

**Qualidade de Dados em Aplicações  
Geográficas**

*Alexandre Carvalho de Alencar*

**Dissertação de Mestrado**

# Qualidade de Dados em Aplicações Geográficas

Alexandre Carvalho de Alencar<sup>1</sup>

Março de 2000

## Banca Examinadora:

- Profa. Dra. Claudia Bauzer Medeiros  
Instituto de Computação - Unicamp (Orientadora)
- Prof. Dr. Juliano Lopes de Oliveira  
Instituto de Informática - UFG
- Prof. Dr. Mario Lúcio Côrtes  
Instituto de Computação - Unicamp
- Profa. Dra. Maria Beatriz Felgar de Toledo  
Instituto de Computação - Unicamp (Suplente)

---

<sup>1</sup>O autor é Bacharel em Ciências Navais com Habilitação na Área de Sistemas de Armas, formado pela Escola Naval.

UNIVERSIDADE DE CAMPINAS  
CHAMADA: UNICAMP  
AL25q  
BO 41521  
278/00  
[ ] [ ] [X]  
R\$ 11,00  
12-07-00  
CPD

CM-00143110-0

FICHA CATALOGRÁFICA ELABORADA PELA  
BIBLIOTECA CENTRAL DA UNICAMP

Al25q Alencar, Alexandre Carvalho de  
Qualidade de dados em aplicações geográficas /  
Alexandre Carvalho de Alencar. -- Campinas, SP :  
[ s.n.], 2000.

Orientador : Claudia M. Bauzer Medeiros.  
Dissertação (mestrado) - Universidade Estadual de  
Campinas, Instituto de Computação.

1. Sistemas de Informações geográficas. 2. Banco  
de dados. I. Medeiros, Claudia Maria Bauzer.  
II. Universidade Estadual de Campinas. Instituto de  
Computação. III. Título.

# Qualidade de Dados em Aplicações Geográficas

Este exemplar corresponde à redação final da Dissertação devidamente corrigida e defendida por Alexandre Carvalho de Alencar e aprovada pela Banca Examinadora.

Campinas, 20 de março de 2000.

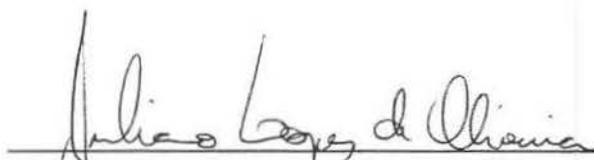


Profa. Dra. Claudia Bauzer Medeiros  
Instituto de Computação - Unicamp  
(Orientadora)

Dissertação apresentada ao Instituto de Computação, UNICAMP, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação.

## TERMO DE APROVAÇÃO

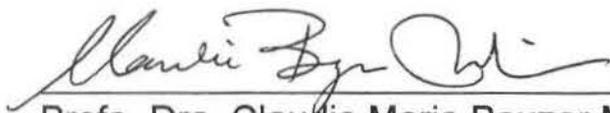
Tese defendida e aprovada em 20 de março de 2000, pela Banca Examinadora composta pelos Professores Doutores:



Prof. Dr. Juliano Lopes de Oliveira  
UFG



Prof. Dr. Mário Lúcio Cortes  
IC-UNICAMP



Profa. Dra. Claudia Maria Bauzer Medeiros  
IC-UNICAMP

“Digno de admiração é aquele que, tendo tropeçado ao dar o primeiro passo, levanta-se e segue em frente.”

Carlos Fox.

# Resumo

Um dos principais objetivos dos Sistemas de Informação Geográfica (SIG) é auxiliar tomadores de decisão no desempenho de suas tarefas para situações onde a dimensão espacial é relevante - por exemplo, no planejamento urbano ou ambiental. A qualidade das decisões, entretanto, é intimamente dependente da qualidade dos dados geográficos utilizados. Isto é em geral ignorado pelos tomadores de decisão, que não levam este tipo de fator em consideração, confiando na correta operação dos equipamentos de coleta ou do SIG em que são desenvolvidas as aplicações.

Esta dissertação tem por objetivo cobrir esta lacuna, apresentando uma análise dos vários fatores que afetam o tema qualidade de dados no contexto de aplicações geográficas. Esta análise engloba desde o processo de captura dos dados até a apresentação do resultado das aplicações e sua interpretação pelo usuário para sua tomada de decisão. Além da revisão bibliográfica realizada, as demais contribuições desta dissertação são a proposta de critérios básicos para a avaliação desta qualidade, e uma análise de como estes critérios podem ser obedecidos. Finalmente, parte desta proposta foi implementada em uma ferramenta acoplada a um SIG, que permite tornar explícita aos usuários a informação sobre a qualidade.

# Abstract

One of the main goals of Geographical Information Systems is to help decision makers in carrying out their tasks for situations where the spatial dimension is relevant - e.g., in urban or environmental planning activities. The quality of the decisions, however, is intimately dependent on the quality of the geographical data used. This is usually ignored by decision makers, who limit themselves to relying on the correct operation of the equipment used to collect data or on the GIS where the applications are developed.

The goal of this dissertation is to fill this gap, by presenting an analysis of the theme data quality in the context of geographic applications. This analysis ranges from the stage of data capture to the presentation of the result of the applications and the interpretation taken by the user for decision making. Besides an extensive bibliographic survey, other contributions of this work include the suggestion of a basic group of criteria to evaluate this quality, and an analysis of how these criteria can be met. Finally, part of these suggestions were implemented in a tool coupled to a GIS, which allows users to visualize data quality information.

# Agradecimentos

Agradeço à Deus por ter me dado força para enfrentar todos os óbices que surgiram ao longo do caminho.

À minha esposa Patrícia pelo amor, compreensão, apoio irrestrito, conselhos, paciência, por me fazer ver as possibilidades e não as limitações e por me fazer acreditar ser possível alcançar mais este objetivo.

Aos meus filhos Lucas e Leonardo agradeço o apoio e compreensão pelas horas de ausência.

Aos meus pais, Roberto e Zuleika, agradeço pelo eterno apoio e incentivo. Às minhas irmãs pelo carinho e pensamento positivo. Aos meus sogros pelo apoio prestado em mais esta fase de transição.

À Marinha do Brasil pela oportunidade.

À minha orientadora, Profa. Claudia Bauzer Medeiros, agradeço pelos inúmeros ensinamentos, pelas cobranças, pela paciência e pela consideração com que me distinguiu.

Agradeço o carinho dos inúmeros amigos que conquistei durante este tempo em que aqui estive. À família Kurka pelo apoio e amizade durante a fase final do trabalho. Um agradecimento especial ao meu companheiro de jornada Aldecir e família, agradeço muito à Deus por tê-los conhecido.

Aos companheiros do grupo de banco de dados, pela ajuda incondicional que recebi de todos sempre que precisei, pelas sugestões de melhoria e pelas críticas construtivas. Um agradecimento especial àqueles que mais diretamente colaboraram para tornar concreto este trabalho: Allysson, Andreia, Daniel Kaster, Fileto, Nathan, Sandro Gatti e Roberta.

Ao Instituto de Computação da Unicamp, seus professores e funcionários agradeço o apoio em nome da Marinha do Brasil.

Este trabalho foi desenvolvido com o auxílio parcial da CAPES e da Marinha do Brasil, e faz parte do Projeto SAI (Sistemas Avançados de Informação) do PRONEX II-MCT.

# Conteúdo

Resumo	vi
Abstract	vii
Agradecimentos	viii
<b>1 Introdução e motivação</b>	<b>1</b>
<b>2 Conceitos básicos</b>	<b>5</b>
2.1 Conceitos de Cartografia . . . . .	5
2.2 Sistema(s) de informação geográfica . . . . .	7
2.3 Qualidade . . . . .	14
2.4 Qualidade em SI e em SGBD . . . . .	15
2.5 Qualidade em sistemas de informação geográfica . . . . .	15
2.5.1 Características de nível macro . . . . .	17
2.5.2 Características de nível micro . . . . .	18
2.5.3 Características de uso . . . . .	21
2.6 Resumo . . . . .	21
<b>3 Qualidade e problemas relativos aos dados</b>	<b>22</b>
3.1 Características das fontes de dados e sensores . . . . .	22
3.1.1 Fontes de Dados espaciais . . . . .	22
3.1.2 Fontes de Dados convencionais . . . . .	27
3.2 Escolha das fontes de dados . . . . .	28
3.2.1 Utilização de fontes de dados em formato digital . . . . .	28
3.2.2 Produção de dados digitais a partir de dados em formato analógico	29
3.2.3 Realização de novo levantamento analógico ou digital . . . . .	30
3.3 Problemas na captura dos dados - coleta . . . . .	30
3.3.1 Parâmetros segundo Montgomery&Schuch [MS93] . . . . .	32
3.4 Problemas no pré-processamento dos dados - conversão . . . . .	35

3.4.1	Parâmetros segundo Montgomery&Schuch [MS93] e Hohl [Hoh98]	36
3.4.2	Controle e garantia de qualidade	39
3.5	Metadados	41
3.5.1	Definição e classificação de metadados	41
3.5.2	Padrões de metadados	43
3.5.3	Qualidade como parte de metadados	44
3.5.4	Produção de metadados geográficos	46
3.5.5	Qualidade dos metadados	48
3.6	CrITÉrios de aceitaÇão de dados	49
3.6.1	MediÇão da qualidade	49
3.6.2	Registro dos processos de controle/garantia de qualidade	50
3.6.3	DocumentaÇão dos requisitos de qualidade	50
3.7	Resumo	51
<b>4</b>	<b>Qualidade e problemas relativos à apresentaÇão</b>	<b>52</b>
4.1	VisualizaÇão de dados geográficos	52
4.1.1	Escalas	53
4.1.2	ProjeÇões	56
4.1.3	SÍmbolos	56
4.2	Projeto de interfaces	58
4.2.1	Fatores humanos	60
4.2.2	CogniÇão espacial em SIG	61
4.3	Resumo	65
<b>5</b>	<b>Aspectos de implementaÇão</b>	<b>68</b>
5.1	AvaliaÇão de qualidade sob múltiplos critÉrios	68
5.2	Uso de metadados	70
5.3	Sistemas Espaciais de Apoio à DecisãO e WOODSS	73
5.4	Introduzindo qualidade no WOODSS	76
5.4.1	Arquitetura do protÓtipo	76
5.4.2	Esquema do banco de dados	80
5.5	Exemplo de uso	82
5.6	Resumo	83
<b>6</b>	<b>Conclusões e extensões</b>	<b>93</b>
6.1	ContribuiÇões	93
6.2	Extensões	94
<b>Bibliografia</b>		<b>97</b>

# Lista de Tabelas

1.1	Possíveis fontes de erro encontradas no uso de SIG segundo Aronoff [Aro95].	3
2.1	Características de qualidade dos dados geográficos segundo Aronoff [Aro95].	17
3.1	Elipsóides de referência segundo Bakker [Bak65]. . . . .	23
3.2	Parâmetros de qualidade utilizados para avaliação da qualidade da captura dos dados segundo Montgomery & Schuch [MS93] e Hohl [Hoh98]. . . . .	31
3.3	Tabela comparativa entre critérios e parâmetros de qualidade identificados e dimensões normalmente informadas em metadados. . . . .	42
5.1	Indicadores de qualidade presentes em padrões de metadados ou propostos por organizações. . . . .	71
5.2	Indicadores de qualidade propostos por alguns autores. . . . .	72
5.3	Indicadores de qualidade presentes no IDRISI [Eas97]. . . . .	72
5.4	Descrição, por tipo, dos atributos dos metadados propostos. . . . .	84
5.5	Continuação da descrição, por tipo, dos atributos dos metadados propostos.	85

# Lista de Figuras

2.1	Sistemas de informação geográfica de acordo com Maguire [Mag91]. . . . .	8
2.2	Arquitetura de sistemas de informação geográfica de acordo com Câmara et al. [CCH+96]. . . . .	9
2.3	Avaliação de precisão posicional usando média e desvio padrão segundo Aronoff [Aro95]. . . . .	20
3.1	Estágios da confecção de mapas segundo Monmonier [Mon91]. . . . .	24
3.2	Informações de qualidade de dados do padrão FGDC [FGD98b]. . . . .	45
4.1	Algumas operações de generalização comuns a polígonos, linhas e pontos segundo Monmonier [Mon91]. . . . .	55
4.2	As seis principais variáveis visuais segundo Monmonier [Mon91]. . . . .	57
4.3	Enfoques no projeto de interfaces segundo Oiveira [Oli97]. . . . .	60
4.4	Ilustração da taxionomia de Pierce [Pie90] utilizando “signos” encontrados em mapas e em cartas náuticas. . . . .	64
4.5	Exemplo de uma possível ordem de precedência entre diferentes classes de aspectos por Barkowsky & Freska [BF97]. . . . .	65
4.6	Possíveis hierarquias de precisão em relação ao aspecto “orientação” por Barkowsky & Freska [BF97]. . . . .	66
4.7	Fragmento de um mapa do metrô de Hamburgo [BF97]. . . . .	67
5.1	O modelo de qualidade de software de acordo com a norma ISO 9126 [BT99, Jun99]. . . . .	69
5.2	O modelo de qualidade de software de acordo com Blin & Tsoukiàs [BT99].	70
5.3	Arquitetura do WOODSS, de acordo com Seffino [Sef98]. . . . .	75
5.4	As relações Atividades, Arquivos e Dependências e o Registro de Metadados do WOODSS. . . . .	77
5.5	Extensão ao Modelo Entidade-Relacionamento do WOODSS. . . . .	78
5.6	Tipos de metadados de qualidade propostos e seus atributos. . . . .	79
5.7	Arquitetura modificada do WOODSS. . . . .	80

5.8	Uma proposta de modelo para qualidade dos dados de entrada usados em uma aplicação geográfica . . . . .	86
5.9	Visualização do processo de geração de um mapa ( <i>workflow</i> ) no WOODSS [Sef98]. . . . .	87
5.10	Cópia da tela de visualização de metadados de dados de entrada. . . . .	88
5.11	Cópia da tela de inserção de metadados adicionais aos dados de entrada. . . . .	89
5.12	Cópia das telas de visualização de metadados do processo e da apresentação. . . . .	90
5.13	Cópia da tela de inserção de metadados adicionais do processo. . . . .	91
5.14	Cópia da tela de inserção de metadados adicionais da apresentação. . . . .	92

# Capítulo 1

## Introdução e motivação

A busca por qualidade tem sido foco de estudo em diversas áreas do conhecimento humano. Muitas vezes confundida com padronização, qualidade é muito mais que isso. Descrições genéricas de qualidade como *traços ou características importantes* normalmente induzem a erro, pois deixam o requisito de avaliação implícito: importantes para que utilização ou sob que pontos de vista? Para [TB98], segurança de dados e qualidade têm a mesma importância. No entanto, é mais fácil especificar níveis de segurança de forma objetiva do que definir qualidade. Daí um maior desenvolvimento da primeira área em comparação à segunda.

O objetivo principal desta dissertação é efetuar um levantamento bibliográfico das diversas abordagens sobre o tema qualidade no contexto de Sistemas de Informação Geográfica - SIG, em especial considerando o uso de aplicações que se utilizam de dados geográficos, e sugerir o uso de uma ferramenta de apoio para auxiliar o usuário destas aplicações a conhecer a qualidade dos dados que estão sendo utilizados.

SIG podem ser definidos como sistemas automatizados para armazenar e manipular *dados geográficos*. Este termo denota dados que representam objetos e fenômenos em que a localização geográfica é uma característica inerente à informação e indispensável para analisá-la [CCH<sup>+</sup>96]. O trabalho a ser desenvolvido nesta dissertação pretende apontar algumas direções em que a noção de qualidade possa ser melhor especificada e, conseqüentemente, avaliada.

Na área de Sistemas de Informações Geográficas, um conceito bastante difundido é o proposto por Chrisman [Chr84] que considera qualidade como "*fitness for use*". Este conceito assegura que qualidade só pode ser conhecida em termos de um propósito, ou seja, qualidade é um conceito que deve ser avaliado dentro de um contexto específico. É fácil observar que, pelo fato do conceito de "*fitness for use*" ser dependente da intenção de uso do dado, não é possível obter uma descrição genérica de qualidade para utilização em SIG.

O aumento em número de fontes e volume de dados disponíveis e o crescimento exponencial de novos sistemas e aplicações tornou crescente também, a preocupação com a qualidade dos dados geográficos. Considerações explícitas sobre a qualidade de dados geográficos oferecem, inclusive, novas perspectivas do ponto de vista de *interoperabilidade*, pois permitem quantificar determinados aspectos que, por sua vez, indicarão quando é adequado intercambiar dados e quais dados podem ser reutilizados [MA99]. A interoperabilidade em SIG tem sido objeto de discussão recente [NCG99] e envolve, dentre outras abordagens, a preocupação com a reutilização e intercâmbio de dados embora a qualidade dos dados manipulados seja, normalmente, relegada a um segundo plano.

Embora não existam padrões de qualidade existe consenso sobre alguns de seus parâmetros. Padrões para SIG usualmente incluem padrões para linguagem de especificação, transferência de dados, geocodificação e documentação de metadados e formatos. Para [CCH+96], em geral, a utilização de padrões facilita o compartilhamento, a integração e a transferência de dados. Já para [Har98] os padrões são importantes para a definição de dados precisos, descrição de metadados e para assegurar consistência entre diferentes plataformas, mas não garantem qualidade.

Além da dificuldade de se definir qualidade há um outro problema que precisa ser considerado: o fato de que o conceito de qualidade pode ser abordado sob diferentes aspectos e aplicado em diferentes fases do processo de implantação de aplicações em um SIG. Segundo [CCH+96], este processo de implantação pode ser dividido em três fases distintas, a modelagem do mundo real, a criação do banco de dados geográfico e a operação. A *modelagem do mundo real* engloba a modelagem de processos e de dados e consiste em selecionar fenômenos e entidades de interesse, abstraindo-os e generalizando-os. A *criação do banco de dados geográfico* pode ser considerada uma das fases mais importantes, exigindo várias etapas, inclusive a coleta dos dados relativos aos fenômenos de interesse identificados na modelagem, a correção dos dados coletados e finalmente o seu georeferenciamento. A fase de *operação* refere-se tanto ao uso do SIG em si, quanto ao desenvolvimento de aplicações específicas por parte dos usuários a partir dos dados armazenados.

Além da idéia de que qualidade depende do uso pretendido, está ligada ao atendimento a padrões e pode ser aplicada a diferentes fases do processo de implantação de um SIG, é preciso, também, para se avaliar qualidade, considerar os erros que são introduzidos a cada passo do processo de geração e uso da informação geográfica, desde a modelagem e coleta dos dados fonte até a interpretação dos resultados de uma aplicação. Apesar disso, apenas alguns poucos autores, segundo [BM98], têm se preocupado com os erros inerentes aos dados geográficos e com a forma como eles se propagam. A maioria dos estudos sobre erros e sobre qualidade ainda encontra-se no nível de pesquisa.

Alguns dos erros mais comuns podem ser vistos na Tabela 1.1 a seguir, adaptada

de [Aro95]. Os estágios de coleta, entrada e armazenamento tratados na tabela, dizem respeito à fase de criação do banco de dados geográfico. Os estágios de manipulação e saída, referem-se à fase de operação do SIG. Erros de modelagem não estão incluídos nesta tabela e o estágio de uso dos dados não faz parte da abordagem de [CCH+96].

Estágio	Fontes de Erro
Coleta	erros na coleta dos dados de campo erros nos mapas usados como fontes de dados erros na análise dos dados de sensoriamento remoto
Entrada (Conversão)	imprecisões na digitalização causadas pelo operador e equipamento imprecisões inerentes à entidade geográfica
Armazenamento	precisão numérica e espacial insuficientes
Manipulação (Processamento)	intervalos de classes não apropriados erros de fronteira de polígonos propagação de erros pela utilização de múltiplos <i>overlays</i> erros causados por <i>overlays</i> de polígonos
Saída (Apresentação)	imprecisões de escala erros causados pela imprecisão inerente aos dispositivos de saída erros causados pela instabilidade do meio
Uso dos Resultados	a informação pode ser incorretamente entendida a informação pode ser utilizada de forma não apropriada

Tabela 1.1: Possíveis fontes de erro encontradas no uso de SIG segundo Aronoff [Aro95].

Na coleta, erros podem ser próprios do material fonte utilizado para entrada no banco de dados ou resultantes de imprecisões de equipamentos e/ou uso de procedimentos incorretos. A interpretação de fotografias aéreas e imagens de satélite, por exemplo, pode introduzir erros de classificação e delineamento de fronteiras de polígonos. No estágio de saída de dados os erros podem ser introduzidos na plotagem dos mapas pelo dispositivo de saída ou pelo encolhimento e esticamento da mídia de saída.

Para avaliar qualidade no contexto de SIG, portanto, é necessário considerar toda a seqüência de processos pela qual os dados passam. Assim, é preciso identificar as fontes de erro durante toda esta seqüência: seleção das fontes de dados, coleta, conversão, armazenamento e se possível estender essa análise para o processamento de aplicações, apresentação dos resultados e interpretação destes resultados pelo usuário.

Tomando por referência a Tabela 1.1, a dissertação pretende abordar, então, aspectos referentes à coleta, entrada e saída dos dados geográficos. Em relação à coleta e entrada, serão abordados diferentes aspectos das possíveis fontes de dados e equipamentos utilizados para captura e o processo de tratamento dos dados até a sua inserção no banco de dados. Sobre a saída serão discutidos alguns aspectos cartográficos e do projeto de interface de SIG, essenciais à apresentação dos dados geográficos. Não fazem parte do escopo desta dissertação considerações sobre a modelagem dos dados, seu processamento

ou o uso dos resultados. Apesar de concentrar-se nestes estágios, algumas considerações serão feitas sobre o processamento de dados (uma vez que a ferramenta de apoio proposta acompanha a interação do usuário com o SIG) e sobre o uso de resultados (abordando aspectos da cognição espacial humana).

As principais contribuições são:

1. levantamento dos diversos níveis em que a qualidade deve ser considerada no contexto de dados e aplicações geográficas;
2. definição de um conjunto básico de critérios de avaliação desta qualidade;
3. proposta de como alguns destes critérios podem ser obedecidos e como esta informação pode ser explicitada ao usuário; e
4. implementação de uma ferramenta que permita aplicar estes conceitos, validando o estudo realizado.

O restante do texto está organizado da seguinte forma. O Capítulo 2 apresenta conceitos básicos necessários ao entendimento do resto do texto, a saber, noções de Cartografia, SIG e qualidade. Os Capítulos 3 e 4 analisam, respectivamente, fatores que influenciam a qualidade no processamento de dados para a entrada e apresentação. O Capítulo 5 apresenta o protótipo de uma ferramenta de apoio para auxiliar a incorporar alguns conceitos de qualidade a aplicações específicas em um SIG e o Capítulo 6 indica as conclusões e propõe algumas extensões à dissertação.

# Capítulo 2

## Conceitos básicos

Um SIG é, antes de tudo, um Sistema de Informação e portanto está sujeito às mesmas considerações aplicadas a sistemas não geográficos, acrescidas da necessidade de se considerar as características peculiares de dados geográficos.

Alguns termos da literatura de SIG têm significados distintos para diferentes autores. Este capítulo apresenta os conceitos utilizados neste trabalho relativos a Cartografia, SIG e qualidade.

### 2.1 Conceitos de Cartografia

A ciência que modernamente trata da construção de cartas é a Cartografia, definida como a arte, ciência e tecnologia de construção de mapas, juntamente com seu estudo científico e trabalho artístico [SE90].

#### Mapas e cartas

Os conceitos de mapas e cartas não possuem uma distinção rígida. Apesar disso algumas definições consideram que um mapa não possui um caráter científico especializado, sendo destinado a fins culturais, ilustrativos ou comerciais. Cartas por sua vez são normalmente interpretadas como representações dos aspectos da Terra destinadas a fins práticos da necessidade humana, permitindo a avaliação precisa de distâncias, direções e localização geográfica de pontos e áreas [Oli88]. Para [Oli95] carta é uma representação similar a mapas, porém, de caráter mais especializado.

De acordo com [Bak75] as cartas podem ser classificadas em: geográficas (topográficas e planimétricas), cadastrais, aeronáuticas, náuticas e especiais (geológicas, geomorfológicas, meteorológicas, de solos, de vegetação, de uso da terra, geofísicas e globos). Neste texto, os termos mapa e carta serão usados indistintamente adotando a definição de caráter

científico.

## Projeções

Todos os mapas são representações aproximadas da superfície terrestre, que projetam cada ponto do globo terrestre em uma superfície plana. Para se obter essa correspondência utilizam-se os sistemas de projeções cartográficas.

Há um grande número de projeções cartográficas, uma vez que há uma variedade de modos de projetar em um plano os objetos geográficos que caracterizam a superfície terrestre. No entanto, é impossível fazer-se uma cópia plana da superfície terrestre sem deformação. O Capítulo 4 apresenta maiores considerações sobre as projeções cartográficas.

## Sistemas de coordenadas

Um objeto geográfico qualquer somente poderá ser localizado, se puder ser descrito em relação a outros objetos cujas posições sejam previamente conhecidas, ou se tiver a localização determinada em um sistema de coordenadas. Os sistemas de coordenadas dividem-se em dois grandes grupos: o de coordenadas geográficas ou terrestres e o de coordenadas planas ou cartesianas.

No sistema de coordenadas geográficas ou terrestres, historicamente mais antigo, cada ponto da superfície terrestre é localizado na interseção de um meridiano com um paralelo. Meridianos são círculos máximos da esfera cujos planos contêm o eixo dos pólos. Paralelos são círculos da esfera cujos planos são perpendiculares ao eixo dos pólos. O meridiano de origem é o que passa pelo antigo observatório britânico de Greenwich e o paralelo de origem é o Equador. A representação de um ponto na superfície terrestre é feita por um valor de latitude ( $\varphi$ ) e longitude ( $\lambda$ ). Latitude é a medida do arco de meridiano entre o paralelo que passa pelo ponto e o Equador. Longitude é a medida do arco de paralelo entre o meridiano que passa pelo ponto e o Meridiano de Greenwich. Pontos que não correspondem ao nível médio dos oceanos podem ter também a altitude como terceiro parâmetro.

O sistema de coordenadas planas ou cartesianas baseia-se na escolha de dois eixos perpendiculares, normalmente denominados eixos horizontal e vertical, cuja interseção é denominada origem e é estabelecida como base para a localização de qualquer ponto do plano. Nesse sistema, um ponto é definido por dois números: um correspondente à projeção sobre o eixo horizontal (eixo  $x$ ), que é associado à longitude, e outro correspondente à projeção sobre o eixo vertical (eixo  $y$ ) que é associado à latitude. Estas coordenadas são associadas matematicamente às coordenadas geográficas de modo que umas podem ser convertidas nas outras.

## Escalas

Escala é a relação entre as dimensões dos elementos representados em um mapa ou carta e a grandeza correspondente, medida sobre a superfície da Terra.

A maioria dos mapas é, obviamente, menor que a realidade que representa, e as escalas nos dão a informação de quão menores eles são. As escalas são representadas usualmente sob a forma de razão (1:10000) ou fração ( $1/10000$ ) mas ambas as formas têm o mesmo significado. Neste caso, 1 unidade de medida no mapa representa 10000 unidades de medida na superfície terrestre. Os termos escala grande e escala pequena normalmente levam a erros mas é fácil entender que como se trata de frações, uma escala é grande quando possui um denominador pequeno na fração e pequena quando ocorre o contrário. Assim, uma escala de 1:10000 é obviamente maior que uma escala 1:50000 [Oli88].

Em geral são considerados mapas de escalas grandes os que possuem escalas de 1:24000 ou maiores e de escalas pequenas os que possuem escala de 1:50000 ou menores [Mon91]. Esta classificação não é rígida e dependendo da aplicação há grande variação das escalas típicas de uso.

## 2.2 Sistema(s) de informação geográfica

Há várias definições para SIG. Classificações baseadas em funcionalidade são as mais populares, mas há outras propostas em função de custo, tamanho, plataforma, aplicação e modelo de dados. Uma razão fundamental para esta dificuldade de definição é o debate acadêmico sobre o foco central das atividades de SIG. Alguns pesquisadores acreditam que o foco principal deve ser o hardware/software enquanto outros acreditam que o elemento chave deve ser o processamento de informações ou mesmo as aplicações.

A Figura 2.1, retirada de [Mag91], apresenta uma maneira de se caracterizar SIG: como resultado da interseção de outros sistemas, de cartografia computacional, de gerenciamento de bancos de dados, de projeto auxiliado por computador e de sensoriamento remoto.

Sistemas de Cartografia Computacional mantêm o foco na recuperação, classificação e representação simbólica automática dos dados geográficos. Utilizam estruturas de dados simples e com pouca informação topológica. Normalmente possuem muitas facilidades para projeto de mapas e produção com qualidade no formato vetorial.

SGBD são otimizados para armazenar e recuperar atributos de dados não-gráficos. Têm capacidade limitada para recuperação e apresentação gráfica e para implementação de operações analíticas espaciais.

Sistemas de CAD são desenvolvidos para projetar e desenhar objetos. Possuem *links* rudimentares a bancos de dados, usam relacionamentos topológicos simples, lidam basi-

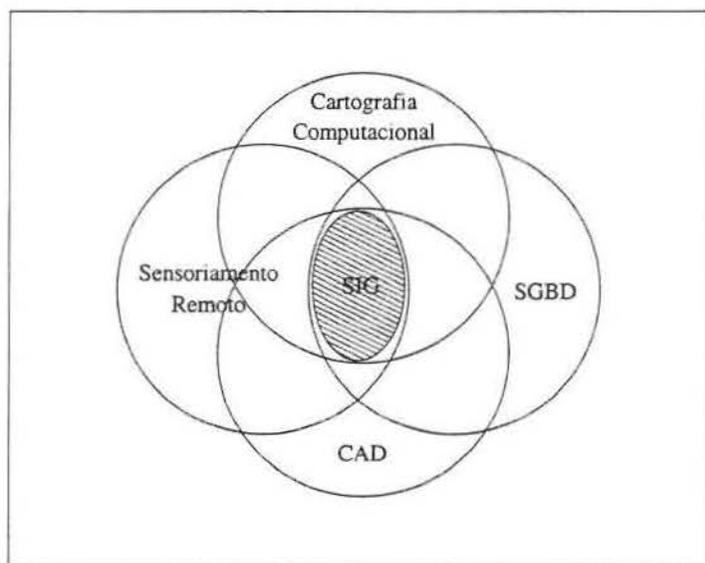


Figura 2.1: Sistemas de informação geográfica de acordo com Maguire [Mag91].

camente com pequenas quantidades de dados, e possuem capacidade analítica limitada.

Sistemas de Sensoriamento Remoto são projetados para coletar, armazenar, manipular e apresentar dados em formato matricial (“*raster*”) tipicamente derivados de *scanners* montados a bordo de aeronaves ou plataformas de satélites, apesar de poderem manipular dados vetoriais. Possuem capacidade bastante limitada para o manuseio de dados não gráficos e *links* para SGBD muito limitados.

SIG assumem algumas características de cada um destes sistemas anteriormente citados. As diversas definições de SIG existentes refletem, cada uma à sua maneira, a multiplicidade de usos e visões possíveis desta tecnologia e apontam para uma perspectiva de utilização interdisciplinar. Esta dissertação adota a definição de Câmara et al. [CCH<sup>+</sup>96]: SIG são sistemas automatizados usados para analisar, armazenar e manipular dados geográficos armazenados em um banco de dados geográfico.

De uma forma abrangente, pode-se considerar que um SIG possui os seguintes componentes principais: interface com o usuário; entrada e integração dos dados; funções de processamento; visualização e plotagem; e armazenamento e recuperação de dados [CCH<sup>+</sup>96]. As várias idéias sobre SIG podem ser sintetizadas e apresentadas sob a forma de três visões que segundo [Mag91] podem ser chamadas de:

- visão de banco de dados - que enfatiza o estudo de aspectos do gerenciamento dos dados;
- visão de análise espacial - que enfoca o processamento dos dados; e
- visão de mapa - que enfatiza o estudo de aspectos de apresentação dos dados.

A Figura 2.2, adaptada de [CCH<sup>+</sup>96] indica o relacionamento entre os componentes de um SIG e mostra quais componentes são englobados por que visões, conforme a classificação de Maguire. A visão de banco de dados cuida da entrada, integração, armazenamento e recuperação dos dados; a de análise espacial das funções de processamento; e a de mapa da interface, visualização e plotagem.

