

## ARQUITETURA DE SISTEMAS DE INFORMAÇÃO GEOGRÁFICA

Gilberto Câmara

Gilberto Ribeiro de Queiroz

### 3.1 DESCRIÇÃO GERAL

O termo *Sistemas de Informação Geográfica* (SIG) é aplicado para sistemas que realizam o tratamento computacional de dados geográficos e recuperam informações não apenas com base em suas características alfanuméricas, mas também através de sua localização espacial; oferecem ao administrador (urbanista, planejador, engenheiro) uma visão inédita de seu ambiente de trabalho, em que todas as informações disponíveis sobre um determinado assunto estão ao seu alcance, interrelacionadas com base no que lhes é fundamentalmente comum -- a localização geográfica. Para que isto seja possível, a geometria e os atributos dos dados num SIG devem estar *georreferenciados*, isto é, localizados na superfície terrestre e representados numa projeção cartográfica.

O requisito de armazenar a geometria dos objetos geográficos e de seus atributos representa uma *dualidade* básica para SIGs. Para cada objeto geográfico, o SIG necessita armazenar seus atributos e as várias representações gráficas associadas. Devido a sua ampla gama de aplicações, que inclui temas como agricultura, floresta, cartografia, cadastro urbano e redes de concessionárias (água, energia e telefonia), há pelo menos três grandes maneiras de utilizar um SIG:

- como ferramenta para produção de mapas;
- como suporte para análise espacial de fenômenos;
- como um banco de dados geográficos, com funções de armazenamento e recuperação de informação espacial.

Estas três visões do SIG são antes convergentes que conflitantes e refletem a importância relativa do tratamento da informação geográfica dentro de uma instituição. Para esclarecer ainda mais o assunto, apresentam-se a seguir algumas definições de SIG:

*“Um conjunto manual ou computacional de procedimentos utilizados para armazenar e manipular dados georreferenciados” (Aronoff, 1989);*

*“Conjunto poderoso de ferramentas para coletar, armazenar, recuperar, transformar e visualizar dados sobre o mundo real” (Burrough, 1986);*

*“Um sistema de suporte à decisão que integra dados referenciados espacialmente num ambiente de respostas a problemas” (Cowen, 1988);*

*“Um banco de dados indexados espacialmente, sobre o qual opera um conjunto de procedimentos para responder a consultas sobre entidades espaciais” (Smith et al., 1987).*

Estas definições de SIG refletem, cada uma à sua maneira, a multiplicidade de usos e visões possíveis desta tecnologia e apontam para uma perspectiva interdisciplinar de sua utilização. A partir destes conceitos, é possível indicar as principais características de SIGs:

- Inserir e integrar, numa única base de dados, informações espaciais provenientes de dados cartográficos, dados censitários e cadastro urbano e rural, imagens de satélite, redes e modelos numéricos de terreno;
- Oferecer mecanismos para combinar as várias informações, através de algoritmos de manipulação e análise, bem como para consultar, recuperar, visualizar e plotar o conteúdo da base de dados georreferenciados.

### 3.2 ESTRUTURA GERAL DE UM SIG

Numa visão abrangente, pode-se indicar que um SIG tem os seguintes componentes:

- Interface com usuário;
- Entrada e integração de dados;
- Funções de consulta e análise espacial;
- Visualização e plotagem;
- Armazenamento e recuperação de dados (organizados sob a forma de um banco de dados geográficos).

Estes componentes se relacionam de forma hierárquica. No nível mais próximo ao usuário, a *interface homem-máquina* define como o sistema é operado e controlado. No nível intermediário, um SIG deve ter mecanismos de processamento de dados espaciais (*entrada, edição, análise, visualização e saída*). No nível mais interno do sistema, um *sistema de gerência de bancos de dados*

*geográficos* oferece armazenamento e recuperação dos dados espaciais e seus atributos.

