ( ou **correntes**) são tipos especiais de segmentos de linha, que correspondem a uma porção do limite de um polígono.

**Escala** é a razão entre as distâncias representadas nos mapas ou fotografias e seus valores verdadeiros na superfície da Terra. Os valores das escalas são normalmente escritos como números adimensionais, indicando que as medidas no mapa e na Terra são na mesma unidade. Uma escala 1:10000 (pronuncia-se um para dez mil) indica que uma unidade no mapa vale dez mil das mesmas unidades na Terra. Uma escala refere-se sempre a distâncias lineares e não medidas de área ou elevação.

Um importante conceito para trabalhar com dados espaciais é o de **resolução**. Tobler (1987) define resolução espacial para dados geográficos como o conteúdo do domínio geométrico dividido pelo número de observações, normalizado pela dimensão espacial. O domínio, para dois conjuntos de dados dimensionais como mapas e fotografias, é a área coberta pelas observações. Portanto, para dados bidimensionais, tome a raiz quadrada da razão para normalizar o valor. Por exemplo, se a área do Brasil é aproximadamente 5,9 milhões de quilômetros quadrados, e há 25 estados, então o elemento resolução média de um mapa retratando os estados será:

$$\begin{aligned}\text{Elemento resolução média} &= \sqrt{\text{area/numero de observações}} \\ &= \sqrt{5,9 \cdot 10^6 / 25} \text{ km} \\ &= \text{aproximadamente } 486 \text{ km.}\end{aligned}$$

Isto nos dá um modo de examinar dados espaciais, e calcular o valor representativo da resolução espacial do conjunto de dados. Se aumentarmos o número de observações, o elemento resolução média decresce em tamanho. Considere um mapa do Brasil que indica cada uma das 8000 cidades.

Elemento resolução média = aproximadamente 28 km.

Como conclusão observamos que o elemento resolução (**resel**) é relacionado ao tamanho dos objetos que conseguimos distinguir em um conjunto de dados.

Outros conceitos geométricos bem como relações e operações geométricas podem ser encontrados em Nagy & Wagle (1979).

### 1.3 - Os quatro M's

Nosso entendimento deste planeta sempre foi limitado pela falta de informação, como também nossa falta de sabedoria e conhecimento. Para coisas muito pequenas desenvolvemos microscópios que permitem descer ao nível molecular. Na outra extremidade do contínuo, para coisas que são muito grandes temos satélites geoestacionários que podem captar uma imagem do hemisfério inteiro.

**Sistemas de Informação Geográfica** são um meio de integrar dados espaciais coletados em diferentes escalas e tempos, e em diferentes formatos.

Basicamente, planejadores urbanos, cientistas, gerentes de recursos, e outros que usam informação geográfica trabalham em diversas áreas principais. Eles observam e **medem** parâmetros ambientais. Eles desenvolvem **mapas** que retratam as características da Terra. Eles **monitoram** alterações ao nosso redor em função do espaço e tempo. Em adição eles **modelam** alternativas de ações e processos para gerenciar o ambiente. Estes são os quatro M's: **medir, mapear, monitorar e modelar** (figura 1.4). Estas atividades chaves podem ser melhoradas através do uso de tecnologias de sistemas de informação, e em particular através de um SIG.



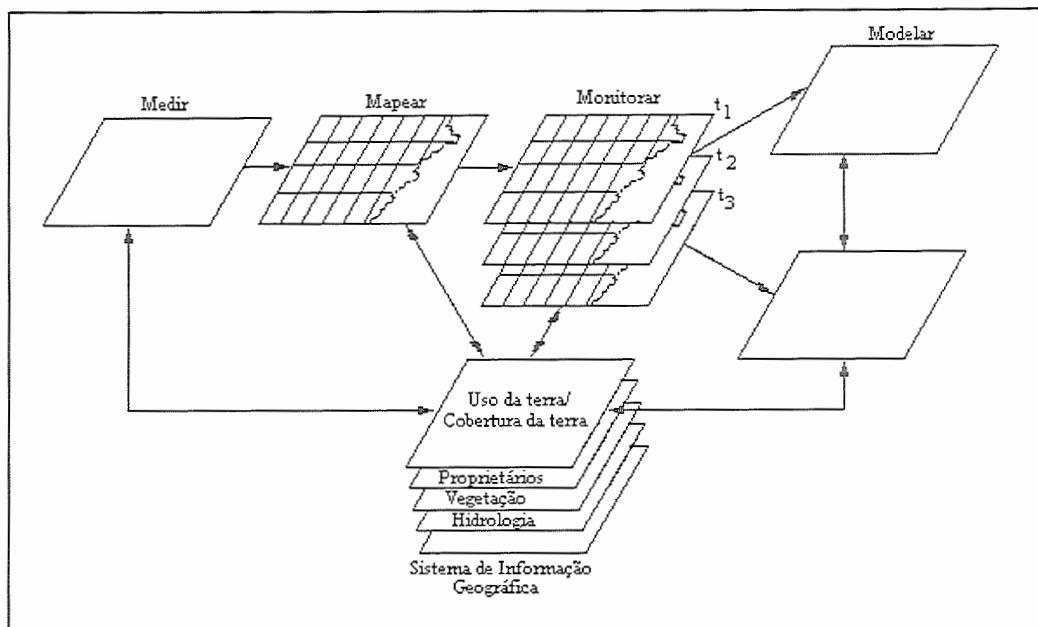


Figura 1.4 - Os 4 M's. Medir, mapear, monitorar e modelar os comportamentos e processos ambientais pode ser melhorado pelo uso de um SIG.

Os **Sistemas de Informação Geográfica** tem o potencial para melhorar o entendimento do mundo à nossa volta. Claro que estes sistemas não controlam a qualidade dos dados, nem fazem o trabalho por nós. O trabalho que podemos fazer com um SIG é claramente dependente da qualidade dos dados que ele contem. Portanto muito cuidado deve ser tomado para entender as fontes potenciais de erros e a grandeza dos erros que podem ocorrer quando se juntam e processam dados espaciais. Deve-se também estar consciente do potencial de erro na interpretação da informação gerada por um SIG.

Na interação com um SIG o usuário deve não somente entender a aplicação mas também a ferramenta e o próprio sistema.

## Capítulo 2

### Elementos essenciais de um SIG

Um sistema de informação é fundamentalmente um sistema que trata com o fluxo de dados e informações desde suas fontes primárias, até às suas informações derivadas e aplicações finais.

**Sistemas de Informações Geográficas** são projetados para tratar informações de locais geográficos de forma espacial. Vamos apresentar os componentes funcionais de um SIG, e discutir alguns conceitos chave em geografia e processamento de dados geográficos.

#### 3.1 Elementos funcionais de um SIG

Existem cinco elementos funcionais que um SIG deve conter: aquisição de dados, preprocessamento, gerenciamento de dados, manipulação e análise, e geração de produtos. Em qualquer aplicação de um SIG, é importante entender estes elementos como interligados em um processo contínuo. Como princípio geral, para resolver um problema, o analista deve, à priori, desenvolver sua modelagem completa. Mesmo quando algum detalhe depende de resultados intermediários do cálculo, um esquema explícito do processo completo é sempre muito útil.

**Aquisição de dados** é o elemento que contém o processo de identificar e coletar os dados requeridos para sua aplicação. Este processo é geralmente incluído diversos procedimentos, por exemplo:

- a coleta de novos dados através da preparação de mapas de larga escala da vegetação natural a partir de observações de campo, ou por delimitação em fotografia aérea.

- pesquisas para determinar comportamentos e interesses, por exemplo, satisfação e preferências dos consumidores em diferentes partes da cidade para ajudar a implantar novos pontos comerciais.

- localizar e coletar dados existentes tais como mapas, fotografias aéreas e terrestres, levantamentos de várias espécies, e documentos de arquivos e repositórios.

Nunca se deve economizar (tempo ou dinheiro) na fase de aquisição de dados. Um SIG não tem utilidade enquanto os dados relevantes para o problema não forem identificados e armazenados. Além disso a acuidade das decisões formadas através da análise espacial é limitada pela acuidade e precisão dos dados básicos.

É freqüente o desconhecimento da qualidade inerente dos dados espaciais, e ainda podemos ser forçados a utilizar mapas e conjuntos de dados cuja qualidade inerente é desconhecida. Sem dispender esforço para assegurar que os vários conjuntos de dados são, além de necessários, também confiáveis corremos o sério risco de enganarmos a nós próprios.

**Preprocessamento:** É o elemento que envolve o tratamento dos dados para que eles possam ser incorporados ao SIG. Duas das principais tarefas de preprocessamento incluem conversão do formato dos dados e identificação da locação dos objetos nos dados originais de forma sistemática. A conversão do formato dos dados originais envolve extrair informação de mapas, fotografias e material impresso (tais como tabelas de censo demográfico) e gravar essas informações numa base de dados no computador. Este processo é muito lento, consumindo muito tempo e muito esforço, traduzindo-se sempre em alto custo para o projeto.

A segunda tarefa chave do preprocessamento é o estabelecimento de um sistema consistente para a gravação e especificação da locação dos objetos no banco de dados. Com esta tarefa completada é possível determinar as características de qualquer localização especificada a partir do conteúdo do banco de dados do sistema. Durante este processo é

muito importante manter um critério de controle de qualidade para acompanhar as operações para que as bases de dados possam ser de máximo valor para o usuário.

**Gerenciamento dos dados:** É o elemento que contém as funções de gerenciamento de dados governam a criação da, e o acesso à, base de dados. Estas funções fornecem métodos consistentes para entrada, atualização, descarte e recuperação dos dados. Os modernos sistemas de gerenciamento de dados isolam o usuário dos detalhes de organização interna do armazenamento de dados, fornecendo uma interface amigável, de forma que o usuário nem chega a notar as diversas operações realizadas. Um sistema pobre exige muito do usuário e é portanto mais sujeito a erros, mais lento e mais difícil de usar. Cuidados no gerenciamento de dados incluem esquemas de segurança, que se traduzem em dispositivos para diferenciar usuários com diferentes níveis de acesso ao sistema e à base de dados. Por exemplo, a atualização da base de dados só pode ser permitida após uma autoridade ter verificado que a alteração é apropriada e correta.

**Tratamento e análise:** Este elemento contém as operações analíticas que trabalham com o conteúdo da base de dados, para derivar novas informações. Muitos usuários acreditam, de forma incorreta, que este módulo é tudo o que constitui um SIG, mas é apenas um elemento dele. Por exemplo nós podemos especificar uma região de interesse e solicitar que a inclinação média da área seja calculada, com base nas curvas de elevação que já estavam armazenadas na base de dados do SIG.

Como nenhum sistema é capaz de incluir todo o campo de possibilidades analíticas que os usuários possam imaginar, nós devemos ter facilidades específicas que permitam mover dados do SIG para um sistema externo onde eles possam ser processados, e levar os resultados de volta para o SIG. Essa espécie de modularidade, que permite outros sistemas de análise e processamento serem conectados ao SIG, é de extremo valor em muitas circunstâncias e permite ao sistema ser facilmente extensível através do acréscimo de outras ferramentas analíticas. Quando se fala em geoprocessamento está-se referindo às componentes de análise e tratamento de um SIG.

