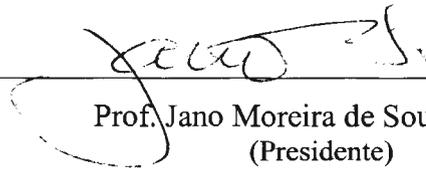


INTEGRAÇÃO DE BASES DE DADOS GEOGRÁFICOS HETEROGÊNEAS E DISTRIBUÍDAS

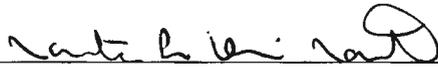
Julia Celia Mercedes Strauch

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO EM ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Aprovada por:



Prof. Jano Moreira de Souza, Ph.D.
(Presidente)



Prof.ª Marta Lima de Queirós Mattoso, D.Sc.



Prof. Geraldo Bonorino Xexéo, D.Sc.



Prof. Gilberto Câmara, Ph.D.



Prof.ª Vania Maria Ponte Vidal, D.Sc.



Prof. Cirano Iochpe, Ph.D.

RIO DE JANEIRO, RJ - BRASIL

DEZEMBRO DE 1998

STRAUCH, JULIA CELIA MERCEDES

Integração de bases de dados geográficos heterogêneas e distribuídas [Rio de Janeiro] 1998

XVII, 248 p. 29,7 cm (COPPE/UFRJ, D.Sc., Engenharia de Sistemas e Computação, 1998)

Tese - Universidade Federal do Rio de Janeiro, COPPE

1. federação de Sistema de Informações Geográficas 2. metodologia para integração de esquemas 3. arquitetura para interoperabilidade entre SIG

I. COPPE/UFRJ II. Título (série)

Aos meus Amores,

Tino e Daniel

Agradecimentos,

Ao Professor Jano Moreira de Souza, pela acolhida em seu projeto de pesquisa, sua orientação e incentivo que possibilitaram o desenvolvimento deste trabalho.

À Professora Marta Lima de Queirós Mattoso, pela sua orientação e valiosas sugestões, que muito acrescentaram ao desenvolvimento deste trabalho.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo suporte financeiro a este trabalho.

Aos membros de minha banca de tese, Prof. Geraldo Bonorino Xexéo, Prof. Gilberto Câmara, Vânia Maria Ponte Vidal e Prof. Cirano Iochpe pelas sugestões ao meu trabalho.

Aos companheiros e ex-companheiros da linha de Banco de Dados do Programa de Engenharia de Sistemas e Computação.

À EMBRAPA/CNPS por me conceder a oportunidade de concretizar este trabalho, proporcionando a minha liberação.

À minha família e, em especial, ao meu marido, Constantino Gonçalves Ribeiro, pelo apoio, paciência e compreensão, sem os quais não teria sido possível realizar o doutorado.

Aos meus amigos Lauro Charlet Pereira, José Carlos Louzada Morelli e Andréa Pereira Ramos, pelo apoio e incentivo, nos momentos de angústia e solidão.

E aos demais professores, colegas e funcionários da COPPE/Sistemas, e a todos aqueles que de forma direta ou indireta contribuíram para a elaboração deste trabalho.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para obtenção do grau de doutor em Ciências (DSc.).

INTEGRAÇÃO DE BASES DE DADOS GEOGRÁFICOS HETEROGÊNEAS E DISTRIBUÍDAS

Julia Celia Mercedes Strauch

Dezembro/1998

Orientador: Jano Moreira de Souza
Marta de Lima Queirós Mattoso

Programa: Engenharia de Sistemas e Computação

O alto custo da aquisição de dados geográficos aliado à complexidade das análises ambientais têm levado as comunidades de informação ao compartilhamento de dados existentes, armazenados por bases de dados de Sistemas de Informações Geográficas (SIG). Os mecanismos utilizados para proporcionar a troca de dados entre estas bases têm sido a conversão de formatos e o estabelecimento de padrões. Estas soluções são inadequadas, uma vez que requerem localizar os dados e compreender os formatos ou padrões envolvidos para estabelecer os processos de conversão e não garantem interoperabilidade, pois parte substancial da informação está contida nos procedimentos de consulta, manipulação e apresentação.

Essa tese apresenta o **Sistema MultiSIG** que tem como proposta criar uma federação de dados aberta. Este Sistema objetiva proporcionar interoperabilidade semântica entre bases de dados geográficos heterogêneas e distribuídas sobre o ambiente da rede global Internet. O desenvolvimento deste Sistema contemplou as diferenças entre: **i)** os modelos de dados usados pelos SIG, **ii)** as linguagens de manipulação de dados oferecidas por estes sistemas, e **iii)** os esquemas decorrentes das diversas áreas de

aplicação através de uma metodologia e de uma arquitetura, denominadas *MultiSIG* e *MultiSIG*, respectivamente. A *MultiSIG* estabelece as atividades necessárias para a integração de esquemas das bases de dados geográficos e a *MultiSIG* oferece um ambiente que proporciona a interoperabilidade entre estas bases.

As principais contribuições decorrentes do desenvolvimento deste Sistema são:

- a proposta de um nível externo sobre o modelo de dados global, gerando um modelo de objetos para um esquema global externo. Este modelo de objetos fornece uma única percepção do significado dos dados armazenados pelas bases componentes da federação, mantendo suas características de temas e regiões;
- a metodologia *MultiSIG* para integração dos esquemas - contempla as diversas características envolvidas na construção de uma base de dados geográficos, de forma que um conjunto compartilhado de termos e definições utilizados pelas comunidades de informação seja definido e apresentado no esquema global; e
- a especificação da arquitetura *MultiSIG* para interoperabilidade: é especificado os serviços desta arquitetura, bem como serviços de suporte a outros trabalhos do Projeto WHIP (*Workbench* de Heterogeneidade, Interoperabilidade e Paralelismo). Deste modo, a arquitetura proposta oferece ferramentas para cadastrar o esquema, realizar consultas, solicitar apresentação de resultados e fornecer suporte à cooperação entre os administradores locais e global para a integração de esquemas.

A metodologia e a arquitetura propostas foram avaliadas na integração de bases de dados voltadas para aplicações na área de hidrologia e ambiental. Os resultados desta implementação evidenciam que a interoperabilidade proporcionada pelo **Sistema MultiSIG** contribui como uma forma alternativa de acessar dados dispostos em outras bases de dados geográficos, preservando o investimento na informação geográfica armazenada pelas bases de dados. A estratégia adotada pelo **Sistema MultiSIG** se mostra eficiente, uma vez que não é necessário saber onde os dados estão localizados, além de promover maior sinergia entre as comunidades de informação.

Abstract of Thesis presented to COPPE/UFRJ as partial fulfillment of the requirements for the degree of Doctor of Science (DSc.).

INTEGRATION OF HETEROGENEOUS AND DISTRIBUTED DATABASE OF GEOGRAPHIC DATA

Julia Celia Mercedes Strauch

December, 1998

Advisors Jano Moreira de Souza
 Marta de Lima Queirós Mattoso

Department: Systems and Computer Engineering

The high cost of the geographical data acquisition combined with to the complexity of the environmental analysis has been driving the information communities to sharing the existent data stored by Geographical Information Systems (GIS) Databases. The mechanisms adapted to provide the data exchanges among these databases have undertaken formats conversions and standards establishment. These solutions are inadequate because they require the location of the data, to understand the formats or standards involved and the establishment of conversion processes. These solutions do not guarantee interoperability, since part of the information is contained in the query, manipulation and presentation procedures of geographical data.

This work develops the **MultiGIS System** that aims at creating an open data federation, which provides semantic interoperability among heterogeneous and distributed geographical databases on the Internet. The development of this System contemplates the differences among **i)** the data models used by GIS; **ii)** the data manipulation languages offered by the GIS; and **iii)** the resultant schema of the several application areas through a methodology and architecture named *MultiGIS* and

MultiGIS, respectively. MultiGIS establishes the required activities for the schema integration of the geographical databases and MultiGIS provides an interoperable environment among these databases.

The main contributions obtained with the development of this System are:

- The object model proposed for the external global schema - it provides a single perception of the meaning of the data stored by the component databases of the federation;
- MultiGIS methodology for schema integration - it encompasses several characteristics involved in the construction of a geographical database, so that a set of shared term and definitions used by the information communities is defined and presented in the global schema; and
- MultiGIS architecture specification for interoperability - it is specified the services of this architecture, as the services to support other works of the WHIP Project (Workbench for Heterogeneity, Interoperability and Parallelism). The proposed architecture offers tools to register the schema, to query, to request presentation of results and to support cooperation among the local and global administrators for schema integration.

The proposed methodology and architecture were evaluated in the integration of databases for hydrology and environmental application areas. The results of this implementation show that the interoperability provided by the **MultiGIS System** contributes significantly as an alternative form of access to stored geographical data in different geographical databases while preserving the investment of the geographical information stored by the original databases. The strategy adopted by the **MultiGIS System** is effective, since it does not require the knowledge of data location and promotes higher synergy among the information communities.

Sumário

CAPÍTULO I - INTRODUÇÃO	1
I.1. Motivação	1
I.2. Objetivo do trabalho	4
I.3. Contexto da tese: Projeto WHIP.....	6
I.4. Organização do trabalho	6
CAPÍTULO II - BASES DE DADOS GEOGRÁFICOS	8
II.1. Introdução	8
II.2. Sistema de Informações Geográficas (SIG).....	9
II.3. Áreas de aplicação de SIG.....	10
II.4. Dados geográficos.....	12
II.5. Conceitos cartográficos implícitos na construção de uma base de dados geográficos.....	13
II.5.1. <i>Data geodésicos</i>	14
II.5.2. <i>Sistemas de coordenadas</i>	16
II.5.3. <i>Sistema de projeção</i>	17
II.5.4. <i>Escala de uma base de dados geográficos</i>	19
II.5.5. <i>Semiologia gráfica</i>	20
II.6. Características das bases de dados geográficos	22
II.7. Considerações sobre os modelos de dados geográficos	24
II.8. Estrutura de dados geográficos e operações	25
II.9. Modelagem dos dados geográficos	28
II.10. Considerações sobre a aquisição de dados geográficos	31
II.11. Qualidade dos dados geográficos.....	32
CAPÍTULO III - INTEROPERABILIDADE ENTRE MÚLTIPLAS BASES DE DADOS	34
III.1. Introdução	34
III.2. Sistemas de suporte à interoperabilidade	35
III.3. Características de um ambiente MBD.....	38

III.3.1. Autonomia no ambiente MBD.....	39
III.3.2. <i>Cooperação no ambiente MBD</i>	41
III.4. Aspectos de representação dos dados em um ambiente MBD.....	43
III.4.1. <i>Semântica dos dados</i>	44
III.4.2. <i>Metadados: representação da semântica dos dados</i>	47
III.4.3. <i>Modelo de dados comum (MDC)</i>	48
III.5. Metodologia para integração de esquemas.....	53
III.5.1. <i>Anatomia dos esquemas envolvidos em um ambiente MBD</i>	55
III.5.2. <i>Heterogeneidades entre os esquemas das bases de dados</i>	58
III.5.3. <i>Etapas de uma metodologia para integração</i>	61
III.5.4. <i>Classificação das metodologias para integração</i>	69
III.6. Arquitetura para múltiplas bases de dados heterogêneas.....	72
III.6.1. <i>Federação rígida: arquitetura do esquema global</i>	74
III.6.2. <i>Federação flexível: arquitetura do mediador</i>	77
III.6.3. <i>Questões relacionadas ao desenvolvimento de uma arquitetura MBD</i>	79
III.6.4. <i>Características da linguagem de consulta para ambiente MBD</i>	83
III.6.5. <i>Considerações sobre o resultado de consultas em ambientes MBD</i>	84
III.6.6. <i>Tecnologia usada para implementar um ambiente MBD</i>	85
CAPÍTULO IV - REQUISITOS DE INTEROPERABILIDADE ENTRE BASES DE DADOS GEOGRÁFICOS E A PROPOSTA MULTISIG.....	88
IV.1. Introdução	88
IV.2. Evolução do compartilhamento de dados geográficos	89
IV.3. Propostas de interoperabilidade entre SIGs	93
IV.3.1. <i>Proposta de interoperabilidade do OGC: OGIS</i>	94
IV.3.2. <i>Propostas de interoperabilidade entre bases de dados de SIGs</i>	99
IV.3.3. <i>Análise comparativa das propostas encontradas na literatura</i>	103
IV.4. Característica das bases de dados geográficos pertencentes a diferentes comunidades de informação	105
IV.5. Requisitos de uma arquitetura para interoperabilidade entre SIG	106
IV.6. Requisitos de uma metodologia para integração de esquemas de bases de dados geográficos	109
IV.7. Requisitos de representação dos esquemas para bases de dados geográficos..	111
IV.7.1. <i>Requisitos dos metadados para um ambiente MBD de SIG</i>	111

<i>IV.7.2. Requisitos de representação do MDC para esquemas de bases de dados geográficos</i>	<i>113</i>
<i>IV.7.3. Requisitos do esquema global para dados geográficos.....</i>	<i>114</i>
<i>IV.8. Proposta de um ambiente MBD: Sistema MultiSIG.....</i>	<i>115</i>

CAPÍTULO V - <i>MULTISIG</i>: METODOLOGIA PROPOSTA PARA INTEGRAÇÃO DE ESQUEMAS DE BASES DE DADOS GEOGRÁFICOS	117
V.1. Introdução	117
V.2. Objetivo da <i>MULTISIG</i>	118
V.3. Característica da metodologia proposta.....	118
V.4. Entradas e saídas da metodologia.....	119
V.5. Política de participação no Sistema MultiSIG	121
V.6. Elementos utilizados pela metodologia	124
<i>V.6.1. Modelo de Dados do esquema global interno.....</i>	<i>125</i>
<i>V.6.2. Metadados utilizados pela metodologia.....</i>	<i>127</i>
<i>V.6.3. Thesaurus utilizado pela metodologia.....</i>	<i>129</i>
<i>V.6.4. Modelo de dados proposto para o esquema global externo</i>	<i>129</i>
V.7. Conceitos utilizados pela metodologia.....	132
<i>V.7.1. Objeto geográfico</i>	<i>133</i>
<i>V.7.2. Contextos.....</i>	<i>133</i>
<i>V.7.3. Definição de unidade territorial.....</i>	<i>135</i>
<i>V.7.4. Granularidade territorial de informação.....</i>	<i>135</i>
V.8. Taxionomia proposta para os conflitos entre os esquemas	136
V.9. Etapas da metodologia proposta para integração	143
<i>V.9.1. Etapa 1: Pré-integração.....</i>	<i>146</i>
<i>V.9.2. Etapa 2: Análise dos esquemas</i>	<i>152</i>
<i>V.9.3. Etapa 3: Conformação.....</i>	<i>175</i>
<i>V.9.4. Etapa 4: Integração no esquema global interno.....</i>	<i>177</i>
<i>V.9.5. Etapa 5: Criação do esquema global externo.....</i>	<i>180</i>
<i>V.9.6. Etapa 6: Validação</i>	<i>183</i>
V.10. Aplicação da metodologia	184
V.11. Comparação com outros trabalhos relacionados na literatura.....	185

CAPÍTULO VI - <i>MultiSIG</i>: ARQUITETURA PROPOSTA PARA A INTEROPERABILIDADE SEMÂNTICA	188
VI.1. Introdução.....	188
VI.2. Objetivo da <i>MultiSIG</i>	189
VI.3. Características da <i>MultiSIG</i>	189
VI.4. Funcionalidades do Sistema MultiSIG.....	191
VI.5. Linguagem de consulta proposta para o ambiente MultiSIG.....	192
VI.6. Considerações sobre os resultados das consultas no ambiente MultiSIG.....	195
VI.7. Componentes da <i>MultiSIG</i>	205
VI.7.1. Primeiro nível: interface com o usuário.....	197
VI.7.2. Segundo nível: de Interoperabilidade.....	199
VI.7.3. Terceiro nível: Servidores de dados.....	203
VI.8. Abordagens da Implementação do protótipo MultiSIG.....	204
VI.9. Serviços do Sistema MultiSIG.....	205
VI.9.1. Serviços de suporte a interface com o usuário.....	206
VI.9.2. Serviços do processador de consulta.....	208
VI.9.3. Serviços do tratador de resultados.....	209
VI.9.4. Serviços do Servidor de dados local.....	211
VI.10. Protótipo MultiSIG.....	211
CAPÍTULO VII - CONCLUSÕES E PERSPECTIVAS FUTURAS	215
VII.1. Resumo.....	215
VII.2. Análise das contribuições.....	217
VII.2. Perspectivas futuras.....	221
REFERÊNCIAS BIBLIOGRÁFICAS	223
7.1. Referências tradicionais.....	233
7.2. Referências bibliográficas <i>on line</i>	238
7.3. Bibliografia complementar tradicional.....	238
7.4. Bibliografia complementar <i>on line</i>	248
APÊNDICE A - ESQUEMAS DE BASES DE DADOS GEOGRÁFICOS UTILIZADOS	A.1

A.1. Introdução	A.1
A.1.1. Projeto Pró-Guaíba	A.1
A.1.2. Projeto SIG-RHidro	A.5
A.1.3. Projeto Pró-Paraíba	A.8
A.1.4. Projeto SIG-Ambiental	A.8
APÊNDICE B - APLICAÇÃO DA MULTISIG	B.1
B.1. Introdução	B.1
B.2. Pré-Integração	B.1
B.2.1. Descrição dos atributos dos esquemas	B.1
B.2.2. Descrição dos esquemas de exportação	B.13
B.3. Análise dos esquemas	B.22
B.3.1 Análise do contexto espacial dos esquemas.....	B.22
B.3.2. Análise do contexto da aplicação dos esquemas.....	B.25
B.3.3. Análise do contexto semântico dos esquemas.....	B.26
B.4. Conformação.....	B.43
B.5. Integração	B.44
B.6. Criação do esquema externo.....	B.71
B.6.1. Definição das instâncias da hierarquia Regions.....	B.71
B.6.2. Definição dos Themes.....	B.72
B.7. Integração do Esquema Ambiental.abs12 ao esquema resultante da integração de (Paraíba.abs1 e 2, RHidro.abs1).....	B.77
B.7.1. Análise do contexto semântico dos esquemas (Integração.abs, Ambiental.abs1).....	B.77
B.7.2. Conformação.....	B.91
B.7.3. Integração	B.92
B.7.4. Criação do esquema externo.....	B.103
B.8. Validação	B.106

Lista de Figuras

<i>Figura II.1 -- Sistemas de coordenadas.</i>	16
<i>Figura II.2 - Organização da base de dados em níveis de informação.</i>	23
<i>Figura II.3 - Exemplo de análise do potencial de erosão.</i>	23
<i>Figura II.4 - Modelagem dos dados.</i>	28
<i>Figura III.1 - Ambiente MBD.</i>	37
<i>Figura III.2 - Analogia para as autonomias de projeto, execução e comunicação (ÖSZU & VALDURIEZ, 1990).</i>	40
<i>Figura III.3 - Semântica do dado.</i>	45
<i>Figura III.4- Esquemas envolvidos em uma arquitetura em cinco níveis.</i>	57
<i>Figura III.5- Taxionomia das metodologias de integração (ÖSZU & VALDURIEZ, 1990).</i>	71
<i>Figura III.6- Arquitetura genérica da federação rígida.</i>	76
<i>Figura III.7- Arquitetura genérica para ambiente MBD.</i>	78
<i>Figura III.8- Processamento de consulta em ambiente MBD (ÖSZU & VALDURIEZ, 1990).</i>	80
<i>Figura III.9- Arquitetura cliente servidor.</i>	86
<i>Figura IV.1- Conversão de dados.</i>	90
<i>Figura IV.2 - Comparação do número de conversores com/sem padrão de dados (as setas significam duas transformações de arquivos).</i>	91
<i>Figura IV.3- Hierarquia de classes do modelo de dados geográficos do OGIS (GARDELS, 1996b).</i>	96
<i>Figura IV.4- Visão geral da arquitetura OGIS segundo GARDELS (1996a).</i>	98
<i>Figura IV.5 - Componentes do Open EIS (GÜNTHER, 1997).</i>	101
<i>Figura V.1. - Esquemas utilizados pela MultiSIG.</i>	120
<i>Figura V.2 - Definição da classe GeographicObject do SAIF.</i>	126
<i>Figura V.3 - Modelo de dados do esquema global.</i>	131
<i>Figura V.4 - Hierarquia Region.</i>	132
<i>Figura V.5 -- MultiSIG.</i>	144
<i>Figura V.6 - Justaposição de esquemas sem relacionamento espacial a hierarquia Region.</i>	156
<i>Figura V.7 - Exemplo de relacionamento espacial parcial.</i>	157
<i>Figura V.8 - Integração e união de esquemas com relacionamento espacial, parcial ao esquema global externo.</i>	158
<i>Figura V.9 - Exemplo das hierarquias contexto de aplicação.</i>	160
<i>Figura V.10- Exemplo de seqüência de integração dos esquemas.</i>	161
<i>Figura V.11- Exemplo: modelagem dos relacionamentos e suas respectivas classes.</i>	166
<i>Figura V.12 - Exemplo: modelagem dos relacionamentos considerando a geometria.</i>	168

<i>Figura V.13- Exemplo: modelagem dos relacionamentos considerando as propriedades.....</i>	<i>170</i>
<i>Figura VI.1-Componentes da arquitetura MultiSIG.....</i>	<i>197</i>
<i>Figura VI.2 - Homepage do Sistema MultiSIG.....</i>	<i>213</i>
<i>Figura VI.3 - Sistema MultiSIG: interface principal</i>	<i>213</i>
<i>Figura VI.4 - Sistema MultiSIG:interface de cadastro do esquema de exportação</i>	<i>214</i>
<i>Figura VI.5 - Interface de visualização do esquema global externo.....</i>	<i>214</i>
<i>Figura VI.6 - Interface de consulta do esquema global externo.....</i>	<i>214</i>
<i>Figura A.1 -Esquema do projeto Pró-Guaíba: Abstração cartográfica em 1:10.000.....</i>	<i>A.2</i>
<i>Figura A.2 -Esquema do projeto Pró-Guaíba: Abstração cartográfica em 1:250.000.....</i>	<i>A.3</i>
<i>Figura A.3- Esquema do projeto SIG-RHidro: Abstração cartográfica em 1:100.000.....</i>	<i>A.5</i>
<i>Figura A.4 - Esquema do projeto SIG-RHidro: Abstração cartográfica em 1:100.000 (parte 2).</i>	<i>A.6</i>
<i>Figura A.5 - Esquema do projeto Teste no modelo de dados do SPRING: 1:100.000.....</i>	<i>A.10</i>

Lista de Tabelas

<i>Tabela II.1 - Taxionomia de área de aplicações</i>	10
<i>Tabela III.1 Comparação das taxionomia de heterogeneidades encontradas na literatura</i>	60
<i>Tabela IV.1 - Comparação das propostas encontradas na literatura</i>	104
<i>Tabela V.1 - Extensões do SAIF.</i>	128
<i>Tabela V.2 - Data nacionais.</i>	137
<i>Tabela V.3- Correspondências entre as estruturas gráficas do ARC/Info e as classes do SAIF.</i>	151
<i>Tabela V.4 - Correspondência entre as classes do SPRING e do SAIF.</i>	151
<i>Tabela V.5- TPC com as propriedades que definem os parâmetros cartográficos.</i>	154
<i>Tabela V.6- TLG com as propriedades que definem a localização geográfica.</i>	155
<i>Tabela V.7- Relacionamentos espaciais.</i>	156
<i>Tabela V.8- Propriedades que descrevem o contexto da área de aplicação.</i>	159
<i>Tabela V.9 - Exemplo de TCA para quatro esquemas.</i>	160
<i>Tabela V.10- Exemplo de TCA para o quinto esquema.</i>	161
<i>Tabela V.11- Comparação entre as denominações das classes.</i>	163
<i>Tabela V.12- Exemplo de TC.</i>	164
<i>Tabela V.13- Grupos Temáticos.</i>	165
<i>Tabela V.14- Relacionamentos das classes de cada grupo.</i>	165
<i>Tabela V.15- Exemplo de TR.</i>	166
<i>Tabela V.16- Geometria das classes com proximidade semântica.</i>	167
<i>Tabela V.17- Exemplo 1 da TG.</i>	167
<i>Tabela V.18- Exemplo 2 da TG.</i>	167
<i>Tabela V.19- Tabela não normalizada das propriedades.</i>	168
<i>Tabela V.20- Exemplos de TNNP.</i>	169
<i>Tabela V.21- Tabela com as correspondências entre as propriedades.</i>	170
<i>Tabela V.22- Exemplo: Tabelas de correspondência entre as propriedades.</i>	171
<i>Tabela V.23- Graus parciais de proximidade semântica.</i>	172
<i>Tabela V.24- Tabela de classes Idênticas.</i>	174
<i>Tabela V.25- Tabela de classes similares.</i>	174
<i>Tabela V.26- Tabelas de classes heterogêneas.</i>	174
<i>Tabela V.27- Heterogeneidades espaciais e de áreas de aplicação entre os esquemas.</i>	184
<i>Tabela A.1 - Estruturas físicas usadas para as estruturas do esquema em 1:10.000.</i>	A.4
<i>Tabela A.2 - Estruturas físicas usadas para as estruturas do esquema em 1:250.000.</i>	A.4
<i>Tabela A.3 - Estruturas físicas usadas para as estruturas do esquema em 1:250.000.</i>	A.7
<i>Tabela A.4 - Definição dos objetos geográficos no SPRING.</i>	A.8

<i>Tabela B.1 - Tabela de Parâmetros Cartográficos (TPC)</i>	B.22
<i>Tabela B.2 - Tabela de Localização Geográfica (TLG)</i>	B.23
<i>Tabela B.3 - Tabela do Contexto da Aplicação(TCA)</i>	B.25
<i>Tabela B.4 - Tabela de Classes (TC)</i>	B.26
<i>Tabela B.5 - Tabela de Relacionamentos (TR)</i>	B.79
<i>Tabela B.6 - Tabela de Geometrias (TG)</i>	B.34
<i>Tabela B.7 - Tabela de Correspondência entre Propriedades (TCP)</i>	B.38
<i>Tabela B.8 -Tabela de Classes Idênticas</i>	B.39
<i>Tabela B.9 - Tabela de Classes Similares</i>	B.40
<i>Tabela B.10 - Tabelas de Classes Heterogêneas do esquema Paraíba</i>	B.40
<i>Tabela B.11 - Tabela de Classes Heterogêneas do esquema RHidro</i>	B.41
<i>Tabela B.12 - Tabela de Classes Heterogêneas do esquema Guaíba</i>	B.42
<i>Tabela B.13 -Tabela de Classes Similares Conformadas</i>	B.44
<i>Tabela B.14 - Tabela de Classes (TC)</i>	B.79
<i>Tabela B.15 - Tabela de Relacionamentos (TR)</i>	B.82
<i>Tabela B.16 - Tabela de Geometrias (TG)</i>	B.86
<i>Tabela B.17 - Tabela de Correspondência entre Propriedades (TCP)</i>	B.89
<i>Tabela B.18 -Tabela de Classes Idênticas</i>	B.90
<i>Tabela B.19 - Tabela de Classes Similares</i>	B.90
<i>Tabela B.20 - Tabela de Classes Heterogêneas do esquema Ambiental</i>	B.91
<i>Tabela B.21 - Tabela de Classes Heterogêneas do esquema Integração</i>	B.92
<i>Tabela B.22 - Tabela de Classes Similares</i>	B.92

Capítulo I - Introdução

I.1. Motivação

O grande volume e diversidade de informações georreferenciadas têm conduzido à necessidade de diversas áreas desenvolverem bases de dados contendo objetos geográficos relacionados a uma grande variedade de aplicações ambientais. Essas bases têm-se consolidado à medida que a tecnologia tem proporcionado ambientes para o desenvolvimento de Sistemas de Informações Geográficas (SIGs), que permitem a captura de dados, análise e apresentação de informações espaciais relacionadas (ARONOFF, 1989; BERNHARDSEN, 1992; LAURINI & TOMPSON, 1992).

Estes sistemas possibilitam as diversas áreas de aplicação combinarem dados que descrevem fatos ou fenômenos, do mundo real, associados a uma representação cartográfica digital, a fim de realizar análises espaciais de acordo com critérios definidos pelos usuários.

O alto custo da aquisição de dados geográficos para a implantação de um SIG aliada à crescente complexidade das análises ambientais, que a cada dia exigem novos parâmetros de diversos domínios de conhecimento não previstos nas questões manifestas e latentes, requerem o acesso a dados multidisciplinares, armazenados em diversas bases de dados, localizadas em diferentes instituições. Isto tem conduzido à necessidade de reutilização destes dados, através do acesso e troca de dados entre estes sistemas (ONSRUD & RUSHTON, 1995; CASSETARI, 1993).

Todavia, o desenvolvimento da tecnologia de informação levou os aplicativos para o desenvolvimento de SIGs a apresentarem uma variedade de modelos de dados e se utilizarem de diferentes plataformas, redes de comunicação e sistemas operacionais (MATUSCHAK^{html}, 1996). Esta diversidade aumenta quando se considera que as bases de dados destes sistemas foram projetadas por instituições independentes, não cooperativas, onde cada uma tem absoluta autoridade e controle sobre seus dados. Assim, estas bases são projetadas de acordo com as necessidades organizacionais e

apresentam aplicações monolíticas, dependentes de plataformas e aplicativos, com facilidades limitadas para compartilhar os dados e os recursos de computação.

Entretanto, o compartilhamento dos dados é desejável, uma vez que nestas bases os objetos têm, em comum, o mesmo espaço geográfico descrito. As soluções atualmente utilizadas para proporcionar este compartilhamento têm sido a conversão de formatos e o estabelecimento de padrões. A primeira solução consiste em converter arquivos de dados de um formato para outro. A complexidade deste processo para os dados geográficos é maior que para os dados convencionais, uma vez que estes são descritos por diferentes componentes gráficos que são relacionados a diversos componentes não gráficos. A segunda solução consiste em estabelecer um denominador comum entre os diferentes SIGs através da especificação de padrões de dados que descrevem objetos geográficos. Estes padrões cobrem os formatos de arquivos, procedimentos administrativos e diversos parâmetros para descrever a proveniência e a qualidade dos dados. Contudo a aplicabilidade destes padrões não garante um compartilhamento de dados adequado, uma vez que parte substancial da informação está contida nos procedimentos de manipulação, consulta e apresentação dos objetos geográficos. Cabe ainda ressaltar que, de modo geral, estas soluções requerem localizar os dados e compreender os formatos ou padrões de dados envolvidos para estabelecer o processo de conversão. Além disto, as soluções não dispõem de facilidades para acomodar novos tipos de dados e exigem à necessidade de conversão e migração reversa de parte dos dados que foram atualizados, para evitar inconsistências entre sistemas (CLEMENT *et al.*, 1997).

Atualmente, no dominante paradigma da globalização, onde as organizações são caracterizadas por serem distribuídas sobre vários locais, é cada vez maior a necessidade de realizar operações globais sem centralização da informação ou perda da autonomia dos sistemas locais. Desta forma, a tendência de compartilhamento de dados entre SIGs tem sido substituir os ambientes fechados, com soluções orientadas a produtos, por ambientes abertos caracterizados por diferentes plataformas, aplicativos e bases de dados que interoperam em tempo real (LAURINI, 1994; GOODMAN, 1994; GARDELS, 1997).

Estas soluções são encontradas no âmbito da tecnologia de Banco de Dados, sob diversas denominações: múltiplas bases de dados (MBD), mediador, base de dados federada e sistemas interoperáveis (SHETH & LARSON, 1990; BREITBART, 1990;

WIEDERHOLD, 1992; AHMED *et al.*, 1995; BUNEMAN, 1997). Elas têm por objetivo preservar os sistemas existentes, mantendo suas autonomias, através de um sistema que possui facilidades para o acesso e compartilhamento dos dados armazenados pelas bases de dados locais.

As pesquisas para o desenvolvimento destas soluções têm sido em duas direções. A primeira tem levado ao desenvolvimento de arquiteturas que permitem a interoperabilidade através de aplicações independentes dos componentes de informática. Estas arquiteturas resolvem as heterogeneidades físicas entre os ambientes caracterizadas por diferentes plataformas, sistemas operacionais e redes de comunicação (PERIZZO *et al.*, 1991; MENG & YU, 1995; KARLPALEM *et al.*, 1995; MANOLA, 1995). Contudo, acima das heterogeneidades físicas existem conflitos semânticos e sintáticos decorrentes das diferenças entre os modelos de dados e do processo de modelagem. Deste modo, a segunda direção das pesquisas tem sido o desenvolvimento de metodologias para capturar a semântica dos esquemas das bases de dados, compreender as diferenças entre eles, os requisitos de consistência, as restrições de integridade e aplicar técnicas desenvolvidas para resolver os conflitos que por ventura existirem (SOUZA, 1986; BATINI *et al.*, 1986; KIM & SEO, 1991; LIM *et al.*, 1994; REDDY *et al.*, 1994; FRANKHAUSER *et al.*, 1995).

As arquiteturas que proporcionam interoperabilidade vêm se beneficiando do paradigma da orientação a objeto aplicado a arquiteturas cliente servidor, visto que esta combinação oferece modularidade. Esta modularidade é alcançada através da utilização de objetos clientes e servidores, distribuídos, que encapsulam as bases de dados locais. A interoperabilidade é obtida através de um protocolo comum para a troca de mensagens entre estes objetos distribuídos (BERTINO *et al.*, 1994; PITOURA *et al.*, 1995; HURSON & BRIGHT, 1996; SILBERCHATZ & ZDONICK, 1997). Uma nova direção na implementação destas arquiteturas é a utilização dos recursos da rede global Internet (MATUSCHAK^{html}, 1996). A utilização desta rede permite o desenvolvimento de um ambiente MBD com maiores proporções. Além disto, ela oferece uniformidade de acesso através da interface com o usuário e dos recursos de comunicação.

As metodologias para resolver as heterogeneidades das bases de dados, nos últimos anos, também vêm se beneficiando do paradigma da orientação a objeto na adoção de um modelo de dados canônico. Um modelo de dados canônico orientado a objeto oferece maior expressividade semântica para capturar melhor o significado dos

dados armazenados (BERTINO *et al.*, 1994).

Estender estas pesquisas para desenvolver um sistema MBD que suporte bases de dados de SIG, localizados em diferentes instituições, representa uma evolução das soluções tradicionais. O desenvolvimento deste ambiente para SIGs oferece inúmeros benefícios, uma vez que preserva os investimentos técnico e organizacional das instituições em seus sistemas. No nível técnico, as bases de dados dos SIGs são bases não convencionais, cujo processo de modelagem e implementação é complexo e dispendioso, envolvendo treinamento de pessoal, aquisição dos dados, calibração do modelo e cuja re-implementação ou modificação é uma tarefa desalentadora ou até mesmo inviável. No nível organizacional, a instalação de um SIG e sua inclusão dentro do processo operacional em uma instituição, exige investimentos científicos e financeiros, cuja realocação e/ou modificação envolve dificuldades organizacionais. Além disto, um sistema MBD oferece um ambiente que difunde os dados geográficos de forma eficiente e a um menor custo, facilitando a aquisição e o intercâmbio de informações, e futuramente expandindo a base de aplicações dos usuários.

1.2. Objetivo do trabalho

O objetivo do presente trabalho consiste em apresentar soluções para a integração de bases de dados geográficas heterogêneas e distribuídas. Essa solução requer a caracterização das heterogeneidades e o desenvolvimento de uma metodologia e de uma arquitetura que possibilite a interoperabilidade no nível semântico e no nível físico, respectivamente.

Para atender este objetivo esse trabalho propõe o desenvolvimento do **Sistema MultiSIG**, que proporciona uma federação aberta de dados geográficos. O desenvolvimento deste sistema demanda: **i)** uma metodologia para capturar a semântica dos esquemas das bases de dados e integrá-los; e **ii)** a especificação de uma arquitetura que possibilite a interoperabilidade entre as bases de dados de SIGs, heterogêneas e distribuídas.

A interoperabilidade, no contexto do **Sistema MultiSIG**, denota que os usuários de duas ou mais bases de dados de SIGs interajam sobre um ambiente distribuído, de forma que recuperem e visualizem dados e o conteúdo de seus metadados.

A metodologia proposta, denominada de **MultiSIG**, visa criar o esquema

global externo, assegurando a inexistência de inconsistências resultantes da redundância de dados e dos conflitos entre esquemas das bases de dados dos SIGs. Para alcançar este objetivo a *MultiSIG* estabelece as atividades para: **a)** capturar a semântica dos esquemas das bases de dados dos SIGs locais; **b)** identificar os conflitos devidos às diferenças entre os espaços geográficos, as áreas de aplicação e a semântica dos esquemas das bases de dados dos SIGs; **c)** resolvê-los de forma que estabeleça uma concordância no conjunto compartilhado de termos e definições; **d)** integrar os esquemas locais no esquema global interno e **e)** derivar o esquema global externo.

A *MultiSIG* estende o metamodelo do padrão de dados SAIF (*Spatial Archive And Interchange Format*) e suas notações para descrever os esquemas das bases de dados locais no nível intencional e extencional. Ela também emprega um modelo de dados baseado em *Themes* e *Regions* proposto para o esquema global externo com o objetivo de permitir o usuário facilmente identificar os dados desejados.

Cabe ressaltar que a *MultiSIG*: *i)* se deteve apenas na semântica espacial apesar dos objetos apresentarem ainda semântica temporal, e *ii)* foi materializada com a extensão proposta para o metamodelo SAIF, todavia outros modelos poderiam ser estendidos.

A arquitetura proposta, denominada de *MultiSIG*, possibilita o compartilhamento de dados entre os usuários da federação. Para isto, a *MultiSIG* utiliza um esquema global que encapsula os dados armazenados pelas bases de dados dos SIGs locais. Este esquema global auxilia os usuários da federação a encontrar os temas e as regiões geográficas de interesse, a pensar sobre os dados que desejam, e a obter informações relacionadas à qualidade dos dados e outras questões relativas aos metadados, sem converter, reestruturar ou localizar os dados que fazem parte da federação. A *MultiSIG*, desenvolvida sobre a rede global Internet, se beneficia da arquitetura cliente-servidor, do paradigma orientado a objeto e emprega uma extensão da OQL (*Object Query Language*) como linguagem de consulta. Ela utiliza os recursos desta rede para estabelecer a comunicação entre as bases de dados de SIG distribuídos em diferentes ambientes computacionais. O protótipo inicial desta arquitetura ficou limitada a realização de consultas às bases de dados dos SIGs, de modo a validar o estudo de caso da *MultiSIG*.

Desta forma, o **Sistema MultiSIG** contempla as seguintes diferenças entre as

bases de dados dos SIGs:

- Diferença de modelo de dados - os dados são armazenados em bases de dados com diferentes modelos. Para contemplar esta diferença a *MultiSIG* propõe a conversão dos esquemas para o SAIF estendido, que é tratado como um modelo de dados comum. Isto permite normalizar as diferenças entre os esquemas e facilita identificar os conflitos entre eles;
- Diferenças entre esquemas decorrentes de diferentes áreas de aplicação - como as bases de dados geográficos foram projetadas e desenvolvidas independentemente, existem incompatibilidades entre os esquemas inerentes às áreas de aplicação. Para contemplar esta diferença, a *MultiSIG* gera um esquema global, sobre o qual o usuário da federação realiza uma consulta que será distribuída e traduzida para os esquemas das bases de dados locais dos SIGs; e
- Diferentes linguagens de consultas - os SIGs utilizam diferentes linguagens de consultas para manipular e consultar objetos geográficos. Para contemplar esta diferença, a *MultiSIG* apresenta o esquema global que disponibiliza temas e as regiões geográficas com os respectivos metadados, e utiliza uma interface de consulta que oferece uma sintaxe gráfica de consulta sobre o esquema global externo.

1.3. Contexto da tese: Projeto WHIP

Este trabalho faz parte de uma das metas do Projeto *Workbench* de Heterogeneidade, Interoperabilidade e Paralelismo (WHIP) do Programa de Engenharia de Sistemas e Computação da COPPE/UFRJ. Este projeto tem como objeto de pesquisa o desenvolvimento de uma plataforma que contemple arquiteturas que são desenvolvidas com o objetivo de abordar os aspectos de heterogeneidade, interoperabilidade e paralelismo nas áreas de aplicações voltadas para bases de dados geográficos e estatísticos.

1.4. Organização do trabalho

O presente trabalho está organizado em sete capítulos. Inicialmente o segundo capítulo, para uma melhor compreensão da metodologia, faz uma revisão na literatura apresentando a tecnologia de Geoprocessamento, a sua principal ferramenta - os

Sistemas de Informações Geográficas (SIGs) - e suas áreas de aplicação; caracteriza os dados geográficos, os principais conceitos cartográficos abordados no desenvolvimento de uma base de dados de SIG e apresenta ainda os principais aspectos do processo de modelagem dos dados geográficos.

No terceiro capítulo é realizada uma revisão na literatura sobre interoperabilidade entre múltiplas bases de dados. Neste capítulo é caracterizado o conceito de interoperabilidade, apresentado os principais aspectos para o desenvolvimento de um ambiente (MBD), as principais etapas de uma metodologia para a integração de esquemas e as duas principais arquiteturas para implementação deste ambiente.

O quarto capítulo faz algumas considerações sobre os requisitos de interoperabilidade entre bases de dados geográficos de diferentes comunidades de informação, e apresenta a proposta do **Sistema MultiSIG**. Neste capítulo também é caracterizada a necessidade de compartilhamento de dados geográficos, dissertando sobre os meios utilizados para a troca de dados geográficos e os esforços do *Open GIS Consortium* (OGC) para estabelecer uma especificação para a interoperabilidade entre SIGs.

No quinto capítulo é apresentada a *MultiSIG* quanto aos conceitos, os elementos utilizados e suas etapas. Neste capítulo é caracterizado ainda a heterogeneidade entre as bases de objetos geográficos, apresentando uma taxionomia para estes conflitos e soluções para tal. Este capítulo faz uma referência ao Apêndice B, onde é ilustrada uma aplicação da *MultiSIG*, a quatro esquemas heterogêneos e comparada com outros trabalhos relacionados na literatura.

No sexto capítulo é apresentada a *MultiSIG* quanto aos seus componentes e sua especificação. São tratados também os aspectos de implementação e um estudo de caso, ilustrando as experiências de consultas sobre esta arquitetura.

No sétimo capítulo são apresentados os comentários conclusivos deste trabalho, descrevendo as considerações e contribuições realizadas e dissertando sobre as conclusões e perspectivas futuras de pesquisa para o **Sistema MultiSIG**, no contexto do Projeto WHIP.

Capítulo II - Bases de Dados Geográficos

II.1. Introdução

Os avanços tecnológicos da cartografia automatizada, dos Sistemas de Gerenciamento de Banco de Dados e do processamento digital de imagens, aliados ao desenvolvimento da Computação, permitiram produzir um conjunto distinto de ferramentas para a captura automática de dados relacionados a superfície terrestre, para o gerenciamento, análise e apresentação das informações geradas.

Estas ferramentas surgiram em resposta à demanda de armazenamento de dados relacionados a superfície terrestre e realizar análises espaciais em ambientes computacionais. Elas visavam reduzir o custo/tempo de realização de serviços, executando-as com mais qualidade e de forma integrada, evitando o serviço moroso para os especialistas de diversas áreas técnicas do conhecimento, com interesse na informação geográfica.

A ligação técnica e conceitual do desenvolvimento destas ferramentas, resultou no desenvolvimento de uma enorme variedade metodológica de processamento de dados geográficos. Este processamento tem sido denominado na literatura nacional como Geoprocessamento (CÂMARA & MEDEIROS, 1996), e mais recentemente como Geomática (CLEMENT *et al.*, 1997; OSTENSEN, 1995). Estes termos têm sido usados para designar o campo da atividade que, sistematicamente, integra todos os meios usados para adquirir, gerenciar dados relacionados à superfície terrestre, e produzir informação geográficas usando os recursos da informática.

Os primeiros usuários desta tecnologia foram as organizações governamentais e empresas relacionadas ao gerenciamento de dados relacionados à superfície terrestre. A partir de meados da última década, o Geoprocessamento tem se popularizado. Isto decorre da redução do custo das plataformas com alto poder de computação e dos aplicativos de Geoprocessamento, além da amigabilidade destes últimos que permitem aos usuários, sem habilidades específicas, rapidamente aprenderem e operarem sistemas informatizados. Nos últimos anos, cabe ressaltar ainda, que esta tecnologia tem ganho

novos impulsos com os avanços proporcionados pela aquisição de imagens de satélites de alta resolução, pela utilização de Sistemas de Posicionamento Global (GPS) e cadernetas eletrônicas (GOODCHILD^{html}, 1996).

Desta forma, este capítulo apresenta o Sistema de Informações Geográficas (SIG) e suas áreas de aplicação, caracteriza os dados geográficos, descreve os conceitos cartográficos empregados no desenvolvimento de uma base de dados para SIG, caracteriza estas bases de dados, discute a necessidade de adoção de um modelo de dados geográficos, descreve o processo de modelagem destes dados, bem como sua aquisição e os seus critérios de qualidade.

II.2. Sistema de Informações Geográficas (SIG)

O desenvolvimento do Geoprocessamento para atender as diversas necessidades dos usuários gerou uma variedade de ferramentas. Dentre elas, cabe destacar o Sistema de Informações Geográficas (SIG) que oferece capacidade de capturar os dados de diversas fontes, armazenar, manipular, analisar e apresentar estas informações em um formato passível de ser compreendido pelo usuário (SOUZA *et al.*, 1993).

Um SIG combina um Sistema Gerenciador de Base de Dados (SGBD) com técnicas de análise e manipulação de dados espaciais. O SGBD proporciona ao SIG facilidades para o compartilhamento de dados entre múltiplos usuários com garantia de integridade, segurança e consistência no acesso concorrente (RAMIREZ, 1994).

O SIG indexa os dados pela localização como forma fundamental de organizá-los e manipulá-los, estabelecendo a associação entre o dado e o local onde ocorre. Ele tem como característica explorar a capacidade dos dados para gerar informações através de procedimentos de análise espaciais, interativas, que fornecem soluções rápidas e precisas para problemas relacionados às distribuições espacial e temporal dos dados, ou ainda, planejar atividades que se relacionam com o espaço físico que habitamos.

Este sistema gerencia uma Base de Dados, definida por MELO (1988) como não convencional. Esta base manipula um grande volume de dados estruturados sobre uma região, no espaço geográfico, e é caracterizada por necessitar de um modelo de dados mais expressivo que capture a semântica espacial das aplicações. Este modelo trata os dados como objetos complexos, cuja estrutura é definida e manipulada logicamente através de operações genéricas, definidas para tipos de dados suportados pelos sistema.

II.3. Áreas de aplicação de SIG

O SIG tem se tornado uma ferramenta de suporte às decisões socioeconômicas que explora a capacidade dos dados para gerar informações que subsidiam a gestão do espaço geográfico, tanto no sentido de preservação do meio ambiente, quanto na racionalização dos recursos naturais. Isto tem consolidado sua aplicação em um grande número de disciplinas, que tem em comum o interesse por objetos do mundo real, com respeito à sua localização e distribuição espacial, ou ainda de seus atributos.

Desta forma, um SIG é um sistema proprietário, que armazena dados voltados para um domínio e finalidades específicas, para uma dada área de aplicação, onde cada grupo de usuários tem absoluta autoridade e controle sobre eles. Cada base de dados de um SIG é desenvolvida de acordo com as necessidades específicas dos usuários, de uma determinada comunidade de informação, de processar os dados geográficos para uma área de aplicação.

Entende-se por comunidade de informação (CI), uma coleção de pessoas que compartilham uma compreensão comum do significado dos dados geográficos. Isto implica numa visão do mundo comum da realidade geográfica, bem como abstrações comuns, representações de feições e metadados. Uma comunidade de informação geográfica pode ser grupos de usuários tanto provedores quanto consumidores de dados geográficos. De modo geral, cada comunidade de informação é caracterizada por ser multidisciplinar, envolvendo diversos tipos de usuários e áreas de aplicação (BUEHLER & MCKEE^{html}, 1996).

Não é o objetivo deste trabalho catalogar as áreas de aplicações, apenas adotar uma taxionomia de aplicações de contexto para as discussões subseqüentes a respeito dos dados geográficos de forma a demonstrar a viabilidade da metodologia proposta. Desta forma, a taxionomia proposta por RAMIREZ (1994), reiterada por LISBOA & IOCHPE, (1996) e RIBEIRO (1996), será utilizada no presente trabalho. Esta taxionomia caracteriza as atuações do homem sobre o meio físico em cinco grandes áreas de aplicação, conforme a Tabela II.1:

Tabela II.1 - Taxionomia de área de aplicações.

Áreas de aplicação	Atuações
Ocupação humana	Área urbana e regional Educação, Ação Social e Análise socioeconômica Saúde

	Transporte
	Segurança
	Turismo, Cultura, Lazer e Desporto
Atividades Econômicas	Marketing
	Indústrias
Uso da Terra	Agroindústria
	Irrigação e drenagem
	Cadastro urbano e rural
Uso de recursos naturais	Extrativismo vegetal
	Extrativismo Mineral
	Energia (gasodutos, oleodutos, poços de petróleo)
	Recursos hídricos
	Oceânica
Meio ambiente	Ecologia
	Climatologia
	Hidrologia
	Gerenciamento florestal
	Poluição
Geociências	Geodésia e cartografia
	Geologia
	Geografia

RAMIREZ (1994) e RIBEIRO (1996) exemplificam para cada área de aplicação as atividades que podem ser desenvolvidas sobre o SIG. Entretanto, a metodologia proposta no presente trabalho necessita de uma classificação genérica das finalidades do desenvolvimento de uma base de dados geográficos, que pode ser adequada a quaisquer área de aplicação. Contudo, vale ressaltar que esta classificação é aberta, uma vez que não tem o propósito de esgotar todas as finalidades do desenvolvimento de uma base de dados, apenas apoiar a metodologia. Assim, são apresentados a seguir estas finalidades, a saber:

- gerenciamento - administrar, gerir os dados de uma determinada área de aplicação;
- estudo - adquirir algum conhecimentos sobre a área de aplicação;
- análise - examinar os fenômenos de uma dada área de aplicação, tendo em vista conhecer sua natureza, suas proporções, suas funções e suas relações;
- planejamento - desenvolver planos ou programas com base técnica, segundo um roteiro e métodos determinados;
- projeto - desenvolver empreendimentos, com base técnica, segundo um roteiro e métodos determinados;

- monitorizar - acompanhar e avaliar dados de uma área de aplicação, de modo a controlar sua existência;
- zoneamento - visam a divisão de um espaço geográfico em zonas reservadas à finalidades específicas de uma área de aplicação;
- cadastro - registrar a ocorrência de um dado fenômeno para uma determinada área de aplicação; e
- mapeamento - representar a existência de fenômenos relacionados a uma área de aplicação.