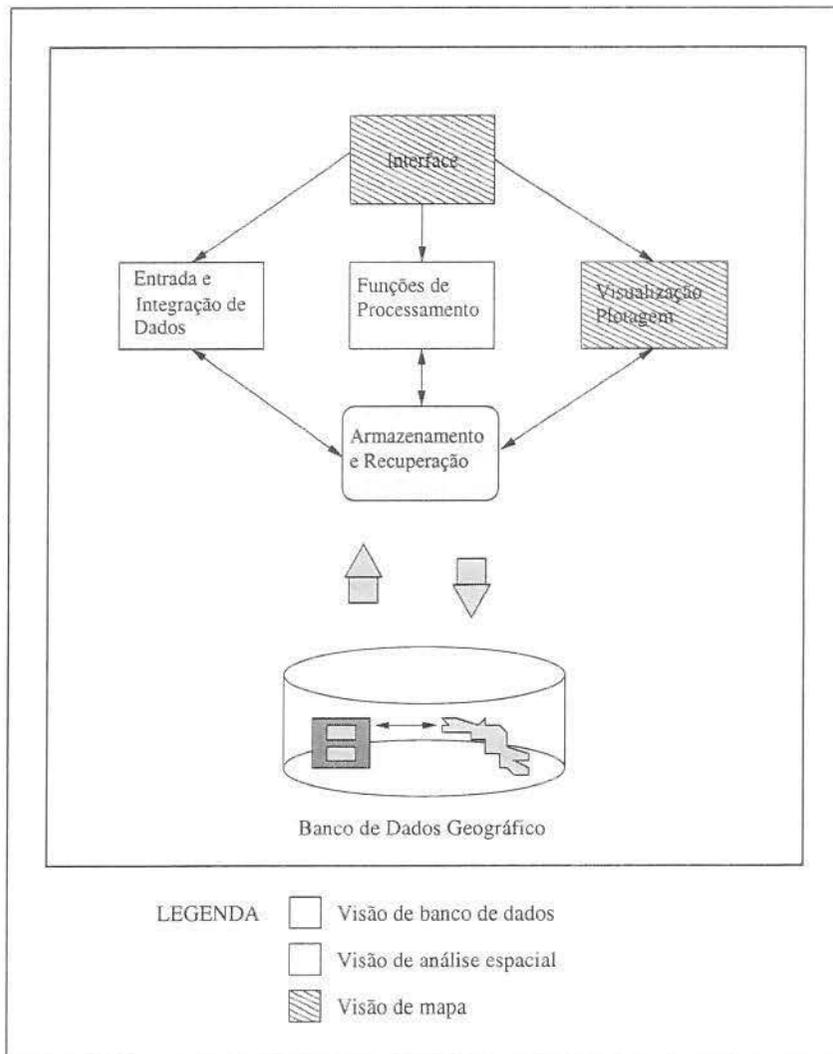


Figura 2.2: Arquitetura de sistemas de informação geográfica de acordo com Câmara et al. [CCH<sup>+</sup>96].

A dissertação está centrada na visão de bancos de dados e na visão de mapa, enfocando assim, aspectos referentes aos dados (Capítulo 3) e à sua apresentação (Capítulo 4). A análise espacial não faz parte do escopo desta dissertação por considerarmos o processamento dos dados geográficos bastante dependente da aplicação. Apesar disso, algumas

considerações sobre o processamento dos dados geográficos são feitas no Capítulo 5, que descreve o protótipo implementado.

## Dados geográficos

O termo *dado espacial* denota qualquer tipo de dado ao qual está associada alguma dimensão espacial. Os dados utilizados em SIG pertencem a uma classe particular de dados espaciais: os dados *georeferenciados* ou dados geográficos. Ambos os termos são encontrados na literatura e denotam dados que descrevem fatos, objetos e fenômenos do globo terrestre associados à sua localização sobre a superfície da Terra, num certo instante ou período de tempo [CCH<sup>+</sup>96]. No restante do texto será adotada apenas a nomenclatura *dados geográficos* para fim de padronização.

Em 1978, Sinton apresentou uma análise fundamental da informação contida em SIG [Sin78]. Os três componentes básicos propostos por ele: posição, tema e tempo, foram largamente aceitos na cartografia temática [Chr91]. Depois de Sinton, outras análises se seguiram, mas quase sempre mantendo esta idéia inicial. Em estudos mais recentes, os dados geográficos são caracterizados a partir de três ou quatro componentes fundamentais, dependendo do autor: posição, tema e tempo [Chr91, Aal96]; espaciais, não espaciais e temporais [CCH<sup>+</sup>96]; ou posição geográfica (coordenadas), atributo (valores de dados), relacionamentos topológicos e componentes temporais [MP94].

A maioria das fontes de informação espacial fixa um dos três componentes fundamentais citados para poder controlar e medir os outros dois. Normalmente o tempo é o componente que é mantido fixo para que os outros dois, posição e tema, possam ser medidos e controlados [Chr91]. Apesar de antiga, a análise fundamental proposta por Sinton ainda permanece como base para trabalhos bem mais recentes sobre qualidade de dados como os de [FB94] e [Aal96].

## Aplicações geográficas

Uma aplicação geográfica é entendida como um conjunto de processos, executados numa seqüência definida, segundo um modelo preestabelecido e sobre uma base de dados geográficos. Alguns desses processos podem ser automatizados (por exemplo, sobreposição de mapas), enquanto em outros a automatização nem sempre é cabível (por exemplo, definição de estratégias) [Pir97].

[CCH<sup>+</sup>96] classifica as aplicações geográficas em três grandes categorias: aplicações sócioeconômicas (manipulação e análise de dados espaciais sobre a terra, o povo e infraestrutura); aplicações ambientais (enfocando o meio ambiente) e aplicações de gerenciamento sobre as duas categorias anteriores. Uma outra divisão canônica (usualmente adotada por fornecedores de SIG) está baseada na escala geográfica tratada pela apli-

cação. [Cif95] classifica as aplicações de SIG em aplicações ambientais/de recursos naturais e urbanas/rurais, baseado fundamentalmente nas visões adotadas (campos ou objetos, definidos a seguir). Nas aplicações ambientais e de recursos naturais predomina a visão de campos enquanto que nas urbanas e rurais predomina a visão de objetos. Estas classificações não são, de forma alguma, absolutas.

### Modelagem de campos e objetos geográficos

No contexto de aplicações de SIG, o mundo real é freqüentemente modelado segundo duas percepções do mundo: a visão de campos e a visão de objetos [MP94, CCH<sup>+</sup>96]. A visão de campos enxerga o mundo como uma superfície contínua, sobre a qual os fenômenos geográficos variam segundo diferentes distribuições. A visão de objetos, por sua vez, representa o mundo como uma superfície ocupada por objetos identificáveis, com geometria e características próprias.

A questão da identidade é fundamental para distinguir os campos dos objetos: podem existir inúmeras áreas no Brasil classificadas como “vegetação arbustiva” (valor do campo) mas certamente só haverá uma “Unicamp” (objeto identificável). Nenhuma das duas visões - campos e objetos - é ideal, e em geral, elas se complementam. Cada visão está, normalmente, associada a uma área de aplicação e a necessidades específicas do usuário.

### Representação de campos e objetos geográficos

A dicotomia de modelagem se reflete no nível de representação, levando a dois formatos básicos, o formato matricial e o formato vetorial. Campos são freqüentemente representados no primeiro formato (normalmente referenciado como “*raster*”) e objetos são tipicamente representados no formato vetorial (“*vector*”). O formato *raster* corresponde a uma matriz cujos elementos são unidades poligonais do espaço (células). No formato *vector* os objetos identificáveis têm sua geometria descrita utilizando-se pontos, linhas e áreas.

A representação de um objeto geográfico não determina completamente sua aparência visual, ou seja, a forma como o objeto será apresentado ao usuário, na tela ou em outro meio. A cada representação podem corresponder uma ou mais apresentações, de acordo com as necessidades da aplicação [JL99].

### Classes de dados geográficos

Alguns tipos de classes de dados geográficos merecem destaque: os mapas temáticos e mapas cadastrais; as redes; os modelos numéricos do terreno e as imagens.

Os mapas temáticos são mapas que mostram uma região geográfica particionada segundo valores relativos a um tema (por exemplo, uso do solo), sendo resultantes de funções

de análise e classificação de dados. Os mapas cadastrais, ao contrário, apresentam objetos identificáveis (por exemplo, loteamento urbano).

O conceito de redes representa informações associadas a serviços de utilidade pública (água, luz, telefone); bacias hidrográficas e estradas. A característica principal é que as informações sobre os recursos estão normalmente armazenadas sob a forma de grafos.

O termo modelo numérico de terreno (MNT) é utilizado para indicar a representação de uma grandeza que varia continuamente no espaço. Esta classe está usualmente associada à altimetria para geração de mapas topográficos.

As imagens, antes utilizadas apenas como um complemento aos mapas, passaram a ser usadas em conjunto ou até mesmo em substituição àqueles. Isto ocorreu, principalmente, devido ao desenvolvimento ocorrido nas áreas de Sensoriamento Remoto e Processamento Digital de Imagens.

### **Operações sobre dados geográficos**

Devido à largura do espectro de domínios de aplicação e à variedade de usuários, muito pouco tem sido feito para identificar e formalizar as necessidades funcionais chave para SIG, em termos de operadores espaciais básicos e relacionamentos suportados.

A ausência de um conjunto básico de operadores e relacionamentos espaciais é refletida, por sua vez, na diversidade de linguagens de consulta em SIG. A definição deste conjunto é uma questão ainda em aberto. De acordo com [MP94] todas as funções necessárias podem ser classificadas de modo geral em: captura, transferência, validação e edição; armazenamento e estrutura; reestruturação, generalização e transformação; e consulta e análise.

### **Processamento de transações**

Em aplicações convencionais de bancos de dados, um usuário modifica os dados através de uma seqüência de operações elementares que deve ser executada como um todo. Os usuários de aplicações geográficas também executam transações desta natureza ao modificar os atributos não geográficos. Na maioria das vezes, entretanto, as transações envolvendo dados geográficos apresentam um padrão de comportamento mais próximo do encontrado em ambientes de desenvolvimento de software, projetos de circuitos VLSI e criação de documentos hipermídia, dentre outros [CCH<sup>+</sup>96, Cur97].

### **Recuperação de dados geográficos**

Segundo [MP94], SIG usualmente suportam dois tipos de mecanismos de interação: linguagens de consulta textuais e manipulação interativa de elementos geográficos.

A maior parte das linguagens de consulta textuais para bancos de dados geográficos baseia-se em extensões da linguagem SQL. Paradoxalmente, no que se refere a SIG, a principal vantagem e fraqueza deste enfoque origina-se do fato de SQL ser baseada no modelo relacional. Se isto por um lado traz vantagens como a facilidade de otimização de consultas, por outro lado, tem sido notado que a abordagem relacional não é completamente apropriada para a modelagem de dados em aplicações geográficas [CCH<sup>+</sup>96].

No grupo de manipulação interativa de elementos situam-se as linguagens visuais e as de consulta sobre imagens. A maioria das linguagens visuais é baseada em visualização gráfica, freqüentemente incorporando recursos de hipermídia, enquanto as linguagens de consulta sobre imagens dividem-se entre dois enfoques. No primeiro, algumas características principais da imagem são extraídas manualmente e armazenadas textualmente em um banco de dados, sendo a recuperação processada através de consultas sobre tais dados textuais. O segundo enfoque é baseado em técnicas de processamento de imagem, que realizam extração de características e determinação de objetos de interesse [CCH<sup>+</sup>96].

## **Apresentação de dados geográficos**

Uma característica importante de SIG é que o resultado de uma consulta a um banco de dados geográfico é, via de regra, apresentado sob a forma de um mapa. A apresentação de dados geográficos é, pois, um aspecto fundamental da interface de um SIG. Para [CCH<sup>+</sup>96] uma interface para SIG deve oferecer facilidades para visualização de campos e objetos geográficos nos espaços 2D e 3D, incluindo comandos para controlar padrões de preenchimento. Deve, ainda, prover formas de combinar os resultados de várias consultas e de construir legendas que ajudem o usuário a examinar o conteúdo apresentado. Para isso, deve considerar ao menos três níveis de visualização: de metadados, dos componentes usados para construir consultas e do resultado propriamente dito, com múltiplas resoluções (vide Subseção 2.5.2 a seguir).

A visualização de metadados propõe-se a facilitar a construção de consultas e permitir estabelecer correlações de dados em um nível mais abstrato. Nos SIG atuais, os metadados, em geral, são limitados ao esquema do banco de dados geográfico. Para [Har98], o grande salto para a melhoria de qualidade em dados geográficos está em tornar os metadados mais acessíveis. Para ele, metadados importantes são freqüentemente subutilizados porque os usuários não podem facilmente encontrá-los, entendê-los e navegar por suas complexas estruturas.

## 2.3 Qualidade

Qualidade é um termo bastante subjetivo, podendo ter inúmeras definições. De acordo com o Novo Dicionário Aurélio da Língua Portuguesa [Fer86], só para citar alguns exemplos, qualidade pode ser entendida como: “*a propriedade, atributo ou condição capaz de determinar a natureza de coisas ou pessoas e de lhes distinguir umas das outras ou como a propriedade que, em uma escala de valores, permite avaliar e, conseqüentemente, aprovar, aceitar ou recusar, alguma coisa*”.

Outras definições, comuns na área de gerência de qualidade da fabricação de produtos ou prestação de serviços incluem, por exemplo, as que se seguem:

- Qualidade é conformidade com especificações [Cro79].
- Qualidade é o nível de satisfação do cliente [Dem86].
- Qualidade é adequação ao uso [Jur88].

De acordo com a atual norma brasileira que trata do assunto (NBR ISO 8402), qualidade é “*a totalidade das características de uma entidade que lhe confere a capacidade de satisfazer às necessidades explícitas e implícitas*”. Por entidade devemos entender o produto do qual se fala, que pode ser um bem ou serviço. As necessidades explícitas devem ser entendidas como as condições e objetivos explicitados pelo usuário, enquanto as necessidades implícitas referem-se aos seus anseios [Jun99].

A partir destas definições, pode-se afirmar que o foco da abordagem sobre qualidade evoluiu a partir de uma preocupação inicial apenas com o produto para uma outra abordagem onde cada vez mais consideram-se o consumidor e a utilização que este fará do produto.

Inicialmente, a qualidade estava associada a um controle final do produto para evitar que produtos defeituosos chegassem às mãos dos consumidores. Com o advento da produção em massa, para atender a mercados em crescimento, foram sendo introduzidas técnicas de controle estatístico da qualidade. Após a Segunda Guerra Mundial, com o desenvolvimento da indústria aeronáutica, começou a ser adotado o controle de processo, englobando desde o projeto até o acabamento. Conseguir qualidade nos processos objetivava, principalmente, aumentar segurança e eliminar defeitos. Este conceito evoluiu para a garantia de qualidade, que consiste na sistematização, por intermédio de normas escritas, dos padrões e requisitos para cada etapa do processo produtivo, com o objetivo de garantir qualidade uniforme para todos os produtos e serviços. A partir da década de 60, com a introdução de mudanças no ambiente de negócios, saturação dos mercados e necessidade de criação de novos bens e serviços para satisfazer à demanda do consumidor, o conceito de qualidade evoluiu para uma idéia de satisfação do cliente e competitividade [SEB94b].

## 2.4 Qualidade em SI e em SGBD

O principal papel de um sistema de informação (SI) é apresentar visões do mundo real de forma que as pessoas em uma organização possam criar produtos e tomar decisões. Estas visões são construídas a partir dos dados gerenciados pelo sistema.

A preocupação com a qualidade de dados visa assegurar que os dados sejam suficientemente precisos, atuais o suficiente e consistentes para que a organização possa tomar decisões razoáveis [Orr98]. Segundo [Orr98], algumas regras gerais para a qualidade de dados podem ser deduzidas:

- dados que não são utilizados não permanecem corretos por muito tempo;
- a qualidade dos dados não poderá ser, em última análise, melhor que seu uso mais rigoroso;
- problemas com a qualidade dos dados tendem a piorar com a idade dos sistemas;
- quanto menor for a frequência de mudanças de um dado atributo, mais impacto ela terá quando ocorrer; e
- as regras de qualidade de dados aplicam-se igualmente a metadados.

No caso de Sistemas de Gerenciamento de Banco de Dados as maiores preocupações em relação aos dados são evitar a redundância e garantir a integridade. A redundância promove altos custos de armazenamento e acesso e pode levar à inconsistência dos dados. A integridade dos dados, por sua vez, é qualidade desejável, pois sua manutenção visa a assegurar que mudanças feitas no banco de dados não resultem na perda da consistência dos mesmos. Para a garantia da integridade, são utilizadas uma série de restrições às quais os dados devem atender. As restrições de domínio, por exemplo, especificam o conjunto de valores possíveis que podem ser associados a um atributo.

Pode-se dizer, então, que a qualidade dos dados sob o ponto de vista de um sistema de gerenciamento de bancos de dados está ligada à existência controlada de redundância e à satisfação a restrições de integridade.

## 2.5 Qualidade em sistemas de informação geográfica

Uma definição geral de qualidade, tal como se encontra em dicionários, não é apropriada para utilização em SIG pelo fato de deixar os requisitos de avaliação implícitos, o que torna seu entendimento possível somente por pessoas que possuam pleno conhecimento de todo o processo de manipulação dos dados, desde a captura até a apresentação dos mesmos.

Definições comuns à gestão empresarial, devidamente adaptadas, podem auxiliar na adoção de um conceito apropriado para o uso em SIG. De fato, pode-se, até mesmo, traçar um paralelo entre as duas áreas e afirmar que a abordagem sobre qualidade no contexto de SIG também evoluiu segundo a mesma linha de pensamento citada na Seção 2.3. Em uma primeira geração destes sistemas, a qualidade era retratada apenas pela reputação de quem produzia os dados, por experiências resultantes do uso do produto e por uma declaração do órgão produtor informando que os dados produzidos estavam de acordo com um padrão cartográfico de precisão. A situação atual de evolução dos SIG, entretanto, requer mais que uma simples precisão posicional, exigida anteriormente pelos cartógrafos e insuficiente como indicador de qualidade [Gup98].

Para Rothemberg [Rot96], a qualidade de dados possui dois diferentes aspectos, um deles envolve a correteza dos dados e o outro sua adequação. Uma definição atual e bastante aceita para uso no contexto de Sistemas de Informação Geográfica é a proposta por Chrisman [Chr84] e citada no Capítulo 1. Segundo ele, qualidade representa o quanto um conjunto de dados se ajusta às necessidades de determinados usuários, considerando-se suas aplicações geográficas. Apesar de geral, esta definição de qualidade como “*fitness for use*” tem sido adotada por diversos autores como uma boa interpretação de qualidade de dados geográficos.

Para tornar possível a avaliação de qualidade, indicadores precisam ser definidos e índices precisam ser padronizados de acordo com as exigências dos consumidores. Os indicadores são as características que se deseja medir, os índices são os resultados das medições e os padrões são os resultados (índices máximo ou mínimo) que se pretende alcançar para aquele indicador. Normalmente, são usados vários indicadores, preferencialmente que possibilitem medidas objetivas [SEB94a].

Assim como na gestão empresarial, é necessário, na área de SIG, definir indicadores e métodos de medição associados, estabelecer padrões de aceitação para cada indicador e tornar explícitos os resultados da avaliação de qualidade. De acordo com [Aro95] as características que afetam a qualidade (utilidade) dos dados geográficos podem ser agrupadas segundo três categorias:

- características de nível macro;
- características de nível micro; e
- características de uso.

As características de nível macro dizem respeito aos dados como um todo, normalmente são variáveis qualitativas e geralmente não são acessíveis para testes, sendo avaliadas por julgamento ou por relatos de informação sobre os dados. As características de nível micro referem-se aos elementos individuais dos dados, normalmente são variáveis quantitativas

e em geral são avaliadas por testes estatísticos dos dados, usando uma fonte independente de qualidade comprovada. As características de uso são aquelas que dizem respeito aos recursos de uma organização em particular.

Dentro destas características podemos identificar vários critérios ou indicadores de qualidade de dados geográficos. O termo *indicador* é amplamente utilizado em gestão empresarial como “aquilo que se quer medir”. Neste texto, entretanto, assume duas definições: critério, mais apropriado quando se trata das características de nível macro, por significar aquilo que serve de base para comparação, julgamento e apreciação; e indicador, propriamente dito, quando se trata de características de nível micro, por significar aquilo que indica, determina, estabelece, mostra a conveniência de alguma coisa. Desta forma, o texto utiliza o termo *critério* quando não é possível se fazer uma medida objetiva, sendo necessário um julgamento. Ao contrário, o termo *indicador* é usado quando uma medida objetiva é utilizada.

A Tabela 2.1 a seguir, apresenta todas as características identificadas por Aronoff [Aro95] bem como os critérios ou indicadores que pertencem a cada categoria de características.

Características	Indicadores/Critérios
Nível Macro	Completude Atualidade Linhagem
Nível Micro	Precisão posicional Precisão de atributo Consistência lógica Resolução
Uso	Acessibilidade Custo

Tabela 2.1: Características de qualidade dos dados geográficos segundo Aronoff [Aro95].

### 2.5.1 Características de nível macro

As características de nível macro incluem três critérios: a completude, a atualidade e a linhagem.

#### Completude

Os diversos aspectos de completude a analisar em função da qualidade fazem com que haja uma divisão em três subcritérios:

1. a completude de cobertura, que é a proporção dos dados disponíveis para a área de interesse em função de um total estimado;
2. a completude de classificação, que é uma avaliação de quão bem a classificação escolhida está apta a representar os dados; e
3. a completude de verificação, que refere-se à quantidade e à distribuição das medidas de campo ou outras fontes independentes de informação utilizadas para validar os dados.

A avaliação de completude é usualmente limitada à informação de cobertura. Normalmente informações de completude de classificação e verificação são omitidas. Enquanto a completude de cobertura pode ser estimada, normalmente em valores percentuais, a completude de classificação e de verificação devem ser julgadas qualitativamente e ter suas avaliações reportadas mais como descrições que como valores numéricos.

Idealmente, a completude de cobertura deve ser 100 %. A completude de classificação e a de verificação são fatores importantes na determinação da conveniência de um conjunto de dados para uma dada aplicação.

### **Atualidade**

O tempo é um fator crítico para muitos tipos de informação geográfica. As informações demográficas, por exemplo, são bastante sensíveis ao tempo. O aspecto tempo em função da qualidade de dados (atualidade) é normalmente reportado como a data em que o material fonte foi obtido ou como período no qual aqueles dados têm validade. O tempo é um critério freqüentemente omitido quando vários conjuntos de dados são coletados independentemente e utilizados em conjunto. Isto pode acarretar problemas de consistência lógica.

### **Linhagem**

Linhagem de um dado é a sua história, da fonte de dados e dos passos de processamento usados para produzi-lo até o seu armazenamento em um banco de dados geográfico. Cada fonte de dados e seus respectivos métodos de processamento introduzem um nível de erro na informação que será produzida, interferindo na sua qualidade final.

## **2.5.2 Características de nível micro**

As características de nível micro incluem quatro indicadores: precisão posicional, precisão de atributo, consistência lógica e resolução.

### Precisão posicional

Precisão posicional é a relação entre a posição real de um objeto geográfico e sua posição registrada. É usualmente testada pela seleção de uma amostra específica de pontos pre-determinados e comparação destas coordenadas de posição com uma fonte independente de qualidade conhecida.

A avaliação é normalmente efetuada utilizando-se dois métodos distintos, o primeiro usa o desvio e a precisão enquanto o segundo, apenas o erro médio quadrático - RMSE (*Root Mean Square Error*).

No primeiro método, o desvio refere-se às discrepâncias sistemáticas entre a posição representada e a real, sendo medido pela média dos erros posicionais de cada ponto da amostra, enquanto, a precisão refere-se à dispersão dos erros posicionais dos elementos de dados, sendo usualmente estimada pelo cálculo do desvio padrão dos erros de posição dos pontos de teste selecionados.

No segundo método de avaliação da precisão posicional, comum em se tratando de coletas de campo e fotogrametria (métodos comuns de coleta de dados), utiliza-se o RMSE, que é calculado a partir da aplicação da função de mínimos quadrados aos erros posicionais relativos aos pontos de teste.

A Figura 2.3, retirada de [Aro95] exemplifica como é avaliada a precisão posicional usando o primeiro método citado.

### Precisão de atributo

A precisão de atributo trata dos componentes não espaciais dos dados geográficos. Estes componentes podem ser variáveis discretas ou contínuas. Categorias como classe de uso do solo, tipo de vegetação ou áreas administrativas assumem valores discretos. Variáveis como temperatura ou valor médio de uma propriedade, por sua vez, assumem valores contínuos.

O método para avaliar precisão de atributos cujo domínio é contínuo é similar ao discutido para precisão posicional, enquanto que para variáveis discretas avalia-se a precisão de classificação. A avaliação da precisão de classificação é um procedimento complexo e controvertido que tem recebido considerável atenção na literatura de sensoriamento remoto [Aro95]. As dificuldades advêm do fato das medidas de precisão serem significativamente afetadas por diversos fatores, tais como: o número de classes, a forma como os pontos de teste são selecionados e de algumas classes serem confundidas com outras.

Como resultado, a avaliação da precisão de classificação não é inteiramente objetiva. Apesar destas dificuldades, um bom número de aproximações de precisão de classificação têm sido desenvolvidas.

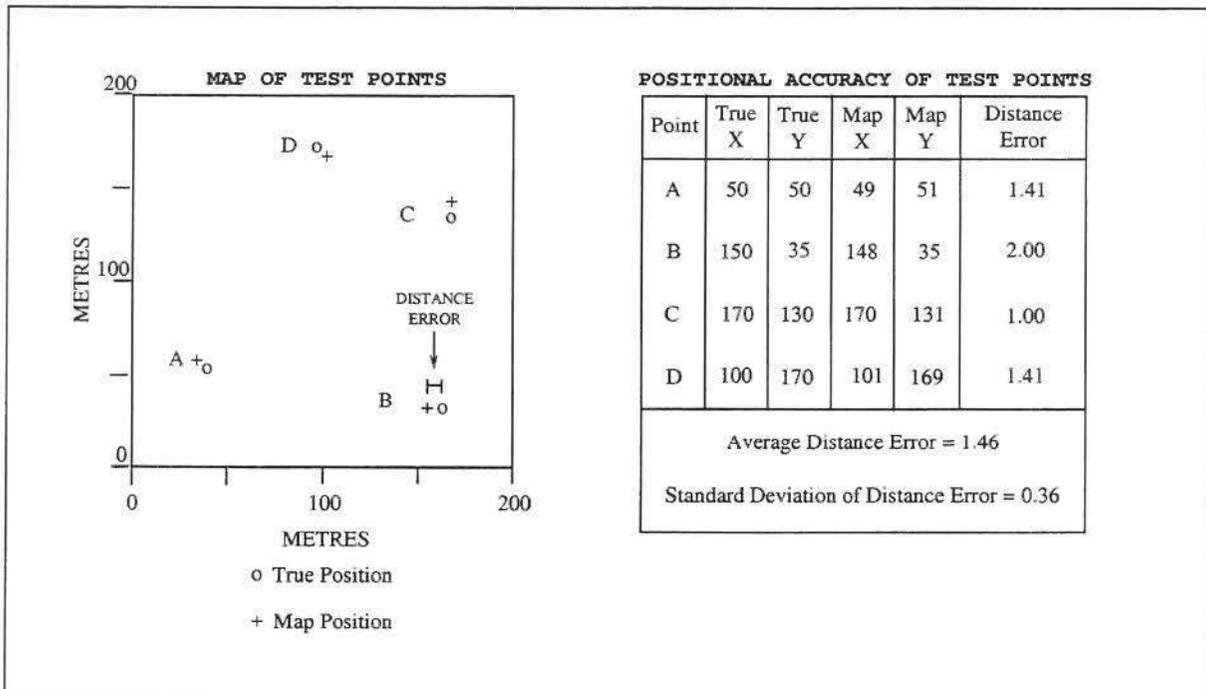


Figura 2.3: Avaliação de precisão posicional usando média e desvio padrão segundo Aronoff [Aro95].

### Consistência lógica

A consistência lógica refere-se à manutenção das regras de consistência entre os objetos geográficos. Um tipo especial de consistência lógica é a que trata dos relacionamentos topológicos entre os dados. Dois conjuntos de dados podem estar corretos quanto ao nível de precisão, e assim mesmo não possuírem consistência lógica. Por exemplo, se polígonos adjacentes forem digitalizados por pessoas ou métodos diferentes, a fronteira comum pode ser mapeada com posições ligeiramente diferentes e ainda manter precisão posicional. Quando estes polígonos são integrados em uma mesma aplicação (ou banco de dados) esta diferença pode ser acentuada.

Não há um padrão único para se medir a consistência lógica, mas normalmente, são utilizadas regras baseadas nas restrições de integridade das entidades envolvidas. Os resultados da avaliação deste indicador podem ser informados através de percentuais de satisfação às regras.

### Resolução

O componente espacial é muitas vezes também avaliado por outro critério, a resolução, que se refere à menor unidade discernível ou apresentável dos dados. Quando se trata

de dados de sensoriamento remoto usa-se o termo *resolução espacial* (linhas/mm para câmaras fotográficas e *pixels* em imagens de satélite) e quando se trata de mapas o termo unidade mínima de mapeamento é mais apropriado. Os dados armazenados em um banco de dados geográfico, podem, em teoria, ser representados em qualquer escala. Entretanto, embora os dados não existam em uma escala específica, na prática existe um compromisso entre o nível de precisão posicional e de resolução que os tornam apropriados para o uso em determinadas escalas.

### 2.5.3 Características de uso

O uso prévio de um conjunto de dados para uma aplicação pode ser um bom indicador, sob o ponto de vista de um novo usuário, do ajuste destes dados a uma outra aplicação. Pode também indicar uma confiança geral no conjunto de dados para diferentes circunstâncias. As características de uso podem incluir inúmeros critérios ou indicadores dentre os quais se pode citar a acessibilidade dos dados e seu custo de aquisição ou produção.

A acessibilidade refere-se a quão fácil é obter e usar os dados. A acessibilidade de dados pode ser restrita pelo fato destes dados serem de propriedade privada.

O presente trabalho enfocará apenas duas das três categorias de características que concernem a qualidade de dados geográficos, as características de nível macro e de nível micro. As características de uso são restritas a cada organização ou aplicação e portanto considerações sobre seus indicadores/critérios não serão abordadas.

## 2.6 *Resumo*

Este capítulo apresentou alguns conceitos e terminologia básicos que serão utilizados no restante do texto.

Alguns aspectos de Cartografia foram abordados, de maneira genérica, por serem fundamentais também em Sistemas de Informação Geográfica. Aspectos mais específicos de Cartografia serão tratados no Capítulo 4.

Foram discutidas as dificuldades de se definir SIG, apresentadas algumas definições, visões e uma arquitetura geral para estes sistemas. Seguiram-se algumas definições com o intuito de caracterizar os Sistemas de Informação Geográfica.

Finalmente, foram abordadas algumas noções sobre qualidade em áreas de pesquisa afins, apresentada uma definição de qualidade no contexto de SIG e discutidas algumas características que afetam a qualidade de dados geográficos, bem como mencionados alguns critérios e indicadores para possibilitar a sua avaliação.

## Capítulo 3

# Qualidade e problemas relativos aos dados

Este capítulo aborda alguns dos problemas de qualidade relativos aos dados geográficos. Para a criação de um banco de dados geográfico é importante definir suas fontes de dados e o tratamento por que estes dados devem passar para serem armazenados. A criação do banco de dados geográfico engloba considerações sobre as características dos diversos tipos de fontes de dados (Seção 3.1); a escolha dentre estas, tanto para dados espaciais quanto convencionais (Seção 3.2); a aquisição (coleta) destes dados brutos (Seção 3.3); seu processamento (conversão) para um formato apropriado (Seção 3.4); e uso de metadados (Seção 3.5), finalizando com considerações sobre documentação do processo de controle e garantia de qualidade (Seção 3.6). Este é um passo inicial para analisar que tipos de problemas podem afetar a qualidade dos dados.

### 3.1 Características das fontes de dados e sensores

As fontes de dados espaciais podem ser, entre outras, mapas, desenhos, fotografias aéreas e imagens de satélite. Já os dados convencionais normalmente podem vir das informações descritivas existentes nos próprios mapas e desenhos, da extração de características de interesse a partir da análise das fotografia e imagens de satélite ou ainda de arquivos em papel ou bancos de dados já existentes.

#### 3.1.1 Fontes de Dados espaciais

##### Mapas

A maioria das organizações que já criaram uma aplicação geográfica usa um grande número de mapas. Virtualmente, todo processo de criação de um banco de dados geo-

gráfico envolve durante a captura dos dados (coleta/conversão) o uso de mapas como documentos fonte. Estes mapas variam amplamente em termos de conteúdo, meio, formato, idade, escala, complexidade e uso. Segundo [CCH<sup>+</sup>96] os mapas utilizados para tratamento de dados geográficos podem ser classificados em temáticos e cadastrais. Os mapas temáticos são aqueles que mostram uma região geográfica particionada em polígonos segundo os valores relativos a um tema (por exemplo, uso do solo). Os mapas cadastrais, ao contrário, apresentam objetos identificáveis (por exemplo, lotes de terreno).

A Figura 3.1 obtida de [Mon91] mostra o processo de confecção de um mapa em dois estágios. No primeiro estágio, a Terra é transformada em um globo onde a escala é válida em qualquer lugar e em todas as direções. No estágio seguinte, o globo é projetado em uma superfície tal como um plano, cone ou cilindro que toca o globo em um ponto, uma ou duas linhas padrão. Nos mapas finais, a escala é usualmente constante somente ao longo destas superfícies de contato.