**Geração de produto:** É a fase do SIG na qual a saída final é gerada. Estes produtos podem incluir relatórios estatísticos, mapas, e gráficos de várias espécies. Como exemplo de cada produto podemos citar:

- tabelas contendo a densidade populacional dos estados brasileiros, ou um rol dos fazendeiros que possuem terras não produtivas, etc.
- mapas indicando tubulações enterradas para delimitar área de escavações, ou delimitações de terras indígenas, etc.
- gráfico de barras indicando a produção de várias colheitas numa região, ou o crescimento populacional para previsão de produção água, etc.

Esses produtos podem ser transientes, com apresentação momentânea na tela do computador, neste caso são denominados **soft copy**. Se o produto é produzido em papel ou filme, mais durável, então ele é chamado **hard copy**. Uma forma, que pode ser considerada **hard copy**, que vem aumentando rapidamente é o produto gerado em formas compatíveis com o computador, tais como disquetes, fitas magnéticas, ou algum sistema de transmissão. A capacidade de um produto ser colocado de volta num SIG, para análise futura, é extremamente importante.

Estes componentes essenciais de um SIG são os mesmos de qualquer outro sistema de informações. Considere os passos tomados em um sistema automatizado para gerenciar as anotações em registros de empregados de uma corporação: informações sobre os indivíduos devem ser juntadas, por questionário ou entrevista quando da contratação (aquisição de dados), as informações devem ser uniformizadas para um vocabulário consistente (preprocessamento), digitação em computador a partir dos dados consistidos, validação dos dados inseridos, uma fase essencial no preprocessamento para assegurar a fidelidade da base de dados resultante. Uma vez que os dados tenham sido convertidos em forma consistente e introduzidos na base de dados do computador, nós cumprimos com uma grande fração dos custos finais. As funções de gerenciamento permitem atualizar a base de dados sempre que ocorrer modificação digna de nota. E finalmente uma grande variedade de operações analíticas podem ser realizadas sobre as informações do banco de

dados para obter alguma informação importante e gerar um relatório como produto. Vemos que estas fases são bastante semelhantes às cinco componentes de um SIG.

### 3.2 - Dados em um SIG

É importante entender as diferentes formas de variáveis que podem ser armazenadas em um SIG. **Variáveis nominais** são aquelas que são descritas por um nome sem ordem específica. Como exemplo de variável nominal podemos citar as categorias de uso da terra (parque, distrito industrial, área residencial etc.). **Variáveis ordinais** são listas de classes discretas, mas com uma ordem inerente. Constitui exemplo desta classe os níveis de instrução (primário, secundário, superior etc.). **Variáveis de intervalo** tem a seqüência natural, mas em adição, as distâncias entre os valores tem significado. Temperatura medida em graus Celsius é uma variável de intervalo, pois a distância entre 10°C e 20°C é a mesma que entre 20°C e 30°C. O último tipo de variável é a **variável de razão**, que tem as mesmas características da variável de intervalo, e ainda tem um ponto de partida ou zero natural. A escala absoluta de temperatura, Kelvin, é baseada em um padrão absoluto e por isso é uma variável de razão. Renda per capita, índice fluviométrico mensal, e fração de massa de solo retida em uma peneira são exemplos comuns de variáveis de razão.

Em adição a estes quatro tipos de dados de informação existem duas diferentes classes de dados encontrados na maioria dos SIG. Vamos considerar um objeto no espaço para armazenar em um SIG: um poço d'água. Para o SIG a peça essencial de informação é sua localização na Terra, que pode ser definida por um par de coordenadas, sua latitude e longitude. Entretanto existe uma infinidade de informações necessárias para completar os dados de informação do poço, por exemplo sua profundidade, produção de água, dados de testes de bombeamento, seqüência temporal da qualidade da água, etc. Este segundo conjunto de **dados não espaciais** ou **atributos**, que são logicamente ligados aos dados espaciais, não pode ser esquecido. Em muitos sistemas de informações geográficas há ferramentas para armazenar e trabalhar os dados de atributos ligados aos dados espaciais. Em algumas aplicações o volume dos atributos pode ser realmente muito maior que o volume de dados espaciais, e as conexões lógicas entre eles muito importante. Embora

ainda não existam normas técnicas para normalização dos dados em um sistema de informação geográficas, foi proposta uma norma com base em objetos reais e um mecanismo para representar esses objetos na base de dados. Essa proposta define um conjunto de objetos espaciais e os divide em três classes:

- **Objeto Sem Dimensão** - constituído por um **ponto**, que é especificado por uma locação geométrica. Os pontos são usados para indicar **posições**, tais como o centro geométrico de um poço, o fim de uma rua, a esquina de um lote. Pontos são também usados como uma posição reservada para rótulos ou símbolos em um mapa. Pontos são também usados para compor objetos espaciais mais complexos, tais como linhas e áreas.
- **Objeto unidimensional** - O mais simples objeto de uma dimensão é a linha reta ligando dois pontos. Formas mais complexas de linhas incluem conjuntos de retas conectados (determinados pela sequência de pontos nos quais o caminho muda de direção), curvas que são baseadas em funções matemáticas, e linhas cuja **direção** é especificada. Uma vantagem do segmento de linha é que ele pode ser orientado, fornecendo um modo de distinguir qual extremidade do segmento de reta é o início e qual é o fim. Este modo tem particular valor em problemas onde a direção de fluxo é importante, por exemplo o estudo do fluxo de água em uma tubulação, e o estudo de modelos para a imigração entre países. Quando um segmento de reta contém informação sobre a direção é também possível distinguir entre os lados direito e esquerdo da linha.
- **Objeto bidimensional** - Os objetos bidimensionais são constituídos por áreas de forma genérica. A descrição da área é normalmente baseada na geometria da linha que a circunda. A área pode ser homogênea ou dividida internamente. Numa aplicação particular podemos nos referir a uma área cercada, ou focalizar a atenção precisamente na linha limite da área, ou ainda só na área interna ao limite. Uma distinção deve ser feita entre conjuntos de regiões bidimensionais limitadas e superfícies verdadeiramente tridimensionais. Em algumas aplicações uma análise baseada em representação planimétrica bidimensional da Terra pode ser plenamente suficiente.

O estudo das conexões entre objetos espaciais é chamado **topologia**. Um dos comportamentos que distinguem um determinado SIG de outros é a presença de mecanismos explícitos para armazenar topologia.

## Capítulo 3

### Estruturas de Dados

Existe um grande número de modos diferentes para organizar os dados em qualquer sistema de informações. A escolha de uma particular **estrutura de dados espaciais** é uma importante decisão no projeto de um SIG. O conhecimento da estrutura dos dados tem grande valor, mesmo que não seja para a finalidade de projeto de um sistema, os usuários devem conhecer as diferentes estruturas de dados, já que diferentes formas padronizadas são comumente usadas, e a escolha de um determinado tipo pode afetar o volume de dados armazenados e/ou a eficiência de processamento. Por outro ponto de vista, deve-se escolher uma estrutura de dados para a coleta de informação para que a eficiência de armazenagem não seja comprometida. Também durante as pesquisas operacionais é freqüente a necessidade de conversão dos dados de uma estrutura para outra, por exemplo, para importar dados para um sistema de tratamento existente. É extremamente importante entender como essas conversões afetam as informações básicas.

Cada diferente tipo de dado espacial ou **tema** em um SIG, refere-se a uma camada ou plano de dados. Cada plano de dados contem informações caracterizadas por um tipo de dado, ou peça de entidade geométricas. Para codificar informações os SIG utilizam três tipos de dados: **pontos**, **linhas**, e **polígonos** ou **planos**. Alguns autores fazem distinção entre a representação de uma superfície tridimensional verdadeira, tal como conjuntos de dados de elevação, e a representação de espaço em duas dimensões, tais como limites legais de um terreno em um mapa plano.



A função essencial de um conjunto de dados espaciais, armazenados e tratados, é subdividir a Terra em entidades significativas ou objetos que podem ser caracterizados. Dessa forma o conteúdo de uma base de dados espaciais é um modelo da Terra. **Pontos**, como a locação de poços de água ou petróleo, **linhas**, tais como a linha de eixo de uma rodovia ou de um rio, são elementos-chaves desta divisão em partes componentes. Quando consideramos regiões limitadas, tais como os limites de subdivisões ou as margens de um lago, temos a atenção focalizada nas linhas de limite, e a região circundada é denominada **polígono**. Estas regiões poligonais não são necessariamente definidas em termos precisos de geometria, na qual um polígono é definido como uma figura planar circundada por uma série de segmentos de reta; no processamento de dados espaciais, é usado o termo polígono mesmo quando os limites são curvos. Muitas aplicações foram desenvolvidas em torno de uma rede de linhas, tais como a rede definida pelas artérias de um sistema de transporte, ou de comunicações; ou ainda sistemas de tubulações como nos esgotos sanitários ou sistema de gás encanado. Igualmente importante é o dado **não espacial** ou **atributo**, que em alguns sistemas pode requerer uma grande quantidade de espaço de armazenamento.

### 3.1 - Estrutura de dados tipo raster

Raster é a estrutura de dados mais simples, nela os dados são organizados de forma celular. Em uma estrutura raster o parâmetro de interesse é definido para cada célula em um arranjo (frequentemente regular) no espaço. Por exemplo, na figura 4.1, a elevação em metros foi gravada em locais definidos por uma grade regular. Os dados originais vieram de um mapa topográfico, do qual foram obtidas as linhas de contorno. Esta espécie de estrutura de dados é intuitiva; pode-se imaginar um topógrafo determinando as elevações em distâncias regulares ao longo de uma latitude constante.

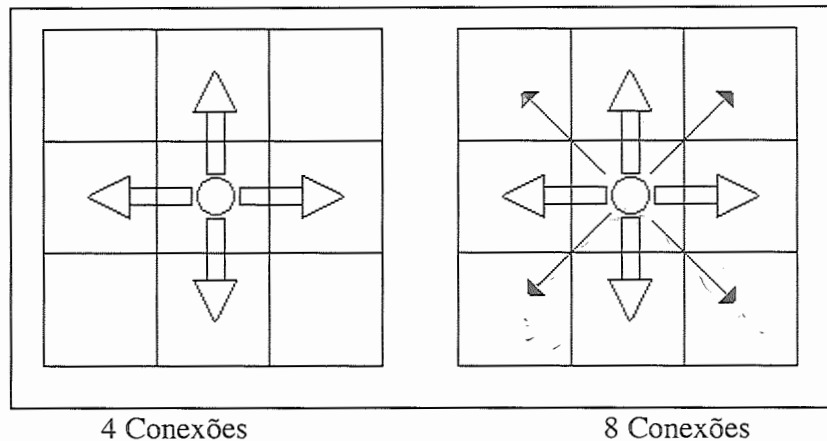


Figura 3.1 Vizinhança espacial. Células que são separadas por um lado são mais próximas que as que são separadas pelo vértice.