De uma forma geral, as funções de processamento de um SIG operam sobre dados em uma *área de trabalho* em memória principal. A ligação entre os dados geográficos e as funções de processamento do SIG é feita por mecanismos de seleção e consulta que definem restrições sobre o conjunto de dados. Exemplos ilustrativos de modos de seleção de dados são:

- "Recupere os dados relativos à carta de Guajará-Mirim " (*restrição por definição de região de interesse*);
- "Recupere as cidades do Estado de São Paulo com população entre 100.000 e 500.000 habitantes" (*consulta por atributos não-espaciais*).
- "Mostre os postos de saúde num raio de 5 km do hospital municipal de S.J.Campos" (*consulta com restrições espaciais*).

A Figura 3.1 indica o relacionamento dos principais componentes ou subsistemas de um SIG. Cada sistema, em função de seus objetivos e necessidades, implementa estes componentes de forma distinta, mas todos os subsistemas citados devem estar presentes num SIG.

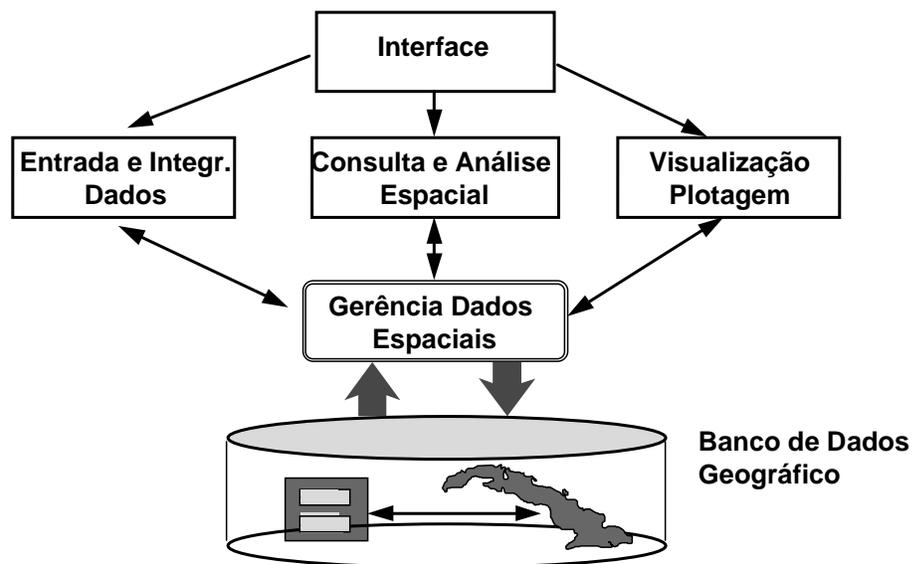


Figura 3.1 - Estrutura Geral de Sistemas de Informação Geográfica

### 3.3 GERÊNCIA DE DADOS EM UM SIG

Atualmente, a principal diferença entre os SIGs é a forma como os dados geográficos são gerenciados. Há basicamente três diferentes arquiteturas de SIGs que utilizam os recursos de um SGBD: *dual*, *integrada baseada em SGBDs relacionais* e *integrada baseada em extensões espaciais sobre SGBDs objeto-relacionais*.

#### 3.3.1 Arquitetura Dual

Um SIG implementado com a estratégia dual utiliza um SGBD relacional para armazenar os atributos convencionais dos objetos geográficos (na forma de tabelas) e arquivos para guardar as representações geométricas destes objetos. No *modelo relacional*, os dados são organizados na forma de uma *tabela* onde as *linhas* correspondem aos *dados* e as *colunas* correspondem aos *atributos*.

A entrada dos *atributos não-espaciais* é feita por meio de um SGBD relacional e para cada entidade gráfica inserida no sistema é imposto um identificador único ou *rótulo*, através do qual é feita uma *ligação lógica* com seus respectivos atributos não-espaciais armazenados em tabelas de dados no SGBD, como ilustrado na Figura 3.2