Já no primeiro estágio ocorre a primeira simplificação que leva à inserção de erro: a escolha da forma da Terra. A superfície da Terra é extremamente irregular, possuindo a forma de um geóide (sólido formado pela superfície do nível médio do mar, suposta prolongada através dos continentes). Entretanto, para atender às necessidades da Cartografia permanece a utilização da forma regular que mais se aproxima: a de um elipsóide de revolução (sólido formado pela rotação de uma elipse em torno do eixo de seus pólos) [Bak65].

Apesar do consenso sobre o uso da forma elipsoidal da Terra, não há consenso sobre as medidas dos elipsóides, que variam conforme o autor, ou sobre qual elipsóide utilizar. O Brasil, por exemplo, para a construção de cartas náuticas, utiliza o elipsóide de Hayford de 1909, recomendado pela Convenção de Madri de 1924. Os Estados Unidos e outros países, diferentemente, utilizam o elipsóide de Clarke de 1866 [Oli95]. A Tabela 3.1, adaptada de [Bak65], apresenta alguns dos elipsóides de referência conhecidos:

<b>Autoridade</b>	<b>Raio Equatorial</b>	<b>Raio Polar</b>
Bessel (1841)	6.377.397,15 m	6.356.078,96 m
Clarke (1866)	6.378.206,40 m	6.356.583,80 m
Clarke (1880)	6.378.249,17 m	6.356.514,99 m
Helmert (1907)	6.378.200,00 m	6.356.818,17 m
Hayford (1909)	6.378.388,00 m	6.356.911,95 m

Tabela 3.1: Elipsóides de referência segundo Bakker [Bak65].

Na realidade, a adoção de um ou outro elipsóide e até mesmo de uma esfera para a forma da Terra, apesar de inserir uma pequena distorção, é perfeitamente aceitável para as finalidades cartográficas [Bak65]. Pode haver, entretanto, um aumento do erro com diminuição da qualidade dos resultados se, inadvertidamente, forem combinados

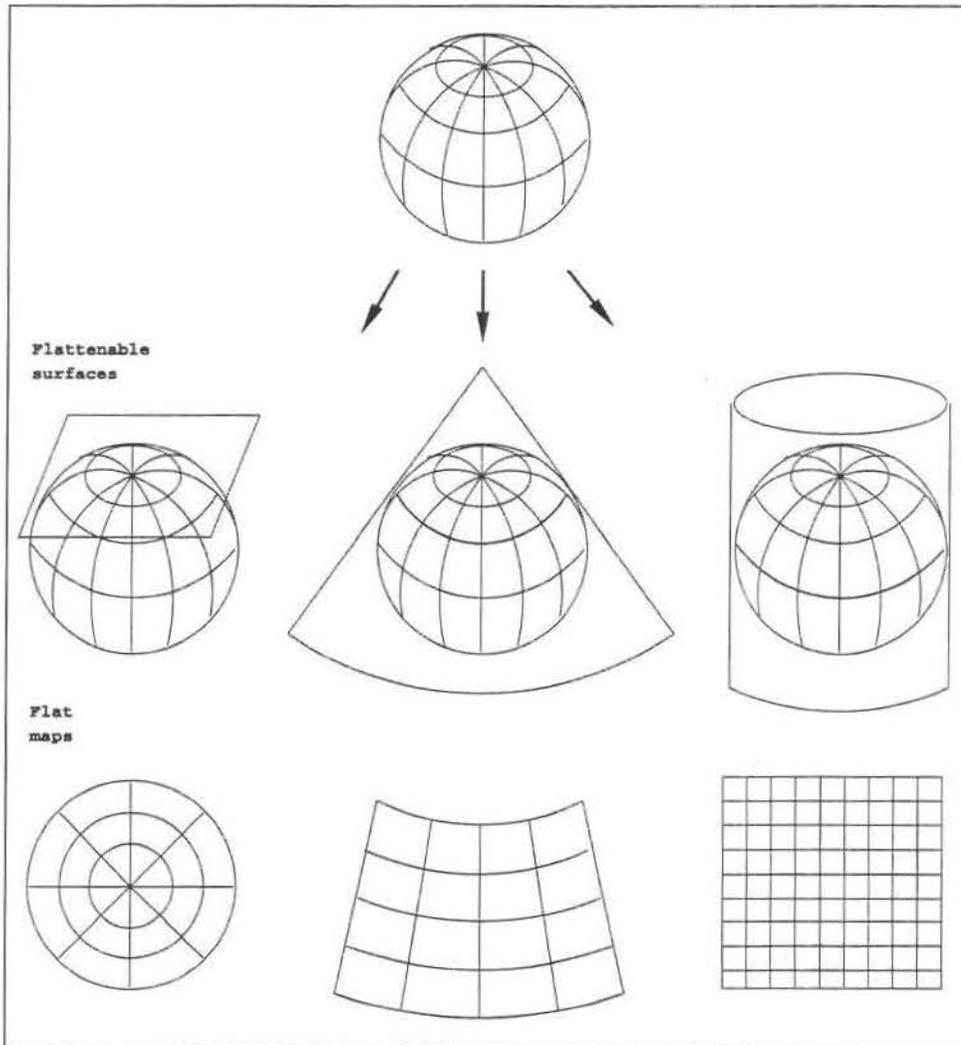


Figura 3.1: Estágios da confecção de mapas segundo Monmonier [Mon91].

dados obtidos a partir de mapas que utilizem elipsóides de referência distintos, projeções que privilegiem diferentes propriedades espaciais e escalas muito diferentes, como será apresentado a seguir.

No estágio seguinte da confecção de um mapa (Figura 3.1), que é o da projeção propriamente dita, as imperfeições advêm de diversos fatores, que servem, inclusive, para classificar os vários sistemas de projeção existentes: da forma como a superfície de projeção toca o globo (tangente, secante ou composta); do tipo de superfície de projeção adotada (plana, cônica, cilíndrica ou poliédrica) e do aspecto (equatorial, polar, oblíqua ou transversa) [CCH<sup>+</sup>96].

Atualmente, a criação de mapas para aplicações em SIG está intimamente ligada ao uso de receptores do sistema GPS. O *Global Positioning System* (GPS) é originalmente

um sistema de navegação militar baseado em satélites desenvolvido pelo Departamento de Defesa dos Estados Unidos e que permite a um usuário em terra obter sua posição com relativa precisão a partir do recebimento e comparação dos sinais de pelo menos quatro satélites [Ber92]. Isto é possível pelo fato de os satélites possuírem uma órbita precisa e conhecida e transportarem relógios atômicos. O sistema possui atualmente cerca de 24 satélites e fornece dois tipos de serviço: o PPS (*Precise Positioning Service*) para uso militar e o SPS (*Standard Positioning Service*), disponibilizado para a população civil. O serviço preciso provê posições com precisão horizontal de cerca de 22 m, vertical de 28 m e temporal de 100 ns. O serviço padrão, cujo sinal é degradado de forma intencional pelo DoD-US (*Department of Defense - US*), tem precisão de cerca de 100 m horizontal, 156 m vertical e 360 ns temporal. Estes resultados podem ser melhorados, ao utilizarmos uma posição conhecida de terra para corrigir o sinal recebido pelo GPS. Este arranjo é normalmente denominado DGPS (*Differential Global Positioning System*) e sua utilização permite uma precisão horizontal de aproximadamente 1 a 10 m quando utilizando o serviço padrão [NAV99]. Os receptores do Sistema de Posicionamento Global variam bastante em precisão e funcionalidade [CCH<sup>+</sup>96].

## Desenhos

A diferença primária entre um mapa e um desenho é que este último não utiliza um sistema de referência de coordenadas geográficas nem uma escala universal. Um desenho pode ser simplesmente uma representação esquemática ou uma representação em escala, mas ambas sem dados geograficamente referenciados [MS93].

De maneira semelhante aos mapas, os desenhos variam amplamente em termos de conteúdo, meio, formato, escala (se houver), complexidade e uso. Normalmente os desenhos são considerados como fontes suplementares aos mapas. Os mapas provêm uma estrutura de referência em escala global, e a informação dos desenhos é inserida nesta estrutura. Os desenhos podem ser usados para extrair informação que não é mostrada ou não possui um nível adequado de detalhes nos mapas.

## Fotografias aéreas

As fotografias aéreas e as imagens de satélite estão entre as tecnologias de aquisição de dados conhecidas como Sensoriamento Remoto. Sensoriamento Remoto é o conjunto de processos e técnicas usados para medir propriedades eletromagnéticas de uma superfície ou objeto, sem que haja contato entre o objeto e o equipamento sensor [DS91, CCH<sup>+</sup>96]. Fotogrametria, por sua vez, pode ser definida como a ciência de se obter medidas acuradas e confeccionar mapas a partir de fotografias aéreas [Wol83].

As fotografias aéreas possuem duas funções principais no processo de criação do banco

de dados geográfico, atuar como fonte de dados ou como fonte de atualização, criando novos arquivos em escala grande ou servindo para controle de dados já armazenados [CCH<sup>+</sup>96].

Um projeto de mapeamento utilizando fotografias aéreas inclui o planejamento de rotas para a aquisição das fotografias, a escolha de pontos de controle para georeferenciamento, a digitalização e a geração automática a partir da fotografia.

As fotografias aéreas são obtidas a partir de um avião voando em linhas paralelas, chamadas de linhas de voo. O planejamento de voo garante que haja sobreposição para assegurar a total cobertura da área a ser estudada. Assim como as outras fontes de dados, as fotografias também variam em termos de conteúdo, meio, formato, idade, escala, complexidade e uso.

### Imagens de satélite

Dentre os principais satélites cujos dados são recebidos no Brasil, destacam-se as séries LANDSAT [NAS99] e SPOT [CNE99]. Os satélites destas séries variam bastante na quantidade, tipo e características dos sensores que utilizam.

Algumas propriedades básicas de um sensor, por exemplo resolução (espectral, radiométrica, espacial e temporal) ou cobertura espectral, afetam a qualidade dos dados produzidos. Nos sistemas de sensoriamento remoto há um compromisso entre cada uma destas características. Para alguns usos, por exemplo um mapeamento cartográfico, o nível de detalhe espacial (precisão posicional) é crucial; em outros casos, a informação deve ser coletada frequentemente (atualidade) mas não há necessidade de um grande nível de detalhe.

O uso dos dados coletados é normalmente dependente do tipo de sensor utilizado e de algumas características específicas do sensor tais como resolução temporal e resolução espectral, que por sua vez dependem do percurso e da altitude da plataforma aonde o sensor está localizado [BM98]. Cada objeto reflete energia eletromagnética em vários comprimentos de onda, determinados pela sua estrutura física e química. Estes dados de diferentes comprimentos de onda são combinados para compor as imagens.

Analisar este tipo de informação envolve técnicas de uma grande variedade de disciplinas incluindo processamento de imagens, processamento de sinais, estatística, fotointerpretação e de uma forma cada vez maior inteligência artificial. Podemos distinguir três fases no processo de tratamento de imagens de satélites de acordo com Günther [Gun98]: processamento da imagem icônica; classificação; e processamento da imagem simbólica.

O primeiro passo corresponde essencialmente ao processamento dos dados brutos. Nesta fase é aplicada uma grande variedade de algoritmos de processamento de imagens aos dados brutos obtidos para proporcionar uma visualização apropriada. Algumas destas técnicas são retificação da geometria das imagens para coincidir com a escala e as coor-

denadas dos mapas a serem utilizados, correções de intensidade espectral para aumento de contraste, eliminação de falhas de sensor e supressão de ruído.

O segundo passo serve para classificar a imagem com respeito a uma taxionomia específica de uma dada aplicação, por exemplo, uso do solo. Esta fase é feita através da análise das assinaturas espectrais e requer considerável interação humana. É basicamente um processo heurístico baseado na experiência, conhecimento técnico e conhecimento do domínio da aplicação. Para o processo de classificação há duas aproximações: a classificação supervisionada e a não supervisionada. Na primeira, dados autênticos de observações efetuadas em campo são usados para identificar os parâmetros usados na classificação. Já a classificação não supervisionada não necessita deste estágio inicial: de acordo com a impressão digital espectral os *pixels* são classificados e organizados em agrupamentos com uma cor associada, representativa da feição desejada.

O processamento da imagem simbólica corresponde aos estágios de validação e interpretação. O dado resultante é condensado, agrupado em objetos, interpretado e associado a entidades geográficas. Usualmente este processo envolve a conversão de dados do formato *raster* para o formato *vector* que é mais compacto. No processamento da imagem simbólica uma atividade importante é a segmentação, que começa com a comparação do valor de um *pixel* com os valores dos *pixels* imediatamente próximos, agregando-os dentro de certos limites estabelecidos. Ao final, a imagem estará separada por regiões de semelhança, que provavelmente indicarão feições geográficas.

### 3.1.2 Fontes de Dados convencionais

Dados convencionais são aqueles usualmente considerados em aplicações de bancos de dados. Sua modelagem, coleta, processamento e armazenamento têm sido objeto de estudo há décadas e portanto não serão tratados nesta dissertação. Problemas antigos, porém ainda sem solução satisfatória envolvem sua integração (quando as fontes são heterogêneas) e, atualmente, a conversão a partir de sistemas legados.

Dados convencionais podem ser inseridos no banco de dados geográfico manualmente ou serem importados usando algum formato padrão de transferência. Após as etapas de preparação e entrada dos dados, é importante que seja realizada uma etapa final de edição, principalmente quando for realizada entrada manual. Existem também, quando se trata de SIG, alguns dados convencionais não usuais como os relativos à produção cartográfica (legendas etc.) como pode ser visto no Capítulo 4.

Algumas organizações que se propõem a criar um banco de dados geográfico já possuem dados geográficos, em forma digital, em algum nível. Dependendo do caso, os dados digitais existentes podem ser completamente traduzidos para o banco de dados ou podem ser tornados acessíveis de forma compartilhada. Qualquer que seja a opção, os dados

digitais existentes devem ter a sua qualidade comprovada, especialmente para os dados que provêm ligação para os dados espaciais.

## 3.2 Escolha das fontes de dados

Como visto na seção anterior, as fontes de dados para aplicações geográficas são variadas e numerosas. Apesar da maioria dos problemas associados à coleta e conversão de dados para SIG ter sido identificada em meados dos anos 70, estes dois processos continuam sendo os aspectos que mais consomem recursos (tempo e dinheiro) para a criação de um banco de dados geográfico.

Pelo fato de os dados geográficos serem encontrados sob diversas formas, o banco de dados pode ser construído de maneiras variadas e não mutuamente exclusivas [BM98]: utilização de dados já existentes em formato digital; produção de dados digitais a partir dos dados em formato analógico já existentes; e realização de novo levantamento analógico ou digital.

A etapa mais crítica da criação do banco de dados geográfico é o planejamento da estratégia de aquisição dos dados e sua posterior manutenção. Cada uma das opções anteriormente citadas e discriminadas a seguir tem vantagens e desvantagens que variam de acordo com os requisitos do projeto sendo desenvolvido. A decisão entre estas diferentes alternativas envolve diferentes aspectos de custo de aquisição e manutenção, dentre outros, e foge ao escopo desta dissertação.

### 3.2.1 Utilização de fontes de dados em formato digital

A principal maneira para se reduzir o custo ao criar um banco de dados para aplicações geográficas é adaptar bancos de dados preexistentes.

Fontes de dados em formato digital existem em número cada vez em maior. Além dos fornecedores comerciais, um grande número de agências governamentais tem dados em formato digital disponíveis ao público, às vezes até disponíveis por intermédio da Internet. A maioria dos dados digitais fornecidos por organizações governamentais é de natureza geral - relevo, estradas, rios - e para uma grande faixa de usuários. Em paralelo, organizações comerciais começaram a produzir conjuntos de dados digitais, mais específicos que os das organizações governamentais, para atender à crescente demanda de informação [BM98].

A aquisição de dados de algum fornecedor deve considerar os metadados associados. Os metadados servirão principalmente para fornecer informação adicional que mostrará se um conjunto particular de dados será apropriado para o projeto em curso e possibilitar a revisão do projeto por quaisquer motivos [Ols97]. Deve-se dar grande atenção à

compatibilidade entre dados fornecidos por diferentes fornecedores e combinados em um mesmo projeto. Eles poderão ser diferentes em projeção, escala e descrição de atributos causando inúmeros problemas. A incorporação de metadados nestes casos também é importante para possibilitar a integração dos dados, conforme citado no Capítulo 1.

De acordo com Burrough & McDonnell[BM98], em um nível mais prático, um usuário deverá considerar, dentre outras, as seguintes características com relação às fontes de dados para assegurar que os mesmos sejam úteis a suas aplicações: atualidade; escala; sistema de georeferenciamento; técnica de coleta e amostragem de dados; e métodos de classificação e interpolação utilizados.

### 3.2.2 Produção de dados digitais a partir de dados em formato analógico

A produção de dados digitais a partir de dados analógicos é uma prática muito comum, seja pela inexistência de levantamentos por meio digital para uma determinada região ou pela falta de dados na escala desejada. Exemplos de fontes de dados analógicos são mapas e desenhos, fotografias aéreas e dados em papel. Para cada tipo de fonte de dados disponível, existem métodos de conversão e entrada de dados no banco de dados.

Os mapas são as principais fontes de dados analógicos espaciais. Usuários de bancos de dados geográficos freqüentemente desconhecem as limitações dos mapas convencionais e, conseqüentemente, fazem considerações impróprias sobre os dados derivados destes, desejando um nível de qualidade resultante melhor que a fonte que os originou [Fis91]. Isto vai de encontro à idéia de qualidade em sistemas de informação proposta por [Orr98] (Seção 2.4).

Existem dois métodos para produzir dados digitais a partir de mapas: a digitalização manual e a varredura (“escaneamento”). Há vantagens e desvantagens no uso de cada uma destas técnicas e além disso ambas introduzem erros. A digitalização pode ser efetuada com equipamentos de baixo custo, não requer muito treinamento e não necessita de mapas de rigorosa qualidade gráfica, sendo em contrapartida um processo lento. O escaneamento é facilmente efetuado e rápido mas requer equipamentos mais caros, envolve pessoal experiente, necessita de considerável edição e precisa de mapas claros e bem definidos [Ber92]. Mapas que contenham grande número de informação adicional, requeiram interpretação e ajustes durante o processo de codificação ou tenham um pequeno número de feições<sup>1</sup> a serem codificadas geralmente são melhores para a digitalização. O processo de escaneamento é mais eficiente para mapas com grande número de elementos espaciais (cerca de 1000 polígonos ou mais) e mapas com grande número de feições com formas irregulares

---

<sup>1</sup>Uma feição pode ser entendida como o fato, objeto ou fenômeno de interesse sobre a superfície terrestre que se quer modelar como dado geográfico.

[Aro95].

Outro importante tipo de fonte de dados analógicos espaciais são as fotografias aéreas, processadas de forma a gerar o mapa digital. Neste caso a mesa de digitalização é substituída por um instrumento fotogramétrico denominado Restituídor Analítico (*Stereoplotter*) [BM98].

Assim como os dados espaciais, os dados descritivos (dados convencionais ou não espaciais) também necessitam de tratamento para a entrada no banco de dados geográfico. Os dados convencionais também podem vir de inúmeras fontes. Eles podem ser inseridos manualmente ou através da importação usando algum formato padrão de transferência ou até mesmo usando OCR (*Optical Character Recognition*). Uma vez efetuada a entrada destes dados é importante efetuar a verificação e a edição dos dados em busca de erros, possíveis omissões, falta de precisão e outros possíveis problemas típicos de gerenciamento de bancos de dados convencionais.

### 3.2.3 Realização de novo levantamento analógico ou digital

Embora a oferta de dados digitais ou analógicos seja crescente, sua qualidade pode apresentar problemas, podem não atender aos requisitos do usuário, estar desatualizados, ter sido digitalizados em escalas não compatíveis, etc. Se as fontes de dados não são adequadas, é necessário realizar levantamentos para uma nova aquisição de dados.

A qualidade dos dados em um banco de dados geográfico é sempre um compromisso entre necessidades e custos. Na prática porém, a escolha é frequentemente uma questão de usar o que está disponível ou o que pode ser adquirido dentro de um certo intervalo de tempo. Como se trata de dados geográficos, o processo de captura pode ter sua rapidez comprometida devido a uma grande variedade de causas: além da heterogeneidade de fontes outros fatores - por exemplo, climáticos - podem afetar as condições de captura. Finalmente, o fornecimento de dados pode sofrer atrasos por problemas de direitos autorais e outras dificuldades legais relativas à propriedade dos dados.

## 3.3 Problemas na captura dos dados - coleta

Por captura de dados podemos entender a fase de criação do banco de dados geográfico que engloba coleta e a conversão dos dados brutos [Gun98].

A precisão final dos dados depende muito da precisão da fonte. Em princípio, o usuário dita a necessidade por maior precisão. Uma precisão mais alta requer qualidade de dados inicial mais alta e processo mais preciso. Por exemplo, se um mapa não mostra a informação exigida com a precisão apropriada, cálculos mais precisos ou um mapeamento aéreo podem ser necessários para prover os dados com a qualidade requerida. Isto mostra

a importância de identificar a qualidade das fontes como parte da definição do banco de dados, antes mesmo dos processos de coleta/conversão começarem. Além disso, as exigências por atualidade afetam a medida de qualidade de dados. Precisão e atualidade são as duas características que mais influenciam a qualidade dos dados nesta fase.

A análise da precisão de uma fonte de dados deve considerar o erro adicional introduzido durante o próprio processo de conversão de dados. Conseqüentemente, o efeito combinado do erro intrínseco da fonte com o erro de conversão de dados deve ser considerado durante a determinação dos níveis de precisão da fonte e o estabelecimento de limites de controle de qualidade.

A Tabela 3.2 a seguir mostra que tipos de parâmetros de qualidade são normalmente analisados para o estabelecimento do limite de qualidade desejado. Os parâmetros citados na tabela foram identificados por Montgomery e Schuch [MS93], no caso da coleta e pelos mesmos autores e Hohl [Hoh98], no caso da conversão. Os parâmetros identificados para a fase de coleta descrevem a qualidade das fontes utilizadas para a obtenção dos dados geográficos, enquanto isso, os parâmetros identificados para a fase de conversão, descrevem a qualidade do processo de transformação dos dados de um formato inicial (mapas, desenhos, arquivos, etc.) para o formato digital. Assim, diversos parâmetros são identificados para avaliar a qualidade das fontes enquanto que a avaliação do processo se restringe à análise de um número menor delas.

Parâmetros	Coleta	Conversão
Precisão posicional	X	X
Cobertura	X	-
Compleitude	X	X
Atualidade	X	X
Precisão de atributo	X	X
Credibilidade	X	-
Validade	X	X
Nível de Confiança	X	-
Conveniência	X	-
Condições físicas	X	-
Legibilidade	X	X
Precedência	X	-

Tabela 3.2: Parâmetros de qualidade utilizados para avaliação da qualidade da captura dos dados segundo Montgomery & Schuch [MS93] e Hohl [Hoh98].

O parâmetro precisão posicional na fase de conversão divide-se em precisão absoluta e precisão relativa. A nomenclatura utilizada para validade na fase de conversão é normalmente integridade. E o parâmetro legibilidade, juntamente com outro denominado simbolização, que consiste na atribuição de símbolos, compõe a qualidade gráfica na fase

de conversão.

### 3.3.1 Parâmetros segundo Montgomery&Schuch [MS93]

De acordo com [MS93], os parâmetros que se seguem devem ser considerados na escolha das fontes de dados pois descrevem sua qualidade. Eles são semelhantes àqueles que avaliam a qualidade do banco de dados:

#### Precisão posicional

É entendida como a conformidade do componente espacial das feições de interesse existentes na fonte em relação a algum padrão de registro espacial, normalmente coordenadas geográficas - latitude ( $\varphi$ ) e longitude ( $\lambda$ ). Na maioria das vezes, é o parâmetro mais importante a considerar sobre as fontes de dados e sem dúvida o mais avaliado deles.

A precisão posicional das fontes precisa ser determinada ou verificada, visando indicar se a fonte terá precisão suficiente para suportar determinada aplicação. A análise de precisão deve considerar também os erros adicionais que serão inseridos durante o processo de conversão.

#### Cobertura

É entendida como sendo a abrangência da fonte em relação à área de interesse para a obtenção dos dados geográficos que povoarão o banco de dados.

A cobertura geográfica dos dados fonte deve ser inventariada e comparada com a extensão cartográfica exigida pelo projeto. Frequentemente, a cobertura geográfica das fontes não cobre toda a área analisada. É comum ter-se que adicionar outras fontes de dados, inclusive inventários de campo, para completar a cobertura planejada.

Quando múltiplas fontes, com níveis de precisão diferentes, são usadas para completar a cobertura de um conjunto de dados, o usuário deve ser alertado sobre o nível de precisão resultante que pode ser esperado. Por exemplo, um projeto com múltiplos participantes pode incluir uma cidade e um município. A cidade pode ter fotografias aéreas atuais e em escala grande para usar como fonte de dados para estradas, enquanto o município pode ter só mapas em uma escala menor, aumentados para atender às suas necessidades cartográficas. Desta forma, a precisão da feição “estrada” será muito diferente dentro e fora dos limites da cidade.

#### Completude

Descreve o relacionamento existente entre os objetos da fonte de dados e o mundo real que se deseja representar.

As fontes devem mostrar a maioria, senão todas as características específicas. Por exemplo, no caso citado anteriormente, estradas não podem ser esquecidas, e todas as estradas mostradas devem ter um nome anexado.

Uma desvantagem de incluir várias fontes é o aumento dos custos envolvidos devido à necessidade de manipulação de documentos adicionais, além do aumento no número de problemas oriundos de conflitos entre as fontes. Havendo conflito entre múltiplas fontes, deve ser indicada qual a fonte preferida durante as especificações do processo de conversão.

### **Atualidade**

É o período decorrido entre a coleta dos dados presentes na fonte e sua utilização.

Ao avaliar a atualidade de uma fonte de dados, o usuário deve perceber que nem todos os itens são atualizados ao mesmo tempo e que pode haver vários ciclos de atualização para uma fonte em particular. Algumas companhias de utilidade pública ("*utility companies*"), por exemplo, podem indicar que atualizam seus mapas a cada duas semanas com informação de novas instalações, mas outras informações municipais podem ser atualizadas só anualmente.

### **Precisão de atributo**

É entendida como a proximidade entre os valores atribuídos aos dados na fonte e os valores das feições correspondentes no mundo real.

A precisão de atributo de uma fonte de dados mede o quanto sua informação condiz com o mundo real, por exemplo, uma rodovia não pode ser identificada como um rio. A precisão de atributo é às vezes referenciada como exatidão.

### **Credibilidade**

É entendida como a probabilidade que os dados de uma determinada fonte têm de serem críveis. A credibilidade é freqüentemente estabelecida por comparação entre múltiplas fontes. Se todas as fontes concordam quanto ao valor de um certo atributo, a credibilidade do valor aumenta. Em casos de conflito, fontes mais recentemente verificadas normalmente prevalecerão em relação a outras. Em geral, cabe aos usuários julgar o nível de credibilidade.

### **Validade**

Indica se valores de atributos obedecem a restrições de integridade. Por exemplo, se uma companhia de utilidade pública compra e instala postes que só estão disponíveis em

alturas múltiplas de 2m, um poste com 11 m não é válido. Este valor incorreto poderia reduzir a validade global da fonte.

### **Nível de confiança**

É uma opinião subjetiva, nem sempre baseada em informações mensuráveis, que uma organização ou usuário pode ter sobre uma fonte em especial.

Com o passar do tempo, uma organização tende a desenvolver uma opinião subjetiva e informal sobre a confiança de um documento fonte em particular. Se as equipes de campo e engenheiros encontram continuamente condições diferentes das mostradas nos mapas de operação, esses mapas são considerados como uma fonte incerta de informação. Esta falta de confiança pode ter uma grande variedade de razões, mas o que importa ao final é que este sentimento ou avaliação subjetiva dos usuários normalmente confirma o fato de que as informações contidas nestas fontes não são confiáveis.

Este parâmetro é bem similar ao parâmetro credibilidade. A diferença básica reside no fato da credibilidade ser um parâmetro mais voltado à avaliação dos dados fornecidos pela fonte enquanto o nível de confiança se reporta mais às fontes de dados propriamente ditas.

### **Conveniência**

Indica a facilidade de utilização da fonte de dados para um determinado fim (normalmente no que concerne à conversão de dados). Este fator simplesmente mostra o quão fácil é localizar e usar uma fonte. Normalmente, quanto mais conveniente uma fonte, mais ela será requerida pelo pessoal da organização durante o período de conversão de dados.

### **Condições físicas**

É entendida como o estado físico da fonte de dados considerando-se o tempo transcorrido desde sua coleta.

Muitos documentos fonte analógicos são tão antigos que requerem manipulação especial. Em alguns casos, pode ser necessário mudar a fonte, transferindo-a para outro meio, por exemplo fotografando ou escaneando o material.

### **Legibilidade**

É a nitidez de uma determinada fonte de dados (facilidade de poder ser “lida”). Este é também um fator de qualidade porque a conversão de dados é normalmente efetuada por pessoal contratado, que nem sempre está familiarizado com a fonte. Infelizmente, o

processo de reprodução freqüentemente reduz a legibilidade global de um determinado tipo de fonte.

### Precedência

Indica a prioridade de utilização de uma certa fonte de dados, em relação às demais disponíveis.

Quando fontes de dados são produzidas a partir de outras, considerações de qualidade indicarão, em uma classificação de precedência de geração, qual fonte deverá ser utilizada como principal para produzir dados. Por exemplo, mapas de utilidade pública poderão mostrar informações da área analisada tais como ruas, mas esta informação foi provavelmente copiada a partir de outros mapas. Durante o processo de cópia, as informações normalmente perdem um pouco em qualidade, sendo quase sempre preferível utilizar a fonte precedente. Isto não vale como regra, pois depende do estado de conservação das fontes.

Outros parâmetros para avaliação de qualidade, como consistência lógica, consistência de ligação (entre dados geográficos e não geográficos) e linhagem, dentre outras, podem ser utilizados durante o processo de conversão dos dados em função dos erros inseridos por ocasião dessa fase.

## 3.4 Problemas no pré-processamento dos dados - conversão

O processo de conversão consiste na transformação das informações existentes para armazenamento em um banco de dados geográfico a ser utilizado por um SIG. Nesta etapa é preciso que haja uma preocupação constante com os erros inseridos na manipulação dos dados fonte. O que se pretende é que, apesar destes erros, seja possível atingir o nível de qualidade desejado que deve ser definido antes do início do processo de conversão [MS93].

Dos fatores citados anteriormente, aqueles que se referem à precisão são agravados pelo processo de conversão, que introduz erros e pode piorar a qualidade geral dos dados.

No caso de atributos disponíveis em mais de uma fonte, o que é comum, pode ainda haver discrepâncias entre as fontes. Se isto ocorrer pode ser útil aos usuários finais saber a fonte que foi utilizada para aquela conversão. O conhecimento e o acompanhamento do processo de conversão torna-se, assim, fator importante para a qualidade do banco de dados geográfico.

A idéia de processos de controle e de garantia de qualidade em SIG aproxima-se bastante das idéias já bastante difundidas no campo da indústria e dos serviços através da GQT (Gerência pela Qualidade Total ou TQM, como é mais conhecida *Total Quality*

*Management*). A área de gerência de qualidade da fabricação de produtos ou prestação de serviços tem uma extensa literatura, com princípios, orientações e técnicas para produzir com qualidade. Baseados na TQM, projetos têm sido desenvolvidos no sentido de criar uma metodologia análoga na área de Sistemas de Informação, que tem sido denominada de TDQM (*Total Data Quality Management*) [Wan94, WSG96, Wan98].

Diversas diferenças podem ser percebidas entre as duas abordagens, principalmente com relação às dimensões ou parâmetros que representam qualidade para os dois tipos de produção (produtos industriais *versus* dados geográficos). Contudo, a idéia básica prevalece: para aumentar a produtividade de uma organização, a informação deve ser gerenciada como um produto.

Uma analogia pode ser feita entre qualidade relacionada a produtos e qualidade relacionada à informação. A fabricação de um produto pode ser vista como um processamento de matéria-prima para gerar produtos físicos. Analogamente, a geração de informação pode ser vista como um processamento de dados brutos para a fabricação de produtos de informação (*Information Product - IP*).

A metodologia proposta para TDQM é baseada em pesquisas recentes e experiências práticas. Alguns dos conceitos básicos da metodologia são o ciclo de TDQM e produtos de informação. O ciclo de TDQM consiste em definir, medir, analisar e melhorar continuamente a qualidade da informação e é essencial para assegurar produtos de informação de alta qualidade. Este ciclo foi desenvolvido a partir de uma adaptação do ciclo de Deming [Dem86], largamente difundido e usado na literatura de TQM (planejar, fazer, verificar e agir). A avaliação repetida deste ciclo pode ser entendida como o uso constante de processos de controle e de garantia de qualidade até chegar-se ao nível de qualidade desejado pelo usuário.