### 3.1.1 Arranjos raster simples.

A área a ser descrita é dividida em células regulares, em um arranjo matricial descrito por linhas e colunas. Por convenção as linhas são orientadas paralelamente ao sentido leste-oeste. A informação obtida para cada célula é denominada de amostragem, e as células são identificadas por um sistema de numeração: as células seguindo uma linha são numeradas da esquerda para a direita, e seguindo uma coluna de cima para baixo. Assim a origem de um raster é o ponto esquerdo superior da imagem. Alguns sistemas definem o ponto origem pelas coordenadas (0, 0), outros pela coordenada (1, 1), devendo ser tomado o devido cuidado com este ponto. Usualmente as células são quadradas, com o tamanho vertical idêntico ao horizontal, e a informação (amostragem) é armazenada no computador em uma matriz bidimensional, (bidimensional array). Embora a estrutura raster retangular seja muito popular ela tem no mínimo duas limitações: primeira, a estrutura raster, para uma escala definida, limita nossa capacidade de definir um objeto porque ficamos limitados ao tamanho da célula, ou estamos em uma célula ou na outra. Segunda, células adjacentes não são igualmente espaçadas. Considerando o espaçamento de 1 unidade, células abaixo ou ao lado estão 1 unidade de distância, enquanto células nas diagonais estão aproximadamente 1,41 unidades de distância uma da outra. Nas pesquisas através dos dados, se forem incluídas apenas as células laterais, acima e abaixo da célula de interesse nós estaremos trabalhando com quatro conexões de vizinhança, e as células são equidistantes, fig. 3.1. As células dividem um lado com as vizinhas.

Incluindo-se os elementos na diagonal estaremos trabalhando com 8 conexões de vizinhança, fig. 3.1, as células não são equidistantes e enquanto algumas células dividem um lado outras apenas o vértice. Como todas as células nessa vizinhança tem formas e tamanhos iguais, nós dizemos que temos uma **similaridade espacial de vizinhança** (Tobler,1979). Na prática muitos casos são constituídos de células diferentes, como por exemplo os tamanhos e formas de terrenos urbanos, mas em SIG o comum é ter células regulares. O uso de quatro ou oito conexões também depende das características do problema, áreas metropolitanas mais recentes tem uma malha retangular de arruamento, e neste caso o estudo da rede de transporte só ocorre em duas direções e o modelo de quatro conexões é suficiente.

O tamanho da célula deve ser tal que o compromisso entre qualidade e quantidade de dados seja bem balanceado. Se a célula for muito pequena o detalhamento será grande porém o número de células crescerá muito exigindo muito espaço de armazenagem. Por outro lado se for muito grande poucas células serão necessárias para cobrir a área, mas os detalhes serão muito pobres. O tamanho de cada célula é frequentemente confundido com a **mínima unidade de mapeamento** ou elemento resolução (**resel**). Enquanto o resel é a menor informação que pode ser representada no mapa, o tamanho da célula deve ser tal que permite identificar a menor informação que se deseja mapear. Uma regra geral, com base em amostragem estatística, é que o tamanho da célula deve ter metade do comprimento, ou um quarto da área, do menor elemento a ser definido.

Para exemplificar esse comportamento vamos imaginar um terreno composto por parte de terra e parte de grama e que contém um canteiro de flores, como mostrado no mapa da figura 3.2 (a), que será transformado para um arquivo raster para estudar a vegetação do terreno. Observa-se que a transformação pode ser feita com qualquer tamanho da célula, na situação mostrada em (b) o tamanho da célula é suficiente para que todos os objetos apareçam no arquivo raster, já no caso (c) um tamanho de raster mal dimensionado fez com que o canteiro de flores (1), desaparecesse por completo no arquivo.

### 3.1.2 - Mapas e Arquivos de valores

Como já mencionado um sistema de informação geográfica manual consiste na coleção de diversos mapas da mesma região, ajustados para a mesma escala, contendo todas as informações a

ela pertinentes. Para agilizar a consulta e a geração de relatórios sobre algum problema sobre a região, é conveniente introduzir os mapas em um banco de dados em computador. Um arquivo de valores é um arquivo que contém dados em forma tabular sobre as características dos mapas. O arquivo de valores armazena o valor de um atributo para cada célula do mapa.

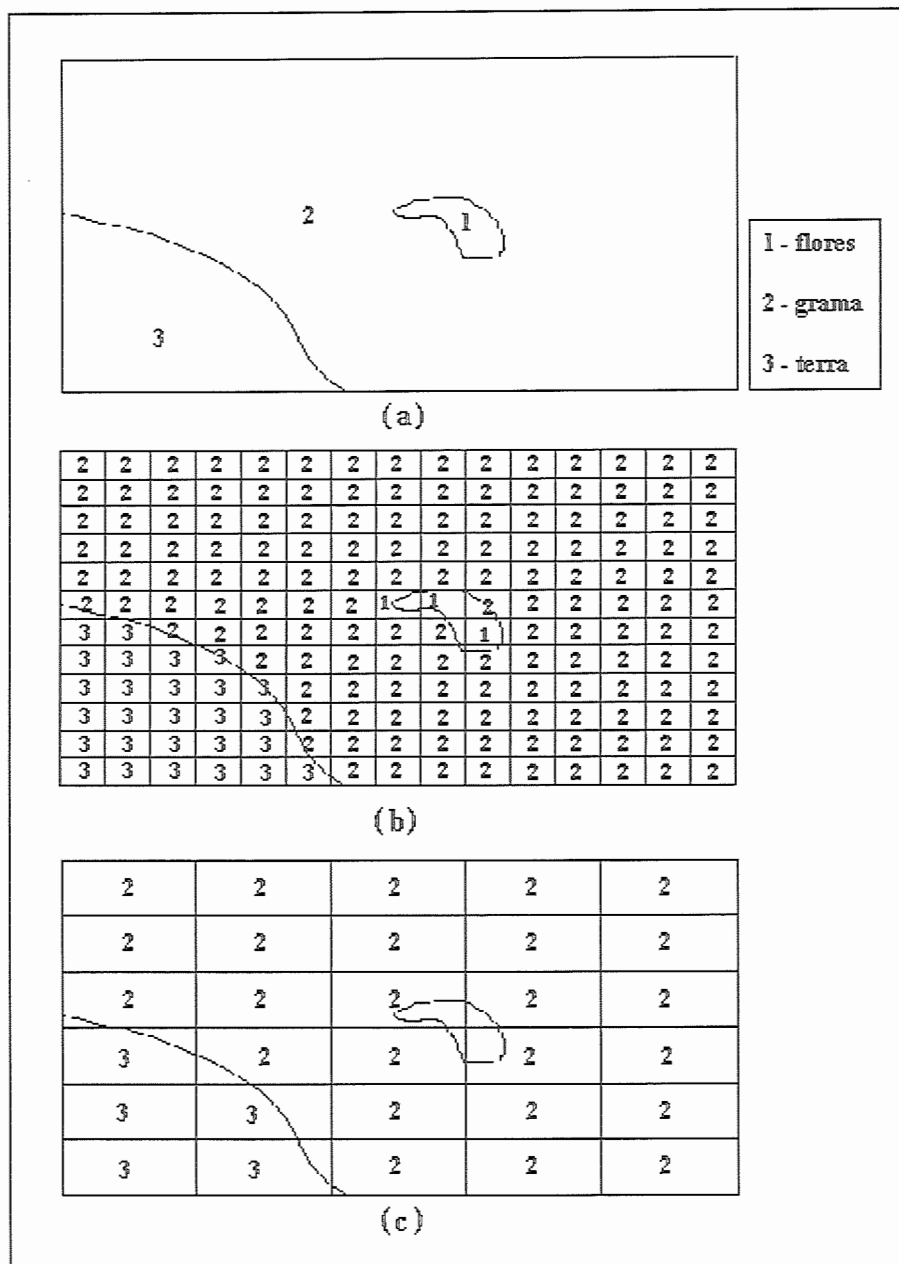


Figura 3.2 - Tamanho da célula raster. (a) Mapa da vegetação no terreno. (b) Conversão para raster. (c) Idem com tamanho de célula mal escolhido.

Como exemplo vamos apresentar um arquivo raster das altitudes de uma região, a partir do mapa de curvas de nível mostrado na figura 3.2. Para construir o arquivo raster do mapa é preciso dividi-lo em quadrados, de tamanho suficientemente pequeno para que não se perca informações do mapa. Para cada célula definida vamos tomar a altitude da célula, um número que indica a diferença de cota do ponto com relação ao nível do mar. Aqui podemos distinguir duas formas de definir o valor da altitude da célula, a altitude do ponto central ou a altitude média da célula. Na realidade nenhum dos dois valores é correto, mas é aceito que o valor médio da célula corresponde a um modelo mais próximo da verdade. Praticamente essa distinção não é importante se escolhermos o tamanho adequado das células.

Na figura 3.2 é apresentado um esquema do mapa de curvas de níveis, e uma tabela representando o valor da altitude em cada célula do arquivo raster arquivado no computador. A representação no computador é feita seguindo uma matriz de  $m$  linhas por  $n$  colunas, e codificado por uma numeração que permite identificar cada célula por um par de números  $(m,n)$  definidos pela posição da célula no arquivo. Na tabela da figura, que representa a área retangular do mapa, podemos observar que a altitude do área, marcada na primeira célula, tem o valor 83 e a célula é identificada pela posição horizontal 1 e pela vertical 25, números que representam a localização da célula.

Para que essas informações sejam armazenadas em um computador, e representem de forma confiável a situação real, é necessário que sejam estabelecidas algumas regras que permitam a recuperação dos dados e a posterior obtenção do mapa original, sem perda de informações. No caso da figura, o retângulo marcado faz parte de uma faixa vertical que pode ser representada por uma matriz de 14 células na horizontal por 50 células na vertical. Assim a altitude de cada ponto da área da faixa é representada por 700 ( 14 x 50 ) números, que serão gravados ordenadamente no computador, e posteriormente podem serem lidos na mesma ordem, e reproduzirem assim o mapa original.