II.4. Dados geográficos

As bases de dados de SIG armazenam dados provenientes de estudos relacionados a um espaço geográfico. Estes dados têm sido denominados, na literatura, de dados geográficos, de geodados (BUEHLER & MCKEE^{html}, 1996; GARDELS, 1997) ou de dados geoespaciais (CLEMENT *et al.*, 1997). Estes dados podem ser compreendidos como sendo um conjunto de abstrações que descrevem elementos geográficos tais como, fenômenos, objetos, fatos físicos ou sociais do mundo real, sejam eles discretos ou contínuos, e suas relações com o meio. Estas abstrações possuem uma existência independente e privilegiam, para fins de análise, propriedades que descrevem sua localização e extensão sobre uma representação gráfica da superfície terrestre. Esta existência pode ser física como por exemplo, um poste, ou uma existência conceitual, a declividade. Estes dados são selecionados e descritos, de acordo com área de aplicação e com o interesse pretendido para aplicação, ressaltando alguns aspectos e omitindo outros (AALDERS, 1996).

Os dados geográficos são descritos em um domínio espacial, uma vez que possuem uma relação direta com a localização de um ponto ou porção da superfície terrestre. Eles são caracterizados por apresentarem dois componentes (HERMELY, 1992; SOUZA *et al.*, 1993, FRIESEN & SONDEHEIM, 1994; AGUIAR, 1995; AALDERS, 1996). O primeiro componente representa as propriedades gráficas dos dados geográficos, denominadas de atributos gráficos. Estes atributos descrevem a localização, a extensão e relacionamento espaciais com outros objetos geográficos sobre uma representação, a qual irá constituir o mapa digital. A localização é descrita pelas coordenadas que caracterizam a posição de um objeto, em relação a um sistema de referência qualquer. A extensão define uma forma, que pode ser tratada

matematicamente como ponto, linha ou polígono, e a qual é associada uma semiologia gráfica apropriada. Os relacionamentos espaciais definem os relacionamentos não métricos entre os objetos geográficos, tais como conectividade, adjacência e proximidade entre os objetos da superfície terrestre.

O segundo componente, denominado de atributos não gráficos, privilegia as características temáticas dos fenômenos do mundo real de forma a permitir sua identificação, classificação e o estabelecimento de relações relevantes. Estes últimos são caracterizados como dados alfanuméricos, normalmente associados à tecnologia de Banco de Dados (SOUZA *et al.*, 1993). Eles são armazenados como dados convencionais que descrevem propriedades qualitativas e quantitativas dos dados geográficos, independente de sua localização (BERNHARDSEN, 1992). Os dados qualitativos expressam valores que distinguem um objeto geográfico do outro. Eles podem representar uma classificação alfanumérica ou numérica, por exemplo, o tipo de solo ou o tipo de população. Os dados quantitativos expressam valores ou medidas em uma determinada unidade e podem ser:

- ordinal - atributos que expressam alguma ordenação entre os valores que descrevem a propriedade de um objeto, por exemplo, forte, médio, fraco;
- nominal ou razão - atributos que expressam uma grandeza em uma determinada unidade, por exemplo: idade, peso, renda per capita; e
- intervalo - expressam um intervalo para um dado domínio.

Os dados geográficos podem ainda ter atributos temporais e atributos descritivos. Os atributos temporais são decorrentes da característica temporal dos dados geográficos, uma vez que os fenômenos do mundo real podem variar sobre o tempo. Desta forma, é necessário também um sistema de referência temporal para definir uma época ou período de tempo. Os atributos descritivos, também denominados de pictóricos, contém uma descrição visual do objeto que auxilia a análise geográfica, como por exemplo, uma foto aérea, um desenho CAD ou uma imagem de satélite que contém uma cidade.

II.5. Conceitos cartográficos implícitos na construção de uma base de dados geográficos

Na construção de uma base de dados geográficos para SIG estão embutidos alguns conceitos da ciência cartográfica para representar os atributos gráficos. Estes

conceitos estão relacionados à forma utilizada para representar a superfície terrestre, a saber: *Data* geodésicos, sistema de coordenadas, sistema de projeção, escala e semiologia gráfica, utilizada para representar os objetos geográficos (ARONOFF, 1989; BERNHARDSEN, 1992; LAURINI & THOMPSON, 1992).

Exceto a semiologia gráfica, os demais conceitos são considerados na aquisição dos dados gráficos e são implicitamente usados durante os processos de análise espaciais, com o objetivo de garantir um correto georreferenciamento entre os componentes gráficos. A semiologia gráfica tem o objetivo de fornecer uma simbologia gráfica para a visualização dos dados geográficos que exprima, de forma adequada, a escala, o correto significado dos objetos geográficos.

Algumas vezes, uma mesma base de dados geográficos pode apresentar diferentes níveis de abstração cartográfica, de acordo com a finalidade da aplicação. Estes diferentes níveis de abstração cartográfica levam os atributos gráficos dos objetos geográficos a utilizarem diferentes parâmetros cartográficos para sua representação. Por exemplo, o Sistema Objetivo de Previsão e Acompanhamento de Safras Agrícolas, desenvolvido pela Fundação Instituto Brasileiro de Geografia e Estatística (IBGE), que adota as escala de 1:10.000 e 1:50.000. A escala de 1:10.000 é adotada no armazenamento dos dados provenientes das pesquisas agrícolas realizadas nos levantamentos de campo, ao passo que a escala de 1:50.000 é adotada no armazenamento dos dados geográficos, resultantes de métodos estatísticos baseados em amostragem probabilística aplicados aos dados levantados (MORELLI *et al.*, 1994).

II.5.1. *Data* geodésicos

Para a representar a superfície terrestre em um plano é necessário arbitrar uma forma para a Terra. Devido a complexidade de representação matemática desta superfície, tem sido utilizado como superfície de referência nos cálculos planimétricos para elaboração de uma representação cartográfica, o elipsóide de revolução. Figura gerada pela revolução de uma elipse em torno de seu menor eixo e, obtendo-se assim um elipsóide achatado nos pólos. Inúmeros elipsóides têm sido utilizados pelos geodestas dos diferentes países e/ou continentes, em posições diferentes de forma a adaptar uma região da superfície terrestre a ser representada. O elipsóide é a superfície de referência usada para a redução e tratamento das observações realizadas na superfície física da Terra.

A posição rígida do elipsóide de revolução em relação a superfície física da Terra, e conseguinte em relação ao geóide, bem como sua forma e tamanho constituem um conjunto de parâmetros que são usualmente denominados de *datum* horizontal. Este *datum* define a origem para as medidas planimétricas. Para definir um *datum* a ser adotado por um país ou continente é necessário uma boa adaptação entre o elipsóide e o geóide ao longo da área.

Assim, os países ou continentes estabelecem seus *data* e a partir destes estabelecem uma rede, denominada de Rede Geodésica Fundamental. Esta rede, materializada no terreno por um conjunto de pontos, estabelece um sistema plano-altimétrico único de pontos de controle, cujas coordenadas, definidas através de métodos geodésicos de alta precisão, servem de base para o desenvolvimento de trabalhos de natureza cartográfica. No Brasil é atribuição do IBGE implementar e manter esta rede de pontos.

Atualmente, o Sistema Geodésico Brasileiro é formado pelos *data* horizontal e vertical. O *datum* horizontal, denominado SAD-69 (*South American Datum of 1969*), com vértice em Chuá, adota o Elipsóide de Referência Internacional de 1967 (SGR-67), cujos parâmetros são:

- semi-eixo maior: $a = 6.378.160$ m
- achatamento: $a = 1:298,25$

Este *datum* passou a ser adotado oficialmente pelo Brasil em 1978 pelo IBGE. Anteriormente, era adotado o elipsóide de Hayford com o *datum* Córrego Alegre, sendo ainda encontradas cartas neste sistema. Cabe ressaltar, que uma região abrangida por dois *data* distintos, terá para um mesmo ponto, dois pares de coordenadas diferentes, cada um referido a um *datum*.

O segundo é o *datum* vertical, relacionado ao referencial altimétrico. Ele é estabelecido pela superfície equipotencial que contém o nível médio dos mares, definido por observações maregráficas tomadas na baía de Imbituba, no litoral de Santa Catarina.

O mapeamento sistemático de um país elabora representações cartográficas contendo informações sobre a planimetria, hidrografia, altimetria e vegetação, com grau de detalhamento adequado a atender fins estratégicos e anteprojetos. Os órgãos nacionais responsáveis pelo mapeamento sistemático, no Brasil, são o IBGE e a Diretoria do Serviço Geográfico do Exército. Entretanto, algumas empresas de aerolevante também armazenam mapeamentos sistemáticos produzidos para finalidades específicas. Os projetos de bases de dados geográficos, em geral, partem das

cartas produzidas por este mapeamento sistemático que se encontram na sua maioria representadas nos *data* SAD-69 e Imbituba.

II.5.2. Sistemas de coordenadas

Para expressar a posição de pontos sobre uma superfície é necessário uma superfície de referência e um sistema de coordenadas. A superfície de referência pode ser o elipsóide, a adotado pela geodésia, o plano, adotado na topografia, ou um sistema esférico, adotado na navegação.

Para a superfície de referência esférica ou elipsoidal os sistemas de coordenadas são curvilíneos e denominados, respectivamente, de Sistema de Coordenadas Geográficas e Sistema de Coordenadas Geodésicas (Figura II.1). O Sistema de Coordenadas Geográficas é um sistema de coordenadas esféricas geocêntricas, cujas linhas básicas são os paralelos e meridianos. O Sistema de Coordenadas Geodésico é um sistema de coordenadas dextrógiro, onde as coordenadas (X,Y,H) são referidas a três eixos cartesianos. O eixo X é definido pela interseção entre o Plano Meridiano de Greenwich e o plano equatorial, e o eixo dos Z coincide com o menor eixo do elipsóide. A terceira coordenada H é dada pela altitude geométrica, definida sobre a normal ao elipsóide como a distância entre o ponto na superfície terrestre e o elipsóide.

Para a superfície de referência plana é adotado um sistema de coordenadas cartesianas denominado sistema topográfico. Este é um sistema de coordenadas topocêntrico, cuja origem é estabelecida pela tangência de um plano à superfície a ser representada. As coordenadas deste sistema, denominadas de (x,y,h) são referidas a um sistema tridimensional, cujos eixos x e y, perpendiculares entre si, representam a abcissa e ordenada no plano, respectivamente. A coordenada h representa a altitude do ponto.

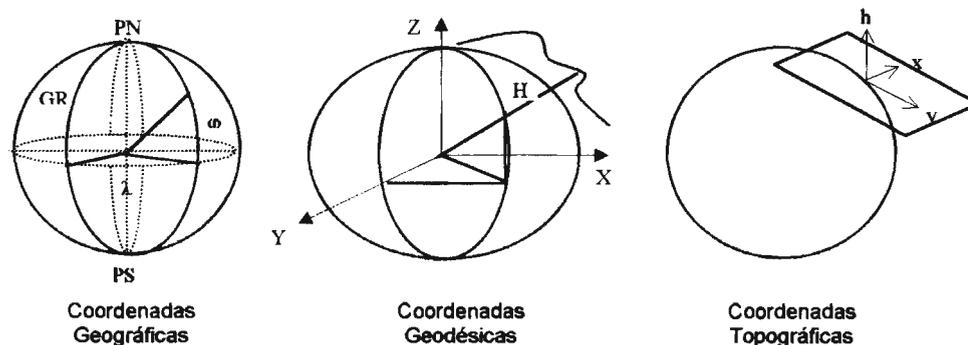


Figura II.1 -Sistemas de coordenadas.

De modo geral, a escolha de um sistema de coordenadas para uma base de dados está relacionada a extensão geográfica e à finalidade da aplicação. As bases de dados adotam o Sistema de Coordenadas Geográficas quando armazenam dados relacionados à grandes extensões geográficas, cujas aplicações manipulam dados genéricos. O Sistema de Coordenadas UTM é adotado quando os dados armazenados são relacionados à médias extensões geográficas. O Sistema de Coordenadas topográfico é adotado quando os dados armazenados são relacionados à pequenas extensões e quando as aplicações requerem apenas um sistema de coordenadas local, não relacionado a outro.

II.5.3. Sistema de projeção

Além da superfície de referência e do sistema de coordenadas se faz necessário adotar um método, segundo o qual cada ponto na superfície terrestre corresponde a um ponto na representação cartográfica e vice-versa. Diversos métodos podem ser empregados para se obter esta correspondência, entre pontos na superfície terrestre e pontos na representação plana. O estudo destes métodos é denominado sistemas de projeção.

Entretanto, ao se tentar representar a superfície terrestre sobre um plano ou sobre uma superfície desenvolvível, ocorrem deformações em área, distância ou ângulos, decorrentes da curvatura terrestre. O estudo dos sistemas de projeção oferece um número ilimitado de possibilidades de se representar a superfície terrestre sobre um plano, de forma a eliminar algumas destas deformações.

Assim, BAKKER (1965) classifica os sistemas de projeção, de acordo com o método de construção adotado, em geométrico, analíticos e convencionais. Os geométricos se baseiam em princípios geométricos e são sub-classificados em perspectivos e pseudo perspectivos. Os perspectivos são novamente classificados em gnômica, esterográfica e ortográfica, de acordo com o ponto de vista. Os analíticos utilizam leis matemáticas para preservar determinada propriedade, e os convencionais baseiam-se em princípios puramente convencionais, em função dos quais se estabelecem expressões matemáticas para definir a correspondência entre pontos na superfície terrestre e pontos sobre a representação cartográfica. Os sistemas de projeção são classificados ainda, segundo a superfície de projeção adotada, em plana ou desenvolvível. Esta última pode adotar como superfície desenvolvível o cilindro, o cone

ou o poliedro. BAKKER (1965) ainda classifica as projeções, de acordo com a propriedade que elas preservam, em:

- equidistantes - são as que não apresentam deformações lineares, isto é, os comprimentos são representados em escalas uniformes;
- equivalentes - são as que não deformam as áreas, isto é, as áreas na carta guardam uma relação constante com as suas correspondentes na superfície terrestre;
- conformes - são as que não deformam os ângulos e decorrente dessa propriedade, não deformam também a forma das áreas; e
- afiláticas - são aquelas em que os comprimentos, as áreas e os ângulos não são conservados. Entretanto podem possuir uma ou outra propriedade que justifique sua construção. A projeção gnomônica, por exemplo, apresenta todas as deformações, mas possui a propriedade de representar as ortodrômicas como retas.

De maneira geral, as projeções são mais conhecidas pelos nomes de seus autores do que, propriamente, pelas designações de suas propriedades ou de suas classificações. Entretanto, é estabelecida uma convenção para designar os diferentes tipos de projeções. Esta convenção especifica primeiramente a natureza da superfície de projeção adotada; em seguida, a classificação da projeção, quanto a situação dessa superfície em relação a superfície terrestre; e finalmente, a sua classificação quanto as propriedades geométricas que conservam, e se for geométrica a sua classificação quanto a posição do ponto de vista.

A escolha de um sistema de projeção para uma representação cartográfica requer que suas propriedades satisfaçam as finalidades impostas pela sua utilização. Em geral, ao se construir uma representação é escolhido um sistema de projeção que, sem possuir todas as condições ideais, possua aquelas que satisfaçam determinados objetivos, considerando a região, o fim a que se destina, as deformações que poderão ser admitidas e que propriedades deverão ser preservadas.

No Brasil, as cartas em escalas maiores que 1:250.000 do mapeamento sistemático são elaboradas na projeção UTM (*Universal Transverse de Mercator*) empregando métodos fotogramétricos e compilação cartográfica. Já as cartas em escalas menores que 1:1.000.000, inclusive, são elaboradas na projeção de Lambert ou na projeção policônica.

O sistema de projeção UTM deriva do sistema de projeção conforme transversa de Gauss, o qual adota o cilindro tangente ao elipsóide transversalmente ao seu semi-eixo menor. Entretanto, a projeção UTM tem como artifício considerar como superfície

de projeção, um cilindro secante ao elipsóide, de forma a reduzir as grandes ampliações que se apresentam nas bordas do fuso no sistema de Gauss. O desdobramento deste cilindro resulta num plano conhecido como UTM, que representa os pontos do elipsóide terrestre contidos no respectivo fuso, segundo um sistema de coordenadas plano retangulares. Trata-se de uma representação plana do elipsóide terrestre composto por vários sistemas parciais, com fusos de 6° de amplitude e meridianos centrais múltiplos ímpares de 3° , numerados de 1 a 60, contados a partir do anti-meridiano de Greenwich, com zonas de superposição de fusos de $30'$. Os eixos de cada sistema plano retangular parcial são as transformadas do equador e do meridiano central, sendo estabelecido o cruzamento de ambas como origem, e com coordenadas convencionadas como sendo $E = 500.000$ m no eixo leste-oeste e $N = 10.000.000$ m, para o hemisfério sul, ou $N = 0$ m, para o hemisfério norte.

II.5.4. Escala de uma base de dados geográficos

Os dados geográficos utilizam uma abstração cartográfica para representar os fenômenos da superfície terrestre na base de dados. Esta abstração requer reduções das dimensões dos objetos geográficos para representá-los no plano.

Essa redução traduz a idéia de escala (E), que pode ser definida como a relação entre o comprimento gráfico (d), na representação cartográfica, e o comprimento equivalente medido na superfície terrestre (D). Desta forma, para uma escala de 1:250.000, 1 centímetro no mapa corresponde a 2500 metros na superfície terrestre.

As representações cartográficas são geralmente associadas às expressões escala grande, escala média ou escala pequena. É considerada escala pequena aquelas entre 1:500.000 a 1:5.000.000, escalas média aquelas entre 1:25.000 a 1:250.000 e escalas grande aquelas entre 1:500 a 1:20.000.

De modo geral, as escalas pequenas representam áreas com a finalidade de subsidiar a execução de estudos e análises de aspectos gerais e estratégicos. As escalas médias para aplicações regionais, representam áreas específicas com grande desenvolvimento, centros políticos ou econômicos, áreas desenvolvidas ou áreas sensíveis para investimentos governamentais, visando subsidiar o planejamento setorial em todos os níveis do governo (Federal, Estadual e Municipal). Estas representações fornecem elementos para planejamentos socioeconômicos e bases para anteprojetos de Engenharia, cartas topográficas para a elaboração de estudos e projetos que envolvam

ou modifiquem o meio ambiente. As escalas grandes são empregadas em aplicações locais para projetos de engenharia que requerem um maior nível de detalhamento cartográfico.

O volume de dados armazenados em uma Base de Dados Geográficos será limitado dentre outros, pela escala de representação dos objetos geográficos. As bases de dados geográficas são armazenadas, em geral, na escala de aquisição dos dados ou em uma escala menor. Cabe ressaltar que os mapas poderão ser impresso em escalas menor que a escala original, uma vez que no caso contrário haverá perda de precisão. As representações impressas em uma escala menor requerem ainda de um processo de generalização cartográfica. Este processo no contexto de um SIG, visa modificar a resolução espacial transformando os dados gráficos através da simplificação destes.

II.5.5. Semiologia gráfica

Na modelagem de uma base de dados para SIG está implícito um processo de generalização cartográfica. Este processo especifica uma semiologia gráfica para os dados geográficos, levando em conta as necessidades de visualização e de operações espaciais a serem realizadas sobre os dados geográficos, bem como as limitações gráficas imposta pela escala adotada para a base de dados.

A semiologia gráfica consiste em símbolos gráficos e cores usados para representar os elementos ou as estruturas, ou os fenômenos que ocorrem sobre a superfície terrestre. Este conjunto de símbolos, resultante do processo de generalização cartográfica da modelagem dos dados, tem por finalidade associar intuitivamente os elementos e seus atributos, com o objetivo de facilitar a identificação e localização dos mesmos nos processos de análises.

A semiologia gráficas pode representar três tipos de informações: espaciais, qualitativas e quantitativas. A informação espacial indica a localização do objeto, a qualitativa distingue um objeto do outro e a quantitativa é usada para expressar quantidades.

A semiologia gráfica é selecionada, visando representar as informações armazenadas de forma clara, precisa, inequívoca e compreensível atendendo as finalidades de uma área de aplicação, a forma de trabalho do usuário e a escala em que se encontram. De modo geral a semiologia gráfica pode ser:

- puntiforme - são usados para identificar e localizar pontos de referências que não tem significação quanto ao tamanho em relação a escala da representação cartográfica, pois a rigor, se desenhados desapareceriam. Um símbolo puntiforme pode representar uma casa em escala grande, ou uma cidade inteira em escala pequena. Esses símbolos também são usados para expressar quantidades como por exemplo, um gráfico de pizza ou de barra associado a uma localização;
- linear - mostram o desenvolvimento de rios, estradas, limites, contornos de mapas, lagos, bosques e divisões administrativas. Estes símbolos indicam a localização de acidentes ou linhas imaginárias, cuja principal característica comum é o comprimento. Eles se utilizam de linhas cheias com várias espessuras, linhas interrompidas, linhas pontilhadas e linhas mistas, contendo uma sucessão de traços e pontos intercalados. Essas linhas podem representar características físicas como cursos d'água e estradas, ou características imaginárias como os limites políticos-administrativos;
- areolar - indicam a distribuição realizada em áreas e permitem a constatação qualitativa e quantitativa. Os símbolos areolares qualitativos discriminam os tipos de áreas, ao passo que, os símbolos quantitativos representam quantidades absolutas, valores médios, densidades, capacidades, coeficientes de resolução, dispersão, etc.

O processo de generalização cartográfica para estabelecer a semiologia gráfica adequada, aplica as técnicas de simplificação, classificação e simbolização. A simplificação é aplicada para fenômenos discretos. Nessa técnica são determinadas as características dos dados a serem retidas, realçadas e a serem eliminadas. A classificação é aplicada a fenômenos contínuos. Nessa técnica são definidas classes com objetivo de reduzir variedades, ordenar e simplificar diferenças irrelevantes, inconseqüentes ou aquelas de magnitude intratável. A simbolização trata da codificação gráfica resultante das técnicas anteriores. Essa técnica consiste no estabelecimento de uma relação semântica entre os entes e fenômeno do mundo real e as representações cartográficas armazenadas pela base de dados geográficos.

De modo geral, a semiologia gráfica está associada a forma de trabalho de uma comunidade de informação. Os aplicativos de SIG oferecem facilidades para criação de uma biblioteca de semiologia gráfica adequada a cada comunidade de informação.

II.6. Características das bases de dados geográficos

As bases de dados dos SIGs reúnem um grande volume de dados geográficos, que são estruturados, de forma apropriada à manipulação integrada dos componentes gráficos e não gráficos para a execução de diversos tipos de análises espaciais.

Tendo em vista que estes componentes apresentam características distintas das bases de dados convencionais, os SIGs têm sido desenvolvidos sobre três tipos de arquiteturas para gerenciar os dados geográficos, a saber (BENHARDSSEN, 1992):

- um sistema que gerencia arquivos com os componentes gráficos e não gráficos, sobre o qual há uma camada com ferramentas para realizar análises espaciais;
- um sistema no qual os componentes gráficos são armazenados em um sistema de arquivos e os componentes não gráficos em um SGBD comercial. Sobre o sistema de arquivos há uma camada de aplicação que gerencia e recupera os dados gráficos para realizar as análises espaciais. Esta arquitetura se utiliza de um modelo de dados, denominado de georrelacional, para associar os componentes gráficos e não gráficos;
- e
- um sistema que gerencia tanto os componentes gráficos quanto os não gráficos. Em geral é utilizado um SGBD que tem suas características estendidas para suportar ambos tipos de dados. Sobre este SGBD há uma camada de aplicação que fornece ferramentas para gerenciar e realizar as análises espaciais.

Estas arquiteturas oferecem modelos de dados adequados a modelar a realidade e sobre o qual gerenciam os dados geográficos. Estes modelos de dados permitem estratificar a informação geográfica em um conjunto de níveis denominados de planos de informação (PIs), *layer* ou *coverage*. Cada nível representa uma mesma área, mas cada qual corresponde a uma perspectiva diferente da realidade geográfica da área. Esta perspectiva corresponde a um aspecto temático independentemente com um conjunto de propriedades que a descrevem, por exemplo, a hidrografia, a vegetação. Cada nível contém elementos geométricos que variam em número, formas e propriedades (Figura II.2). Estes níveis podem ser distinguidos de acordo com a finalidade deles na base de dados em:

- níveis de informação básica - é um conjunto mínimo de feições físicas e culturais que fornecem informações necessárias ao reconhecimento topográfico da região. Eles têm por objetivo fornecer a base cartográfica para o espaço geográfico, de forma a proporcionar uma melhor compreensão da região, dentro dos limites impostos pela

escala, por exemplo, limite político administrativo, hidrografia, rede viária, cidades, etc..

- níveis de informação temática ou especial - fornecem a representação policromática do espaço geográfico de interesse da área de aplicação da base de dados. Estes níveis armazenam as descrições dos temas necessários para o SIG atingir o objetivo proposto. Eles descrevem as ocorrências, distribuições e relações espaciais de entes e feições do contínuo espacial para cada tema.

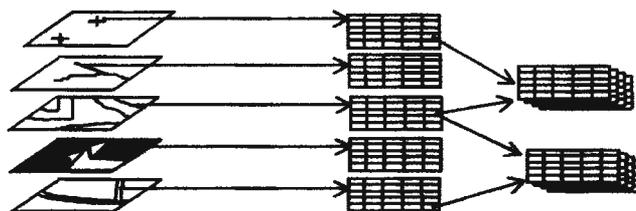


Figura II.2 - Organização da base de dados em níveis de informação.

Esta organização em níveis permite ao usuário combinar conceitualizações simples da modelagem de dados para derivar níveis de informação para produzir conjuntos de características que representam relações complexas do mundo real. Por exemplo, sejam os níveis de informação simples: uso da terra, solos, geologia, topografia e divisão municipal, ilustrados na Figura II.3. Através de operações aplicadas ao nível de informação topografia é gerado o nível de informação de declividade e através da combinação dos demais níveis de informação, com a aplicação de operações espaciais, é derivado o nível de informação de erodibilidade. A partir de análises espaciais sobre estes níveis gerados é criado o nível de informação que retrata as áreas com potencial de erosão. Este último nível pode ser combinado com o nível de informação, contendo a divisão municipal, de forma a proporcionar a visualização dos distritos com áreas mais propícias à erosão.

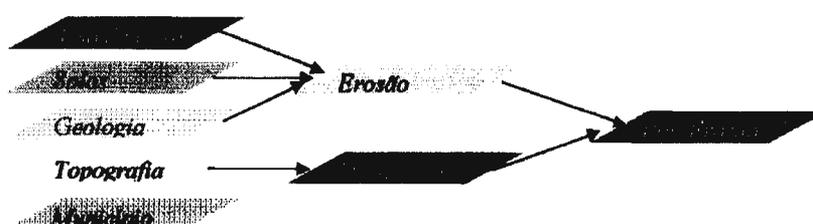


Figura II.3 - Exemplo de análise do potencial de erosão.

Nestas arquiteturas a associação entre os atributos gráficos e não gráficos é estabelecida por um identificador comum, denominado de geocódigo, que vincula a

representação gráfica em uma dada localização aos atributos não gráficos armazenados em sistemas de arquivos, tabelas internas e tabelas externas ao SIG.

A execução de uma consulta nestas arquiteturas é dividida em sub-consultas aos componentes gráficos e não gráficos. Devido a isto, alguns autores (SOUZA *et al.*, 1993; KUMAR *et al.*, 1997) caracterizam estas arquiteturas como sistemas híbridos, cujo gerenciamento e processo de consulta são complexos.

II.7. Considerações sobre os modelos de dados geográficos

Segundo TSICHRITZIS & LOCHOVSKY (1982), um modelo de dados é uma abstração usada para descrever a realidade formalmente, através de um conjunto de conceitos rigorosamente estabelecidos, de modo que os dados são estruturados e possam ser compreendidos conceitualmente por todos, da mesma forma. Estes conceitos fornecem subsídios para especificar um conjunto de estruturas de dados e um conjunto de restrições e operadores sobre estas estruturas. As estruturas de dados capturam as propriedades dos dados através de atributos, tipos de entidades e tipos de relacionamentos entre as instâncias. As restrições capturam propriedades comuns invariantes para todas as instâncias do modelo. As operações definem a natureza dos operadores sobre as estruturas dos dados e o modo como elas serão usadas.

Os modelos de dados usados em SIG são fortemente tipados, uma vez que são baseados em estruturas que fornecem tipos para representar os dados de acordo com a sua geometria, e operações específicas que podem ser aplicadas sobre estes tipos. Por exemplo, o ARC/Info (ARC, 1991) e o SPRING (CÂMARA & MEDEIROS, 1996) apresentados em STRAUCH, SOUZA & MATTOSO (1998d), oferecem, respectivamente, estruturas de dados e um modelo orientado a objeto com tipos de dados que fornecem as estruturas e operações apropriadas a cada estrutura.

Desta forma, os SIGs utilizam um modelo de dados, denominado de modelo de dados geográficos, que oferece estruturas e operações associadas a uma linguagem coerente para a descrição e manipulação dos dados geográficos. Este modelo utiliza elementos para abstrair ou representar as entidades/fenômenos que têm uma localização e extensão espacial (AGUIAR, 1995; BORGES & FONSECA, 1996; LISBOA & IOCHPE; 1996).

II.8. Estrutura de dados geográficos e operações

Os SIGs requerem estruturas de dados que fornecem uma representação matemática para o espaço geográfico, uma linguagem para a definição das estruturas suportadas e uma linguagem de manipulação que oferece funções para recuperar os dados geográficos e realizar análises espaciais. Deste modo, os SIGs utilizam dois tipos de estruturas que modelam o espaço geográfico, em um plano bidimensional, segundo duas estruturas (WORKBOYS & DEEN, 1991).

O primeiro tipo de estrutura, denominado de matricial ou *raster*, modela o espaço geográfico como uma coleção de camadas e referenciais espaciais. Cada camada representa uma superfície contínua sobre a qual a variação de um determinado fenômeno sobre a região é apresentada de acordo com uma perspectiva temática.

Esta estrutura trata o componente gráfico como uma superfície plana cartesiana, que subdivide o espaço geográfico regularmente em um conjunto de células. Este conjunto constitui uma matriz retangular, composta de m linhas e n colunas, que associa um sistema de coordenadas à superfície terrestre onde a região que cada célula compreende é associada a um valor/código correspondente a uma classe referente ao fenômeno representado (ARONOFF, 1989; WORKBOYS & DEEN, 1991; LAURINI, 1992; BERNHARDSEN, 1992).

Cabe ressaltar que a unidade básica desta matriz é a célula e indica a resolução de cada nível. Esta resolução é dada pela relação entre o tamanho da célula e a área por ela representada no terreno. Quanto menor a área que cada célula representar, maior o nível de detalhamento obtido para o conjunto de dados.

Neste tipo de representação um ponto é representado por uma célula. Uma linha por um conjunto de células arranjado numa determinada direção, e uma área por um conjunto de células que representam uma geometria. Por exemplo, uma estrada não existe como uma entidade distinta e sim como um grupo de células com a codificação de estrada. As células que representam a estrada formam a entidade.

Esta estrutura utiliza uma álgebra que é especificada sobre os níveis de informação, implementando operações lógicas, aritméticas, geométricas e estatísticas (BERNHARDSEN, 1992). Estas operações, segundo WORKBOYS & DEEN (1991), podem ser:

- locais - são aquelas cujos argumentos da operação são pontos;
- focais - são aquelas cujos argumentos da operação são um ponto e sua vizinhança; e

- zonais - são aquelas cujos argumentos da operação são um ponto e uma região que o contém.

O segundo tipo de estrutura, denominado de vetorial, modela o espaço geográfico como uma coleção de objetos espaciais genéricos, distintos e identificáveis, que corresponde à representação gráfica das entidades do mundo real e sua respectiva localização espacial, definidas com referência a um sistema de coordenadas relacionadas à superfície terrestre. Esta estrutura é empregada quando há interesse na informação sobre a distribuição de objetos no espaço ou as condições que se aplicam a uma área.

A unidade básica desta estrutura é o vetor representado por seqüência ordenada de pares de coordenadas (X,Y) que permite representar os objetos geográficos como (ARONOFF, 1989; WORKBOYS & DEEN, 1991; LAURINI, 1992; BERNHARDESEN, 1992):

- ponto - destina-se a representar objetos pontuais em uma determinada escala. É definido por um par de coordenadas (X,Y). Por exemplo, a localização de objetos que de acordo com a escala possui uma representação discreta, tais como indústrias, escolas, poços, torres, postes;
- linha - é um conjunto de segmentos contínuos, cujas extremidades são pontos. Seus segmentos são definidos por uma lista de pontos associadas a uma função. Representam objetos que possuem uma distribuição linear, tais como rodovias; hidrovias ou linhas de transmissão; e
- área - é um conjunto de uma ou mais linhas que representam a geometria de objetos planares. Podem estar associadas a objetos tais como propriedades; reservatórios de água e polígonos de tipos de solo. Este último constitui um exemplo de unidades espaciais de áreas, as quais são associadas a valores agregados correspondentes a distribuição espacial de valores de atributos dos objetos geográficos.

Esta estrutura pode ser espaguete ou topológica. A estrutura espaguete armazena o componente posicional como uma representação fiel do mapa e associa seus respectivos atributos convencionais. Esta estrutura não reconhece relacionamentos espaciais entre os dados gerando redundância e ineficiência. Ela não é adequada à consultas recursivas para obtenção de informações sobre os relacionamentos espaciais. Portanto, não é muito utilizada pelos SIGs.

A estrutura topológica surgiu como solução à ineficiência na realização de análises espaciais complexas que envolvessem o relacionamento entre os objetos. Esta

estrutura armazena, além do componente posicional, os atributos dos dados e os relacionamentos topológicos entre os objetos. Os relacionamentos topológicos referem-se às relações espaciais, essencialmente não métricos. Estes relacionamentos permitem que os objetos espaciais sejam armazenados de modo claro e não ambíguo, definidos pelas suas localizações espaciais com referência a um sistema de coordenadas. Estes tipos de relacionamentos são intuitivos para o cérebro humano quando visualiza o mapa, mas ao serem armazenados em um computador devem ser criados subsídios, de modo a permitir determiná-los.

Deste modo, na estrutura topológica as relações espaciais são representadas de forma explícita, através de tabelas que armazenam as relações topológicas entre objetos genéricos definidos como um conjunto de nós, arcos e polígonos. O elemento básico desta estrutura é o arco. Um arco é composto por uma série de pontos que começam e terminam em um nó. Um nó é o ponto de interseção onde dois ou mais arcos se interceptam podendo ocorrer também no fim de um arco pendente. Nós isolados, não conectados aos arcos, representam pontos. Um polígono é constituído por uma cadeia fechada de arco que representam seus limites. Esta estrutura trata os objetos geográficos como uma série de pares de coordenadas (x,y) e associa, a cada geometria formada, às linhas de uma tabela que armazenam os relacionamentos topológicos que explicitam:

- conectividade - áreas que se conectam às outras através de nós, também denominado de topologia arco-nó;
- definição de áreas - arcos que conectados definem uma área fechada, também denominada de topologia de arco-polígono; e
- contigüidade - arcos que possuem direções fornecendo a noção de direito e esquerdo, também denominado de topologia direito-esquerdo.

Esta estrutura permite realizar operações no espaço Euclidiano, contemplando operações espaciais que tratam os objetos geográficos como conjunto, por exemplo, operações de união, interseção, diferença de regiões; operações que envolvem cálculos aritméticos entre objetos, por exemplo, o cálculo do azimute e distância entre dois pontos, e operações que extraem o relacionamento espacial entre os objetos, por exemplo, a proximidade, a contigüidade entre objetos.

II.9. Modelagem dos dados geográficos

A tecnologia de Banco de Dados tem como principal característica a habilidade de armazenar representações sobre uma realidade a ser recuperada e interpretada. Estas representações armazenam uma única conceitualização que compreende todas as demais, de forma a evitar redundâncias e inconsistências (TSICHRITZIS & LOCHOVSKY, 1982).

A realidade geográfica, na verdade, apresenta uma infinidade de fenômenos, objetos, ou fatos geográficos, que não se apresentam de forma ordenada, podendo cada comunidade de informação perceber e conceber a realidade geográfica de diversas formas (Figura II.4). Entretanto, para que esta realidade possa realmente fornecer subsídios às análises, os seus elementos (fenômenos, objetos, ou fatos geográficos) devem ser abstraídos, capturados e estruturados segundo uma forma sistemática. Este processo, especificado durante o projeto da base de dados, é denominado de modelagem de dados (DBOUK *et al.* 1996).

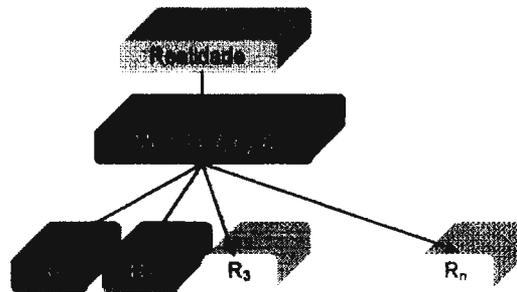


Figura II.4 -Modelagem dos dados.

Este processo de modelagem dos dados objetiva organizar e estruturar os dados de uma base de modo as suas aplicações alcançarem resultados esperados. Ele está condicionado às necessidades latentes e manifestas de uma comunidade de informação, de acordo com a finalidade, a área de aplicação e os recursos disponíveis para capturá-la. Cabe ressaltar ainda que este processo de modelagem reflete restrições organizacionais, econômicas, financeiras e técnicas da comunidade de informação, uma vez que é influenciada pelos procedimentos necessários a implantação da base de dados. Estas restrições envolvem estudos de viabilidade, requisitos lógicos da instituição, análise de processos envolvidos, a proposição conceitual dos dados e os recursos tecnológicos disponíveis.

No processo de modelagem para o desenvolvimento de bases de dados geográficos, esta percepção da realidade requer um conhecimento mais próximo dela,

na medida que as entidades do mundo real são percebidas com uma localização e uma representação gráfica em um domínio espaço-temporal (BORGES & FONSECA, 1996). Este processo de modelagem utiliza um modelo de dados que permite, ao analista, entender a gênese das necessidades de informação e representar todos os aspectos envolvidos na modelagem da área de aplicação (CARR, 1994). Esses modelos de dados fornecem ainda uma notação gráfica formal e uma terminologia usada para modelar a realidade.

De modo geral, o processo de modelagem para bases de dados de SIG tem empregado o modelo relacional e o modelo orientado a objeto. No modelo relacional a percepção da realidade geográfica é expressa sobre três conceitos (ELSMARI & NAVATHE, 1994):

- entidade - representação abstrata de um objeto, fato ou fenômeno do mundo real que possui uma existência independente, sobre a qual se deseja armazenar e recuperar informações;
- relacionamento - representa uma associação entre duas ou mais entidades. Esta associação, utiliza o conceito de cardinalidade para expressar quantas entidades de um tipo são associadas a outras. Esta cardinalidade pode ser de 1:1; 1:n e n:m; e
- atributos - são propriedades que descrevem uma entidade ou um relacionamento.

Este modelo utiliza uma notação gráfica simples que tem sido largamente usada em geoprocessamento, uma vez que os SIGs implementam um modelo georrelacional. Este modelo consiste fisicamente em um conjunto de tabelas contendo registros e atributos definidos por tipos primitivos de dados que descrevem as características temáticas da entidade e relacionamentos entre as tabelas.

Mais recentemente, a modelagem de dados orientado a objeto tem sido utilizada para descrever a realidade. Neste contexto, as entidades do mundo real são modeladas como objetos pertencentes a uma classe. Uma classe possui um conjunto de propriedades e operações que definem, respectivamente, a estrutura e o comportamento de seus objetos. As classes são fabricas de instâncias de objeto de um mesmo tipo que tem o mesmo comportamento. Os detalhes da estrutura e do comportamento podem ser encapsulados, isto é escondidos do usuário. As classes se organizam em uma hierarquia de generalização/especialização de classes que utiliza o conceito de herança, no qual as subclasses herdam a estrutura e o comportamento da superclasse. Os objetos de uma classe podem ser agregados para constituir um outro objeto (WORBOYS *et al.*, 1990; MILNE *et al.*, 1993; WORBOYS, 1994; SMAALEN, 1996).

Os conceitos utilizados pelo modelo orientado a objeto se ajustam bem à modelagem do dados geográficos e implementação do banco de dados geográficos (STRAUCH *et al.*, 1997). Nos últimos anos alguns modelos de dados orientados a objetos têm sido propostos para modelar os objetos geográficos, como por exemplo, os modelos propostos por TIMES & SALGADO (1994) e o modelo proposto por LISBOA & IOCHPE (1996).

O processo de modelagem de dados geográficos consiste em capturar os aspectos essenciais da realidade geográfica através da abstração conceitual e cartográfica (SMAALEN, 1996). O processo de abstração conceitual consiste em estabelecer a semântica das entidades geográficas através do estabelecimento de uma denominação, propriedades e seus respectivos valores, e relacionamentos com outros objetos. Estes relacionamentos diferenciam dos relacionamentos convencionais, uma vez que estão também associados a uma referência espaço temporal. CÂMARA & MEDEIROS (1996) distinguem os seguintes tipos de relacionamentos:

- relacionamentos espaciais - são relacionamentos que associam uma entidade geográfica a sua localização espacial. Esta associação é materializada pelo geocódigo;
- relacionamentos topológicos - são relacionamentos não métricos, que expressam como uma entidade geográfica se relaciona com as demais;
- relacionamentos temporais - são relacionamentos que expressam como as entidades relacionam umas com as outras sobre o domínio temporal; e
- relacionamentos temáticos - são relacionamentos que expressam como as entidades formam uma totalidade interrelacionada;

A abstração cartográfica na modelagem envolve um processo de generalização conceitual e generalização geométrica. A generalização conceitual envolve a seleção dos objetos a serem representados, o estabelecimento de uma classificação e da semiologia gráfica adequada. A generalização geométrica, estabelece para esta última, uma forma que leva em consideração a resolução espacial definida pela escala e referência espacial a ser empregada na base de dados. Assim, o processo de a generalização cartográfica estabelece as propriedades que descrevem graficamente a entidade geográfica quanto a sua geometria, apresentação (semiologia gráfica) e relacionamentos espaciais com outros objetos, considerando o domínio espacial da base de dados. Por exemplo, um conjunto de áreas que apresenta a taxionomia para os padrões de florestas e suas propriedades tais como: altura, volume de madeira, tipo, etc,

cuja geometria pode ser definidas por polígonos com uma semiologia gráfica de acordo com o tipo da floresta.

Este processo de modelagem da realidade resulta na definição de uma única conceitualização que compreende todas as demais necessárias à comunidade de informação para uma dada área de aplicação. Esta representação é acordada no modelo da base de dados e expressa através de um esquema, também denominado de intenção da base de dados. Este esquema, em geral, é enriquecido com uma descrição, em meio analógico, contendo o significado da informação geográfica, os aspectos geométricos e topológicos, a forma de aquisição e a qualidade dos dados geográficos.

Assim, o esquema representa uma estrutura lógica que utiliza os conceitos e formalismo de um modelo de implementação dos dados para expressá-los. Este esquema facilita o gerenciamento dos dados descrevendo as estruturas das entidades, os tipos de dados, os relacionamentos, as operações dos usuários e as restrições para aquela comunidade de informação. Ele esconde os detalhes físicos das estruturas armazenadas e é vital para os usuários de uma base de dados realizar operações, bem como programar aplicações.

Na modelagem relacional o esquema descreve as definições de tabelas, enquanto que na modelagem orientado a objeto o esquema descreve as definições de classes ou estruturas que descrevem os dados.

Os valores dos dados são denominados de instâncias ou ocorrências e algumas vezes de extensão da base de dados.

O esquema é especificado durante o projeto da base de dados e mantido durante o seu ciclo vida, no qual é muito difícil haver mudanças significativas, ao passo que a extensão do esquema pode mudar sobre o tempo (ELMASRI & NAVATHE, 1994).

II.10. Considerações sobre a aquisição de dados geográficos

Os dados geográficos são modelados para propostas específicas de área de aplicação. Eles são modelados e representados na base de dados de acordo com o processo de abstração e generalização cartográfica especificados no domínio espaço temporal.

Não existem diretrizes para a coleta destes dados. Em geral, essas diretrizes são específicas para cada área de aplicação. Segundo ARONOFF (1989), a coleta de dados é resultante da observação direta ou indireta do espaço geográfico. As observações

diretas são aquelas provenientes de medições realizadas em trabalho de levantamentos de campo ou da interpretação analógica ou digital de fotografias e imagens de satélite. Por exemplo, a determinação da localização e forma do objeto geográfico, medição de valores de atributos, classificações de imagens. As observações indiretas dos dados são aquelas extraídas de meios existentes, tal como mapas analógicos ou digitais, estudos estatísticos e bibliografia existente, contendo textos, documentos gráficos, registros históricos e cadernetas de campo.

Os dados gráficos são convertidos para o formato digital, através de processos de digitalização manual ou automática. A digitalização manual pode ser realizada sobre uma mesa digitalizadora, com um dispositivo de apontamentos que permite o operador selecionar os pontos que serão convertidos para o formato digital, ou através do processo de restituição digital de um par de imagens estereoscópicas. A digitalização automática emprega um sistema de varredura ótica que transforma os dados para o formato digital. Os dados não gráficos são introduzidos no SIG através de digitação via teclado ou conversão de arquivos existentes.

II.11. Qualidade dos dados geográficos

A qualidade dos dados refere-se ao grau de fidelidade à realidade atual dos mesmos. A qualidade dos dados geográficos é fundamental para conduzir os processos de análises espaciais em uma base de dados geográficos. Ela deve ser controlada ao longo de todo o processo de coleta e armazenamento dos dados, uma vez que quando se encontram incorretos não há informações confiáveis, o que conseqüentemente ocasiona erros gerenciais e desperdícios em todas as dimensões de um processo público ou privado.

Desta forma, ARONOFF (1989), BERNHARDSEN (1992), AALDERS (1996) e COCKCROFT (1997) propõem uma apreciação sobre a qualidade dos dados, destacando os seguintes aspectos que devem ser considerados na coleta dos mesmos:

- Acurácia posicional - estabelece a qualidade dos dados geográficos quanto a sua posição em um sistema de referência. Esta acurácia é estabelecida através dos seguintes conceitos:
 - precisão - refere-se a qualidade da operação pela qual um resultado qualquer é obtido; um par de coordenadas é considerado preciso se atender as tolerâncias preestabelecidas; e

- exatidão - é um conceito estatístico que diz respeito a probabilidade de um dado se aproximar do seu valor real.
- Acurácia temporal - estabelece a qualidade dos dados geográficos em relação ao tempo quanto a exatidão do tempo e atualização dos dados:
 - época - indica a data de referência para os dados coletados. Na construção de uma base de dados é importante conhecer a idade da informação. Isto ajuda ao usuário avaliar a legitimidade da informação e quaisquer mudanças que possam ter ocorrido desde a sua inclusão na base; e
 - atualidade - mede o quão recente são os dados. Este conceito não implica no fato de um dado ter sido coletado há muito tempo estar desatualizado. O tipo de dados e a aplicação a que se destina devem determinar os critérios deste julgamento.
- Acurácia temática - descreve a qualidade dos atributos gráficos e não gráficos que tratam das características temáticas. Este critério vale tanto para os valores quantitativos quanto para os qualitativos. Esta acurácia é fornecida pela:
 - integridade dos dados - em relação os componentes gráficos refere-se à capacidade do atributo gráfico representar de maneira clara e correta o elemento geográfico descrito. A avaliação da integridade objetiva eliminar inconsistências espaciais, tais como linhas que não se encontram. No que se refere aos dados não gráficos, a integridade diz respeito à regra que lhes impõem restrições de valores e de relacionamentos; e
 - consistência - mede se uma mesma informação armazenada em mais de um arquivo tem o mesmo valor, a qualquer momento, em todos os seus locais de ocorrência.

Capítulo III - Interoperabilidade entre múltiplas Bases de dados

III.1. Introdução

Interoperabilidade é um conceito aberto e amplo que tem sido usado em diferentes contextos para expressar o compartilhamento de dados em um ambiente composto de vários sistemas. PERINE (1995) define, genericamente, a interoperabilidade como a habilidade de dois ou mais sistemas, sejam computadores, dispositivos de comunicação, base de dados, rede e outras tecnologias de informação, interagirem uns com os outros e trocarem dados, de acordo com um método prescrito a fim de alcançar resultados previsíveis. Para CHOMICKI & REVESZ (1997), a interoperabilidade não é somente a habilidade de trocar dados, mas também serviços.

A interoperabilidade subentende que diferentes sistemas se comunicam de modo que os usuários não levem em conta as diferenças entre seus produtos ou serviços. Para obtê-la não é necessário estabelecer padrões de dados ou ferramentas de conversão. Ela implica em estabelecer mecanismos para troca de dados e serviços através da interação entre sistemas com protocolos aceitos entre eles.

HEILER (1995) e MANOLA (1995) distinguem interoperabilidade de interoperabilidade semântica. A primeira é definida como a habilidade dos sistemas trocarem serviços e dados uns com os outros, ao passo que a segunda assegura que esta troca de serviços e dados faz sentido e que o sistema requisitante e o sistema requisitado têm uma mútua compreensão dos significados destes. Para alcançar esta interoperabilidade é necessário o entendimento dos sistemas, não somente no nível de sistema ou programas, mas também ao nível de tipos de dados. Um sistema deve saber requisitar o serviço e os demais sistemas compreender o serviço requisitado e gerar uma resposta.

Desta forma, a interoperabilidade entre Sistemas Gerenciadores de Banco de Dados (SGBD) envolve um componente semântico e um componente técnico que pressupõe a compatibilidade entre sistemas em dois níveis específicos de interações. O primeiro através de uma metodologia para integração de esquemas das bases de dados e

o segundo através de uma arquitetura em rede que encapsula os sistemas envolvidos e permite a interconexão física entre múltiplas bases de dados, respectivamente.

A metodologia para integração de esquemas consiste em um conjunto de técnicas e processos formais para detectar objetos similares. Esta metodologia emprega uma estratégia para compreender os esquemas, aplicar critérios de comparação que exploram a semântica e a sintaxe dos esquemas, de forma a detectar os conflitos e inconsistência entre eles e a seguir estabelecer um conjunto de termos compartilhados. Em geral, estes passos são complexos, existindo diferentes metodologias que contemplam vários critérios que podem ser aplicados para detectar as heterogeneidades e a diversidade de tipos de conflitos encontrados.

A arquitetura suporta o acesso à múltiplas bases de dados heterogêneas e distribuídas, projetadas independentemente e operando autonomamente, através de um aplicativo que oferece um ambiente que esconde as heterogeneidades físicas e lógicas entre os sistemas, esquemas de dados ou aplicações. Sobre este ambiente, o usuário submete consultas usando a definição dos dados e/ou uma linguagem de manipulação comum. Este ambiente proporciona a troca de dados e expande a base de aplicações dos usuários, permitindo que dados residindo em um sistema possam ser efetivamente usados por outro (GOH *et al.*, 1994).

Este capítulo inicialmente caracteriza, no âmbito das bases de dados convencionais, os sistemas que fornecem um ambiente interoperável entre múltiplas bases de dados e apresenta os conceitos necessários para seu desenvolvimento. A seguir, faz uma revisão na literatura a respeito das metodologias e arquiteturas, considerando os aspectos mais significativos.

III.2. Sistemas de suporte à interoperabilidade

Os avanços nas tecnologias de rede e de comunicação proporcionaram um rápido crescimento na direção de ambientes de computação distribuída sobre vários locais que são interconectados via rede. Estas propostas têm levado uma ampla variação de soluções de Sistemas de Gerenciamento de Banco de Dados Distribuídos (SGBDDs) que oferecem mecanismos para a troca ou o compartilhamento de dados armazenados (BRIGHT *et al.*, 1994).