Para a realização dos processos de controle e garantia de qualidade na fase de conversão, mais uma vez é importante definir parâmetros que possam ser utilizados. Normalmente, escolhe-se alguns dos parâmetros identificados na fase de coleta. Várias abordagens podem ser encontradas na literatura; entretanto, qualquer que seja a abordagem, haverá sempre uma divisão em dois grandes grupos básicos a serem analisados: os aspectos que afetam a qualidade do componente geográfico do dado e os aspectos que afetam a qualidade do componente não geográfico.

### 3.4.1 Parâmetros segundo Montgomery&Schuch [MS93] e Hohl [Hoh98]

Segundo Montgomery&Schuch, as discussões sobre qualidade durante o processo de conversão envolvem a análise de dois tipos de parâmetros: qualidade cartográfica e qualidade de informação. De forma semelhante, Hohl identifica dois parâmetros principais a serem

analisados no processo de conversão: precisão posicional e precisão de atributo. As considerações a seguir englobam as opiniões dos dois trabalhos.

### **Qualidade cartográfica ou precisão posicional**

A qualidade cartográfica, ou precisão posicional, envolve os seguintes parâmetros: precisão relativa, precisão absoluta e qualidade gráfica.

**Precisão relativa** A precisão relativa expressa, em unidades de distância, o desvio máximo entre o intervalo entre dois objetos em um mapa e o correspondente intervalo entre os mesmos objetos no campo. Normalmente, quando se opta por precisão relativa não se considera a precisão absoluta.

**Precisão absoluta** A precisão absoluta, por sua vez, é entendida como sendo o desvio máximo entre a localização aonde a feição é mostrada no mapa e sua posição real sobre a superfície da Terra. A precisão absoluta é normalmente expressa em faixas de distância.

A precisão dos dados em um mapa base dá o limite superior para a precisão posicional de todo o banco de dados geográfico.

**Qualidade gráfica** A qualidade gráfica refere-se a uma medida de legibilidade dos dados, consistência das representações gráficas, estética e conformidade com os padrões gráficos normalmente requeridos pela aplicação. Pontos, linhas e áreas devem ser convertidos corretamente para utilizar simbolização apropriada, estilo de linha, preenchimento de área, localização de texto e seleção de fonte.

Legibilidade é a clareza com que os dados são apresentados e um aspecto de qualidade que deve ser definido no início do esforço de conversão de dados. Pelo fato de os dados em SIG poderem ser representados em várias escalas a definição de legibilidade deve incluir as escalas onde os dados sejam legíveis.

Simbolização correta é também parte da especificação do projeto de uma aplicação. Esta definição inclui símbolos utilizados para representar cada feição, fontes e seus tamanhos, cores e outras características que devem ser exibidas em diferentes escalas. Por exemplo, pode ser importante mostrar válvulas de água em uma escala 1:100 mas na escala de 1:2000 as mesmas válvulas podem criar distração ou até mesmo ocultar outras feições mais importantes.

### **Qualidade de informação ou precisão de atributo**

Montgomery&Schuch consideram ao menos quatro características como componentes da qualidade de informação, sendo a precisão de atributo uma delas:

**Completude** A completude pode ser entendida como uma medida do quanto as feições foram incluídas no banco de dados como resultado da conversão. Por exemplo, um mapa em que se esperam 200 feições está 75% completo se mostrar apenas 150.

Uma especificação típica de completude de conversão concernente à completude pode ser: "Todas as facilidades representadas nas fontes a serem convertidas devem ser capturadas. Não mais que 1% dos objetos e atributos, por mapa, podem ser perdidos a partir dos dados originais". Esta especificação é um simples exemplo; critérios de aceitação são aplicados à parte gráfica, aos atributos, às anotações, à conectividade, à topologia etc. Na maioria dos casos, não é possível conseguir uma fonte de dados 100% completa em relação ao mundo real porque este muda diariamente e em alguns casos a cada hora. A completude deve ser considerada sempre em relação à fonte de dados e não ao mundo real.

**Precisão de atributo** Um mapa que mostre uma plantação de cana-de-açúcar como sendo de café pode estar correto quanto à parte posicional mas o tipo de objeto estará obviamente incorreto, não tendo portanto precisão de atributo. Tipicamente, tanto equipes de conversão quanto usuários executam métodos de validação automática para assegurar que valores de atributo tenham um nível especificado de precisão.

Uma especificação típica em relação à precisão de atributo para os dados convertidos pode ser: "Todas as facilidades devem estar em conformidade com as especificações gráficas e relacionamentos do banco de dados. Não mais de 1% das facilidades e atributos, por mapa, podem ser incorretamente representados".

Um dos problemas de conversão quanto a precisão de atributo se refere a imagens de sensoriamento remoto e fotos aéreas em geral, em especial para dados ambientais (vegetação, uso do solo). Os dados relativos a um levantamento aerofotogramétrico, satélite ou radar precisam ser validados a partir de dados locais em pontos predefinidos. O processo de interpretação é baseado na extração de informações baseadas na forma, tamanho, padrão, sombra, tons de cinza ou cor e textura [Ber92].

A definição dos valores de atributos de dados de sensoriamento remoto é ainda mais delicada porque depende da localização para a sua especificação. Por exemplo, duas imagens de satélite que contenham regiões com as mesmas características espectrais (e portanto mesmos valores de atributos em termos de bandas) podem se referir a valores de atributo distintos do mundo real se as imagens foram capturadas de lugares distintos ou em diferentes épocas do ano. As bandas espectrais são combinadas automaticamente e recebem valores de cor e textura definidas por usuários, resultando em imagens coloridas, que a seguir são fotointerpretadas para determinar os valores de atributos a serem armazenados no banco de dados. O erro de atributo durante a conversão pode ocorrer neste caso tanto no processamento das bandas espectrais quanto na fotointerpretação - é

comum por exemplo, confundir tipos de plantação.

A qualidade de atributo deve ser medida em relação aos documentos fonte usados para a conversão dos dados. Onde a informação for inexistente o valor de atributo é definido como nulo, usando um valor *default* ou em branco. A forma de manusear dados inexistentes deve ser definida durante a fase de projeto. Normalmente o uso de valores *default* é preferível ao uso de campos em branco para propósitos de controle de qualidade.

**Atualidade** A atualidade pode também ser entendida como uma medida de precisão. Dentro de um contexto específico de conversão os dados podem ser correntes (atuais) ou estarem dentro de uma época específica. Após a conversão, a manutenção da atualidade torna-se responsabilidade do usuário.

**Integridade** A integridade do banco de dados é essencial para demonstrar e confirmar se os dados são úteis. Especificações de conversão de dados normalmente indicam que erros não são aceitáveis nesta categoria. Este critério é normalmente descrito em detalhes nas especificações do processo de conversão de dados.

### 3.4.2 Controle e garantia de qualidade

Hohl define dois novos processos dentro do processo de conversão: controle e garantia de qualidade [Hoh98]. Na verdade, controle e garantia de qualidade são conceitos bem estabelecidos na literatura sobre qualidade. Contudo, garantia de qualidade, neste contexto parece ter um significado distinto.

#### Controle de qualidade

Controle de qualidade é um processo que monitora a qualidade de dados que serão armazenados no banco de dados e atua de forma corretiva para assegurar que os dados atinjam os padrões de qualidade definidos no início do projeto. A qualidade dos dados obtida nesta fase é apenas parte do processo maior de conversão. O ideal é prever os possíveis erros e planejar métodos para corrigi-los. Além da identificação e correção de erros, o processo de controle de qualidade deve verificar se as correções foram efetuadas de forma correta.

Pode-se dividir o processo de controle de qualidade em duas fases: o controle efetuado através de *software* e o controle manual. Na primeira fase, o *software* verifica os dados convertidos para assegurar que os valores dos atributos são aceitáveis, de acordo com uma tabela predefinida de valores válidos. O *software* também pode verificar relacionamentos entre atributos. Na segunda fase, é feita a verificação da correspondência entre os dados

convertidos e os dados fonte, para verificar se os dados foram convertidos corretamente conforme, por exemplo, os indicadores da Seção 3.4.

Para um efetivo controle de qualidade os seguintes aspectos devem ser observados:

**Erros de transcrição** Para evitar erros na transcrição dos dados, a mesma pessoa que os coletou deve, se possível, efetuar a transcrição para o banco de dados. Caso contrário, será necessário incluir também uma etapa para verificar se os dados foram corretamente inseridos.

**Tarefas estruturadas e rotatividade** Outro aspecto importante é que o trabalho seja estruturado de forma que a equipe que faz a conversão fique especializada em cada tarefa desempenhada. A rotatividade de pessoal é também um fator de controle de qualidade pois expõe os membros da equipe a todo o processo de conversão.

**Realimentação** A realimentação sobre o processo de controle de qualidade deve ser estimulada sempre que possível para os membros da equipe fornecerem sugestões criativas e colaborem para a melhoria do mesmo.

**Verificação independente** Membros da equipe que fazem a conversão não devem verificar a qualidade de seu próprio trabalho. A revisão do trabalho deve ser feita por um outro membro da equipe.

### Garantia de qualidade

A garantia de qualidade é a verificação final dos dados convertidos antes de serem armazenados no banco de dados geográfico. Os dois principais objetivos da garantia de qualidade são monitorar o nível de qualidade final dos dados convertidos e assegurar que todo o processo de controle de qualidade esteja sendo desenvolvido apropriadamente.

Assim como no controle de qualidade, no processo de garantia de qualidade os detalhes devem ser conhecidos desde o início do processo de conversão. A garantia de qualidade também pode incluir verificações automáticas e manuais. A diferença básica entre controle e garantia de qualidade automático ou manual é que nesta última a verificação manual é normalmente efetuada através de amostragens.

Nesta fase do processo de conversão, o nível de qualidade é definido como sendo o número de erros permitidos para que o conjunto de dados não seja rejeitado. Se possível, o software de controle de qualidade deve ser usado como parte do processo de garantia de qualidade. Isto pode assegurar que nenhum erro foi introduzido desde a última etapa de verificação do controle de qualidade. Verificações feitas por software normalmente

têm uma boa relação custo-benefício e portanto é interessante aplicá-las a todos os dados convertidos e não somente a amostras como no caso da garantia de qualidade manual.

Erros aleatórios identificados durante o processo de garantia de qualidade não são significativos se não indicarem uma tendência e não excederem um total de erros permitidos e predefinidos. Embora, idealmente, todos os erros encontrados devam ser corrigidos. Todos os conjuntos de dados que contenham erros além do valor permitido devem ser reprovados durante o processo de garantia de qualidade. A reprovação na etapa de garantia de qualidade indica falhas nas etapas anteriores do controle de qualidade.

## 3.5 Metadados

Todas as informações sobre os processos anteriores (escolha das fontes de dados, coleta e conversão), interferem no julgamento da qualidade final dos dados armazenados em um banco de dados geográfico. Como a qualidade depende do uso pretendido, e um banco de dados pode ser usado por muitas aplicações, não basta a garantia de qualidade na coleta e conversão. É preciso, também, fornecer a futuras aplicações informação sobre os diversos fatores que afetam a qualidade para que os usuários possam avaliar a qualidade do conjunto de dados e sua possível adequação ao uso pretendido. O uso de metadados facilita a realização desta tarefa.

A Tabela 3.3, identifica as principais dimensões de qualidade de dados que são normalmente apresentadas aos usuários por intermédio de metadados (segundo será verificado na Subseção 3.5.3), em comparação com os critérios/indicadores identificados por Aronoff (vide Tabela 2.1) e com os parâmetros identificados por Montgomery, Schuch e Hohl (vide Tabela 3.2) para os processos de coleta e conversão.

### 3.5.1 Definição e classificação de metadados

Metadados são normalmente definidos como “dados sobre dados” [KS95]. Metadados vêm sendo usados em grandes bancos de dados, inclusive geográficos, para auxiliar a descrição, operação e armazenamento dos dados; e para facilitar a interoperabilidade entre SIG [KPD<sup>+</sup>95].

É cada vez mais crescente o número de trabalhos que têm utilizado metadados para a descrição de dados em SIG. Não existe, porém, um padrão único que seja capaz de atender a todos os aspectos necessários ao sistema, principalmente se forem considerados campos de conhecimento específico, o que é bastante comum em aplicações destes sistemas [Fag99]. Isto contribui para que a pesquisa na área de metadados, em banco de dados, se torne cada vez mais alvo de congressos especializados.

Nomenclatura	Critérios [Aro95]	Parâmetros coleta [MS93]	Parâmetros conversão [MS93, Hoh98]	Dimensões
Precisão posicional	X	X	X	X
Precisão de atributo	X	X	X	X
Consistência lógica	X	-	-	X
Completude	X	X	X	X
Linhagem	X	-	-	X
Cobertura	-	X	-	-
Atualidade	X	X	X	-
Credibilidade	-	X	-	-
Validade	-	X	X	-
Confiança	-	X	-	-
Conveniência	-	X	-	-
Condições físicas	-	X	-	-
Legibilidade	-	X	X	-
Precedência	-	X	-	-
Resolução	X	-	-	-
Acessibilidade	X	-	-	-
Custo	X	-	-	-

Tabela 3.3: Tabela comparativa entre critérios e parâmetros de qualidade identificados e dimensões normalmente informadas em metadados.

O uso de metadados em SIG é de natureza recente [Rad91]. Em termos de consequências práticas, cada vez mais os metadados têm sido incorporados a SIG comerciais embora ainda de uma maneira bastante rudimentar. O uso ideal de metadados, entretanto, excede esta forma rudimentar de apresentação que existe nos SIG comerciais. Uma tendência atual é a coleta de informações mais detalhadas sobre o conteúdo dos dados. Outras tendências no gerenciamento de metadados incluem: a descrição, em mais detalhes, tanto da história quanto da qualidade de conjuntos de dados e de suas fontes; a inclusão de mais elementos espaciais e a descrição de modelos e algoritmos.

Uma dimensão de metadados ainda pouco explorada é seu uso como forma de avaliar a qualidade dos dados. Para [Har98] existe uma conexão essencial entre qualidade de dados e o uso de metadados. Em muitos casos, padrões de metadados incluem informações explícitas sobre qualidade, enquanto que em outros casos a qualidade pode ser inferida diretamente a partir do conteúdo dos mesmos. No primeiro bloco incluem-se os padrões [FGD98b], [USG98] e [Gro99], dentre outros, e as propostas de [Gup98] e [Fra98]. No segundo bloco incluem-se outros padrões de metadados espaciais que incorporam um menor número de informações sobre qualidade de dados, normalmente de forma textual e referindo-se basicamente à descrição do processamento histórico dos dados (o GCMD da

NASA - EUA e o IBRA - *Interim Biogeographic Regionalisation of Australia, Land Cover e Marine Protect Areas* são exemplos disto) [NAS98, On199].

Podemos identificar, desta forma, quatro grandes vertentes de estudos sobre metadados como forma de possibilitar um ganho de qualidade para bancos de dados geográficos: padrões de metadados; informações de qualidade como parte de metadados de forma explícita ou implícita; produção de conjuntos de metadados geográficos para áreas específicas do conhecimento; e avaliação da qualidade dos próprios metadados.

### 3.5.2 Padrões de metadados

A utilização de padrões de metadados é plenamente justificada uma vez que auxilia o tratamento de dados heterogêneos, facilitando o acesso e a atualização destes dados. Assim, o uso de metadados possibilita uma maior consistência, uniformidade e qualidade dos dados, além de melhorar a integração das atividades e reduzir as perdas de informação [Lum96].

Um grande esforço vem sendo desenvolvido para a padronização de metadados geográficos. Nos EUA, por exemplo, foi desenvolvida, a partir de 1980, uma infra-estrutura nacional de dados espaciais (NSDI - *National Spatial Data Infrastructure*) com o intuito de abolir incompatibilidades entre formatos internos como o DLG (do *U. S. Geological Survey*), o DIGEST (da *Defense Mapping Agency*) e o DX-90 (do *National Ocean Service*). Além disso, em um nível mais elevado, os objetivos principais deste esforço podem ser vistos como o desenvolvimento de capacidade para [CCH<sup>+</sup>96]:

1. achar conjuntos de dados espaciais e identificar usos apropriados deste conjuntos de dados;
2. transferir ou mover dados sem perda de conteúdo ou estrutura; e
3. coletar dados para suportar múltiplos propósitos.

Nos EUA criou-se o padrão *FGDC Content Standard for Digital Geospatial Metadata* [FGD98b] a ser adotado por todas as agências governamentais, contemplando o objetivo (1). Outro padrão, o SDTS (*Spatial Data Transfer Standard*) [USG98] contempla o objetivo (2).

Apesar do SDTS e do *FGDC Content Standard* referirem-se a metadados como “dados sobre dados” eles têm funções distintas: o STDS é uma linguagem para comunicação de dados espaciais sem perda de informação enquanto o *FGDC Content Standard* especifica o tipo de metadado que as agências federais devem armazenar para o conjunto de dados que elas mantêm. A única similaridade entre os dois padrões está no fato de possuírem duas seções em comum: a que concerne a qualidade dos dados e uma outra que consiste basicamente de um dicionário de dados [Gun98].

O trabalho de [MM98] discute este compromisso que existe entre qualidade e quantidade de metadados, afirmando que apesar da riqueza das informações que podem ser catalogadas em um metadado, o que é essencial é saber a medida certa de informação para atender o usuário sem sufocá-lo e sem exigir que ele tenha grande conhecimento em SIG. A idéia geral é que encontrar o metadado correto é mais importante que prover grande quantidade de metadados. Os autores concluem que existem elementos comuns nas diversas categorias de metadados avaliadas e que novas tecnologias como, por exemplo, a mineração de dados, aparentam ser uma solução para minimizar os custos associados à criação de metadados.

### 3.5.3 Qualidade como parte de metadados

Alguns padrões de metadados incluem, além de outras informações, informações sobre qualidade de dados em suas estruturas.

O padrão mais completo neste sentido é o FGDC [FGD98b]. A parte que se refere à qualidade de dados inclui informações sobre cinco dimensões de qualidade de dados já citadas nesta dissertação, dispostas nos seguintes campos: linhagem; precisão posicional; precisão de atributo; consistência lógica; e completude. Além destes campos o padrão FGDC possui ainda um campo opcional denominado “cobertura de nuvens” que expressa a área do conjunto de dados obstruída por nuvens, como uma porcentagem da extensão espacial, para imagens de satélite. A Figura 3.2 apresenta a parte deste padrão FGDC que se refere à qualidade.

O campo linhagem descreve a fonte de dados, as fontes auxiliares, os métodos de derivação, as transformações ocorridas e outros processamentos históricos. Os campos precisão posicional e precisão de atributo correspondem às definições já dadas neste capítulo. O campo consistência lógica refere-se à qualidade de codificação dos relacionamentos na estrutura dos dados espaciais. Por fim, o campo completude inclui informações sobre a área geográfica e a cobertura [Gun98].

O campo linhagem pode incluir uma descrição das fontes de dados e dos métodos de derivação, incluindo todas as transformações envolvidas para a produção dos arquivos finais. A descrição deve incluir as datas de coleta do material fonte e as datas das informações auxiliares usadas para atualização. No caso de não se conhecer a data de coleta, deve-se usar a data de publicação. Se os dados foram criados através da fusão de fontes distintas, a fusão deve ser descrita de forma detalhada para possibilitar a identificação da fonte para cada elemento. Neste campo deve-se descrever também as transformações matemáticas de coordenadas usadas em cada passo desde o material fonte até o produto final. Os métodos usados para efetuar as transformações devem ser documentados.

O campo precisão posicional deve incluir o grau de conformidade com um padrão de registro espacial. As informações de transformações formam uma parte do relato de



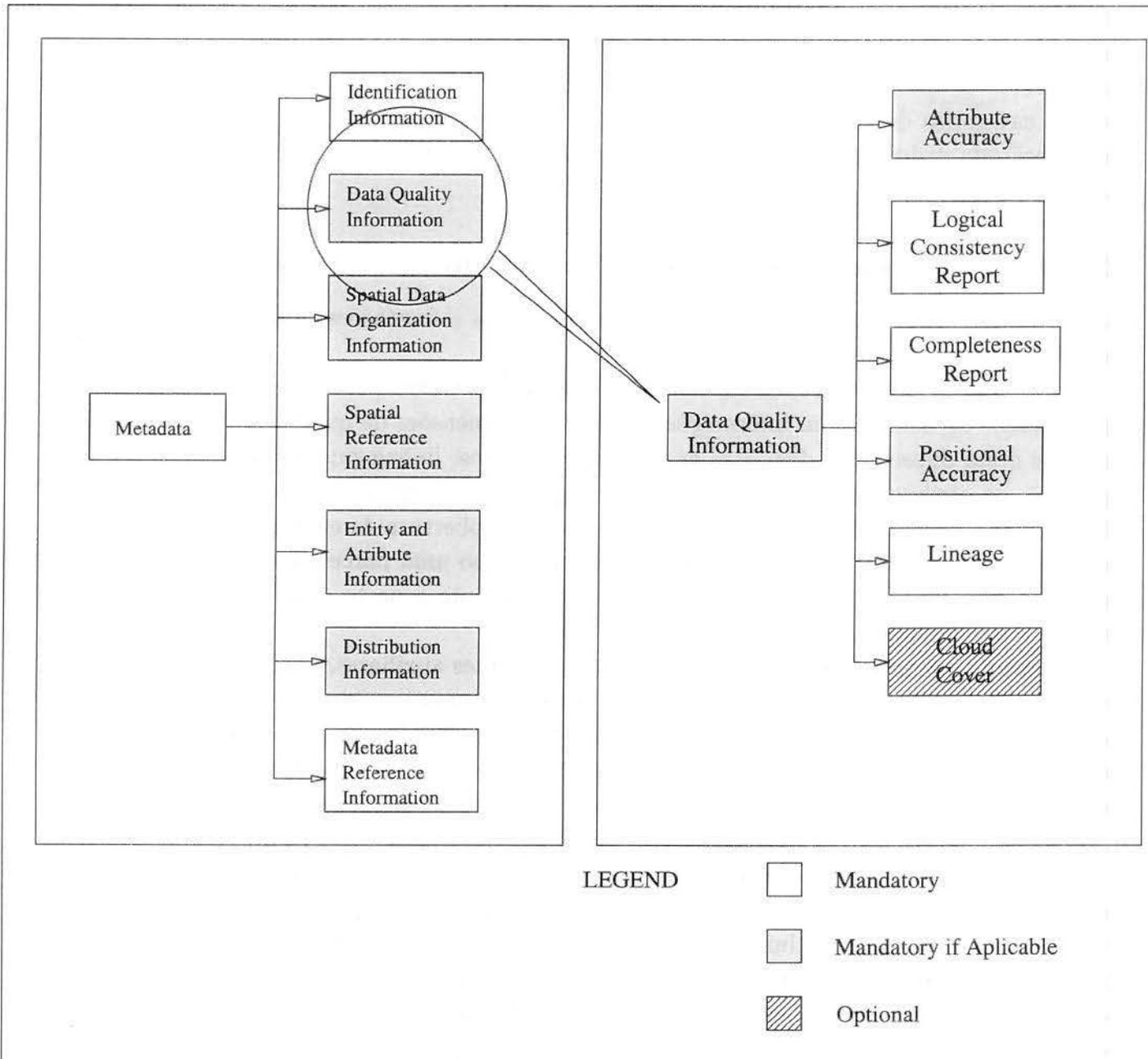


Figura 3.2: Informações de qualidade de dados do padrão FGDC [FGD98b].

qualidade descrito no campo linhagem. Relatórios de qualquer teste de precisão posicional devem incluir a data do teste. Medidas de precisão posicional podem ser obtidas por algum dos métodos a seguir, de acordo com [USG98]: avaliação dedutiva - qualquer declaração dedutiva baseada no conhecimento de erros em cada passo da produção; evidência interna - realização de testes exaustivos efetuados pelo setor de controle; comparação com a fonte; ou utilização de uma fonte independente de maior precisão.

O campo precisão de atributo deve ser relacionado a uma avaliação de precisão, medida em uma escala contínua, que deve ser desenvolvida utilizando procedimentos semelhantes àqueles usados para precisão posicional. O resultado de testes de precisão de atributo devem incluir a data do teste e os materiais usados. No caso de diferentes datas, deve ser descrita a razão da mudança do fenômeno classificado. Segundo o padrão SDTS [USG98] podem ser realizados testes de precisão por algum dos seguintes métodos: avaliação dedutiva - testes baseados em amostras independentes; testes baseados em *overlay* de polígonos; e estimativa, mesmo uma suposição baseada na experiência, explicando a base usada para a suposição. Declarações como “bom” ou “pobre” devem ser explicados de uma maneira tão quantitativa quanto possível.

O campo consistência lógica deve descrever a fidelidade dos relacionamentos codificados na estrutura de dados do dado espacial. Deve detalhar os testes realizados, os seus resultados e as respectivas datas. Os testes podem incluir: testes de valores válidos; testes gerais para dados gráficos; e testes topológicos específicos.

O campo completude deve incluir o critério de seleção, definições utilizadas e outras regras de mapeamento relevantes. Valores limites tais como área mínima ou largura mínima devem ser relatados. O relatório de completude deve descrever o relacionamento existente entre os objetos representados e o universo abstrato de todos os objetos.

Outros padrões de metadados espaciais incorporam um menor número de informações sobre qualidade de dados, nem sempre identificadas de forma explícita, mas que possibilitam, de alguma forma, uma avaliação de qualidade pelo usuário mais experiente. Estas informações normalmente são apresentadas de forma textual e referindo-se basicamente à descrição do processamento histórico dos dados.

### 3.5.4 Produção de metadados geográficos

A criação de uma coleção de metadados é normalmente um processo manual, o que gera um grande problema: diferentes concepções do conceito de qualidade entre “produtores” e “consumidores” além de diferentes avaliações de qual seu nível mínimo aceitável. Isto acontece principalmente pelo fato das pessoas mais qualificadas serem os produtores de dados, que usualmente não têm grande motivação em desempenhar este processo de criação de metadados pois sabem exatamente como fazer para obter e interpretar os dados desejados.

Outro aspecto importante é que as descrições de qualidade de dados são, normalmente parte destes metadados que o produtor torna disponível aos usuários. Segundo [Fra98] isto pode acarretar grandes distorções pois os metadados podem espelhar um padrão de qualidade que não é o desejado pelo usuário. Além disto há o agravante que a qualidade é assegurada apenas pelos produtores.

Basicamente, duas abordagens são encontradas na literatura para solução deste problema: uma solução hierárquica que é tipicamente usada em organizações governamentais e que consiste na criação de um *framework* que requeira o atendimento a padrões estabelecidos por um órgão de controle; e uma solução mais voltada ao usuário que inclui, inclusive, a extração automática de metadados (técnica ainda pouco desenvolvida mas que começa a tornar-se importante).

É crescente o número de abordagens que criticam a adoção de uma solução puramente hierárquica e pode-se notar uma tendência na adoção de soluções mais voltadas para os usuários.

O trabalho de [Fra98] por exemplo, sugere que as atuais descrições de qualidade em metadados não são convenientes para uma tomada de decisão se um certo conjunto de dados é apropriado ou não para uma dada tarefa. De acordo com o autor, as descrições de qualidade de dados são freqüentemente apresentadas apenas como linhagem. Esta descrição, é normalmente bastante detalhada, mas deixa a responsabilidade da tomada de decisão de "*fitness for use*" a cargo do usuário, o que em teoria é ideal. No entanto, a descrição de qualidade apenas como linhagem não é suficiente para dar suporte ao usuário para que este tome uma decisão fundamentada. É preciso que os metadados proporcionem um maior suporte, aumentando o número de parâmetros de qualidade de dados para facilitar a decisão do usuário.

Outra dificuldade é que a descrição de qualidade de dados como linhagem não pode ser operacional, ou seja, não possui procedimentos padronizados que possam ser usados para determinar valores de qualidade de dados. Por exemplo, conjuntos de dados com características bastante semelhantes mas resultantes de diferentes processos de coleta podem ter descrições de linhagem bem diferentes.

Além de tudo isso, as descrições de qualidade de dados deveriam ser, sempre que possível, mensuráveis para que o usuário possa decidir sobre a utilização do conjunto de dados para uma tarefa específica. Entretanto, como já foi visto anteriormente, isto nem sempre ocorre.

O uso de metadados é também uma importante maneira de possibilitar a avaliação da integração entre dados provenientes de diferentes fontes. [Gup98] apresenta alguns dos problemas que ocorrem em termos de qualidade de dados, quando informações geográficas são produzidas por vários produtores e sugere o uso de um *framework* de dados geográficos como forma de integrar estes dados coletados por vários produtores. Este *framework*

seria montado a partir da coleta e análise de inúmeros dados geográficos característicos e viria a favorecer a colaboração entre produtores de dados, estimulando organizações a contribuir com sua construção e manutenção. Para Guptill, a chave para a construção deste *framework* está na definição de qual o melhor dado disponível e encontrá-lo significa ter um conjunto de padrões, elementos e medidas de qualidade de dados.

Embora seja uma forma usual de apresentar avaliações sobre a qualidade dos dados, há críticas à adoção deste método. [Har98], por exemplo, questiona se a apresentação das medidas e indicadores de qualidade como metadados é a melhor maneira de manter os usuários informados sobre qualidade. Sua argumentação é que, freqüentemente, estes metadados são de difícil acesso. Sua proposta de melhoria inclui algoritmos que permitem incorporar qualidade em um nível mais baixo de estruturas de dados. Para [Har98] o grande salto para melhoria de qualidade em dados geográficos está em tornar os metadados mais acessíveis. Para ele, metadados importantes são freqüentemente subtilizados porque os usuários não podem encontrá-los facilmente, entendê-los e navegar por suas complexas estruturas. Ele considera duas abordagens para a solução deste problema: embutir os metadados diretamente nos conjuntos de dados geográficos e prover ferramentas para navegar mais facilmente nos metadados já existentes.

Embutir os metadados é uma estratégia que requer reformulação de estruturas de dados e algoritmos em SIG. Melhorar a navegação, por outro lado, sugere uma maneira potencial de tornar conjuntos de metadados existentes mais acessíveis e úteis para grandes grupos de usuários.

Para [BBS97], a extração automática de metadados de conjuntos de dados é possível apenas para certos tipos de metadados como, por exemplo, tamanho de arquivo. A análise, a síntese e a representação de conteúdos chave em elementos de metadados tais como relatos de qualidade de dados são, entretanto, melhor gerados por especialistas.

### 3.5.5 Qualidade dos metadados

A idéia de avaliação de qualidade de dados deve ser estendida também aos metadados. Apesar de haver considerável discussão sobre metadados, métodos e ferramentas para testá-los são menos evidentes que o consenso sobre sua utilização como forma de influenciar positivamente a qualidade dos dados. Pelo menos dois níveis de avaliação de qualidade podem ser definidos para metadados. Primeiro, dadas certas necessidades documentadas para a composição de metadados, verificações podem ser conduzidas para testar conformidade. Segundo, a saída resultante do uso dos metadados pode ser avaliada pela utilidade e relevância dos elementos em termos de se eles suportam o propósito e os objetivos do esquema do metadado. O trabalho de [MSM97] faz uma análise de metadados baseada no primeiro nível. Para isso são levantados critérios de avaliação de registros bibliográficos, que têm sido objeto de pesquisa significativa por 20 anos e é traçado um paralelo para

avaliação de metadados.

## 3.6 Critérios de aceitação de dados

Os critérios de aceitação dos dados devem atender ao nível de precisão estabelecido pelo usuário. Pode ocorrer que por falta de dados mais precisos ou pela urgência em se realizar a conversão de dados, se inicie o processo de conversão com fontes de dados que não atendam aos requisitos do usuário. Entretanto, durante o processo de conversão alguns erros e anomalias das fontes podem ser identificados e corrigidos. Isto tende a compensar um pouco os erros inseridos no processo de conversão, melhorando a qualidade dos dados [Hoh98].

Durante a conversão dos dados é importante identificar que atributos serão efetivamente utilizados. Atributos que certamente serão utilizados nas aplicações geográficas devem ser obrigatoriamente convertidos, enquanto que outros atributos, opcionais, podem ser deixados em branco ou preenchidos com valores *default*.

O custo da coleta e conversão de dados também precisa ser considerado quando do estabelecimento do critério de aceitação. Com tempo e recurso suficientes, todos os conjuntos de dados podem buscar atingir a precisão máxima estabelecida mas em muitos casos isto não é necessário de imediato.

### 3.6.1 Medição da qualidade

A especificação da qualidade deve incluir os instrumentos para medi-la. Como visto no capítulo, há vários parâmetros da qualidade de dados que podem ser avaliados durante a coleta e a conversão de dados. Alguns são objetivos, mas a maioria é subjetivo e portanto mais difícil de definir. Em se tratando do componente posicional dos dados geográficos a qualidade é focada na precisão posicional, legibilidade e simbolização. Destes três parâmetros citados, somente a simbolização pode ser medida de forma decisiva: uma feição é apresentada usando o símbolo correto ou não, em função de símbolos predefinidos<sup>2</sup>. Em muitos sistemas, os símbolos e as feições são armazenados de forma conjunta durante o processo de inserção no banco de dados. Assim, embora estes sejam relevantes apenas na fase de apresentação, acarretam um impacto direto sobre os dados.