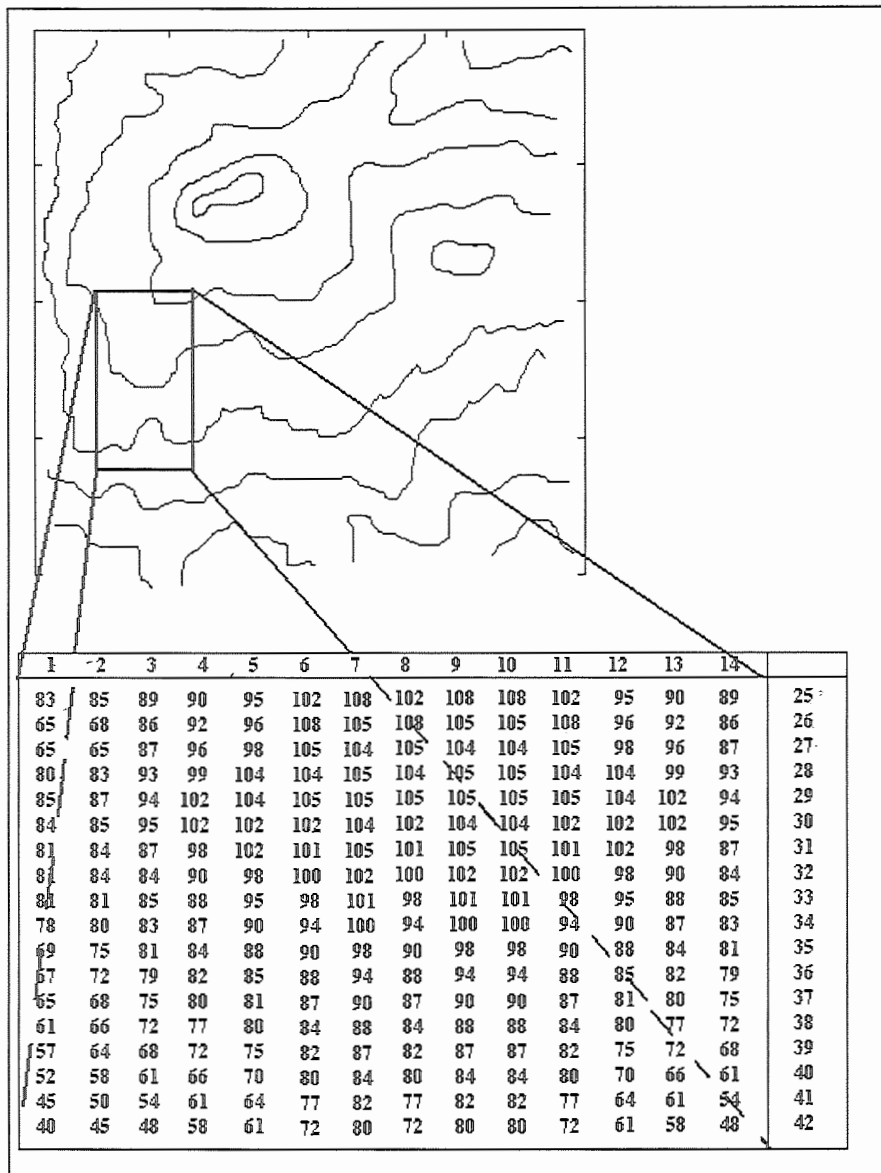


Figura 3.3 - Mapa e sua representação em arquivo raster.

### 3.1.3 - Organização dos arquivos Raster

No exemplo anterior percebe-se que se os pontos forem gravados por linhas, de quinze números cada linha, a recomposição da área é evidente, pois sua organização já apresenta os valores na posição correta, como mostrado na tabela da figura 3.3.

Programas de análise de imagens digitais são construídos com base em determinada forma de organizar estas informações. O Idrisi, por exemplo, coloca esses números gravados em seqüência, como uma única coluna de números. Uma parte da tabela, da figura 3.3, pode ser usada como

exemplo da organização dos arquivos. A parte tomada para exemplo, da tabela anterior é constituída de 3 linhas de 5 números cada, como mostrado abaixo.

83 85 89 90 95  
 65 68 86 92 96  
 65 65 87 96 98

O arquivo do Idrisi apresenta estes mesmos números como uma longa coluna de números, como apresentado na figura 3.4, à esquerda, e para recompôr a forma da tabela usa um outro arquivo com as informações de formatação.

Arquivo de dados	Arquivo de documentação	
83	File title	Curva de nível
85	Data type	integer
89	File type	ASCII
90	Columns	5
95	Rows	3
65	ref. system	plane
68	ref. units	m
86	unit dist.	1
92	min. X	0
96	max. X	45000
65	min. Y	0
65	max. Y	15000
87	resolution	30
96	minimum value	0
98	maximum value	250
	error	.25
	comment	dados digitalizados de mapa de
	comment	curvas de níveis.

Figura 3.4 - Par de arquivos utilizados por SIG para definir um arranjo tipo raster.

Usualmente essa tarefa é feita através de um arquivo auxiliar, ou de **documentação**, que informa, além do número de colunas e de linhas, qual é o tamanho de cada célula, qual a unidade de medida utilizada, a estrutura do arquivo, e outras informações necessárias para recomposição do mapa original. Um exemplo de arquivo de documentação, para o arquivo de dados do, é mostrado na figura 3.4, à direita.

### 3.1.4 - Tipos de arquivos

Ao armazenar dados, que representam os atributos de uma imagem, em um arquivo os computadores podem usar tres tipos de números: inteiros, bytes ou reais.

1- Inteiros (**integer**) são números que não tem parte fracionária e, nos computadores de 16 bits, podem variar de -32768 até +32768. Os inteiros ocupam 2 bytes (16 bits) de memória. Os inteiros podem ser usados para representar o valor real da variável, como no exemplo das curvas de nível, ou como código para algumas categorias de variáveis, como no caso do tipo de vegetação. Usualmente o arquivo de documentação inclui uma legenda para identificação do código, armazenando a relação entre o código inteiro e o tipo de atributo, no exemplo tipo de vegetação. Os programas de computador também podem usar o inteiro longo (**longint**), que varia entre -2147483647 e +2147483647, não usados nos SIG por ocupar 4 bites de memória, portanto o dobro do espaço necessário para o inteiro.

2- Valores em bytes são números inteiros positivos entre 0 e 255. Tem utilidade em casos onde o campo de valores é mais limitado. Esse campo de valores pode ser definido com apenas 8 bits de memória, 1 byte, e por isso ocupa metade do espaço do tipo inteiro. Este tipo é o mais utilizado por arquivos de SIG, pois fornece um campo adequado de valores para descrever a maioria dos mapas qualitativos e virtualmente todos os dados de sensoriamento remoto.

3- Reais (**real, single**) são números com parte fracionária, e são usados quando uma variável tem comportamento contínuo e deve ser armazenada com grande precisão. Tem aplicação também quando o campo excede o dos inteiros. Os dados reais podem ser utilizados dentro do campo  $\pm 1 \times 10^{38}$ , com precisão de 7 algarismos significativos. Como consequência eles ocupam mais posições de memória exigindo 4 bytes para cada número, sendo preferível com relação ao inteiro longo. Existe também o real com precisão estendida (**double**) que chegam a 14 algarismos significativos, mas não são necessários em SIG.



Os arquivos podem ser armazenados como ASCII, binário ou binário compactado. A forma deve ser informada no arquivo de documentação.

O arquivo ASCII é também conhecido como arquivo texto, e pode ser visto diretamente por qualquer editor de texto, como também pelo comando TYPE do DOS. É um formato adequado para transferir dados de um programa para outro, pois o ASCII (American Society Code for Information Interchange) é um formato padronizado e de ampla utilização. Apesar de muito fácil de ser lido, o padrão ASCII não é muito eficiente para armazenamento de dados produzindo arquivos muito grandes, e de acesso mais lento devido à necessidade de conversão do formato ASCII para o binário, formato utilizado pelos processadores.

O arquivo binário armazena os dados na mesma forma em que são tratados pelo processador, sendo portanto de maior eficiência para escrita e leitura, apresentando-se ainda com menor espaço de armazenamento. A maior dificuldade é que os arquivos binários não podem ser examinados diretamente, necessitando de um programa de conversão, normalmente fornecido com a maioria dos programas de tratamento de dados.

O arquivo binário compactado é o arquivo binário, inteiro ou byte, tratado por alguma forma de compactação que reduz o tamanho do arquivo. De maneira geral a compactação só reduz o tamanho se houver alta frequência de células, com igual valor, vizinhas uma das outras. Nesse caso no lugar de armazenar, por exemplo, dez vezes o valor de dez células, o arquivo recebe o número 10, um sinal que indica repetição e uma vez só o valor da célula. Em situações favoráveis o tamanho do arquivo pode ser reduzido de 80%.

Uso de arquivos compactados é extremamente útil quando se trabalha com muitos dados. Por exemplo fotografias de satélites de sensoriamento remoto, frequentemente usados para distinguir coberturas vegetais em grandes áreas, cobrem área de aproximadamente 30 000 km<sup>2</sup>, com tamanho nominal de célula, ou elemento de imagem, de 30 m.

Uma tomada de dados corresponde à 35 milhões de células raster, também denominada **pixel**, uma contração de “**picture element**”. Com uma base de dados tão grande os algoritmos de compactação são de alta importância. Alguns algoritmos são perfeitamente reversíveis, o que significa que reproduzem exatamente as condições originais. Outros minimizam a quantidade de dados perdendo de forma controlada pequena taxa da informação original.

### 3.2 - Estruturas de dados do tipo Vetor

Para um matemático um vetor é uma entidade com um ponto de início (ou de aplicação), um deslocamento associado e uma direção (ou rotação). Na descrição de dados espaciais baseada em vetores fazemos a hipótese de que um elemento pode estar localizado em qualquer posição, sem os vínculos de posição de um arranjo raster.

Estruturas vetoriais de dados são baseadas em pontos elementares cuja localização pode ser fixada com precisão arbitrária, contrastando com a estrutura raster. Como um exemplo simples para armazenar um círculo, em uma estrutura raster, nós devemos encontrar e codificar todas as células raster cuja localização corresponde ao limite do círculo, como mostrado na figura 3.5 (a). Temos nesse caso uma descrição de baixo-nível do círculo.

Numa estrutura vetorial o círculo seria armazenado gravando a localização do ponto que é o centro do círculo, e o seu raio. Essa descrição de alto nível é muito mais eficiente em termos de espaço, para descrever o círculo, do que a estrutura raster anterior, e ainda é muito mais precisa, figura 3.5 (b).

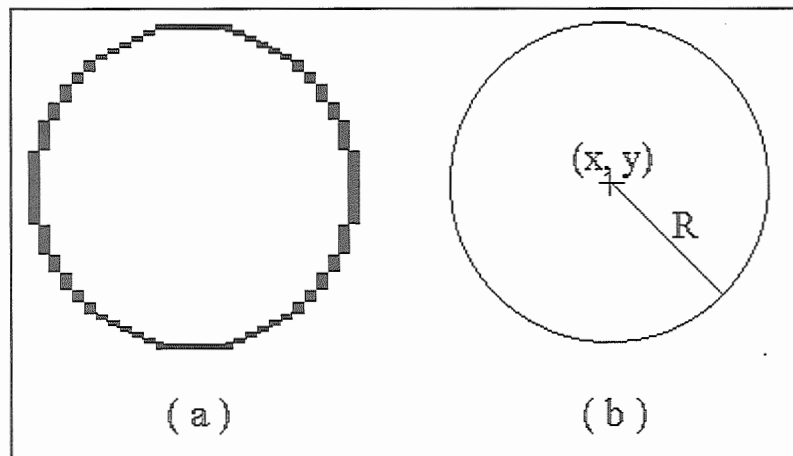


Figura 3.5 - Representação de um círculo. (a) raster, e (b) vetorial.

Muitos sistemas gráficos e CAD's usam modelos do tipo vetorial como sua organização interna de dados, usando formas primitivas tais como pontos, linhas e círculos. A vantagem da

eficiência e precisão pode desaparecer se o círculo precisar ser armazenado como uma sequência conectada de segmentos de reta.

Várias formas de estruturas de vetor estão em uso hoje, tanto para representar tipos de bases de dados em SIG, como para padrão para troca de informação entre sistemas:

- Estrutura de polígono inteiro,
- Dual Independent Map Encoding (DIME) file structure,
- Estrutura arco-nó.
- Estrutura relacional, e
- Gráficos de linhas digital.

### 3.2.1 - Estrutura de polígono inteiro,

Nesta estrutura cada camada, da base de dados, é dividida em um conjunto de polígonos, fig. 3.6. Cada polígono é codificado como uma sequência de coordenadas que define os limites de cada área fechada em um dado sistema de coordenadas.