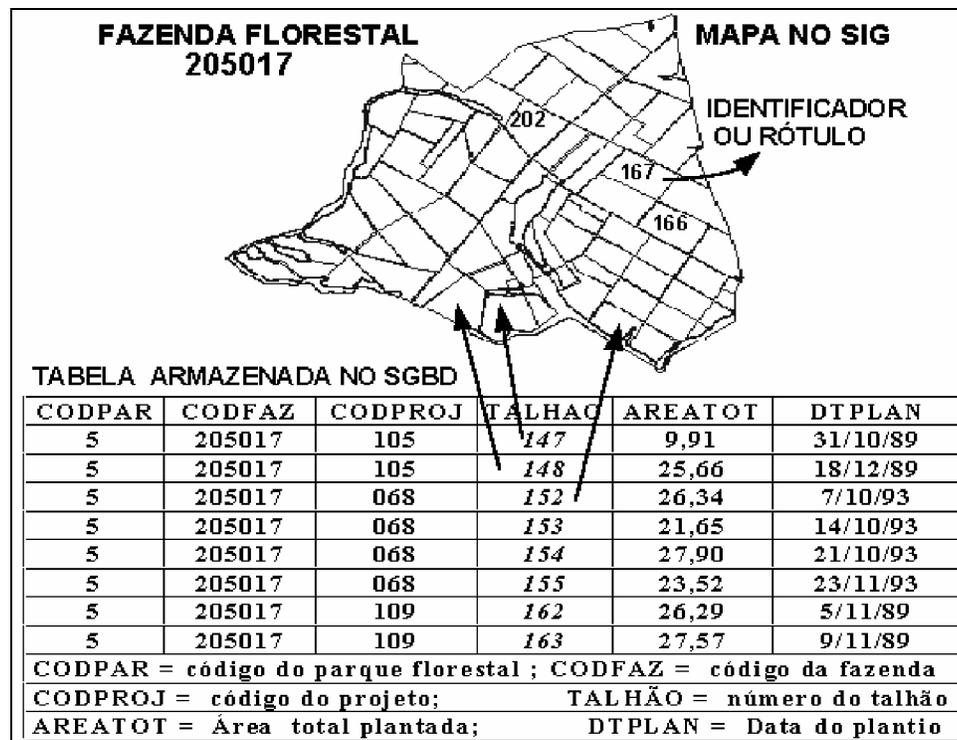


Figura 3.2 Estratégia dual para bancos de dados geográficos.

A Figura 3.2 exemplifica as ligações lógicas criadas entre os *rótulos* dos talhões de um mapa florestal e seus atributos correspondentes (registros no “campo” ou *coluna* TALHÃO) numa tabela de banco de dados. O mesmo tipo de relacionamento lógico pode ser feito em outros casos, como por exemplo: moradores em um lote, lotes em uma quadra, quadras em bairro, bairros em uma cidade; hidrantes de segurança ou telefones públicos ao longo de uma avenida; postos de serviço e restaurantes ao longo de uma rodovia.

A principal vantagem desta estratégia é poder utilizar os SGBDs relacionais de mercado. No entanto, como as representações geométricas dos objetos espaciais estão fora do controle do SGBD, esta estrutura dificulta o equacionamento das questões de otimização de consultas, gerência de transações e controle de integridade e de concorrência. Exemplos de sistemas comerciais baseados em estratégia dual são o ARC/VIEW, MGE e o SPRING (Câmara et al., 1996b). As principais desvantagens desta arquitetura são:

- § Dificuldades no controle e manipulação dos dados espaciais;
- § Dificuldade em manter a integridade entre a componente espacial e a componente alfanumérica;
- § Consultas mais lentas, pois são processadas separadamente. A parte convencional da consulta é processada pelo SGBD separado da parte espacial, que é processada pelo aplicativo utilizando os arquivos proprietários;
- § Falta de interoperabilidade entre os dados. Cada sistema produz seu próprio arquivo proprietário sem seguir um formato padrão, o que dificulta a integração destes dados.