O grau de precisão posicional atingido durante a conversão de dados é determinado pelas técnicas de coleta de dados e pela precisão dos dados do documento fonte utilizado. Como a o nível exigido de precisão posicional varia de acordo com cada aplicação, este conceito, apesar de poder ser medido de forma objetiva, torna-se subjetivo.

---

<sup>2</sup>Cabe ressaltar que isto não significa que os símbolos sejam apropriados ao correto entendimento. Detalhes sobre a utilização de símbolos serão tratados no Capítulo 4.

A legibilidade é um fator totalmente subjetivo. Em aplicações de SIG, a legibilidade está freqüentemente relacionada à escala e será discutida no Capítulo 4.

### 3.6.2 Registro dos processos de controle/garantia de qualidade

Metadados permitem associar informação sobre qualidade referente aos dados principalmente quanto à coleta. A documentação de qualidade requer informação adicional sobre os procedimentos de conversão e sobre a própria verificação de qualidade.

Cada passo dos processos de controle e garantia de qualidade deve ser arquivado sob a forma de um registro histórico para permitir consultas posteriores. As seguintes informações, segundo Hohl, podem fazer parte deste registro [Hoh98]: a etapa do processo de controle/garantia de qualidade que está sendo registrado; as pessoas que efetuaram as verificações; a data; os resultados que foram verificados, incluindo o número de erros encontrados e quantas feições ou atributos foram verificados; e o nome ou um designador do conjunto de dados que está sendo verificado.

Os softwares de controle/garantia de qualidade devem produzir algum tipo de saída que informe sobre o resultado da conversão de dados. Estes resultados devem ser arquivados, para que o usuário aumente o seu nível de confiança nos dados e conheça o processo de conversão. Os relatórios resultantes de verificação automática devem incluir ao menos as seguintes informações: nome do conjunto de dados verificado; data que o programa foi executado; nome dos técnicos que executaram o programa e corrigiram erros; lista descritiva de cada erro encontrado; e resumo dos resultados.

Quando da identificação de erros durante a conversão, seja durante o processo de controle ou de garantia de qualidade, há normalmente duas abordagens comuns para sua correção: correção imediata ou diferida. Nesta última o erro é corrigido somente quando o conjunto de dados final for estabelecido. A esta fase dá-se o nome de solução de conflitos. Ela é também parte importante nos processos de controle/garantia de qualidade pois efetivamente melhora os dados e agrega valor ao banco de dados geográfico.

### 3.6.3 Documentação dos requisitos de qualidade

Os requisitos de qualidade de dados devem ser definidos durante a fase de projeto do banco de dados. Em particular deve-se usar os seguintes fatores para guiar o processo [Hoh98]:

- que nível de precisão/qualidade pode ser alcançado dentro do objetivo e orçamento do projeto?
- que nível de precisão/qualidade de dados deve ser requerido para atender às futuras necessidades dos usuários finais?

É uma tarefa difícil estabelecer este limite de qualidade. Um erro comum é estabelecer padrões acima do necessário. Isto pode trazer diversos prejuízos, dentre os quais não conseguir atingir o final do projeto por falta de verba, tornar difíceis de atender as crescentes expectativas dos usuários e até mesmo tornar impossível o trabalho de conversão. Da mesma forma, o estabelecimento de um limite abaixo do desejado pode levar a displicência da equipe em relação a erros, insatisfação dos usuários com os novos sistemas, fazendo-os migrarem de volta aos métodos tradicionais e obviamente a tomada de decisão, por parte dos usuários, baseada em informações imprecisas.

A conversão de dados usualmente considera os aspectos de completar o projeto em tempo e dentro do orçamento. Ao contrário, a ênfase em qualidade é, na prática, muitas vezes deixada em segundo plano.

### **3.7 Resumo**

Este capítulo discorreu sobre algumas características das principais fontes de dados e equipamentos utilizados em sua captura (coleta e conversão).

Foram apontadas algumas opções de escolha dentre as possíveis fontes de dados para a construção de um banco de dados geográfico e algumas particularidades, vantagens e desvantagens de cada uma delas.

O capítulo também identificou parâmetros de qualidade a serem utilizados durante as fases de coleta e conversão de dados. Além da identificação dos parâmetros, foram sugeridas algumas linhas de ação para produção de dados dentro de um padrão de qualidade.

Finalmente, discutiu-se sobre a crescente utilização de metadados em aplicações geográficas, como forma de auxiliar a descrição da qualidade dos dados e, desta maneira, facilitar, principalmente, a sua reutilização em outras aplicações. Estes conceitos foram utilizados como base para a ferramenta proposta no Capítulo 5.

## Capítulo 4

# Qualidade e problemas relativos à apresentação

Neste capítulo serão descritos os principais aspectos que envolvem a apresentação de dados geográficos e que podem ter influência sobre o julgamento de qualidade efetuado pelos usuários.

A apresentação dos resultados causa problemas se a linguagem utilizada para comunicação não for clara, as cores e sombreamentos não forem apropriados, a escala de utilização for muito diferente da escala de origem, o tipo de projeção não privilegiar a proporção espacial desejada pela aplicação ou se o resultado apresentado estiver tão cheio de informações que seja difícil ao usuário identificar o que ele deseja. Na apresentação destes resultados, a interface tem um papel fundamental.

Para a maioria dos usuários, a interface é o sistema, ou seja, para eles o comportamento e as funcionalidades do sistema se restringem àquilo que a interface apresenta. Problemas na apresentação se refletem na interpretação e portanto causam erros. Assim, é importante analisar tanto aspectos relativos à apresentação dos dados geográficos quanto aspectos da interface que podem afetar a interpretação da apresentação.

Este capítulo está dividido da seguinte maneira: a Seção 4.1 aborda aspectos da visualização de dados geográficos tecendo considerações sobre a produção cartográfica e a Seção 4.2 aborda aspectos importantes sobre o projeto de interfaces, em especial os que dizem respeito a fatores humanos, à semiótica e à cognição espacial em SIG. A Seção 4.3 resume o capítulo.

### 4.1 Visualização de dados geográficos

A visualização de dados geográficos é bastante complexa. Em aplicações tradicionais, não existem representações múltiplas e os relacionamentos entre os dados são estabelecidos

durante a modelagem dos dados. No caso de sistemas geográficos, uma determinada representação pode ser apresentada de diversas formas a fim de atender aos diversos requisitos do usuário. Além disso, a maioria dos relacionamentos espaciais é calculada a partir do contexto e da representação adotados. O resultado de uma consulta depende da representação sobre a qual é calculada que por sua vez determina como e quais relacionamentos são suportados [CCH<sup>+</sup>96].

A demanda por múltiplas representações em SIG decorre de dois requisitos básicos:

1. da necessidade de variar a representação de fenômenos dependentes da escala, eliminando detalhes e simplificando a aparência e a densidade de objetos; e
2. da necessidade de acomodar diferentes percepções sobre o mesmo fenômeno do mundo real.

Algumas considerações básicas sobre a apresentação de mapas aplicam-se igualmente a SIG. Para [Mon91], são utilizados três elementos básicos para a construção de mapas: escala, projeção e símbolos. Cada um destes elementos pode ser visto como uma fonte de distorções.

#### 4.1.1 Escalas

A representação da superfície terrestre sob a forma de mapas implica na representação de uma superfície muito grande sobre uma outra de dimensões reduzidas. Como visto no Capítulo 2, é necessário reduzir a primeira superfície para que o mapa possa ser reproduzido na segunda. Esta redução importa na idéia de escala. Do ponto de vista teórico, no entanto, o conceito de escala é irreal porque, via de regra, ela é variável em um mesmo mapa de lugar para lugar e não pode ser usada sem restrições. Surge daí a primeira distorção do uso deste elemento básico.

Além disto, em teoria, um mapa digital pode ser apresentado em qualquer escala independente da escala de origem em que os dados foram armazenados. O que ocorre, no entanto, é que se a escala escolhida for menor que a original, alguns detalhes poderão ser perdidos; ao contrário, se a escala escolhida for maior, não haverá um ganho natural destes detalhes. Ou seja, operações de mudança de escala, muitas vezes essenciais para uma dada aplicação, não podem ser implementadas de forma automática sem perda de qualidade.

#### Generalização Cartográfica

Como a apresentação de um mapa em escala 1:1 é em geral inviável, qualquer apresentação cartográfica corresponde a uma redução da realidade, efetuada a partir do que se chama

“generalização cartográfica”. A generalização cartográfica pode ser definida como “a adaptação dos elementos qualitativos e quantitativos a um mapa derivado, por meio da seleção e simplificação de detalhes oriundos do mapa básico original” [Oli88]. Este método cartográfico é erroneamente considerado como executável de forma automática a partir de operações matemáticas de mudança de escala. Muito pelo contrário, a automação completa de procedimentos de generalização raramente é bem sucedida pois estes exigem envolvimento do usuário quanto a decisões de apresentação [CCH<sup>+</sup>96].

Na Cartografia, operações de generalização cartográfica são realizadas com base no conhecimento empírico do cartógrafo, incluindo o seu senso estético. Em SIG, praticamente não existem recursos automáticos de generalização cartográfica, o que limita as possibilidades de compartilhamento de informações entre grupos de usuários, uma vez que os níveis de detalhe e precisão exigidos por uma aplicação podem não ser adequados para outras [JL99].

Segundo [Mon91] um bom mapa tem que suprimir um pouco da verdade para auxiliar o usuário a ver somente o que precisa ser visto. A generalização cartográfica, entretanto, não é usada somente para propiciar uma diminuição de escala sem tornar a apresentação saturada, mas também porque os símbolos utilizados na apresentação normalmente ocupam mais espaço (proporcionalmente) do que as entidades geográficas realmente ocupam na superfície da Terra.

A generalização cartográfica se aplica em dois contextos: no processamento numérico dos dados (definindo valores derivados e agregando dados) e na apresentação (relacionando visualmente as características relevantes) [CCH<sup>+</sup>96].

No contexto do processamento numérico dos dados a generalização cartográfica é realizada através de duas operações essenciais: seleção e classificação. A seleção promove a generalização através da escolha das entidades mais importantes, enquanto a classificação reconhece similaridades entre elas de forma que um único símbolo possa representar o conjunto.

No contexto da apresentação, a generalização varia de acordo com a geometria das entidades geográficas a serem mapeadas. Se considerarmos as representações vetoriais de objetos geográficos (pontos, linhas e polígonos) conforme citado no Capítulo 2 poderemos verificar diferentes operações de generalização. Algumas das mais comuns são ilustradas pela Figura 4.1 adaptada de [Mon91].

As operações de generalização comuns para pontos são: seleção, associação gráfica, agregação, deslocamento, abreviação e conversão para polígono. Para linhas, as operações possíveis são: seleção, deslocamento, simplificação, suavização e realce. Finalmente, para os polígonos são usadas operações de seleção, agregação, simplificação, deslocamento, suavização, realce, dissolução, segmentação, conversão para ponto e conversão para linha.

A operação de seleção consiste na supressão de pontos, linhas ou polígonos para a

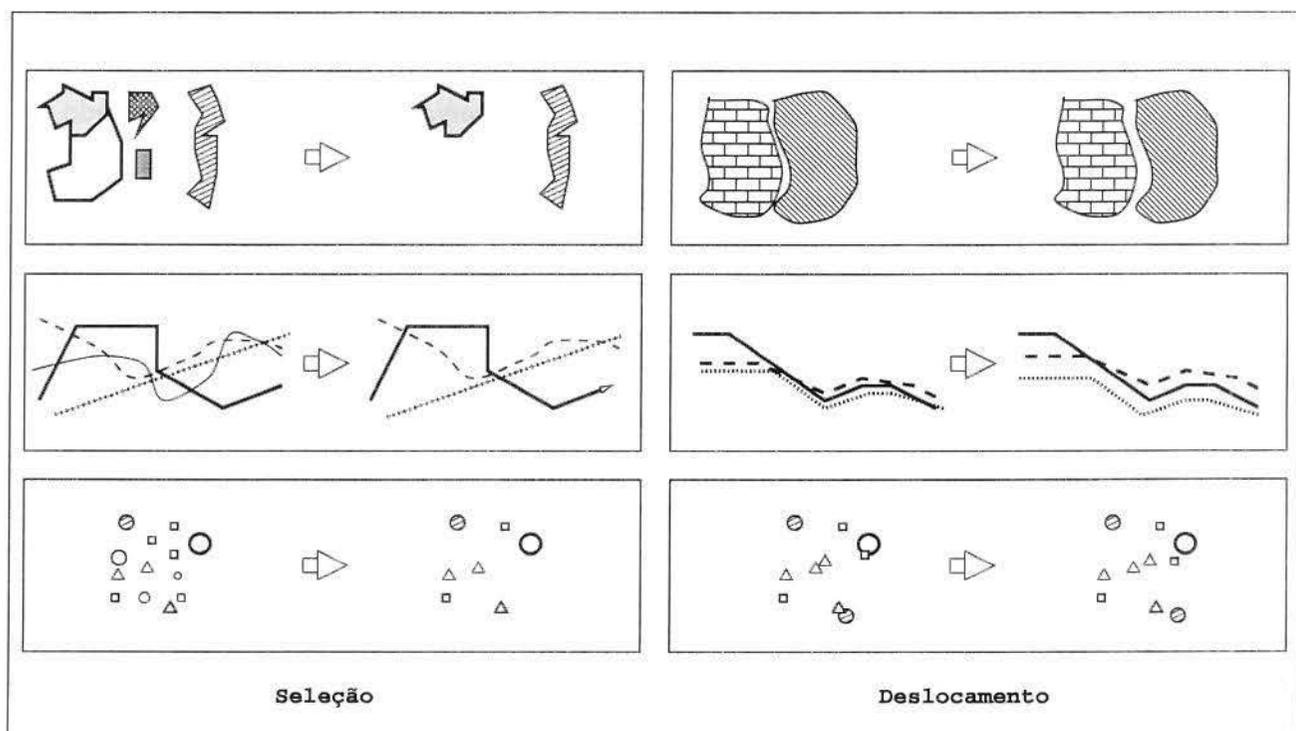


Figura 4.1: Algumas operações de generalização comuns a polígonos, linhas e pontos segundo Monmonier [Mon91].

apresentação no mapa. A associação gráfica é uma operação característica sobre pontos e consiste em associar uma linha fina ou código numérico ao ponto para identificá-lo, evitando ambigüidade no caso de pontos muito próximos. A agregação consiste em juntar vários objetos equivalentes (pontos ou polígonos) em um único que os represente. O deslocamento afasta pontos, linhas ou polígonos para evitar interferência entre eles. A abreviação consiste na diminuição dos nomes das legendas. A conversão para polígono converte vários pontos próximos em um polígono que os represente. A simplificação consiste na redução de detalhes de linhas e polígonos pela eliminação de pontos. A suavização é também uma operação de redução de detalhes de linhas e polígonos, só que feita pela movimentação e/ou adição de pontos. O realce, ao contrário da simplificação e suavização, adiciona detalhes visando uma aparência mais realista. As operações de dissolução e segmentação são operações similares à agregação, com as seguintes particularidades: na dissolução, o polígono que agrega não muda de forma pois o agregado está contido naquele e na segmentação ocorre uma separação de polígonos após a agregação. Finalmente, as operações de conversão para ponto e linha transformam polígonos pequenos ou estreitos em pontos ou linhas, respectivamente.

### 4.1.2 Projeções

As projeções cartográficas transformam as superfícies curvas e tridimensionais do planeta em outras planas e bidimensionais [Mon91]. Assim, são responsáveis pela distorção da realidade geográfica tanto no armazenamento quanto na apresentação das informações. A Figura 3.1 obtida de [Mon91] (vide Capítulo 3) trata a confecção de um mapa como um processo de dois estágios e ajuda a explicar o significado e as limitações que as escalas e as projeções possuem.

O primeiro estágio corresponde à fase de transformação dos dados de acordo com a projeção escolhida e o segundo é a projeção. Para a escolha do sistema de projeção, o ideal é analisar que tipo de propriedade espacial (área, forma e distância, por exemplo) o sistema pode representar melhor e confrontar esta característica com a necessidade da aplicação. A partir desta análise, surge uma outra classificação importante, a que é feita em função das propriedades dos sistemas de projeção em garantir fidelidade às proporções espaciais: áreas (isométricas ou equivalentes), forma (conformes ou isogonais), distância (equidistantes, azimutais ou zenitais) ou nenhuma das anteriores (afiláticas ou arbitrarias) [CCH<sup>+</sup>96, Oli88, Bak65].

Estas propriedades não podem ser atendidas simultaneamente, pois a preservação de uma propriedade normalmente implica na distorção das demais. A solução é, portanto, construir uma carta que, sem possuir todas as condições ideais, possua aquelas que satisfaçam determinado objetivo. A escolha da projeção é, assim, diretamente ligada à carta que se deseja construir [Bak65].

### 4.1.3 Símbolos

Os símbolos gráficos complementam a escala e a projeção na composição de um mapa, tornando visíveis entidades, lugares e outras informações de localização representadas no mapa. Através da descrição e diferenciação das diversas entidades, os símbolos em mapas servem como um código gráfico para armazenar e recuperar dados [Mon91].

Em um mapa, os símbolos normalmente podem diferir em: tamanho, forma, escalas de cinza, textura, orientação e cor. Cada uma destas seis variáveis visuais distingue-se em retratar apenas um tipo de diferença geográfica. Para diferenças quantitativas, por exemplo, o tamanho é normalmente mais eficiente. Forma, textura e cor são normalmente mais eficientes para mostrar diferenças qualitativas. Escalas de cinza, por sua vez, são preferíveis para demonstrar intensidade de ocorrência. A Figura 4.2, obtida de [Mon91], apresenta estas seis variáveis visuais.

De acordo com [CCH<sup>+</sup>96], tamanho, forma e cor são as que permitem um melhor entendimento do resultado de uma consulta pelo usuário.

A variação do tamanho dos componentes de um mapa permite avaliar suas grandezas

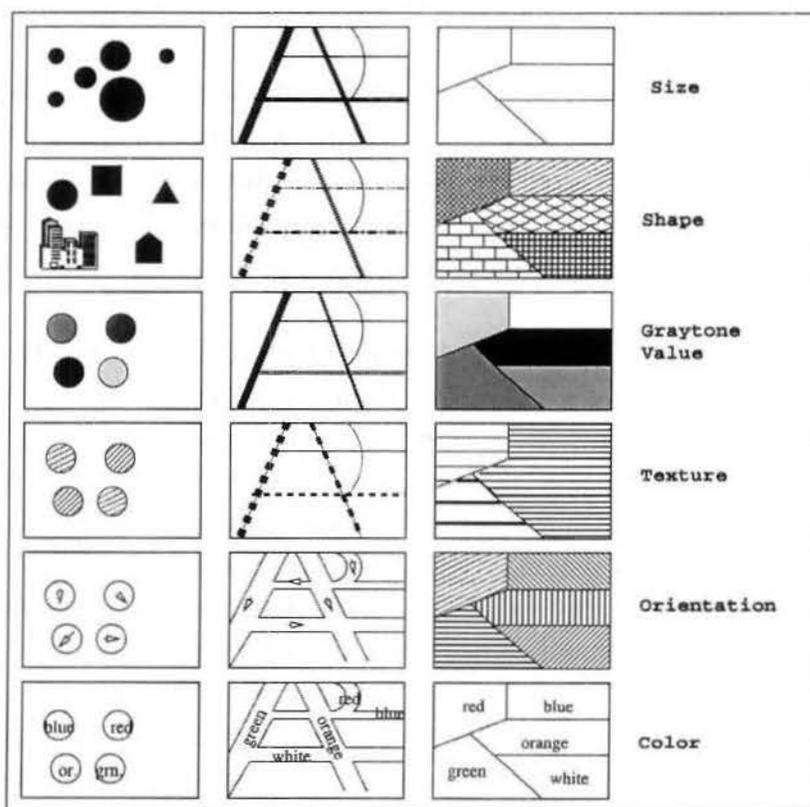


Figura 4.2: As seis principais variáveis visuais segundo Monmonier [Mon91].

relativas e absolutas. Na prática, esta variação é limitada pelas dimensões físicas observáveis pelo olho humano e pelo conjunto de dados apresentados (por exemplo, havendo um grande conjunto de fenômenos a serem apresentados, o uso do tamanho para avaliar a ordem de grandeza relativa de um fenômeno pode inviabilizar a apresentação de outros fenômenos na região). Se um conjunto de cidades é apresentado por meio de pontos (forma) em um mapa, a variação do tamanho destes pontos permite ao usuário visualizar sua importância relativa em termos de população, poder econômico, índice de criminalidade, etc., dependendo do objetivo da apresentação.

Ainda segundo [CCH<sup>+</sup>96], a variável forma só permite ao usuário determinar informação qualitativa, precisando estar associada a outras variáveis (cor, tamanho, textura) para permitir derivar informação quantitativa. Em uma carta náutica, por exemplo, o estilo e espessura de uma linha permitem determinar o tipo de fenômeno apresentado: batimetria, região de visibilidade de um farol ou contorno de ilhas [Oli95].

A cor é a variável que melhor permite separar fenômenos de natureza diferente. Apesar de representar melhor diferenças qualitativas, a variação de tom e contrastes das cores pode possibilitar também a determinação de diferenças relativas a valores de um

fenômeno. A apresentação exige também que haja um contraste suficiente entre cores de regiões contíguas, para permitir determinar seus limites. Por exemplo, em cartas náuticas, diferenças de tons de azul indicam a variação na profundidade [Oli95].

Para informações qualitativas sobre objetos pontuais ou lineares, a apresentação deve ser feita variando a forma, a orientação ou a cor; se as informações são sobre objetos poligonais, varia-se a cor. Se, por outro lado, a informação é quantitativa, varia-se o tamanho se o objeto é pontual ou linear, enquanto que para objetos poligonais varia-se o tamanho, o valor ou a cor.

A utilização de símbolos pode desta forma inserir distorções de duas maneiras: através da utilização de variáveis visuais que sejam inadequadas para transmitir a informação desejada e pelo fato dos símbolos utilizados na apresentação ocuparem mais espaço proporcional no mapa que as entidades de interesse realmente ocupam na Terra. A utilização de variáveis visuais apropriadas está contida na discussão de como os seres humanos entendem os símbolos e pode ser vista como parte do estudo sobre cognição espacial humana, um campo de pesquisa de grande importância para SIG e que vem sendo discutido em diversas conferências internacionais [FK95, HF97].

Em imagens processadas fornecidas pelo INPE (Instituto Nacional de Pesquisas Espaciais), diferentes cultivos recebem tonalidades de verde, para facilitar a associação vegetação - verde. Já no processamento realizado por organismos ligados à agricultura e uso do solo, cultivos recebem tonalidades de vermelho, indicando correspondência com a banda de satélite utilizada.

É importante salientar que textos e símbolos que são claros e fáceis de ler em uma escala podem ser completamente ilegíveis em outra. Ainda assim, pode-se considerar as seguintes questões para ajudar em uma tentativa de estabelecimento de um critério de legibilidade do mapa [Hoh98]: definir todas as escalas de saída e os tamanhos de texto utilizados em cada escala; definir as camadas de dados que irão aparecer em primeiro plano e portanto podem ter permissão para ocultar outras camadas; especificar os locais aonde poderão aparecer setas para indicar qual feição está ligada a que texto sem prejuízo da legibilidade; considerar a omissão de rótulos em áreas congestionadas, estabelecendo uma hierarquia para determinar que texto pode aparecer e em que local em diferentes camadas; e considerar que várias escalas de saída podem ser requisitadas por diferentes usuários e cada usuário pode ter necessidades únicas para visualizar e analisar os dados.

## 4.2 Projeto de interfaces

A diversidade de aplicações e de usuários constitui um grande desafio para o projeto de interfaces. De um modo geral, as interfaces atuais de SIG refletem a arquitetura interna do sistema e o seu modelo de representação de dados que nem sempre é compatível com

o modelo mental do usuário.

O desenvolvimento de interfaces para SIG precisa tratar de uma série de dificuldades oriundas das características do uso de dados geográficos. Oliveira identificou algumas destas dificuldades, dentre elas [Oli97]:

- Ligação entre a interface e o SIG - Existem duas abordagens para solução deste problema: na primeira (sistemas fechados), a interface faz parte do SIG e se beneficia de conhecimentos sobre estruturas físicas de armazenamento e na segunda (sistemas abertos) a interface é um módulo externo que se comunica com o sistema sendo independente de sua implementação.
- Diversidade de linguagens de manipulação de dados - Existe uma grande dificuldade para adaptação de uma interface a diferentes SIG pois cada um implementa suas próprias linguagens, como visto na Seção 2.2. Mais ainda, a diferença não se restringe a aspectos sintáticos, mas inclui também semântica de operações. Além disso, as linguagens são bastante dependentes das representações dos dados.
- Modelo mental do usuário - O projeto de interfaces se baseia em suposições feitas sobre seus potenciais usuários, expressas através de um modelo mental do usuário. Os problemas para a definição de modelos mentais envolvem vários aspectos, que vão desde a definição de conceitos espaciais básicos até a proposta de metáforas adequadas para a representação destes conceitos.

Para [Oli97] o projeto de interfaces deve contemplar, então, três enfoques básicos: arquitetura, linguagens e fatores humanos, ligados às três dificuldades identificadas. A Figura 4.3, obtida de [Oli97], apresenta estes três enfoques do projeto de interfaces.

O enfoque relativo à arquitetura se ocupa em definir a interface funcionalmente, como um conjunto de blocos que se comunicam entre si e com o SIG. Cada bloco é responsável por um conjunto de tarefas. O enfoque de linguagens se preocupa em definir e implementar uma linguagem utilizada pelo usuário para definição e manipulação dos dados. A linguagem possui, neste caso, mecanismos que traduzem comandos do usuário na linguagem proposta (na Figura 4.3 - Le) em comandos para o SIG (na Figura 4.3 - Li). O enfoque em fatores humanos estuda o modelo mental do usuário para, a partir deste, especificar o modelo de interface apropriado, buscando minimizar o esforço do usuário em definir e manipular dados geográficos.

O estudo dos enfoques relativos à arquitetura e às linguagens está intimamente ligado à implementação de um SIG e não será desenvolvido neste texto. O enfoque em fatores humanos, ao contrário, é menos dependente do SIG utilizado e tem uma real influência sobre a qualidade dos dados, do ponto de vista de sua apresentação, sendo por este motivo, tratado a seguir.

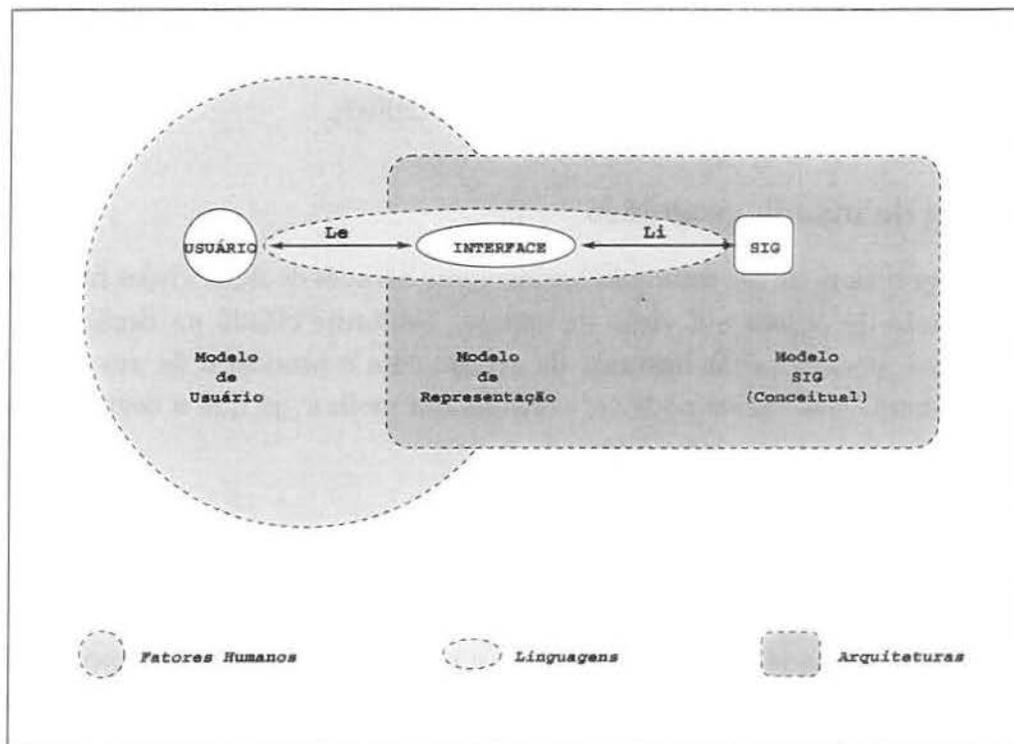


Figura 4.3: Enfoques no projeto de interfaces segundo Oiveira [Oli97].

### 4.2.1 Fatores humanos

Segundo [Oli97], do ponto de vista de fatores humanos, os principais aspectos avaliados para a construção da interface são o uso de metáforas de interface geográfica, o entendimento do mundo geográfico, a utilização e construção de mapas e principalmente a criação de um modelo mental de usuário.

#### Metáfora de interface

Uma metáfora de interface é um mapeamento (matemático) que parte de um domínio de origem e define uma estrutura para um domínio alvo. A escolha da metáfora em SIG especifica: conceitos que o usuário pode manipular, divisão de tarefas entre sistema e usuário e o tipo de comunicação que é adotado entre eles. Grande parte das interfaces propostas para SIG baseia-se na metáfora de mapa. O sucesso desta metáfora em SIG se deve, em grande parte, à popularidade de mapas em papel [Oli97].

Entretanto, mapas em papel tendem a esconder a incerteza e as imprecisões inerentes aos dados geográficos, privilegiando os aspectos de apresentação. Em SIG o objetivo principal é, normalmente, proporcionar conhecimento correto (pressupondo que os dados armazenados possuam algumas das características de qualidade, citadas no Capítulo 2),

enquanto em um mapa o objetivo principal é, em geral, proporcionar conhecimento relevante. Apesar destas diferenças, examinar a interação humana com mapas é ainda fundamental para o entendimento da cognição espacial [BF97].

### **Entendimento do mundo geográfico**

A realidade geográfica pode ser estudada basicamente através de duas visões fundamentais do espaço: a visão de objeto e a visão de campo, conforme citado na Seção 2.2. Estas visões, entretanto, podem variar bastante de acordo com o propósito de uso da realidade geográfica. Nenhuma das visões pode ser considerada melhor, já que a cognição humana utiliza ambas de acordo com o propósito.

### **Utilização e construção de mapas**

A interação com mapas em papel é diferente da interação com mapas em SIG. Um mapa em papel representa uma realidade permanente e tangível. Em SIG os mapas podem ser: visíveis mas não permanentes (mapas na tela) e não diretamente visíveis mas permanentes (representação dos mapas nos bancos de dados). O processo de construção de mapas é uma atividade que tem sido objeto de pesquisa por muito tempo. O processo de uso computacional de mapas, por outro lado, é objeto de pesquisa recente [Oli97].

### **Modelo mental do usuário**

Um modelo mental do usuário (ou simplesmente modelo de usuário) representa o conjunto de hipóteses sobre o comportamento dos usuários e sobre a sua forma de raciocinar. Em se tratando de SIG, pode ser entendido como uma representação explícita das propriedades de um certo usuário, em função de uma aplicação geográfica. Dada a diversidade de aplicações e particularidades dos usuários de uma mesma aplicação, [CCH<sup>+</sup>96] conclui que um SIG deveria ter uma interface diferente para cada usuário o que obviamente seria bastante complexo. Uma recomendação próxima para esta solução é a idéia da criação de um banco de regras que descreveria o perfil do usuário direcionando-o para a apresentação mais adequada [CCH<sup>+</sup>96, Oli97].

## **4.2.2 Cognição espacial em SIG**

Apesar de tratadas sob a ótica do projeto de interfaces, por serem mais comuns nesta área, as considerações sobre cognição espacial afetam também e principalmente, a visualização dos dados geográficos e por conseguinte a qualidade das conclusões obtidas desta visualização. A pesquisa sobre este tópico é inerentemente multidisciplinar, envolvendo áreas

como a psicologia, a lingüística e a filosofia, só para citar algumas, além de geografia, cartografia e ciência da computação [FK95].

Várias idéias a respeito da percepção humana sobre o espaço geográfico, algumas das quais são citadas a seguir, foram sintetizadas por [Ege95] mas ainda precisam ser analisadas para possibilitarem um real auxílio no desenvolvimento de SIG e um ganho de qualidade na interpretação dos dados geográficos: apesar de tridimensional, a interpretação humana mais comum sobre o espaço geográfico é a de que ele é bidimensional; as pessoas normalmente não fazem considerações sobre a curvatura da Terra; existe uma íntima ligação entre espaço e tempo; a informação geográfica é freqüentemente incompleta; as pessoas utilizam múltiplas conceituações para o espaço geográfico, com múltiplos níveis de detalhe; as informações topológicas são consideradas de maior importância que as métricas; e as distâncias não são simétricas.