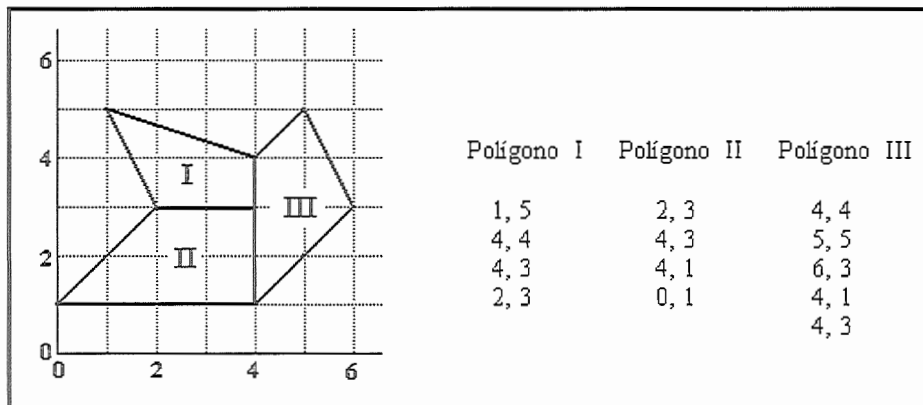


Figura 3.6 - Estrutura de polígono inteiro. Os nós de cada polígono são armazenados separadamente.

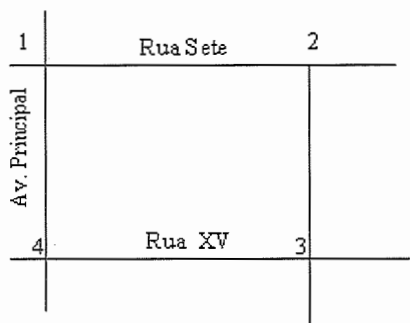
Cada polígono é então armazenado como um propriedade independente. Não existe meio explícito, neste sistema, de referenciar áreas que são adjacentes. Cada área discreta é armazenada separadamente, e os atributos que definem as propriedades do polígono podem ser armazenadas junto com a lista de coordenadas.

Manter cada polígono como uma entidade separada desta forma não mantém a organização topológica dos polígonos. Por topologia entendemos a relação entre diferentes objetos espaciais, isto é, quais polígonos são vizinhos, ou qual limite pertence a que área, etc. Em uma estrutura de polígonos inteiros, os segmentos de reta que definem os lados dos polígonos são gravados duas vezes, uma para cada polígono em cada lado da linha. Da mesma forma, vértices são gravados em vários polígonos.

### 3.2.2 - Estrutura DIME

A estrutura DIME ( do inglês Dual Independent Map Encoding ) foi desenvolvida para uso pelo U.S. Bureau of the Census. Ele foi projetado para incorporar informação topológica sobre áreas urbanas, para uso em análises demográficas. Os arquivos DIME não correspondem a uma estrutura de base de dados usada em SIG, mas são adequados para transferência de dados entre sistemas. O elemento básico do DIME é o segmento de reta definido por duas extremidades ou nós. A estrutura do DIME é complicada por que armazena, além dos nós, todos os atributos tais como nome do segmento, direção e sentido, código de área, código endereçamento postal, etc.

A maior desvantagem do DIME é a dificuldade de tratar linhas complexas, e os tamanhos de arquivos requeridos pelo particionamento de uma reta em diversos segmentos, exigindo grande esforço computacional. A figura 3.6 representa, com alguma simplificação, a estrutura de um arquivo DIME representando uma quadra definida por tres ruas.



Itens Título: Grupos de Segmentos			
Número do Título	CEP	DDD	Ident.
1000	13560-250	016	14
1001	13563-970	016	14

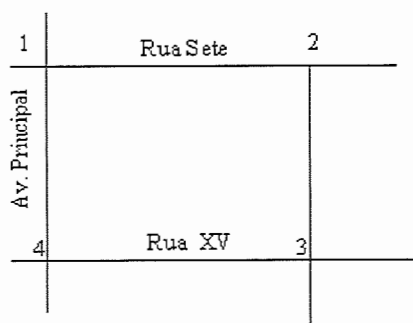
Códigos dos Segmentos: Cada segmento de reta							
Segmento	nós	Polígono	Endereços				
Nome	de para	Esq. Dir.	Esq. Início	esq. Fim	Dir. Início	Dir. Fim	Título
Rua Sete	1	-	101		100		1000
	2	Quadra		175		174	
Rua XV	3	-	103		102		1000
	4	Quadra		179		176	
Av. Principal	4	-			2450		1000
	1	Quadra	2451	2699		2698	

Locação dos nós		
Nó	Leste	Norte
1	127 251	1 340 600
2	127 352	1 340 601
3	127 350	1 040 589
4	127 249	1 040 598

Figura 3.6 - Estrutura simplificada da Estrutura DIME

### 3.2.3 - Estrutura Arco-Nó

Em uma estrutura de dados tipo Arco-Nó os objetos na base de dados são estruturados hierarquicamente. Neste sistema os pontos são os elementos básicos. Os arcos são formados por segmentos de reta, e são definidos por uma série de pares de coordenadas x,y. Os nós são as extremidades dos arcos e constituem os pontos de intersecção entre arcos. Existem ainda nós que não estão associados com linhas, devendo ser diferenciados dos nós nos extremos dos arcos. Polígonos são as áreas completamente cercadas por arcos. Esta estrutura é utilizada por diversos SIG comerciais. Na figura 3.7 é apresentado um arquivo com estrutura arco-nó.



Nós:				
Números	Leste	Norte	Controle de tráfico	Calçada
1	127 251	1 340 600	semáforo	sim
2	127 352	1 340 601	placas	sim
3	127 350	1 340 489	nada	sim
4	127 249	1 340 498	placas	sim

Arcos:					
Nome	de	para	Comprimento	Pavimento	faixas
Rua Sete	1	2	100	asfalto	2
s/nome	2	3	112	terra	2
Rua XV	3	4	101	asfalto	2
Av. Principal	4	1	102	concreto	4

Polígonos:					
Nome	proprietário	Arcos	perímetro	área	zoneamento
Q7-L1	José Silva	1-2,2-3,3-4,4-1	415	10 000	Residencial

Figura 3.7 - Estrutura arco-nó.

A estrutura arco-nó permite codificar a geometria dos dados sem redundância e, diferentemente da estrutura de polígonos inteiros, os pontos são armazenados somente uma vez. A facilidade de incluir dados de atributos é ampla, e eles são completamente ligados à geometria. No exemplo percebe-se prontamente que a sinalização de tráfico é ligada aos nós e condições das ruas e avenidas ligadas aos arcos apropriados.

Nota-se, no exemplo, que foram incluídos na base de dados foram incluídos os comprimentos dos arcos, o que é uma redundância pois eles são facilmente calculados das posições dos nós. Essa inclusão é muito útil se os comprimentos são muito usados nas aplicações, pois eles são calculados só uma vez na criação do banco, no lugar de calcular-los todas as vezes que necessários. Este é um exemplo de balanceamento entre custo de armazenamento e de processamento que deve ser continuamente gerenciado.

#### 3.2.4- Estrutura relacional

Uma base de dados com **estrutura relacional** é semelhante à **estrutura arco-nó** tendo como principal diferença o tratamento dado aos atributos, que são armazenados em arquivos separados. Este comportamento tornou-se uma estratégia muito popular para vários sistemas de informação geográfica comerciais. Os dados topológicos em uma estrutura relacional são organizados do mesma forma que no arco-nó, residindo a principal diferença nos valores dos atributos, que são armazenados em tabelas relacionais, que são acessadas através de métodos de indexação usando uma das colunas como índice, ou ponteiros. Dessa forma na estrutura relacional é mantido um maior número de arquivos separados, e um maior controle dos ponteiros.

#### 3.2.5- Estrutura de gráfico de linha digital

O U.S. Geological Survey investiu muito em dados de cartografia digital, e desenvolveu uma série de padrões para estes dados. Um dos mais conhecidos é o padrão de formato de arquivo vetorial, conhecido como Gráfico de Linha Digital (DLG). O conteúdo dos arquivos DLG é subdividido em diferentes camadas temáticas. A primeira camada consiste na informação de limites, incluindo ambos os limites político e administrativo da região. Uma segunda camada é para os comportamentos hidrológicos. Uma terceira para a rede de transporte da área. Finalmente a quarta camada é baseada no **Public Land Survey System**, que tem seu foco no sistema de topografia administrado pelo U.S. Bureau of Land Management.

Os elementos de dados essenciais nos três níveis do DLG são semelhantes aos discutidos anteriormente, com nós, linhas, linhas orientadas permitindo determinar começo e fim, além de definir lados da linha. É incluído um tipo especial de linha degenerada, que começa e termina no mesmo local, usada para definir locais representados por pontos no mapa. As áreas no formato DLG são completamente limitadas por segmentos de linhas. Cada área tem um ponto associado, que representa as características da área; a localização do ponto é arbitrária, e não precisa nem mesmo ser interna à área. Os códigos dos atributos são baseados nas propriedades existentes nos mapas topográficos do USGS. Os códigos são estruturados em duas partes, com um componente

majoritário, e um minoritário. O código majoritário tem três dígitos, com os dois primeiros definindo a categoria geral do elemento e o terceiro fornecendo algum detalhe especial. Alguns exemplos dos códigos majoritários são:

Código majoritário	Categoria
020	Hypsography
050	Hydrography
070	Cobertura da superfície

Os códigos minoritários são de quatro dígitos, o primeiro sempre zero, e os outros detalham o elemento.

Código minoritário	Categoria
001 - 099	nós
100 - 199	áreas
200 - 299	linhas
300 - 399	linha degenerada
400 - 499	códigos de propósito geral
600 - 699	Código descritivos

Vários códigos de atributos tem significado especial, por exemplo o código 000 0000 é reservado para áreas fora de uma dada prancha de mapa. Existem códigos reservados para dados revisados fotograficamente e para as características que não podem ser identificadas das fontes disponíveis. Para ilustrar alguns destes detalhes vamos apresentar alguns códigos da camada de dados de hidrografia, cujo código majoritário é 050.

<u>Nós</u>	<u>Áreas</u>
050 0001 Nascente de um rio	050 1010 Reservatório
050 0004 Entrada em reservatório	050 0103 Geleira
050 0005 Saída de reservatório	050 0106 Criação de peixes
<u>Linhas</u>	<u>Linhas degeneradas</u>
050 0200 Praia	050 0300 Fontes
050 0201 Praia artificial	050 0302 Poço jorrante
<u>Atributos de ordem geral</u>	<u>Atributos gerais descritivos</u>
050 0400 Corredeiras	050 0601 Subterrâneo
050 0401 Queda d'água	050 0602 Elevados
050 0406 Barragem ou vertedores	050 0604 Túnel

Um arquivo de dados DGL de 3 níveis contém algumas gravações de cabeçalho, seguida pela gravação dos dados. O cabeçalho fornece informações sobre a data de criação, tipo de projeção do



mapa e sistemas de coordenadas, e o número de pontos, linhas e áreas armazenadas no arquivo. Gravação de dados de nós incluem uma descrição (que contém a localização do nó), os códigos de atributo e um texto associado. Gravação de dados de linhas incluem uma descrição (que contém a identificação do início e fim, e as áreas à esquerda e direita), uma seqüência ordenada de coordenadas x, y ao longo da linha, os códigos de atributo e um texto associado. Gravação de dados de áreas incluem uma descrição (que contém a localização do ponto representativo), os códigos de atributo e um texto associado.