Um Sistema de Base de Dados Distribuída (SGBDD) é um sistema que proporciona visão lógica de dados armazenados em múltiplas Bases de Dados

fisicamente distribuídas sobre uma rede de comunicação (BATINI *et al.*, 1986). Sobre esta visão lógica os usuários podem realizar consultas e atualizações. O SGBDD oferece acesso aos dados através da transparência na gerência dos dados, e de mecanismos que proporcionam o controle de concorrência distribuído na execução das transações nas bases de dados dos sistemas componentes.

Este sistema tem sido classificado na literatura de SGBDD homogêneo ou heterogêneo (ELMAGARMID & PU, 1990). O SGBDD homogêneo é aquele cujas bases de dados componentes apresentam esquemas não ambíguos, que se utilizam do mesmo modelo de dados e linguagem de consulta. O SGBDD heterogêneo é aquele cujas bases de dados locais apresentam esquemas diferentes e empregam uma diversidade de modelos de dados e linguagem de consulta. Nesta classificação não é contemplada as heterogeneidades decorrentes da diversidade de plataformas e sistemas operacionais, uma vez que estas têm sido resolvidas através de protocolos de comunicação.

Na literatura são encontradas diversas propostas de arquiteturas que contemplam SGBDDs heterogêneos (SGBDDH). Estas arquiteturas têm sido denominadas na literatura de Banco de Dados heterogêneos e distribuídos, Múltiplas Bases de Dados (MBD) e Banco de Dados Federados (BDF). Na verdade não existe um consenso entre as definições destes termos, por exemplo:

- GOH *et al.* (1994) e HWANG *et al.* (1994), BRIGHT *et al.* (1994), LITWIN *et al.* (1990) tratam as denominações MBD, BDF e SGBDDH como sinônimos;
- ÖSZU & VALDURIEZ (1990) distinguem MBD e BDF, de acordo com o grau de autonomia entre as bases de dados;
- TARI *et al.* (1996), WANG & ZHANG (1996), FORST & BUKHRES (1995), BREITBARTH *et al.* (1995), ROSENTHAL & SELIGMAN (1994), CHEN (1990), KIM (1991) e KIM & SEO (1991) tratam MBD como uma modalidade de SGBDDH;
- PIRES (1997), MARRS^{html} *et al.* (1996), ZHAO *et al.* (1996), DUPONT & SPACCAPIETRA (1996), UCHÔA (1995), AGUIAR (1995), BUKHERS & ELMAGARMID (1996), DEACON *et al.* (1994), GOODMAN (1994), RADEKE & SCHOLL (1994a e 1994b) tratam BDF como uma modalidade de SGBDDH;
- KIM *et al.* (1995) tratam MBD como uma federação de bases de dados;
- BREITBARTH (1990), ELMAGARMID & PU (1990) e MILLINNER *et al.* (1995) distinguem MBD e BDF pelo processo que tornam os dados disponíveis. MBD

seria um sistema com um esquema global e o BDF um sistema com esquemas parciais e um processo de negociação para compartilhar os dados; e

- COLLET *et al.* (1991) considera tanto MBD como BDF como sistemas que apresentam um esquema global.

Neste trabalho será considerada uma Federação de Dados a associação que proporciona um ambiente MBD. Esta associação é materializada por um sistema MBD que suporta operações de acesso a múltiplas bases de dados heterogêneas e distribuídas, denominadas de bases de dados locais. Estas bases de dados mantêm coleções de dados e aplicações específicas. Elas participam do ambiente MBD, apresentado na Figura III.1, de forma a permitir o compartilhamento parcial/total e controlado de seus dados/serviços.

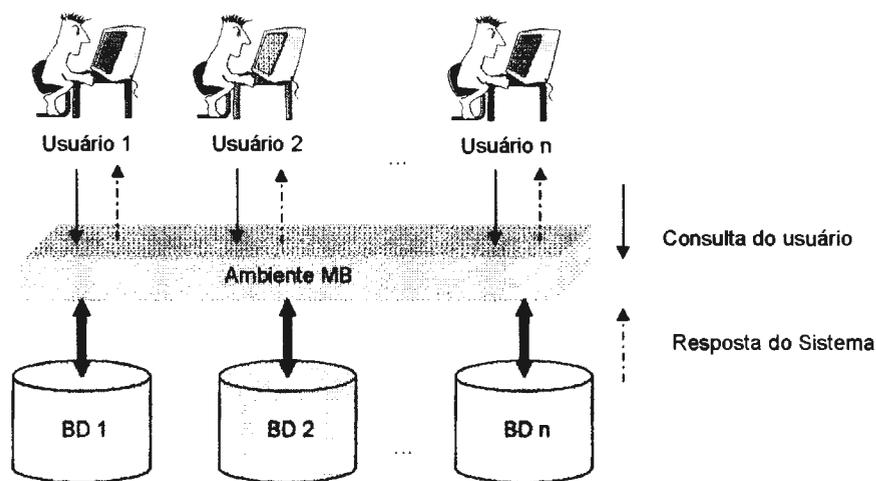


Figura III.1 - Ambiente MBD.

O sistema MBD proporciona um ambiente, também denominado de MBD, que oferece suporte a interoperabilidade, empregando os recursos de computação distribuída sobre uma rede que torna prático o acesso remoto às bases de dados locais. Neste ambiente nenhuma base de dados local deve exercer controle sobre a outra. As bases de dados locais são interconectadas através de redes de comunicação e inter-relacionadas de forma controlada e coordenada para permitir os usuários da Federação compartilharem seus dados. Contudo, o sistema deve preservar as aplicações existentes em cada base de dados local componente do ambiente MBD (STRAUCH *et al.*, 1997). Assim, o seu desenvolvimento deve atender a três requisitos básicos que tratam do gerenciamento de rede, a saber (ÖSZU & VALDURIEZ, 1991):

- as bases de dados devem ter autonomia local, ou seja as bases de dados compartilham os seus dados com as demais bases, sem afetar as aplicações existentes

nas bases de dados locais ou comprometer suas restrições de integridade, aplicações específicas ou mecanismos de segurança;

- os usuários devem ter acesso transparentemente aos dados dos demais sistemas, ou seja, os usuários devem acessar diferentes sistemas como se estivessem acessando a uma aplicação à qual já estão acostumados; e
- o ambiente de interoperabilidade deve possuir mecanismos de controle de concorrência eficientes, de forma que as consultas realizadas pelos usuários sejam executadas corretamente.

III.3.Características de um ambiente MBD

Um ambiente MBD é formado por bases de dados distribuídas, autônomas e provavelmente heterogêneas. A distribuição decorre do fato destas bases de dados estarem geograficamente distribuídas sobre uma rede de comunicação. A autonomia decorre do fato de elas serem projetadas, desenvolvidas e mantidas independentemente com seus respectivos SGBDs. A heterogeneidade é decorrente destas bases de dados serem projetadas para diferentes aplicações e gerenciadas por diversos tipos de SGBDs.

Segundo DBOUK *et al.* (1996), neste ambiente as heterogeneidades podem ser classificadas em heterogeneidades de baixo e alto nível. As heterogeneidades de baixo nível tratam das questões relativas a diversidade de plataformas, sistemas operacionais, protocolos de comunicação, sistemas de gerenciamento de dados e linguagens de consulta. Estas heterogeneidades causam ineficiência quando se faz necessária a troca de dados e são solucionadas por processos que consomem tempo. Os esforços para obter estas soluções no ambiente MBD têm se concentrado no estabelecimento de padrões como TCP/IP, RPC, SQL e mais recentemente os padrões propostos pelo *Object Management Group* (OMG) e *Object Database Management Group* (ODMG). As heterogeneidades de alto nível tratam das questões relativas às diferentes percepções da realidade, das necessidades de aquisição dos dados de cada usuário local e de sua representação nas bases de dados (SCHIMITT & SAAKE, 1996). Estas heterogeneidades são decorrente das bases de dados serem projetadas independentemente e atenderem diferentes aplicações. Elas causam os conflitos, no nível intencional, e inconsistências, no nível extencional, entre os esquemas das bases de dados.

Estes dois níveis de heterogeneidades devem ser consideradas no seu

desenvolvimento para se alcançar um eficiente compartilhamento de dados. Eles são tratados no desenvolvimento de um ambiente MBD através de questões táticas e operacionais abordadas pela metodologia para integrar os esquemas das bases de dados e por uma arquitetura para acessar as bases de dados locais.

Deste modo, o ambiente MBD é um sistema híbrido entre um SGBD centralizado e distribuído, uma vez que as bases de dados locais que compõem este ambiente são autônomas e centralizadas para os usuários locais e ao mesmo tempo para os usuários do ambiente MBD elas são distribuídas. Ele tem como particularidades permitir que as aplicações existentes permaneçam operacionais e ser orientados a dados. Estas particularidades requerem a compreensão de dois conceitos fundamentais, a saber: a autonomia e a cooperação entre as bases de dados locais.

III.3.1. Autonomia no ambiente MBD

A autonomia das bases de dados locais é a principal característica de um sistema MBD. ÖZSU & VALDURIEZ (1990) definem autonomia como o grau que indica quanto cada sistema local pode operar independentemente na integração de bases de dados heterogêneas e distribuídas. A determinação deste grau envolve os seguintes fatores: i) troca de informações entre as bases componentes da integração; ii) execução das transações de forma independente; e iii) permissão para um sistema local modificar outro sistema local.

Da perspectiva do desenvolvimento de um ambiente MBD, são identificados na literatura (ÖZSU & VALDURIEZ, 1990; SCHEUERMANN *et al.*, 1990 e BREITBART *et al.*, 1995) três tipos de autonomia, cada uma correspondendo a um tipo de cooperação:

- autonomia de projeto - refere-se à capacidade de cada SGBD local escolher seu próprio modelo de dados e procedimentos de implementação, que devem ser mantidos na integração das bases de dados heterogêneas, pois além de ser caro alterá-los, geralmente causa problemas de manutenção das aplicações;
- autonomia de execução - refere-se a habilidade de um sistema local reter o completo controle sob a execução de transações emitidas por outro sistema, decidindo quando e como executá-las, sem nenhum ou com um mínimo impacto na execução e desempenho das transações dos SGBDs locais. Esta autonomia fornece aos sistemas a capacidade de abortar a execução das transações locais a qualquer momento

durante sua execução, inclusive quando uma transação do ambiente MBD está em processo de salvamento; e

- autonomia de comunicação - refere-se a liberdade de cada SGBD local decidir com que SGBDs locais irão se comunicar e que informações serão trocadas entre eles, bem como a capacidade dos SGBDs locais não compartilharem sua informação de controle com os outros sistemas ou com o próprio ambiente MBD.

Nessa classificação está presente uma analogia apresentada por ÖSZU & VALDURIEZ (1990) entre três aspectos relevantes a serem observados no desenvolvimento de arquiteturas propostas para alcançar a interoperabilidade no ambiente MBD, a saber: a) heterogeneidade; b) autonomia propriamente dita e c) distribuição entre as várias bases de dados e seus SGBDs locais, respectivamente (Figura III.2). Os autores dispõem estes aspectos em três eixos ortogonais, e neste contexto a heterogeneidade reflete as diferenças entre modelos de dados, plataformas, aplicativos para o gerenciamento dos dados e protocolos de comunicação; a autonomia reflete a distribuição do controle, ou seja quanto cada sistema pode operar independentemente; e finalmente, a distribuição que reflete a distribuição física dos dados sobre as várias bases componentes de um ambiente MBD.

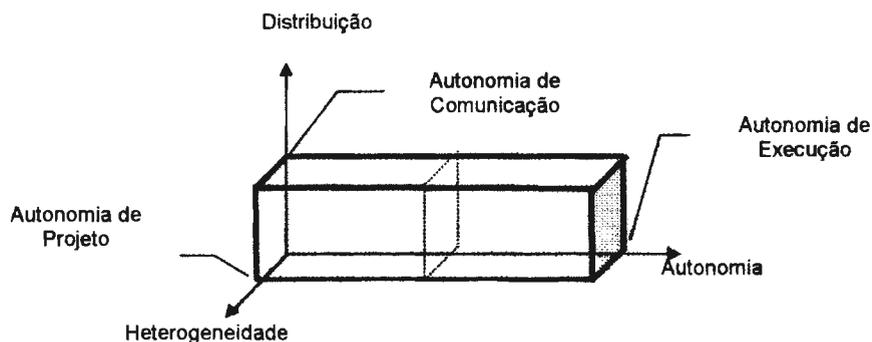


Figura III.2 - Analogia para as autonomias de projeto, execução e comunicação (ÖSZU & VALDURIEZ, 1990).

Assim, no desenvolvimento de uma arquitetura para criar um ambiente MBD é necessário ter definida a autonomia para o ambiente MBD e para os sistemas locais, observando os graus de heterogeneidade, distribuição e a autonomia, propriamente dita.

Segundo BREITBART *et al.* (1995) deve-se levar em conta na definição da autonomia que quanto mais os sistemas componentes do ambiente MBD são autônomos mais difícil é garantir uma consistência global. Por outro lado, definir a autonomia do ambiente MBD mínima, também pode não satisfazer as necessidades básicas dos sistemas locais de independência, tornando o ambiente MBD inaceitável para os

usuários.

No desenvolvimento de um ambiente MBD deve-se assegurar a autonomia de projeto, de execução e de comunicação dos sistemas locais, de forma que estes não sejam modificados e que mantenham o controle sobre seus dados e processamento. Isto permite que as organizações preservem seus investimentos em SGBD.

Deste modo, a autonomia de projeto é mantida pela adoção de um modelo de dados comum e através de regras de mapeamento entre os dados, a autonomia de execução e de comunicação é suportada pelo desenvolvimento de arquitetura em rede, que oferece serviços de acesso às bases de dados locais sem afetar suas aplicações locais.

III.3.2. Cooperação no ambiente MBD

Antes de abordar a cooperação, se faz necessário compreender os papéis e atores encontrados no ambiente MBD. Assim, neste ambiente existem dois tipos de papéis fundamentais: os fornecedores e os consumidores de dados. Os fornecedores de dados dispõem todo/parte do seu esquema para federação. Os consumidores de dados desejam acessar e interpretar a informação, isenta de erros e problemas operacionais, através de recursos dispostos pelo ambiente MBD (GARDELS, 1997). Assim os fornecedores de dados são responsáveis pela qualidade e integridade dos dados no ambiente MBD e sua utilização é responsabilidade dos consumidores de dados da federação.

Estes papéis são realizados pelos seguintes atores, que podem ser encontrados no ambiente MBD (DUPONT & SPACCAPIETRA, 1996):

- os usuários locais - têm acesso somente dados locais, sendo a federação transparente para eles;
- os usuários da federação - têm acesso às bases de dados componentes da federação manipulando a informação compartilhada, através de consultas realizadas sobre um ou mais esquemas federados;
- os administradores locais - são responsáveis pela administração das bases de dados locais; e
- o(s) administrador(es) da federação - trata-se de um ou mais responsáveis pela visão lógica das bases de dados componentes da federação.

A cooperação no ambiente MBD pode ser alcançada através da integração ao nível de esquema ou de aplicação. Esta cooperação não requer conhecimento do usuário

da federação sobre a localização dos dados ou as características das diferentes bases e seus correspondentes SGBDs. Ela é refletida pelo controle e limitada no processo de integração das bases de dados.

Segundo ZHAO *et al.* (1995), no desenvolvimento do ambiente MBD as bases de dados se associam sob uma administração e, a seguir, estabelecem um grau de cooperação entre as bases de dados locais, que definem a proposta de compartilhar dados e ao mesmo tempo manter um forte controle sobre os dados locais. Este grau de cooperação determina os dados que as bases de dados desejam compartilhar, com quem compartilhar e como compartilhar.

Assim, cada base de dados pode cooperar em diferentes graus, de acordo com um critério de segurança estabelecido. Este critério é estabelecido através de permissões de leitura e escrita locais ou, ainda, através de critérios para determinar as bases a serem acessadas na realização da consulta. Cada administrador local negocia as partes do seu esquema que estarão disponíveis à federação de acordo com a permissão de leitura e escrita na base de dados local.

Segundo SHETH & LARSON (1990) este processo de negociação pode ser de duas maneiras. Na primeira, o administrador local permite ao(s) administrador(es) da federação ler o esquema local, identificar as necessidades de acesso aos dados e enviar uma solicitação ao administrador local para definir um esquema exportável com acesso apropriado. Esta solicitação descreve os tipos de acesso e usuários da federação que poderão acessar os dados. Na segunda, todos os administradores locais definem as partes dos esquemas locais a serem dispostos para a federação à priori e oferecem acesso aos administrador(es) e a qualquer usuário da federação, limitado apenas à leitura.

Em qualquer um dos procedimentos acima, o grau de cooperação é estabelecido por um processo de negociação entre o(s) administrador(es) da federação e os diferentes administradores locais. O processo de negociação consiste na troca de mensagens entre os administradores, de forma a: i) tornar disponível todo/parte dos esquemas locais à federação de dados; ii) identificar as partes comuns e as heterogeneidades entre os esquemas; e iii) desenvolver soluções para resolver estas heterogeneidades e fixar os mapeamentos necessários para cada comunidade de informação.

Em arquiteturas mais rígidas, o processo de negociação é realizado entre um administrador da federação, o qual é o responsável pelo esquema no ambiente MBD, e os administradores locais responsáveis pelos esquemas exportáveis das bases de dados

locais (HEIMBGNER & MCLEOD, 1985). Em arquiteturas mais flexíveis não há administrador(es) da federação, apenas administradores locais, entre os quais é realizado o processo de negociação.

Tanto a metodologia quanto a arquitetura para a integração devem suportar este processo de negociação, de uma forma segura e eficiente, através de um comportamento cooperativo. A metodologia estabelece critérios para ajudar os administradores a identificar o conteúdo dos esquemas de exportações e operações a serem realizadas, de forma a contribuírem coletivamente para a solução de um domínio comum do problema. A arquitetura deve fornecer ferramentas para a troca de mensagens entre os administradores.

Cabe ressaltar que este processo de negociação deve ser flexível, de modo que a federação não seja fechada, mas sim aberta. O conceito de federação aberta está associado à facilidade com que outras bases de dados, no futuro, possam entrar/sair da federação.

III.4. Aspectos de representação dos dados em um ambiente MBD

O desenvolvimento do ambiente MBD é considerado um projeto ascendente, uma vez que, a partir das bases de dados existentes é desenvolvido um sistema que as integra. Este desenvolvimento compreende um processo de modelagem que tem por objetivo compreender os esquemas, identificar as heterogeneidades, resolver as incompatibilidades e superposições entre os esquemas e integrar os recursos heterogêneos e distribuídos, através de aplicações globais. Na realidade, este processo compreende uma modelagem da informação a ser disponível e uma modelagem da aplicação de consulta sobre o ambiente MBD (CHORAFAS & STEINMANN, 1993).

A modelagem da informação possibilita o uso efetivo dos dados no ambiente MBD, através da criação de uma estrutura abstrata de dados comuns. Para isto, se faz necessário, primeiro, o projetista de um ambiente MBD aprender que fontes de informações desejam participar da federação, onde elas estão localizadas, identificar os tipos de gerências (hierárquico, relacional, orientado a objeto) e o que elas contêm (SPACCAPIETRA & PARENT, 1991).

Desta forma, os sistemas locais devem adotar um formalismo para descrever, ou permitir descrever as informações, denominadas de metadados, que revelam a semântica, a sintaxe, os requisitos de consistência e as restrições de integridade de seus

esquemas e aplicações relacionadas (BONJOUR & FALQUET, 1994).

Entretanto, os esquemas das bases de dados locais podem estar representados por modelos de dados heterogêneos e/ou o usuário utilizar diferentes denominações, ou representação dos elementos de dados. Desta forma, se faz necessário estabelecer um Modelo de Dados Comum (MDC) para o qual os esquemas são traduzidos. Posteriormente, os esquemas descritos no MDC são analisados e identificados os conflitos semânticos e sintáticos que por ventura existirem. A seguir, são estabelecidas técnicas para resolver os conflitos e regras de mapeamentos entre os elementos dos esquemas, de forma a pré-integrá-los ou integrá-los.

No processo de modelagem da aplicação são identificadas as necessidades de serviços globais para satisfazer os usuários da federação. Por exemplo: a) serviços de interfaces com o usuário que permitam descrever as bases de dados locais, traduzir de um modelo de dados local para um MDC, ou seja, submeter uma consulta; e b) serviços de rede que permitam o acesso aos esquemas das bases de dados locais e a integração de recursos heterogêneos e distribuídos através de aplicações globais.

Deste modo as próximas seções discorrem onde encontrar a semântica dos dados nas bases de dados locais, a importância dos metadados nestas arquiteturas, e a utilização de um modelo de dados canônico.

III.4.1. Semântica dos dados

Segundo SHETH & RUSINKIEWICZ (1990), no desenvolvimento de um ambiente MBD estão presentes múltiplas interpretações dos dados, dadas as necessidades dos diversos usuários, nos diferentes contextos, em modelarem o mundo real. As múltiplas interpretações levam as bases de dados a apresentarem semânticas diferentes.

As propostas iniciais de integração de esquemas consideravam o problema de detecção da proximidade semântica entre esquemas de forma rudimentar. Elas comparavam apenas as sintaxes dos esquemas e detectavam os conflitos estruturais que surgem quando a mesma abstração da realidade é representada de forma diferente, devido a utilização de vários construtores estruturais. Todavia, este tipo de comparação se mostrou inadequado, uma vez que determinava as equivalências apenas entre os atributos, sem considerar o significado (semântica) dos dados. Esta comparação é necessária, mas não suficiente para determinar o relacionamento semântico entre os

esquemas.

Assim, os métodos passaram a considerar a semântica dos dados de cada comunidade de informação no processo de integração, de modo a ajudar na interpretação dos esquemas, na compreensão das disparidades de representação e das características, mesmo as implícitas, de cada sistema existente (KASHYAP & SHETH, 1996).

A semântica dos dados consiste no significado do dado e está relacionada a seu uso. Ela é definida, segundo BONJOUR & FALQUET (1994), por um termo que projeta um conceito, o qual explicita uma definição dentro de um domínio (Figura III.3). O conceito é a representação de um aspecto isolado da realidade e o domínio é a atividade, problema ou disciplina que possui uma aplicação particular.

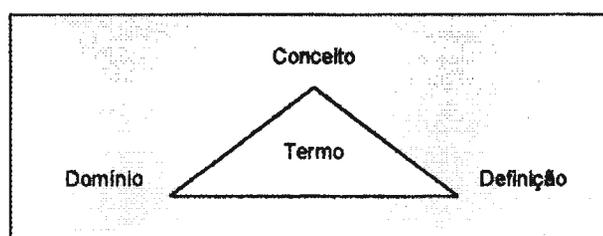


Figura III.3 - Semântica do dado.

Em uma definição mais abrangente a semântica é tratada como um mapeamento entre um objeto modelado, representado e/ou armazenado em um sistema de informação, e o objeto do mundo real representado. Este mapeamento representa a semântica do objeto modelado pela descrição ou identificação da perspectiva do significado e de uso (SHETH, 1995).

O valor semântico é associado a um contexto e este é modelado, usando-se os metadados (SCIORE *et al.*, 1994). SCIORE *et al.* (1994) definem contexto como uma parte do dado a ser relacionada aos metadados: seu significado, propriedades e organização. Esta definição pode variar, abrangendo desde unidades, qualidade da informação a formato dos dados. GOH *et al.* (1994) definem o contexto como uma suposição que esquematiza o modo no qual um agente interoperável representa ou interpreta os dados. Contexto de troca é definido por MADNICK (1995) como uma coleção de suposições implícitas sobre o significado e as características (qualidade) da informação. A representação do contexto, bem como os metadados e vocabulários são elementos importantes para comparar e integrar os esquemas. O contexto é definido ainda por KASHYAP & SHETH (1996) como o conhecimento necessário ao raciocínio

sobre o sistema, de forma a auxiliar a identificação do conteúdo, da organização e das propriedades dos dados. Estes últimos autores propõem ainda a utilização do contexto, da ontologia com conhecimento extra para compor a base da taxionomia semântica.

As recentes pesquisas do processo de integração passaram a considerar os relacionamentos semânticos entre os esquemas. Para isto, se faz necessário capturar informações que auxiliarão a detectar a semântica associada às entidades/classes e aos seus atributos. Entretanto, tornar disponível a informação semântica é um processo particularmente difícil. Para capturá-la é necessário que as convenções e características dos esquemas das bases de dados sejam expostas e explicadas. Isto requer a leitura de esquemas, entrevista com os usuários e desenvolvedores, a análise dos ambientes de execução e documentação das observações, de modo que possam ser compreendidas por outras pessoas (HEILER, 1995).

Dois técnicas têm sido empregadas. A primeira consiste em adquirir documentação disponível dos dados e amostras destes, realizar entrevistas com os administradores locais e os usuários através de um processo iterativo. Esta técnica além consumir tempo, não é acurada, uma vez que os sistemas são pobremente documentados e nem sempre toda a semântica dos esquemas é percebida por todos. A segunda técnica consiste em deixar os administradores locais, responsáveis pelo desenvolvimento e manutenção dos esquemas, utilizarem ferramentas para descreverem seus esquemas, os metadados e para disporem estas descrições aos outros administradores locais ou a um administrador central.

RIBEIRO & OLIVEIRA (1996) sugerem que a extração da semântica de um esquema seja realizada através da análise dos códigos das aplicações, visto que muitas vezes estes contêm a semântica estrutural e, principalmente, a comportamental do esquema. Em geral, as aplicações explicitam os dados a serem usados nos relatórios e análises locais. Entretanto, nem todos os dados de um esquema se encontram nestas aplicações.

A captura da semântica é um fator crítico na integração de esquemas, uma vez que é difícil especificar completamente a semântica de todos os esquemas, seja manual ou automaticamente.

A detecção de relacionamentos semânticos entre esquemas das bases de dados também é uma tarefa crítica no processo de integração, pois exige um excessivo número de comparações para identificar as similaridades e heterogeneidades existentes entre os esquemas. Isto torna esta comparação, complexa além de requerer aplicação de

inteligência e julgamento humanos.

Assim, se faz necessário o desenvolvimento de uma metodologia que limite e conduza as comparações a serem empregadas, bem como estabeleça um processo de cooperação para validá-las.

III.4.2. Metadados: representação da semântica dos dados

Os metadados constituem pré requisitos para implementar um ambiente MBD e vão além de sua definição: dados sobre os dados. CHORAFAS & STEINMANN (1993) propõem que os metadados, para o desenvolvimento de um ambiente MBD, definam os dados, seu escopo de relevância, estratégia de atualização, frequência e identificação de acesso e regras para mecanismos de troca de dados.

Eles são utilizados no ambiente MBD como entidades que encapsulam unidades de informação física que descrevem uma visão estruturada e organizada da coleção dados e do sistema. A descrição da coleção de dados visa documentar o conteúdo e o contexto, de forma a representar o significado dos dados. O conteúdo descreve a estrutura de cada arquivo: os itens de dados quanto ao seu tipo, formato, domínio e restrições. O contexto descreve informações sobre o histórico, a qualidade e outras informações associadas à coleta, ao processamento e à utilização dos dados. E quanto ao sistema, eles descrevem informações sobre a localização, a plataforma, o aplicativo, a documentação, os usuários, bem como outras informações relevantes para os administradores envolvidos no ambiente MBD. Por exemplo, informações sobre o mapeamento e restrições associadas a cada base de dados local.

Os metadados são armazenados e gerenciados por um catálogo e sua utilização no ambiente MBD tem três objetivos: *i)* auxiliar o processo de integração, *ii)* como elemento de pesquisa dos dados que estão disponíveis na federação, e *iii)* na recuperação dos dados. No processo de integração, os metadados oferecem uma forma padronizada para descrever as bases de dados e são usados para encontrar objetos com proximidade semântica, ajudando a responder as seguintes questões: O que? Aonde? Qual o domínio dos dados? Neste processo, os metadados fornecem ainda subsídios para determinar as transformações a serem aplicadas, de modo a resolver as heterogeneidades. No processo de pesquisa, alguns sistemas permitem ao usuário visualizá-los de forma a encontrar os dados desejados, conhecer suas estruturas e formular a consulta. Na recuperação dos dados, o sistema MBD consulta os metadados

para encontrar os mapeamentos necessários e para processar, otimizar e distribuir a consulta. Este catálogo pode se apresentar nas arquiteturas MBD como um catálogo único e centralizado ou múltiplos catálogos distribuídos na rede, mas associado a cada base de dados local.

III.4.3. Modelo de dados comum (MDC)

O modelo de dados é o mecanismo da tecnologia de Banco de Dados para abstrair uma realidade e representá-la. Em geral, os esquemas das bases de dados locais podem empregar modelos de dados heterogêneos. Todavia, no processo de integração de esquemas para um ambiente MBD é importante que o projetista compreenda e raciocine sobre as semânticas dos esquemas locais para encontrar as similaridades. Desta forma, para facilitar este processo se faz necessário adotar um modelo de dados único, denominado de modelo de dados canônico (MDC).

Segundo SOUZA (1986) a adoção deste modelo no desenvolvimento de uma federação tem por objetivo oferecer uma compreensão normalizada do significado dos dados armazenados por cada esquema, facilitar a comparação dos esquemas no processo de integração, e ser um mecanismo comum sobre o qual é realizado uma consulta.

O MDC consiste em um modelo de dados para o qual todos os esquemas das bases de dados componentes são transformados, de modo a minimizar as diferenças sintáticas conseqüentes do uso de diferentes modelos de dados. Esta transformação dos esquemas locais para o MDC não é uma simples tradução de um modelo para o outro. Isto inclui um processo de enriquecimento semântico para capturar a semântica implícita nas aplicações locais.

O MDC é usado como referência padrão para o qual todos os esquemas devem ser traduzidos e analisados no processo de integração. A escolha de um modelo de dados para expressar o modelo conceitual dos esquemas é um fator crítico no processo de integração de esquema. Esta decisão influencia na qualidade da etapa de integração. Além disto, o conceito de federação aberta, no qual outras bases de dados podem entrar na federação no futuro, introduz a necessidade do modelo de dados, a ser adotado como MDC, assegurar que as novas bases de dados tenham modelos de dados mais pobres ou iguais a ele.

Um modelo simples, ou semanticamente pobre, isto é, com poucos construtores de modelagem, é uma ferramenta fraca nas mãos de um projetista do ambiente MBD

para descobrir as similaridades/heterogeneidades. Apesar de um MDC com esta característica empregar operações de tradução e integração simples que envolvem poucas operações primitivas, ele tem a desvantagem de não expressar toda a semântica dos esquemas locais, apenas um subconjunto, limitando a representação do esquema local e as informações necessárias para a integração. Em geral, a sua utilização requer um processo de enriquecimento semântico dos esquemas, através de um processo de aquisição de conhecimento pelo qual todas as semânticas extraídas são explícitas em uma representação formal para cada base de dados local.

Um modelo semanticamente rico, isto é, um modelo que oferece construtores para modelar a semântica capturando os aspectos estruturais e comportamentais dos esquemas locais, bem como qualquer semântica adicional obtida do processo de enriquecimento. Um modelo de dados semanticamente rico, como MDC, facilita a detecção de similaridades/heterogeneidades entre esquemas em um nível mais alto de abstração, todavia o número de heterogeneidades representadas é maior. A operação de integração de esquemas é facilitada com o uso dos construtores do modelo de dados e operações de integração.

Um modelo de dados para MDC deve oferecer uma riqueza de representação que permita detalhar os esquemas e compreender a semântica representada, de forma a capturar as diversidades de representações. O poder de representação deste modelo deve capturar as características específicas dos diversos modelos de dados, oferecer estruturas adequadas para o tratamento de heterogeneidade de informação e suportar a descrição de semânticas adicionais explícitas através do processo de enriquecimento semântico.

As características que um modelo de dados deve conter para ser adotado como MDC tem sido bastante pesquisadas na literatura. Dentre estas características, destacam-se (VENTRONE & HEILER, 1991 e OLIVEIRA & MAGALHÃES, 1993):

- capacidade de expressar diferentes níveis de abstração para representar as diversas semânticas contidas em outros modelos, de forma simples e clara;
- possibilidade de acrescentar informações relevantes, não contidas nos esquemas das bases de dados, que possam facilitar a comparação dos esquemas; como por exemplo, restrições de integridade implícitas ou mantidas pelas aplicações;
- facilidades para representar interdependências e diferenças entre os esquemas e suportar mecanismos para reforçá-las; e
- habilidade para expressar a informação semântica comum a todos os usuários do

ambiente MBD.

Este trabalho se detém nas características propostas por SALTOR *et al.* (1991) que desenvolveram uma estrutura de características que o modelo de dados deve possuir para atender a conveniência do MDC no processo de integração. Esta estrutura é baseada em dois fatores de habilidade de representação de um modelo:

- expressividade semântica - corresponde ao grau com que o modelo representa uma conceituação expressa de forma natural. Esta expressividade pode ser decomposta em:
 - expressividade estrutural - diz respeito ao poder das estruturas do modelo para representar e interpretar os conceitos; e
 - expressividade comportamental - trata do poder do modelo para representar comportamento dos conceitos.
- relativismo semântico - é o grau com que o modelo de dados pode acomodar diferentes conceituações do mundo real. Expressa o poder das operações do modelo para derivar esquemas externos, a partir do esquema da base de dados.

As características da expressividade são descritas através dos conceitos de classificação/instanciação, generalização/especialização, agregação/decomposição e da capacidade de definir e encapsular operações e restrições de integridade (SALTOR *et al.*, 1991). O conceito de classificação/instanciação suporta o conceito de classes, facilitando distribuir os elementos no processo de integração. O conceito de generalização/ especialização suporta a superposição entre classes, através do conceito de hierarquia de herança. Isto permite que as hierarquias de especialização sejam integradas tornando-se uma única hierarquia de generalização. O conceito de agregação/decomposição permite agregar objetos e a associação destes para formar um objeto complexo. A definição e encapsulamento de operações e restrições de integridade permite mais naturalmente esconder as bases de dados locais, operações e as soluções para resolver os conflitos entre os esquemas.

As características do relativismo semântico são descritas através das seguintes habilidade do modelo: *i)* suportar a definição de operadores para a integração de esquemas, *ii)* produzir qualquer tipo de estrutura necessária para derivar esquemas externos, *iii)* e facilitar a integração criando uma estrutura básica para a comparação (SALTOR *et al.*, 1991). Segundo SPACCAPIETRA & PARENT (1994), suportar relativismo semântico é a chave para capacitar o usuário a construir sua própria visão, de acordo com a percepção do mundo real e independente de outros usuários.

Inicialmente a tecnologia de Banco de Dados empregou os modelos hierárquico e rede, que empregavam estruturas simples para representar a realidade a ser modelada pelos esquemas. Estes modelos se mostraram inadequados, tanto para representação do esquema quanto para MDC, uma vez que se encontravam mais próximos da implementação do Banco de Dados, não representando adequadamente a semântica dos dados.

Visando suprir as necessidades dos modelos anteriores vieram, uma diversidade de modelos denominados de semânticos que eram caracterizados por representar a realidade, modelada, partindo de um ponto de vista mais significativo desta. Entretanto, apesar destes modelos oferecerem mecanismos para representarem a realidade com maior riqueza semântica, eles se mostraram inadequados por se distanciarem da implementação do Banco de Dados. Nesta fase encontra-se a iniciativa de SOUZA (1986) em utilizar o esquema conceitual abstraído como MDC para expressar a sintaxe e quantificar a similaridade entre construtores dos esquemas.

Mais tarde vieram os modelos que ofereciam um formalismo para representar a semântica do mundo real: o modelo relacional, o entidade-relacionamento, o relacional estendido e o orientado a objeto.

Inicialmente foi utilizado o modelo relacional que, apesar de usar toda a álgebra relacional para derivar esquemas externos (visões), possui pouca expressividade estrutural para ser utilizado como MDC, uma vez que ele não possui os conceitos de generalização/especialização, agregação/decomposição e a definição de operadores para integração. A seguir, foi utilizado o modelo entidade relacionamento que atuava como um padrão de fato, na área, com métodos e ferramentas adequados a sua modelagem conceitual. Entretanto, seu poder de expressão satisfaz precariamente o conceito de agregação/decomposição, a definição de operações e a criação de uma estrutura básica para comparação. Logo depois vieram as extensões do modelo relacional. Estes satisfazem quase todas as características, porém não suportam adequadamente a definição de operadores para integração. Por exemplo, BREITBART *et al.* (1986), para o desenvolvimento de MBD ADDS, utiliza o modelo relacional estendido para se beneficiar das operações da álgebra relacional.

Com o advento do paradigma da orientação a objeto, as pesquisas têm sido conduzidas nesta direção. O modelo orientado a objeto permite capturar a semântica e as funções de outros modelos de dados, com pouca ou nenhuma perda de semântica durante a transformação dos esquemas locais para o MDC. Para SALTOR *et al.* (1991)

o modelo orientado a objeto satisfaz todas as características essenciais de um MDC, visto que possui uma maior expressividade e relativismo semânticos do que os outros modelos de dados.

Neste modelo, um objeto representa a percepção de uma entidade abstrata e concreta do mundo real. Cada entidade do mundo real é modelada como objeto. Cada objeto é associado a um único identificador que o torna distinguível dos outros. Ele possui um conjunto de propriedades e métodos que representam sua estrutura e seu comportamento, respectivamente. O valor de cada propriedade pode ser um objeto ou um conjunto de objetos. Os valores das propriedades representam o estado do objeto. Este estado é acessado ou modificado, enviando mensagens para o objeto que invoca o método correspondente. O conjunto de mensagens que pode ser enviado para um objeto é a interface deste. Um objeto satisfaz uma interface se ele fornece métodos para responder as mensagens. Os objetos que compartilham a mesma estrutura e comportamento são agrupados em classes. Uma classe representa um conjunto de objetos similares. Uma classe pode ser definida como uma especialização de uma ou mais classes. Uma classe definida como especialização é chamada de subclasse e herda propriedades e métodos das suas superclasses.

Todavia, cabe ressaltar que existem diferentes propostas para os modelos orientados a objetos. Elas variam nos conceitos do paradigma da orientação a objeto de forma a refinar as abstrações para representar o mundo real.

Na literatura são encontrados algumas aplicações de modelos de dados orientados a objetos como MDC. Por exemplo, HAMMER (1994) utiliza um modelo denominado *Core Object Data Model (CODM)* para descrever a estrutura, restrições e operações para compartilhar dados. Este é um modelo de objetos funcional. Ele suporta as abstrações da orientação a objeto e a definição de funções estabelecidas pelos usuários que determinam as transformações entre os objetos. Um segundo exemplo é o modelo orientado a objeto OMS-EITIS proposto por UCHÔA (1995) que suporta meta-informação para descrever um esquema local. Um terceiro exemplo é uma extensão semântica do modelo orientado denominado de BLOOM, proposto por SOLACO *et al.* (1995), para capturar relacionamento semântico de agregação e especialização entre as classes. Estes relacionamentos são utilizados no modelo BLOOM para identificar a proximidade semântica entre as classes.

Dentre os principais benefícios da adoção de um MDC orientado a objeto para a integração pode-se citar (BERTINO *et al.*, 1994; ELIASSEN & KARLSEN, 1991;

PITOURA *et al.*, 1995; BRODIE & STONEBRAKER, 1995):

- A utilização do identificador do objeto fornece uma identificação única no ambiente MBD que independe do valor do objeto e de sua localização;
- Os conceitos de classificação, generalização e composição de objetos são ferramentas essenciais na integração de esquemas para capturar a semântica dos dados e modelar objetos próximos;
- O mecanismo de herança oferece mais flexibilidade para acomodar as necessidades de diferentes aplicações e simplifica a definição de visões através da reutilização de mapeamentos definidos em outras visões;
- A utilização da parte comportamental dos objetos (métodos) permite definir um conjunto de funções de mapeamentos e de acesso a serem aplicadas aos objetos globais.
- O polimorfismo proporciona aos projetistas mais liberdade na definição de funções globais; e
- A separação da interface dos objetos de sua implementação através do encapsulamento dos métodos oferece um meio mais natural para esconder as diferenças entre os sistemas e realizar transformação entre modelos;

Cabe ressaltar que após a tradução dos esquemas para o MDC ainda existirá as heterogeneidades decorrentes das diferenças semânticas resultantes das várias conceituações abstraídas no processo de modelagem. Estas heterogeneidades são tratadas pela metodologia utilizada para integrar os esquemas expressos no MDC (ELIASSEN & KARLSEN, 1991).

III.5. Metodologia para integração de esquemas

A integração de esquemas é uma atividade que tem sido bastante utilizada pela tecnologia de Banco de Dados. Segundo BATINI (1986) esta atividade ocorre em dois contextos.

O primeiro, denominado de integração de visões, consiste em estabelecer um filtro específico para cada usuário sobre a base de dados, de forma a assegurar a privacidade dos dados (VIDAL, 1994). A integração de visões é um processo *top down*, para produzir um esquema conceitual global a partir das especificações de visões do usuário sobre uma base de dados centralizada ou distribuída.

O segundo ocorre no desenvolvimento de um ambiente MBD. Trata-se de uma

etapa de um projeto MBD que ajuda a produzir um esquema global ou esquemas federados, a partir de um conjunto de esquemas locais existentes, formalmente definidos (SPACCAPIETRA & PARENT, 1994; DUPONT & SPACCAPIETRA, 1996). O processo de integração de esquemas é um processo *botton up*, que tem por finalidade integrar, a posteriori, os esquemas de várias bases de dados preexistentes e independentes, contendo representações incompletas e heterogêneas, de modo a fazê-las cooperarem ou operarem umas com as outras, comportando-se como se elas fossem parte de uma única base de dados. Neste processo é criado um ou mais esquemas resultantes que as integram. Este esquema esconde os detalhes das estruturas fisicamente armazenadas nas bases de dados componentes da federação:

$$\text{ESQUEMA} = \int \text{F(ESQUEMA}_1) \cdot \text{F(ESQUEMA}_2) \cdot \dots \cdot \text{F(ESQUEMA}_n)$$

Diversas metodologias têm sido amplamente investigadas para conduzir o processo de integração de esquemas das bases de dados tradicionais (SOUZA, 1986; BATINI *et al.*, 1986; GOTTHAR *et al.*, 1992; BONJOUR & FALQUET, 1994; HAMMER, 1994; FRANKHAUSER *et al.*, 1995; PARENT *et al.*, 1996). Elas têm por objetivo harmonizar os esquemas das bases de dados locais e construir uma representação consistente destes. Estas metodologias requerem a compreensão das diferenças entre as bases de dados. Para isto, se faz necessário capturar a semântica dos dados, os requisitos de consistência e as restrições de integridade. A seguir estas metodologias aplicam um conjunto de critérios, segundo uma estratégia de comparação entre os esquemas, para explorar as estruturas sintáticas e semânticas, de modo a identificar as similaridades e os conflitos. Identificados os conflitos entre os esquemas, são desenvolvidas soluções para resolvê-los ou atenuá-los e criar um ou mais esquemas federados (STRAUCH, 1997).

Segundo VERMEER & APERS (1996), a integração de esquemas é complexa, visto que apresenta um caráter objetivo e outro subjetivo. A objetividade é fornecida pelas declarações que descrevem a modelagem do esquema, como por exemplo, as propriedades, as restrições e as regras de mapeamento. A subjetividade é por causa da necessidade de capturar a semântica real dos dados representados nos esquemas das bases de dados, o que por vezes é obtida através de suposições implícitas no contexto de uma particular aplicação. Deste modo, uma metodologia para integração de esquemas envolve o desenvolvimento de técnicas e processos para assegurar que a informação não seja perdida nem corrompida.

De maneira geral, as metodologias são baseadas no modo como os esquemas são integrados. Assim, se faz necessário compreender a anatomia dos esquemas que poderão estar envolvidos no processo de integração. Este trabalho antes de apresentar as taxionomias encontradas na literatura, para melhor compreensão, apresenta a anatomia dos esquemas que podem estar envolvidos no desenvolvimento de uma arquitetura para o ambiente MBD.

III.5.1. Anatomia dos esquemas envolvidos em um ambiente MBD

A integração de esquema no ambiente MBD pode ser realizado em diferentes níveis. SHETH & LARSON (1990), PITOURA, *et al.* (1995), HEIMBIGNER & MCLEOD (1985), RADEKE (1995) apresentam a classificação dos esquemas que podem estar envolvidos no desenvolvimento de um ambiente MBD (Figura III.4), a saber:

- Esquema local - Cada base de dados componente da federação possui um esquema local que define todos os dados locais. Este esquema representa a implementação do esquema conceitual expresso no modelo de dados do sistema local. Ele utiliza os construtores disponíveis para descrever as características físicas da estrutura lógica dos dados e para executar operações. Dentre estas características se encontram informações sobre a localização dos registros no sistema de arquivo físico; a representação física de relacionamentos entre os registros lógicos; e a utilização de indexes;
- Esquema componente - Corresponde a um esquema equivalente ao esquema local, de cada base de dados, componente expresso no modelo de dados do ambiente MBD. Este esquema também pode conter informações que estão implícitas no sistema local. Ele é criado através de um processo de tradução que gera mapeamentos, entre as representações em ambos os modelos. As principais razões de criar este esquema são: facilitar a integração e utilizar, posteriormente, os mapeamentos para tradução de consultas expressas no modelo do ambiente MBD para o modelo local. Este esquema desempenha o primeiro estágio no processo de integração de esquemas;
- Esquema de exportação - trata-se de uma porção do esquema componente que estará disponível na federação. Ele contém os dados de cada esquema local que estarão disponíveis no ambiente MBD. Sua elaboração consiste em desmembrar o

esquema componente em um subconjunto de esquemas que podem ser usados para controlar o acesso à informação, considerando-se o uso de cada usuário da federação. Este esquema tem a finalidade de restringir o grau com que cada base de dados participa do ambiente MBD. Ele oferece liberdade de compartilhamento e adaptabilidade para situações específicas, tais como:

- adição ou remoção de esquemas componentes à federação; e
- utilização de construtores de lingüística associados com cada SGBD;
- Esquema global ou esquema federado - trata-se de um esquema lógico que integra todos os esquemas de exportação das bases de dados componentes da federação. Ele representa a normalização das bases de dados locais em um ambiente global, fornecendo o relacionamento entre elas. Este esquema é obtido pelo processo de integração de todos os esquemas de exportação, através de operações de mapeamento entre esquemas e informações sobre a distribuição dos dados pelas bases de dados, componentes da federação. Os mapeamentos ajudam a resolver os conflitos entre os esquemas, especificação de objetos integrados, propriedades e relacionamentos. Este esquema viabiliza a decomposição das consultas globais formuladas em sub-consultas locais; e
- Esquema externo ou de importação - O usuário final da federação pode desejar ter acesso a todo o esquema ou às diferentes porções do esquema global. Estas porções são subconjuntos do esquema global, denominados de esquemas externos. Estes esquemas representam visões que são derivadas de partes do esquema global que os diferentes grupos de trabalho e suas aplicações podem ter acesso. Ele é criado para restringir o acesso às bases de dados componentes. As motivações para esta restrição incluem privacidade dos dados, manutenção de integridade do esquema local e fornecimento de visões lógicas ou de operações de alto nível.

Para desenvolver um ambiente MBD se faz necessário percorrer todos, ou partes destes esquemas, através de uma série de operações para criar um esquema global ou vários esquemas federados, a saber: operações de transformação, operações construtoras e operações de filtragem (SHETH & LARSON, 1990).

As operações de transformação executam ações para traduzir um esquema em um modelo de dados local para um esquema equivalente, expresso no MDC. Estas operações devem ainda implementar ações para validar os esquemas, de forma a assegurar conformidade e reforçar regras de restrições de integridade dos esquemas locais. As operações construtoras executam ações de mapeamentos dos dados e

comandos no ambiente MBD, para um ou mais sistemas locais. Estas ações visam resolver os conflitos, agrupar os dados em entidades usáveis, definir uniões necessárias e visões. As operações de filtragem são ações executadas ao nível dos esquemas ou dos sistemas para implementar restrições em dados e comandos.

Uma arquitetura que utiliza estes cinco níveis, citados acima, permite que os sistemas locais sejam traduzidos para os esquemas componentes. Os esquemas componentes são filtrados em um ou mais esquemas de exportação. A partir dos esquemas de exportações um esquema federado é construído. Finalmente, a partir de um esquema federado, um número de esquemas externos é construído (Figura III.4).

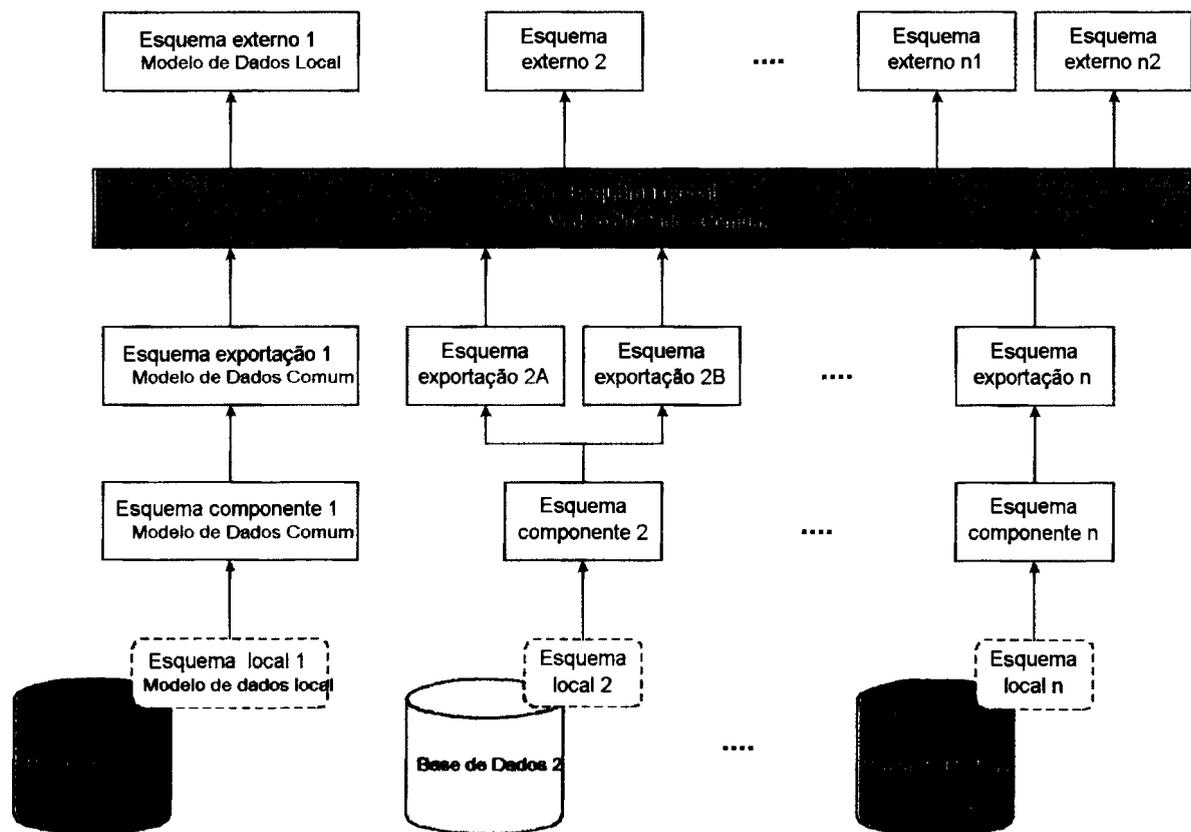


Figura III.4- Esquemas envolvidos em uma arquitetura em cinco níveis.