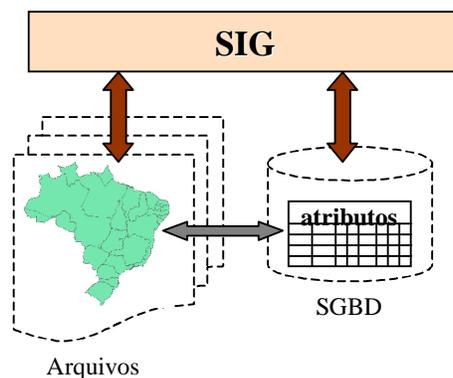


Figura 3.3a- Arquitetura Dual

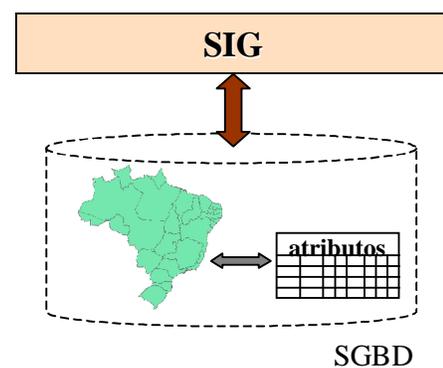


Figura 3.3b - Arquitetura Integrada

### 3.3.2 Arquitetura Integrada para Gerência de Dados

A arquitetura Integrada, mostrada na Figura 3.3b, consiste em armazenar todo o dado espacial em um SGBD, tanto sua componente espacial como a parte alfanumérica. Sua principal vantagem é a utilização dos recursos de um SGBD para controle e manipulação de dados espaciais, como gerência de transações, controle de integridade e concorrência. Sendo assim, a manutenção de integridade entre a componente espacial e alfanumérica é feita pelo SGBD. Há duas alternativas para a arquitetura integrada: (a) *baseada em SGBDs relacionais*; (b) *baseada em extensões espaciais sobre SGBDs objeto-relacionais*

A arquitetura integrada *baseada em um SGBD relacional* utiliza campos longos, chamados de BLOBs, para armazenar a componente espacial do dado. Suas principais desvantagens são:

- § Não é capaz de capturar a semântica dos dados espaciais: como o SGBD trata o campo longo como uma cadeia binária, não é possível conhecer a semântica do seu conteúdo;
- § Métodos de acesso espacial e otimizador de consultas devem ser implementados pelo SIG: como o SGBD trata os dados espaciais como uma cadeia binária, não possui mecanismos satisfatórios para o seu tratamento;
- § Limitações da linguagem SQL<sup>1</sup> para a manipulação dos dados espaciais: a SQL padrão oferece recursos limitados para o tratamento de campos longos.

O outro tipo de arquitetura integrada consiste em utilizar *extensões espaciais desenvolvidas sobre SGBDs objeto-relacionais* (SGBDOR). Estas extensões contêm funcionalidades e procedimentos que permitem armazenar, acessar e analisar dados espaciais de formato vetorial. Como desvantagens dessa arquitetura podem ser citadas as faltas de mecanismos de controle de integridade sobre os dados espaciais e a falta de padronização das extensões da linguagem SQL. Os SGBDs objeto-relacionais, também chamados de SGBDs extensíveis, oferecem recursos para a definição de novos tipos de dados e de novos métodos ou operadores para manipular esses tipos, estendendo assim seu modelo de dados e sua linguagem de consulta. Por isso, um SGBDOR é mais adequado para tratar dados complexos, como dados geográficos, do que um SGBDR, o qual não oferece esses recursos.

---

<sup>1</sup> SQL (*Structured Query Language*): linguagem padrão para manipular bancos de dados relacionais. Incluem recursos para definir estruturas de dados; consultar, inserir e modificar dados do banco de dados e especificar restrições de segurança.

Um SGBDOR que possui uma extensão para tratar dados espaciais deve ter as seguintes características:

- § Fornecer tipos de dados espaciais (TDEs), como ponto, linha e região, em seu modelo de dados e manipulá-los assim como os tipos alfanuméricos básicos (inteiros, string, etc);
- § Estender a linguagem de consulta SQL para suportar operações e consultas espaciais sobre TDEs;
- § Adaptar outras funções de níveis mais internos para manipular TDEs eficientemente, tais como métodos de armazenamento e acesso (indexação espacial) e métodos de otimização de consultas (junção espacial).

Portanto, além dos TDEs, as extensões espaciais fornecem operadores e funções que são utilizados, juntamente com a linguagem de consulta do SGBD, para consultar relações espaciais e executar operações sobre TDEs. Além disso, fornecem métodos de acesso eficiente de TDEs através de estruturas de indexação, como R-tree e QuadTree.