Embora não consensual, um estudo de [Mon95] concluiu que muitos aspectos da cognição espacial podem ser considerados universais pois a variação em função de diferenças culturais não se mostrou substancial.

Embora haja um grande interesse em automatizar o processo de construção de mapas, desde a aquisição dos dados até a geração dos mapas propriamente ditos, produzir mapas com a qualidade de apresentação conseguida pelos especialistas humanos em Cartografia ainda é uma tarefa sem solução para SIG. Mesmo que isto ocorra, o “leitor” do mapa jamais será liberado da tarefa de interpretação [BF97].

A “leitura” dos significados em mapas vai além da identificação literal das entidades geográficas e das relações apresentadas entre elas. Assim como diferentes usuários de um mapa podem “ler” coisas diferentes, a representação de uma entidade geográfica pode ser interpretada de várias formas por diferentes pessoas. Dessa maneira, a “leitura” de um mapa não é simplesmente uma questão de reconstituir mentalmente a realidade geográfica subjacente mas de ser capaz de construir uma significação que possa ser relacionada aos interesses e preocupações particulares de um usuário [CG97].

Segundo [BF97] deveria existir idealmente uma correspondência entre dados no banco de dados geográfico e o mapa produzido a partir destes dados. Esta correspondência não é facilmente obtida por dois motivos principais: em primeiro lugar porque a geração de um mapa envolve perda de informação devido às diversas simplificações citadas na Seção 4.1, principalmente as ocorridas por intermédio da generalização cartográfica; e em segundo lugar porque os mapas geralmente são fornecidos sem informação de como foram gerados a partir dos dados e porque o seu processo de geração não pode ser descrito totalmente de maneira formal, o que torna a interpretação uma tarefa difícil mas muito importante [BF97].

Nesta última década, algumas pesquisas têm sido desenvolvidas com enfoque na construção da significação necessária ao processo de interpretação através do uso da semiologia

(ou semiótica) [BF97]. Semiologia ou semiótica é a ciência geral dos signos que estuda todos os fenômenos culturais como se fossem sistemas de signos, ou seja, sistemas de significação [Fer86].

Um signo é definido de forma simplista como “*alguma coisa que representa algo para alguém*” e surge como resultado do relacionamento ternário entre um objeto, um representante e um interpretante. O objeto é a parte do signo que se deseja referenciar, o representante é a entidade efetivamente utilizada na tentativa de comunicação e o interpretante é o processo de significação criado na mente pensante, que ao se deparar com o representante produz uma idéia que remete ao objeto [San95, Pie90].

Para Pierce [Pie90] os signos podem ser divididos em símbolos<sup>1</sup>, índices e ícones segundo uma ordem crescente de especificidade. Nos símbolos, a associação entre o objeto e seu representante é estabelecida socialmente ou de forma imposta. Nos índices o representante se associa ao objeto por uma relação natural de pressuposição ou dependência. Finalmente, no caso do ícone, a representação se dá de forma direta, por semelhanças das características entre o objeto e seu representante.

Além desta classificação inicial, há uma subclassificação para os ícones em: imagem, diagrama e metáfora, aqui em ordem decrescente de especificidade. As imagens reproduzem as características do objeto. Os diagramas, por sua vez, têm sua representatividade baseada em relação a partes do objeto referenciado. Por último, na metáfora a representação se dá através de um paralelismo com alguma outra coisa. A Figura 4.4 utiliza a classificação de [Pie90] e exemplifica alguns signos que podem ser encontrados usualmente em mapas, segundo [Oli88, Ago75, Mar91] ou em cartas náuticas de acordo com [DHN95].

A utilização de signos mais específicos, em geral, facilita o entendimento dos usuários mas leva normalmente à necessidade de grupos mais restritos de usuários. Há um claro compromisso entre a especificidade dos signos e a amplitude de comunicação. Na Cartografia convencional um mapa é utilizado para comunicar um conhecimento específico e portanto destina-se a um grupo exclusivo, o que possibilita o uso de signos com maior grau de especificidade tais como os ícones. Em SIG, por sua vez, dadas as suas características de possibilitar múltiplas utilizações, é difícil utilizar signos mais específicos, sendo freqüente o uso de símbolos. Apesar disso, é importante que se busquem alternativas balanceadas no uso destes signos para ao mesmo tempo aumentar a compreensão dos usuários sem perda da generalidade de uso.

A interpretação de mapas envolve a transformação das relações perceptíveis entre os símbolos utilizados para representação dos objetos geográficos em informação espacial sobre os mesmos. Assim, uma consideração interessante é possibilitar a criação de uma ordem de precedência de classes de aspectos (para facilitar a decisão de quais aspectos

---

<sup>1</sup>Cabe ressaltar que a noção de símbolo de acordo com a classificação de Pierce é menos genérica que a utilizada na Subseção 4.1.3.

SIGNO	ICONE	IMAGEM		PINUS ITALICO	NORMALMENTE NAO ENCONTRADO	
		DIAGRAMA		PINUS		BOIA PILAR
		METAFORA		LAGO INTERMITENTE		RIO TEMPORARIO
	INDICE		ANCORADOURO		FUNDEADOURO	
	SIMBOLO		GRAFITA		EDIFICAÇÃO	
			MAPAS	CARTAS NAUTICAS		

Figura 4.4: Ilustração da taxionomia de Pierce [Pie90] utilizando “signos” encontrados em mapas e em cartas náuticas.

podem ser relaxados para uma dada apresentação) e a criação de uma hierarquia de precisão dentro de cada classe (para explicitar ao usuário a real intenção do cartógrafo em relação àquele aspecto).

As Figuras 4.5 e 4.6, extraídas de [BF97], exemplificam uma possível ordem de precedência entre diferentes classes de aspectos e uma possível hierarquia de precisão dentro do aspecto “orientação”. Note-se que variações na ordem de precedência podem mudar totalmente uma apresentação e, por conseguinte, a interpretação dela decorrente.

A má interpretação de um mapa, segundo [BF97], se dá, na maioria das vezes, quando um aspecto apresentado é entendido pelo usuário de maneira mais ou menos rigorosa que a intenção do cartógrafo. No exemplo da Figura 4.7, uma *superinterpretação* do aspecto “orientação” ocorreria se os símbolos utilizados para a apresentação das estações do metrô fossem interpretados como tendo a orientação precisa das reais estações e esta orientação fosse, na verdade, feita de acordo com uma aproximação em 8 setores conforme mostra a Figura 4.6. Assim, a *super ou subinterpretação* ocorre quando se escolhe um nível de precisão superior ou inferior, respectivamente, ao nível correto de correspondência dentro da hierarquia de aspecto.

A Figura 4.7 estaria associada, então, às Figuras 4.5 e 4.6 para proporcionar ao

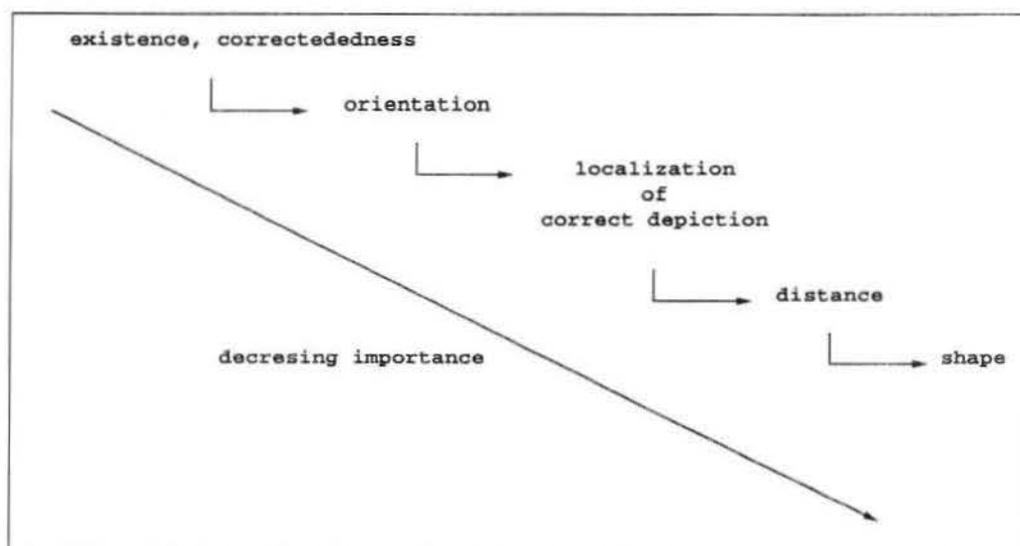


Figura 4.5: Exemplo de uma possível ordem de precedência entre diferentes classes de aspectos por Barkowsky & Freska [BF97].

usuário um modelo do processo de interpretação de um mapa.

### 4.3 Resumo

Este capítulo tratou de dois aspectos fundamentais no julgamento de qualidade de dados, efetuado pelo usuário de SIG: a visualização dos resultados de consultas, que se dá normalmente através da apresentação de um mapa na tela, a partir da representação dos dados armazenados em um banco de dados geográfico; e o processo de interpretação deste resultado realizado pelo usuário, que depende de aspectos do projeto de interfaces para aqueles sistemas.

A apresentação é tão importante quanto a especificação e controle da qualidade dos dados durante a coleta e conversão discutida no Capítulo 3. É nesta fase que o usuário tem a oportunidade de visualizar os dados e de ter conhecimento de possíveis erros que afetam sua qualidade.

Os aspectos da Cartografia abordados (escalas, generalização cartográfica, projeções e símbolos) continuam presentes na cartografia usando SIG. A apresentação de dados geográficos deve levar em consideração todas estas distorções, para tentar reduzir o erro na produção do mapa que será mostrado ao usuário em resposta à sua consulta. É importante realçar que as regras de uma boa produção cartográfica variam de acordo com os objetivos da aplicação.

O uso de interfaces gráficas não garante a clareza ou qualidade dos resultados, pois

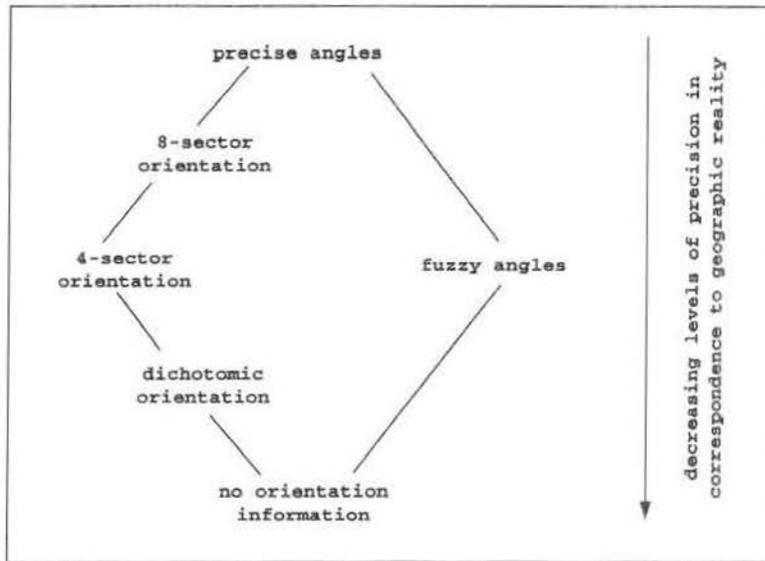


Figura 4.6: Possíveis hierarquias de precisão em relação ao aspecto “orientação” por Barkowsky & Freska [BF97].

os usuários “enxergam” o resultado de uma consulta não apenas vendo o que é mostrado explicitamente, mas também em função de alguma notação gráfica implícita, transmitida, por exemplo, por noções de adjacência, cor e outras afetadas por particularidades do perfil cognitivo do usuário. Este fato, alerta para a necessidade de se entender como acontece o processo de interpretação das informações indiretas existentes em um mapa. A Semiótica, surge, neste contexto, como uma ferramenta para auxiliar a compreensão deste processo. Além disso, estas notações gráficas implícitas são derivadas da experiência do usuário no domínio da aplicação e seu conhecimento pessoal da região geográfica estudada, ou seja, a familiaridade do usuário com o problema tem influência na escolha da estratégia de apresentação.

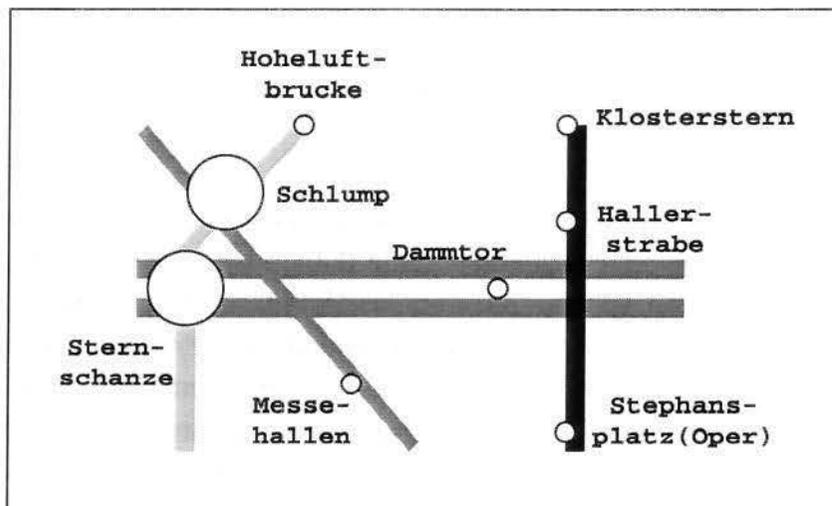


Figura 4.7: Fragmento de um mapa do metrô de Hamburgo [BF97].

# Capítulo 5

## Aspectos de implementação

O Capítulo 3 ressaltou a importância de metadados como forma de auxiliar o usuário de SIG a tomar conhecimento de algumas informações sobre a qualidade dos dados utilizados. Este capítulo apresenta um protótipo que foi desenvolvido para auxiliar os usuários a definir e avaliar a qualidade dos mapas produzidos com o auxílio de SIG. Esta qualidade é decorrente de três tipos de fatores: dados de entrada (fonte e conversão), processo de geração do mapa e apresentação do mapa. O protótipo considera a qualidade associada a um mapa como calculada a partir de vários indicadores. Os resultados das avaliações destes indicadores são armazenados em metadados. O protótipo foi desenvolvido como extensão a um sistema criado em uma dissertação de mestrado no IC-Unicamp. O sistema, denominado WOODSS (*WO*rkf*l*Ow-based *s*patial *D*ecision *S*upport *S*ystem) [Sef98], é um sistema espacial de apoio ao processo decisório acoplado ao SIG IDRISI [TIP97]. Os metadados gerenciados pelo protótipo correspondem aos vários tipos de indicadores de qualidade descritos nos Capítulos 3 e 4, cujos valores são fornecidos pela equipe que produz e converte os dados, pelos usuários e pelo SIG.

O capítulo está organizado como se segue. A Seção 5.1 discute aspectos da avaliação de qualidade utilizando-se múltiplos critérios. A Seção 5.2 realiza uma análise comparativa dos diversos padrões e propostas de uso de metadados que dizem respeito à qualidade. A Seção 5.3 fornece os conceitos básicos de Sistemas de Apoio ao Processo Decisório e apresenta o WOODSS e o IDRISI. A Seção 5.4 estende o WOODSS, a partir dos padrões da Seção 5.2 para incorporar metadados de qualidade. A Seção 5.5 traz um exemplo de uma sessão no protótipo e por último a Seção 5.6 resume o capítulo.

### 5.1 Avaliação de qualidade sob múltiplos critérios

Esta dissertação aborda nos capítulos precedentes alguns indicadores necessários para caracterizar a qualidade de dados geográficos. Entretanto, além da dificuldade de se

estabelecer o conjunto mínimo de indicadores para estimar a qualidade, há outra questão importante a analisar: como realizar uma avaliação global de qualidade ao se utilizarem múltiplos critérios.

Diversos autores reconhecem que os dados são melhor caracterizados via múltiplos atributos ou dimensões. [BT99], por exemplo, trata da avaliação da qualidade de pacotes de softwares comerciais usando a metodologia de Avaliação sob Múltiplos Critérios (*Multi Criteria Evaluation - MCE*).

A metodologia de Avaliação sob Múltiplos Critérios adaptada à avaliação de “softwares de prateleira” utiliza o modelo de processo proposto pelas normas ISO 9126 e IEEE 1061 [BT99]. De acordo com estas normas, a qualidade de software é definida como um conjunto de atributos a avaliar que são organizados em uma árvore e aos quais estão associados pesos. A qualidade global é obtida pela soma ponderada das medidas. As Figuras 5.1 e 5.2 a seguir representam, respectivamente, o modelo hierárquico de qualidade definido pelas normas ISO 9126 e a adaptação de [BT99].

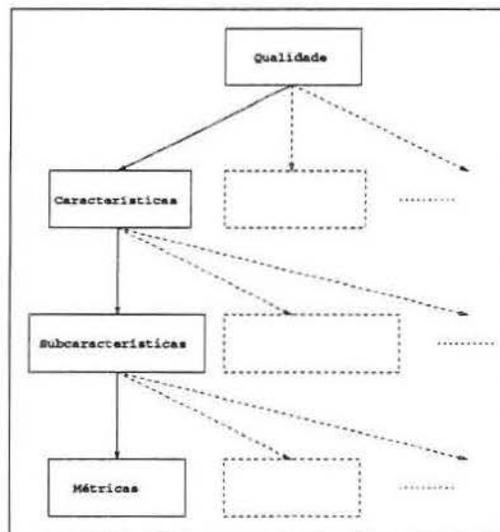


Figura 5.1: O modelo de qualidade de software de acordo com a norma ISO 9126 [BT99, Jun99].

De acordo com estas figuras, a qualidade pode ser analisada a partir de várias características (fatores). Cada característica (fator), por sua vez é composta de uma ou mais subcaracterísticas (critérios) que são analisadas por uma ou mais métricas. [BT99] discute qual a melhor forma de se agregar as informações em cada nível da hierarquia de qualidade, alertando para o fato de existirem diferentes tipos de medidas (booleanas, escalas, numéricas etc.) e portanto para a necessidade destas medidas serem normalizadas. Conclui que o processo de avaliação é dependente do uso pretendido e portanto sugere a definição de modelos de qualidade para diferentes grupos de usuários.

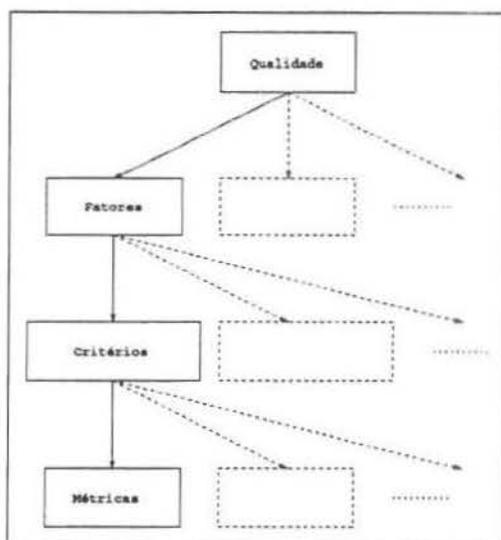


Figura 5.2: O modelo de qualidade de software de acordo com Blin & Tsoukiàs [BT99].

Os princípios básicos da metodologia de MCE aplicada à avaliação de qualidade compreendem estabelecer: os fatores componentes da qualidade; os critérios componentes de cada um destes fatores; as métricas que serão utilizadas para medir cada critério; os métodos de normalização a usar para permitir a agregação de diferentes tipos de métricas, critérios ou fatores; os pesos a serem atribuídos em cada nível (métricas, critérios e fatores); a forma de agregação a usar em cada nível e por fim a realização da avaliação. Esta metodologia pode ser adaptada aos objetivos da dissertação e usada em qualquer nível do julgamento de qualidade, desde a avaliação global da qualidade de um dado de entrada, usado em uma sessão de trabalho para construir um mapa no SIG, até a avaliação de qualidade de um modelo, auxiliando o usuário na decisão de qual modelo melhor se ajusta ao seu problema dentre vários recuperados de um banco de modelos.

## 5.2 Uso de metadados

A base do desenvolvimento da ferramenta é considerar os múltiplos critérios para cálculo da qualidade armazenados em metadados.

Esta seção apresenta, de forma comparativa, alguns dos parâmetros/indicadores de qualidade presentes em padrões de metadados ou propostos por diversas organizações (Tabela 5.1) visando identificar os metadados que devem ser incorporados ao WOODSS para permitir ao usuário avaliar a qualidade do trabalho desenvolvido no IDRISI. Também são apresentadas as análises de diversos autores sobre o assunto (Tabela 5.2) e por último o que existe implementado no IDRISI (Tabela 5.3) em sua Versão 2.0.

Indicadores/ Padrões ou Organizações	FGDC	SDTS	ICA	CNIG	CEN/ TC 287	ISO/ TC 211
Precisão posicional	X	X	X	X	X	X
Precisão atributo	X	X	X	X	X	X
Precisão semântica	-	-	X	-	-	-
Consistência lógica	X	X	X	X	X	X
Compleitude	X	X	X	X	X	X
Linhagem	X	X	X	X	-	-
Informação temporal	-	-	X	X	X	X
Resolução	-	-	-	-	-	X
Cobertura de nuvens	X	-	-	-	-	-
Fonte	-	-	-	-	X	X
Uso	-	-	-	-	X	X
Meta qualidade	-	-	-	-	X	-
Homogeneidade	-	-	-	-	X	-
Testes e conformidade	-	-	-	-	-	X

Tabela 5.1: Indicadores de qualidade presentes em padrões de metadados ou propostos por organizações.

A Tabela 5.1 foi criada a partir da documentação disponibilizada pelas organizações. Vale ressaltar que, nesta tabela:

a) Os indicadores “Meta qualidade” e “Homogeneidade” descrevem, respectivamente, a confiança no indicador de qualidade (ou seja, a qualidade do indicador de qualidade) e quão bem as informações de qualidade são aplicáveis ao conjunto de dados;

b) Dependendo da organização, o indicador informação temporal tem nomes diferentes tais como data, tempo ou precisão temporal; e

c) As siglas utilizadas têm o significado a seguir:

FGDC - *Federal Geographic Data Committee* [FGD98a] e SDTS - *Spatial Data Transfer Standard* [USG98] - descritos no Capítulo 3;

ICA - *International Cartographic Association* - E. U. A - expressa a opinião da organização sobre quais os metadados de qualidade necessários a dados geográficos [ICA99];

CNIG - *Conseil National de l'Information Géographique* - França - expressa a opinião da organização sobre os tipos de erros que afetam a qualidade de dados geográficos [CNI99]. É o padrão seguido, por exemplo, por Faiz & Boursier [FB94] (vide Tabela 5.2);

CEN/TC287 - *European Committee for Standardization, Technical Committee on Geographic Information* - expressa o resultado dos estudos do comitê sobre os princípios gerais para a descrição da qualidade das informações geográficas. Estes princípios foram transformados, em 1998, em padrões provisórios (*European Prestandards - ENVs*) a serem adotados pelos países participantes do grupo [CEN99]; e

ISO/TC211 - *International Organization for Standardization, Technical Committee on*

*Geographic Information and Geomatics* - expressa o resultado dos estudos do Grupo de Trabalho 3 (*Geospatial Data Administration*) em seu projeto 13 que trata da definição do esquema para qualidade aplicada a dados geográficos [ISO99].

A Tabela 5.2 apresenta a visão de alguns autores sobre: os diferentes indicadores de erros presentes em dados geográficos que devem ser medidos e incorporados aos dados como metadados [FB94]; as características que afetam a qualidade dos dados geográficos [Aro95] (vide Capítulo 2); e os elementos de qualidade que devem ser armazenados sob a forma de metadados, para permitir aos usuários decidirem se um conjunto de dados se ajusta ou não ao uso pretendido [Aal99].

Indicadores/Autores	[FB94]	[Aro95]	[Aal99]
Precisão posicional	X	X	X
Precisão atributo	X	X	X
Precisão semântica	-	-	X
Consistência lógica	X	X	X
Compleitude	X	X	X
Linhagem	X	X	X
Precisão temporal, data ou atualidade	X	X	X
Resolução	-	X	-
Fonte	-	-	X
Uso	-	X	X
Homogeneidade	-	-	X

Tabela 5.2: Indicadores de qualidade propostos por alguns autores.

O IDRISI também utiliza metadados para informar sobre a qualidade dos dados geográficos que manipula. Para isso, o software possui associado a cada plano de informação um arquivo de documentação. Estes arquivos de documentação possuem, dentre outras informações, algumas que dizem respeito à qualidade dos dados e que são listadas na Tabela 5.3.

Indicadores/Formato dos dados	Imagem matricial	Arquivo vetorial
Precisão posicional	X	X
Precisão atributo	X	X
Consistência lógica	X	X
Compleitude	X	X
Linhagem	X	X
Resolução	X	-

Tabela 5.3: Indicadores de qualidade presentes no IDRISI [Eas97].

A análise destas três tabelas nos permite identificar que tipo de informações são normalmente utilizadas sob a forma de metadados para avaliar qualidade de dados geo-

gráficos. Entretanto, permanecem as considerações identificadas na Seção 3.5: há uma grande dificuldade de uso destas estruturas pelos usuários de SIG e estes metadados representam a visão somente dos produtores de dados que nem sempre é a visão dos usuários.

Além disso, é importante perceber que ainda há uma grande diferença entre as informações que são propostas e as que realmente existem nos metadados. Existem menos informações sobre as fontes de dados e sobre o processo de conversão do que o que seria desejável, praticamente não existem informações sobre o uso dos dados, a informação temporal é muitas vezes omitida e não há indicadores para a qualidade da apresentação.

Esta dissertação pretende reduzir estas deficiências incorporando funcionalidades ao WOODSS e usando-o como uma ferramenta de manipulação dos metadados de qualidade. Os tipos de metadados propostos na Seção 5.4 refletem estas inovações. Estes tipos consideram o estudo de [WSG96] sobre qualidade de dados do ponto de vista dos usuários. Consideram igualmente o uso dos dados, permitindo ao usuário incluir informações sobre o *workflow* (qualidade do processo) e sobre a apresentação dos resultados, incorporando algumas informações importantes discutidas no Capítulo 4.

### 5.3 Sistemas Espaciais de Apoio à Decisão e WOODSS

Um Sistema de Apoio ao Processo Decisório (Decision Support System - DSS) é um software projetado para aumentar a efetividade dos tomadores de decisões, fornecendo mecanismos que facilitam a interação com modelos de análise e dados [Sef98]. A decisão é tomada em um processo iterativo, por refinamentos sucessivos. O processo de decisão em tal contexto consiste em gerar e avaliar soluções alternativas, alcançadas através do estudo dos compromissos entre objetivos em conflito, e em identificar características indesejáveis nas soluções.

Um Sistema Espacial de Apoio ao Processo Decisório (*Spatial Decision Support System - SDSS*), assim como um DSS, tem por objetivo oferecer ao usuário um ambiente flexível para a tomada de decisões, só que em um contexto onde a dimensão espacial é fundamental para a análise das decisões. Frequentemente, este processo consiste em criar uma série de mapas com o uso de SIG, onde cada mapa corresponde a uma alternativa de solução de um problema, e a seguir escolher o mapa que fornece a solução mais adequada.

Considere-se por exemplo que o problema de decisão de encontrar uma área adequada para a instalação de uma fábrica no estado de São Paulo deve satisfazer aos três critérios a seguir: o local deve apresentar uma declividade máxima específica; estar suficientemente afastado de reservas ambientais e não ser uma área de florestas. Podem ser encontradas

várias soluções que atendam de forma satisfatória à combinação destes requisitos. Entretanto, caberá ao tomador de decisão, usando um SIG e criando mapas alternativos, definir qual solução será melhor, de acordo com a maior ou menor importância do critério analisado.

O WOODSS (*WorkfLOw-based spatial Decision Support System*) é uma ferramenta desenvolvida no IC-Unicamp [Sef98], usada em conjunto com o SIG IDRISI [Eas97], para apoio ao processo decisório no contexto espacial. O objetivo é documentar as interações do usuário com o SIG durante o processo de geração de cada mapa. Esta interação é armazenada sob a forma de *workflows científicos*.

Um *workflow* pode ser definido como uma seqüência de passos necessários para atingir um determinado objetivo. Cada passo deste processo é chamado atividade ou tarefa, e pode ser executado por um ou mais agentes. Um agente ou executor é uma pessoa ou componente de software capaz de executar uma ou mais destas tarefas [Bar96].

Um *workflow científico* é um *workflow* adaptado aos procedimentos e necessidades de aplicações científicas, por exemplo em geoprocessamento [MVW96, WVM98]. As peculiaridades de *workflows científicos* e seu uso em situações deste tipo são definidas em [WVM96] e escapam do contexto desta dissertação.

A arquitetura interna do WOODSS é apresentada pela Figura 5.3 a seguir, obtida de [Sef98]. O protótipo implementado é baseado em modificações ao WOODSS. O Banco de Dados é formado por dados das aplicações do usuário, manipulados em um SIG, e por *workflows científicos* que usam estes dados. Estes *workflows* descrevem a parte do processo decisório relativa à confecção dos mapas no SIG. O SGBD representa, na realidade, dois sistemas diferentes, um gerenciando o banco de dados da aplicação e o outro o Banco de *Workflows* gerado pelo WOODSS.

Nesta figura podem ser vistos cinco módulos principais. O módulo Monitor monitora as atividades executadas pelo usuário no SIG e traduz esta informação para o formato usado por outro módulo, o Gestor de *Workflows*, que é responsável por criar e gerenciar *workflows* fazendo a ligação entre os módulos de Consulta e de Atualizações e o SGBD. O módulo de Atualizações permite ao usuário atualizar os *workflows* e o módulo de Consultas permite navegar no Banco de *Workflows*. Finalmente, o módulo Interface realiza a comunicação do usuário com os módulos internos do WOODSS.

O WOODSS funciona acoplado ao software IDRISI. Este software é um SIG desenvolvido na Universidade de Clark e considerado o mais disseminado dentre os SIG para microcomputadores baseados em estrutura matricial. O IDRISI é formado por um programa principal que atua como interface para uma coleção de mais de cento e cinquenta módulos. Estes módulos provêm facilidades para a entrada, apresentação e análise de dados geográficos.

No IDRISI, os dados geográficos são descritos sob a forma de planos de informação

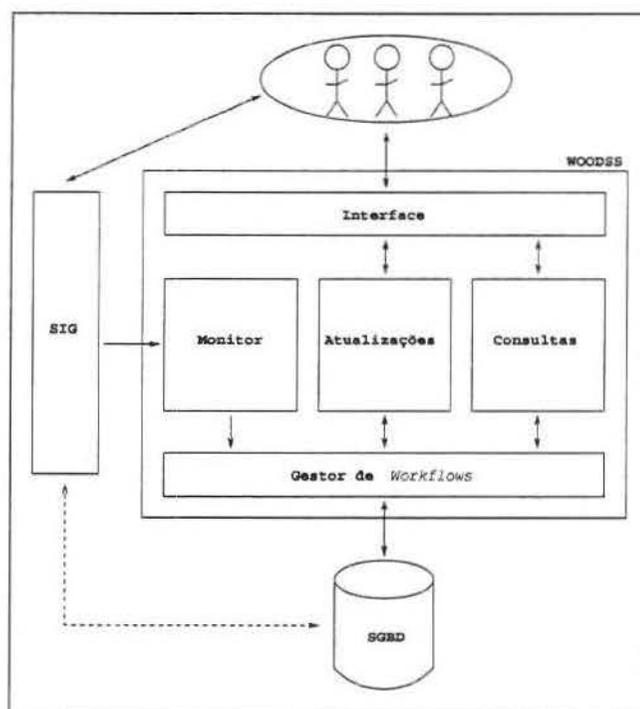


Figura 5.3: Arquitetura do WOODSS, de acordo com Seffino [Sef98].

(*map layers*). Toda a análise é realizada sobre estas estruturas e o sistema suporta a sobreposição de até 16 delas, que podem ser armazenadas sob a forma de um único arquivo, representando toda a composição. O IDRISI permite planos de informação quer no formato de imagem matricial, quer como arquivo vetorial. Embora existam algumas facilidades para a análise vetorial e para conversões entre tipos, as funções de análise são principalmente orientadas ao formato matricial [Eas97].

As principais operações normalmente realizadas sobre os dados geográficos são identificadas por [CCH<sup>+</sup>96] como pertencendo a quatro categorias de operações primitivas: reclassificação de categorias, sobreposição de mapas, cálculo de vizinhança e medidas de distância e conectividade.