Em cada arquitetura a atividade de integração pode ser realizada em momentos diferentes e com finalidades diferentes. Por exemplo, uma arquitetura rígida é caracterizada por um esquema global que apresenta os dados disponíveis na federação. O processo de integração nesta arquitetura é realizado a priori, envolvendo todos os esquemas componentes ou exportáveis locais, expressos no MDC. Por outro lado, uma arquitetura flexível permite a existência de vários esquemas federados. O processo de integração é realizado no contexto de uma consulta sobre os esquemas exportáveis,

expressos no MDC, escolhidos pelo usuário ou administrador local. A integração pode ser realizada ao nível dos modelos de dados ou ao nível de aplicação para acessar os dados.

III.5.2. Heterogeneidades entre os esquemas das bases de dados

As heterogeneidades entre as bases de dados de diferentes comunidades de informação são decorrentes de três causas (SOUZA, 1986; BATINI *et al.*, 1986 e HAMMER, 1994). A primeira, trata das diferenças entre os esquemas conceituais. As conceituações armazenadas pelas bases de dados são resultantes do processo de abstração, e como tal capturam informações, parciais e relativas, de pontos de vista distintos, uma vez que cada usuário percebe e concebe a realidade de modo individualizado para sua aplicação. A segunda é decorrente de cada comunidade de informação utilizar diferentes modelos de dados para representar os esquemas conceituais e, como cada modelo de dados oferece uma diversidade de construtores que permite inúmeras possibilidades de modelagem; o processo de modelagem resulta em uma variedade de estruturas conceituais da base de dados. A terceira é decorrente das necessidades da especificação do projeto de uma base de dados em remover dados redundantes, reduzir possibilidades de inconsistências durante atualizações e melhorar a eficiência de avaliações de consultas.

Estas causas levam as bases de dados, desenvolvidas independentemente para atender às necessidades de seus usuários, a não compartilharem do mesmo conjunto de dados sobre a realidade que elas tratam, ou se compartilham os mesmos dados, estes não são representados da mesma forma, tornando os seus esquemas incompatíveis (diferentes) ou inconsistentes (falta informações) (GARCIA *et al.*, 1995 e KIM *et al.*, 1995).

Desta forma, uma metodologia para integração de esquemas deve identificar e tratar todas as possíveis heterogeneidades que podem ser encontradas entre os esquemas das bases de dados. Para isto é adequado o estabelecimento de uma taxionomia das heterogeneidades presentes entre os esquemas a serem integrados. Esta taxionomia facilita a análise de soluções para conformação e união de esquemas. A ausência desta pode ocasionar falhas do processo de integração.

Na literatura são encontradas diversas abordagens sobre as heterogeneidades (CHATTERJEE & SEGEV, 1991; DREW *et al.*, 1993; CASTELLANOS *et al.*, 1994;

LIM *et al.*, 1994; GOH *et al.*, 1994; RIBEIRO & OLIVEIRA, 1995; KIM *et al.*, 1995; KASHYAP & SHETH, 1996; ZHAO *et al.*, 1996). Entretanto, este trabalho se detém nos conflitos mais significativos no âmbito das Bases de Dados Convencionais e os compara com os diversos trabalhos (Tabela III.1), a saber:

- conflitos semânticos - ocorrem quando as conceituações sobre uma realidade comum são percebidas de formas diferentes, levando as bases de dados a apresentarem divergências semânticas (diferentes significados). Este tipo de heterogeneidade se propaga através dos modelos de dados, das aplicações, dos procedimentos e restrições de integridade. As principais fontes desta incompatibilidade são:
 - ✓ conflitos de denominação - são devidos aos esquemas incorporarem denominações para os objetos representados. Este tipo de conflito pode ser classificado ainda em sinônimos e homônimos. Os sinônimos ocorrem quando entidades ou propriedades semanticamente idênticas são denominadas diferentemente; e os homônimos ocorrem quando entidades e propriedades diferentes semanticamente compartilham a mesma denominação;
 - ✓ conflitos de abstração - são devidos as conceituações a respeito da realidade serem representadas em níveis de abstração diferente, por exemplo, um elemento de algum tipo (classe de objeto, atributo, relacionamento) corresponde a outro tipo em outro esquema;
 - ✓ conflitos de propriedades - ocorrem quando as propriedades que descrevem os objetos não estão completas, ou estão implícitas em outras propriedades;
- conflitos sintáticos - também denominados de conflitos estruturais. Eles ocorrem quando, apesar das conceituações serem as mesmas e adotarem o mesmo modelo de dados, elas possuem diferentes especificações, restrições e domínios, fazendo com que a realidade comum seja representada de forma diferente nas bases de dados a serem integradas. As principais fontes deste conflito são:
 - ✓ conflitos de tipos de dados - ocorre quando as propriedades semanticamente equivalentes apresentam definições de tipos incompatíveis, em diferentes bases de dados;
 - ✓ conflitos de chaves - as chaves são usadas por várias razões, como por exemplo, facilitar o acesso à informação ou otimizar espaço. Eles geralmente não são uniformes entre as bases, uma vez que atendem às necessidades específicas dos usuários;
 - ✓ conflitos de formatos - a mesma propriedade pode usar diferentes formatos nas

bases de dados a serem integradas; e

- ✓ conflito de unidades - ocorre quando são definidas unidades diferentes para a mesma conceituação, devido as necessidades operacionais dos usuários. Isto leva os domínios das propriedades serem diferentes.

Alguns autores tratam ainda das inconsistências entre os dados, na ausência de conflitos a nível extencional. Estes conflitos são devido a (LIM *et al.*, 1994; GOH *et al.*, 1994 e KASHYAP & SHETH, 1996):

- atualizações assíncronas - acontecem quando as propriedades replicadas entre as diferentes bases são atualizadas em diferentes épocas, tornando-as inconsistentes. Isto é realçado se os itens de dados variarem no tempo;
- erros de registros - são erros tipográficos ou variações nos processos de medição, ou digitação; e
- valores *default* - os SGBDs e aplicações podem proporcionar diferentes valores *default* para propriedades semanticamente iguais.

Tabela III.1- Comparação das taxionomia de heterogeneidades encontradas na literatura.

Conflitos	ChSe	RePr	Lim	Goh	RiOI	KiSe	KaSh	Sola
Semânticos		x		x	x	x	x	x
Denominação	x	x	x		x	x	x	x
Abstração			x		x	x	x	x
Propriedades	x				x	x	x	x
Estruturais				x		x	x	x
Tipos	x	x			x	x	x	x
Formatos	x	x			x	x	x	x
Chaves	x	x	x		x	x	x	x
Domínios	X	x			x	x	x	x
Entre valores			x			x	x	
Erros de registros	X	x				x	x	
Valores <i>default</i>	X	x				x		
Atualização	X	x				x		

Onde:

ChSe - CHATTERJEE & SEGEV, 1991

RiOI - RIBEIRO & OLIVEIRA, 1995

RePr - REDDY *et al.*, 1994

KiSe - KIM & SEO, 1991

Lim - LIM *et al.*, 1994

KaSh - KASHYAP & SHETH, 1996

Goh - GOH *et al.*; 1994

Sola - SOLACO *et al.*, 1996

III.5.3. Etapas de uma metodologia genérica para integração

As operações citadas na seção III.5.1 são realizadas através de um metodologia que formaliza as etapas, de modo que elimine a subjetividade e facilite o processo de integração. Alguns autores (MARRS^{html} *et al.*, 1996, QIAN, 1995) generalizam estas operações em etapas para a tradução e integração, enquanto outros autores a subdividem em mais etapas, a saber:

- SOUZA (1986) propõe uma metodologia baseada na sintaxe dos esquemas onde são comparados nome das entidades, chaves e atributos, restrições e combinação de hierarquia de generalização. Ele divide o processo de integração em oito etapas: abstração, representação, integração de visão homogênea, mapeamento do esquema, integração do esquema conceitual abstraído, definição da visão, mapeamento para o esquema externo e integração de dados físicos;
- BATINI *et al.* (1986), por exemplo, investigaram doze metodologias baseadas na sintaxe e as compararam fundamentando-se nas seguintes atividades de integração: pré-integração, comparação de esquemas, conformação de esquemas, união e reestruturação.
- GOTTARD *et al.* (1992) propõem uma metodologia dividida em três fases inter-relacionadas e interativas para a integração, isto é: comparação de esquemas; conformação de esquemas e união de esquema;
- BONJOUR e FALQUET (1994) dividem as etapas em pré-integração, comparação dos elementos dos esquemas, integração e transformação de esquemas;
- HAMMER (1994) propõe uma metodologia com três fases operacionais: localização e identificação de informação não local relevante para uma consulta; determinação do relacionamento semântico entre o esquema local e não local e solução da heterogeneidade semântica e acesso aos objetos compartilhados;
- FRANKHAUSER *et al.* (1995) dividem o processo de integração em identificação de correspondência entre os esquemas heterogêneos e geração de um esquema unificado. Ele utiliza uma função que transforma cada esquema original e suas instâncias no esquema unificado;
- PARENT *et al.* (1996) sugerem três etapas: preparação dos esquemas, investigação de correspondências e integração.

Este trabalho para tornar a metodologia um processo mais compreensível divide-

a em seis etapas apresentadas a seguir:

III.5.3.1. Pré-integração

Esta etapa consiste em estabelecer as regras para o processo de integração. Nesta etapa são observados os esquemas, de forma que se defina uma política de integração quanto à metodologia. São identificados os esquemas a serem integrados.

Nesta pré-análise dos esquemas são identificados as localizações dos esquemas e seus ambientes quanto a plataforma e aplicativos. São identificadas também as prioridades entre esquemas, as dependências implícitas entre eles, a necessidade de transformação entre as representações e estabelecida uma ordem entre os esquemas para aplicar as etapas seguintes. Nesta etapa também pode ser efetuada uma coleta de informação adicional relevante para o processo de integração, por exemplo, enriquecimento semântico ou restrições entre visões.

III.5.3.2. Tradução

Esta etapa tem por objetivo fornecer homogeneidade entre os modelos, minimizando a diferença semântica relativa à heterogeneidade dos modelos de dados usados. Ela consiste em reescrever, a nível intencional, o esquema local expresso no modelo de dados M_1 para um esquema equivalente, no caso do esquema global, ou para esquemas equivalentes, expressos no modelo de dados M_2 . O esquema criado por este processo de tradução é denominado de esquema componente e equivale ao esquema local expresso no MDC (FRANKHAUSER *et al.*, 1995). Para este processo de tradução se faz necessário definir regras de tradução que transformam as representações dos modelos de dados locais para o MDC.

Este processo de tradução é um processo estrutural que normaliza tanto a sintaxe quanto a semântica dos esquemas. Ele proporciona, ao integrador, uma visão geral de todas as informações sobre as bases de dados. Isso facilita ao integrador identificar a presença de conflitos entre os diferentes esquemas e determinar que representação tem precedência sobre as outras.

Esta etapa pode ser manual ou automática. Ela deve ser de responsabilidade do administrador das bases de dados locais, uma vez que estes conhecem os seus esquemas e podem melhor expressar a semântica das suas bases.

III.5.3.3. Análise comparativa dos esquemas

Esta etapa determina quais as partes dos esquemas traduzidos representam as mesmas conceituações do mundo real, a fim de estabelecer uma correspondência entre eles. Para isto é analisada a descrição intencional dos esquemas exportáveis das bases de dados, de forma que se identifique a existência de correspondências e de conflitos semânticos e estruturais entre as representações dos mesmos objetos, em diferentes esquemas. Esta análise envolve as denominações de entidades/classes, suas propriedades e relacionamentos com conhecimento adicional sobre os metadados. Ela resulta em diferentes associações entre os esquemas, conforme listadas abaixo (CHORAFAS & STEINMANN, 1993):

- uma exata combinação das definições dos objetos, de esquemas diferentes, sem nenhuma perda de informação;
- embora não haja uma combinação exata da semântica, entre as definições dos objetos nos esquemas, pode ser usado um função sobre uma ou mais propriedades de uma ou mais entidades/classes; ou pode ser usado um construtor de modelo de dados para associá-las; e
- não há combinação entre as diferentes entidades/classes;

Estas situações estabelecem um grau de proximidade entre os esquemas de forma que determine quão próximos são os conceitos representados pelos diferentes esquemas, a saber:

- identidade semântica - ocorre quando dois objetos representam os mesmos conceitos, utilizam os mesmos construtores e não há incoerência entre as representações;
- similaridade semântica - ocorre quando dois objetos representam os mesmos conceitos, mas eles são representados nos esquemas de forma diferente. Este grau pode ser classificados em:
 - ✓ equivalência semântica - neste caso, os objetos representam os mesmos conceitos mas utilizam diferentes construtores para modelá-los.
 - ✓ compatibilidade semântica - neste caso, os objetos representam os mesmos conceitos que não são nem idênticos nem equivalentes, mas também não são contraditórios; e
- heterogeneidade semântica - ocorre quando as definições não são relacionadas por definição, isto é, são incoerentes com respeito um ao outro e que não existe contexto associado com os objetos, tal que eles possuam o mesmo papel.

Na literatura (BATINI *et al.*, 1986; SHETH, 1991; HAMMER, 1994) são encontrados diferentes taxionomias para este grau. BATINI *et al.* (1986) e HAMMER (1994) apresentam estes conceitos nas suas taxionomias, entretanto BATINI *et al.* especializam o conceito de equivalência comportamental, transformacional e mapeamento. A equivalência comportamental ocorre quando as instâncias nos dois esquemas respondem a uma mesma pergunta. A equivalência transformacional ocorre quando é necessário aplicar um conjunto de transformações para preservar a equivalência. A equivalência de mapeamento ocorre quando há uma correspondência 1:1 entre as instâncias. SHETH (1991) levam em consideração contexto e papel, estabelecendo além destes graus, os seguintes:

- relacionamento semântico - ocorre quando dois objetos são semanticamente relacionados e existe um mapeamento parcial de um valor para muitos, ou abstrações de generalização, agregação entre os domínios dos objetos. Esta expressão denota que há um relacionamento entre dois objetos do mundo real, que entretanto não são exatamente os mesmos;
- pertinência semântica - ocorre quando dois objetos podem ser relacionados através de alguma abstração, em algum contexto. Este conceito é dependente do contexto, isto é eles podem ser semanticamente relevantes em um contexto, mas não em outro;
- semelhança semântica - ocorre quando dois objetos não são relacionados um ao outro, por qualquer abstração, mas eles não podem ser associados a um contexto em que eles têm o mesmo papel e cujas definições de contexto são coerentes com respeito um ao outro. Este conceito refere-se ao relacionamento entre um objeto e o contexto semântico a qual ele pertence;
- heterogeneidade semântica - ocorre quando dois objetos que representam a mesma entidade do mundo real são representados diferentemente. Isto ocorre por causa da discordância sobre o significado, interpretação, ou uso pretendido do mesmo;
- discrepância semântica - ocorre quando dois objetos que representam o mesmo fenômeno do mundo real ou a mesma definição têm informações inconsistentes ao nível extencional.

Em virtude dos esquemas serem projetados individualmente é necessário identificar este grau de proximidade. FRANKHAUSER *et al.* (1995) propõem que sejam especificados construtores do MDC para representar os objetos do mundo real e estabelecer o relacionamento entre os objetos. Este relacionamento é estabelecido pela equivalência de conceitos, termos e definições de um domínio em outro, ou pela

concordância de um conjunto compartilhado de termos e definições utilizadas em ambos os domínios. SOUZA (1986) propõe aplicar funções que determinam peso para os elementos, de acordo com a atividade de comparação, de modo a identificar a proximidade semântica entre os esquemas. CHATTERJEE & SEGEV (1991) propõem estimar um vetor de valores probabilístico que estima a similaridade encontrada entre tuplas comparáveis dos esquemas. SOLACO *et al.* (1995) propõem aplicar penalizações no grau de proximidade semântica entre esquemas, de acordo com o relaxamento de abstrações de agregação e especialização encontrados na etapa de análise. YU *et al.* (1991) propõem que esta equivalência seja estabelecida através de conceitos comuns e técnicas baseadas no processamento de texto, para mapear as denominações para um conjunto de conceitos comuns e, ao mesmo tempo, encontrar a similaridade entre pares de denominações e arranjá-las em ordem descendente de similaridade. KIM (1991) propõe usar um catálogo para estabelecer a correspondência entre as denominações. Entretanto, esta técnica é viável quando as bases de dados são pequenas ou autonomia não é crucial. FRANKHAUSER *et al.* (1992) propõem determinar a similaridade das classes combinando a teoria *fuzzy* e conhecimento da terminologia usada no esquema. SCHMITT & SAAKE (1996) propõem uma análise intencional dos esquemas através de uma matriz composta de linhas e colunas, onde são dispostas as classes e suas propriedades, respectivamente. Através da comparação da matriz formada são detectados os atributos comuns e os não comuns. O esquema integrado é derivado da detecção de retângulos verticais com classes e propriedades comuns. FRANKHAUSER *et al.* (1995) e RIBEIRO & OLIVEIRA (1996) propõem conhecimento sobre o código das aplicações locais, visto que muitas vezes a semântica estrutural e, principalmente, a semântica comportamental dos esquemas das bases estão embutidas nele. SALTOR *et al.*, 1996 propõem que nesta etapa as similaridades e os relacionamentos encontrados entre os esquemas sejam confirmados com os administradores das bases de dados.

III.5.3.4. Conformação ou resolução de conflitos

Uma vez identificada a existência de entidades/classes de objetos em dois esquemas com algum grau de proximidade, é conduzida à conformação dos esquemas. Esta etapa tem por objetivo solucionar os conflitos de forma a tornar os esquemas compatíveis para que posteriormente os esquemas possam ser integrados. Ela prescreve operações padronizadas para o estabelecimento de soluções para os conflitos entre os

esquemas levando em consideração os metadados, a semântica e a sintaxe destes.

Esta etapa requer do projetista compreensão do relacionamento semântico, entre os conceitos envolvidos, para estabelecer operações de mapeamentos, também denominadas de isomorfismo. O isomorfismo estabelece correlações entre termos e definições de um domínio em outro, através da correspondência biunívoca entre elementos, sejam objetos, propriedades ou operações, de dois ou mais esquemas (KASHYAP & SHETH, 1996).

Dentre as técnicas mais utilizadas para o isomorfismo tem-se a renomeação, as abstrações proporcionadas pelo modelo de dados usado para integrar e a homogeneização das propriedades.

A renomeação é aplicada para resolver conflitos de denominação de entidades e suas propriedades. A técnica consiste em padronizar as denominações, estabelecendo um conjunto de conceitos comuns ou criar novos conceitos, se necessário. A seguir, deve-se criar artificios, de modo a evitar que as denominações nas bases não sejam alteradas, uma vez que isto implicaria na necessidade de uma recompilação das aplicações existentes (SALTOR *et al.*, 1996).

O emprego desta técnica introduz uma certa dificuldade para encontrar uma denominação apropriada para representar todas as semânticas dos esquemas. Por um lado, arbitrar uma denominação não é justo, visto que uma denominação é mais apropriada para uma área e para outra não. Uma denominação não apropriada leva a perda da semântica verdadeira. Além do que, os usuários podem não concordar em encontrar os elementos de seus esquemas na federação com novos nomes ou com nomes modificados. Por outro lado, os usuários da federação poderiam apreciar o acesso aos nomes locais, de modo a compreender os pontos de vistas sobre o mesmo item do mundo real. Assim, tem sido proposta três outras soluções, analisadas em HAMMER (1994). A primeira, denominação completa (*fullnaming*), estabelece que o nome é formado pelo prefixo seguido do nome dos elementos originais. O prefixo pode ser o nome do esquema, do local ou do administrador local. Esta técnica permite identificar qualquer item de informação e evita conflitos de homônimos. A segunda técnica proposta, denominada de multi denominação (*multinaming*), mantém na federação todos os nomes usados pelos esquemas componentes, de forma que possa ser compreendida por qualquer usuário da federação e pelo administrador. Isto evita perda de informação, além do que reforça a precisão da semântica. Nesta técnica, quando dois objetos são unidos, o elemento resultante fornece ambos os nomes e todos estes fornecem nomes

que podem ser usados como referência para seus elementos. Isto torna o esquema global mais flexível. A terceira técnica consiste em combinar as duas técnicas de renomeação, em conjunto.

As técnicas de conformação, empregando as abstrações proporcionadas pelo modelo de dados, consistem em criar entidades ou classes virtuais que apontam para as bases de dados locais. Entretanto, estas abstrações dependem das propriedades desejadas para o esquema integrado e dos relacionamentos oferecidos pelo modelo de dados. Por exemplo, a classificação permite definir classes; a abstração de generalização permite que as classes com alguma proximidade semântica sejam integradas, através de hierarquias de generalizações; a abstração de agregação permite tratar a composição de partes e a abstração de associação permite associar propriedades. Desta forma, esta técnica depende da expressividade e do relativismo semântico do modelo selecionado para representar o esquema integrado.

Na realidade estas abstrações são empregadas com diversas finalidades na resolução dos conflitos, a saber: remover redundância de propriedades, tratar a ausência de propriedades ou propriedades implícitas, impor restrição de integridade além de remover os conflitos de abstrações e conflitos de cardinalidade entre entidades e classes.

A homogeneização das propriedades é aplicada quando as propriedades necessitam de transformações para serem integradas, a fim de resolver conflitos de definição de tipos de dados incompatíveis; formatos distintos; diferenças nas unidades e nos metadados; uso de códigos e valores *default*; e conversão de domínios ou propriedades derivadas. E pode ser através da:

- criação de uma LUT (*Look Up Table*) que define o isomorfismo;
- definição de operadores apropriados na linguagem de consulta que convertem formatos, valores *default* e valores numéricos de uma unidade para outra. Nesta última, deve-se ressaltar que pode haver perda de informação, devido às aproximações numéricas e limitações na representação de máquinas para os valores reais; e
- definição de uma transformação entre os domínios das propriedades equivalentes semanticamente.

A resolução automática de conflito geralmente não é praticada, pois requer interações entre os projetistas e os administradores locais para chegar ao esquema integrado. Ainda assim, algumas vezes os conflitos podem não ser resolvidos devido alguma inconsistência. Neste caso, o conflito é reportado para os usuários que guiam o

projetista para uma solução.

Esta etapa desenvolve os esquemas exportáveis para cada base de dados local com os respectivos mapeamentos entre os esquemas.

III.5.3.5. Integração

Esta etapa consiste em um processo operacional, no qual os esquemas exportáveis são reestruturados e unidos em um ou mais esquemas federados. Ela requer que as entidades/classes sejam compatíveis, através da solução dos conflitos representados e tratados no esquema exportável.

Nesta etapa o(s) esquema(s) federado(s) é (são) estabelecido(s) através da superposição de conceitos comuns que são reunidos em uma única entidade ou classe virtual. Esta entidade/classe virtual une todas as instâncias materializadas em cada base de dados componente (KIM *et al.*, 1995).

Para a criação desta entidade/classe virtual são empregados as abstrações do modelo de dados utilizado pelo esquema federado para permitir classificar, generalizar, especializar, agregar ou associar conceitos comuns.

O esquema federado representa todos os dados disponíveis na federação e os respectivos mapeamentos estruturais entre ele e os esquemas locais. Estes mapeamentos consistem em funções que implementam operações de transformação e recuperação dos dados locais. As operações de transformação são responsáveis por tornar compatíveis os objetos do esquema global com os esquemas locais, de forma a serem acessados adequadamente quando da execução de uma consulta. As operações de acesso são responsáveis pela recuperação dos dados nas bases de dados locais. Estas operações transformam uma consulta global em sub-consultas equivalentes para cada bases de dados local.

III.5.3.6. Avaliação e reestruturação

A etapa final do processo de integração consiste em avaliar o esquema integrado e reestruturar, a fim de alcançar qualidades desejáveis. Esta avaliação tem por objetivo assegurar que o esquema resultante seja facilmente compreendido por qualquer usuário; que não houve perda de informação, além de remover qualquer redundância ou complexidade desnecessária. Desta forma, ÖSZU & VALDURIEZ, (1990) e SCHMITT & SAAKE (1996) propõem aplicar os seguintes critérios sobre o esquema integrado:

- **completude** - Este critério avalia se o esquema integrado contém todos os conceitos presentes nos esquemas componentes, corretamente. O esquema integrado deve ser a representação da união dos domínios das aplicações, associadas com os esquemas. Para isto é utilizado os conceitos de generalização e especialização, de forma a verificar a sub-definição de uma entidade em outra;
- **exatidão** - mantém cada entidade/classe local no esquema integrado com uma política que adiciona todos os grupos de entrada e cria grupos de subclasses com o mesmo nome da classe;
- **minimalidade** - este critério avalia se os mesmos conceitos, apresentados em mais de um esquema componente, são representados somente uma vez no esquema integrado, isto é, se não há redundância; e
- **compreensão** - é um conceito subjetivo, mas consiste em avaliar se o esquema integrado é facilmente compreendido pelo projetista e pelo usuário. Este critério leva em conta que, dentre as possíveis representações do resultado de integração, uma é mais facilmente compreendida pelos usuários do que as outras.

III.5.4. Classificação das metodologias para integração

Os primeiros processos de integração foram inteiramente **manual**. Um operador era responsável por dimensionar toda a complexidade dos esquemas envolvidos e produzir uma imagem global das estrutura dos dados envolvidos. Todavia, com a automatização de processos algumas ferramentas têm sido propostas para cobrir uma fase específica ou todo o espectro do processo de integração, passando os processo a serem classificados em **semi-automáticos** e **automáticos**.

Os processos semi-automáticos apresentam uma interface gráfica com ferramentas que ajudam os administradores a identificar alguma proximidade semântica entre os esquemas. Estas ferramentas assistem ao administrador na tarefa de identificar semelhança entre os esquemas. Elas agem como um investigador, tentando encontrar que elementos de dois ou mais esquemas apresentam similaridades (de nomes, domínios, relacionamentos). Em geral, estas ferramentas possuem mecanismos para definir os esquemas componentes e uma linguagem de manipulação com operadores especiais para realizar as comparações e capturar a estrutura da base de dados em questão e o relacionamento entre os esquemas. Por exemplo, SOUZA (1986) propõe uma ferramenta para integração de esquema conceitual, denominada *Schema*

Integration System (SIS) que emprega métodos para medir a similaridade através da comparação sintática dos esquemas.

Nos processos automáticos, os esquemas são convertidos para um modelo de dados formal que proporciona um modelo de classificação, fornecendo regras formais para determinar o relacionamento entre os esquemas. Este processo é recente, e se utiliza do paradigma orientado a objeto e de técnicas de inteligência artificial para identificar automaticamente uma boa percentagem dos relacionamentos. Por exemplo, WEISHAR & KERSCHBERG (1991) propõem técnicas de inteligência artificial para construir um modelo de domínio no qual são introduzidos dados e conhecimento sobre as bases de dados; COLLET *et al.* (1991) propõem uma ferramenta que se utiliza de uma base de conhecimentos que permite os esquema locais serem mapeados após serem traduzidos para o modelo de dados usado pelo Projeto Carnot, denominado Cyc; NICOLLE & CULLOT (1996) propõem uma metodologia que emprega um metamodelo chamado TIME, uma máquina de inferência e uma base de conhecimento, contendo um conjunto de regras de tradução de esquema dos modelos locais para este metamodelo.

Os processos semi-automáticos e automáticos são classificados por SPACCAPITERA & PARENT (1991) e DUPONT & SPACCAPIETRA (1996) em **metodologia procedural** e **metodologia declarativa**. A metodologia procedural é orientada a instrução. Ela manipula diretamente os esquemas cadastrados no mesmo modelo de dados. O administrador local escreve um conjunto de instruções, sendo que a ferramenta de integração tem a finalidade de verificar apenas a integridade das sintaxe das instruções. Esta metodologia deixa a cargo do usuário resolver os conflitos. A metodologia declarativa permite que os administradores locais declarem a existência de correspondência e incompatibilidades entre os esquemas, sem dizer como a integração será trabalhada. Cada declaração afirma um relacionamento entre os elementos de dois esquemas heterogêneos. O conhecimento expresso por cada declaração ou pela combinação destas é usado pelo sistema, tanto para gerar o mapeamento entre os elementos correspondentes quanto para unir estes elementos em um esquema integrado. Este mecanismo é complementado por uma ferramenta de investigação semântica e sintática para automaticamente detectar similaridades entre os esquemas.

A metodologia procedural é simples de implementar e conveniente para integração de esquemas, mas a tendência tem sido em favor da declarativa, pois ela é mais flexível ao fornecer serviços para detectar a existência de correspondência entre os

itens que podem ser unidos para formar o esquema integrado.

Na realidade este processo de integração não é completamente automático, devido a necessidade da semântica dos dados ser especificada, os modelos de dados correntes não capturam adequadamente a informação semântica das múltiplas visões e interpretações do mundo real, além destas variarem sobre o tempo. Desta forma, as metodologias para a integração de esquemas requer a presença dos administradores locais que possuam conhecimentos a respeito dos esquemas das bases de dados locais, de modo que ele possa especificar a semântica correspondente entre os esquemas componentes, e validar esta correspondência (confirmar ou negar semelhanças) com os administradores locais e estabelecer soluções para os conflitos (SPACCAPIETRA & PARENT, 1991).

Uma outra classificação encontrada em SOUZA (1986) e ÖSZU & VALDURIEZ (1991) distinguem as metodologias para integração de esquemas, de acordo com o número de esquemas envolvidos no processo de integração em **metodologias binárias ou n-ária** (Figura III.5).

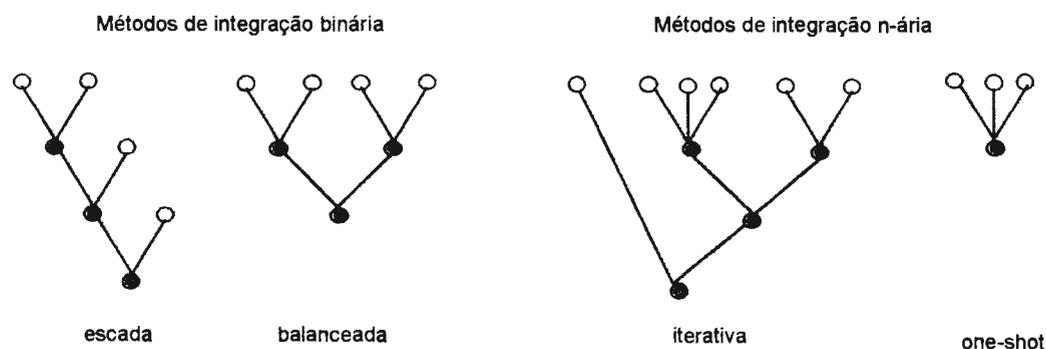


Figura III.5- Taxionomia das metodologias de integração (ÖSZU & VALDURIEZ, 1990)¹

As metodologias de integração binária envolvem a manipulação de dois esquemas de cada vez. Ela pode ser desenvolvida de duas maneiras: os esquemas intermediários são criados e a seguir integrados com um subsequente esquema intermediário, ou de modo puramente binário, onde cada esquema é integrado a um outro, criando um esquema intermediário para ser integrados com outros esquemas intermediários. Esta estratégia é denominada em BATINI (1986) de estratégia escada (*ladder*), quando um

¹ Cada nó corresponde a um esquema componente. Os nós cheios correspondem a nós intermediários resultantes da integração.

novo esquema componente e criado como um resultado intermediário. Esta estratégia é balanceada quando os esquemas são divididos em pares e são integradas de modo simétrico.

As metodologias para integração n-ária integra mais de dois esquemas por vez. Ela também pode ser desenvolvida de duas formas: integrar todos os esquemas de uma só vez, ou de acordo com a preferência do integrador, selecionar os esquemas mais semelhantes e integrá-los por um processo iterativo e interativo. A primeira forma não prioriza uma forma de integração, oferecendo a vantagem de tornar disponível a informação completa sobre todas as bases de dados a serem integradas, contudo aumenta a complexidade do processo. Por outro lado, a segunda forma oferece maior flexibilidade na seleção dos esquemas a serem integrados.

As metodologias binárias são um caso específico das n-árias e têm a vantagem de diminuir a complexidade do processo de integração nas etapas de comparação e integração de esquemas, podendo ser automatizada, uma vez que um número menor de esquemas permite um melhor gerenciamento do processo. A sua desvantagem é devido ao aumento do número de operações de integração e a necessidade de uma análise final para adicionar propriedades perdidas.

SHETH (1995) considera a compreensão da semântica importante para o processo de integração. Desta forma, ele apresenta uma outra classificação das metodologias levando em consideração a exposição da semântica, a saber:

- metodologia com pouca semântica - são metodologias que utilizam informações simples, clara e determinista, envolvendo um formalismo e serviços de mediação simples;
- metodologia com fraca semântica - são metodologias que capturam a semântica dos esquemas baseada na estrutura, sintática e informação extencional do esquema; e
- metodologia baseada na semântica - são metodologias que tratam a semântica através de questões da cognição, percepção, ou interpretação humana.

Como pode ser observado, as diferentes formas de integrar os esquemas resultam em diferentes expectativas da metodologia, técnicas e ferramentas.

III.6. Arquitetura para múltiplas bases de dados heterogêneas

Atualmente, são encontradas na literatura diversas propostas de arquiteturas para um ambiente MBD. Alguns autores, para melhor compreender as arquiteturas para o

ambiente MBD, propõem que elas sejam classificadas, porém cada autor apresenta uma perspectiva diferente. Por exemplo, SHETH & LARSON (1991) propõem uma taxionomia para estes sistemas, de acordo com a cooperação e o número de esquemas disponíveis à federação, ou seja:

- **Sistemas não federado** - nesta arquitetura as bases de dados não são autônomas. Existe um único esquema e um único nível de gerenciamento, no qual todas as operações são realizadas uniformemente. Não há distinção entre os usuários locais e globais;
- **Sistema Federado** - Nesta arquitetura as bases de dados são autônomas e cooperam, participando de uma federação para permitir o compartilhamento de seus dados. De acordo com o grau de cooperação, estas arquiteturas podem ser classificadas em:
 - ✓ **federada com fraco acoplamento** - nesta arquitetura os administradores das bases e dados são responsáveis por criar e manter a federação, tornando disponível parte/todo o seu esquema para a federação. Suporta sempre múltiplos esquemas federados;
 - ✓ **federada com forte acoplamento** - nesta arquitetura as bases de dados também disponibilizam parte/todo o seu esquema, entretanto estes são integrados em um único esquema controlado por um administrador central, que soluciona os problemas de integração dos dados. Pode ter um ou mais esquemas federados podendo ser classificado em federação única ou múltiplas federações.

ÖSZU & VALDURIEZ (1990) distinguem três tipos de arquitetura, de acordo com os níveis de autonomia:

- **alta integração** - neste nível, os dados são logicamente centralizados em um esquema, que integra várias base de dados de usuários que desejam compartilhar informação. O gerenciamento dos dados é implementado por um sistema de gerenciamento que possui o controle para realizar processamento através dos SGBDs locais. Este sistema de gerenciamento não opera independente dos SGBDs locais, embora ele possa ter funcionalidade para fazê-lo;
- **semi-autônoma** - neste nível, os SGBDs operam independentemente, mas foram reprojatados para participarem de uma federação, a fim de compartilhar seus dados locais. Cada SGBD local possui autonomia para determinar que parte da sua base de dados estará acessível. Contudo, eles não são completamente autônomos, uma vez que precisam ser modificados para capacitar a troca de informações com os outros SGBDs locais; e

- completamente autônomo - neste nível os SGBDs locais são completamente *stand-alone*, desconhecendo a existência de outros SGBDs ou de como se comunicar com eles. Os ambientes com estas características possuem dificuldades no processamento das transações de acesso às múltiplas bases de dados, devido a ausência de um controle global para gerenciar a execução das transações através dos SGBDs locais.

As taxionomias propostas por DUPONT & SPACCAPIETRA (1996), GOH *et al.* (1994) e QIAN (1993) levam em consideração o grau de acoplamento entre as bases de dados, classificando-as em fortemente e fracamente acopladas. Este grau determina quanto centralizada e coordenada é o processamento da consulta. Ele é medido pela razão entre a quantidade de dados trocados e a quantidade de processamento local realizado para executar uma consulta.

BRIGH *et al.* (1994) distinguem as arquiteturas, de acordo com o esquema ao nível global, a saber: mono federação, com um único esquema (global), ou uma multi-federação, com múltiplas superposição de esquemas federados. KÜHN & LUDWIG, (1988) classificam a aproximação do esquema global como aproximação estática e as arquiteturas com múltiplos esquemas e que estendem uma linguagem de consulta como aproximação dinâmica.

De modo a ilustrar estas arquiteturas, este trabalho se detém, nas duas seções a seguir, nos dois extremos das arquiteturas MBD que proporcionam uma federação, do ponto de vista de integração existente entre os esquemas componentes: uma arquitetura de federação rígida e uma arquitetura de federação flexível. A arquitetura de federação rígida utiliza um esquema global obtido previamente, através do processo de integração dos esquemas componentes da federação, ao passo que a arquitetura da federação flexível disponibiliza várias esquemas exportações, sobre os quais os usuários realizam suas consultas. Na primeira arquitetura existe uma autoridade central denominada de administrador da federação. Este administrador é responsável pelo projeto e manutenção do processo de integração. Na segunda arquitetura não há uma autoridade central e sim os administradores locais que criam visões locais e utilizam os mecanismos oferecidos pela linguagem para integrar e facilitar o acesso.

III.6.1. Federação rígida: arquitetura do esquema global

Esta arquitetura constitui em uma federação rígida, uma vez que os esquemas das bases de dados componentes da federação são integrados em um esquema global

que serve como base para o compartilhamento e troca de dados. Ela é também denominada de integração total, base de dados virtual ou fortemente acoplada (DUPONT & SPACCAPIETRA, 1996; BOUGUETTAYA, 1995). Nesta arquitetura os sistemas locais são integrados às aplicações globais que encapsulam as heterogeneidades, preservando suas autonomias. Ela oferece um ambiente que suporta o gerenciamento de atividades distribuídas, no qual um único método de acesso proporciona ao usuário a ilusão de que todos os componentes estão sob um mesmo sistema.

O esquema global funciona como uma interface uniforme entre o usuário e as bases de dados que compõem a federação. Ele é construído a priori e mantido pelo administrador da federação através do processo de integração de esquemas. Este processo requer a troca de informações com os administradores locais para chegarem a um consenso sobre o esquema global. A informação disponível no esquema global é compatibilizada e padronizada de forma a assegurar a consistência dos esquemas locais. Sobre este esquema o usuário submete consultas que são processadas e distribuídas para serem executadas nas bases de dados que participam da federação.

Esta arquitetura, ilustrada na Figura III.6, é composta basicamente por uma interface global que apresenta um conjunto de ferramentas, um gerenciador de consultas, um gerenciador de transação, um dicionário de dados e um tradutor de consultas locais.

A interface global permite que um usuário submeta uma consulta usando uma definição de dados e uma linguagem de manipulação comum. O gerenciador de consulta e o gerenciador de transação transformam a consulta em um plano de execução que acesse corretamente todas as bases de dados locais. O dicionário de dados oferece uma forma padronizada de armazenar os metadados do ambiente MBD e de auxiliar o sistema a localizar os dados no processamento e otimização da consulta global. Ele é logicamente centralizado, mas pode estar fisicamente centralizado ou distribuído sobre a rede. O tradutor de consulta local transforma subconsulta global em uma consulta local e a submete ao SGBD local (BREITBART, 1990; CHATTERJEE & SEGEV, 1991; KIM *et al.*, 1995, MENG & YU, 1995 e SCHEUERMANN *et al.*, 1990).

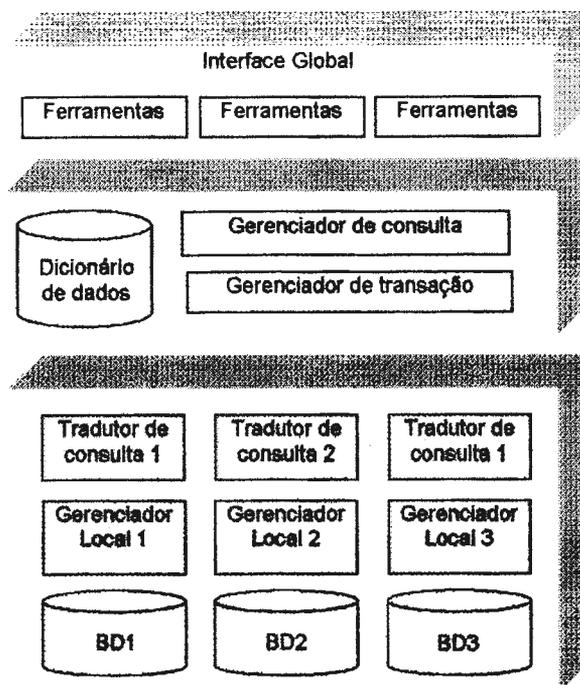


Figura III.6- Arquitetura genérica da federação rígida.

Algumas arquiteturas oferecem outros componentes. Por exemplo, DUPONT & SPACCAPIETRA (1996) oferecem ferramentas de administração que resolve questões complexas de negociação entre esquemas e tradutores sintáticos e unificação semântica, e mecanismos de integração dinâmicos utilizando regras; utiliza um base de conhecimento onde são armazenadas as informações necessárias ao comportamento cooperativo do sistema e emprega um mecanismo de inferência para gerar consultas alternativas para a obtenção de respostas aproximadas, no caso das consultas sem sucesso.

Em geral esta arquitetura é fechada, reservada a um conjunto de usuários com necessidades previamente definidas, não permitindo adicionar ou remover novas bases de dados componentes à federação. Contudo, atualmente as proposta destas arquiteturas tendem tratar a evolução do esquema global para inserir novas bases de dados.

A vantagem desta abordagem está na transparência fornecida pelo acesso uniforme às bases de dados. O esquema global promove uma compreensão única e correta dos dados, familiariza o usuário com a nomenclatura utilizada pela federação, simplifica a tarefa de construir uma consulta e esconde dos usuários, do ambiente MBD, as diferenças entre os esquemas, as linguagens de consulta, os formatos de arquivos, a complexidade e distribuição dos dados.

A desvantagem é que o processo de integração de esquemas é uma tarefa difícil de ser realizada, e quando há alteração de um esquema local, ou um novo esquema é

adicionado à federação, se faz necessário que este processo de integração seja realizado novamente. Outra desvantagem é não permitir ao usuário selecionar a qual esquema ele quer consultar e como transformar os dados.

São exemplos destas arquiteturas os sistemas: VIP (*Vienna Integrated Prolog*) (KÜHN & LUDWIG, 1988); HODFA (*HOmogenized Database Federation Architecture*) (KARLPALEM *et al.*, 1995); HEROS desenvolvido pelo Grupo de Pesquisas em Banco de Dados da PUC-RIO (UCHÔA, 1995); PEGASUS (AHMED *et al.*, 1991; SHAN *et al.*, 1995); ADDS (*Amoco Distributed Database System*) (BREITBART & REYES, 1995) e MERMAID/InterViso (TEMPLETON *et al.*, 1995).

III.6.2. Federação flexível: arquitetura do mediador

Esta arquitetura também denominada de mediadores (WIERDHOLDD, 1992), base de dados cooperativas ou fraco acoplamento (SHETH & LARSON, 1990), é caracterizada pela falta de um esquema global e um administrador central.

As bases de dados, a fim de participar da federação, dispõem todo/parte de seus dados em esquemas exportáveis, apresentados em um MDC, a outros usuários da federação. Os esquemas exportáveis das bases de dados locais são integrados no contexto de uma seção para construir um esquema federado, de forma a atender às necessidades dos usuários em realizar uma consulta na federação.

Cada usuário ou administrador local tem a tarefa de localizar os dados desejados nos diversos esquemas exportáveis, especificar a forma de resolver os conflitos e construir um esquema federado local. Para isto os usuários ou administradores locais devem conhecer suas localizações, compreender semântica dos esquemas e as diferenças de representação dos dados.

Não existe controle imposto à federação. Cada usuário ou administrador é responsável pela criação e gerenciamento do seu próprio esquema federado local. Este esquema descreve a porção dos dados distribuídos que é do seu interesse, tornando o processo de integração dinâmico e personalizado (DUPONT & SPACCAPIETRA, 1996). Geralmente, neste esquema federado os dados são reorganizados de forma a atender o interesse dos usuários da federação.

Esta arquitetura suporta várias semânticas, desde que diferentes usuários possam importar ou integrar diferentemente os esquemas de exportação e manter diferentes mapeamentos entre os seus esquemas federados e os esquemas de exportação. Os

usuários usam livremente sua própria compreensão dos sistemas de BD componentes para projetar um esquema federado e para especificar consultas.

Nesta arquitetura, a metodologia de integração de esquemas se encerra na etapa de conformação deixando a cargo do usuário ou administrador local unir e integrar as bases de dados, de acordo com a sua necessidade. A integração é realizada no contexto de uma consulta. Desta forma esta arquitetura é uma multi-federação com múltiplas superposições entre os esquemas federados. Sobre cada esquema federado local pode ainda ser derivadas visões ou esquemas externos para atender as necessidades locais, construindo uma arquitetura em cinco níveis de esquemas.

Esta arquitetura, ilustrada na Figura III.7, consiste em aplicações do usuário, mediadores e tradutores para os sistemas locais. As aplicações do usuário se encontram ao nível de interface com o usuário. Nesta, o usuário observa a disponibilidade dos dados nos esquemas exportáveis e consulta um dicionário de dados que descreve a distribuição, as estruturas dos dados e as técnicas para integração que mapeam as diversas representações locais. Os mediadores permitem aos usuários manipularem os esquemas exportáveis com operadores especiais para selecionar os dados nas bases de dados locais, capturar suas estruturas e determinar o relacionamento que os unirá. Os tradutores oferecem mecanismos para o acesso aos esquemas das bases de dados locais através de aplicações que residem nos sistemas locais.

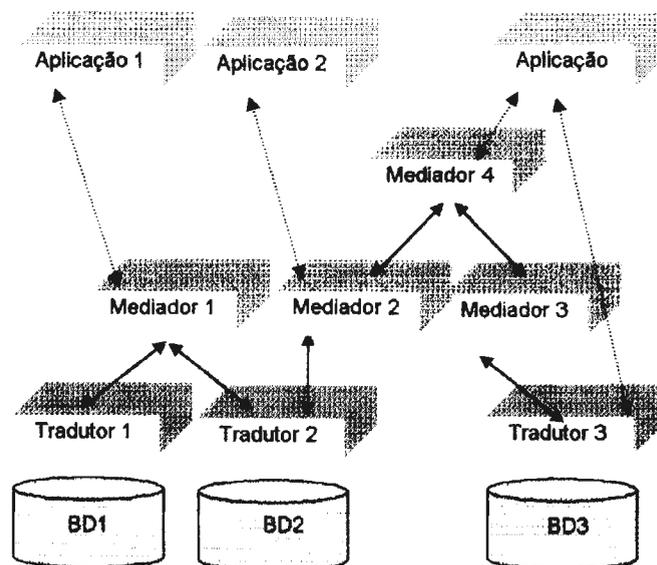


Figura III.7- Arquitetura genérica para ambiente MBD.

A vantagem apresentada por esta abordagem é que não é necessário um administrador central antecipar as necessidades de integração dos usuários, uma vez que

estes conseguem ter acesso aos dados desejados quando necessário. Além disto esta abordagem é mais flexível, permitindo tratar evolução dos esquemas locais, bem como a adição ou remoção de novos esquemas exportáveis, sem causar impactos aos demais componentes (DUPONT & SPACCAPIETRA, 1996).

A desvantagem desta abordagem é que não fornece suporte adequado a compreensão semântica dos esquemas locais, a identificação e resolução dos conflitos são delegadas ao usuário ou administrador local, podendo haver interpretações incorretas. Outra desvantagem é quanto a sincronização global e a eficiência que não são tão fortes como no caso dos sistemas fortemente acoplados.

São exemplos desta arquitetura os sistemas: METU (*Metu Interoperable Database System - MIND*) (DOGAC *et al.*, 1995), TSIMMIS (GARCIA *et al.*, 1995); HERMES (*Heterogeneous Reasoning and Mediator System*) (SUBRAHMANIAN *et al.*, 1995); VIEW System, VODAK (VODAK^{html}, 1996) e HIMPAR (PIRES, 1997).

III.6.3. Questões relacionadas ao desenvolvimento de uma arquitetura MBD

No desenvolvimento da arquitetura MBD existem duas questões que têm sido objeto de estudo, a saber: problemas de representação de dados das bases de dados e problemas de processamento de consulta. Os primeiros têm sido tratados através de metodologias para integração de esquemas. Os segundos são decorrentes de cada base de dados local ser gerenciada por um SGBD autônomo, que possui seu próprio processador de consulta. Isto faz com que a execução de uma consulta, neste ambiente, possua características distintas dos ambientes que tratam apenas de dados homogêneos e distribuídos ao longo de uma rede de computação, a saber (KIM *et al.*, 1995; MENG & YU, 1995; SILBERSCHATZ *et al.*, 1991):

- a forma de processamento de cada SGBD local é diferente, o que impede um tratamento uniforme das consultas;
- custo de processamento das consultas é diferente nos diversos SGBDs, aumentando a complexidade do cálculo do custo total da consulta; e
- as diferenças na capacidade de otimização de cada SGBD local.

III.6.3.1. Processamento da consulta no ambiente MBD

A estratégia para execução da consulta requer um fluxo de controle em níveis, tal como nos SGBDDs, que gerencia desde a especificação da consulta à integração dos resultados parciais das consultas às bases de dados heterogêneas (Figura III.8).

O primeiro nível consiste no ambiente MBD sobre um controle local de um dos sistemas componentes que recebe uma consulta global. Esta consulta é expressa em uma linguagem global, com denominações globais sobre um esquema, cujas divergências entre as bases de dados são harmonizadas.

O segundo nível consiste em um processador de consultas global do ambiente MBD. O processador decompõe a consulta global em um conjunto de subconsultas, também expressas na linguagem de consulta global. Esta decomposição é realizada de acordo com a informação sobre a distribuição dos dados através dos vários locais, armazenada em um catálogo que contém a localização dos dados.

Estas subconsultas globais podem ser executadas de diversas maneiras e cada uma delas pode produzir execuções com diferentes custos, causando um impacto significativo no custo total da execução da consulta global. Assim, o controle passa a um otimizador de consultas global no nível do ambiente MBD, que será o responsável pela otimização global da consulta.