Existem hoje basicamente três extensões comerciais disponíveis no mercado: Oracle Spatial (Ravada e Sharma, 1999), IBM DB2 Spatial Extender (IBM, 2001) e Informix Spatial Datblade (IBM, 2002). Há ainda um projeto de implementar uma extensão espacial chamada PostGIS (Ramsey, 2002) para o SGBD PostgreSQL, que é objeto-relacional, gratuito e de código fonte aberto. Todas essas extensões baseiam-se nas especificações do OpenGIS (OGC, 1996), porém, possuem variações relevantes entre os modelos de dados, semântica dos operadores espaciais e mecanismos de indexação. O OpenGIS é uma associação formada por organizações públicas e privadas envolvidas com SIGs, dedicada à criação e gerenciamento de uma arquitetura padrão para geoprocessamento. Seu objetivo técnico é definir e manter:

- § Um modelo universal de dados espaço-temporais e de processos, chamado modelo de dados OpenGIS;
- § Uma especificação para cada uma das principais linguagens de consulta a banco de dados para implementar o modelo de dados OpenGIS;
- § Uma especificação para cada um dos principais ambientes computacionais distribuídos para implementar o modelo de processo OpenGIS.

### 3.4 UMA VISÃO GERAL DA TECNOLOGIA DE SIG

O conceito "sistemas de informação geográfica" está relacionado hoje a diferentes alternativas. Atualmente, há uma grande diversificação de oferta de SIG, com pelo menos quadro grandes tecnologias complementares:

- Os "GIS desktop", com interfaces amigáveis e crescente funcionalidade.
- Os "Gerenciadores de Dados Geográficos", que armazenam os dados espaciais em ambiente multi-usuário.
- Os "Componentes GIS", ambientes de programação que fornecem insumos para que o usuário crie seu próprio aplicativo geográfico.
- Os "Servidores Web de Dados Geográficos", utilizados para publicação e acesso a dados geográficos via Internet.

Os "GIS desktop" são sistemas herdeiros da tradição de Cartografia, com suporte de bancos de dados limitado e cujo paradigma típico de trabalho é o mapa (chamado de "cobertura" ou de "plano de informação"). Desenvolvidos a partir da início da década de 80 para ambientes da classe VAX e - a partir de 1985 - para sistemas PC/DOS, esta classe de sistemas é utilizada principalmente em projetos isolados, sem a preocupação de gerar arquivos digitais de dados. Esta geração também pode ser caracterizada como sistemas orientados a projeto ("project-oriented GIS").

A *segunda geração* de SIGs ("banco de dados geográfico") chegou ao mercado no início da década de 90 e caracteriza-se por ser concebida para uso em ambientes cliente-servidor, acoplado a gerenciadores de bancos de dados relacionais e com pacotes adicionais para processamento de imagens. Esta geração também pode ser vista como sistemas para suporte à instituições ("enterprise-oriented GIS").

A *terceira geração* de SIGs ("bibliotecas geográficas digitais" ou "centros de dados geográficos"), caracterizada pelo gerenciamento de grandes bases de dados geográficos, com acesso através de redes locais e remotas, com interface via WWW (World Wide Web). Para esta terceira geração, o crescimento dos bancos de dados espaciais e a necessidade de seu compartilhamento com outras instituições requer o recurso a tecnologias como *bancos de dados distribuídos e federativos*. Estes sistemas deverão seguir os requisitos de *interoperabilidade*, de maneira a permitir o acesso de informações espaciais por SIGs distintos. A terceira geração de SIG pode ainda ser vista como o desenvolvimento de sistemas

orientados para troca de informações entre uma instituição e os demais componentes da sociedade ("society-oriented GIS").

Um aspecto fundamental das diferentes tecnologias apresentadas é sua *complementaridade*: os "GIS desktop" podem utilizar "gerenciadores de dados geográficos", que podem estar ligados a "servidores web", e os usuários destes dados podem ter interfaces personalizadas, construídas a partir de "componentes GIS". Neste capítulo, estaremos dando ênfase à organização dos gerenciadores de dados geográficos, pois são estes que permitem a organização de grandes bancos de dados em ambientes corporativos. Apresentamos a seguir uma breve visão do contexto geral das geotecnologias.