Pode existir uma gravação opcional com informações sobre a acuidade e precisão dos dados.

Já existe um novo desenvolvimento da estrutura DLG e da DIME, denominada TIGER (Topologically Integrated Geographic Encoding and Referencing system). Uma descrição do TIGER foge dos nossos objetivos.

## Capítulo 4

### Aquisição de dados

É universalmente aceito que o grande valor de um SIG é devido em grande parte a qualidade dos dados contidos no sistema. Essa afirmação traz a tona a particular importância da **aquisição de dados** para um SIG.

#### 4.1 Introdução

Uma base de dados para informação geográfica inicia-se a partir da coleta de dados para colocá-los no sistema. Um SIG deve ser capaz de aceitar uma ampla variedade de espécies e formatos de dados de entrada. Em alguns casos um usuário precisa gerar suas próprias bases de dados, embora os bancos de dados já são bastante amplos, e mesmo assim coletar dados é um dos maiores problemas operacionais e de custos. A importância dos dados em sistema automatizado de informações espaciais é descrito por Kennedy & Guinn (1975) da seguinte forma:

*Embora modelos que usam dados sejam importantes nas atividades de tomada de decisão daqueles que usam o sistema, uma grande porção de investimento deverá ser feita na obtenção, conversão e armazenamento de novos dados.*

A fonte de dados para a entrada de um SIG é variada, podendo ser proveniente de gráficos ou de tabelas. Estas fontes incluem mapas, fotografias, relatórios de visitas de especialistas dos mais variados campos, informações não espaciais relacionadas em arquivos digitais ou impressos. Existem informações em forma digital, que incluem dados

espaciais digitais de relatórios demográficos, ou de propriedades imóveis, ainda fitas magnéticas com informações de topografia, e ainda imagens de sensoriamento remoto.

A inclusão desses dados em um SIG requer um pré-processamento, manual ou automatizado, antes de codificação. Por exemplo, dados tabelados precisam ser introduzidos no computador, fotografia aérea requer análise de fotointerpretação para extrair os objetos importantes e sua localização, um processo de digitalização para converter os dados para forma digital, e algoritmos de retificação numérica para converter locações de dispositivos significantes para um sistema padrão de geo-referenciamento.

Um item muito importante e as vezes esquecido durante a aquisição de dados é obter informação da precisão, acuidade, custo e características espaciais dos próprios dados.

São fontes importantes de dados para um SIG os mapas, fotografias, aerofotogrametrias, tabelas etc. As propriedades geométricas são muito importantes, e isto vem da capacidade de definir como os dados foram locados, qual sistema de coordenadas e que tipo de projeção foram usadas. Isto implica em métodos de alterar o arranjo espacial de um conjunto de dados para corresponder a um desejado arranjo de outro conjunto correspondente à mesma posição.

Muitos conjuntos de dados espaciais de interesse são efetivamente um conjunto de pontos. Estes pontos incluem a locação de poços de água, gás e petróleo, a locação de representativa para grupos de edificações, e os endereços de membros de um determinado grupo demográfico alvo. Outras bases de dados podem ser considerados uma forma de rede; o conjunto das estradas em uma área, desde estradas secundárias não pavimentadas até as modernas rodovias de alta velocidade e diversas faixas de tráfego constituem uma rede de transporte. Outras redes incluem as hidrovias, as redes de distribuição de água de abastecimento, os sistemas coletores de águas residuárias, o sistema de estradas de ferro, as redes elétricas.

Outras espécies de dados espaciais podem ser descritos como meio contínuo, onde podemos teoricamente calcular ou medir um valor em qualquer locação. Exemplos desta classe de dados são as descrições de elevação, biomassa vegetal, ou densidade de população (junto com muitas espécies de de variáveis demográficas). Finalmente, uma importante classe de dados espaciais envolve a divisão de uma porção da superfície da

Terra em regiões discretas relativamente homogêneas. Um mapa de divisões geo-político é talvez o exemplo mais comum deste tipo, desde que ele subdivide uma porção da Terra em países, estados e municípios. Um modo similar de dividir a Terra seria desenvolver classes de coberturas vegetais ou uso da terra, indicando os limites das diferentes classes e as características dentro dos limites.

Cada uma destas diferentes espécies de dados espaciais pode ser armazenado e apresentado aos usuários de diversas maneiras. Produtos cartográficos tradicionais fornecem um meio fácil de entender para armazenar e comunicar essas várias espécies de dados espaciais. Em muitos casos fotografia pode ser usada em combinação com camadas cartográficas para destacar fenômenos e relações na Terra de maneira eficiente. Textos e tabelas são também de uso comum para arquivar informações sobre objetos espaciais.

Quando pensamos sobre base de dados espaciais, seja examinando um conjunto de dados existente ou propondo especificações para desenvolver um futuro banco de dados, há um número muito grande de elementos de dados não espaciais a considerar. A data do processo de coleta de dados é um item natural para ser armazenado. Ele é um dos elementos adicionais mais importantes para nos fornecer indicações do valor do conjunto de dados. A indicação temporal permite avaliar a evolução do comportamento daquele conjunto de dados, e permite também quando necessário escolher o mais atual levantamento de dados para determinada região. Outra consideração importante é sobre a procedência e critérios de um conjunto de dados. Por exemplo tem mais valor um mapa de cobertura vegetal desenvolvido por um especialista a partir de observações “in loco” do que mapas desenvolvidos a partir de hipóteses com base em localização e clima.

Portanto um usuário deve avaliar as características da melhor fonte disponível, e ainda considerar a acuidade geométrica dos dados, sua consistência lógica e sua abrangência.

Escala e resolução são dois itens separados em dados espaciais. Por exemplo duas fotos tiradas com a mesma câmara, da mesma posição e do mesmo local, terão a mesma escala, mas se um filme for de melhor qualidade, granulometria mais fina, e de maior contraste ele terá maior capacidade de resolver detalhes, fornecendo portanto uma melhor resolução, isto é, ela terá uma unidade mínima de mapeamento melhor.

As propriedades geométricas de um conjunto de dados espaciais são importantes. Nós devemos ser capazes de determinar os métodos particulares usados para especificar as coordenadas armazenadas nos dados. Isto usualmente implica que devemos ser capazes de identificar os sistemas de coordenadas e as projeções usados ou sermos capazes de dividir métodos para modificar o arranjo espacial original para que ele corresponda a um desejado arranjo.

De um ponto de vista bem simplificado podemos distinguir duas diferentes famílias de base de dados: **Conjunto de dados existentes**, aqueles que já estão compilados e disponíveis de alguma forma, não esquecendo que eles podem necessitar algum esforço para torná-los apropriados para algum uso particular; e os conjuntos de dados que **devem ser desenvolvidos** para uso próprio, neste caso temos controle total sobre o processo de coleta dos dados, mas também muito mais trabalho para prepará-los.

#### 4.2 Conjuntos de dados existentes

Usualmente ganha-se muito tempo através da procura de dados disponíveis, em vez de mergulhar de cabeça na produção de novos dados. Existe uma grande quantidade de dados espaciais disponível no domínio público para algumas partes do mundo. Deve-se tomar o devido cuidado já que diferentes países tratam diferentemente os dados espaciais. Nos Estados Unidos da América a maioria dos dados espaciais coletados por agências governamentais são considerados de domínio público, incluindo mapas, fotografias, e muitas espécies de dados digitais. Por outro lado, em alguns países esses dados são controlados, por questões econômicas ou de segurança.

A forma mais comum de dados espaciais é o mapa. Mapas de vários tipos são de uso comum para muitas espécies de análise espacial. No Brasil a fonte de mapas é o IBGE, que possui mapas cartográficos na escala de 1:50.000, e que podem ser introduzidos no computador através de um processo de digitalização. O INPE também tem mapas provenientes de fotografias de satélites, e que podem ser encontrados através da Internet.

O formato padrão de fotografias aéreas é impressão em papel com aproximadamente 9"x9". A inclinação com relação à vertical não excede 3°. A escala

dessas fotografias depende do comprimento focal da lente e da altitude em que ela foi tirada. São mais comuns as escalas de 1:20.000 e 1:40.000.

#### 4.3- Desenvolvendo seus próprios dados

A construção de um novo banco de dados envolve trabalho de campo de diversas espécies. É necessário criar mapas do terreno e da localização de certos comportamentos culturais. A localização de um poço vai exigir o conhecimento das tendências do nível das águas subterrâneas, bem como o ciclo anual de descarga e recarga do aquífero, e também para situar um terreno para disposição de resíduos sólidos.

O projeto de amostragem é um dos elementos importantes para qualquer plano de coleta de dados, onde decisões devem ser tomadas sobre como juntar os dados de interesse.

Na figura 4.1 são mostrados três diferentes disposições espaciais. Essas disposições representam um exercício no mapeamento da localização de certo tipo de planta. Na figura 5.1-a o conjunto pode ser descrito como **agregado** (clumped), pois os objetos estão agrupados em certas áreas; essa distribuição pode ser descrita como tendo **auto-correlação positiva** os objetos são tipicamente encontrados próximos uns dos outros da mesma espécie. Falando de plantas esta situação pode ocorrer em um ambiente favorável ao tipo de planta. Na figura 4.1-b os objetos estão em um conjunto **disperso**, ou de **auto-correlação negativa**, significando que os objetos são encontrados longe uns dos outros de mesma espécie. Voltando ao caso das plantas, esta situação ocorre com as plantas de deserto que competem por água, ou em florestas formadas onde as plantas competem por luz do sol e por espaço para raízes. Uma distribuição **aleatória** dessas plantas implica no posicionamento ao acaso das plantas. Podem existir plantas próximas umas das outras, como também bastante afastadas com a mesma probabilidade. (figura 5.1-c).

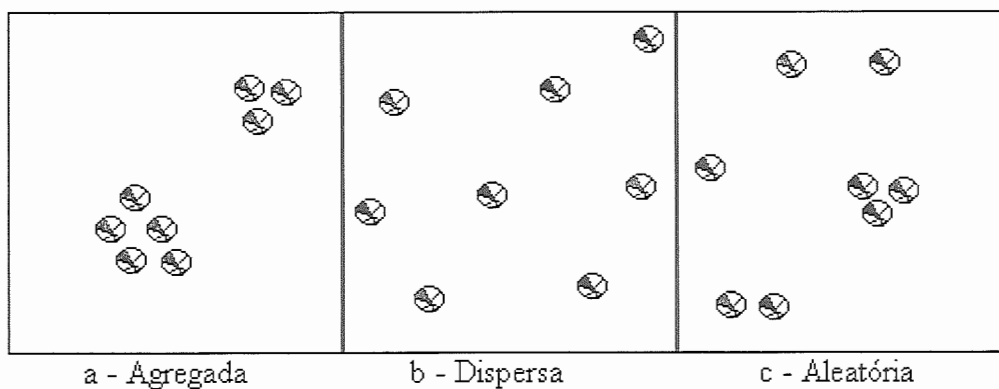


Figura 4.1- Tipos de distribuição de objetos no espaço.