As operações de reclassificação reorganizam os dados de um mapa de forma a criar um novo mapa, onde cada faixa de valores (classe) corresponde a um dado: por exemplo, em um mapa de vegetação, classes podem ser vegetação rasteira, arbustiva, savana etc. As operações de sobreposição envolvem vários mapas e criam novas regiões que resultam da combinação, para cada posição geográfica, dos valores dos mapas sobrepostos naquela posição. As operações de vizinhança geram mapas que indicam zonas de influência de algum tipo de fenômeno. Finalmente, as operações de distância e conectividade geram mapas que em geral apresentam traçado de rotas. A visualização das diferentes regiões em um mapa é promovida através da variação dos atributos visuais (cor, textura e outros

- vide Capítulo 4).

A maioria destas operações primitivas é tratada no IDRISI pela utilização de módulos isolados ou em conjunto. As operações de reclassificação, por exemplo, correspondem ao uso do módulo *Reclass*; operações de sobreposição podem ser efetuadas com a utilização do módulo *Overlay*; o módulo *Distance* permite a realização de operações de distância, quando utilizado sozinho, e de operações de vizinhança, quando utilizado em conjunto com o módulo *Reclass*. Estas noções serão importantes quando da apresentação do problema exemplo utilizado para demonstrar uma sessão completa de utilização do WOODSS na Seção 5.5.

## 5.4 Introduzindo qualidade no WOODSS

### 5.4.1 Arquitetura do protótipo

O WOODSS armazena cada *workflow* sob a forma de três relações (Atividades, Arquivos e Dependências) e um arquivo registro de metadados (vide Figura 5.4). Cada *workflow* recebe um nome definido pelo sistema que é igualmente atribuído às três relações e ao registro de metadados que descreve o *workflow*. O conjunto Metadados, Atividades, Arquivos e Dependências de um *workflow* é acessado pelo nome deste *workflow*.

A primeira tupla da relação Atividades, por exemplo, mostra que o *workflow* tem uma atividade correspondente ao acionamento do módulo *Surface* do IDRISI, que o usuário denominou “*Slope Calculation*”, com entrada “arquivo 1#” e saída “arquivo 2#” (respectivamente, arquivos “*relief*” e “*slopes*” da relação Arquivos). A relação Dependências indica quais atividades são conectadas entre si: sua primeira tupla, por exemplo, mostra que a “atividade 1” e a “atividade 2” possuem uma dependência de dados (*Type=D*) referente ao “arquivo 2” (*Name=2* - arquivo *slopes*).

Observando-se a figura, pode-se perceber o quanto é limitada a estrutura dos Registros de Metadados associados aos *workflows*. O protótipo é baseado em modificações nesta parte do WOODSS. Trata-se de um módulo, acoplado ao WOODSS que possibilita o armazenamento de informação sobre qualidade sob a forma de metadados, permitindo ao usuário avaliar a qualidade da informação geográfica segundo três fatores: a qualidade dos dados de entrada, a qualidade do processo e a qualidade da apresentação do resultado (mapa gerado).

A Figura 5.5 apresenta a extensão ao Modelo Entidade Relacionamento do WOODSS após a incorporação de metadados de qualidade e a Figura 5.6 detalha os atributos destes metadados.

A Figura 5.7 mostra a arquitetura do WOODSS modificada para incluir o novo módulo - Qualidade - responsável pela interação com os demais módulos do WOODSS para

**Aw19980501235253**

Id	Name	Origin	Module	Input	Output	Others
1	Slope Calculation	I	surface	1#	2#	1@d@1.0@
2	UserDefined Classification	I	reclass	2#	3#	i@2@1 @0@3@0@3@999@-9999@
3	UserDefined Classification	I	reclass	4#	5#	i@2@0@0@2@1@2@2@0@3@99@-9999@
4	Distance Calculation	I	distance	5#	6#	NULL
5	UserDefined Classification	I	reclass	6#	7#	i@2@0@0@250@1@250@99999@-9999@
6	UserDefined Classification	I	reclass	4#	8#	i@2@0@0@9@1@9@11@0@11@99@-9999@
7	Image Multiplication	I	overlay	3#7#	9#	3@
8	Image Multiplication	I	overlay	9#8#	10#	3@
9	Level Curves Calculation	M	NULL	11#	1#	NULL

Relação Atividades

**Xw19980501235253**

Id	Name	Extension	Title
1	relief	img	Sao Paulo State - DEM
2	slopes	img	Sao Paulo State - Slopes
3	slopeok	img	Sao Paulo State - Slopes > 3 degrees
4	landuse	img	Sao Paulo State - Land Use map
5	reserves	img	Sao Paulo State - Reserves
6	disres	img	Sao Paulo State - Distances from a reserve
7	bufres	img	Sao Paulo State - 250 m. buffer area around reserves
8	forests	img	NULL
9	tmp	img	NULL
10	areasok	img	Sao Paulo State - Appropriate areas for installing a factory
11	level	vec	Sao Paulo State - Level Curves

Relação Arquivos

**Dw19980501235253**

Id1	Id2	Type	Name
1	2	D	2
3	4	D	5
4	5	D	6
2	7	D	3
5	7	D	7
7	8	D	9
6	8	D	8
9	1	D	1

Relação Dependências

**Mw19980501235253 - Bloco de Notas**

Arquivo Editar Pesquisar Ajuda

-----  
 Meta-data file generated by WOODSS  
 -----

**#Author:**  
 Laura Andrea Seffino

**#Objectives:**  
 1. Find appropriate areas for installing a Factory  
 in Sao Paulo State

**#Area:**  
 Sao Paulo State - Brazil

**#Other:**  
 |

Registro de Metadados

Figura 5.4: As relações Atividades, Arquivos e Dependências e o Registro de Metadados do WOODSS.

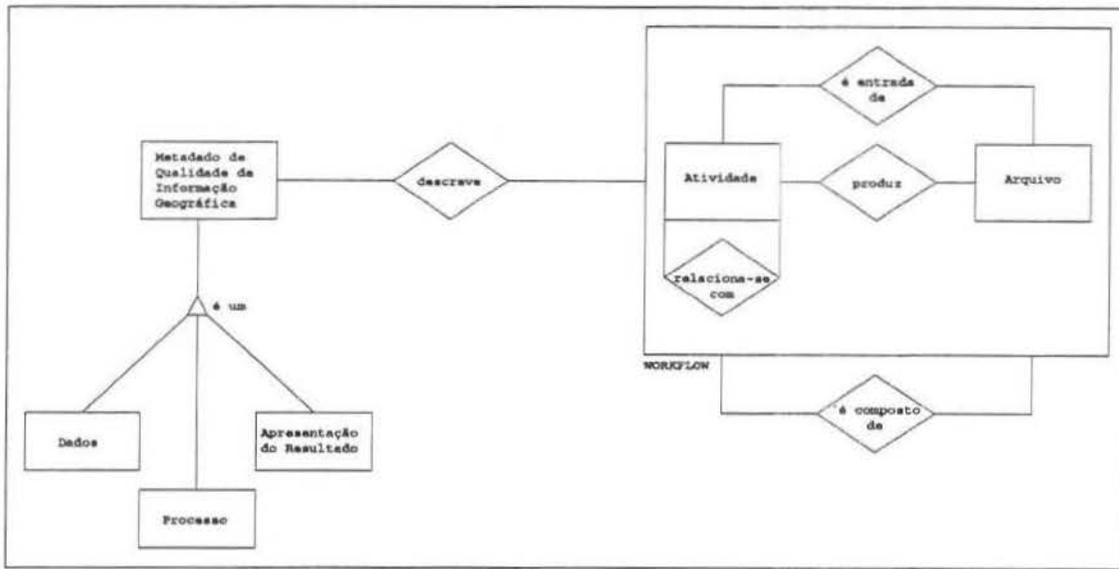


Figura 5.5: Extensão ao Modelo Entidade-Relacionamento do WOODSS.

gerenciamento da informação de qualidade. As seguintes alterações foram realizadas no WOODSS:

- O módulo Consultas passou a permitir consultas aos metadados de qualidade associados aos *workflows*; e
- O módulo Atualizações passou a permitir a atualização dos metadados de qualidade associados aos *workflows*. As modificações são restritas aos dados de qualidade fornecidos pelo usuário (atributos de origem u nas Tabelas 5.4 e 5.5).

As alterações no WOODSS foram implementadas em JAVA<sup>TM</sup>. A interação de JAVA<sup>TM</sup> com o Visual FoxPro<sup>TM</sup> 5.0 é realizada a partir de chamadas a comandos SQL [EN89] dentro do código JAVA<sup>TM</sup>.

Os metadados de qualidade dos dados de entrada descrevem as diversas informações afetas à qualidade dos arquivos de entrada do *workflow*.

Os metadados de qualidade do processo se referem ao processo de construção do mapa correspondente ao *workflow*. Detalhes sobre este tipo de metadado de qualidade escapam ao escopo da dissertação, pois é preciso considerar itens como o modelo utilizado para gerar o mapa, as funções do SIG escolhidas para esta implementação e até mesmo como as operações são implementadas no SIG. A hipótese feita neste estudo é que a qualidade de um processo pode ser avaliada de uma forma inicial pelo usuário ao visualizar o *workflow* correspondente, sendo apenas indicados alguns atributos padrão (área geográfica, objetivos, data) e poucos atributos de qualidade (autor, relevância e adequabilidade).

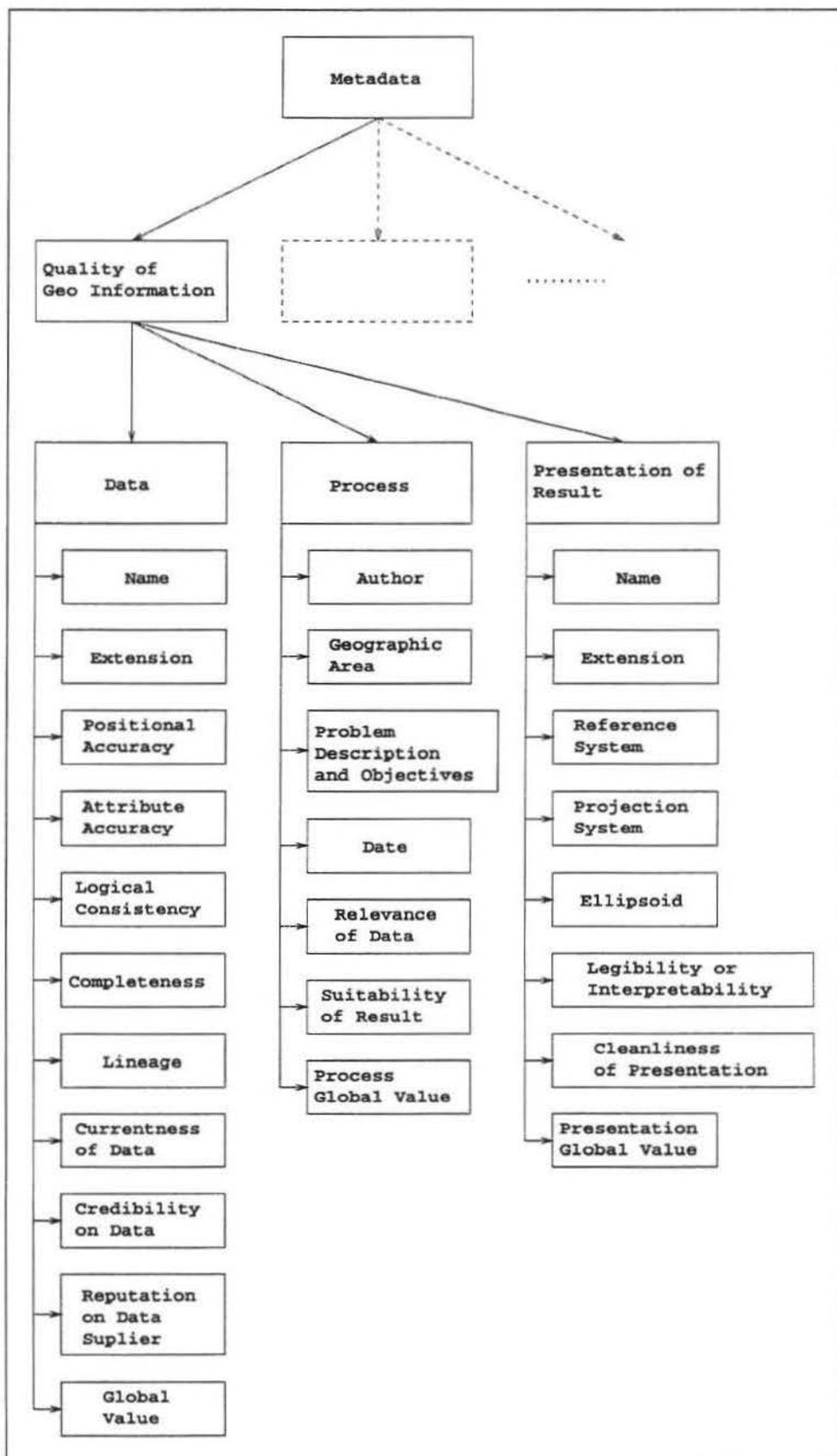


Figura 5.6: Tipos de metadados de qualidade propostos e seus atributos.

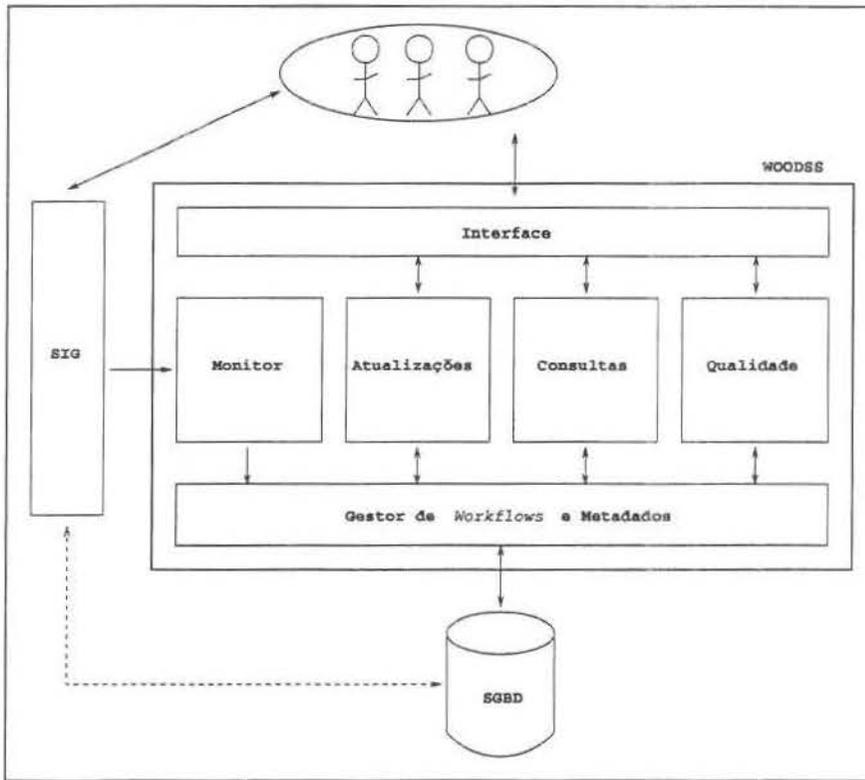


Figura 5.7: Arquitetura modificada do WOODSS.

Finalmente, a qualidade da saída considera alguns dos fatores de apresentação levantados no Capítulo 4 e se refere ao mapa final produzido pelo *workflow*. A apresentação deste mapa é aquela gerada pelo IDRISI.

#### 5.4.2 Esquema do banco de dados

O banco de dados foi modificado para substituir o arquivo Registro de Metadados por três relações, correspondendo às três entidades da Figura 5.5 (dados de entrada, processo e apresentação do resultado) com os atributos da Figura 5.6.

As relações implementadas foram: Dados (que engloba atributos de qualidade dos arquivos de entrada do *workflow*), Processo (que engloba além dos atributos anteriormente cobertos pelo registro de metadados, atributos de qualidade do processo) e Apresentação do Resultado (que engloba os aspectos de qualidade da apresentação do resultado).

As Tabelas 5.4 e 5.5 apresentam os atributos de cada relação, indicando ainda uma classificação de cada atributo conforme seu tipo - se resultante de cálculo (c), decorrente de avaliação (a) ou não (na) - e conforme sua origem - se obtido a partir dos sistemas envolvidos (s) ou solicitada aos usuários (u).

Por exemplo, o atributo “*positional accuracy*” da tabela Dados de Entrada (Vide Tabela 5.4) é classificado como “tipo=na” e “origem=s”, indicando que não é decorrente de avaliação e é fornecido pelo sistema. Já o atributo “*reputation on data supplier*” da mesma tabela é classificado como “a, u” denotando que é avaliado pelo usuário. Finalmente os atributos “*global value*” de cada tabela são calculados pelo sistema, a partir de pesos fornecidos pelo usuário associados a cada fator de qualidade. A Seção 5.5 mostra como o usuário pode realizar o cálculo destes atributos.

O protótipo implementado segue a metodologia MCE para avaliar a qualidade dos dados de entrada, do processo e da saída mas somente para alguns critérios. A árvore parcial que define a qualidade da informação geográfica para os dados de entrada é a mostrada na Figura 5.8. Nesta árvore a qualidade das informações geográficas pode ser avaliada de acordo com três fatores, dentre os quais o fator dados de entrada. O fator dados de entrada pode ser avaliado sob vários critérios dentre os quais a sua precisão posicional (neste caso dividida em dois subcritérios - precisão absoluta e relativa) e cada um deles pode ser medido por pelo menos duas métricas, média em conjunto com desvio padrão ou RMSE. Na atual implementação, somente alguns critérios são avaliados e não existem tipos diferentes de métricas, portanto não é utilizado qualquer fator de normalização.

A idéia básica do cálculo dos valores globais de qualidade é utilizar a soma ponderada de fatores, considerando múltiplos critérios. O usuário considera os metadados de qualidade associados a cada tipo de informação (arquivo de entrada, processo ou apresentação do resultado) e efetua o cálculo dos fatores globais.

Os pesos são atribuídos pelo usuário por avaliação direta como pode ser visto nas Figuras 5.11, 5.13 e 5.14 e a soma ponderada é utilizada para agregar as diversas medidas. Ao final, o usuário tem, além das informações constantes dos metadados, valores associados aos dados de entrada, ao processo e à apresentação do resultado (respectivamente *global value*, *process global value* e *presentation global value* nas Tabelas 5.4 e 5.5).

A soma ponderada é realizada de acordo com a fórmula a seguir, onde,  $x_i \in [0, 10]$  é o valor atribuído ao fator, critério ou métrica e  $w_i \in [0, 1]$  o peso correspondente.

$$WS = \sum_{i=1}^n x_i w_i$$

Nesta implementação,  $x_i$  é o valor dado pelo usuário aos atributos credibilidade nos dados, reputação do fornecedor de dados, relevância dos dados, adequabilidade do resultado, legibilidade ou interpretabilidade e clareza da apresentação. Os valores destes atributos serão atribuídos de acordo com a avaliação que o usuário faça, dentro de uma escala cujos valores principais são: *poor*, *fair*, *good* e *excellent*. Este modelo segue o proposto por [BT99]. Com relação aos pesos, cabe ressaltar que quando se usa este tipo de fórmula para agregação a soma dos valores dos pesos atribuídos aos fatores deve ser igual

a 1 [Eas97]. Nesta primeira implementação os valores de pesos possíveis são: 0,  $\frac{1}{4}$ ,  $\frac{1}{2}$ ,  $\frac{3}{4}$  e 1.

## 5.5 Exemplo de uso

Considere-se o seguinte problema de decisão extraído de [Sef98] e que consiste em encontrar uma área adequada para a instalação de uma fábrica no estado de São Paulo. A área escolhida deve satisfazer os três critérios seguintes:

- (a) Apresentar uma declividade menor que 3 graus;
- (b) Estar fora de uma área de 250 metros ao redor de reservas ambientais; e
- (c) Não ser atualmente uma área de florestas.

A resolução deste problema no IDRISI pode se realizar segundo quatro passos: um para cada critério mencionado, e um último passo combinando os resultados anteriores. Para isso considera-se um conjunto inicial de arquivos, “*relief*” e “*landuse*” que representam, respectivamente, o relevo e o uso da terra da região escolhida

(a) Para o cálculo da declividade é preciso desenvolver um mapa de declividades (em graus) da área geográfica a analisar e efetuar uma reclassificação para criar uma imagem booleana que represente áreas de declividades menores e maiores que 3 graus. Para a criação do mapa de declividades é preciso usar no IDRISI o módulo *Surface* - opção *Slopes* - e para a reclassificação o módulo *Reclass*;

(b) A área de 250 metros ao redor das zonas de reservas é calculada em três passos: (1) reclassificação de um mapa para criar uma imagem booleana correspondente às áreas de reservas; (2) cálculo da distância de cada ponto da área em estudo à reserva mais próxima; e (3) reclassificação do mapa de distâncias em distâncias maiores e menores que 250 metros. Nos passos (1) e (3) é usado o módulo *Reclass* e no passo (2) o módulo *Distance* do IDRISI;

(c) As zonas que não correspondem a florestas são obtidas a partir de outra operação de reclassificação (módulo *Reclass*); e

(d) As três imagens geradas nos passos anteriores são a seguir combinadas em uma operação de sobreposição (módulo *Overlay* do IDRISI).

Suponha que um usuário ao terminar a produção deste mapa deseja inserir as informações de qualidade associadas. Inicialmente, solicita ao WOODSS a visualização do processo de geração do mapa, que lhe é mostrado como o *workflow* da Figura 5.9. Esta figura corresponde às relações da Figura 5.4.

Neste *workflow*, atividades são representadas por caixas - por exemplo, a caixa *Reclass* indica uma operação de reclassificação. Arquivos são arcos entre caixas. Os arquivos de entrada são *relief* e *landuse* e a saída é um mapa de nome *areasok*.

A seguir, o usuário acessa o Menu “*Metadata*” (vide Figura 5.9) para ativar o módulo

Qualidade que lhe permite manipular os metadados de entrada, processo ou saída. Inicialmente o sistema mostra os dados que são fornecidos pelo próprio IDRISI e portanto não podem ser modificados; a seguir, o usuário pode informar metadados do tipo u (vide Tabelas 5.4 e 5.5). O usuário pode decidir que tipo de metadados deseja verificar escolhendo os itens do Menu “*Metadata*”.

Há um par de telas de qualidade para cada arquivo de entrada: dados fixos (fornecidos pelo IDRISI, “origem=s”) e complementares (fornecidos pelo usuário, “origem=u”). A Figura 5.10 mostra uma tela do primeiro tipo, relativa aos dados não modificáveis do arquivo de entrada *relief*, onde se vê, por exemplo, que este arquivo foi inserido no banco de dados geográfico em 14/04/95. Se o usuário clica o botão “*Comp. Data*”, surge a tela de metadados adicionais (Figura 5.11). Esta figura mostra, por exemplo, que o usuário indicou que a credibilidade dos dados é boa, e que este fator deve ter um peso de  $\frac{1}{4}$  no cálculo do fator global de qualidade do arquivo. Se o usuário desejar retornar à janela principal do WOODSS ou à caixa de diálogo anterior basta acionar o botão “*Cancel*” da janela. As informações sobre outros arquivos de entrada (apenas *landuse*, neste caso) podem ser visualizadas e inseridas acionando-se botão “*Next*”.

De forma semelhante, as Figuras 5.12, 5.13 e 5.14 mostram cópias de tela que permitem a visualização e inserção de informações da qualidade do processo e da qualidade da apresentação.

Quando o usuário decide inserir as informações de qualidade associadas, aperta o botão “*Insert*” em cada uma das janelas de inserção. Neste instante os valores globais são calculados utilizando a soma ponderada e de acordo com os valores e pesos atribuídos pelo usuário e as relações correspondentes são atualizadas.

## 5.6 *Resumo*

Este capítulo apresentou um protótipo que permite a um usuário do IDRISI avaliar a qualidade de um mapa gerado. Este protótipo modifica o WOODSS, um sistema desenvolvido no IC - Unicamp para documentar a interação de um usuário com SIG sob a forma de *workflows científicos*.

As modificações consistem em adicionar informações sobre qualidade sob a forma de metadados - para entrada, processo e visualização da saída - e calcular um valor de qualidade usando múltiplos critérios. Estes metadados foram definidos a partir da análise de padrões de metadados de qualidade existentes. O gerenciamento destes metadados é realizado por um novo módulo - Qualidade - incorporado ao WOODSS.

Metadado	Atributo	Descrição	Tipo	Origem
de Dados	id (NUMERIC)	Um identificador do <i>workflow</i> gerado pelo WOODSS e usado como chave estrangeira para as relações metadados.	na	s
	ida (INTEGER)	Um identificador do arquivo de entrada gerado pelo WOODSS.	na	s
	name (CHARACTER)	Nome do arquivo de entrada armazenado no IDRISI.	na	s
	extension (CHARACTER)	Extensão que define o tipo de arquivo no IDRISI (.img para matricial e .vec para vetorial).	na	s
	positional accuracy (NUMERIC)	Uma medida de precisão de posição do arquivo (imagem matricial ou arquivo vetorial). Normalmente apresenta o resultado do cálculo do RMSE. O valor 9999 representa "desconhecido".	na	s
	attribute accuracy (NUMERIC)	Uma medida de precisão dos atributos do arquivo. Para atributos qualitativos é informado como erro proporcional e para quantitativos como RMSE. O valor 9999 representa "desconhecido".	na	s
	logical consistency (CHARACTER)	Uma descrição da consistência lógica do arquivo.	na	s
	completeness (CHARACTER)	Informação sobre o quão completo o arquivo é em relação àquele assunto.	na	s
	lineage (CHARACTER)	Descrição da história de como os arquivos foram registrados/derivados.	na	s
	currentness of data (DATE)	Data em que os arquivos foram inseridos no banco de dados geográfico.	na	s
	credibility of data (NUMERIC)	Avaliação da credibilidade que os dados possuem junto ao usuário.	a	u
	reputation on data supplier (NUMERIC)	Avaliação da reputação que o fornecedor dos dados tem junto ao usuário.	a	u
	global value (NUMERIC)	Valor calculado e atribuído ao arquivo de entrada em função da avaliação e pesos selecionados pelo usuário para os atributos credibilidade nos dados e reputação do fornecedor de dados.	c	s/u
do Processo	id (NUMERIC)	Um identificador do <i>workflow</i> gerado pelo WOODSS e usado como chave estrangeira para as relações metadados.	na	s
	author (CHARACTER)	Nome do usuário que foi o autor do <i>workflow</i> e fez o julgamento dos atributos de qualidade.	na	u
	geographic area (CHARACTER)	Nome da área geográfica sobre a qual foi realizado o estudo do problema (município, estado, país, região etc.).	na	u
	problem description and objectives (CHARACTER)	Descrição sucinta do problema sendo estudado e dos objetivos que se deseja alcançar.	na	u
	date (DATE)	A data da criação do <i>workflow</i> .	na	s
	relevance of data (NUMERIC)	Avaliação da importância dos dados para o sucesso ou insucesso do <i>workflow</i> .	a	u
	suitability of result (NUMERIC)	Avaliação do nível de sucesso ou insucesso do <i>workflow</i> .	a	u
	process global value (NUMERIC)	Valor calculado e atribuído ao <i>workflow</i> em função da avaliação e pesos selecionados pelo usuário para os atributos relevância dos dados e adequabilidade do resultado.	c	s/u

Tabela 5.4: Descrição, por tipo, dos atributos dos metadados propostos.

Metadado	Atributo	Descrição	Tipo	Origem
da Apresentação do Resultado	id (NUMERIC)	Um identificador do <i>workflow</i> gerado pelo WOODSS e usado como chave estrangeira para as relações metadados.	na	s
	name (CHARACTER)	Nome do arquivo de saída que corresponde à apresentação do resultado do <i>workflow</i> .	na	s
	extension (CHARACTER)	Extensão que define o tipo de arquivo no IDRISI (.img para matricial e .vec para vectorial).	na	s
	reference system (CHARACTER)	Sistema de referência utilizado.	na	s
	projection system (CHARACTER)	Indicação do tipo de projeção cartográfica utilizado na apresentação.	na	s
	ellipsoid (CHARACTER)	Nome do elipsóide utilizado como referência.	na	s
	legibility or interpretability (NUMERIC)	Avaliação da facilidade de entendimento do resultado com relação aos aspectos gráficos.	a	u
	cleanliness of presentation (NUMERIC)	Uma avaliação de quão clara e limpa é a apresentação do resultado de acordo com a avaliação do usuário.	a	u
presentation global value (NUMERIC)	Valor calculado e atribuído ao arquivo de saída em função da avaliação e pesos selecionados pelo usuário para os atributos legibilidade ou interpretabilidade e clareza da apresentação.	c	s/u	

Tabela 5.5: Continuação da descrição, por tipo, dos atributos dos metadados propostos.

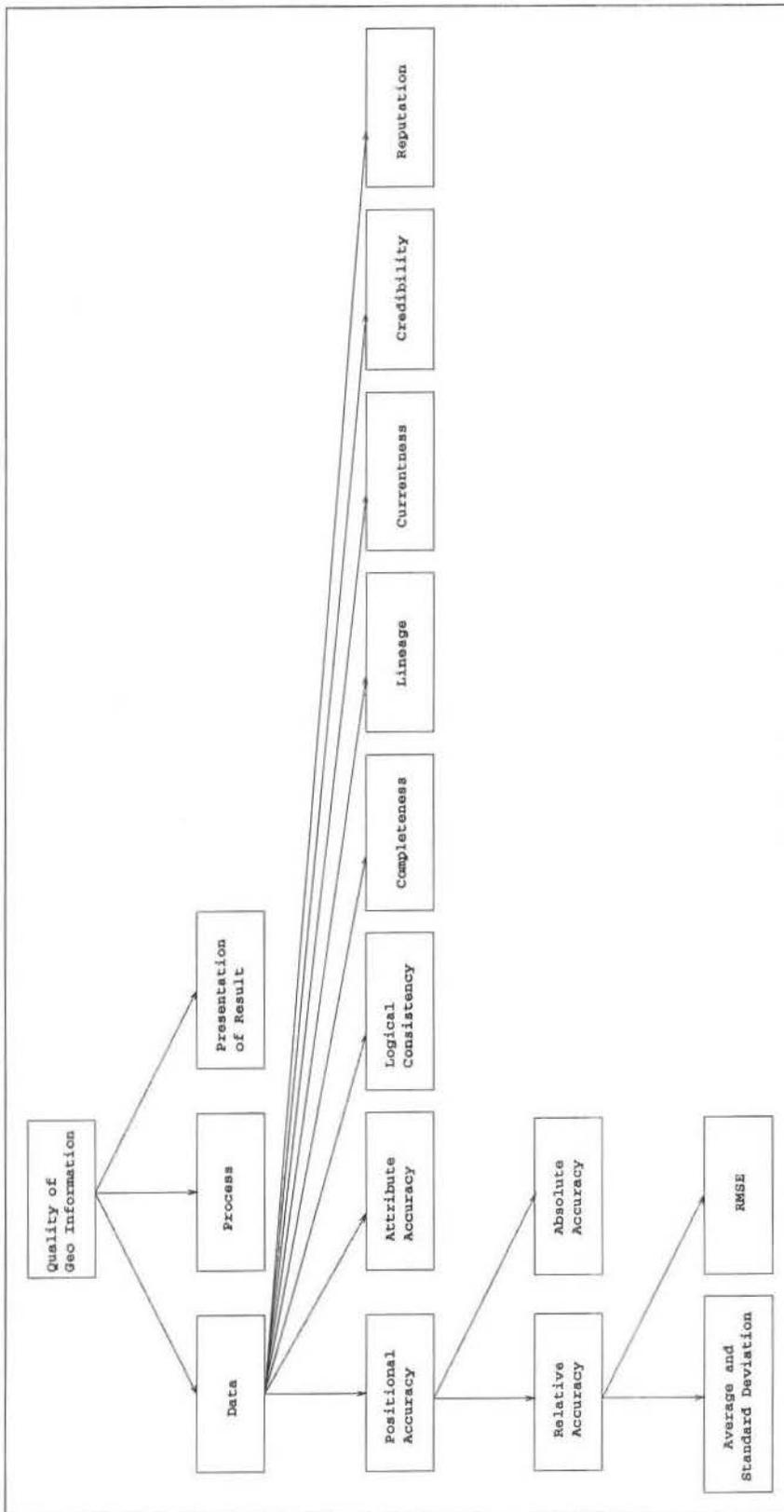


Figura 5.8: Uma proposta de modelo para qualidade dos dados de entrada usados em uma aplicação geográfica

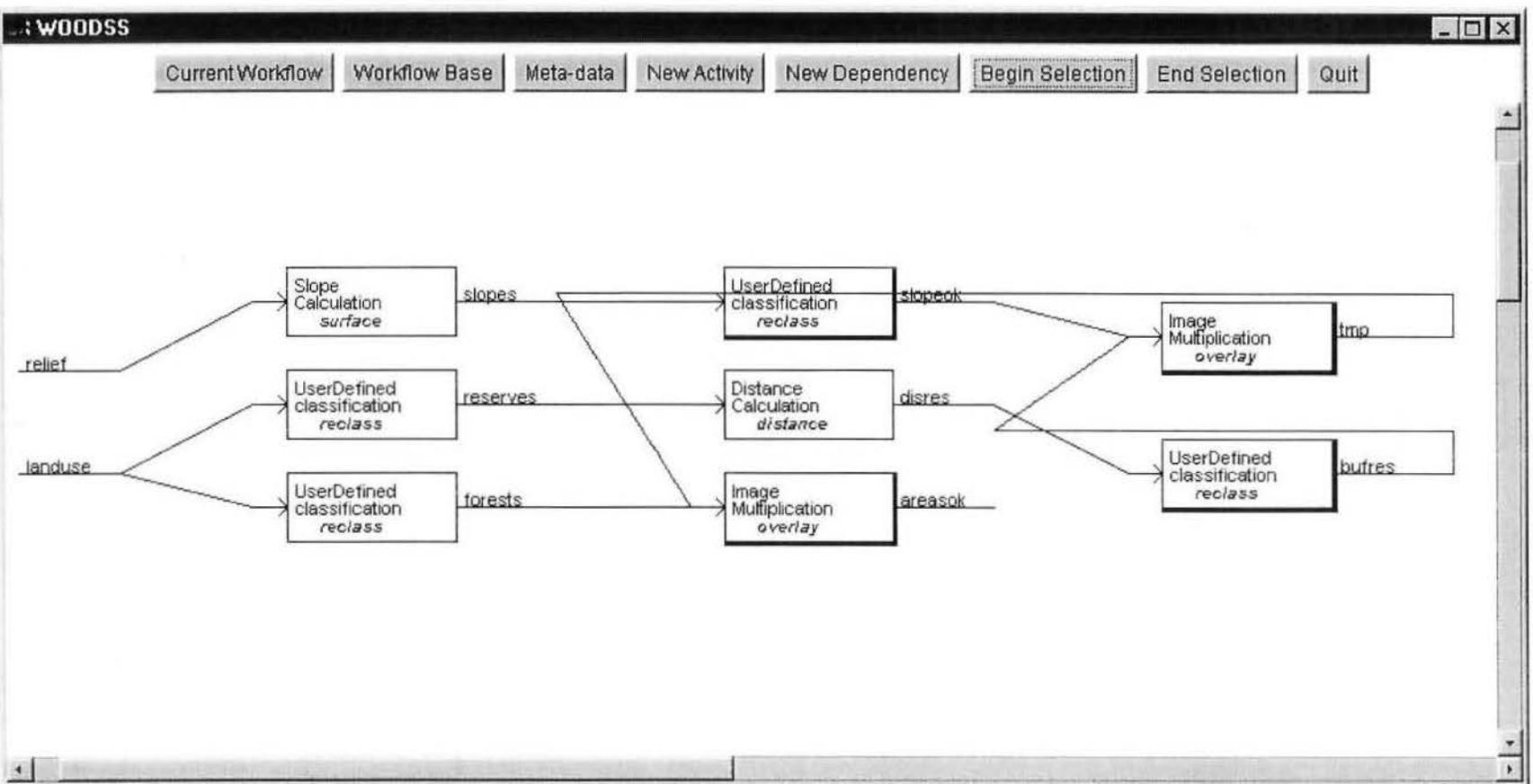


Figura 5.9: Visualização do processo de geração de um mapa (*workflow*) no WOODSS [Set98].