### 3.4.1 GIS DESKTOP

A primeira geração de SIG caracteriza-se por sistemas com operações gráficas e de análise espacial sobre arquivos ("flat files"). Sua ligação com gerenciadores de bancos de dados é parcial (parte das informações descritivas se encontra no sistema de arquivos) ou inexistente. Mais adequados à realização de *projetos de análise espacial* sobre regiões de pequeno e médio porte, estes sistemas enfatizam o aspecto de mapeamento. O sistema permite a entrada de dados sem definição prévia do esquema conceitual, assemelhando-se assim a ambientes de CAD que possuem a capacidade de representar projeções cartográficas e de associar atributos a objetos espaciais. Por força de sua concepção, tais ambientes não possuem suporte adequado para construir grandes bases de dados espaciais.

Num "GIS desktop" tradicional, os dados geográficos são armazenados de forma separada, com os atributos descritivos guardados em tabelas (usualmente no padrão xBase) e as geometrias em formatos proprietários (como os "shapefiles" do ARC/VIEW). Originalmente sistemas simples de consulta e apresentação de dados, os "GIS desktop" tem evoluído para oferecer uma crescente gama de funcionalidade, incluindo:

- A combinação de tratamento de dados vetoriais e matriciais ("raster") no mesmo ambiente, com uma integração maior entre Processamento de Imagens e GIS, a exemplo do IDRISI.
- Linguagens de programação de "scripts", em que as variáveis refletem os tipos de dados geográficos suportados pelo sistema (e.g., AVENUE do ARC/VIEW e MAPBASIC do MAPINFO), e que permitem ampliar a funcionalidade disponível.

- Ferramentas sofisticadas de Análise Espacial, como os módulos de Geoestatística disponíveis nas novas versões do IDRISI, ARC/INFO E SPRING e funções de Álgebra de Mapas como as disponíveis no módulo SPATIAL ANALYST do ARC/VIEW.
- Uma integração do "desktop" com os gerenciadores de dados geográficos, como no caso da ligação entre GEOMEDIA com ORACLE SPATIAL e AUTODESK WORLD com VISION\* e TerraView com mySQL e PostgreSQL.
- O aumento da potencial de interoperabilidade e da conversão automática de formatos de dados geográficos, como suportado pelo GEOMEDIA.
- O uso de conceitos de orientação-a-objetos, que permitem uma aproximação melhor entre os problemas do mundo real e sua representação computacional, como no ARC/GIS-9 e no SPRING.

### 3.4.2 A SEGUNDA GERAÇÃO: BANCOS DE DADOS GEOGRÁFICOS

Nos anos recentes, a interesse pelo uso de GIS no ambiente corporativo levou ao aparecimento de *gerenciadores de dados geográficos*, que armazenam tanto a geometria como os atributos dos objetos dentro de um sistema gerenciador de bancos de dados (SGBD). As principais vantagens desta estratégia são: (a) evitar os problemas de controle de integridade típicos do ambiente "desktop", permitindo o acesso concorrente aos dados; (b) facilitar a integração com as bases corporativas já existentes, como sistemas legados, que já utilizam SGBDs relacionais. Um SGBD apresenta os dados numa visão independente dos sistemas aplicativos, além de garantir três requisitos importantes: *eficiência* (acesso e modificações de grandes volumes de dados); *integridade* (controle de acesso por múltiplos usuários); e *persistência* (manutenção de dados por longo tempo, independentemente dos aplicativos que acessem o dado). O uso de SGBD permite ainda realizar, com maior facilidade, a interligação de banco de dados já existente com o sistema de Geoprocessamento.

Estes gerenciadores possuem dois componentes, usualmente de distintos fabricantes: um SGBD com suporte a dados geográficos e uma "camada de acesso", que fornece um ambiente de armazenamento e recuperação, visível externamente ou integrado a um "GIS desktop". As "camadas de acesso" interagem com estes servidores, para fornecer um ambiente de armazenamento e recuperação, através

de uma interface de programação (API). Dentre os gerenciadores de bancos de dados com capacidades de armazenamento de dados geográficos, estão o ORACLE, PostgreSQL e MySQL. Dentre as API para acesso a estes bancos, estão o ArcSDE da ESRI e a TerraLib desenvolvida pelo INPE.