**Amostragem de ponto** envolve a determinação da informação desejada em um único ponto. Por exemplo a procura de um ponto com determinada posição do lençol subterrâneo para a correta determinação da profundidade de perfuração. **Amostragem de linha** onde dois pontos são determinados e faz-se o censo da informação desejada ao longo da linha. Esta é uma prática comum na ecologia de vegetação, onde as espécies são contabilizadas ao longo de linhas escolhidas ou **transects**. A terceira opção é locar áreas de tamanho adequado ou **quadrats** na região de interesse, e determinar a informação necessária em cada área. Quadrats são tipicamente quadrados ou circulares, e uma das suas vantagens é que fornecem informação sobre a disposição espacial dentro das amostras e entre elas.

A estratégia de amostragem aplicada a um contínuo é usada para através de poucas medidas inferir o comportamento de toda a região. Consiste na distribuição de quadrats de forma a cobrir uma porcentagem da área e das medidas nos quadrats calcular a medida da região. Quando o quadrat é definido de forma a cobrir toda a área temos uma **amostragem exaustiva** que oferece a melhor informação, claro que a custa de enormes espaços de arquivos.

A distribuição dos quadrats, na área de interesse, pode ser **aleatória**, **sistemática** ou **estratificada**, cada uma delas com vantagens e restrições. Uma amostragem aleatória é feita com um número de quadrats suficiente para cobrir uma porcentagem determinada da área de interesse e distribuí-los segundo um critério de números aleatórios. Na figura 4.2 a é mostrada uma distribuição aleatória para contagem de objetos em uma área. A amostragem sistemática é feita através de uma distribuição regular dos quadrats, como

apresentada na figura 4.2 **b** para a mesma função anterior. A amostragem estratificada foi desenvolvida para uso em áreas não homogêneas, com distribuição de quadrats de acordo com a parte da área de interesse.

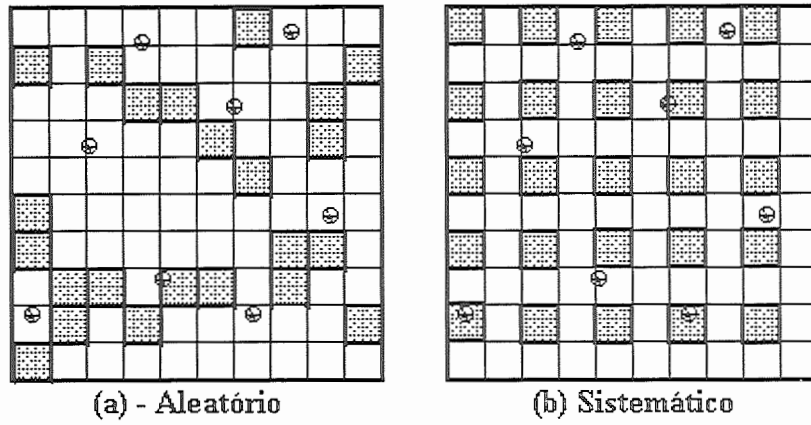


Figura 4.2 - Estratégias de amostragem



## Capítulo 5

### Pré-processamento

Os procedimentos de pré-processamento são usados para converter um conjunto de dados de tipo genérico, em uma forma adequada para armazenamento permanente em uma base de dados de um SIG. Os pré-processamento mais comuns são:

- conformação de tipos de dados
- sistema de geo-referenciamento
- estrutura de dados

O resultado final é um conjunto coordenado de camadas de dados temáticos. Os procedimentos de pré-processamento considerados essenciais incluem:

- conversão de formato
- redução de dados e generalização
- detecção de erros e edição
- transformação de pontos em linhas e de pontos e linhas em polígonos onde apropriado
- ajustagem de limites
- retificação e registro
- interpolação e
- fotointerpretação.

#### 5.1- Conversão de formato

Os dois principais problemas de conversão de formato referem-se a duas famílias: conversão entre diferentes estruturas de dados digitais e conversão entre diferentes médias de armazenamento de dados.

A primeira refere-se à conversão entre diferentes tipos de dados, por exemplo um arquivo binário convertido para um arquivo texto para facilitar o exame dos dados.

A segunda é ligada à conversão de uma fonte de material, tal como mapas, fotografias e tabelas, em uma forma compatível com o computador. O inverso desta conversão consiste na geração de produtos a partir do SIG.

#### 5.1.1- Conversão de estruturas de dados

Na armazenagem de dados alguns tem natureza mais compatível com a estrutura de dados ASCII, e outros são mais facilmente armazenados de forma binária, causando uma falta de homogeneidade no banco de dados. Este problema vem se tornando mais agudo à medida que crescentes quantidades de dados são criadas e mantidas em forma digital enquanto que os dados históticos são, quasi que universalmente, armazenados em papel ou filme.

Outro tipo de problema que aparece frequentemente é devido aos arquivos raster, tais como os obtidos por fotografias digitalizadas ou dados de escanerização multi-espectral, e o programa de SIG é do tipo de estrutura vetorial. Nestes casos é necessário inter-converter de um tipo de dados para o outro, tanto para processamento quanto para desenvolvimento de produto final.

Os tipos mais simples de conversão são entre membros de uma mesma família de estrutura. Por exemplo, existem diversos tipos de formatos raster para dados raster. Conjuntos de dados típicos de estrutura raster incluem mapas de elevação, de índices fluviométricos, dos tipos de cobertura vegetal. Os arquivos de formato raster incluem também os arquivos gerados por scanerização multi-espectral, que são sensores comuns em plataformas de aerofotogrametria e de satélites. Estes sistemas geram conjuntos de dados que são comparáveis a qualquer outra coleção multivariada de dados raster, incluindo problemas de registro geométrico entre as bandas de comprimento de onda ou entre diferentes datas de aquisição.

Existem diversas formas de organizar tais bases de dados. Pode-se manter cada variável (elevação, índice fluviométrico, etc.) como um arquivo separado, em método denominado de **band sequencial (BSQ)**, pelo fato de cada conjunto de dados da variável é mantido em um arquivo separado na fita magnética ou disco. Uma alternativa comum, chamada **band interleaved by pixel (BIP)**, coloca junto todas as diferentes medidas de um único pixel. Por exemplo, se o banco de dados contém tres informações do mesmo local (zoneamento, vegetação e elevação) o primeiro elemento conteria o valor da elevação para o pixel, o segundo o tipo de vegetação do mesmo pixel e o terceiro o código de zoneamento, a seguir reiniciaria a informação do segundo pixel com a elevação, a vegetação e o zoneamento e assim sucessivamente.



da topologia, tornando necessário diversas operações de referenciamento cruzado para identificar os vértices comuns e em seguida definir as áreas adjacentes e gerar os ponteiros para finalmente definir completamente as áreas.

A conversão de um arquivo de dados vetorial em estrutura raster é conceitualmente fácil, mas, praticamente, é difícil de implementar. Para elementos de dados pontuais a conversão vai transferir o atributo do ponto à célula raster cujo centro estiver mais próximo da coordenada geográfica do ponto. Portanto, por exemplo, o valor de uma referência de altitude é atribuído à célula raster cuja locação é mais próxima da do ponto original. É fácil perceber que essa operação, usualmente, muda a locação armazenada do ponto, pois é improvável que a coordenada original do ponto coincida exatamente com a do centro da célula raster. Este enfoque também ignora o problema de diferentes objetos ocupando a mesma célula. Devido a estas importantes limitações a conversão de uma estrutura vetorial para uma estrutura raster não é reversível: é impossível recuperar os dados originais dos pontos da estrutura raster sem erros. Para certos tipos de operação isto pode ser uma falha fatal.

Para elementos de dados lineares, a conversão de estruturas pode ser visualizada sobrepondo-se o vetor e o arranjo de células raster, como mostrado na figura 5.2. A estratégia de conversão mais simples seria identificar as células raster cruzadas pela linha, e codificar essas células com o atributo associado com a linha. Percebe-se da figura 5.2 que elementos inclinados com relação a orientação das linhas e colunas do arranjo raster a representação mostra uma distorção em degraus.

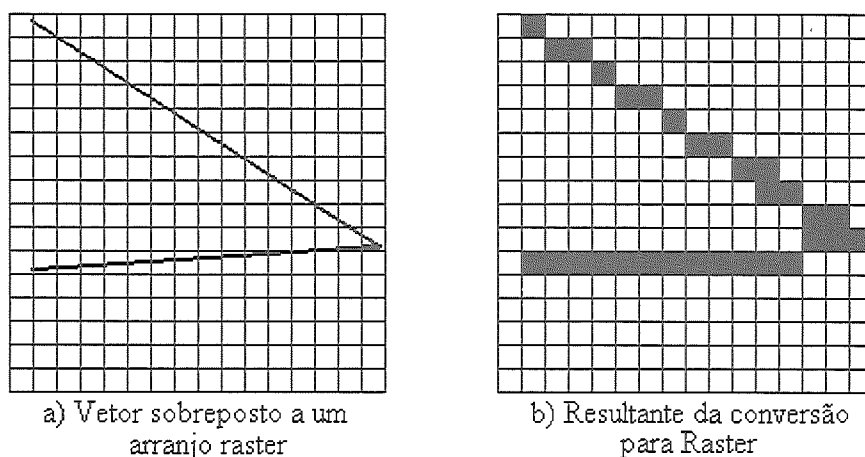


Figura 5.2- Conversão de elemento linear vetorial para estrutura raster.

A conversão de polígonos pode ser feita em dois passos. Primeiro os segmentos de linhas são convertidos como no exemplo anterior, formando um esqueleto do polígono. Segundo os elementos raster contidos no polígono são codificados para o valor do atributo apropriado.

A conversão de raster para vetor já é muito mais complexa. Uma forma simplista é considerar que cada célula raster é representada por um ponto no seu centro, e então são traçados os vetores entre as células não nulas, seguindo a lógica de quatro conexões. Este algoritmo modela o arranjo raster em vetores ortogonais, paralelos às linhas ou colunas, e com comprimento de uma célula raster, em uma restrição muito fora da realidade. Na figura 5.3 é apresentado esse comportamento.

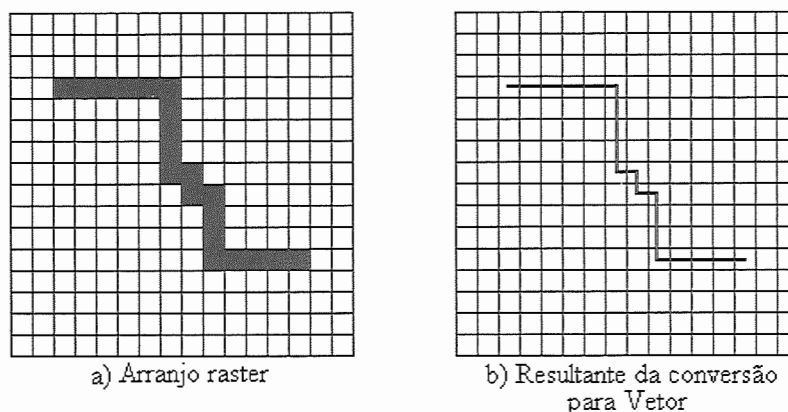


Figura 5.3- Conversão simples de raster para vetor.