Quality Information Metadata - Data Entry

Name	relief
Extension	img
Positional accuracy	unknown
Attribute Accuracy	unknown
Logical Consistency	unknown
Completeness	unknown
Lineage	unknown
Currentness of Data	14/04/95

Next  
Cancel  
Comp. Data

Figura 5.10: Cópia da tela de visualização de metadados de dados de entrada.

**Quality Complementary Data - Data Entry**

**Credibility of Data**

*Evaluation*      *Weight*

Excellent  
Good  
Fair  
Poor

0     1/4     1/2     3/4     1

**Reputation on Data Supplier**

*Evaluation*      *Weight*

Excellent  
Good  
Fair  
Poor

0     1/4     1/2     3/4     1

Next  
Cancel  
Insert

Figura 5.11: Cópia da tela de inserção de metadados adicionais aos dados de entrada.

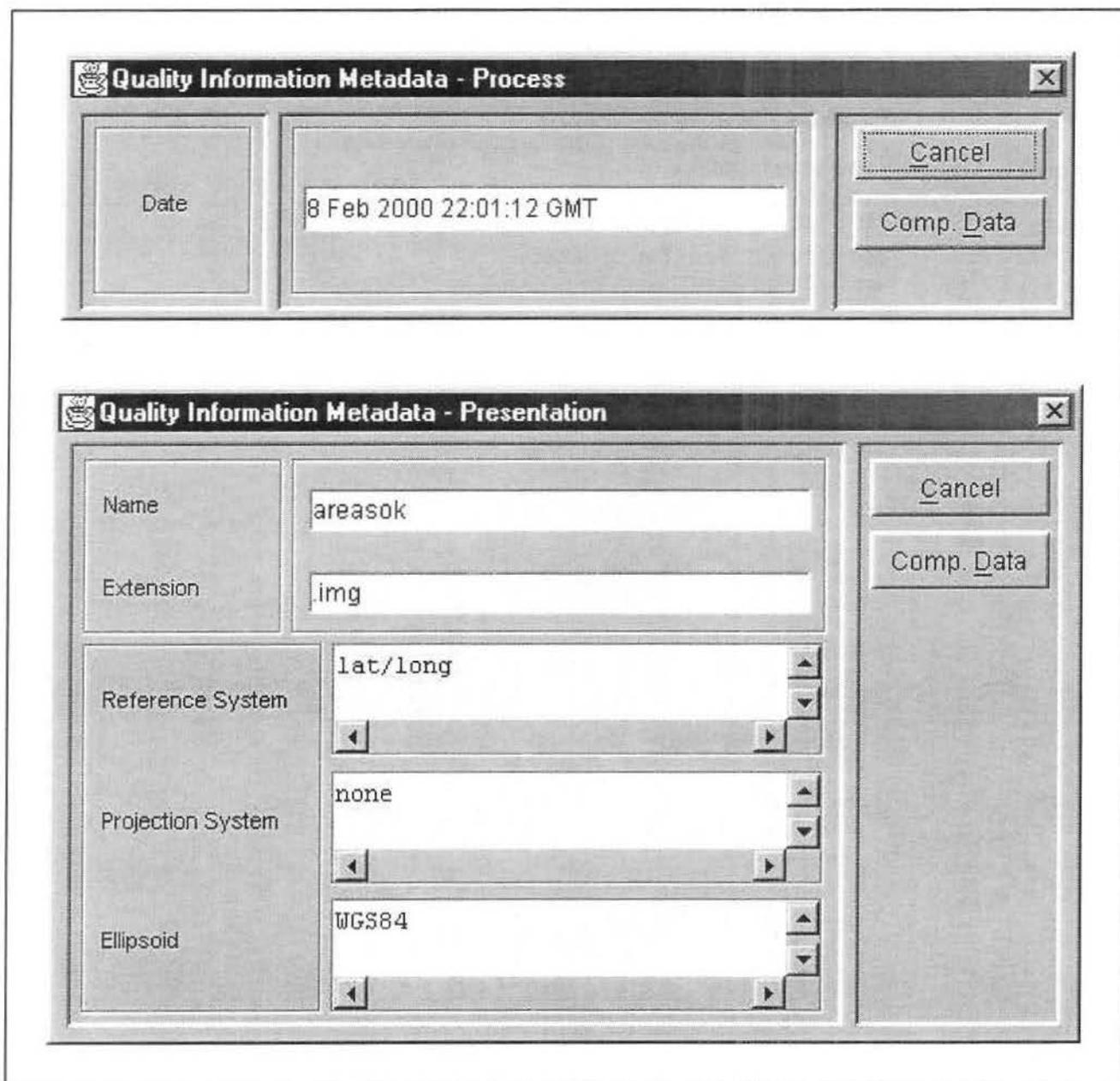


Figura 5.12: Cópia das telas de visualização de metadados do processo e da apresentação.

**Quality Complementary Data - Process** [X]

Author	Alexandre Carvalho de Alencar
Geographic Area	Sao Paulo - State
Problem Description and Objectives	Find an appropriate area to install a factory in Sao Paulo state

**Relevance of Data**

Evaluation	Weight
Excellent	<input type="radio"/> 0 <input type="radio"/> 1/4 <input checked="" type="radio"/> 1/2 <input type="radio"/> 3/4 <input type="radio"/> 1
Good	
Fair	
Poor	

**Suitability of Result**

Evaluation	Weight
Excellent	<input type="radio"/> 0 <input type="radio"/> 1/4 <input checked="" type="radio"/> 1/2 <input type="radio"/> 3/4 <input type="radio"/> 1
Good	
Fair	
Poor	

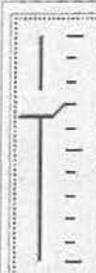
Cancel

Insert

Figura 5.13: Cópia da tela de inserção de metadados adicionais do processo.

**Quality Complementary Data - Presentation** [X]

Legibility or Interpretability

Evaluation	Weight
	<input type="radio"/> 0 <input type="radio"/> 1/4 <input type="radio"/> 1/2 <input checked="" type="radio"/> 3/4 <input type="radio"/> 1

Cleanliness of Presentation

Evaluation	Weight
	<input type="radio"/> 0 <input checked="" type="radio"/> 1/4 <input type="radio"/> 1/2 <input type="radio"/> 3/4 <input type="radio"/> 1

Cancel

Insert

Figura 5.14: Cópia da tela de inserção de metadados adicionais da apresentação.

# Capítulo 6

## Conclusões e extensões

### 6.1 Contribuições

Esta dissertação discutiu diferentes aspectos do tema “qualidade” no contexto de informação geográfica. O uso dos dados é um aspecto fundamental citado desde a definição de qualidade como “*fitness for use*” até as discussões de como pode variar o entendimento do resultado de uma aplicação geográfica sob a forma de mapas.

Alguns aspectos importantes de Cartografia e conceitos básicos de SIG foram apresentado e conceitos de “qualidade” em áreas afins foram discutidos. Seguiu-se uma análise de aspectos relevantes a serem considerados durante as diversas fases da manipulação de dados geográficos, desde sua captura, passando por seu pré-processamento até a apresentação dos dados como resultado de uma aplicação geográfica. Em cada uma destas fases foram identificados aspectos que de alguma forma precisam ser considerados para melhorar a “qualidade” da informação geográfica como um todo. Identificados estes aspectos e algumas maneiras de avaliá-los, discutiu-se a melhor forma de tornar os resultados destas avaliações acessíveis aos usuários. A solução adotada foi incorporar metadados a um sistema espacial de apoio ao processo decisório baseado em *workflows*. Este sistema passou a ser uma ferramenta de apoio para documentação de metadados de “qualidade” e para avaliação de alguns aspectos de “qualidade” considerados importantes pelos usuários.

Esta dissertação considerou como grupo de trabalho inicial pesquisadores da área agro-ambiental, uma vez que teve como ponto de partida para a ferramenta de apoio um sistema desenvolvido para ser usado neste contexto [Sef98] e que por sua vez usou uma metodologia de trabalho específica desta área [Pir97]. Embora inicialmente voltada para a pesquisa agro-ambiental, as idéias discutidas na dissertação são suficientemente gerais para serem aplicadas em outras áreas de aplicações geográficas.

As principais contribuições desta dissertação foram:

1. levantamento dos diversos níveis em que a qualidade deve ser considerada no con-

texto de dados e aplicações geográficas;

2. definição de um conjunto básico de critérios de avaliação desta qualidade;
3. proposta de como alguns destes critérios podem ser obedecidos e como esta informação pode ser explicitada ao usuário; e
4. implementação de uma ferramenta que permite aplicar estes conceitos, validando o estudo realizado.

## 6.2 Extensões

Dentre as possíveis extensões pode-se identificar as que se seguem:

- inclusão de outros tipos de metadados de qualidade após validação efetiva por usuários da área inicial e uso destes metadados como parâmetros de acesso para recuperar “*Workflows*”;
- abordagem mais completa dos problemas relacionados ao processamento dos dados, incluindo o estudo da adequação de modelos matemáticos;
- avaliação de necessidade de alteração dos atributos constantes de cada tipo de relação de metadados, também após a validação inicial;
- testes de avaliação da eficiência da fórmula de agregação utilizada para composição dos valores globais e qualidade global; e
- extensão da especificação do protótipo.

A incorporação de metadados de qualidade ao WOODSS o torna uma ferramenta importante para documentação e avaliação da qualidade, mas ainda não é suficiente para atender a outras necessidades do próprio sistema.

Na verdade, o ideal seria que o sistema tivesse em sua estrutura (Vide Figura 5.3) um Banco de Metadados, além do Banco de *Workflows* já existente, suportando não só metadados de qualidade mas metadados de uma forma geral. A utilização de um Banco de Metadados auxiliaria, inclusive, a recuperação e análise de interações já realizadas armazenadas no Banco de *Workflows*.

Considerações sobre as classes de metadados de uso geral e suas estruturas necessitariam de uma análise mais aprofundada (semelhante à realizada por esta dissertação para qualidade) e portanto não fazem parte do escopo deste trabalho. Há diversas pesquisas sendo desenvolvidas em relação à criação de bancos de metadados para aplicações geográficas. Uma análise de diversos padrões de metadados existentes, seguida de adequação

a uma finalidade específica pode ser vista, por exemplo, no trabalho de Fagundes [Fag99]. Neste trabalho é criado um Banco de Metadados para mapeamento da biodiversidade no Estado de São Paulo. A análise desenvolvida por esta dissertação, entretanto, restringiu-se ao estudo de um conjunto mínimo de metadados necessários a auxiliar o usuário na avaliação da qualidade dos dados.

Este conjunto inicial mínimo obviamente pode não atender completamente às necessidades de um grupo de usuários. Estudos posteriores podem indicar a necessidade de aumentar este conjunto inicial de metadados.

Um aspecto não abordado nesta dissertação foi o da modelagem matemática associada à produção de mapas em SIG. A qualidade de uma simulação está fortemente associada à qualidade do modelo usado. Para cálculo de erosão, por exemplo, existem inúmeros modelos matemáticos para análise. Isto corresponderia, no WOODSS, a associar uma métrica de qualidade distinta a cada *workflow*. Esta métrica indicaria, de alguma forma, a confiança do usuário na escolha do modelo específico para a simulação em questão. Considerações sobre este tipo de qualidade não foram abordadas pela dissertação, pois são do domínio de cada grupo de pesquisa. Usou-se no entanto uma aproximação que considerou que o especialista que usa o SIG pode avaliar todo o processo.

Os problemas relacionados ao processamento dos dados incluem portanto o estudo de modelos matemáticos mais apropriados para uma tarefa específica, uma vez que a qualidade de uma simulação está bastante ligada à qualidade do modelo associado e até mesmo de aspectos do SIG.

A fórmula utilizada para agregação dos valores resultantes da avaliação feita pelos usuários e de sua atribuição de pesos em valores globais deve ser testada para analisar sua eficiência em termos de diferenciar dados que possuam qualidade distinta conhecida. Esta análise pode levar à adoção de outras fórmulas para agregação dos resultados.

O protótipo pode ser melhor especificado incorporando, dentre outros, aspectos de controle de acesso de usuários para permitir diferentes níveis de permissões de acordo com uma hierarquia de usuários. Assim, alterações nos metadados de qualidade só seriam permitidas aos autores dos *workflows*, por exemplo. O sistema de consulta pode ser mais flexível no sentido de permitir consulta aos *workflows* por intermédio dos diversos atributos das relações de metadados, ao contrário da implementação atual que só permite o acesso aos metadados a partir dos *workflows* armazenados, mas não a consulta inversa.

Uma vez armazenados os metadados, o usuário poderia avaliar também o modelo como um todo em função da qualidade. Neste contexto, a qualidade poderia ser calculada em função de todos os arquivos de entrada, da qualidade do processo (*workflow*) e a qualidade da apresentação do resultado. Estes valores poderiam ser agregados em um único valor que poderia servir de base para o julgamento da adequação de um modelo dentre vários analisados considerando sua qualidade.

Finalmente, o sistema poderia ter sido aplicado a casos reais, visando inclusive verificar sua adequação em apontar problemas de qualidade.

# Bibliografia

- [Aal96] H. J. G. L. Aalders. Quality Metrics for GIS. In M. J. Kraak and M. Molenaar, editors, *Advances in GIS Research II*, pages 5B1–5B10. 7th International Symposium on Spatial Data Handling, Delft University, august 1996.
- [Aal99] H. G. L. Aalders. The Registration of Quality in a GIS. *International Symposium on Spatial Data Quality*, july, 18th-20th 1999. Department of Land Surveying and Geo-Informatics.
- [Ago75] A. Agostini. *Disegno Topografico*. Editore Ulrico Hoepli Milano, 1975.
- [Aro95] S. Aronoff. *Geographic Information Systems: A Management Perspective*. WDL Publications, 1995.
- [Bak65] M. P. R. Bakker. *Cartografia - Noções Básicas*. Diretoria de Hidrografia e Navegação - Marinha do Brasil, 1965.
- [Bak75] M. P. R. Bakker. *A Projeção de Mercator*. Diretoria de Hidrografia e Navegação - Marinha do Brasil, 1975.
- [Bar96] P. Barthelmeß. Sistemas de Workflow: Análise da Área e Proposta de Modelo. Master's thesis, Universidade Estadual de Campinas - DCC, 1996.
- [BBRS97] R. H. Bordeau, C. J. Burley, G. S. Roseberry, and K. C. Sullivan. Institutions, Languages, Content Standards, Records Sintaxes and Protocols: Fitting It All Together. *Second IEEE Metadata Conference*, september 1997.
- [Ber92] T. Bernhardsen. *Geographic Information Systems*. Viak IT and Norwegian Mapping Authority, 1992.
- [BF97] T. Barkowsky and C. Freska. Cognitive Requirements on Making and Interpreting Maps. In S. C. Hirtle and A. U. Frank, editors, *Spatial Information Theory: A Theoretical Basis for GIS - International Conference COSIT'97*, pages 347–361. Springer-Verlag, october, 15–18 1997. Lecture Notes in Computer Science.

- [BM98] P. A. Burrough and R. A. McDonnell. *Principles of Geographical Information Systems - Spatial Information Systems and Geostatistics*. Oxford University Press, 1998.
- [BT99] M. Blin and A. Tsoukiàs. Evaluation of COTS using Multicriteria Methodology. In *Proceedings of 6th European Conference on Software Quality*, pages 429–438, april, 12–16 1999.
- [CCH<sup>+</sup>96] G. Camara, M. A. Casanova, A. S. Hemerly, G. C. Magalhães, and C. B. Medeiros. *Anatomia de Sistemas de Informação Geográfica*. Instituto de Computação - IC - UNICAMP, 1996.
- [CEN99] CEN. The Geographic Information European Prestandards and CEN Reports. <http://forum.afnor.fr/afnor/WORK/AFNOR/GPN2/Z13C/PUBLIC- /WEB/ENGLISH% - /pren.html>, novembro 1999.
- [CG97] H. Couclelis and J. Gottsegen. What Maps Mean to People: Denotation, Connotation, and Geographic Visualization in Land-Use Debates. In S. C. Hirtle and A. U. Frank, editors, *Spatial Information Theory: A Theoretical Basis for GIS - International Conference COSIT'97*, pages 151–162. Springer-Verlag, october, 15–18 1997. Lecture Notes in Computer Science.
- [Chr84] N. R. Chrisman. The Role of Quality Information in the Long Term Functioning of a GIS. *Cartographica*, 21(2):79–87, 1984.
- [Chr91] N. R. Chrisman. The Error Component in Spatial Data. In M. F. Goodchild, D. J. Maguire, and D. W. Rhind, editors, *Geographical Information Systems - Principles and Applications - Vol. I*, chapter 12, pages 165–174. Longman Scientific & Technical, 1991.
- [Cif95] R. R. Ciferri. Um Benchmark Voltado à Análise de Desempenho de SIGs. Master's thesis, Universidade Estadual de Campinas - DCC, junho 1995.
- [CNE99] CNES. Spot. [http://www.spot4.cnes.fr/spot4\\_gb/index.htm](http://www.spot4.cnes.fr/spot4_gb/index.htm), maio 1999.
- [CNI99] CNIG. Conseil National de l'Information Géographique. <http://www.cnig.fr>, novembro 1999.
- [Cro79] P. B. Crosby. *Quality is Free - The Art of Making Quality Certain*. McGraw-Hill Company, 1979.

- [Cur97] L. M. Del Val Cura. Tratamento de Versões em Bancos de Dados para Sistemas Informações Geográficas. Master's thesis, Universidade Estadual de Campinas - IC, março 1997.
- [Dem86] W. E. Deming. *Out of Crisis*. Center for Advanced Engineering Study - MIT - Cambridge, MA, 1986.
- [DHN95] DHN. *Carta 12000 (INT 1) Símbolos, Abreviaturas e Termos, usados nas Cartas Náuticas Brasileiras*. Diretoria de Hidrografia e Navegação - Marinha do Brasil, 1995.
- [DS91] F. W. Davis and D. S. Simonett. GIS and Remote Sensing. In M. F. Goodchild, D. J. Maguire, and D. W. Rhind, editors, *Geographical Information Systems - Principles and Applications - Vol. I*, chapter 14, pages 191–213. Longman Scientific & Technical, 1991.
- [Eas97] J. R. Eastman. *IDRISI for Windows - Tutorial Exercises, Version 2.0*. Clark Labs for Cartographic Technology and Geographic Analysis - Clark University, 950 Main St., Worcester, MA, USA, june 1997.
- [Ege95] M. J. Egenhofer. Naive Geography. In A. U. Frank and W. Kuhn, editors, *Spatial Information Theory: A Theoretical Basis for GIS - International Conference COSIT'95*, pages 01–15. Springer-Verlag, september, 21–23 1995. Lecture Notes in Computer Science.
- [EN89] R. Elmasri and S. B. Navathe. *Fundamentals of Database Systems*. The Benjamin/Cummings Publishing Company Inc., 1989.
- [Fag99] A. S. Fagundes. Projeto e Implementação de um Banco de Metadados para o Sistema de Informação de Biodiversidade do Estado de São Paulo. Master's thesis, Universidade Estadual de Campinas - IC, dezembro 1999.
- [FB94] S. Faiz and P. Boursier. Geoqual: a Data Model for Handling the Quality of Geographic Information. In ACM, editor, *Proceedings of 2nd ACM Workshop on Advances in Geographic Information Systems*, pages 107–114. 2nd ACM Workshop on Advances in Geographic Information Systems, Gaithersburg, MD, USA, december 1994.
- [Fer86] A. B. H. Ferreira. *Novo Dicionário da Língua Portuguesa*. Ed. Nova Fronteira, 1986.
- [FGD98a] FGDC. Federal Geographic Data Comitee. <http://www.fgdc.gov/metadata-/metahome.html>, julho 1998.

- [FGD98b] FGDC. The FGDC Content Standard for Digital Geospatial Metadata. <http://www.fgdc.gov/fdgc/fgdc.html>, setembro 1998. Version 1.0 june/8/1994.
- [Fis91] P. F. Fisher. Spatial Data Sources and Data Problems. In M. F. Goodchild, D. J. Maguire, and D. W. Rhind, editors, *Geographical Information Systems - Principles and Applications - Vol. I*, chapter 13, pages 175–189. Longman Scientific & Technical, 1991.
- [FK95] A. U. Frank and W. Kuhn, editors. *Spatial Information Theory: A Theoretical Basis for GIS - International Conference COSIT'95*. Springer-Verlag, september, 21–23 1995. Lecture Notes in Computer Science.
- [Fra98] A. U. Frank. Metamodels for Data Quality Description. In Michael Goodchild and Robert JeanSoulin, editors, *Data Quality in Geographic Information - From Error to Uncertainty*, chapter Quality Concepts and Models, pages 15–29. Hermes, 1998.
- [Gro99] ANZLIC Working Group. ANZLIC Guidelines: Core Metadata Elements. <http://anzlic.org.au/metaelem.htm>, agosto 1999. Version 1.
- [Gun98] O. Gunther. *Environmental Information Systems*. Springer-Verlag, 1998.
- [Gup98] S. C. Guphill. Building a Geospatial Data Framework - Finding the Best Available Data. In Michael Goodchild and Robert JeanSoulin, editors, *Data Quality in Geographic Information - From Error to Uncertainty*, chapter Quality Concepts and Models, pages 31–36. Hermes, 1998.
- [Har98] F. Harvey. Quality Needs More than Standards. In Michael Goodchild and Robert JeanSoulin, editors, *Data Quality in Geographic Information - From Error to Uncertainty*, chapter Quality Concepts and Models, pages 37–42. Hermes, 1998.
- [HF97] S. C. Hirtle and A. U. Frank, editors. *Spatial Information Theory: A Theoretical Basis for GIS - International Conference COSIT'97*. Springer-Verlag, october, 15–18 1997. Lecture Notes in Computer Science.
- [Hoh98] P. Hohl. *GIS Data Conversion; Strategies, Techniques and Management*. Onword Press, 1998.
- [ICA99] ICA. International Cartographic Association. <http://www.geog.psu.edu/ica/-ICAunsc.html>, novembro 1999.

- [ISO99] ISO. ISO/TC 211 Geographic Information/Geomatics. <http://www.statkart.no/isotc211/scope.htm>, novembro 1999.
- [JL99] C. A. Davis Jr. and A. H. F. Laender. Múltiplas Representações em Aplicações Urbanas de Sistemas de Informação Geográficos. In INPE, editor, *Proceedings of the I Brazilian Workshop on Geo Informatics*, pages 35–39, october 20-21 1999. Unicamp - Campinas - SP.
- [Jun99] J. B. Junior. Qualidade de Software. <http://www.barreto.com.br/qualidade>, setembro 1999.
- [Jur88] J. M. Juran. *Juran on Planning for Quality*. Juran Institute Inc., 1988.
- [KPD<sup>+</sup>95] K. Kelly, T. Pardo, S. Dawes, A. DiCarerino, and W. H'erad. Sharing the cost, sharing the benefits: The NYS GIS Cooperative Project,. Technical report, New York Department of Environmental Conservation Center for Technology in Government, december 1995.
- [KS95] H. F. Korth and A. Silberschatz. *Sistema de Bancos de Dados*. MAKRON Books - SP, 1995.
- [Lum96] J. Lummaux. A normalização em informação geográfica: porquê? *Le Géométre*, fevereiro 1996. <http://snig.cnig.pt/standards>.
- [MA99] C. B. Medeiros and A. C. Alencar. Qualidade dos Dados e Interoperabilidade em SIGs. In INPE, editor, *Proceedings of the I Brazilian Workshop on Geo Informatics*, pages 45–49, october 20-21 1999. Unicamp - Campinas - SP.
- [Mag91] D. J. Maguire. An Overview and Definition of GIS. In M. F. Goodchild, D. J. Maguire, and D. W. Rhind, editors, *Geographical Information Systems - Principles and Applications - Vol. I*, chapter 1, pages 9–20. Longman Scientific & Technical, 1991.
- [Mar91] M. Martinelli. *Curso de Cartografia Temática*. Editora Contexto, 1991.
- [MM98] R. Musick and C. Miller. Report on the Second IEEE Metadata Conference (Metadata'97). *SIGMOD Record*, 27(1):11–15, march 1998.
- [Mon91] M. Monmonier. *How to Lie with Maps*. The University of Chicago Press, 1991.
- [Mon95] D. R. Montello. How Significant are Cultural Differences in Spatial Cognition? In A. U. Frank and W. Kuhn, editors, *Spatial Information Theory: A Theoretical Basis for GIS - International Conference COSIT'95*, pages 485–500. Springer-Verlag, september, 21–23 1995. Lecture Notes in Computer Science.

- [MP94] C. B. Medeiros and F. Pires. Databases for GIS. *ACM SIGMOD Records*, 23(1):107-115, 1994.
- [MS93] G. E. Montgomery and H. C. Schuch. *GIS Data Conversion Handbook*. GIS World Books, 1993.
- [MSM97] W. E. Moen, E. L. Stewart, and C. R. McClure. The Role of Content Analysis in Evaluation Metadata for the U. S. Government Information Locator Service (GILS): results from Exploratory Study. *Second IEEE Metadata Conference*, september 1997.
- [MVW96] C. B. Medeiros, G. Vossen, and M. Weske. GEO-WASA - Combining GIS Techniques with Workflow Management. Technical Report 02/96-I, University of Munster, Germany, 1996.
- [NAS98] NASA. Nasa's Global Change Master Directory. <http://gcmd.gsfc.nasa.gov>, setembro 1998.
- [NAS99] NASA. Landsat. <http://www.landsat.gsfc.nasa.gov>, maio 1999.
- [NAV99] NAVCEN. U. S. Coast Guard Navigation. <http://www.navcen.uscg.mil/gps>, novembro 1999.
- [NCG99] NCGIA. International Conference and Workshop on Interoperating Geographic Information Systems, 1997. <http://www.ncgia.ucsb.edu/conf/interop97>, outubro 1999.
- [Oli88] C. Oliveira. *Curso de Cartografia Moderna*. Editora Fundação Instituto Brasileiro de Geografia e Estatística, 1988.
- [Oli95] C. M. M. Oliveira. Cartas Náuticas Eletrônicas: Operações e Estruturas de Dados. Master's thesis, Universidade Estadual de Campinas - DCC, novembro 1995.
- [Oli97] J. L. Oliveira. *Projeto e Implementação de Interfaces para Sistemas de Aplicações Geográficas*. PhD thesis, Universidade Estadual de Campinas - IC, dezembro 1997.
- [Ols97] J. Olsenholler. Find the Right Fit - How to Select the Best Data Collection Strategy. *GIS World*, 10(7):52-53, 1997.
- [Onl99] Environment Australia Online. Distributed Spatial Data Library. <http://www.environmental.gov.au/>, setembro 1999.

- [Orr98] K. Orr. Data Quality and Systems Theory. *Communications of the ACM*, 41(2):66–71, 1998.
- [Pie90] C. S. Pierce. *Semiótica*. Editora Perspectiva S. A., 1990. Tradução do original "The Collected Papers of Charles Sanders Pierce".
- [Pir97] F. Pires. *Um Ambiente Computacional para Modelagem de Aplicações Ambientais*. PhD thesis, Universidade Estadual de Campinas - IC, dezembro 1997.
- [Rad91] F. J. Radermacher. The Importance of Metaknowledge for Environmental Information Systems. In O. Gunther and H. J. Scheck, editors, *LNCS 525*, chapter Advances in Spatial Databases, pages 35–44. Springer-Verlag, 1991.
- [Rot96] J. Rothemberg. Metadata to Support Data Quality and Longevity. *First IEEE Metadata Conference*, abril 1996.
- [San95] L. Santaella. *A Teoria Geral dos Signos - Semiose e Autogeração*. Editora Ática, 1995.
- [SE90] J. Star and J. Estes. *Geographic Information Systems - An Introduction*. Prentice Hall Inc., 1990.
- [SEB94a] SEBRAE. Como medir os resultados e melhorar ainda mais. *Folha de São Paulo*, Qualidade Total(Fascículo 8):1, maio 1994.
- [SEB94b] SEBRAE. Uma história que conquistou o mundo. *Folha de São Paulo*, Qualidade Total(Fascículo 2):3, março 1994.
- [Sef98] L. A. Seffino. WOODSS - Sistema Espacial de Apoio ao Processo Decisório baseado em Workflows. Master's thesis, Universidade Estadual de Campinas - IC, julho 1998.
- [Sin78] D. Sinton. The Inherent Structure of Information as a Constraint to Analysis: Mapped Thematic Data as a Case Study. In Dutton G., editor, *Harvard Papers on Geographic Information Systems*, volume 6. Addison-Wesley, 1978. Reading MA.
- [TB98] G. K. Tayi and D. P. Ballou. Examining Data Quality. *Communications of the ACM*, 41(2):54–57, 1998.
- [TIP97] Clark University The IDRISI Project. Idrisi for windows academic, version 2.00.03. CD, june 1997.

- [USG98] USGS. The Spatial Data Transfer Standard. <http://mcmcweb.er.usgs.gov/sdts/standard.html>, setembro 1998.
- [Wan94] R. Y. Wang. Towards a Natural Language User Interface: on Approach of Fuzzy Query. *Journal of Geographic Information Systems*, 08(2):143-162, 1994.
- [Wan98] R. Y. Wang. A Product Perspective on Total Data Quality Management. *Communications of the ACM*, 41(2):58-65, 1998.
- [Wol83] P. Wolf. *Elements of Photogrammetry*. McGraw Hill Book Co., 1983.
- [WSG96] R. Y. Wang, D. M. Strong, and L. M. Guarascio. Beyond Accuracy - What Data Quality means to Data Consumers. *Journal of Management Information Systems*, 12(4):05-34, 1996.
- [WVM96] M. Weske, G. Vossen, and C. B. Medeiros. Scientific Workflow Management: WASA Architecture and Applications. Technical Report 03/96-I, University of Munster, Germany, 1996.
- [WVM98] M. Weske, G. Vossen, and C. B. Medeiros. Workflow Management in Geoprocessing Applications. Technical Report 02/98-I, University of Munster, Germany, 1998.