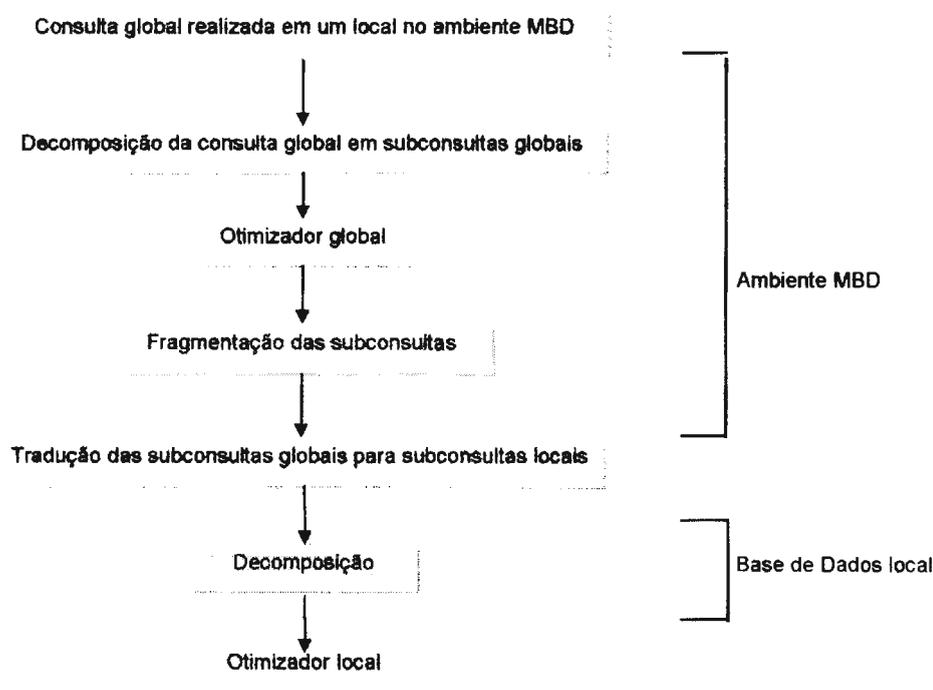


Figura III.8- Processamento de consulta em ambiente MBD (ÖSZU & VALDURIEZ, 1990).

Esse otimizador tem por objetivo maximizar paralelismo e minimizar a transferência de dados na rede. Para tal, ele gera planos potenciais para a execução da consulta global a partir da consulta fragmentada, levando em conta a combinação de técnicas de otimização heurística e sistemática. A seguir são estimados seus custos e selecionado o menor.

Após traçar o plano de execução as subconsultas são enviadas para as bases de dados locais. Contudo, o esquema e a linguagem de consulta global nem sempre tem correspondência direta aos esquemas e às linguagens de consulta das bases de dados locais, necessitando serem traduzidas.

Assim o ambiente MBD deverá possuir diferentes tradutores, a fim de transformar as subconsultas globais para os esquemas e linguagens de consultas originais dos SGBDs locais. Esta tradução requer informações quanto às transformações entre os esquemas global e local e entre as linguagens. Essas informações são armazenadas em uma base de dados auxiliar contendo dados que descrevem o tipo de transformação entre os esquemas participantes e o esquema global.

Algumas vezes este processo de tradução encontra dificuldades devido às diferenças de sintaxe e expressividade das linguagens de consulta envolvidas na tradução, não permitindo que algumas operações na linguagem de consulta global sejam suportadas eficientemente pela linguagem de consulta original das bases de dados locais. Nessa situação é necessário desenvolver uma linguagem hospedeira da linguagem de consulta original usando algumas facilidades de uma linguagem de programação de alto nível. Neste processo de tradução há dois aspectos a serem considerados: i) a tradução correta da consulta e ii) o desempenho da consulta traduzida (MENG & YU, 1995).

A seguir as subconsultas, corretamente traduzidas para os esquemas e linguagens de consulta locais, são enviadas para as bases de dados locais, passando o controle para o terceiro nível, o SGBD local, que será o responsável pela otimização e execução da consulta local. Normalmente, cada SGBD local tem seu próprio processador de consulta que executa as consultas de acordo com um plano de execução local. Deste modo, ao receber a subconsulta local, o SGBD constrói este plano para o processamento da consulta e o executa para produzir a resposta. Este plano tem por finalidade estabelecer a seqüência de operações ordenadas, bem como a seleção da trajetória de acesso a ser usada no acesso aos registros físicos.

Após a execução das consultas em cada local pelo processador local, as respostas

serão retornadas ao usuário. Entretanto, primeiro elas serão pós-processadas pelo próprio ambiente MBD, ou pelo sistema local projetado pelo ambiente MBD, para resolver inconsistências e perda de valores de dados, reestruturar e/ou para padronizar os dados usados em diferentes aplicações das bases de dados, de forma a combinar os resultados retornados pelas consultas locais e uni-los dentro de uma única resposta a ser apresentada aos usuários.

III.6.3.2. Gerenciamento de transações no ambiente MBD

O gerente de transação no ambiente MBD tem por objetivo fornecer acesso integrado à bases de dados heterogêneas e distribuídas, de modo que suas aplicações e das bases componentes manipulem os dados concorrentemente, preservando a autonomia local de cada SGBD e sem prejudicar a consistência dos dados entre o ambiente e as bases (GARCIA *et al.*, 1995; KIM *et al.*, 1995 e SILBERSCHATZ *et al.*, 1991).

Para alcançar um bom desempenho, as arquitetura de um ambiente MBD se utilizam de duas propriedades ACID (Atomicidade, Consistência, Isolamento e Durabilidade) para o gerente de transação. A primeira é relativa à propriedade de atomicidade, que garante que na execução de uma transação do ambiente MBD, executada em vários SGBDs locais, possa tanto ser simultaneamente gravada como abortada, no caso de falha, em todos os locais. A segunda é relativa à propriedade de isolamento de uma transação do ambiente MBD, que garante ao gerente de transação submeter transações globais aos SGBDs locais, de maneira isolada de outras transações concorrentes.

Entretanto, estas propriedades são difíceis de serem asseguradas e a origem destas dificuldades pode ser observada de duas perspectivas: autonomia e heterogeneidade. Da primeira perspectiva, cada SGBD componentes da federação, ao executar transações pertencentes a este ambiente, deverá preservar um certo grau da autonomia já existente. A segunda perspectiva é decorrente da primeira, uma vez que a autonomia de cada SGBD local assegura a utilização de diferentes mecanismos de controle de concorrência e protocolos *commit* que não podem ser alterados, tornando difícil a execução de uma transação isolada de outras transações concorrentes.

Assim, tem sido realizado esforços para projetar uma arquitetura para o gerenciamento de transação do ambiente MBD, que permita flexibilidade na execução

da transação para garantir a atomicidade e isolamento das transações globais.

III.6.4. Características da linguagem de consulta para ambiente MBD

Um ambiente MBD utiliza linguagem de consulta uniforme e de alto nível que deve estar acima das diferenças entre os sistemas das bases de dados locais. Esta linguagem de consulta deve ser capaz de expressar qualquer consulta às várias bases de dados componentes deste ambiente. Ela deverá possuir facilidades para definir o esquema que integra as bases de dados heterogêneas; para interrogar sobre a estrutura das bases de dados relacionadas, o domínio e a natureza do processo que unirá os dados; bem como possuir mecanismos para otimizar a execução da consulta. Segundo BRIGHT *et al.* (1994) e LITWIN *et al.* (1990) KRISHNAMURTHY *et al.* (1991), uma linguagem de consulta para suportar um ambiente MBD deve possuir as seguintes características:

- habilidades para definir alias e abreviações para as denominações das propriedades, de forma a referir-se a múltiplos objetos semanticamente equivalentes, em diferentes bases de dados locais e para tratar duplicação e fragmentação de dados;
- dispositivos para que uma consulta se mantenha válida, apesar das alterações no ambiente provocadas pelo acesso a diversas bases de dados locais com esquemas distintos;
- permitir que, em diferentes pontos na execução de uma consulta, a operação possa ser aplicada a um único objeto ou a múltiplos objetos, em diferentes bases de dados;
- e
- permitir que uma operação em um determinado objeto possa ser desmembrada em várias operações a múltiplos objetos.

Além dessas características, a linguagem de consulta deve suportar estruturas temporárias para armazenar resultados intermediários ou novas representações da informação. Estas estruturas são denominadas por BRIGHT *et al.* (1994) e LITWIN *et al.* (1994) de atributos dinâmicos, e consistem em atributos temporários, definidos pela transformação dos atributos existentes nas bases de dados locais, componentes do ambiente MBD. Eles podem ser usados para acompanhar transformações no formato dos dados, abstrair valores dos atributos das diversas bases de dados em um único conjunto de valores e realizar uniões das relações.

As propostas para esta linguagem em ambiente MBD geralmente estendem as

tradicionais DDL (*Data Definition Language*) e DML (*Data Manipulation Language*). Segundo BREITBART (1990) estas propostas podem ser distinguidas de acordo com a arquitetura de ambiente MBD que se propõe atender, isto é:

- linguagem de manipulação de dados geradora de uma base de dados virtual - é empregada em arquiteturas federadas. Ela é caracterizada pelos meios que oferece ao usuário para especificar consultas que integram as várias bases de dados ou para definir uma visão temporária de seu interesse; e
- linguagem de definição e manipulação de dados Multi Base - é aplicada em arquiteturas de sistemas MBD. Ela é caracterizada por fornecer meios para realizar consultas a um esquema global que integra as várias bases de dados, de forma transparente ao usuário.

Em qualquer uma destas propostas, as linguagens exigem mais poder que as linguagens de consultas tradicionais, uma vez que elas deverão possuir mecanismos para definir exportação de esquemas, processar consultas globais transformando-as em um conjunto de sub-consultas na linguagem original das bases de dados e unir os resultados locais, além de suportar novos operadores para realizar transformações dinâmicas que resolvem os conflitos de heterogeneidade semântica e sintática dos dados.

III.6.5. Considerações sobre o resultado de consultas em ambientes MBD

Em um ambiente MBD os resultados retornados podem ser classificados em respostas certas (isométricas) e respostas prováveis.

As respostas prováveis ocorrem por causa da existência de inconsistências entre as tuplas resultantes das consultas locais. Elas são caracterizadas como tuplas possíveis que nem sempre no processo de combinação das respostas podem ser comparáveis. Isso, às vezes, traz alguma confusão ao usuário que não consegue distinguir qual delas é a resposta correta a consulta (CHEN, 1990).

ZHAO *et al.* (1995), por exemplo, propõem que seja utilizado um enriquecimento semântico de forma a apresentar junto com o resultado da consulta, as seguintes informações: as bases de dados fornecedoras dos resultados, a localização na rede desta, a especificação de atributo não existente e o domínio do valor do atributo.

Os autores (SOUZA, 1986; CHATTERJEE e SEGEV, 1991) que propõem na integração estimar o grau de proximidade semântica entre os elementos dos esquemas,

propõem também que, no processamento da consulta, estes graus sejam processados e apresentados junto com os resultados da consulta. Isto tem por finalidade permitir que os usuários possam comparar as tuplas prováveis com uma melhor compreensão destes.

III.6.6. Tecnologia usada para implementar um ambiente MBD

Atualmente, a implementação do ambiente MBD tem se beneficiado da arquitetura cliente-servidor e do paradigma orientado a objeto, uma vez que estes oferecem recursos para o processamento distribuído e heterogêneo.

A arquitetura cliente-servidor divide uma aplicação em processos executados em dois módulos independentes e localizados em plataformas diferentes, conectadas a uma rede, formando um único sistema, a saber: módulo cliente e módulo servidor (Figura III.9).

Na arquitetura MBD, o módulo cliente é composto por uma camada interface com o usuário e uma camada de processos. A camada interface é responsável pelo gerenciamento do diálogo com o usuário, permitindo realizar consultas ou comandos para a recuperação de dados e análises. Também é o meio pelo qual são apresentados os resultados. A camada de processos se utiliza de mecanismos para localizar os dados nos servidores, verificar a validade da consulta, decompor a consulta de acordo com a distribuição dos dados, acessar os servidores de forma otimizada e integrar o resultado da consulta.

O módulo servidor é responsável pela recepção da sub-consulta global, sua execução na base de dados local e a devolução de uma resposta ao módulo cliente que o requisitou. No módulo servidor estão localizados serviços que podem ser adquiridos através da rede por aplicações localizadas no módulo cliente, enquanto neste último estão localizados processos que disparam requisições do usuário a um/vários módulo(s) servidor(es) e a seguir aguarda o retorno do resultado destas requisições.

Esta arquitetura se utiliza ainda de um protocolo, também independente de plataforma, que fornece comunicação para transmitir requisições/resultados entre os clientes e os servidores e vice-versa.

O aspecto fundamental desta arquitetura é a separação dos processos em disparadores e processos reativos, correspondentes aos módulos cliente e servidor, respectivamente. Esta separação permite que o desenvolvimento da arquitetura MBD seja projetada como sistemas independentes, cada qual com seu próprio ambiente, não

importando o sistema operacional, o tipo de plataforma, ou que tipo de SGBD local.

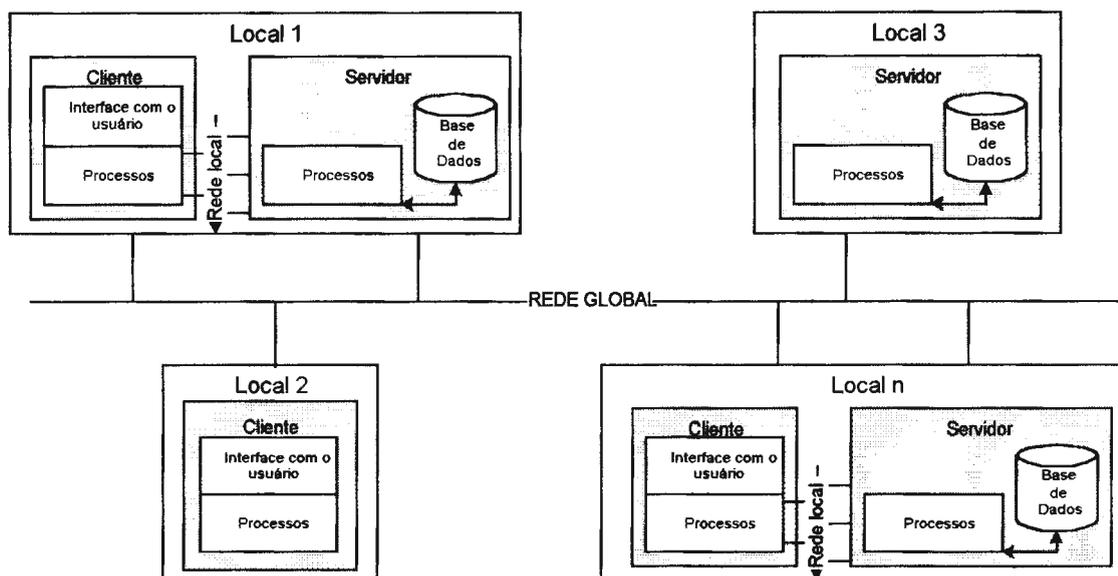


Figura III.9- Arquitetura cliente servidor.

A arquitetura cliente-servidor tem se beneficiado da tecnologia de objeto distribuído, proposto pelo paradigma da orientação a objeto (BRODIE & STONEBRAKER, 1995; ELMAGARMID *et al.*, 1990). O paradigma orientado a objeto permite definir tipos de dados, organizados em classes, de acordo com a sua estrutura e comportamento. O comportamento é implementado através de métodos que executam funções específicas e comunicação entre objetos. A estrutura e comportamento do objeto das classes são encapsulados, isto é, a modificação do estado de um objeto é realizada única e exclusivamente pelos métodos internos de sua classe. Estas classes são organizadas em uma hierarquia que permite herança entre elas.

Neste paradigma os recursos dos vários sistemas são modelados como objetos clientes e servidores que interagem através da troca de mensagens entre interfaces. Um objeto cliente solicita um serviço ao objeto servidor e este retorna ao primeiro os dados requisitados ou uma função de processamento.

A aplicabilidade do paradigma de orientação a objeto para o desenvolvimento de um ambiente MBD, usando arquitetura cliente-servidor, tem se mostrado uma forma mais natural de projetar a sua construção. Além da reutilização dos métodos dos objetos como unidades de programas, das facilidades para desenvolvimento do sistema usando o mecanismo de herança e do polimorfismo, este paradigma garante a autonomia e suporta a heterogeneidade. A autonomia é fornecida pela troca de mensagens através das interfaces. As operações enviadas pelas mensagens são executadas localmente

independente dos outros sistemas. A heterogeneidade entre os sistemas é suportada pela troca de mensagens entre objetos, uma vez que esta troca depende somente das interfaces componentes e não da sua implementação. As interfaces formam protocolos que são compreensíveis pelos sistemas componentes que precisam interoperar; eles podem residir em várias plataformas; ser implementados de diferentes maneiras; e terem outras características que são distintas uma das outras.

O mecanismo de herança, o encapsulamento do objeto e o conceito de polimorfismo tornam a programação mais fácil. Os mecanismos de herança permitem gerar novos objetos que automaticamente herdam todas as capacidades dos objetos existentes. O encapsulamento controla a modificação do estado de um objeto, ditando que mudanças nos dados são feitas via métodos dos objetos. Além disto, o encapsulamento permite modificar métodos ou adicionar novos, aproveitando os métodos já existentes e os reutilizando como unidades de programa. O conceito de polimorfismo, permite que um programa possa utilizar objetos de diferentes classes de forma transparente, pela interpretação de suas características durante o processamento (CATTEL, 1994).

Nos últimos anos, as pesquisas no desenvolvimento da computação de objetos distribuídos geraram produtos, tais como CORBA (*Common Object Request Broker Architecture*) e OLE, que introduzem novas perspectivas para o desenvolvimento destas arquiteturas. Estes produtos oferecem um conjunto de ferramentas, classes e serviços para comunicação estruturada entre objetos-aplicação e objetos-serviços (BERTINIO *et al.*, 1994; PITOURA *et al.*, 1995; HURSON & BRIGHT, 1996; DOGAC *et al.*, 1995; SILBERCHATZ & ZDONICK, 1997).

Outra inovação tecnológica que tem trazido novas dimensões, tornando o conceito de interoperabilidade mais amplo é a emergente expansão da *World Wide Web* (WWW). A rede global INTERNET consiste em um conjunto de redes heterogêneas de computadores interligadas pelo mundo inteiro com um conjunto de protocolos e serviços comuns. A utilização dos recursos desta rede para o desenvolvimento de ambientes MBD oferece vantagens, uma vez que, proporciona uma interface com o usuário, comum, baseada em hipertexto e protocolos de comunicação para programas e dados heterogêneos, além de permitir acessar repositórios de dados distribuídos em diferentes ambientes computacionais (MANOLA, 1995; FELLERS *et al.*, 1995; KONOPNICKI & SHMUELI, 1995; MALAIKA, 1996 e ABEL, 1997).

Capítulo IV - Requisitos de interoperabilidade em tre bases de dados geográficos e a proposta MultiSIG

IV.1. Introdução

A necessidade de compartilhar dados de várias fontes sempre esteve presente durante a evolução da tecnologia de Geoprocessamento, devido à existência de um grande volume de dados geográficos, já armazenados em diversos SIGs, e ao alto custo da aquisição de dados.

Entretanto, o maior obstáculo encontrado tem sido os diferentes formatos de arquivos, proporcionados pelas diversas estruturas de dados utilizadas para representar os objetos geográficos e as diferentes semânticas associadas a estas estruturas (FROMENT, 1994). Os SIGs, em geral, utilizam estruturas e formatos próprios, que foram projetados para maximizar o desempenho dos objetos geográficos, modelados de acordo com as necessidades das aplicações e considerando as operações oferecidas pela linguagem de manipulação (CLEMENT *et al.*, 1997).

As soluções propostas a esta necessidade de compartilhamento dos dados têm sido a conversão de estruturas e formatos de arquivos gerados pelos SIGs. Todavia, atualmente, a tendência para o compartilhamento de dados entre SIGs consiste em substituir estas soluções por um ambiente formado por Múltiplas Bases de Dados (MBD) de SIGs, que interoperam em tempo real (LAURINI, 1994; GOODMAN, 1994; GARDELS, 1997).

A implementação de um ambiente MBD para SIGs proporcionará o efetivo compartilhamento de dados geográficos, através de consultas a diversas bases de dados de SIGs. Este ambiente permite que dois ou mais SIGs heterogêneos interajam uns com os outros e troquem dados geográficos e serviços, através de consultas que acessam às bases de dados de forma transparente, independente da estrutura de dados, do modelo e formato de dados, da modelagem, do aplicativo e da plataforma.

Um ambiente MBD para SIGs atende às necessidades atuais de troca de dados geográficos e introduz vantagens no processo de intercâmbio de informações

geográficas e aquisição de dados. Dentre estas vantagens, destacam-se: *i)* maior eficácia no acesso às bases de objetos que oferecem uma maior acurácia ou estão mais atualizadas; e *ii)* redução do custo e do tempo de desenvolvimento da base de dados geográficos, uma vez que o processo de aquisição é o mais caro e trabalhoso na implantação de um projeto de SIG.

Este ambiente proporciona ainda a expansão das bases de dados geográficos e das aplicações dos usuários de SIGs, possibilitando obter novos tipos de análises, considerando dados multidisciplinares, não previsto pelas questões manifestas e latentes destas instituições (GOH *et al.*, 1994).

Todavia, a adaptação de um ambiente MBD para bases de dados convencionais não seria apropriada às bases de dados desenvolvidas, com finalidades de Geoprocessamento. Isto porque as metodologias convencionais não consideram a diversidade e riqueza de representação semântica dos dados geográficos, armazenados pelos SIGs, e nem as arquiteturas MBD contemplam as necessidades de consulta e manipulação espacial dos dados geográficos, nem tampouco as diferentes possibilidades de representá-los.

Este capítulo, inicialmente, apresenta a evolução do compartilhamento de dados geográficos, as propostas de interoperabilidade entre SIGs e uma análise comparativa destas. A seguir, faz algumas considerações sobre as bases de dados geográficos de diferentes comunidades de informação, disserta sobre os requisitos de uma metodologia para integração de esquemas de bases de dados geográficos, sobre os requisitos de representação dos esquemas destas bases para atender a metodologia e de uma arquitetura MBD. E, finalmente, apresenta a proposta do ambiente MBD, denominado de **Sistema MultiSIG**.

IV.2. Evolução do compartilhamento de dados geográficos

A estratégia inicial para promover o compartilhamento de dados geográficos entre SIGs é derivada dos sistemas convencionais. Ela consiste no desenvolvimento de aplicações orientadas a produtos que têm por finalidade transformar as estruturas de dados de um sistema para outro, através da conversão de formatos de arquivos de dados processados em *batch* (Figura IV.1). Esta estratégia, no âmbito da tecnologia de Banco de Dados, não é uma tarefa difícil, entretanto, quando aplicada a SIG se torna mais complexa, pois além da conversão de dados não gráficos é necessário considerar a conversão de dados gráficos e o relacionamento entre eles.

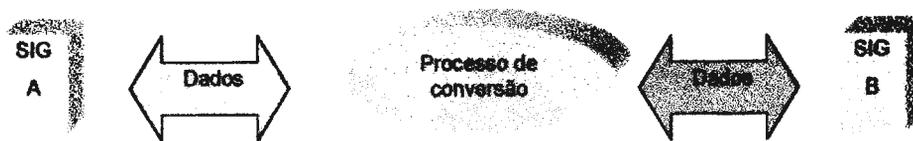


Figura IV.1- Conversão de dados.

Deste modo, a implementação destas aplicações exige o conhecimento prévio dos formatos dos arquivos de dados gráficos e não gráficos, levando em consideração as características dos *softwares*, das organizações dos dados, dos modelos dos dados, bem como a compreensão das necessidades dos usuários, em ambos os sistemas, precisando algumas vezes realizar considerações bidirecionais.

Esta estratégia se mostra relativamente eficiente quando se sabe onde os dados estão localizados. Porém ela é pontual, proporcionando interoperabilidade entre pares específicos de SIGs e apresenta a necessidade de conversão e migração reversa de parte dos dados, que foram atualizados em um deles, de forma a não gerar inconsistências entre os sistemas.

Uma outra solução é decorrente da consolidação de alguns formatos de dados, mais utilizados para troca de dados entre SIGs. De modo geral, ela resulta do amadurecimento dos *softwares* para SIGs disponíveis no mercado. Eles passaram a incorporar os processos de conversão através de aplicações. Estas aplicações permitem importar/exportar formatos de arquivos de dados que se consolidaram no mercado como padrão, tal como formato DXF (*Drawing eXchange Format*) e SIF (*Standard Interchange Format*) definidos, respectivamente, pela *Autodesk* e *Intergraph* como formatos para importação/exportação de gráficos para o seus sistemas de desenho assistido por computador (CASSETARI, 1993; ZAJAC, 1994). Todavia, esta solução se torna ineficiente quando há necessidade de trocar dados com um novo tipo de sistema não previsto por estas aplicações.

Atualmente, uma das soluções propostas pela comunidade usuária dos SIGs é o estabelecimento de um padrão para descrever os dados, de forma a facilitar o intercâmbio e a transferência de dados entre os diferentes sistemas. Esta solução representa um denominador comum entre diferentes sistemas e viria a reduzir o número de aplicações de conversão de $n(n-1)$ para $2n$, onde n é o número de formatos de arquivos de dados, conforme apresentado na Figura IV.2.

Estes padrões atendem às necessidades de intercâmbio e transferência de dados com diferentes formatos. Em alguns países, órgãos com poder normativo têm proposto inúmeros padrões para o intercâmbio de dados geográficos. Dentre estes, destacam-se

os padrões americanos *Spatial Data Transfer Standard* (STDS) e o *Digital Geographic Information Exchange Standard* (DIGEST), o padrão canadense *Spatial Archive Interchange Format* (SAIF) e o padrão europeu proposto pelo *Technical Committee* (TC) 287 do *European Standardisation Committee* (CASSETARI, 1993; KUCERA *et al.*, 1993; ARMSTRONG & CARR, 1994; SAIF, 1995; TIMPF *et al.*, 1997).

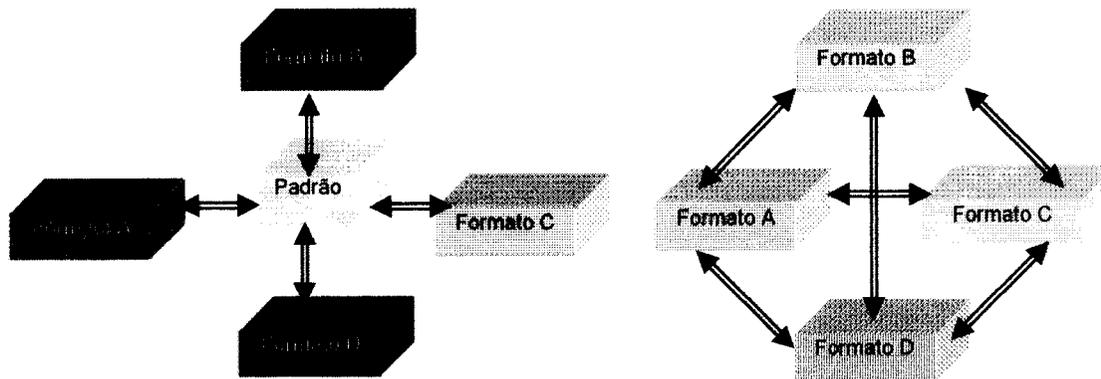


Figura IV.2 - Comparação do número de conversores com/sem padrão de dados (as setas significam duas transformações de arquivos).

O SDTS é um padrão americano, aprovado em 1992 como uma *Federal Information Processing Standard* (FIPS), *Publication 173*. Este padrão suporta a representação lógica, física e sobre o conteúdo dos diferentes modelos de dados usados pelos SIGs. Ele oferece componentes para representar os dados nas estruturas vetorial e matricial, e para definir referenciais, especificações para troca de dados, especificações sobre a qualidade dos dados e características cartográficas (CASSETARI, 1993).

O DIGEST é um padrão desenvolvido pelo *Digital Geographic Information Working Group* (DGIWG) e aprovado como um padrão NATO (*North Atlantic Treaty Organisation*). Ele foi desenvolvido com a finalidade de atender às necessidades das agências militares, em trocar dados cartográficos digitais. Este padrão emprega um modelo de dados hierárquico que contempla estruturas vetoriais e matriciais. Ele permite empregar diferentes níveis de descrição da topologia e dos atributos alfanuméricos, além de informações sobre a organização, a estrutura e a qualidade dos dados (CASSETARI, 1993).

O SAIF é um metamodelo proposto pelo *Survey and Resource Mapping Branch of the Ministry of Environment Lands and Parks* (MELP) do Canadá. Este metamodelo emprega o paradigma da orientação a objetos para definir uma hierarquia de classes que oferece estruturas de dados, independente de *software*, para descrever dados geográficos em um nível mais abstrato do que os outros padrões de dados geográficos. Este

metamodelo oferece ainda duas notações sintáticas, a saber: uma para descrever as classes e outra para descrever os objetos. Este padrão foi projetado para modelar e promover a troca de dados espaço-temporais, utilizando ferramentas, tal como *Feature Manipulation Engine* (FMB) (KUCERA *et al.*, 1993; ARMSTRONG & CARR, 1994; SAIF, 1995).

O padrão proposto pelo TC 287 define um esquema conceitual para metadados baseado em dois padrões, a saber: um para expressar a qualidade e outro para expressar as referenciais espaciais e o posicionamento dos objetos geográficos do esquema. Este padrão utiliza a linguagem *Express*, desenvolvida *pele International Standards Organization*, para a definição formal de entidades e relações dos esquemas (TIMPF *et al.*, 1997).

Um padrão para descrever dados geográficos deve oferecer simplicidade de manuseio e alto poder de expressão para representar qualquer modelo de dados. Além disto, o seu estabelecimento requer três considerações. A primeira delas consiste em definir como o padrão pretende atender o requisito heterogeneidade. A segunda é relativa à neutralidade deste padrão, que deve ser estabelecida através de um consenso entre os fabricantes e usuários. E a última, consiste em estabelecer um padrão capaz de representar os sistemas atuais e evoluir com os novos conceitos que poderão surgir no âmbito da tecnologia de Geoprocessamento. Cabe ressaltar o cuidado com a taxa de inovação da informática, que torna os padrões obsoletos mesmo antes de serem completamente estabelecidos.

Em geral, as especificações de padrões para descrever dados geográficos fornecem meios para descrever além de formatos, parâmetros relacionados a qualidade dos dados, o esquema semântico, o esquema espacial, e os metadados do esquema. Entretanto, o compartilhamento de dados empregando padrões focaliza a parte estrutural dos dados, não contemplando a semântica destes.

Nos últimos anos, a utilização da rede Internet pela tecnologia de SIG tem proporcionado inovações tecnológicas para o compartilhamento de dados (FELLERS *et al.*, 1995), a saber: o desenvolvimento de catálogos e bibliotecas digitais.

Os catálogos são mecanismos utilizados por comunidades de informação para colocar à disposição os dados armazenados para outras comunidades de informação. Estes catálogos empregam os padrões propostos para descrever os dados de forma a encapsular a origem dos dados. Eles oferecem interfaces para requisitar dados e mecanismos para os servidores SIGs gerarem as respostas com componentes espaciais (BLOTT & VCKOVSKI, 1995; KRAMER *et al.*, 1997). Por exemplo, o

Umweltdatenkatalog (UDK) (KRAMER *et al.*, 1997) é um sistema de meta-informação para dados ambientais de uso institucional, usado na Alemanha e Áustria, disponível na rede Internet e Intranet. Este catálogo possui um dicionário de termos ambientais que permite o acesso, *on line*, para compreender os dados. Um outro exemplo é o Projeto GeoScope, desenvolvido pelo consórcio entre *Belcore Chamber Corporation*, *California Resource Agency*, *Rutgers University Center for Remote Sensing and Spatial Analysis*, NASA e o consórcio OPENGIS (CORNELIO & THOMAS, 1995; SHKLAR, *et al.*, 1995). Este projeto possui uma arquitetura para Internet que localiza, acessa e permite a visualização de dados provenientes de imagens de satélite, das cartas digitais americanas e dados de setores censitários americanos. Ele possui uma terminologia comum, usando os metadados definidos pelo FGDC. Os dados são recuperados nos *Hierarchical Data Format* (HDF) ou *Spatial Data Transfer Standard* (SDTS).

As bibliotecas digitais tratam de uma coleção organizada de informação digital, cuja natureza da informação é essencialmente multimídia. Estas bibliotecas oferecem ferramentas para busca e acesso à informação. Elas se utilizam da tecnologia de *data mining* para pesquisar e recuperar informação relevante, sem qualquer conhecimento a priori da corrente organização dos servidores de dados. Esta pesquisa é feita por palavras chaves, agrupadas em catálogos de metadados. Um exemplo é o Projeto da Biblioteca Digital Alexandria (ADL), desenvolvida pelo *Map and Imagery Laboratory* da Universidade de Santa Barbara em consórcio entre US *National Science Foundation* (NSF), *Advanced Research Projects Agency* (ARPA) e o *National Aeronautics and Space Administration*. Este projeto objetiva construir uma biblioteca para dados digitais espacialmente indexados e distribuídos sobre a rede Internet. A ADL utiliza um subconjunto do FGDC para especificar os metadados e oferece mecanismos para localizar, catalogar, pesquisar e visualizar mapas e imagens.

Nestas propostas verifica-se que o intercâmbio de dados não garante interoperabilidade semântica entre as bases de dados geográficos, uma vez que, parte substancial da informação está contida nos procedimentos de consulta, manipulação e apresentação dos dados geográficos.

IV.3. Propostas de interoperabilidade entre SIGs

No âmbito da Tecnologia de Banco de Dados Convencionais, as soluções se concentram no desenvolvimento de ambientes MBD, apresentados no capítulo 3. Na

verdade, esta solução traz novas perspectivas para o compartilhamento de dados entre SIGs e caracteriza a tendência atual de substituir as soluções orientadas ao produto do SIG por arquiteturas abertas que permitam o compartilhamento de dados. Ela oferece mecanismos para substituir o atual ambiente monolítico, no qual cada SIG controla o acesso restrito à sua base de dados e sua aplicação, por um ambiente aberto, composto por outros SIGs que interoperam em tempo real, a partir de aplicações independentes fortemente encapsuladas (GARDELS^{html}, 1996c; IGRAS, 1994; GOODMAN, 1994; LAURINI, 1994; CLEMENT *et al.*, 1997; ABEL *et al.*, 1997; LECLECQ *et al.*, 1996, GOODCHILD^{html}, 1997).

Nos últimos anos, os esforços da tecnologia de Geoprocessamento têm se concentrado em especificações de mecanismos para alcançar a interoperabilidade semântica, preservando o vasto investimento nas bases de dados geográficos já existentes e assegurando a oportunidade de introduzir novos métodos para acessar a informação geográfica distribuída. Dentre estes esforços encontram-se a especificação proposta pelo OGC (*Open GIS Consortium Inc.*) e algumas propostas de interoperabilidade entre SIGs.

O OGC é um consórcio fundado em 1994 como resposta ao reconhecimento do problema de não interoperabilidade entre diferentes comunidades de informação de Geoprocessamento. Este consórcio propõe o OGIS (*Open Geodata Specification*), de forma a estabelecer um consenso entre a indústria, fornecedores e consumidores de dados geográficos (GARDELS, 1997).

Assim, as seções a seguir apresentam a especificação proposta pelo OGC, algumas propostas de interoperabilidade entre bases de dados de SIGs, encontradas na literatura, e, ao final, uma análise comparativa destas propostas.

IV.3.1. Proposta de interoperabilidade do OGC: OGIS

OGIS é uma especificação baseada na tecnologia de distribuição e no paradigma da orientação a objeto. Ela estabelece fundamentos para o desenvolvimento de uma tecnologia comum que tem por objetivo auxiliar os desenvolvedores de sistemas a construir aplicações para o compartilhamento de dados geográficos e aplicações de Geoprocessamento (GARDELS, 1997). Ela define uma estrutura detalhada para o desenvolvimento de aplicativos disponíveis em rede, que através de um fluxo de trabalho permite aos usuários acessarem e processarem dados geográficos armazenados em diversas fontes de Geoprocessamento distribuídas e heterogêneas. Segundo

BUEHLER & MCKEE^{html} (1996), esta especificação oferece as seguintes características para o desenvolvimento de interoperabilidade entre aplicações SIG:

- portabilidade - independência do ambiente do aplicativo SIG, da plataforma e da rede;
- cooperação - suportar computação distribuída e recursos de dados compartilhados;
- compatibilidade - preservar investimentos de usos nos dados geográficos e aplicativos locais; e
- extensibilidade - assimilar novos aplicativos para Geoprocessamento e tipos de dados geográficos, bem como acomodar novas tecnologias.

A especificação proposta pelo OGC para uma arquitetura interoperável possui três modelos conceituais (BUEHLER & MCKEE^{html}, 1996; GARDELS^{html}, 1996a; GARDELS^{html}, 1996b; YASER, 1997). O primeiro, consiste em um modelo de dados que estabelece objetos (entidades, atributos e relacionamentos) e eventos que são manipulados pelas comunidades de informação. O segundo, consiste em modelo de serviços a serem desenvolvidos para requisitar dados, acessar as bases e gerar as respostas. O terceiro modelo estabelece o fluxo de trabalho entre as comunidades de informação, considerando as limitações impostas pelas plataformas e aplicativos SIGs. Desta forma, o OGIS descreve a especificação para o desenvolvimento de um sistema para interoperabilidade através dos seguintes componentes:

- *Open Geodata Model (OGM)* - fornece um modelo de dados comum para representar a superfície terrestre e seus fenômenos espaço-temporais, matematicamente e conceitualmente;
- *Open Geoprocessing Services (OGS)* - trata-se da especificação para implementação de serviços para acessar, gerenciar, manipular, representar e compartilhar dados geográficos entre comunidades de informação; e
- *Information Communities Model (ICM)* - oferece uma estrutura para usar um modelo de dados geográficos e o modelo de serviços OGS, de forma a solucionar não somente o problema de interoperabilidade técnica, mas o problema de interoperabilidade institucional.

O OGM modela os objetos geográficos como entidades e fenômenos da superfície terrestre, em um domínio espacial. Para isto, o OGM estabelece uma hierarquia de tipos de dados genéricos (Figura IV.3) que permite modelar o espaço geográfico necessário a qualquer domínio de aplicação, independente do modelo de dados, da estrutura de dados e do formato dos arquivos. Sobre esta hierarquia são

definidas operações para análise e apresentação dos dados. Esta hierarquia é formada pelas seguintes classes básicas:

- *Metadata* - oferece uma estrutura para descrever informações que descrevem a qualidade e o histórico de uma coleção de dados geográficos, no contexto da aplicação de uma comunidade de informação;
- *Spatial Component* - fornece subsídios para descrever a referência espacial e a geometria. A geometria é contemplada pelas seguintes especializações:
 - ✓ *Feature* - oferece uma hierarquia de classes para modelar as entidades do mundo real como objetos discretos, com limite e extensão bem definidos em um domínio espacial, de acordo com a geometria de ponto, linha, polígono e superfície; e

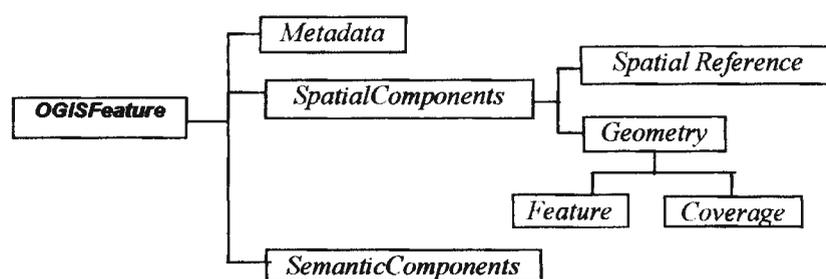


Figura IV.3- Hierarquia de classes do modelo de dados geográficos do OGIS (GARDELS, 1996b).

- ✓ *Coverage* - oferece uma hierarquia com um conjunto de tipos que modelam fenômenos variantes sobre o espaço e não possuem extensão espacial específica. Estes fenômenos são tratados como um conjunto de pontos em um domínio espaço-temporal, no qual cada ponto é associado a um valor simples ou complexo, tal como imagem, mapas temáticos e *Triangular Irregular Networks* (TINs); e
- *Semantic Components* - descreve as propriedades simples ou complexas dos objetos geográficos, em termos de tipos de dados primitivos e restrições sobre estes tipos.

O OGS emprega a modelagem orientada a objeto para definir um conjunto de serviços de rede para acessar e processar aplicações que gerenciam uma variedade de dados geográficos, armazenados em uma ampla variedade de Bases de Dados, relacionais e não relacionais. O acesso e processamento, neste contexto, significa que os usuários de dados geográficos podem consultar bases de dados, usar recursos de processamento remotos e obter vantagens de outras tecnologias de computação distribuídas. Por exemplo, o *download* de aplicativos, a partir de um ambiente remoto para uso temporário em um ambiente local.

O OGS é dividido em três categorias de serviços, a saber: *a)* serviços para identificar, interpretar e representar o conjunto de dados armazenados em um servidor de dados geográficos; *b)* serviços para capacitar as aplicações a interagirem umas com as outras; e *c)* serviços para gerenciar a entrega de dados geográficos processados. Estes serviços são apresentados por BUEHLER & MCKEE ^{html} (1996), como:

- serviços de catálogo e esquema - são serviços para cadastrar um esquema que deseja compartilhar dados geográficos e documentar seu conteúdo, bem como consultar o catálogo de esquemas;
- serviços de transformação - são serviços para realizar transformações semânticas e estruturais, de forma a possibilitar a troca de dados geográficos. Para as transformações semânticas são oferecidos serviços para capturar a semântica e traduzi-la para outras comunidades que usam diferentes definições dos dados geográficos. Os serviços de transformação semântica oferecem uma estrutura para interpretar os dados quanto ao seu significado, sua estrutura e seu nível de qualidade. Os serviços de transformações estruturais tornam compatíveis ou integram os resultados das consultas; e
- serviços de consulta - são serviços que permitem realizar uma consulta e recuperar a informação distribuindo-a nos servidores de dados geográficos. Este serviços são modelados como o *OMG Query Service*.

O OGS propõe estender a *Object Query Language* (OQL) para consultar os dados geográficos, uma vez que esta linguagem suporta a definição de dados e operações na *Object Data Language* (ODL). Desta forma, algumas operações geométricas e topológicas são definidas, tais como *intersect*, *containedIn*, *equal*, *bufferZone*.

A especificação destes serviços pode ser implementada pela tecnologia de objetos distribuídos, baseada em arquiteturas cliente-servidor, tal como *OMG CORBA*, *Microsoft OLE/COM*, *SQL* via interface *Object Database Connectivity* (ODBC) ou pela linguagem de programação *Java* (VCKOVSKI & BUCHER, 1996), conforme apresentado em GARDELS (1996a) e ilustrado na Figura IV.4.

O ICM estabelece um meio para as comunidades de informação, fornecedoras e consumidoras de dados geográficos, que desejam compartilhá-los, cooperarem conforme a especificação proposta pelo OGIS. Isto requer o estabelecimento de um sistema de comunicação e um fluxo de trabalho, de forma que as comunidades colaborem para informar sobre as características dos dados, no que diz respeito ao

conteúdo da informação, à qualidade, ao histórico dos dados e também sobre os serviços envolvidos no seu uso.

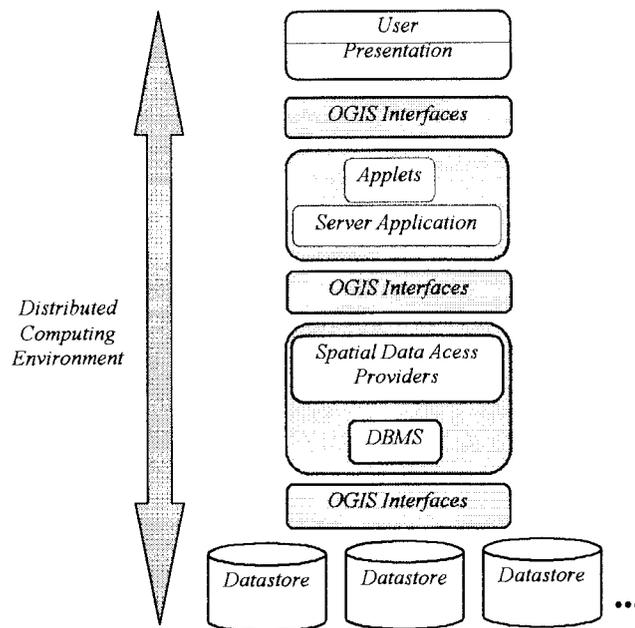


Figura IV.4- Visão geral da arquitetura OGIS segundo GARDELS (1996a).

Como resultados dos esforços da especificação proposta pelo OGIS (GARDELS, 1997), têm-se: o *Open Geospatial Datastore Interface* (OGDI) (CLEMENT *et al.*, 1997), o Projeto CERES (CERES^{html}, 1996) e o *Environment Information System* (EIS) (GÜNTHER, 1997); que são baseados nela.

O OGDI é uma Interface de Programação de Aplicação (IPA), de domínio público, que facilita o acesso a produtos de dados geográficos distribuído sobre a rede INTERNET. Esta IPA encapsula as tarefas de acesso a aplicativos SIGs, através de métodos padronizados que permitem a leitura de formatos sem tradução. Trata-se de uma arquitetura cliente-servidor desenvolvida sobre uma rede TCP/IP (*Transmission Control Protocol/Internet Protocol*), em linguagem C, que emprega um modelo de dados do padrão DIGEST, com estruturas matricial e vetorial para ponto, linha, área e texto.

Tanto o Projeto CERES quanto o EIS são iniciativas para o desenvolvimento de arquiteturas que possibilitam o compartilhamento de dados geográficos capazes de localizar, acessar, visualizar e consultar dados de SIG distribuídos sobre uma rede. O Projeto CERES utiliza o FGDC para definir os metadados, ao passo que o EIS utiliza o modelo proposto pelo OGIS.

IV.3.2. Propostas de interoperabilidade entre bases de dados de SIGs

Na literatura são encontradas algumas iniciativas de implementação de sistemas que proporcionam interoperabilidade ou ambientes MBD para bases de dados geográficos. Dentre elas, cabe destacar os esforços de:

- KUMAR *et al.* (1997) propõem um sistema denominado Do-GIS para SIG distribuídos. Este sistema não preconiza a integração das bases de dados. Ele possui uma hierarquia orientada a objeto, que é utilizada como modelo de dados para representar os objetos disponíveis nas bases de dados locais. Esta hierarquia é composta pelas classes *Point*, *Line* e *Polygon*, que são especializadas de acordo com a geometria do objeto geográfico nas bases de dados dos SIG. A linguagem de consulta é extensão da *Structured Query Language* (SQL), denominada de GEOSQL. Os autores adicionam às cláusulas básicas da SQL construtores para suportar operadores e relações espaciais, a saber:

```
FROM <classeDeEntidade> <NomeVariavel_1> <NomeVariavel_2>
SELECT [qualificador] <NomeVariavel_1>. <atributo>
WHERE
    <NomeVariavel_1>. <atributoNãoEspacial> = <constante>
    <operadorLógico>
    <NomeVariavel_1>. <operadorEspacial> <NomeVariavel_2>
```

A arquitetura é desenvolvida sobre plataforma UNIX local e composta por:

- *Object base* - onde se encontram os arquivos e estruturas dos SIG locais;
 - *Object Base Server* - trata-se de um servidor que tem a função de requisitar dados no *Object Base*;
 - *Meta Knowledge Base* - trata-se de um repositório que mantém informação sobre a localização dos dados nas bases de dados locais;
 - *Query Translator* - analisa a consulta e constrói uma consulta que é usada pelo *Query Server*;
 - *Query Server* - avalia as consultas introduzidas pelo usuário e a distribui para os *Object Base Server*; e
 - *Mom* - monitora e informa sobre o estado do sistema, acompanhando os módulos ativos e tratando a ocorrência de falha no sistema.
- LECLECQ *et al.* (1996, 1997a, 1997b) propõem um SIG interoperável que fornece um SIG virtual. Este sistema se utiliza de um modelo de dados orientado a objeto. A parte estrutural deste modelo é usada como modelo de dados canônico para

representar os conceitos espaciais, armazenados pelos esquemas das bases de dados dos SIGs locais; a parte comportamental é usada como um modelo funcional para definir operadores espaciais genéricos para cada tipo de dados.

A arquitetura é desenvolvida sobre o *Object Request Broker* (ORB) de forma a proporcionar interconectividade ao nível de serviço. O ORB, definido na arquitetura CORBA (PIRES,1997), fornece transparência de modo que os diferentes serviços residindo em plataformas distribuídas sobre uma rede local possam ser implementadas em diferentes linguagens. Esta arquitetura é composta por um tradutor (*wrapper*) para cada base de dados do SIG local. Este tradutor utiliza regras de transformação que convertem a consulta global para a linguagem do SIG local, proporcionando conexão ao SIG local; submete a execução da consulta e compõe as instâncias do resultado. Esta arquitetura é composta ainda por *Database Type Services* que possui os seguintes componentes:

- ◆ *Schemas Dictionary* - descreve os esquemas locais das bases de dados dos SIGs, locais no modelo de dados canônico;
 - ◆ *Operators Dictionary* - consiste em um conjunto de operadores presentes nos diferentes SIGs locais e um esquema de operadores integrados expressos no modelo funcional;
 - ◆ *Global Query Manager* - gerencia a execução da consulta, decompondo-a em subconsultas de acordo com a distribuição dos dados; e
 - ◆ *Global Query Transaction* - é responsável pelo controle de concorrência das transações.
-
- YASER (1997) propõe um sistema para identificação e transferência de dados entre SIGs heterogêneos e distribuídos. O autor não utiliza um MDC, e sim uma hierarquia semântica para determinar similaridade entre aplicações. A especificação da semântica descreve o conteúdo da base de dados, de cada comunidade de informação, e o relacionamento semântico entre elas. Sobre esta hierarquia são definidos tipos de contextos e operadores semânticos, baseados na especificação do OGIS. Os tipos de contextos especificam a semântica das bases de dados dos SIG locais, onde contexto é o conhecimento sobre um domínio de um problema. A descrição dos metadados utiliza a especificação proposta pelo FGDC e está no mesmo nível da hierarquia semântica.

A arquitetura é implementada sobre uma rede local com dois tipos de componentes: um servidor global e os sistemas locais. O servidor global é composto

pelo esquema global, os metadados, um otimizador de consulta associado a uma engenharia de pesquisa e um FGIS *driver*. Este último fornece uma plataforma e interface, independente de sistema operacional, que permite o acesso aos serviços dos SIGs locais. Os sistemas locais são compostos por uma interface gráfica com o usuário, que torna disponível os recursos da federação, e também por um FGIS *driver* similar ao global, que acessa o esquema local e os metadados.

- GÜNTHER (1997) propõe uma arquitetura baseada no OGIS, denominada de *Open EIS (Environmental Information System)*, que tem por objetivo o desenvolvimento de um sistema de integração centrado em conceitos de subsistemas autônomos que interagem. As interfaces são orientadas para o modelo proposto pelo *Open GIS*. Este sistema tem como componentes (Figura IV.5):
 - ❖ *viewer*- apresenta todas as funcionalidades do sistema permitindo ao usuário interagir;
 - ❖ *analyser* - realiza operações de geoprocessamento sobre os dados;
 - ❖ *fuser* - decompõe a consulta e integra os resultados da consulta; e
 - ❖ *atomiser* - responsável por extrair os dados dos servidores de dados.