#### 5.1.2- Conversão de mídia

A maioria dos dados espacializados disponíveis hoje não estão em formato compatível com computadores. Estes dados incluem mapas de muitas espécies e escalas, manuscritos impressos, e imagens baseadas em processos fotográficos ou geradas por outros instrumentos. A conversão destes materiais em formato compatível com computador é muito caro e consome muito tempo. O meio mais comum de converter mapas e outros dados gráficos em formato digital é através de mesas digitalizadoras.

Uma mesa digitalizadora é um dispositivo eletrônico composto por uma mesa, onde os mapas são fixados, e um cursor com uma mira, para localizar pontos nos mapas. A parte eletrônica da mesa converte a posição do cursor, relativa à mesa, em um sinal digital com precisão de 4 a 40 pontos por milímetro. Existem várias tecnologias usadas em mesas digitalizadoras comerciais, mas a mais aceita é a eletromagnética devido à sua precisão e acuracidade. O cursor da mesa geralmente é equipado com botões (1, 4 ou 16), que são utilizados para definir comandos para o programa sem que o operador necessite desviar sua atenção da mesa, para acessar o keyboard.

Os dados armazenados em papel ou filmes apresentam problemas de variação nas dimensões em função da temperatura ou umidade, e se o serviço exigir alta precisão deve-se dar

preferência aos filmes fotográficos devido a sua melhor estabilidade às condições climáticas, permitindo dados com distorção abixo de 0,2%. O uso de papel em condições não controladas pode elevar as distorções à 3% dependendo do tipo e gramatura do papel.

Ao iniciar um processo de digitalização o usuário deve especificar diversos atributos do mapa e a sua localização na mesa digitalizadora. Tipicamente o usuário é inquirido pelo sistema para fornecer informações sobre escala e tipo de projeção do mapa, e é solicitado a digitalizar, usando o cursor pontos de locação conhecida (latitude e longitude) especificando desta forma a informação de georeferenciamento da área de interesse. Usualmente este processo permite que qualquer posição subsequente do cursor seja transformada, sem ambiguidade, em locação geodésica.

O método de digitalização é pode adquirir tres formas de operação:

- O **modo ponto** permite a entrada de um único ponto quando o botão do cursor é acionado. É muito útil na digitalização de pontos como cruzamentos, poços, referências de altitude etc.
- O **modo linha** permite a entrada de linhas pela digitalização dos dois pontos extremos dessa linha. O sistema converte os dois pontos no vetor apropriado. É útil na entrada de longas retas como estradas, limites territoriais etc.
- O **modo contínuo** permite a entrada de dados de forma automática separados por curto espaço de tempo ou por por um deslocamento especificado, enquanto o botão estiver acionado. Os dados adquiridos neste modo apresentam-se com espaçamento aproximadamente regular. É de grande utilidade para adquirir dados de formas irregulares como cursos d'água ou margens de lagos e represas. O grande problema do uso deste modo é o tamanho dos arquivos, pois os dados são adquiridos velozmente. Além disso exige muito cuidado do operador.

Outro método muito importante para conversão de mídia é o scanner, que transforma o mapa em um arquivo raster de forma automática através de um sistema de varredura (scan) ótica. O sistema eletrônico varre o mapa e um sensor focalizado sobre o mapa percebe variações no brilho (ou cor) e as transforma em valores digitais que são armazenados. Os scanners tem alta resolução gráfica e podem ser em tons de cinza ou coloridos, com definição de tons de cinza (ou cores) de 16 a  $3 \times 10^6$ . A resolução gráfica em termos de pixel pode chegar a 1200 ppi (pontos por polegada) no entanto o tamanho do gráfico é proporcional à resolução gráfica e de cores, e os arquivos tendem a serem muito grandes.

Os arquivos scannerizados são também extremamente redundantes, pois não tem embutidos nenhum método de codificar a entrada para compactação do arquivo.

## 5.2- Generalização e redução de dados

Um registro de entrada pode requerer redução de dados de diversas espécies. Por exemplo, a identificação de uma árvore pode conter várias medidas ao redor da circunferência da árvore. Em uma dada aplicação pode ser necessário apenas a média dos dados de uma única árvore para a base de dados. Como outro exemplo conjuntos de dados podem definir a vegetação por espécies, e nossos requerimentos podem ser satisfeitos simplesmente por grupo de espécies. Uma forma mais complexa de redução de dados envolve mudanças de escala em dados espaciais. Por exemplo, necessitamos gravar os proprietários de terras em uma área, e as gravações do levantamento original são mais detalhadas que o requerido. Temos duas opções óbvias:

- ou aceitamos o nível de detalhe, portanto incorporando um volume de dados maior que o necessário com os custos de armazenagem e operação inerentes,
- ou desenvolvemos uma representação menos precisa da fonte de dados original. Esta última é denominada de generalização.

Para dados vetoriais, considere o lugar geométrico dos pontos, ligados por segmentos de reta, que descrevem a linha costeira. Neste caso um procedimento de generalização simples pode ser conseguido pela substituição de pares de pontos vizinhos pela média de suas coordenadas geográficas. Este processo reduz o número de pontos por um fator dois, e no processo reduz o nível de detalhes na base de dados.

Um ataque mais sofisticado envolve a modelagem do comportamento de um modesto número de pontos sucessivos, por um algoritmo numérico. Isto é, uma curva polinomial pode ser ajustada, pelo método dos mínimos quadrados, e então os pontos originais são substituídos por um menor número de pontos calculados pela aproximação numérica. Uma outra alternativa é eliminar pontos que são muito próximos para um determinado propósito, ou que caem ao longo de uma linha reta (dentro de certa tolerância). Essas operações são muito usadas em sistemas gráficos quando os pontos são muito pertos para serem distinguidos pela resolução gráfica do dispositivo de impressão. A repetição dessas operações gera uma representação de baixa resolução da base de dados vetorial original. Este enfoque é denominado “**thinning**”.

Em arquivos de dados raster pode-se criar um conjunto de dados generalizado substituindo o valor de quatro células vizinhas pelo valor médio dos atributos das quatro células em uma nova célula com o tamanho de quatro células originais. Este procedimento é chamado de **resampling** (re-amostragem).

### 5.3- Detecção de erros e correção

Os sistemas de base de dados sofrem a influência de diversos tipos de erros, dependentes do tipo de fonte que os gerou. Para ilustrar essa variedade vamos discutir alguns tipos de erros que ocorrem na geração de uma base de dados vetorial.

Se um objeto gráfico foi codificado como polígono, seu limite deve ser uma curva contínua, isto é, deve ser fechada. Os softwares devem ser capazes de detectar se um polígono não está fechado. As causas para este tipo de erro incluem falhas de codificação, o objeto é codificado como vetor em vez de polígono, ou de digitalização, pontos ao longo do polígono ou a conexão entre pontos são perdidos. Na figura 5.4 a é mostrada uma representação desse tipo de erro.

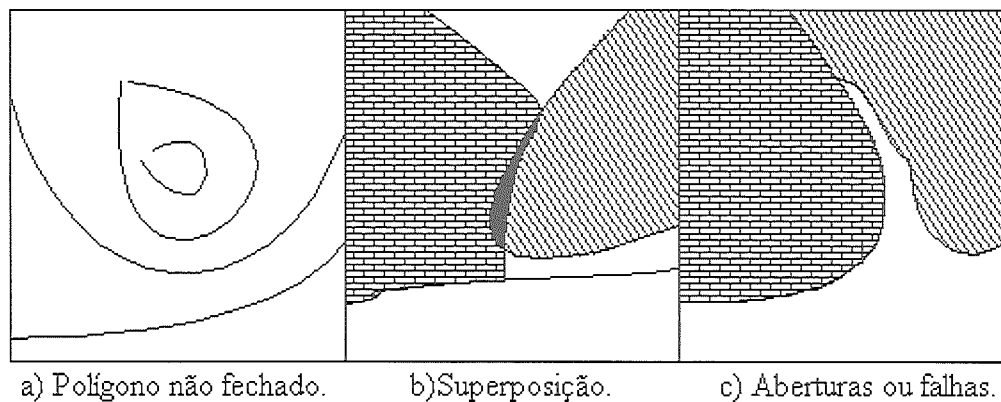


Figura 5.4- Erros comuns em arquivos vetoriais.

Em arquivos vetoriais é comum a digitalização duplicada de alguns limites, um para cada polígono que divide aquele lado. Tais procedimentos podem criar **superposição** de áreas ou **aberturas** (ou **falhas**) entre áreas adjacentes. Na figura 5.4 são mostrados esses comportamentos em (b) e (c). As superposições e aberturas ocorrem com frequência quando os documentos não são inteiramente digitalizados numa única sessão.

### 5.4- Fusão

É o processo de construir objetos mais complexos a partir dos pontos elementares. Com base nos dados adquiridos durante uma sessão de digitalização, vetores são construídos pela conexão dos pontos apropriados, e os polígonos pela ligação dos vetores apropriados.



## 5.5- Ajustagem de descontinuidades

É frequente a necessidade de utilizar mais de um mapa para cobrir a região de interesse. O mesmo problema ocorre com conjuntos de dados digitais, armazenados em diversos arquivos para cobrir a região de interesse. Nestes casos devemos ser capazes de extrair informações de cada uma das partes relevantes e processar a informação de forma que os limites dos mapas se ajustem sem erros.

Normalmente quando duas folhas de mapa são digitalizadas ou escaneadas separadamente, os objetos que cruzam o limite entre as duas folhas não se alinham corretamente. Mesmo quando os mapas são impressos sem erro discernível, as condições de umidade e temperatura podem alterar os tamanhos de forma significativa. De forma geral os problemas de desalinhamento devem ser solucionados antes dos dados serem introduzidos na base de dados do SIG. Na figura 5.5 são mostrados uma junção de mapas com problemas e a mesma junção corrigida.

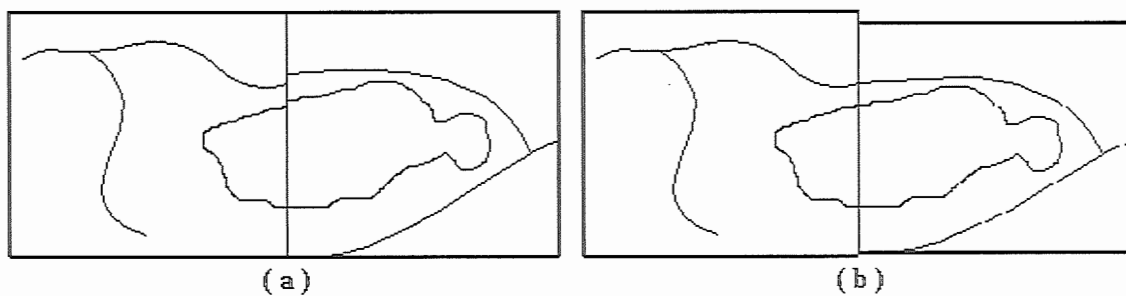


Figura 5.5 - Erros de ajustagem: (a) Mapas com erros no limite. (b) Mapas ajustados.