### 3.4.3 A TERCEIRA GERAÇÃO: BIBLIOTECAS GEOGRÁFICAS DIGITAIS

Uma *biblioteca geográfica digital* (ou um “centro de dados geográfico”) é um banco de dados geográfico compartilhado por um conjunto de instituições. Esta biblioteca deve ser acessível remotamente e armazenar, além dos dados geográficos, descrições acerca dos dados (“metadados”) e documentos multimídia associados (texto, fotos, áudio e vídeo). Este novo paradigma é motivado pelo aguçar da nossa percepção dos problemas ecológicos, urbanos e ambientais, pelo interesse em entender, de forma cada vez mais detalhada, processos de mudança local e global e pela necessidade de compartilhar dados entre instituições e com a sociedade. O núcleo básico de uma biblioteca geográfica digital é um grande banco de dados geográficos. Para ilustrar, imaginemos dois cenários: um grande centro ambiental brasileiro e uma secretaria de planejamento de uma prefeitura de médio porte (cerca de 1.000.000 de habitantes):

- Cenário 1 (“*Banco de Dados da Amazônia*”): Este banco conteria informações detalhadas sobre a ocupação humana na região, incluindo temas como temas básicos (vegetação, pedologia, geomorfologia), ocupação humana (desmatamento, grandes projetos agropecuários, áreas de prospecção mineral), temas derivados (zoneamentos econômicos) e imagens de satélite atualizadas. Este banco de dados poderia ser mantido por uma instituição central (como o IBAMA) e permitir o acesso de forma concorrente por pesquisadores de todo o país.
- Cenário 2 (“*Prefeitura de Curitiba*”): Este banco conteria todas as informações necessárias para planejamento da cidade, incluindo: lotes, quadras, ruas, equipamentos urbanos (hospitais, escolas), redes de água, esgoto e luz. Poderia ser consultado on-line pelas diversas secretarias municipais, por concessionárias de serviço e por cidadãos.

Nos dois cenários, o ambiente deve garantir acesso concorrente a uma comunidade de usuários, com diferentes métodos de seleção, incluindo folheamento (“browsing”) e linguagem de consulta. Para facilitar a difusão de informação geográfica através da Internet, a solução mais adotada é acoplar um servidor de dados geográficos. A solução mais adotada é o uso de servidores de imagens, que, respondendo a pedidos remotos, enviam uma imagem (matriz) de

tamanho fixo nos formatos GIF ou JPEG. Esta solução permite configurar o servidor para responder a diferentes tipos de consulta, sem requerer que todos os dados a ser transmitidos sejam pre-computados. Como exemplo, temos o "Internet Map Server", da ESRI e o MapServer (produto de software livre). O protocolo WMS (Web Map Server) do consórcio OpenGIS encapsula esta funcionalidade.

#### 3.4.4 BIBLIOTECAS DE COMPONENTES GIS

Nenhum GIS nasce pronto e muitas vezes, há necessidade de desenvolver aplicativos dirigidos especificamente para um cliente. Para tanto, uma tendência crescente destes últimos anos é fornecer um *ambiente de componentes*, com tipos de dados geográficos básicos e métodos de acesso e apresentação. A linguagem de programação mais comum é VISUALBASIC, como no caso dos produtos MAPOBJECTS da ESRI e MAPX da MAPINFO. A comunicação com outras aplicações pode ser conseguida utilizando recursos do Windows, como OLE (*Object Linking and Embedding*) e ODBC (*Open Database Connectivity*). A alternativa (utilizada usualmente por projetos de software livre) é oferecer uma biblioteca de rotinas em código aberto num ambiente de programação em linguagens como C ou C++. Este é o caso da biblioteca TerraLib. Neste caso, toda a funcionalidade do sistema está disponível, mas precisa ser remontada pelo programador.