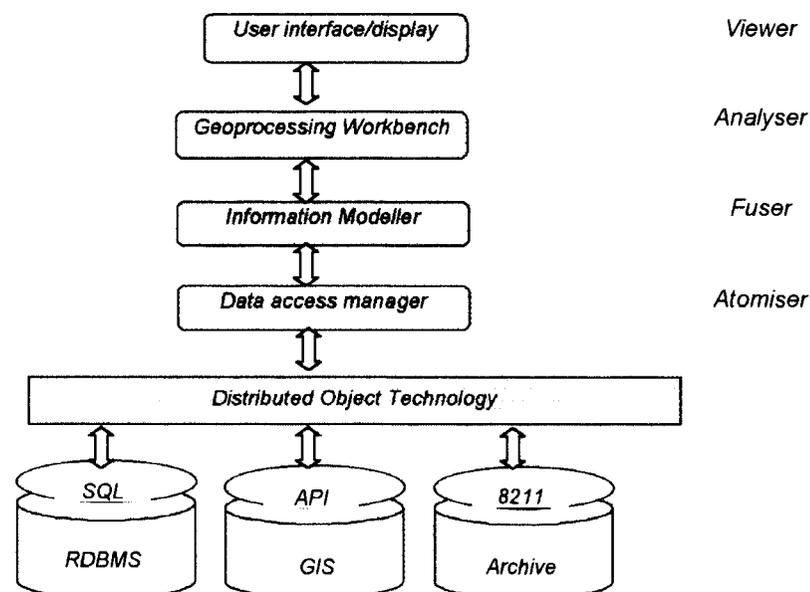


Figura IV.5 - Componentes do *Open EIS* (GÜNTHER, 1997).

- DEVOGELE *et al.* (1996) não propõem uma arquitetura para MBD, mas abordam o problema de múltiplas representações utilizadas em bases de dados geográficos. Os autores propõem o conceito de relacionamento de escala de transição. Este conceito conecta dois conjuntos de elementos (classes ou objetos) que representam o mesmo

fenômeno no mundo real, através de uma seqüência de operações entre escalas para ir de uma representação para outra, a saber:

- ◆ Generalização cartográfica - eliminação de detalhes geométricos ao representar objetos em uma escala menor que a original;
 - ◆ Operações tomadas emprestadas da teoria de Banco de Dados para encontrar correspondência e conflitos entre os esquemas. Estas correspondências e soluções dos conflitos são armazenadas no relacionamento da escala de transição; e
 - ◆ Combinação dos dados com ajuda da informação geométrica e topológica.
-
- PARENT *et al.*(1996) propõem uma metodologia para integração de esquemas de bases de dados de SIGs composta de três fases: preparação dos esquemas, investigação de correspondência e integração. Na pré-integração os esquemas são descritos e enriquecidos com informações complementares sobre o esquema. Na investigação de correspondência são identificados os conflitos existentes entre os esquemas das bases de dados geográficos. Nesta etapa, os autores apontam os principais tipos de conflitos semânticos e estruturais que podem ser encontrados entre os esquemas e propõem soluções para os conflitos identificados. Na integração é criado um esquema virtual, acima das fontes de dados, contendo as soluções dos conflitos.
 - AGUIAR (1995) e AGUIAR & MEDEIROS (1996) propõem uma solução para integrar aplicações de planejamentos urbanos de SIG, empregando um modelo de dados proposto por PIRES (1995). A solução proposta desenvolve atividades para integrar SGBDs relacionais e orientados a objetos, a partir dos modelos *Object Modelling Technique* (OMT) e Entidade-Categoria-Relacionamento (ECR). Esta solução consiste em mapear os modelos de dados locais para um esquema global implementado no SGBD orientado a objeto (SGBDOO) O₂.
 - IGRAS (1994) e GOODMAN (1994) apresentam o projeto *Land Related Information System* (LRIS), que implementa um sistema de Banco de Dados Federados para SIG sobre uma rede local. Este sistema traduz os esquemas dos SIGs locais para esquemas componentes definidos no modelo relacional. Ele utiliza como linguagem de manipulação, uma extensão da SQL denominada *Spatial Query*

Language (SpQL). Esta linguagem estende as cláusulas da SQL para suportar operadores espaciais e habilidade para especificar formato de saída dos dados:

```
SELECT <ListaSeleção1>... <ListaSeleção1N>  
FROM <NomeTabela1>... <NomeTabelaN> OR <IdRequisitado>  
WHERE <operadoresRelacionais>.AND/OR <operadoresEspaciais>  
REFORMAT <ListaSeleção1> AS <formatoSaida1>  
....  
REFORMAT <<ListaSeleção1> AS <formatoSaidaK>
```

A arquitetura deste sistema é composta de uma interface gráfica na qual o usuário submete uma consulta, e esta é passada para um *Distributed Query Processing Engine* (DQPE). Este tem a função de validar, estimar o custo da consulta e enviá-la para os *Local Query Processor* (LQP). O DQPE é composto por uma *Standard Network Interface* (SNI), um *Global Data Dictionary* (GDD), um *Global Query Processor* (GQP), e ferramentas para calcular o custo da consulta e estabelecer a estatísticas. Os LQP recuperam os dados nos SIG locais e enviam para um *Global Result Database* (GRDB) que armazena os dados temporários recuperados pelo LQP e os devolve ao DQPE que integra os resultados e os coloca no formato solicitado pela cláusula REFORMAT da SpQL.

IV.3.3. Análise comparativa das propostas encontradas na literatura

A especificação proposta pelo OGIS contempla a necessidade de representação através do modelo de dados proposto, a interoperabilidade física através do conjunto de serviços especificados, e o fluxo de trabalho entre as comunidades de informação. Entretanto, esta especificação não garante interoperabilidade semântica, uma vez que, não leva em consideração a compreensão das bases de dados dos SIG, não somente ao nível de sistema ou programas, mas também ao nível de tipos de dados armazenados por estas bases. Esta interoperabilidade semântica é obtida através de uma metodologia para a integração lógica dos esquemas, conforme desenvolvido para os sistemas convencionais, apresentada na seção III.5.

Contudo, o desenvolvimento de interoperabilidade semântica, entre bases de dados geográficos, é um pouco mais complexo que os sistemas convencionais. Esta complexidade decorre dos SIGs armazenarem dados não convencionais que são representados nestes sistemas por uma diversidade de modelos de dados. Estes modelos oferecem uma variedade de possibilidades de representar os dados geográficos nos

SIGs. Além disto, os dados geográficos são semanticamente mais ricos e caracterizados por apresentarem diferenças espaço-temporais, uma vez que podem apresentar diversas interpretações geométricas e variar quanto à sua localização, extensão espacial e sobre o tempo, respectivamente (GÜNTHER, 1997).

Quanto às propostas dos sistemas encontrados na literatura, observa-se que elas enfatizam, ora a interoperabilidade semântica através de uma metodologia e de um modelo de representação dos esquemas, ora a interoperabilidade física através de uma arquitetura, de serviços ou de uma linguagem de consulta, conforme apresentado na Tabela IV.1.

KUMAR *et al.*(1997), LECLEQC *et al.* (1997a, 1997b) GÜNTHER (1997), IGRAS (1994) e GOODMAN (1994) enfatizam arquiteturas que possibilitam acessar bases de dados geográficos. Estas arquiteturas recuperam as características temáticas e espaciais dos dados geográficos, através de operadores espaciais. Contudo, estas soluções não contemplam as heterogeneidades dos esquemas de bases de dados geográficos, como por exemplo, diferenças semânticas e sintáticas decorrentes da diversidade de representação dos dados geográficos.

Tabela IV.1 - Comparação das propostas encontradas na literatura

	REPRESENTAÇÃO DOS ESQUEMAS	METODOLOGIA	ARQUITETURA	LINGUAGEM DE CONSULTA	CONJUNTO DE SERVIÇOS
KUMAR <i>et al.</i> (1997)	X		X	X	
LECLEQC <i>et al.</i> (1996)	X		X		X
GÜNTHER (1997)	X		X		
YASER (1997)	X	X	X		X
DEVOGELE <i>et al.</i> (1996)		X			
PARENT <i>et al.</i> (1996)		X			
AGUIAR & MEDEIROS (1996)	X	X		X	
IGRAS(1994) & GOODMAN(1994)			X	X	

DEVOGELE *et al.* (1996) apresentam soluções para tratar múltiplas representações dos dados geográficos, sem considerar uma metodologia para integrar esquemas ou arquitetura MBD.

PARENT *et al.* (1996) e AGUIAR (1995) enfatizam metodologias para integração de esquemas de bases de dados geográficos. A metodologia proposta por PARENT *et al.* (1996), apesar de enfatizar os conflitos semânticos e estruturais que podem ser encontrados entre os esquemas de bases de dados geográficos, é difícil de ser aplicada para o desenvolvimento de um ambiente MBD, uma vez que não leva em

consideração a afinidade entre esquemas que possuem a mesma área de aplicação. A metodologia proposta por AGUIAR (1995), por tratar da integração de dois esquemas pré-definidos com a mesma área de aplicação, não contempla os aspectos espaciais e nem a representação gráfica dos objetos em sua plenitude. Isto torna estas metodologias restritas. A primeira, a aplicações que possuem a mesma área de aplicação e empregam as mesmas referências para tratar o espaço geográfico. A segunda, a aplicações semelhantes àquelas para as quais foi desenvolvida.

YASER (1997), dos trabalhos citados, é o único que contempla a heterogeneidade de alto e baixo níveis através de uma metodologia e uma arquitetura, respectivamente, para múltiplas bases de dados geográficos. Na solução proposta pelo autor, apesar de considerar a diversidade das áreas de aplicação, a semântica dos objetos geográficos é definida sobre uma sintaxe para representar os objetos geográficos e tratada sobre um modelo hierárquico que representa os esquemas e os metadados. Esta solução não integra os esquemas, apenas cria um meio de tornar disponível os dados de esquemas que compõem uma federação.

De modo geral, as soluções analisadas acima são fechadas a comunidades de informação preestabelecidas, atendem a aplicações específicas e não contemplam toda a riqueza da representação semântica dos dados geográficos, armazenados pelas bases de dados dos SIGs. Assim este trabalho, a seguir, faz um levantamento das características das bases de dados geográficos pertencentes a diferentes comunidades de informação, define os requisitos da arquitetura, da representação dos esquemas locais e da metodologia, e, finalmente, apresenta a solução proposta.

IV.4. Característica das bases de dados geográficos pertencentes a diferentes comunidades de informação

As bases de dados dos SIGs são heterogêneas, distribuídas e gerenciadas por diferentes comunidades de informação. Elas armazenam dados geográficos que muitas vezes descrevem a mesma região geográfica de perspectivas diferentes, considerando os diversos aspectos temáticos, associados a uma dada realidade de uma área de aplicação, tais como: físico, cultural e econômico.

Isto é consequência dos requisitos de cada comunidade de informação que modela os dados geográficos, de acordo com a semântica necessária à área de aplicação e à finalidade da instituição a qual pertence a informação. Na coleta são empregadas estratégias de amostragem espacial e temporal, com acurácias e precisões diferentes,

bem como podem ser usadas diversas abstrações cartográficas para representar o espaço geográfico amostrado. Além disto, as bases de dados geográficos possuem características decorrentes do ambiente de implementação e da adequação da modelagem dos dados ao aplicativo SIG, utilizado para o acesso, gerência e manipulação dos dados. Isto influencia no formato e na estrutura de armazenamento dos componentes gráficos e não gráficos, dos dados geográficos (STRAUCH *et al.*, 1995).

Assim, a especificação de um esquema para uma base de dados geográficos de uma dada comunidade de informação reflete a área de aplicação, a finalidade, a modelagem da realidade geográfica, o modelo de dados utilizado, a coleta de dados, o aplicativo e a plataforma utilizada.

Desta forma, analisando diferentes bases de dados geográficos de diversas comunidades de informação, averigua-se que elas possuem: *i)* um conhecimento implícito da área de aplicação; *ii)* podem tratar de diferentes regiões geográficas, empregar diferentes parâmetros cartográficos para representar os dados gráficos; *iii)* utilizar critérios variados de qualidade para sua elaboração; e, *iv)* para uma mesma região podem ter seus fenômenos e características descritas de diversas formas. Esta última decorre das bases de dados de SIGs empregarem diversos tipos de construtores de dados e operações oferecidos pela grande variedade de modelos de dados existentes no âmbito da Tecnologia de Geoprocessamento.

Esta informação precisa estar de alguma forma preservada e ao mesmo tempo estar disponível no ambiente MBD. Nesse sentido o modelo de representação dos esquemas locais deve ter mecanismos para armazená-la e torná-la disponível no ambiente MBD.

IV.5. Requisitos de uma arquitetura para interoperabilidade entre SIG

Para o desenvolvimento de um ambiente MBD é necessário selecionar uma arquitetura e estabelecer uma metodologia que integre os esquemas das bases de dados. Desta forma, a autora faz as considerações abaixo, de forma a analisar as arquiteturas estudadas na seção III.6, e definir uma arquitetura MBD para SIG heterogêneos e distribuídos.

A **primeira**, diz respeito às bases de dados que, em geral, são desenvolvidas e mantidas com limitadas oportunidades de modificação, uma vez que re-implementar ou modificar estas bases de dados é uma tarefa, que envolve significativas dificuldades

organizacionais. Isto decorre dos inúmeros tipos de dados e dos elevados investimentos técnicos e financeiros necessários para sua implementação.

A **segunda**, refere-se aos diferentes modelos e estruturas de dados usados, bem como as diferentes linguagens de consulta utilizadas pelos diversos aplicativos SIGs, que requerem para cada sistema treinamento de pessoal específico.

A **terceira**, diz respeito à riqueza semântica do processo de modelagem das bases de dados de SIG, que introduz uma maior heterogeneidade a ser resolvida para proporcionar uma eficiente troca de dados.

A **quarta**, refere-se ao usuário final do SIG, cujo nível de formação é voltado para uma área de aplicação e limitado a algum tipo de aplicativo SIG. Em geral, estes usuários apresentam falta de conhecimento sobre outras áreas com que eles possam interagir. Além disto, deve-se considerar que uma arquitetura que exija conhecimentos mais específicos de computação, para especificar transformação entre os esquemas, suprime a característica dos aplicativos amigáveis, tão desejáveis nos SIGs.

A **quinta**, diz respeito à principal característica do SIG, que é a sua capacidade de análise. Um ambiente MBD para SIG deveria oferecer esta capacidade. A solução ideal seria selecionar um SIG e implementar uma arquitetura sobre este, de modo a torná-lo interoperável com os outros. Todavia, adotar esta solução encontra algumas dificuldades devido a:

- as bases de dados geográficos são projetadas para aplicativos e plataformas específicas. Um único SIG não atende a todos os diferentes ambientes que podem ser encontrados nas comunidades de informação; e
- a passagem de um SIG para outro é extremamente dispendiosa devida a necessidade de conversão de dados e aplicações.

Concluindo, uma arquitetura MBD para SIG deve contemplar os diversos modelos, estruturas de dados, e linguagens de consultas utilizadas pelos aplicativos SIGs, além de preservar as autonomias de projetos, de execução e de comunicação, apresentadas na seção III.3.1.

Estas considerações determinam os requisitos de uma arquitetura para integração das bases de dados dos SIGs existentes e na identificação de mecanismos que podem facilitar esta tarefa.

Assim, considerando estes requisitos e as arquiteturas estudadas na seção III.6 observa-se por um lado, que a arquitetura do esquema global captura as várias semânticas locais em um único esquema, e este suporta uma linguagem de consulta comum a todos os usuários da federação. Desta forma, esta arquitetura suporta as várias

semânticas locais, expressas no esquema global, e encapsula a diversidade de modelos de dados e linguagens de consultas locais dos SIGs, preservando as bases de dados locais.

Esta arquitetura requer um único administrador responsável pela criação e manutenção do esquema global. Este administrador deverá estar capacitado a compreender os esquemas das diferentes comunidades de informação, independente da área de aplicação, e interagir com os administradores, locais de modo a estabelecer a cooperação para a integração dos esquemas. Ele será o único que deverá possuir conhecimentos sobre o acesso as bases de dados dos SIGs locais.

Por outro lado, a arquitetura de mediadores, apesar de suportar várias semânticas, contemplar uma diversidade de modelo de dados e linguagens de consultas, não é adequada para proporcionar a interoperabilidade entre SIGs. Nesta arquitetura se faz necessário um usuário/administrador de dados em cada base de dados. Este determina sua própria compreensão dos sistemas e especifica os relacionamentos e mapeamentos entre os objetos no esquema de exportação. Isto requer do usuário/administrador local conhecimento especializado em vários sistemas para elaborar esquemas externos, necessários a cada comunidade de informação.

Ademais, este administrador deve compreender a semântica dos esquemas das bases de dados locais e estar familiarizado com técnicas para resolver os conflitos semânticos que por ventura existirem, além de navegar entre estruturas de dados para elaborar uma consulta. Isto torna o sistema extremamente complexo, uma vez que existe uma diversidade de modelos, e a semântica dos dados geográficos é extremamente rica, podendo facilmente levar à interpretações erradas. A utilização adequada deste ambiente dependeria de um administrador local, especializado em informática e em vários SIGs, para elaborar esquemas externos necessários a cada comunidade de informação.

Assim, tendo em vista que: *a)* os SIGs são sistemas estáveis, resistentes à evolução dos esquemas; *b)* os usuários de SIG atuarem em para áreas de aplicação específicas; e *c)* disporem de pouco conhecimento de informática; a abordagem selecionada neste trabalho para a interoperabilidade entre os SIG é da arquitetura com o esquema global.

Dentre os benefícios de aplicar esta abordagem para o desenvolvimento de uma arquitetura que contemple bases de dados de SIGs, podem ser citados:

- as várias semânticas dos dados são integradas no esquema global, de forma a facilitar a compreensão e manipulação dos dados na especificação da consulta;

- permite qualquer usuário compartilhar os dados geográficos, de forma transparente, sem precisar conhecer suas localizações e os esquemas locais, quanto a modelagem e estruturas locais utilizadas; e
- isola os fornecedores de dados do sistema. Se por ventura houver mudanças locais, implicará em mudanças apenas no esquema global.

Além disto, esta arquitetura não impõe obrigações aos administradores locais para manter o esquema global. A responsabilidade de manter a integração pertence ao administrador da federação.

IV.6. Requisitos de uma metodologia para integração de esquemas de bases de dados geográficos

As questões relacionadas a semântica e sintaxe dos dados geográficos são, geralmente, bem compreendidas dentro de um domínio específico de uma base de dados. Todavia, para construir um ambiente MBD que proporciona uma federação de dados, estas questões necessitam ser também entendidas pelo administrador da federação e estar disponível para os usuários da federação, de forma que estes compreendam a limitação dos dados e sua conveniência de uso nos vários contextos das aplicações.

Desta forma, para o desenvolvimento de um ambiente MBD não basta definir uma arquitetura e um modelo de dados para representar os esquemas. É necessária uma metodologia para integração de esquemas, e no caso de base de dados geográficos, esta metodologia deve contemplar as características dos componentes gráficos e não gráficos, de diferentes comunidades de informação.

Esta metodologia consiste em um processo de modelagem que estabelece as atividades para derivar um esquema conceitual a partir dos esquemas de exportação das diferentes comunidades de informação. Estas atividades têm por objetivo identificar os dados geográficos que apresentam proximidade semântica, bem como os conflitos e inconsistências que por ventura existirem, aplicar técnicas para resolver ou atenuar estes conflitos e integrar os esquemas locais.

Na verdade, a integração de bases de dados geográficos herda todos os conflitos semânticos e estruturais das bases de dados convencionais, heterogêneas e distribuídas, e a estes são adicionados os conflitos decorrentes da componente gráfica dos dados geográficos. Para tratar adequadamente estes conflitos, a metodologia deve considerar, além dos aspectos da modelagem referentes a representação da semântica, as afinidades

entre as aplicações e os conceitos de espaço associados a cada comunidade de informação.

A semântica, neste contexto, é denominada por BUEHLER & MCKEE^{html} (1996) como interpretação de seu conteúdo, e está associada a cada comunidade de informação. A correta compreensão desta semântica capacita um conjunto de dados de uma comunidade de informação a ser efetivamente usado por outras comunidades de informação. Ela é capturada pela descrição do conjunto de informações relacionadas ao esquema e à descrição da coleção de dados, da base de dados do SIG, também denominado de metadados.

A metodologia deve estabelecer uma forma padronizada de descrever os metadados, de maneira a captar a semântica dos esquemas das bases de dados e as características, mesmo as implícitas, de cada sistema existente. Esta descrição dos metadados tem por objetivo normalizar as diferenças entre os modelos e captar a informação semântica.

Assim, os metadados do esquema da base de dados geográficos devem descrever: as informações relacionadas a localização do esquema na rede; as características do aplicativo e o tipo de plataforma; as características cartográficas associadas ao esquema, tais como: *data* geodésicos, escala, sistema de coordenadas, e sistema de projeção utilizados; e, ainda, a própria coleção de objetos. Para esta última devem ser descritos para cada objeto geográfico, além das suas propriedades temáticas e os seus relacionamentos com outros objetos, o seu significado, as propriedades relacionadas às componentes gráficas, tais como o tipo, formato, método de coleta, data de coleta, qualidade, unidades e semiologia gráfica, usada para representar os objetos geográficos.

Entretanto, cabe ressaltar que, de modo geral, os aplicativos de SIG não possuem mecanismos para descrever os metadados e partes destes estão implícitos nas bases de dados, podendo ser estes inferidos de outras fontes. Ademais, há uma diversidade de modelos de dados utilizados para tratar os dados geográficos de SIG.

Desta forma, a metodologia para integração de bases de dados de SIG deve adotar um modelo de dados comum (MDC) que capture as particularidades lingüísticas/culturais da informação geográfica e suporte todas as conceitualizações dos dados geográficos, armazenadas pelas bases de dados dos SIGs, de cada comunidade de informação. A tradução dos esquemas das bases de dados geográficos para este modelo facilita a identificação dos conflitos entre os esquemas (VEMEER & APERS, 1996).

Na metodologia para integração de esquemas, o MDC tem a finalidade de normalizar as diferenças entre os modelos das bases de objetos participantes da

federação. Este modelo deve proporcionar, ao desenvolvedor das etapas da metodologia, uma correta compreensão dos dados quanto a sua estrutura, o domínio, a aplicação e a terminologia usada para as diferentes representações, nos esquemas das bases de dados projetadas independentemente. Em suma, ele deverá permitir a visualização das diferenças semânticas e estruturais entre os esquemas e ajudar a estabelecer relacionamentos entre objetos com diferentes semânticas, relacionadas a mesma região geográfica.

A metodologia deve oferecer etapas para: *i)* a tradução dos esquemas descritos nos modelos locais para o MDC; *ii)* para interpretar os dados representados no esquema quanto ao seu significado, sua estrutura e seu nível de abstração, referência espaço-temporal; *iii)* para resolver os conflitos entre os dados; e *iv)* integrar os esquemas de exportação.

IV.7. Requisitos de representação dos esquemas para bases de dados geográficos

Uma metodologia de integração de esquemas que suporte uma arquitetura do esquema global requer três níveis de representação dos esquemas, conforme apresentado na seção anterior: os **metadados** necessários à compreensão dos esquemas de exportação das bases de dados geográficos, um **MDC** para descrever o esquema de exportação e o esquema resultante da integração, denominado de **esquema global**.

Assim, esta seção descreve a seguir os requisitos dos metadados, do MDC e do esquema global para assistir uma metodologia para integração de bases de dados geográficos.

IV.7.1. Requisitos dos metadados para um ambiente MBD de SIG

Os metadados são empregados no desenvolvimento de um ambiente MBD para atender uma metodologia para integração de esquemas e execução da consulta neste ambiente. No processo de integração, os metadados capturam as informações necessárias para subsidiar o processo de integração e construção do esquema global. Na execução da consulta os metadados fornecem subsídios para localizar informações sobre as estruturas dos dados solicitados.

A metodologia deve especificar um conjunto mínimo de metadados que descrevam as bases de dados participantes da federação, de modo a permitir uma

compreensão de seus esquemas e possibilitar a interoperabilidade lógica e física. Esta especificação deve considerar os aspectos da modelagem espaço-temporal, o conteúdo semântico dos dados geográficos e aspectos relativos a qualidade dos dados de cada comunidade de informação. Assim, este trabalho considera os seguintes tipos de metadados apresentados em BUNEMAN *et al.* (1997):

- *Metadados estruturais* - tratam das estruturas para representar os dados geográficos nas bases participantes da federação. Eles são utilizados pela metodologia de integração para análise sintática dos dados armazenados e pelo tratador de resultados para integrar os dados retornados do processo de consulta;
- *Metadados semânticos* - fornecem significados aos dados armazenados pelas estruturas de dados mencionadas acima. Estes metadados especificam a informação sobre o tipo semântico e o relacionamento entre os tipos. Eles ajudam a encontrar o grau de proximidade semântica entre os esquemas;
- *Metadados relacionados à localização da base de dados* - descrevem a localização da base de dados, quanto a localização na rede e informações sobre os aplicativos utilizados para o seu gerenciamento;
- *Metadados históricos* - referem-se à informação genérica sobre a origem dos dados. Devido à necessidade de considerar os aspectos relativos a qualidade dos dados, estes metadados são sub-classificados em:
 - ◆ *Atualidade da base de dados* - descrevem como os dados armazenados nas bases de dados, variam sobre o tempo e se estes são atualizados. Eles são importantes para avaliar o grau de atualização da base de dados; e
 - ◆ *Confiabilidade da base de dados* - descrevem a proveniência dos dados gráficos e não gráficos. Eles auxiliam o usuário a determinar a confiabilidade da fonte de dados para fornecer certos tipos de informações.

Entretanto, quando trata-se de dados geográficos, os metadados fornecem ainda subsídios para consolidar os resultados provenientes das bases de dados locais e fornecer informações para os usuários analisarem e interpretarem os resultados. Estes metadados, denominados no presente trabalho de metadados de representação, têm por objetivo controlar a aparência dos temas e gerenciar meta informação sobre a apresentação dos dados. Eles contêm as seguintes informações: escala, identificadores de temas (atributos descritivos) e semiologia gráfica para representar os objetos.

IV.7.2. Requisitos de representação do MDC para esquemas de bases de dados geográficos

Conforme observado na seção III.4.3., a metodologia para integração de esquemas deve adotar para o MDC, um modelo de dados semanticamente rico para capturar as similaridades e diferenças semânticas e sintáticas das comunidades de informação, armazenadas por diferentes modelos de dados geográficos. Contudo, um modelo de dados para descrever os esquemas das bases de dados de SIG, deve ter componentes espaço-temporais, semânticos, estruturais e de qualidade dos dados. Os componentes espaciais tratam da geometria e a referência espaço-temporal, definida pela projeção, sistema de coordenadas e transformações permitidas. Os componentes semânticos definem o significado de elementos de objeto, em termos de modelo do mundo real. O componente estrutural trata de uma ampla variedade de estruturas utilizadas pelas bases de dados geográficos. O componente da qualidade descreve o histórico e qualidade dos dados armazenados pelo esquema. Este último, adicionado à descrição do esquema, determina a influência semântica em relação a outros esquemas durante o processo de comparação.

O MDC para dados geográficos também deve oferecer expressividade e relativismo semântico, de forma a capacitar os dados a serem modelados em um domínio específico. Esta modelagem deve ser extensível de forma a suportar extensões definidas pelo usuário para representar qualquer fenômeno ou feição da superfície terrestre, em termos de propriedades e relacionamentos. Além disto, ele não deve estar preso aos ambientes locais (plataformas/aplicativos) ou a uma arquitetura de comunicação (SAIF, 1995, BUEHLER & MCKEE^{html}, 1996).

Na literatura são encontradas algumas propostas do modelo de dados para aplicações SIG, mas infelizmente eles não trazem soluções para as questões de múltiplas representações dos planos de informação e aspectos cartográficos. Um exemplo é o modelo de dados proposto por DBOUK *et al.* (1996) baseado na estrutura de objetos. Neste modelo, qualquer entidade ou objeto em um universo observado pode ter um conjunto de atributos descritivos. O objeto geográfico é formalmente o conjunto que capacita o processamento relacional de objetos (seleção com critério quantitativo). Esta estrutura é considerada como um esquema geral que contém dois componentes:

- um conjunto de representação de objetos do mundo real; e
- um conjunto de ligações explícitas que expressam relacionamentos de composição lógica e semântico entre as entidades.

IV.7.3. Requisitos do esquema global para dados geográficos

O esquema global é parte central da abordagem selecionada para o ambiente MBD proposto. O esquema global esconde dos usuários a distribuição dos dados e as diferenças entre as linguagens de consulta, estrutura de dados e formatos de arquivos nos SIGs locais.

Ele é a base para o compartilhamento e troca de dados, proporcionando transparência sobre o conteúdo semântico dos objetos geográficos e uniformidade para consulta. Este esquema é apresentado na interface do ambiente MBD. Sobre esta interface o usuário: *a)* visualiza a informação disponível no esquema da federação; *b)* estabelece relacionamentos entre objetos com diferentes semânticas, relacionadas à mesma região geográfica; e *c)* realiza consultas, sem precisar conhecer ou compreender a organização original e a distribuição dos dados nos SIGs locais. Sobre este esquema devem ser oferecidos serviços. Dentre estes serviços pode estar uma linguagem de consulta, não necessariamente uma linguagem textual, mas um meio para o usuário expressar uma consulta.

Este esquema resulta da aplicação da metodologia para integração dos esquemas das bases de dados componentes de uma federação. Ele contém uma visão lógica, consistente da informação geográfica de modo a permitir, ao usuário, compreender o exato significado dos objetos geográficos. Cada definição de objeto neste esquema compreende todas as representações incompletas e heterogêneas entre os esquemas, considerando as diversas semânticas dos dados e as diferentes representações para um mesmo fenômeno geográfico. Assim, um modelo de dados geográficos para um esquema global deve ter as seguintes características:

- oferecer formalidade para representar de forma não ambígua os dados disponíveis na federação;
- refletir o universo conceitual disponível na federação, através de conceitos mais próximos de uma realidade comum;
- esconder os detalhes de representação dos dados geográficos, nas bases de dados locais, e de implementação do acesso a estas bases; e
- minimizar a complexidade envolvida nos diferentes tipos de representação geométrica.

Por outro lado, o esquema global deve ter um modelo de dados que permita a representação uniforme dos objetos. Uma forma de alcançar este objetivo é através de hierarquias de especialização/generalização de conceitos. Esta hierarquia pode ser absoluta, isto é, seja válida para todos os casos, ou relativa quando associada a um

determinado contexto. Ele deve oferecer mecanismos para estabelecer os mapeamentos entre os conceitos apresentados no esquema global e os esquemas das bases de dados componentes do sistema, solucionando aspectos relativos a heterogeneidade.

Tendo em vista as características acima e as necessidades de um modelo de dados para MDC, exploradas na seção III.4.3., o paradigma orientado a objetos também é adequado a modelagem abstrata dos dados geográficos. A adoção de um modelo de dados orientado a objeto, para o esquema global, permite a aplicação de técnicas deste paradigma para a criação do esquema global. Ele fornece mecanismos para armazenar e gerenciar dados abstratos (objetos geográficos) representados por um conjunto de classes. Estas classes podem encapsular os métodos que implementam os mapeamentos entre os esquemas locais e o esquema global, e operações distribuídas de acesso aos dados nos esquemas locais. Além disto, um modelo orientado a objeto é extensível, de modo que novos tipos de dados e métodos possam ser acomodados sem modificar o sistema existente.

Todavia, cabe ressaltar que o modelo proposto para o esquema global deve ser simples, de modo que qualquer usuário acostumado com um determinado ambiente SIG possa compreendê-lo e realizar consultas, sem precisar conhecer a localização ou a estrutura dos dados locais.

IV.8. Proposta de um ambiente MBD: Sistema MultiSIG

Este trabalho tem por objetivo desenvolver o **Sistema MultiSIG**, que oferece um ambiente MBD no qual as bases de dados dos SIGs formam uma federação de objetos geográficos. Esta federação deve ser aberta, permitindo que a qualquer momento novas comunidades de informação possam participar, ou sair da federação, através de negociações estabelecidas entre os administradores locais e o administrador da federação.

O ambiente MBD proposto pelo **Sistema MultiSIG** proporciona interoperabilidade semântica entre bases de dados geográficos, usando um esquema global. Assim, para o seu desenvolvimento são propostas uma metodologia e uma arquitetura denominadas, respectivamente, **MultiSIG** e **MultiSIG**.

A **MultiSIG** fornece suporte para a obtenção da interoperabilidade semântica no nível lógico. Isto é alcançado através da integração dos esquemas de exportações das bases de dados locais cadastrados na federação. Esta metodologia especifica um conjunto de atividades para normalizar os esquemas locais, analisá-los, aplicar soluções

para resolver os conflitos entre eles e derivar um esquema global consistente. Para isto o MDC adotado deve poder ser estendido para suportar as várias áreas de aplicações. Este MDC deve atender aos seguintes requisitos:

- ser independente de plataformas e aplicativos de SIG;
- descrever as características das aplicações locais;
- descrever sistemas de referência espaço-temporal utilizados; e
- descrever os objetos dos esquemas quanto à extensão espacial, dimensão temporal, propriedades e relacionamentos.

Cabe ressaltar que o MDC utilizado pela *MultiSIG* deve suportar o corrente estado da arte da tecnologia de Geoprocessamento no que diz respeito aos paradigmas de modelagem de dados geográficos.

A *MultiSIG* implementa a interoperabilidade física, através de componentes que tornam disponível o esquema global. Ela permite aos usuários submeter uma consulta que será distribuída, seus resultados recuperados corretamente nas bases de dados dos SIGs e integrados para apresentá-los ao usuário da federação. Com estes componentes, a *MultiSIG* tem por objetivo esconder a heterogeneidade entre os sistemas e aplicações geográficas, bem como evitar mudanças locais nos SIGs componentes da federação. Assim, esta arquitetura tem como requisitos:

- preservar a autonomia das bases de dados locais quanto à estrutura e semântica dos dados geográficos;
- ser independente do sistema operacional, da plataforma, do aplicativo SIG e da linguagem de consulta local;
- utilizar uma arquitetura cliente-servidor que opere com protocolos de comunicação distribuída consistentes com os SIGs.

Esta arquitetura deverá ainda adotar uma interface intuitiva e flexível, acessível pela rede e comum a todos os usuários da federação, de forma a promover uniformidade de apresentação do **Sistema MultiSIG** sobre cada plataforma.

Os capítulos a seguir apresentarão a *MultiSIG* e a *MultiSIG*.

Capítulo V - MultiSIG: Metodologia proposta para integração de esquemas de bases de dados geográficos

V.1. Introdução

A MultiSIG é uma metodologia que tem por objetivo suportar a integração das diversas representações de dados encontradas em SIGs independentes que precisam de interoperação. Para isto a MultiSIG integra os esquemas das bases de dados geográficos gerando um esquema global.

A MultiSIG especifica as atividades necessárias para gerar os esquemas de exportação na federação, analisá-los de forma a identificar os conflitos semânticos, resolvê-los e derivar o esquema global.

Tendo em vista que os esquemas locais, em geral, apresentam-se em modelos de dados diferentes, usando uma terminologia própria da área de aplicação, a MultiSIG, para desenvolver suas atividades corretamente, utiliza: *i)* um MDC para representar os esquemas de exportação; *ii)* um conjunto mínimo de metadados necessários para descrever os esquemas locais; e *iii)* um vocabulário para adequadamente expressar a semântica dos objetos geográficos e suas propriedades.

Visando facilitar o desenvolvimento de suas atividades de análise, a MultiSIG propõe ainda uma política de participação na federação MultiSIG e a utilização de conceitos relacionados aos dados geográficos. A política de participação determina os procedimentos para os administradores negociarem a participação dos esquemas na federação MultiSIG e estabelecerem cooperação, de modo a criarem um conjunto de definições e termos comuns entre os esquemas. Os conceitos visam facilitar a compreensão dos esquemas locais e auxiliar a estabelecer uma acurada estrutura de correlação entre os objetos geográficos armazenados pelos esquemas locais.

Assim, este capítulo apresenta a MultiSIG. Inicialmente são expostos os seus objetivos, são descritas as principais características da metodologia, as entradas e saídas

da metodologia, a política de participação proposta para a federação MultiSIG, os elementos e os conceitos principais utilizados pela metodologia, e finalmente as suas atividades distribuídas em seis etapas. Ao final, este capítulo apresenta uma comparação da *MultiSIG* com outros trabalhos relacionados na literatura.

V.2. Objetivo da *MultiSIG*

A *MultiSIG* tem por objetivo suportar o processo de integração e criar um esquema global, a partir dos esquemas locais das bases de dados geográficos participantes da federação, contemplando os requisitos levantados no capítulo IV, seção IV.6. Para isto, a *MultiSIG* contempla as diferenças entre os modelos de dados dos SIGs e as diferenças entre os esquemas decorrentes da modelagem das diferentes áreas de aplicação.

O objetivo é inicialmente alcançado através da conversão dos esquemas locais para um modelo de dados comum (MDC) orientado a objeto. Esta conversão tem por finalidade homogeneizar a representação dos esquemas locais e facilitar a identificação dos conflitos entre eles.

O resultado final é realizado através da especificação detalhada do conjunto de atividades necessárias para interpretar os dados geográficos, representados nos esquemas de exportação, e de soluções para resolver os conflitos e criar o esquema global. Desta forma, o esquema global leva em consideração além dos aspectos da modelagem referentes às representações semânticas, as afinidades entre as áreas de aplicações e os conceitos de espaço associado a cada comunidade de informação.

V.3. Características da metodologia proposta

A principal característica da *MultiSIG* é a ênfase dada às atividades de análise que tratam dos aspectos relativos à descrição do espaço geográfico, à descrição da área de aplicação e à semântica dos esquemas, no nível extencional e intencional. Esta característica é um dos diferenciais da *MultiSIG* em relação as outras metodologias encontradas na literatura, discutidas na seção IV.11.

A análise extencional é realizada de modo n-ário sobre os atributos implícitos do esquema. Esta análise contempla o contexto espacial quanto à localização geográfica, os

parâmetros cartográficos implícitos na representação cartográfica do esquema e as heterogeneidades e similaridades entre as áreas de aplicação. Esta última comparação determina quais esquemas possuem maior proximidade semântica e estabelece a seqüência dos pares de esquemas a serem integrados. A análise intencional é realizada de modo binário sobre as especificações das classes dos pares de esquemas de exportação, expressos no MDC selecionado. Esta análise estabelece atividades, de modo a determinar uma acurada estrutura de correlação que determina uma proximidade semântica entre elementos dos esquemas. Isto evita comparar objetos geográficos inconsistentes. Esta análise resulta em uma série de conflitos, devido às heterogeneidades semânticas que são tratadas pelas etapas seguintes da *MultiSIG*.

A *MultiSIG* para facilitar a identificação da proximidade semântica entre as classes dos esquemas propõe duas taxionomias. A **primeira**, denominada de **taxionomia de conflitos**, visa facilitar a identificação dos conflitos entre os esquemas. Esta taxionomia considera os conflitos relacionados à definição do contexto espacial do esquema local, os conflitos semânticos, os conflitos estruturais e os conflitos entre os valores dos dados. Os três primeiros tipos de conflitos são detectados e tratados pela *MultiSIG*. Os últimos são detectados e tratados pelas operações que elaboram o resultado das consultas, a ser apresentado no próximo capítulo. A **segunda**, denominada de **taxionomia de proximidade semântica**, visa facilitar o reconhecimento dos objetos geográficos que representam conceitos similares ou relacionados.

Cabe ressaltar ainda, que estas atividades são desenvolvidas, atualmente, de forma manual, entretanto, futuramente poderão ser desenvolvidas ferramentas para automatizar as etapas da metodologia proposta.

V.4. Entradas e saídas da metodologia

Segundo a classificação proposta por SHETH & LARSON (1990), a metodologia se enquadra na categoria de representação das bases de dados em quatro níveis. Esses quatro níveis representam diferentes situações de integração, a saber: o esquema local, o esquema de exportação, o esquema global interno e o esquema global externo (Figura V.1).

O esquema local representa o esquema existente em cada base de dados geográficos dos SIGs, componentes da federação. Ele é expresso no modelo de dados do SIG local e usado pelas aplicações locais.

Os esquemas de exportação correspondem à descrição de todo ou parte dos esquemas locais que farão parte da federação, expressos no MDC selecionado pela MultiSIG. Estes esquemas restringem o grau com que as bases de dados participam da federação. Cabe ao administrador local selecionar todo/parte do seu esquema e cadastrá-lo na federação. Os esquemas de exportação são armazenados pelo Sistema MultiSIG para serem analisados pelas etapas da metodologia proposta.

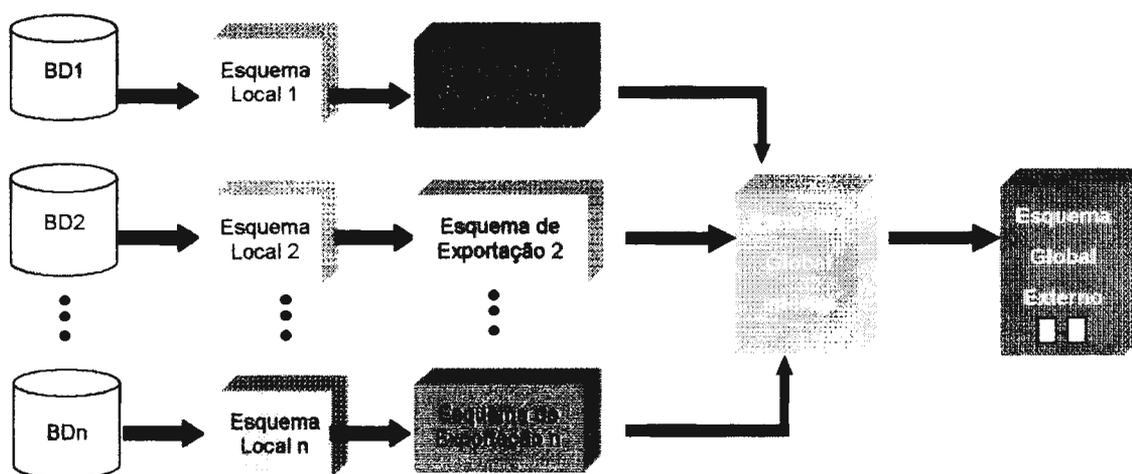


Figura V.1- Esquemas utilizados pela MultiSIG.

O esquema global interno representa a visão integrada de todos os esquemas das bases de dados componentes da federação, expressas no MDC. Neste modelo os objetos são resultantes das soluções aplicadas aos conflitos semânticos e sintáticos, entre os esquemas, devido às heterogeneidades entre os modelos de dados e a modelagem das bases de dados dos SIGs componentes da federação.

Para facilitar a compreensão do esquema global interno a MultiSIG sugere descrever este esquema em um esquema global externo. O esquema global externo foi criado em analogia ao nível externo de representação da tradicional arquitetura em três níveis da ANSI/SPARC. A idéia é permitir que usuários que trabalham sobre temas ou regiões identifiquem facilmente os dados requisitados na federação. Este esquema corresponde ao esquema global e estará disponível na interface com o usuário. Sobre este esquema é que os usuários da federação realizarão as consultas no ambiente MultiSIG.

Em geral as metodologias param no esquema global interno. Entretanto, observa-se que o conceito de temas sobre os quais os usuários costumam trabalhar, fica diluído e difícil de ser identificado no esquema tradicional de integração.

A metodologia tem, como entrada, os esquemas locais. Estes esquemas são traduzidos para o MDC selecionado, gerando o esquema de exportação. A seguir os esquemas de exportação são analisados e comparados aos demais, de acordo com a seqüência de atividades estabelecidas pelas etapas da *MultiSIG* para encontrar as similaridades e diferenças entre os esquemas. A partir desta análise, os esquemas de exportação são integrados no esquema global interno. A *MultiSIG* mapeia este esquema global interno no modelo de dados temático, criando o esquema global externo da federação.

V.5. Política de participação no Sistema MultiSIG

Tendo em vista que: *i)* a criação de um esquema global requer um processo de julgamento e validação entre os administradores locais e o administrador da federação, de forma a definir uma terminologia comum a todos os usuários da federação, e que *ii)* o **Sistema MultiSIG** propõe uma federação aberta, permitindo a qualquer momento novas comunidades de informação participarem ou saírem da federação, se faz necessário estabelecer uma política de participação.

Uma metodologia para integração requer a participação dos administradores locais para definir um consenso na integração dos seus esquemas. A *MultiSIG* para alcançar seus objetivos desenvolve uma política de participação que determina um fluxo de trabalho, usando procedimentos preestabelecidos para os administradores combinarem os esquemas de exportação na federação; cooperarem nas atividades relacionadas ao estabelecimento de soluções para os conflitos encontrados e na aplicação das abstrações da orientação a objeto, adequadas a integrar os objetos geográficos no esquema global. Ao final de cada etapa, o administrador da federação estabelece a comunicação com os administradores locais, através destes procedimentos, de modo a consolidar suas análises.

Esta política de participação define um conjunto de procedimentos que caracterizam a existência da federação e garantem a sua estrutura. Estes procedimentos estabelecem os mecanismos para proporcionar a cooperação entre o administrador da

federação e os administradores locais. A cooperação, neste contexto, consiste em ações a serem realizadas, de modo a possibilitar que os administradores negociem o compartilhamento dos dados e obtenham sucesso no processo de integração de esquemas.

Todavia, para definir estes procedimentos, se faz necessário, inicialmente, definir os papéis a serem representados pelos usuários da federação, a saber: o administrador da federação, o administrador local e o usuário leitor da federação. Para cada papel, a MultiSIG estabelece os seguintes níveis de segurança da federação:

- administrador da federação ou global - tem a responsabilidade de criar, alterar e manter o esquema global, estabelecer a comunicação com os administradores locais e implementar as funções de acesso e mapeamento para os esquemas locais;
- administrador local - é responsável pelo cadastro do esquema de exportação na federação. Ele deve conhecer o esquema local, de forma a fornecer informações ao administrador da federação para a integração e ter disponibilidade para trocar informações com os outros administradores locais. A MultiSIG preconiza que os administradores locais e o administrador global estabeleçam um consenso entre as semânticas representadas nos esquemas locais, de modo a definir uma terminologia a ser usada no esquema global externo; e
- usuário leitor - é o usuário da federação, o qual poderá realizar consultas sobre o esquema global.

A cooperação no ambiente MultiSIG estabelece três procedimentos administrativos para a participação de uma base de dados geográficos na federação, proporcionada pelo ambiente MultiSIG: *procedimentos de participação, procedimentos de cadastro e procedimentos de cooperação*.

O *procedimento de participação* tem por objetivo permitir que um administrador local manifeste o desejo de participar da federação proposta pelo **Sistema MultiSIG**. Para isto, este procedimento estabelece a comunicação entre o administrador da federação e o administrador local. O administrador local, através de navegação no sistema, conhece os objetivos da federação e encaminha ao administrador da federação o desejo de participar da federação.

Tendo em vista que o administrador local deverá concordar com o trabalho em cooperação com outros administradores, a fim de negociarem o relacionamento dos seus esquemas com os dos outros. Será formalizada a participação do administrador local

através de um termo de compromisso com a federação. Neste termo, o administrador local se compromete a: *i)* assegurar a qualidade dos dados; *ii)* se responsabilizar pela veracidade das informações para integração; e *iii)* concordar em trabalhar em cooperação para integrar o esquema local no esquema global.

O *procedimento de cadastro* tem por objetivo permitir que os administradores locais descrevam seus respectivos esquemas locais, segundo a metodologia, no MDC selecionado, e forneçam as informações necessárias para a integração. Estes procedimentos devem ser implementados pelos administradores locais, uma vez que estes compreendem todos aspectos semânticos subjetivos e objetivos, armazenados pelo esquema local. Além disto, os administradores locais se responsabilizam pelo o grau de transparência dos dados locais.

Ao cadastrarem os esquemas na federação, os administradores locais traduzem toda/parte da semântica e sintaxe dos esquemas das bases de dados geográficos que desejam compartilhar, dos seus modelos de dados locais para o esquema de exportação.

A metodologia, visando facilitar a comparação dos esquemas, propõe a utilização de um *Thesaurus* com a finalidade de fornecer um vocabulário padronizado de termos. O *Thesaurus* fornece uma terminologia para os usuários definirem as classes de objetos e suas propriedades. Os termos são selecionados pelos administradores locais, de maneira a refletir a semântica associada ao objeto geográfico e representar suas propriedades cadastradas. Estes termos capturam a informação semântica que não pode ser expressa pelas definições sintáticas, usadas pelo modelo da base de dados.

Cabe ressaltar que o conjunto destes termos reflete o sistema conceitual de um domínio de aplicação de cada esquema e permite criar um inventário de termos do domínio da aplicação que serão utilizados, posteriormente, pela metodologia, para estabelecer a equivalência entre os elementos do esquema.

Este cadastro é armazenado em um dicionário global, utilizado pelo administrador da federação para tomar ciência de quais dados estão disponíveis, seus significados e as estruturas locais utilizadas para armazená-los. Futuramente, este cadastro deverá estar disponível às outras comunidades de informação.

Na verdade, este procedimento permite que cada administrador local gerencie a participação da sua base de dados local na federação, através da inserção, alteração ou remoção de seu esquema da federação. A inserção e alteração de esquemas iniciarão um novo processo de cooperação, de forma a estabelecerem como o novo esquema será integrado a federação ou como a alteração se refletirá no esquema global,

respectivamente. A remoção do esquema implicará no processo a ser realizado pelo administrador global da federação. Este processo consiste em remover do esquema global externo todos os mapeamentos para os objetos do esquema local removido.

O *procedimento de cooperação* consiste na operação que tem por finalidade auxiliar o administrador da federação a aplicar a metodologia para construção manual dos mapeamentos semânticos entre os esquemas, em parceria com os administradores locais.

Através das atividades propostas pela *MultiSIG*, o administrador da federação determina: *a)* quais esquemas de exportação apresentam o mesmo contexto territorial da informação; *b)* quais esquemas de exportação apresentam maior proximidade semântica de áreas de aplicação; e *c)* quais os administradores locais devem estabelecer um fluxo de trabalho, de forma a definirem os mapeamentos para os objetos, para os quais é identificada alguma proximidade semântica.

Através dos mecanismos propostos pela arquitetura *MultiSIG*, os administradores locais comunicam-se de forma a estabelecer a concordância quanto ao significado de cada elemento. Os administradores locais devem determinar interpretações dos termos que apresentam similaridades e diferenças (conflitos) e estabelecer definições comuns. Este procedimento cooperativo resulta em uma série de definições semânticas e funções de mapeamento entre os esquemas de exportação e o esquema global.

Resumindo, iniciado o procedimento de participação, o administrador da federação fornece permissão de cadastro no **Sistema MultiSIG**. O administrador local toma conhecimento das regras de tradução, seleciona todo/parte do seu esquema, e realiza o procedimento de cadastro do esquema de exportação. O administrador da federação através das etapas da *MultiSIG* analisa os esquemas cadastrados e estabelece o fluxo de cooperação entre os administradores locais.

V.6. Elementos utilizados pela metodologia

A *MultiSIG* para alcançar a integração de esquemas adota um modelo de dados comum (MDC) para o esquema global interno, especifica os metadados necessários para descrever os esquemas de exportação, utiliza um *thesaurus* para

eliminar a subjetividade da descrição semântica, e propõe um modelo de dados para o esquema global externo sobre o qual os esquemas das bases de dados serão integrados.

V.6.1. Modelo de dados do esquema global interno

A utilização de um MDC possibilita que o administrador do ambiente MultiSIG obtenha familiaridade com os esquemas das bases de dados dos SIGs, de maneira a compreender e interpretar informação semântica, bem como identificar as similaridades e conflitos que por ventura existam. Isto requer uma forma apropriada para descrever os metadados dos esquemas das bases de dados geográficos.

No estabelecimento de um MDC para a metodologia proposta foi levado em consideração que as informações relativas à área de aplicação, à finalidade e às questões cartográficas (sistema de referência, sistemas de projeção), estão, em geral, implícitas no esquema e na aplicação. Assim, a MultiSIG selecionou um modelo capaz de representar os metadados explícitos e os que são derivados de dados privados ou inferidos de outras fontes.

A estratégia adotada para descrever os esquemas das bases de dados dos SIGs e informações relevantes para a integração não contidas nestes esquemas foi a adoção de um padrão de dados para objetos geográficos. Os padrões fornecidos pelos esforços de padronização, comentados na seção IV.2, em geral, cobrem vários modelos de dados, descrevem parâmetros para avaliação dos dados, requisitos dos dados, restrições de integridade, formatos de arquivos, procedimentos administrativos, plataforma e aplicativo para implementação.

O padrão selecionado foi o metamodelo SAIF (SAIF, 1995). Trata-se de um padrão de dados neutro, independente de plataforma/aplicativo e de área de aplicação. Este padrão foi proposto com o objetivo de padronizar um modelo de dados de modo que todos os dados, em um ambiente aberto, possam ser descritos e armazenados de modo coerente.

O padrão de dados SAIF usa um modelo orientado a objeto, que oferece um conjunto de classes básicas para descrever o conteúdo das bases de dados e estas podem ser especializadas pelos usuários para descrever seus esquemas locais. Este modelo é centralizado no conceito representado pela classe *GeographicObject*, qual possui uma hierarquia de classes que oferece uma linguagem para descrever, no espaço e no tempo, entidades e fenômenos que tem uma localização e extensão espacial. As classes

permitem representar os dados geográficos de acordo com as estruturas utilizadas pela tecnologia dos SIGs (STRAUCH, SOUZA & MATTOSO, 1998; SAIF, 1995).

Neste modelo, um objeto do tipo *GeographicObject* possui três componentes (Figura V.2): *GeometricObject*, *Metadata* e *Relationship*. O componente *GeometricObject* é definido por classes que definem primitivas geométricas de pontos, linhas, áreas, volumes e imagens. O componente *Metadata* é definido por classes que armazenam os parâmetros para descrever as referências espaço-temporal, a localização geográfica, a qualidade, o histórico, os aspectos administrativos, os procedimentos de atualização e os produtos gerados. O componente *Relationship* define os relacionamentos entre os objetos através de uma hierarquia de classes que expressa as associações espaciais, topológicas e temporais.

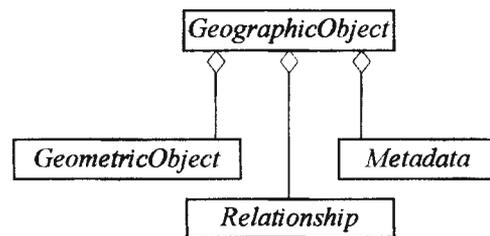


Figura V.2 - Definição da classe *GeographicObject* do SAIF.

O metamodelo SAIF não suporta herança múltipla diretamente, contudo, os objetos geográficos podem ser agregados em objetos geográficos compostos, isto é, o SAIF aceita a recursão do objeto geográfico de modo que objetos possam ser compostos por suas subclasses. O metamodelo SAIF não oferece métodos, entretanto, ele permite a especificação destes pelos usuários para incluir operações ou para definir comportamentos específicos.

O SAIF possui também uma linguagem denominada *SAIF Talk*, apresentada no apêndice B, que oferece uma estrutura formal para expressar a definição de classes e objetos. Ela é baseada em texto e apresenta dois componentes: *Class Syntax Notation* (CSN) e *Object Syntax Notation* (OSN). A primeira é utilizada para descrever classes e a segunda utilizada para construir representações legíveis de instâncias de objetos que são descritas pelo metamodelo de dados do SAIF.

O SAIF foi analisado em relação aos modelos apresentados na seção IV.2 e adotado como MDC, uma vez que seu modelo oferece classes distribuídas em três níveis de abstração. Isto permite, além de descrever a sintaxe do esquema, capturar parte da semântica de cada esquema. Dentre as vantagens de se adotar este padrão como

MDC, cabe destacar que ele cobre várias estruturas de SIG para representar os objetos geográficos e pode ser estendido para qualquer área de aplicação, independente do ambiente e da plataforma em que se encontram os dados. Ele suporta a descrição do conteúdo temático e espaço temporal, captura o domínio dos metadados, incluindo associação de entidades, cardinalidade, dependência de existência, restrições e nomes dos usuários. A sua notação de classes possui cláusulas que permitem descrever restrições (cláusulas *restricted* e *constraints*), associar valores a classes, subclasses e as suas propriedades (cláusulas *default* e *ClassAttributeDefault*), expor o domínio das propriedades das classes (cláusula *ClassAttributeValue*), e expressar informações que comentam o significado das classes e suas propriedades (cláusula *comments*). Além destas vantagens, este padrão pode evoluir com a tecnologia.

Assim, os esquemas das bases de dados participantes da federação são traduzidos para o modelo de dados do padrão SAIF, através do procedimento de cadastro do esquema na federação. Este procedimento é realizado de acordo com a atividade estabelecida por uma das etapas da *MultiSIG*.

V.6.2. Metadados utilizados pela metodologia

Conforme apresentado no capítulo três e reiterado no capítulo quatro, os metadados são de relevante importância no desenvolvimento do ambiente MBD, auxiliando o processo de integração e construção do esquema global. Os metadados consistem em um conjunto mínimo de informações necessárias às atividades de integração da *MultiSIG*. Neste trabalho, os metadados além de descreverem o esquema, descrevem informações sobre a localização das bases de dados na rede, as estruturas de dados usadas, o significado dos dados armazenados por estas estruturas, sobre a procedência, a atualidade e confiabilidade dos dados.

Deste modo, a *MultiSIG* sugere a utilização da hierarquia *Metadata* proposta pelo SAIF, apresentada em STRAUCH, SOUZA & MATTOSO (1998) e descrita no SAIF (1995), a qual é formada pelas hierarquias de classes que descrevem informações relacionadas, a saber:

- *SpatialReferencing* - referência espacial;
- *TemporalReferencing* - referência temporal;
- *GeneralLocation* - localização geográfica;
- *Quality* - a qualidade do *GeographicObject*;

- *Lineage* - o histórico do *GeographicObject*;
- *Source* - a fonte e localização do *GeographicObject*;
- *ProductionDescription* - os tipos de produtos gerados; e
- *UpdateOperation* - os mecanismos de atualização do *GeographicObject*.

Entretanto, para atender ao objetivo de integração, a MultiSIG propõe estender as hierarquias *SpatialReferencing*, *GeneralLocation* e *Source* inserindo as seguintes propriedades nestas classes, conforme apresentado na Tabela V.1. Estas hierarquias, para distingui-las do modelo original do SAIF, passarão a ser denominadas de *SpatialReferencingMM*, *GeneralLocationMM* e *SourceMM*.

Tabela V.1 - Extensões do SAIF.

ESPECIALIZAÇÃO DO SAIF	PROPRIEDADES PROPOSTAS	TIPO
<i>SpatialReferencingMM</i>	<i>Base.ScaleInverse</i>	<i>Integer</i>
<i>GeneralLocationMM</i>	<i>GeographicRegion</i>	<i>Enumeration</i>
<i>SourceMM</i>	<i>DBInfo.queryLanguage</i>	<i>Enumeration</i>
	<i>DBInfo.operationSystem</i>	<i>Enumeration</i>
	<i>DBInfo.hardwareType</i>	<i>Enumeration</i>
	<i>DBInfo.platformName</i>	<i>Enumeration</i>
	<i>Project.applicationArea</i>	<i>Enumeration</i>
	<i>Project.projectObjective</i>	<i>Enumeration</i>
	<i>Project.projectSubsystem</i>	<i>List(String)</i>
	<i>CollectUnit</i>	<i>Enumeration</i>

A propriedade *SpatialReferencingMM.baseScaleInverse* descreve a escala em que se encontram os dados armazenados pelo esquema local. O SAIF descreve a escala, para cada mapa, utilizada na aquisição dos dados, através da *GeneralLocationMM.tile.inverseScale*. Todavia, esta escala pode ser usada para a aquisição e posteriormente alterada em decorrência da aplicação de generalização cartográfica aos dados a serem armazenados.

A propriedade *GeneralLocationMM.geographicRegion*, permite à metodologia verificar o grau de relacionamento espacial com as outras bases que fazem parte da federação. Este grau de relacionamento permite compreender como o esquema fará parte da federação e determinar com quais esquemas, o novo esquema deverá ser integrado.

As propriedades propostas à classe *SourceMM.DBInfo* visam descrever a localização da base de dados na rede, as características do gerenciamento local do esquema, de modo a, posteriormente, fornecer informação para o acesso ao esquema local na rede. As propriedades propostas para a classe *SourceMM.Project* têm por objetivo ajudar a metodologia proposta a identificar os esquemas que possivelmente

terão algum grau de proximidade semântica. Este grau de proximidade determina o quanto as classes dos esquemas representam o mesmo aspecto da realidade.

A *MultiSIG* propõe que os metadados do esquema sejam descritos, usando as notações sintáticas de classes e objetos proposta pelo SAIF, através do procedimento de cadastro do esquema na federação.

V.6.3. *Thesaurus* utilizado pela metodologia

Grande parte do conhecimento de uma base de dados é adquirido no processo de modelagem e este conhecimento é subjetivo. Para explicitar esta subjetividade e capturar a semântica implícita nos esquemas, a metodologia propõe, no processo de tradução para o SAIF, a utilização de um *Thesaurus*, a ser apresentado na interface com o usuário do **Sistema MultiSIG**.

Este *Thesaurus* é formado por um dicionário ambiental e pelo estudo da análise de domínio da aplicação de SIG, proposta por RAMIREZ (1994) e reiterada por RIBEIRO (1995). O dicionário ambiental é empregado para descrever cada elemento no nível intencional e a análise de domínio da aplicação é usada para descrever as características do domínio da aplicação.

Este *Thesaurus* oferece um vocabulário padronizado de termos que expressam a semântica associada à descrição de cada elemento, no nível intencional, e as características do domínio da aplicação. Ao descrever cada elemento do esquema no modelo de dados do SAIF, o usuário seleciona o termo que melhor descreve o elemento. Este termo enriquece e complementa as descrições, fornecendo informação adicional para que os elementos descritos sejam apropriadamente interpretados.

Estes termos constituirão um inventário da terminologia utilizada pelos esquemas das bases de dados dos SIGs. A utilização deste termo visa enriquecer e complementar as descrições dos elementos dos esquemas, bem como facilitar a identificação de similaridade/conflito entre os esquemas e os mapeamentos entre eles.

V.6.4. Modelo de dados proposto para o esquema global externo

O **Sistema MultiSIG** tem por objetivo desenvolver um ambiente MBD para usuários de SIG. Sobre este ambiente é apresentado um esquema global que fornece uma única definição que compreende todas as representações heterogêneas dos

esquemas locais. Ele tem a finalidade de familiarizar o usuário do ambiente MultiSIG com a nomenclatura utilizada pela federação. Sobre este modelo, o usuário da federação realiza consultas, sem precisar conhecer a localização dos dados, compreender os esquemas locais e as linguagens de consulta de cada SIG.

Um modelo de dados para o esquema global tem duas funcionalidades: *i)* apresentar os dados que estão disponíveis aos usuários da federação e *ii)* encapsular as funcionalidades necessárias para mapear os objetos geográficos para os esquemas locais e recuperar os dados nos esquemas locais.

Apesar das vantagens do metamodelo SAIF para descrever o esquema integrado, este modelo adiciona complexidade ao esquema global, uma vez que, não é um modelo que oferece simplicidade suficiente para o usuário, não especialista, compreender e usar seus conceitos. Desta forma, a metodologia propõe que o SAIF seja usado como MDC apenas no processo de integração, e na interface com o usuário seja utilizado um modelo mais simples que forneça maior compreensão dos dados disponíveis na federação. Sobre este modelo será expresso um outro nível de esquema global, o qual passará a ser denominado de esquema global externo. Sobre este esquema o usuário poderá formular consultas usando uma linguagem comum a todos os usuários da federação.

Para o estabelecimento de um modelo adequado à representação do esquema global externo, foi levado em consideração o fato que os usuários de Geoprocessamento estão acostumados a pensar sobre uma coleção de planos de informações e a realizar consultas sobre as bases de dados que, de modo geral, expressam um caráter espacial ou temático. A consulta de caráter espacial permite a localização de regiões pela denominação ou por suas coordenadas geográficas, ou ainda através da coincidência com outro fenômeno geográfico ou relacionamento topológico. A consulta de caráter temático pode relacionar um tema, o conteúdo do tema e agregações ou associações de temas. Um modelo de dados para o esquema global que não considere estes aspectos, levará o usuário da federação a encontrar dificuldades em localizar o que deseja ou a realizar consultas mais complexas.

Assim, para estabelecer um modelo de dados para o esquema global externo foi levado em consideração que, além de apresentar os objetos geográficos disponíveis na federação, ele deve exprimir as diferentes semânticas que estes podem ter e suportar a distribuição destes sobre várias regiões geográficas. Deste modo, o modelo de dados do

esquema global externo deve permitir ao usuário responder a duas questões básicas, a saber:

- ♦ Que temas estão disponíveis na federação e como eles serão apresentados?
- ♦ Que regiões são tratadas na federação e como elas serão relacionadas aos temas?

A *MultiSIG* propõe um modelo de dados orientado a objeto, apresentado na Figura V.3. Este modelo é formado por duas hierarquias de classes abstratas, *Theme* e *Region*, que representam os objetos geográficos em termos da região geográfica e o aspecto temático da realidade que representam.

As classes *Theme* e *Region* fornecem conhecimento sobre os temas dos objetos geográficos disponíveis na federação e sua distribuição geográfica, respectivamente. Elas são classes virtuais, e estão no nível mais alto da hierarquia do esquema global externo. Elas podem ser especializadas em outras classes, agregadas e instanciadas de acordo com os conceitos de classificação, aplicados aos objetos geográficos, no processo de integração dos esquemas.

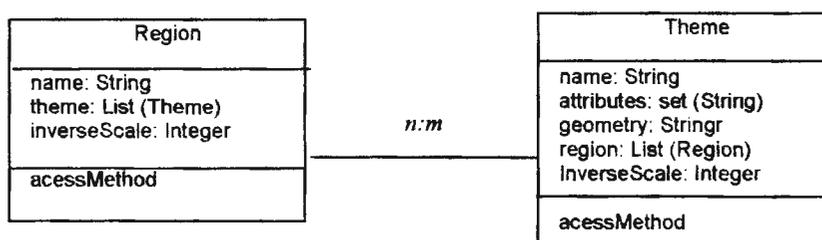


Figura V.3 - Modelo de dados do esquema global.

As instâncias de *Theme* são derivadas dos conceitos que representam as semânticas mais fortes, identificadas nas etapas da metodologia. Elas são denominadas, segundo a convenção, de denominação lógica dos objetos geográficos, conforme apresentado na seção V.9.4. Este nome lógico resulta da cooperação entre os administradores locais para estabelecer uma terminologia comum entre os esquemas. Ele garante unicidade de nomes para todos os objetos geográficos e suas propriedades.

As instâncias e classes especializadas de *Region* representam os espaços tratados pelos esquemas locais a serem disponibilizados no esquema global externo. Para facilitar, este trabalho propõe uma hierarquia para *Region*, apresentada na Figura V.4. Esta hierarquia fornece a noção dos limites políticos administrativos, tais como: *Pais*, *Estado*, *Mesorregiões*, *Município*, *Microrregiões*, *Distrito*, *Cidades* e *BaciasHidrográficas*, pois estes conceitos são usados para tomadas de decisão nos

projetos de SIG. Entretanto, a hierarquia *Region* pode ser especializada para inserir novos objetos geográficos, não previstos por esta hierarquia básica.

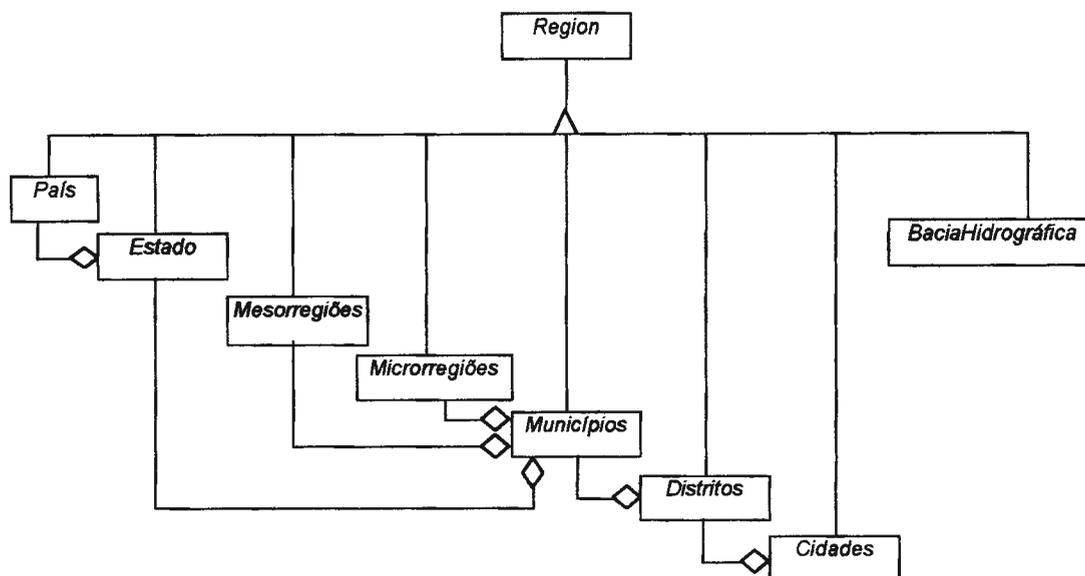


Figura V.4 - Hierarquia *Region*.

As classes do esquema global externo são virtuais, pois elas contêm a definição de instâncias virtuais. Esta definição descreve as propriedades estruturais, os relacionamentos com os outros objetos geográficos e um ou mais métodos que encapsulam as funções de acesso para os esquema global interno. As propriedades e os relacionamentos de cada classe são também identificados pela metodologia e confirmada a sua correta definição, através da cooperação entre os administradores locais. Os métodos que realizam as funções de acesso implementam o acesso e a recuperação dos dados no esquema global interno. Estes métodos são implementados pelo administrador global.

V.7. Conceitos utilizados pela metodologia

No desenvolvimento de uma base de dados geográficos existem dois elementos a serem considerados, quando se modela o mundo real: o universo observado e as necessidades das comunidades de informação quanto à área de aplicação, à finalidade da base de dados, e os recursos utilizados para a implementação. Isto faz com que estas bases de dados apresentem diferentes semânticas, tratem de diversos espaços geográficos e utilizem diferentes granularidades territoriais para armazenar o dado geográfico.

A *MultiSIG* propõe a utilização de quatro conceitos durante as suas atividades, a saber: **objeto geográfico**, **contexto**, **unidade territorial de informação** e **granularidade da informação**. Os dois primeiros para associar a comparação semântica entre as bases de dados, e os dois últimos para associar o espaço geográfico, que está sendo tratado, e o nível da informação.

V.7.1. Objeto geográfico

O objeto geográfico é um conceito lógico que representa a abstração ou definição intencional dos fenômenos/entidades geográficas existentes no mundo real, que estão representadas nos esquemas.

Este objeto incorpora conhecimentos sobre o domínio de aplicação, contemplando os aspectos semânticos, aspectos espaciais e aspectos de qualidade dos dados geográficos. Os aspectos semânticos descrevem o fenômeno que está sendo representado. Os aspectos espaciais definem sua posição e forma em relação a um sistema de referência espaço-temporal. Os aspectos de qualidade indicam seu potencial.

O esquema global interno apresenta um conjunto de classes de objetos geográficos. Cada classe de objeto geográfico é resultante da integração das classes dos objetos geográficos locais, representados pelos esquemas de exportação de cada base de dados local. Cada classe de objeto geográfico é descrita por um nome (N) e um termo (T) que representa um aspecto da realidade. Ela possui os relacionamentos (R_k) que fornecem a ligação com objetos de outras classes e propriedades estruturais (P_j), as quais descrevem as características temáticas e a representação gráfica (G). Estas propriedades possuem um nome (n), uma descrição representada por um termo (t), um formato (f), um domínio (d) e uma representação gráfica (g) que fornece a posição e geometria em relação a um sistema de referência espaço-temporal. Sobre esta representação estão implícitos a escala, os *Data*, o sistema de projeção e o sistema de coordenadas.

V.7.2. Contextos

Segundo KASHAP & SHETH (1996) e SCIORE *et al.* (1994), a semântica dos dados é associada a um contexto. Um contexto é o conjunto de descritores que fornecem conhecimento para auxiliar o raciocínio no processo de integração quanto à identificação do conteúdo, da organização, do significado e das propriedades dos dados.

A *MultiSIG* objetivando facilitar a identificação da referência espacial tratada pela base de dados, da semântica da área de aplicação da base de dados, e a determinação do grau de proximidade semântica entre os esquemas, propõe a utilização de três tipos de contexto, a saber:

- *Contexto espacial* - conjunto de descrições que tratam das referências espaciais;
- *Contexto da aplicação* - conjunto de descrições que caracterizam a área e a finalidade da aplicação da base de dados; e
- *Contexto semântico* - conjunto de descrições e relacionamentos que caracterizam a semântica do objeto geográfico.

As bases de dados dos SIGs podem tratar do mesmo assunto em regiões geográficas diferentes ou tratar da mesma região, mas utilizar parâmetros de representação cartográfica diferentes (escala, sistema de projeção, *Data* e sistema de coordenadas). A *MultiSIG* tem como ponto de partida a análise do contexto espacial. Esta análise determina o relacionamento espacial entre os esquemas. Ela determina quais os esquemas que compartilham o mesmo espaço geográfico e como serão tratados pela metodologia para a integração.

A semântica e a sintaxe dos dados armazenados pelas bases de dados possuem um conhecimento implícito da área de aplicação. A análise deste conhecimento ajuda a determinar o grau de proximidade entre os esquemas e seus objetos geográficos. A *MultiSIG* propõe analisar este grau de proximidade através da observação do contexto das aplicações dos esquemas e materializá-lo em uma hierarquia de contextos que representa os relacionamentos existentes entre as aplicações dos esquemas. Isto facilita o administrador da federação a determinar os esquemas que serão integrados e estabelecer uma ordem para tal. Esta análise é a chave para o processo de integração semântica do objeto geográfico.

A representação do contexto semântico proporciona um resumo semântico que fornece a compreensão dos objetos de um esquema e determina os seus relacionamentos com os objetos dos demais esquemas. A análise deste contexto ajuda a identificar o grau de proximidade semântica entre as classes dos esquemas. Esta análise resulta na detecção dos conflitos semânticos e sintáticos entre os esquemas.

V.7.3. Definição de unidade territorial

A  MultiSIG contempla o espaço geográfico tratado pelas diversas bases de dados geográficos de diferentes comunidades de informação. Para isto é introduzido o termo unidade territorial de informação. Este termo expressa o nível de representação dos dados armazenados pelas bases de dados geográficos participantes da federação. Ele representa uma porção do espaço geográfico, no qual os objetos geográficos são armazenados pelas bases de dados.

Na federação, as unidades territoriais de informação geográfica constituirão a hierarquia *Region* (Figura V.4). Esta hierarquia denota os espaços geográficos manipulados pelas bases de dados componentes da federação e serve como suporte geométrico para localização dos objetos geográficos na federação.

Assim, as diferentes unidades territoriais detectadas pela metodologia se tornarão instâncias ou levarão à especialização das classes da hierarquia *Region*. A especialização ocorrerá quando a unidade territorial de informação dos esquemas não se enquadrarem nas classes prevista pela hierarquia básica de *Region*. Neste caso, poderão ser introduzidas novas classes como especialização ou agregação das classes básicas desta hierarquia.

V.7.4. Granularidade territorial de informação

As bases de dados geográficos são desenvolvidas usando diferentes granulações de unidades territoriais de informação. Por exemplo, uma base de dados com finalidade de prever safra tem como menor unidade territorial de informação a unidade de coleta por município e para uma outra base de dados com finalidades cadastrais, a menor unidade territorial de informação é a propriedade.

Além disto, uma mesma área de aplicação pode apresentar mais de uma granulação da unidade territorial da informação, como por exemplo, a área de saúde que pode ter como unidade de coleta o setor censitário, as unidades hospitalares e ambulatoriais.

A granularidade territorial da informação fornece a unidade para a representação geométrica dos objetos geográficos. Este conceito será utilizado pela metodologia para a comparação dos relacionamentos semânticos entre os objetos geográficos.

V.8. Taxionomia proposta para os conflitos entre os esquemas

Os esquemas apresentam diferenças de representação do mesmo conceito do mundo real. Estas diferenças, denominadas de conflitos entre esquemas ou heterogeneidades semânticas, são decorrentes: *i)* dos projetistas modelarem a mesma perspectiva do mundo real de forma diferente; *ii)* dos projetistas utilizarem diferentes modelos de dados ou diferentes construtores para modelar o mesmo aspecto da realidade; e *iii)* das diferenças na especificação dos projetos das bases de dados, considerando as denominações empregadas, a especificação dos tipos e as restrições de integridade empregadas. Para integrar os esquemas, estas heterogeneidades precisam ser compreendidas com profundidade, a fim de detectá-las e solucioná-las.

As pesquisas em Bancos de Dados Convencionais e seus desenvolvimentos vêm ao encontro das necessidades de integração de informações multidisciplinares das diversas bases de objetos geográficos heterogêneas distribuídas, uma vez que na integração destas bases estão presentes os conflitos semânticos e estruturais das bases de dados convencionais, heterogêneas e distribuídas. Entretanto, a estes conflitos adicionam-se conflitos devido à forma de representar o objeto geográfico (STRAUCH *et al.*, 1995).

Desta forma, a *MultiSIG*, visando facilitar a identificação destas heterogeneidades, divide os conflitos em quatro grandes grupos, apresentados a seguir:

Grupo 1: Conflitos de definição do contexto espacial

Uma base de dados ao ser implementada adota parâmetros cartográficos que fornecem suporte para a representação da base de dados sobre uma cartografia adequada a sua finalidade. Dentre estes parâmetros se encontram os *Data* vertical e horizontal, o sistema de coordenadas, o sistema de projeção e a escala.

Estes parâmetros são selecionados em função das necessidades de projeto, do custo da aquisição dos dados, da qualidade pretendida para a base de dados e da disponibilidade do material a ser digitalizado. Eles estão implícitos em cada esquema de base de dados geográficos.

No presente trabalho, estes parâmetros são tratados em conjunto com outras informações que descrevem a localização, a qualidade, o histórico dos produtos gerados pelo esquema e são denominados de atributos do esquema.

No processo de integração, onde os esquemas apresentam uma multiplicidade de representações gráficas e discrepâncias cartográficas, estes parâmetros precisam ser considerados para que os elementos das diversas bases de dados sejam corretamente integrados e não apresentem inconsistências, quando da apresentação dos resultados de uma consulta sobre uma representação cartográfica.

Assim, o primeiro tipo de conflito, denominado de conflito de definição de contexto espacial, são decorrentes das diferentes definições dos parâmetros empregados pela base de dados para representação gráfica dos seus elementos. Estes conflitos são classificados em:

- *Diferença de Data* - devido à grande extensão territorial e ao alto custo da atualização do mapeamento sistemático, o Brasil tem seu território mapeado em dois Sistemas Geodésicos, conforme apresentado na Tabela V.2:

Tabela V.2 - Data nacionais.

DATA	ANTIGO	ATUAL
Horizontal	Córrego Alegre	SAD -69
Vertical	Torres	Imbituba

A adoção de Sistemas Geodésicos diferentes leva o mesmo espaço geográfico a apresentar, para um mesmo ponto, pares de coordenadas diferentes. Isto é considerado pela metodologia, uma vez que em virtude da falta de conhecimento cartográfico é possível encontrar bases de dados digitais usando o sistema antigo, principalmente o *Data* horizontal Córrego Alegre. Além disto, o Sistema Geodésico deverá futuramente ser alterado para o novo sistema, SIRGAS;

- *Diferença de sistema de projeção* - Embora no Brasil o sistema de projeção utilizado no mapeamento sistemático seja o UTM, algumas vezes, as bases de dados, por abrangerem grandes extensões, são armazenadas sem uma projeção, ou adotam outro sistema de projeção. Cada projeção representa a superfície terrestre, preservando uma característica em detrimento de outras, e adota convenções para o seu sistema de coordenadas. Um mesmo ponto representado por sistemas de projeções diferentes, também apresentará pares de coordenadas diferentes;
- *Diferença de escala* - A representação da superfície terrestre sobre o plano, implica na adoção de uma escala para representação dos elementos. A granularidade da informação e as necessidades das áreas de aplicação levam as bases de dados a apresentarem escalas diferentes. Um mesmo fenômeno, para uma mesma área de

aplicação representado por escalas diferentes, pode apresentar diferenças em sua geometria;

Grupo II: Conflitos semânticos

Os conflitos semânticos ocorrem quando as conceitualizações, sobre uma realidade comum, são percebidas de formas diferentes, levando as bases de dados a apresentarem divergências semânticas na definição dos elementos dos esquemas. Este tipo de conflito se propaga através dos modelos de dados, da organização, das aplicações, dos procedimentos e restrições de integridade.

Estes conflitos cobrem um amplo espectro, contemplando as denominações empregadas, a abstração utilizada para modelar, as propriedades que descrevem a representação dos objetos geográficos e os conflitos geométricos destes.

Para facilitar a exposição dos exemplos, sejam duas bases de dados, BD_1 e BD_2 , com seus respectivos esquemas de exportação E_1 e E_2 . Estes esquemas tratam do mesmo contexto espaço-temporal (R) e possuem um dado contexto de área de aplicação (A). Para formalizar os esquemas de exportação, o contexto espaço-temporal e o contexto de área de aplicação são tratados como conjunto e definidos os operadores \subseteq_e e \subseteq_a para representar o relacionamento espacial e de área de aplicação entre os conjuntos, respectivamente.

Cada esquema de exportação possui um conjunto de classes ($E_1 = \{C_{1a}, C_{1b}, \dots, C_{1m}\}$, $E_2 = \{C_{2a}, C_{2b}, \dots, C_{2n}\}$), formalizadas por: $C_{1a} :: E_1, C_{1b} :: E_1, \dots, C_{1m} :: E_1$ e $C_{2a} :: E_2, C_{2b} :: E_2, \dots, C_{2n} :: E_2$, onde m e n representam o número de classes destes esquemas, respectivamente. Cada classe é formada por um conjunto de propriedades $\{P_{j1}, P_{j2}, \dots, P_{jo}\}$ que representam os dados gráficos e não gráficos. Cada classe é formalizada pela tupla $C_{ij}(P_{j1}, P_{j2}, \dots, P_{jo}) :: E_i$, onde i varia de acordo com o número de propriedades de cada classe. Então:

$$C_{1a} :: E_1, C_{1b} :: E_1, \dots, C_{1m} :: E_1 \in E_1 \mid E_1 \subseteq_e R \wedge E_1 \subseteq_a A$$

$$C_{2a} :: E_2, C_{2b} :: E_2, \dots, C_{2n} :: E_2 \in E_2 \mid E_2 \subseteq_e R \wedge E_2 \subseteq_a A$$

Assim, os conflitos semânticos, neste trabalho, classificam-se em:

- *Conflitos de denominação* - estes conflitos também são encontrados nas bases de dados convencionais. Eles podem ocorrer quando um objeto geográfico e suas características temáticas, representados em esquemas que tratam de áreas de aplicação diferentes, apresentam designações desiguais em função das nomenclaturas

associadas a estas áreas. Além disto, um mesmo objeto e suas características temáticas para uma mesma área de aplicação, em esquemas diferentes, podem apresentar denominações diferentes por facilidades de manipulação ou interpretação. Isto resulta na proliferação de diferentes denominações para o mesmo objeto geográfico e suas características temáticas, bem como possíveis inconsistências entre nomes empregados pelos esquemas. Assim, estes conflitos são classificados ainda em:

- ♦ *sinônimos* - ocorrem quando os elementos dos esquemas, as classes ou suas propriedades, semanticamente idênticas no mundo real, são denominadas diferentemente; e

Exemplo em classe

HIDROGRAFIA(ID#, NOME, TIPO, VAZÃO)::E₁
 RIO(NOME, TIPO, VAZÃO)::E₂
 HIDROGRAFIA::E₁ = RIO::E₂

Exemplo entre as propriedades

ALTIMETRIA(ID#, ALTITUDE)::E₁
 ALTIMETRIA(COTA)::E₂
 ALTIMETRIA.ALTITUDE::E₁ = ALTIMETRIA.COTA::E₂
 A cota é referida ao nível médio dos mares

- ♦ *homônimos* - ocorrem quando a mesma denominação é usada para representar conceitos diferentes, isto é, os elementos (classes e propriedades) dos esquemas semanticamente diferentes compartilham a mesma denominação. Este conflito pode ser devido às semânticas das áreas de aplicação e inconsistências conceituais de modelagem.

Exemplo em classe

RECURSOATENDIMENTO(NSALAS, NPROFESSORES)::E₁
 RECURSOATENDIMENTO(NLEITOS, NMÉDICOS)::E₂
 RECURSOATENDIMENTO::E₁ ≠ RECURSOATENDIMENTO::E₂

Exemplo entre as propriedades

ALTIMETRIA(ALTITUDE)::E₁
 ALTIMETRIA(COTA)::E₂
 ALTIMETRIA.ALTITUDE::E₁ ≠ ALTIMETRIA.COTA::E₂
 A cota é referida a uma altitude arbitrária

- *Conflitos de abstração* - estes conflitos também são encontrados nas bases de dados convencionais. Eles ocorrem quando, no processo de modelagem, as conceitualizações a respeito da realidade são representadas em níveis de abstração diferente. Isto decorre das diversas interpretações da realidade e das diferentes necessidades das áreas de aplicação. Estes conflitos podem ser classificados em:

- ◆ *Conflitos de generalização* - ocorrem quando uma classe é mais geral em um esquema, sendo representada por uma propriedade em um esquema, enquanto em outro ela é mais específica sendo representada como uma classe;

Exemplo classe X propriedade
 INFRAESTRUTURA(ID#, TIPO, NOME, ADMINISTRAÇÃO)::E₁
 RODOVIAS(NOME, ADMINISTRAÇÃO, NMÃOS, LARGURA)::E₂
 FERROVIAS(NOME, ADMINISTRAÇÃO, BITOLA)::E₂
 INFRAESTRUTURA.TIPO::E₁ = RODOVIAS::E₂
 INFRAESTRUTURA.TIPO::E₁ = FERROVIA::E₂

- ◆ *Conflitos de agregação* - ocorre quando um objeto corresponde a um conjunto de objetos em outro esquema. Este conjunto pode ser homogêneo, no caso de todos os objetos terem o mesmo tipo, ou heterogêneo, no caso dos objetos terem vários tipos. Isto ocorre devido a diferentes granularidades geográficas;

Exemplo de conjunto heterogêneo
 CIDADES(ID#, NOME, POPULAÇÃO)::E₁
 NÚCLEOSURBANOS(NOME, POPULAÇÃO, RENDAPERCAPITA)::E₂
 NÚCLEOSRURAIIS(NOME, POPULAÇÃO, PRODUTIVIDADE)::E₂
 CIDADES::E₁ é composto pelas instâncias de NÚCLEOSURBANOS::E₂ e NÚCLEOSRURAIIS::E₂

Exemplo de conjunto homogêneo
 DIVESTADUAL(CÓDIGO, NOME)::E₁
 MUNICÍPIO(CÓDIGO, NOME, ESTADO)::E₂
 Uma instância de DIVESTADUAL::E₁ corresponde a várias instâncias de MUNICÍPIO::E₂

- ◆ *Conflitos de relação* - ocorrem quando uma relação entre as classes de objetos, com proximidade semântica, é representada por um relacionamento diferente em outro esquema, quanto ao nome, estrutura, e restrição;

Exemplo de conjunto heterogêneo
 MUNICÍPIO(CODESTADO, CODMUN, NOME)::E₁ POSSUI CIDADES(CODESTADO, CODMUN, CODCID, NOME)::E₁
 CIDADES (MUN, CÓDIGO, NOME)::E₂ PERTENCE MUNICÍPIO.(MUN, NOME)::E₂
 CIDADE::E₁ e CIDADES::E₂ possuem proximidade semântica, mas o relacionamento ente as classes possui nome e estruturas diferentes, a saber: POSSUI(CODESTADO,CODMUN)::E₁ e PERTENCE(MUN)::E₂

- ◆ *Conflitos entre propriedades dos objetos* - ocorrem quando as propriedades que descrevem os objetos não estão completas, ou estão implícitas nas outras propriedades, ou ainda são inconsistentes. Em geral, tratam de propriedades similares ou não equivalentes, que descrevem objetos semanticamente equivalentes;

RODOVIA(ID#, NOME, ADMINISTRAÇÃO)::E₁
 RODOVIA(ID#, NOME, CÓDIGO)::E₂
 RODOVIA.CÓDIGO::E₂ encapsula o tipo de administração ∴ RODOVIA::E₁ = RODOVIA::E₂

PROPRIEDADE (ID#, PROPRIETÁRIO, LARGURA, COMPRIMENTO)::E₁
 PROPRIEDADE (ID#, PROPRIETÁRIO, ÁREA)::E₂

PROPRIEDADE.LARGURA::E₁ * PROPRIEDADE .COMPRIMENTO::E₁ = PROPRIEDADE .ÁREA::E₂

CLIMA (PRECIPITAÇÃO, TEMPERATURA, UMIDADE)::E₁

CLIMA (PRECIPMÍN,PRECMAX, TEMPMMIM, TEMPMAX, UMIÍMÍN, UMIÍMAX)::E₂

As propriedades do E₁ são as médias das precipitações, temperaturas e medidas de umidade.

As propriedades do E₂ são as mínimas e máximas da precipitação, temperatura e umidade

- *Conflitos geométricos* - estes conflitos são próprios dos esquemas das bases de dados geográficos. Eles ocorrem quando um mesmo objeto do mundo real é apresentado nos esquemas com diferentes representações geométricas. Eles são classificados em:

- ♦ *Conflitos de abstração cartográfica* - ocorrem quando diferentes abstrações cartográficas são utilizadas para representar o mesmo objeto do mundo real. Este conflito decorre do fato das abstrações cartográficas serem realizadas de acordo com a finalidade da aplicação da base de dados. Isto determina uma escala e uma representação geométrica que pode ser diferente nos vários esquemas. Por exemplo:

CIDADE (NOME, POPULAÇÃO, GEOMETRIA)::E₁ ∴ CIDADE .GEOMETRIA::E₁ = ponto

CIDADE (NOME, POPULAÇÃO, GEOMETRIA)::E₂ ∴ CIDADE .GEOMETRIA::E₂ = área

- ♦ *Conflitos entre as granularidades territoriais de informação* - os objetos geográficos são armazenados, de acordo com uma unidade territorial de coleta necessária a aplicação. Portanto, a definição da granularidade territorial de informação pode variar de esquema para esquema. Por exemplo:

Pesquisa para saúde: SOCIAIS(NMORTALIDADE, CAUSAS, GEOMETRIA)::E₁

SOCIAIS.GEOMETRIA::E₁ = área(Setor censitário)

Secretaria de saúde: SOCIAIS (NÓBITOS, MOTIVO, GEOMETRIA)::E₂

SOCIAIS.GEOMETRIA::E₂ = ponto (Hospitais e Ambulatórios)

Grupo III: Conflitos estruturais

Ocorrem quando as conceitualizações, apesar de serem as mesmas e adotarem o mesmo modelo de dados, possuem diferentes especificações, restrições e domínios, fazendo com que a realidade comum seja representada de forma diferente nas bases de dados a serem integradas. Estes conflitos são classificados em:

- *Conflitos de tipo* - ocorrem quando as propriedades semanticamente equivalentes de dois esquemas apresentam definições de tipo incompatíveis;

MUNICÍPIO(CÓDIGO, NOME)::E₁ ∴ MUNICÍPIO.CÓDIGO::E₁ é definido como inteiro

MUNICÍPIO(CÓDIGO,NOME)::E₂ ∴ MUNICÍPIO.CÓDIGO::E₂ é definido como char(6)

- *Conflitos de formatos* - ocorrem quando a mesma propriedade emprega formatos diferentes nas bases de dados a serem integradas;

PROPRIEDADERURAL(ID#, PROPRIETÁRIO, DATACADASTRO)::E₁
 PROPRIEDADE.DATACADASTRO::E₁ é expressa em dia/mês/ano
 PROPRIEDADERURAL(ID#, PROPRIETÁRIO, DATACADASTRO)::E₂
 PROPRIEDADE.DATACADASTRO::E₂ é expressa em mês/dia/ano

- *Conflitos de unidades* - ocorrem quando as unidades das propriedades são diferentes para a mesma conceitualização, em virtude das necessidades operacionais dos usuários;

NÚCLEORURAL(ID#, NOME, PRODUTIVIDADE)::E₁
 NÚCLEORURAL.PRODUTIVIDADE::E₁ é expressa pelo peso total em toneladas
 CIDADERURAL(ID#, NOME, POPULAÇÃO, PRODUTIVIDADE)::E₂
 CIDADERURAL.PRODUTIVIDADE::E₂ é expressa em reais

- *Conflitos de domínios* - ocorrem quando os objetos do mundo real são percebidos no mesmo nível de abstração e classificados de acordo com o mesmo critério, mas a propriedade temática assume diferentes conjuntos de valores de domínio. Isto decorre do fato de que os valores de algumas propriedades não sejam intrínsecos aos objetos, mas dependam da proposta de aplicação;

USODATERRA(TIPO)::E₁ ∴ USODATERRA.TIPO::E₁ = {área cultivada, área construída}
 USO(TIPO)::E₂ ∴ USO.TIPO::E₂ = {cultivada(cereal(feijão, milho, arroz), flores(rosa, Íris..), área construída)}

DECLIVIDADE(TIPO)::E₁ ⇒ DECLIVIDADE.TIPO::E₁ = {Fraca, Moderada, Alta}
 DECLIVIDADE(TIPO)::E₂ ⇒ DECLIVIDADE.TIPO::E₂ = {A[0-10], B[11-20], C[21-30], D[31-45]} ∴

Os intervalos estão em graus

DECLIVIDADE(TIPO)::E₃ ⇒ DECLIVIDADE.TIPO::E₃ = {A[0-3], B[4-8], C[8-20], D[20-45]} ∴

Os intervalos estão em percentagem

- *Conflitos de restrição* - ocorrem quando alguma propriedade possui uma restrição lógica aos valores que pode assumir em um esquema;

CLIMA (TEMPDIA, TEMPNOITE, TEMPERATURA)::E₁

A propriedade temperatura é menor igual a 50° C

CLIMA (TEMPERATURA)::E₂

A propriedade temperatura é menor igual a 45° C

- *Conflitos de códigos* - os códigos são usados por várias razões, tal como facilitar o acesso à informação e otimizar espaço. Eles geralmente não são uniformes entre as bases, uma vez que eles atendem às necessidades específicas dos usuários. Assim o uso de código introduz diferença de valores;

FERROVIA(NOME, ADMINISTRAÇÃO, BITOLA CÓDIGO)::E₁ ∴ O código é um número seqüencial

FERROVIA(NOME, CÓDIGO)::E₂ ∴ O código é o número da ferrovia

- *Conflitos entre chaves* - ocorrem quando as chaves primárias nos diversos esquemas são diferentes, ou diferentes chaves são usadas para o mesmo conceito;

MUNICÍPIO(CÓDIGO, NOME)::E₁ ∴ A chave é o código da divisão político-administrativa fornecido pelo IBGE

MUNICÍPIO(CEP,NOME)::E₂ ∴ A chave é o código de endereçamento postal do município

Grupo IV: Conflitos entre valores

Estes conflitos ocorrem no nível extencional das propriedades. Eles podem ser:

- *Conflitos de valores default* - os SIGs e aplicações podem proporcionar diferentes valores *default* para propriedades semanticamente iguais;

SOLO(TIPO)::E₁ ∴ SOLO.TIPO::E₁ é assumido 0
SOLO(TIPO)::E₂ ∴ SOLO.TIPO::E₂ é assumido latossolo

- *Conflitos de valores* - alguns SIGs oferecem facilidades para automaticamente calcular valores geométricos, os quais podem assumir valores diferentes decorrentes da qualidade dos dados espaciais, da precisão adotada para as coordenadas ou mesmo do algoritmo usado pelo SIG;

FERROVIA(ID#, NOME, PERIMETER)::E₁ ∴ FERROVIA.PERIMETER::E₁ é calculado pelo SIG

FERROVIA(NOME, PERÍMETRO)::E₂ ∴ FERROVIA::E₂ é medido em campo

- *Conflitos de atualização* - acontecem quando as propriedades de diferentes bases de dados são atualizadas em diferentes épocas;

MUNICÍPIO(CÓDIGO, NOME, POPULAÇÃO)
MUNICÍPIO::E₁(33,Rio de Janeiro, 200.000) em 1995
MUNICÍPIO::E₂(33,Rio de Janeiro, 150.000) em 1975

- *Conflitos de erros de registros* - podem ser erros tipográficos ou variações nos processos de medição, ou erros provenientes da aquisição de dados nos processos de digitação/digitalização.

MUNICÍPIO(CÓDIGO, NOME, POPULAÇÃO)
MUNICÍPIO::E₁(33,Rio de Janeiro, 200.000) em 1995
MUNICÍPIO::E₂(33,Rio de Janeiro, 50.000) em 1975

V.9. Etapas da metodologia proposta para integração

A ~~o~~ MultiSIG tem por objetivo elaborar um esquema global, a partir dos esquemas locais cadastrados com os respectivos mapeamentos para as bases de dados locais, que participam da federação proposta pelo **Sistema MultiSIG**.

Para melhor compreensão a metodologia preconiza **seis etapas**, a saber: pré-integração, análise dos esquemas, conformação de esquemas, integração no esquema global interno, definição do esquema global externo e validação da integração (Figura V.5).

Estas seis etapas desenvolvem uma série de atividades que visam facilitar a integração dos esquemas de exportação. A primeira etapa, **pré-integração**, consiste na descrição dos esquemas das bases de dados locais. Esta etapa implica na execução dos procedimentos de participação e no procedimento de cadastro, apresentados na seção V.5. O primeiro procedimento implica na negociação entre os administradores locais e o administrador da federação, de forma a conhecerem a filosofia da federação e assumirem os compromissos para a participação na federação. O segundo procedimento, consiste em descrever os esquemas locais no modelo de dados do SAIF, o qual foi selecionado como MDC, de acordo com as regras de mapeamento proposta pela metodologia, apresentadas na seção V.9.1.2.

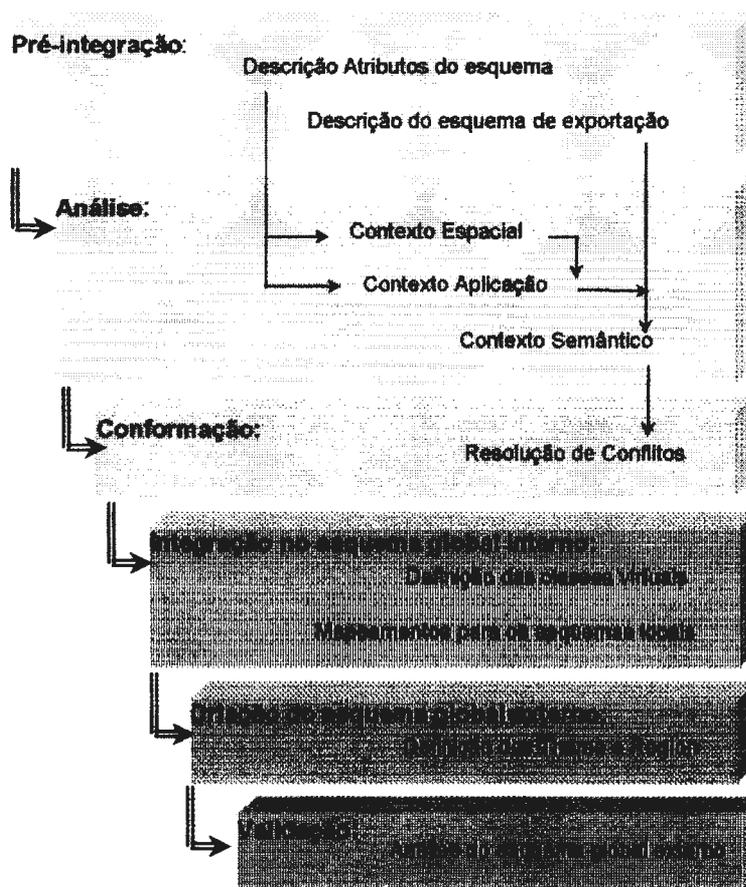


Figura V.5 -- MultiSIG.

A segunda etapa, **análise dos esquemas**, consiste em realizar sobre os esquemas de exportação uma série de análises, no nível extencional dos atributos dos esquemas e no nível intencional, objetivando identificar os esquemas a serem integrados e os relacionamentos existentes entre eles.

Nesta etapa são identificados os conflitos propostos pela taxionomia, apresentada na seção anterior. Os conflitos de definição do contexto espacial são tratados na etapa de análise do contexto espacial, na qual são detectadas as bases de dados que tratam do mesmo espaço geográfico. A metodologia considera estes conflitos, propondo na integração de resultados que sejam realizadas as transformações de parâmetros cartográficos necessárias para integrar os dados recuperados, resultantes das consultas as bases de dados locais. Os conflitos semânticos e estruturais são identificados na análise do contexto semântico e tratados na etapa de conformação. Os conflitos entre os valores serão analisados na integração de resultados da consulta pelo sistema.

A etapa de **conformação** tem por objetivo harmonizar as classes comuns que possuem proximidade semântica discrepantes detectadas na etapa de análise. Nesta etapa é estabelecido o processo de cooperação entre o administrador global e os administradores locais, de modo a estabelecerem os conceitos comuns e mapeamentos entre eles.

Na etapa de **integração dos esquemas** são criadas as classes virtuais, a partir das classes dos esquemas de exportação utilizando, uma notação estendida a partir da SAIF CSN. Nesta etapa, são definidos os métodos de acesso para as bases de dados locais e os métodos de mapeamento entre as classes com proximidade semântica. As definições destas classes serão armazenadas pelo **Sistema MultiSIG** em um diretório denominado de Metaesquema.

A seguir, a etapa de **criação do esquema global externo**, faz uma modelagem das classes virtuais do esquema integrado no modelo de dados proposto pela MultiSIG. Esta modelagem consiste em criar instâncias e especializações de *Theme* e *Region*, de acordo com as classes detectadas na etapa de análise, conformadas e integradas.

E finalmente, a etapa de **validação** avalia o esquema global, de forma a verificar se a informação está consistente e compreensível.

V.9.1. Etapa 1: Pré-integração

Os esquemas das bases de dados geográficos representam o resultado da modelagem lógica do mundo real. Eles possuem características implícitas relacionadas às construções de suas bases de dados e as suas finalidades. Estas características, denominadas de atributos do esquema, definem as referências espaço-temporais, o histórico dos dados, a proveniência e a qualidade dos dados.

Assim, a primeira etapa da metodologia proposta consiste em descrever os esquemas das bases de dados de SIGs, no modelo de dados do SAIF, usando as notações sintáticas proposta por este modelo. Entretanto, a metodologia propõe que esta descrição seja realizada em dois níveis, dividindo esta etapa em duas atividades, a saber: *descrição dos atributos do esquema e descrição do esquema de exportação*.

A descrição dos atributos dos esquemas consiste em descrever os atributos implícitos, relacionados à construção e a finalidade da base de dados geográficos, utilizando a notação sintática de objetos proposta pelo SAIF. Esta descrição tem por objetivo permitir extrair informações sobre o domínio da aplicação, que muitas vezes não se apresentam nos esquemas.

A descrição do esquema de exportação consiste em traduzir a modelagem do esquema local para o modelo de dados do SAIF estendido pela metodologia, usando a notação sintática de classes. Contudo, as regras estabelecidas pelo SAIF para a criação de esquema no seu modelo de dados não são adequadas para a tradução dos esquemas na federação MultiSIG, uma vez que estas regras criam classes com novas denominações para agregar as especializações de *GeographicObject*, que não existem no esquema local. Isto além de não retratar adequadamente o esquema local, no mínimo duplica a quantidade de classes de cada esquema, aumentando a complexidade no processo de comparação. Assim, a metodologia propõe regras especiais de tradução dos esquemas para o SAIF.

V.9.1.1. Descrição dos atributos do esquema

Tendo em vista que o esquema de uma base de dados geográficos é criado para atender uma finalidade de aplicação e que um mesmo esquema pode apresentar, de acordo com as necessidades dos usuários locais, mais de uma abstração cartográfica para apresentar os planos de informação, esta atividade tem por objetivo caracterizar o domínio do esquema de exportação quanto a área de aplicação e a finalidade, e

explicitar os diferentes sistemas de referência espaço-temporal. Estes atributos do esquema são comuns aos planos de informações e, em geral, não são armazenados pelos esquemas. Eles são do conhecimento dos administradores e usuários locais.

Para descrever estes atributos, a *MultiSIG* utiliza a definição da classe do SAIF *GeographicObject* e o artifício, proposto pela notação sintática de classe do SAIF, para associar valores às classes e as suas especializações através da cláusula *default*.

A classe *GeographicObject* na sua definição é composta pela hierarquia *Metadata*. Esta hierarquia contém as classes e propriedades que descrevem os atributos do esquema. Através da cláusula *default* os valores instanciados para as classes da hierarquia *Metadata* são definidos e herdados pelas subclasses de *GeographicObject*.

Desta forma, a *MultiSIG* propõe que, para cada esquema local, seja criado uma ou mais classes de *GenericGeographicObject* (GGO) especializadas de *GeographicObject*, que devem ser denotadas por:

GGO::nomeDoEsquemaLocal. Abstração

Para esta classe deve ser instanciada a propriedade *Metadata*, a qual é representada pelas classes que compõem a hierarquia *Metadata* do SAIF. As classes do esquema que possuem os mesmos atributos de esquema deverão ser especializadas a partir desta. Se houverem classes no esquema local que possuem outros níveis de abstração cartográfica, devem ser criadas tantas classes de GGO quantas forem as diferenças.

Os metadados definidos pelas classes da hierarquia *Metadata* que descreve *SpatialReferencingMM*, *TemporalReferencing*, *SourceMM* e *GeneralLocationMM* tratam de propriedades comuns a todas as classes do esquema que estão no mesmo nível de abstração, portanto estas classes devem obrigatoriamente ser instanciadas para a classe de GGO. O metadado referente a identificação da fonte, descrito pela propriedade *SourceMM.database.DbInfo.dbIdentifier*, deve ser instanciado para cada classe do esquema de forma a fornecer a localização das instâncias daquela classe. Os demais metadados, *Quality*, *Lineage*, *ProductDescription* e *UpdateOperation*, referentes a qualidade, histórico, produto e atualização, respectivamente, podem ser instanciadas para as classes do esquema, de acordo com a necessidade do usuário.

Para ilustrar, seja um esquema (E) com objetos geográficos que empregam dois níveis de abstrações cartográficas. O primeiro, ABSTRAÇÃO¹, objetiva registrar os dados levantados em campo e o segundo, ABSTRAÇÃO², divulgar a informação ao nível de

município. Estes dois níveis de abstrações implicam no esquema utilizar dois sistemas de referência espacial, escalas e metodologias de aquisição dos dados adequadas a finalidade dos planos de informação. Assim, neste exemplo, ao descrever este esquema no SAIF serão criados duas classes de GGO, de acordo com o nível de abstração e instanciados os seus metadados, a saber: GGO::E.ABS1 e GGO::E.ABS2. As classes deste esquema serão descritas no SAIF como especializações do GGO, de acordo com o nível de abstração cartográfica utilizado por elas.

Assim, a *MultiSIG* preconiza que para cada esquema local deve ser criado uma ou mais classes que representam os *GenericGeographicObject* (GGO), especializados de *GeographicObject*, e genéricos para todo o esquema, conforme as abstrações cartográficas empregadas no esquema. A partir destas classes as demais classes do esquema traduzidas para o SAIF serão especializadas. Este objeto que tem a propriedade *metadata*, vincula os atributos da classe *Metadata* a todas as classes especializadas desta, através da cláusula *default*, que é herdada pelas suas subclasses.

V.9.1.2. Descrição do esquema de exportação

A descrição do esquema de exportação envolve a reconstrução semântica e mapeamento de um esquema, expresso no modelo de dados do SIG, para um esquema orientado a objeto do metamodelo SAIF. Isto requer a identificação do construtor apropriado para melhor descrever cada elemento expresso no modelo de dados do SIG.

Todavia, cada SIG oferece um modelo de dados próprio, ao qual está associado estruturas para representar o objeto de forma que o sistema possa manipulá-lo, através das operações oferecidas pela sua linguagem de consulta. Assim, a descrição do esquema no SAIF está extremamente associada à identificação da estrutura para armazenar o objeto geográfico. É preciso reconhecer as estruturas oferecidas por cada SIG, para armazenar o objeto geográfico, e identificá-las no SAIF.

A descrição do esquema de exportação, portanto, consiste em: *i)* distinguir as entidades do esquema local que possuem atributos gráficos das que não têm; *ii)* identificar a estrutura de dados, usada pela base de dados geográficos para representar o objeto geográfico; e *iii)* especificar a semântica da classe transcrevendo seu nome, suas propriedades e respectivos domínios, relacionamentos e termos no glossário disponibilizado pelo **Sistema MultiSIG** que melhor descreva o objeto geográfico. Esta descrição utiliza as cláusulas proposta pela CSN, apresentadas no apêndice A.

Assim, cada entidade (E), do esquema original com chave (K), propriedades ($P_i | i > 0$) de domínio (D) e relacionamentos (R), será mapeada em uma classe especializada de GGO ou de *AbstractObject*. A classe será criada no SAIF com o mesmo nome, mesma chave, mesmos atributos e relacionamentos, de acordo com as regras de mapeamentos (M) apresentadas abaixo, na ordem em que devem ser aplicadas:

- M 1: entidades que não possuem atributos gráficos serão modeladas como classes especializadas de *AbstractObject*, na qual o administrador local transcreverá todas as suas propriedades;
- M 2: entidades que possuem representação gráfica serão definidas como uma especialização da classe GGO, de acordo a abstração cartográfica usada na representação do esquema local.
- M 3: tendo em vista que as entidades com representação gráfica são armazenadas na base de dados local, como planos de informações que contêm um conjunto de objetos geográficos de um tipo, cada plano de informação será traduzido para classes especializadas que representam conjuntos de objetos no SAIF, isto é: *Coverage* ou *SpatialDataSet*;
- M 4: a geometria de cada objeto geográfico será escolhida de acordo com estrutura usada para representar o objeto geográfico no esquema local. Cada SIG oferece um modelo de dados próprio, ao qual estão associadas estruturas para representar o objeto geográfico. Desta forma, esta regra de mapeamento será especializada, posteriormente, para cada SIG, em função das estruturas oferecidas pelo modelo de dados por ele utilizado;
- M 5: as denominações de cada classe serão formadas por seu nome no esquema local, seguido pelo nome do esquema de exportação, concatenado com a *string* "::";
- M 6: as propriedades originais da classe selecionada no SAIF, para representar o objeto geográfico deverão ser preservadas e a estas deverão ser acrescentadas as propriedades definidas para os objetos geográficos no esquema local, inclusive as propriedades pré-estabelecidas pelo SIG, se fizerem parte da modelagem de dados e aplicações locais;
- M 7: para cada propriedade deverão ser descritos o nome, o formato, o domínio e o valor *default*. As propriedades chaves deverão ser indicadas com um asterisco na notação, após a sua denominação. O nome e o formato são descritos na cláusula *classAttributes*, o domínio na cláusula *classAttributeValue*, e o valor *default* na cláusula *default* da CSN. Quando o domínio da propriedade for um atômico ele

- será de um tipo básico *Integer*, *Real* ou *String*. Quando o domínio da propriedade tratar de uma classificação, isto é, o domínio for representado no SIG como uma tabela, esta propriedade será do tipo *Enumeration*. Esta classe poderá ser uma enumeração própria do SAIF ou uma nova especialização da classe *Enumeration*;
- M 8: caso alguma das propriedades da entidade, no esquema original, apresente alguma restrição de integridade, ela deverá ser descrita usando as cláusulas *restricted* e *constraints* da CSN. A cláusula *restricted* deve ser usada para expressar restrições simples, aos valores dos atributos, e a cláusula *constraints* deve ser usada para expressar restrições que envolvam alguma lógica entre os atributos;
- M 9: todos os relacionamentos entre as classes deverão ser transformados em classes de objetos especializadas da classe *Relationship*. Esta classe é estendida pela *MultiSIG* com a propriedade *relationAttribute*. Assim, cada relacionamento no esquema original é descrito pelo seu nome no esquema local, pelas classes envolvidas e suas respectivas cardinalidades, e pelos atributos do relacionamento. Os relacionamentos temporais deverão ser expressos por especialização da classe *TemporalRelationship*. Os relacionamentos espaciais deverão ser especializados a partir da hierarquia de *SpatialRelationship*; e
- M 10: a cláusula *comments* da SAIF CSN será utilizada pelo **Sistema MultiSIG** para permitir ao usuário introduzir os termos escolhidos no *Thesaurus*, que descrevam o objeto geográfico e suas propriedades.

V.9.1.2.1. Regras específicas

Tendo em vista o objetivo de testar a metodologia em esquemas (Apêndice A) desenvolvidos sobre os SIGs ARC/Info, produzido pela *Environmental Systems Research Institute*, e SPRING, produzido pelo Instituto Nacional de Pesquisas Espaciais, as estruturas destes SIGs foram estudadas em STRAUCH, SOUZA e MATTOSO (1998d) e estabelecidas as correspondências entre suas estruturas para representar os objetos geográficos e as classes do SAIF, de modo a adequar a regra quatro; a saber.

- *Regras de tradução do ARC/Info para o SAIF*

Os objetos geográficos armazenados com as estruturas gráficas do ARC/INFO são definidos no MDC como classes especializadas de GGO e de uma classe da hierarquia de *GeographicObject*, com a geometria que será definida de acordo com o tipo de estrutura usada no ARC/Info, conforme a Tabela V.3:

Tabela V.3 - Correspondências entre as estruturas gráficas do ARC/Info e as classes do SAIF.

ARC/INFO	Classes no SAIF
<i>Coverage Label point</i>	<i>PartitionedCoverage: position.Geometry: Point</i>
<i>Coverage Arc</i>	<i>PartitionedCoverage: position.Geometry: VectorLine</i>
<i>Coverage Node</i>	<i>PartitionedCoverage: position.Geometry: Point Segment(Arc)</i>
<i>Coverage Route</i>	<i>Graph: AcyclicNetwork</i> <i>Graph: SingleLineNetwork:</i>
<i>Coverage Section</i>	<i>Graph: SingleLineNetwork</i>
<i>Coverage Polygon</i>	<i>PartitionedCoverage: position.Geometry: VectorArea</i>
<i>Coverage Annotation</i>	<i>Annotation</i> <i>AbstractObject</i>
<i>Grid e Lattice</i>	<i>Raster: position.Geometry: Grid</i> <i>CategoricalRaster: position.Geometry: CellGrid2D</i>
<i>TIN</i>	<i>Field: position.Geometry: VectorSurface Grid</i> <i>IrregularDem: position.Geometry: MeasuredSurface</i> <i>IrregularDem: position.Geometry: masspoints{:Point DEMPoint</i>
<i>Image</i>	<i>GeneralRaster: position.Geometry: ImageGrid2D</i>

- *Regras de tradução do SPRING para o SAIF*

Os objetos geográficos que possuem representação gráfica são definidos no modelo de dados do SPRING como objetos pertencentes a classes derivadas de *PlanoDeInformação* e *GeoObjeto*. Estas classes são traduzidas para o MDC, de acordo com a classe e a sua estrutura, conforme a Tabela V.4 abaixo:

Tabela V.4 - Correspondência entre as classes do SPRING e do SAIF.

SPRING	Classes no SAIF
<i>GeoObjeto:</i>	<i>GeographicObject:</i>
<i>Point2D ou ConjuntoDePontos2D</i>	<i>position.Geometry: Point</i>
<i>Linha2D</i>	<i>position.Geometry: VectorLine</i>
<i>Poligono:</i>	<i>position.Geometry: VectorArea</i>
<i>MapaCadastral</i>	
<i>SubdivisãoPlanar:</i>	<i>SpatialDataSet:</i>
<i>Point2D</i>	<i>geoComponents{:position.Geometry: Point</i>
<i>Linha2D</i>	<i>geoComponents{:position.Geometry: VectorLine</i>
<i>Poligono</i>	<i>geoComponents{:position.Geometry: VectorArea</i>
<i>ConjuntoDePontos2D :</i>	<i>SpatialDataSet:</i>
	<i>geoComponents{:position.Geometry: Point</i>

<p><i>GrafoOrientado:</i></p> <p><i>GeoCampoNumérico:</i> <i>Grade Regular</i> <i>ConjuntoDeIsolinhas</i> <i>GradeTriangular:</i> <i>MapaPontos3D:</i></p> <p><i>GeoCampoTemático:</i> <i>ImagemTemática</i> <i>SubdivisãoPlanar:</i> <i>Point2D</i> <i>Linha2D</i> <i>Poligono</i></p> <p><i>GeoCampoDadosSensor</i> <i>ImagemTonsDeCinza</i> <i>ImagemSintética</i></p>	<p><i>SpatialDataSet:</i> <i>geoComponents{}.position.Geometry:Point</i> <i>geoComponents{}.position.Geometry:VectorLine</i> <i>geoComponents{}.position.Geometry:VectorArea</i> <i>Field \IrregularDem</i> <i>Field: position.Geometry: Grid</i> <i>Field: position.Geometry: VectorSurface.^Isolines</i> <i>Field: position.Geometry: VectorSurface.^TIN</i> <i>IrregularDem:position.Geometry:masspoint{}:^DEMpoint^Point</i></p> <p><i>CategoricalRaster: positiongeometry: CellGrid2D</i> <i>PartitionedCoverage:</i> <i>geoComponents{}.position.Geometry: Point</i> <i>geoComponents{}.position.Geometry: VectorLine</i> <i>geoComponents{}.position.Geometry: VectorArea</i></p> <p><i>GeneralRaster: position.Geometry: ImageGrid2D</i> <i>GeneralRaster: position.Geometry: ImageGrid2D</i></p>
--	--

V.9.2. Etapa 2: Análise dos esquemas

Para estabelecer a proximidade semântica entre os esquemas é necessário estabelecer a convergência das descrições dos esquemas. Para isto, se faz necessário uma etapa de análise dos esquemas. Esta etapa tem por objetivo definir os esquemas a serem integrados, estabelecer uma ordem para a integração dos esquemas, identificar o grau de proximidade semântica entre as classes dos esquemas e os conflitos que porventura existirem.

A metodologia proposta sugere que esta análise seja dividida em três tipos:

- *análise do contexto espacial dos esquemas* - compara os parâmetros utilizados para representar o espaço geográfico, com o objetivo de estabelecer espaços geográficos comuns tratados pelos esquemas;
- *análise do contexto das aplicações dos esquemas* - compara os esquemas que tratam do mesmo espaço geográfico, quanto as áreas de aplicação, com a finalidade de estabelecer uma ordem entre os esquemas para a integração, de acordo com a proximidade do domínio da área de aplicação; e
- *análise do contexto semântico dos esquemas* - compara os elementos do esquema que tratam do mesmo espaço geográfico e do mesmo domínio da aplicação, de forma a identificar os relacionamentos entre as classes dos esquemas.

A definição da correspondência espacial e do contexto da aplicação são responsabilidades do administrador da federação. Esta correspondência é estabelecida com base nas informações fornecidas pela descrição dos atributos do esquema. A responsabilidade da análise do contexto semântico do esquema é dividida entre o administrador da federação e os administradores locais, uma vez que ao encontrar

algum grau de proximidade semântica, este grau deve ser validado através de cooperação com os administradores locais.

V.9.2.1. Análise do contexto espacial dos esquemas

Esta etapa realiza uma análise dos parâmetros espaciais, utilizados para a abstração cartográfica dos esquemas e a análise da unidade territorial de informação, com a finalidade de estabelecer as regiões a serem disponibilizadas pelo esquema global externo, através da hierarquia *Region*.

- **Análise dos parâmetros espaciais**

Esta análise identifica os conflitos do Grupo I da taxionomia apresentada na seção V.8. Ela visa assegurar a correta comparação dos espaços geográficos tratados pelos esquemas e identificar a necessidade de transformação entre os parâmetros cartográficos utilizados para representar seus elementos gráficos nos esquemas de exportação.

Para isto, a *MultiSIG* propõe inicialmente detectar os conflitos entre os metadados relacionados ao contexto espacial. Isto é realizado através da comparação da notação sintática, utilizada para descrever os metadados relacionados a *SpatialReferencingMM*.

Sobre os metadados de *SpatialReferencingMM* serão analisados os parâmetros de representação utilizados pelas entidades gráficas das bases de dados geográficos. Estes parâmetros são os atributos do esquema descritos pelas propriedades *SpatialReferencingMM.coordSystem*, *SpatialReferencingMM.projection.System*, *SpatialReferencingMM.coordSystem.horizontalReference*, *SpatialReferencingMM.CoordSystem.verticalReference* e *SpatialReferencingMM.inverseScale*, que descrevem o sistema de coordenadas, o sistema de projeção, os *Data* horizontal e vertical e a escala, respectivamente.

Para facilitar esta comparação deve ser criada a Tabela de Parâmetros Cartográficos (TPC), apresentada abaixo (Tabela V.5), dispondo os parâmetros espaciais utilizados para cada esquema (E_1, E_2, \dots, E_n).

A diferença de parâmetros cartográficos entre os esquemas, dependendo do espaço geográfico tratado, indica a necessidade de aplicar operações de transformações para a comparação e integração dos esquemas, a saber:

- Diferença de sistema de coordenadas - requer a transformação de sistema de coordenadas para os metadados relativos a descrição da unidade territorial do esquema, de modo a determinar o relacionamento espacial entre os esquemas. No caso dos esquemas apresentarem a mesma unidade territorial de informação e a metodologia indicar a proximidade semântica entre as classes, também se faz necessário indicar a necessidade de aplicar transformações posteriormente na consulta para a integração de resultados;
- Diferença de *Data* horizontal e vertical - este caso é semelhante ao anterior, entretanto, deve ser aplicado uma transformação de *Datum*; e
- Diferença de escala - esta diferença traduz uma diferença na granularidade da informação armazenada. Se os esquemas apresentarem a mesma unidade territorial de informação e a metodologia indicar a proximidade semântica entre as classes, será necessário analisar com cuidado a geometria. No caso da diferença de escala não alterar a geometria, poderá ser empregada uma generalização semântica no processo de conformação, e, conseqüentemente, uma generalização cartográfica na apresentação dos resultados da consulta. No caso da diferença de escala alterar a geometria, as classes serão integradas à hierarquia *Theme*, de acordo com a sua geometria.

Tabela V.5- TPC com as propriedades que definem os parâmetros cartográficos.

ATTRIBUTES	E ₁	E ₂	...	E _n
SpatialReferencingMM.coordSystem				
SpatialReferencingMM.projectionSystem				
SpatialReferencingMM.coordSystem.HorizontalReference				
SpatialReferencingMM.coordSystem.VerticalReference				
SpatialReferencingMM.baseInverseScale				

- ***Determinação do relacionamento espacial entre os esquemas***

A metodologia proposta aceita esquemas que não descrevem o mesmo espaço geográfico, porém, é necessário estabelecer os relacionamentos entre os espaços geográficos. Desta forma, é realizada uma análise do espaço geográfico, tratado pelas bases de dados geográficos, com a finalidade de encontrar o relacionamento espacial entre elas.

Esta análise é realizada através da comparação da descrição sintática dos metadados, fornecidos pelas propriedades que descrevem a localização do espaço geográfico tratado pelos esquemas, a saber: *GeneralLocationMM.geographicRegion* e/ou *GeneralLocationMM.range.minPoint.coord.c1*, *GeneralLocationMM.range.minPoint.coord.c2*, *GeneralLocationMM.range.maxPoint.coord.c1* e *GeneralLocationMM.range.maxPoint.coord.c2*.

Visando facilitar esta comparação, estas propriedades devem ser dispostas em uma Tabela de Localização Geográfica (TLG), apresentada abaixo (Tabela V.6), contendo estas propriedades, para cada esquema (E_1, E_2, \dots, E_n).

No caso de ter sido detectado uma diferença nos parâmetros espaciais, na etapa anterior, as propriedades que definem as coordenadas máximas e mínimas devem ser compatibilizadas em um único sistema de referência cartográfico, de forma a possibilitar a correta comparação

Tabela V.6- TLG com as propriedades que definem a localização geográfica.

ATTRIBUTES	E_1	E_2	...	E_n
GeneralLocationMM.geographicRegion				
GeneralLocationMM.range.minPoint.coords.c1				
GeneralLocationMM.range.minPoint.coords.c2				
GeneralLocationMM.range.maxPoint.coords.c1				
GeneralLocationMM.range.maxPoint.coords.c2				

A determinação do relacionamento espacial define como o esquema de exportação fará parte dos esquemas global e do externo. O esquema de exportação poderá ser associado ou integrado. A identificação deste relacionamento espacial é estabelecida pela operação de interseção espacial entre as propriedades descritas na Tabela V.7, comparando-se cada par de esquemas de exportação e procurando identificar os relacionamentos espaciais próximos.

Para ilustrar a análise do relacionamento espacial, seja o formalismo apresentado na seção V.8 para o par de esquemas de exportação E_1 e $E_2 \mid E_1 \subseteq_e R_1 \wedge E_2 \subseteq_e R_2$, onde R_1 e R_2 são os contextos espaço-temporais destes esquemas. Esta comparação pode resultar nos seguintes de relacionamentos espaciais (RE):

- RE vazio - as bases de dados não possuem nenhum relacionamento espacial, uma vez que tratam de espaços geográficos diferentes;

- RE parcial - as bases de dados apresentam dois tipos de relacionamentos espaciais entre os espaços geográficos dos esquemas. Eles apresentam interseção entre os espaços geográficos, mas eles não são coincidentes, havendo uma região de não interseção dos espaços geográficos e outra de interseção; e
- RE total - as bases de dados apresentam interseção total entre os espaços geográficos não vazios e eles são coincidentes.

Tabela V.7- Relacionamentos espaciais.

RE	INTERSEÇÃO ESPACIAL ENTRE OS ESQUEMAS		PARTICIPAÇÃO NO ESQUEMA GLOBAL INTERNO
VAZIO	$R_1 \cap R_2 = 0$		Justaposição de esquemas
PARCIAL	$R_1 \cap R_2 \neq 0$	$R_1 \subset R_2$	Integração de esquemas
		$R_1 \cap R_2$ $R_2 - R_1$	Justaposição de esquemas
TOTAL	$R_1 \equiv R_2$		Integração de esquemas

O RE vazio ocorre quando as comunidades de informação manifestam vontade de participar da federação, apesar de seus esquemas tratarem de espaços geográficos diferentes. Neste caso, será verificada a existência, na hierarquia *Region*, de uma classe que contemple o espaço geográfico, de forma a inserir uma instância que contemple o espaço geográfico. A esta instância será definido a lista de *Theme*, onde cada *Theme*, corresponderá a uma classe do esquema de exportação descrita no modelo de dados do esquema global externo.

Para ilustrar o RE vazio, seja um esquema E_1 que trata do estado de São Paulo e um esquema E_2 que trata do estado do Rio de Janeiro. A classe *Estado* da hierarquia básica proposta para *Region* serão instanciados RioDeJaneiro e SãoPaulo (Figura V.6). Para cada uma destas instâncias serão associadas as respectivas listas de *Theme*. Cada *Theme* corresponde a uma classe no esquema de exportação.

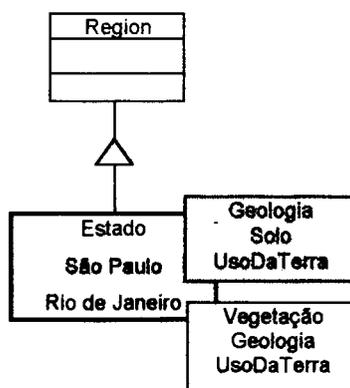


Figura V.6 - Justaposição de esquemas sem relacionamento espacial a hierarquia *Region*.

O segundo, RE *parcial*, ocorre quando, as bases de dados apresentam regiões de interseção espacial, porém, não total. Neste caso, a metodologia é aplicada de modo a identificar o relacionamento semântico entre as classes dos esquemas de exportação. De acordo com o diagnóstico do relacionamento espacial e semântico poderão ser tomadas as seguintes medidas para a região do espaço geográfico do esquema de exportação:

- As classes tem relacionamento espacial e semântico - as classes serão integradas, conforme a metodologia e inseridas na hierarquia *Region* em uma classe que contemple o espaço geográfico de interseção; e
- As classes possuem relacionamento espacial, mas não tem relacionamento semântico - As classes dos esquemas de exportação serão descritas e associadas às classes da hierarquia *Region* que melhor descrevam seus espaços geográficos, sem integração, apenas com os métodos de acesso para as respectivas bases de dados locais.
- As classes não possuem relacionamento espacial nem semântico - neste caso as classes serão tratadas como no caso anterior.

Para ilustrar o RE *parcial*, seja um esquema E_1 , que armazena para os municípios do estado do Rio de Janeiro a geologia, uso do solo e vegetação e seja um esquema E_2 , que trata dos municípios da Bacia Hidrográfica do Rio Paraíba do Sul, abrangendo municípios do Rio de Janeiro, de São Paulo, de Minas Gerais e do Espírito Santo, com os planos de informações relacionados à geologia, florestas e declividade (Figura V.7). Neste exemplo, para a hierarquia *Region* serão criadas as seguintes instâncias de *Region*, conforme apresentado na Figura V.8:

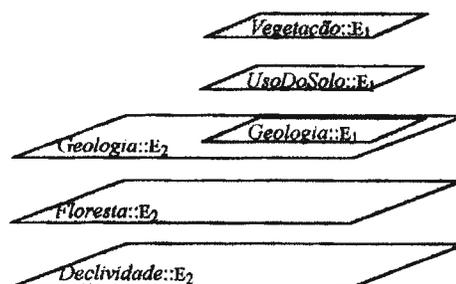


Figura V.7 - Exemplo de relacionamento espacial parcial.

- Estado: RioDeJaneiro, que será associada aos temas geologia, uso do solo, vegetação, floresta e declividade, onde a geologia será resultante da integração de geologia::E1 e geologia::E2, uso do solo e vegetação terá o método de acesso para o esquema E_1 e floresta e declividade para o esquema E_2 ; e

- *BaciaHidrográfica*: *BaciaParaíbaDoSul* com os temas *geologia::E₂*, *florestas::E₂* e *declividade::E₂*

Para ilustrar RE total, ocorre quando há coincidência do espaço geográfico tratado pelas bases de dados locais. No caso de ser uma região já contemplada pelas classes da hierarquia *Region*, será criada a respectiva instância, e a ela associada a lista de *Theme*, resultantes do processo de integração das classes dos esquemas de exportação. Para ilustrar, sejam os esquemas E_1 e E_2 que tratam dos Municípios do Rio de Janeiro. Os esquemas destas bases de dados serão integrados conforme as atividades da *MultiSIG*.

Cabe ressaltar, quando a federação já existir e um novo esquema desejar participar, as regiões geográficas resultantes da análise do relacionamento espacial devem ser comparadas com a hierarquia *Region*, na etapa de criação do esquema global externo. Esta comparação visa verificar a existência de outro esquema, já pertencente a federação, que trate do mesmo espaço geográfico. Caso este espaço já exista em alguma das classes da hierarquia *Region*, o esquema federado, expresso no MDC, deve ser recuperado. A seguir, a metodologia deve ser aplicada a parte do esquema federado e do esquema de exportação, de modo a detectar o relacionamento semântico entre os esquemas para integrá-los. Caso o espaço geográfico não seja contemplado pela hierarquia *Region*, será tratado como o relacionamento espacial vazio, isto é, criada uma especialização e uma instância que descreva este espaço geográfico e associada a lista de *Theme*, onde cada *Theme* corresponderá a uma classe do esquema de exportação, expresso no modelo de dados do esquema global externo.

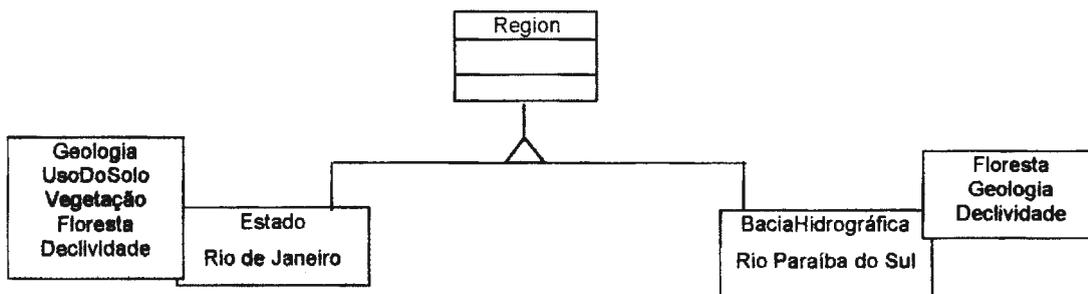


Figura V.8 - Integração e união de esquemas com relacionamento espacial, parcial ao esquema global externo.

V.9.2.2. Análise do contexto da aplicação dos esquemas

A MultiSIG propõe uma integração binária. Assim, nesta etapa são analisados os esquemas que possuem relacionamentos espaciais, parcial e total. Para determinar a seqüência de pares de esquemas a serem integrados, esta etapa leva em consideração que cada área de aplicação possui um vocabulário próprio e que esquemas que tratam do mesmo espaço geográfico e do mesmo domínio de aplicação, possuem uma maior proximidade semântica entre as suas classes.

Desta forma, esta etapa analisa o contexto da aplicação dos esquemas, visando estabelecer uma hierarquia de áreas de aplicação e identificar quais os pares de esquemas que possuem maior proximidade semântica. Esta análise define uma prioridade de integração entre os esquemas, em função da proximidade de área de aplicação e finalidade do projeto.

A análise é realizada sobre a descrição sintática dos metadados fornecidos pelas propriedades que descrevem os projetos associados aos esquemas, quanto a finalidade e área de aplicação. Assim, para facilitar esta análise deve ser criada a Tabela do Contexto das Aplicações (TCA) contendo as propriedades que descrevem os projetos para cada esquema (E_1, E_2, \dots, E_n), conforme a Tabela V.8.

Em função desta análise serão estabelecidas hierarquias de contextos de áreas de aplicação para cada esquema. Estas hierarquias materializam árvores, em cuja estrutura os nós raízes são as áreas de aplicação. Estes nós podem ser subdivididos de acordo com a finalidade da aplicação de cada esquema, em um ou mais nós filhos. Cada nó filho será novamente dividido, de acordo com os subsistemas que compõem a finalidade de cada esquema.

Tabela V.8- Propriedades que descrevem o contexto da área de aplicação.

ATTRIBUTES	E_1	E_2	...	E_n
SourceMM.Terminus.Agency.agencyName				
SourceMM.Project.projectIdentifier				
SourceMM.Project.objective				
SourceMM.Project.subsystem				
SourceMM.Project.aplicationArea				

A análise das áreas de aplicação, quanto a finalidade e seus subsistemas representados nas hierarquias, define o relacionamento genérico existente entre os esquemas. Este relacionamento genérico entre as áreas de aplicação estabelece a

prioridade de análise semântica entre os esquemas e, conseqüentemente, determina uma seqüência para os esquemas serem integrados. Para ilustrar, sejam os esquemas E_1 , E_2 , E_3 e E_4 , que possuem RE total, descritos abaixo na TCA (Tabela V.9) com suas respectivas propriedades.

Tabela V.9 - Exemplo de TCA para quatro esquemas.

ATTRIBUTES	E_1	E_2	E_3	E_4
SourceMM.terminusInfo.agencyName	Embrapa	White Martins	CPRM	Cataguazes
SourceMM.project.projectIdentifier	RJ2000	Reflares do Vale	Despoluição	Ger. de Rede de Luz
SourceMM.project.Objective	Estudo Ambiental	Extrati. vegetal	Monit. Ambiental	Planej. E monit
SourceMM.Project.subsystem	Ambiental Social	Ambiental Econômicas	Ambiental Social Econômico	Cadastral Equipamentos Social
SourceMM.project.aplicationArea	Meio Ambiente	Recursos naturais	Meio Ambiente	Ocupação humana

As hierarquias que expressam o relacionamento entre os contextos das áreas de aplicação destes esquemas são representadas na Figura V.9. Neste exemplo, possivelmente os esquemas E_1 e E_3 apresentam proximidade semântica entre suas classes, uma vez que elas apresentam a mesma área de aplicação e subsistemas compatíveis. Estes dois esquemas possuem uma superposição maior entre os subsistemas do esquema E_2 e E_4 . Desta forma, será estabelecida a seguinte ordem de pares de esquemas para a integração $((E_1, E_3), E_2), E_4$.

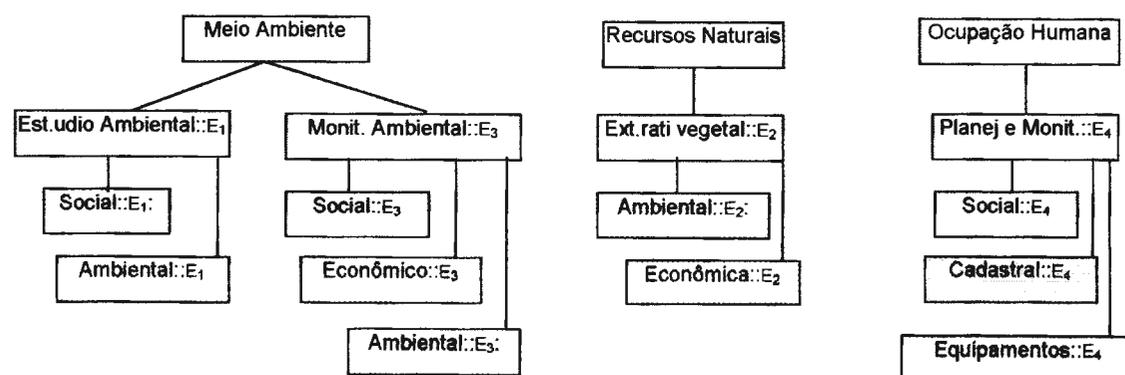


Figura V.9 - Exemplo das hierarquias contexto de aplicação.

Esta seqüência de integração cria os esquemas globais intermediários dos E_{G13} e E_{G123} e o esquema global interno E_{G1234} . O esquema E_{G13} representa a integração das classes relacionadas ao contexto de aplicação voltado, para meio ambiente. O esquema E_{G13} é integrado às classes relacionadas aos recursos ambientais derivando o esquema E_{G123} e, finalmente, a este esquema é integrado as classes relacionadas ao contexto de

aplicação (Figura V.10). As classes de contextos diferentes podem apresentar proximidade semântica, sendo tratadas pelas demais etapas da metodologia proposta.

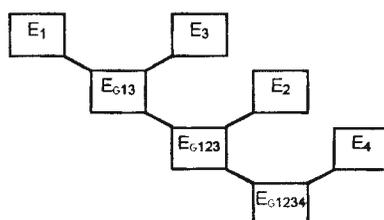


Figura V.10- Exemplo de seqüência de integração dos esquemas.

Sendo o objetivo do **Sistema MultiSIG** criar uma federação aberta, o exemplo considerará um esquema E_5 , abaixo descrito (Tabela V.10), a ser inserido na federação após o esquema global (E_G) interno ter sido derivado dos esquemas anteriores.

A parte do esquema global interno que trata da mesma região será analisada, buscando encontrar proximidades entre o esquema global intermediário com maior proximidade de contexto de aplicação, no caso E_{G13} . O par de esquema (E_5, E_{G13}) é analisado e integrado no esquema E_{G135} . Este esquema é analisado com o esquema E_{G123} , de acordo com a metodologia, procurando identificar a existência de classes que possuem proximidade semântica.

Tabela V.10- Exemplo de TCA para o quinto esquema.

ATTRIBUTES	E_5
SourceMM.terminusinfo.agencyName	UERJ
SourceMM.project.projectIdentifier	KM-Ambiental
SourceMM.project.Objective	Estudo Ambiental
SourceMM.Project.subsystem	Ambientais Sociais
SourceMM.project.aplicationArea	Meio Ambiente

V.9.2.3. Análise do contexto semântico dos esquemas

Esta etapa objetiva estabelecer o grau de proximidade semântica entre as classes dos pares de esquemas (E_m, E_n), identificando as similaridades e os conflitos do grupo II e III da taxionomia apresentada na seção V.8.

A **MultiSIG**, visando facilitar o tratamento das classes, propõe classificar os pares de classes com proximidade semântica de acordo com os graus abaixo, os quais

serão ilustrados pelos esquemas de exportação, formalizados por (E_m, E_n) , onde $E_m = \{C_{ma}, C_{mb}, \dots, C_{mi}\}$ e $E_n = \{C_{na}, C_{nb}, \dots, C_{nj}\}$

- Identidade semântica: $(C_{mi} \equiv C_{nj})$ - ocorre quando os esquemas utilizam os mesmos construtores de modelagem e os mesmos descritores para representar objetos que tratam da mesma semântica da realidade. Neste caso, os objetos de duas classes provenientes de esquemas diferentes são idênticos. Estes pares de classes raramente serão encontrados, devido as heterogeneidades inerentes às diferentes perspectivas do mundo real no processo de modelagem dos dados;
- Similaridade semântica $(C_{mi} \approx C_{nj})$ - ocorre quando os esquemas empregam diferentes construtores e diferentes especificações de formatos, tipos, unidades e domínios para descrever o mesmo aspecto da realidade. Isto leva os objetos de duas classes de diferentes esquemas a apresentarem diferentes representações para o mesmo aspecto da realidade, sejam decorrentes de discordância sobre o significado, interpretação, ou uso pretendido do mesmo; e
- Heterogeneidade semântica $(C_{mi} \neq C_{nj})$ - ocorre quando as classes não tratam dos mesmos aspectos da realidade. Neste caso as classes são heterogêneas e descrevem objetos geográficos diferentes.

Este grau de proximidade semântica é identificado através da análise e comparação da denominação e descrição semântica das classes, considerando seus atributos gráficos e não gráficos. Desta forma, a análise do contexto semântico (CS) determina o grau de proximidade semântica (G_{PS}) entre os objetos geográficos, denotado por:

$$G_{PS}:CS (C_{mi}::E_m, C_{nj}::E_n) = \langle T(N), R_k, G, P_k(n_k, d_k, f_k, r_k, d_k) \rangle$$

- O primeiro componente, G_{PS} , denota o grau de proximidade semântica que pode ser a identidade, similaridade e heterogeneidade;
- O segundo componente denota o par de classes (C_{mi}, C_{nj}) dos esquemas E_m e E_n , respectivamente, que estão sendo comparados; e
- O terceiro componente denota os elementos das classes a serem comparados quanto: a denominação (N) e o termo (T) que definem os objetos, os relacionamentos (R_k), a geometria (G) e as propriedades das classes quanto a denominação (n_k), descrição (d_k), formato (f_k), restrição (r_k) e domínio (d_k) da propriedade.

Assim, esta etapa coordena as análises, abaixo descritas, realizadas sobre cada par de esquemas com proximidade de área de aplicação, a saber:

- **Análise das denominações das classes dos esquemas**

Esta análise tem por objetivo identificar as semelhanças e as diferenças de representação do mesmo objeto do mundo real. Para isto é observado no nível intencional das classes especializadas de GGO e de *AbstractObject*, de cada par de esquema (E_m, E_n). As classes são dispostas em uma Tabela de Classes (TC), com sua respectiva lista de termos utilizados para descrever as classes ($C_{mi}::E_m, C_{nj}::E_n$) e os seus respectivos tipos (Tabela V.11). A seguir, é realizada uma pesquisa sobre os termos de modo a determinar a proximidade semântica (G_{PS}) entre as classes e avaliar os tipos usados para representar as classes.

Tabela V.11- Comparação entre as denominações das classes.

E_m			E_n		
C_{mi}	LISTA(TERMO _i)	TIPO _i	C_{nj}	LISTA(TERMO _j)	TIPO _j

Esta pesquisa na TC é descrita pelo seguinte formalismo:

$$G_{PS} : CS (C_{mi}::E_m, C_{nj}::E_n) \Rightarrow \exists \text{TERMO} \in TC \mid \text{TERMO}_i \in C_{mi}::E_m \wedge \text{TERMO}_j \in C_{nj}::E_n \wedge$$

$$C_{mi} \cdot \text{TERMO}_i::E_m \cong C_{nj} \cdot \text{TERMO}_j::E_n \vee C_{mi} \cdot \text{TERMO}_i::E_m = C_{nj} \cdot \text{TERMO}_j::E_n$$

Nesta comparação são detectadas as classes que podem apresentar proximidade semântica quanto a denominação e os termos usados para descrevê-la. Em geral, são encontrados os conflitos de denominação. Apesar das denominações serem diferentes, os termos utilizados para descrever os objetos geográficos, facilitam identificar as classes que apresentam uma proximidade semântica.

A descrição dos tipos utilizados para representar as classes de objetos geográficos no SAIF, na Tabela V.11, tem a finalidade de verificar se eles foram descritos como objetos que possuem representação gráfica. Isto visa facilitar, posteriormente, a criação da tabela de geometria, na atividade de análise da geometria.

Para simplificar a manipulação das classes com indicação de proximidade semântica esta etapa emprega o conceito de grupo temático. O grupo temático define um contexto semântico que representa as classes de objeto geográfico, que possuem uma indicação de proximidade semântica estabelecida, através dos termos selecionados

para defini-las. Estes grupos temáticos estabelecem uma terminologia para a comparação que facilita a análise dos pares de classes com proximidade semântica.

Nesta etapa, as classes que possuem esta proximidade são organizadas nestes grupos temáticos. Para este grupo temático é adotada uma denominação geral, que melhor represente o conjunto de classes. Visando facilitar as etapas a seguir, cada grupo temático deve ser colocado em uma tabela, denominada de Tabela de Grupo Temático (TGT), onde as classes são dispostas em ordem ascendente de similaridade, conforme a indicação de proximidade semântica.

Estes grupos temáticos devem ser encaminhados aos administradores das bases de dados locais, os quais devem confirmar o agrupamento temático e estabelecer o relacionamento entre as suas classes com as dos outros esquemas. Isto é realizado através do procedimento de cooperação. Para ilustrar, sejam os esquemas E_1 e E_2 apresentados abaixo, na Tabela V.12.

As classes $RioPrin::E_1$ e $Hidrografia::E_2$ apresentam os mesmos termos para defini-las. As classes $Afluente::E_1$ e $Drenagem::E_2$ apresentam termos diferentes, mas com uma possibilidade de proximidade semântica relacionada a hidrografia. As classes $Alti::E_1$, $PontosCotados::E_2$, $Cmestre::E_2$ e $Declivi::E_1$, apesar de apresentarem termos diferentes para defini-las, descrevem a topografia da região.

Tabela V.12- Exemplo de TC.

E ₁			E ₂		
C _{mi}	LISTA(TERMO _i)	TIPO _i	C _{nj}	LISTA(TERMO _j)	TIPO _j
RioPrin	Rio principal	GGO	Rio	Hidrografia	GGO
Alti	Topografia	GGO	Drenagem	Canais de drenagem	GGO
Declivi	Declividade	GGO	PontosCotados	Relevo	GGO
Afluente	Rios Secundários	GGO	CurvaMestre	Relevo	GGO

Através do procedimento de cooperação, os administradores locais discutem a separação das classes nos grupos temáticos (Tabela V.13), com denominações que melhor exprima o conjunto de classes próximas, por exemplo, Hidrografia e Topografia. O grupo temático de hidrografia será formada pelas classes $RioPrin::E_1$, $Afluente::E_1$, $Hidrografia::E_2$ e $Drenagem::E_2$. O grupo temático de topografia será formado pelas classes $Alti::E_1$, $Declivi::E_1$, $PontosCotados::E_2$ e $CurvaMestre::E_2$.

O grupo temático, nas etapas a seguir, poderá ser especializado conforme a necessidade de categorizar as classes. Assim, ao final da análise do contexto semântico,

cada grupo temático será tratado como uma coleção de classes de objetos que possuem alguma associação entre elas, através de suas denominações, relacionamentos e propriedades comuns.

Tabela V.13- Grupos Temáticos.

Grupo temático: Hidrografia	
RioPrin::E ₁	Rio::E ₂
Afluente::E ₁	Drenagem::E ₂
Grupo temático: Topografia	
Alti::E ₁	PontosCotados::E ₂
Declivi::E ₁	CurvaMestre::E ₂

- **Análise dos relacionamentos semânticos**

Esta análise visa detectar as classes dos esquemas que, além de uma proximidade semântica, encontrada na etapa anterior, apresente o mesmo contexto de relacionamento.

Para cada grupo temático (GT_p), detectado na análise anterior, deve ser criada uma Tabela de Relacionamentos (TR) contendo os relacionamentos entre as suas classes e as classes de outros grupos temáticos. Nesta tabela (Tabela V.14) são observados então o contexto do relacionamento quanto a denominação dos relacionamentos (REL) e das classes (C) envolvidas. Nesta análise são observadas ainda as cardinalidades (CAR), a estrutura do relacionamento quanto aos atributos (AT. DO REL) e, se necessário, quanto a utilização de restrições aplicadas a estes atributos.

Tabela V.14- Relacionamentos das classes de cada grupo.

E _m : GT _p					E _n : GT _p				
REL _{mi}	C _{me}	C _{mf}	CAR _{mi}	AT. DO REL _{mi}	REL _{nj}	C _{nk}	C _{nl}	CAR _{nj}	AT. DO REL _{nj}

Esta pesquisa na TR é descrita pelo seguinte formalismo:

$$G_{PS}: CS (REL_{mi}::E_m, REL_{nj}::E_n) \Rightarrow \exists REL \in TR |$$

$$C_{me}, C_{mf} \in REL_{mi}::E_m \wedge C_{nk}, C_{nl} \in REL_{nj}::E_n |$$

$$C_{me}.TERMO::E_m \cong C_{nk}.TERMO::E_n \vee C_{me}.TERMO::E_m \cong C_{nl}.TERMO::E_n \vee$$

$$C_{mf}.TERMO::E_m \cong C_{nk}.TERMO::E_n \vee C_{mf}.TERMO::E_m \cong C_{nl}.TERMO::E_n \wedge$$

$$REL_{mi}.CAR::E_m = REL_{nj}.CAR::E_n \wedge REL_{mi}.ATDOREL_{mi}::E_m \cong REL_{nj}.ATDOREL_{nj}R::E_n$$

Nesta análise são detectados os conflitos de abstração causados pelas diferentes formas de representar a realidade, pelo uso dos diversos construtores dos modelos usados para modelá-la e quanto a estrutura do relacionamento.

Esta análise resulta em uma modelagem preliminar entre o pares de classes dos esquemas que possuem alguma proximidade semântica. Para ilustrar, seja o grupo temático Hidrografia, citado anteriormente, com os relacionamentos apresentados abaixo na Tabela V.15. O nome dos relacionamentos e as cardinalidades são diferentes, entretanto, ele apresenta uma proximidade que pode ser modelada para integrar as classes, a saber:

Tabela V.15- Exemplo de TR.

REL _{mi}	C _{me}	C _{mf}	CAR _{mi}	AT. DO REL _{mi}	REL _{nj}	C _{nk}	C _{nl}	CAR _{nj}	AT. DO REL _{nj}
Contém	RioPrin	Afluente	1:n	cod	Pertence	Drenagem	rio	1:1	dren

Neste exemplo, a classe *Drenagem::E₂* pertence a um *Rio::E₂*, não importando se é rio principal ou afluente. Por outro lado, o esquema *E₁* modela os vários afluentes de um rio. Desta forma, através de consenso entre os administradores locais é proposta uma modelagem das classes e seus relacionamentos de modo a integrá-los, como apresentado na figura abaixo (Figura V.11):

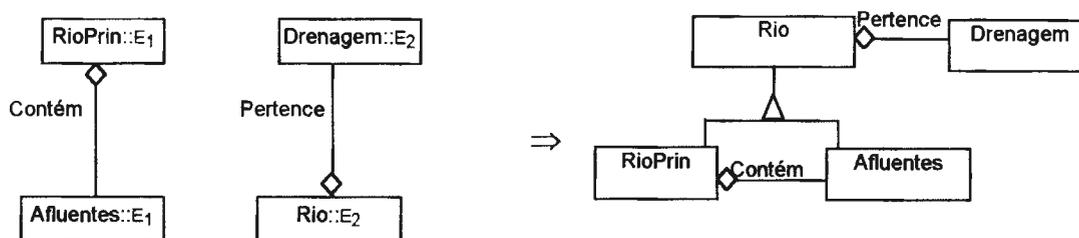


Figura V.11- Exemplo: modelagem dos relacionamentos e suas respectivas classes.

- **Análise da geometria das classes**

No caso das classes, para as quais foram detectadas uma proximidade semântica, e são especializações de *GeographicObject*, devem ser analisadas as geometrias (GEO). Esta análise visa identificar os conflitos de representação geométrica causados pelo uso de diferentes granularidade da informação e representação gráfica. Assim, para as classes, as quais além da denominação foram detectadas o mesmo tipo de

relacionamento, devem ser organizada em uma nova tabela (Tabela V.16) de geometrias (TG), contendo a geometria das classes.

Tabela V.16- Geometria das classes com proximidade semântica.

E _m :GT _p		E _n :GT _p	
C _{mi}	GEO _i	C _{nj}	GEO _j

Esta pesquisa na TG é descrita pelo seguinte formalismo:

$$G_{PS} : CS (C_{mi}::E_m, C_{nj}::E_n) \Rightarrow C_{mi}.GEO::E_m = C_{nj}.GEO::E_n \vee C_{mi}.GEO::E_m \cong C_{nj}.GEO::E_n$$

Para ilustrar, seja as classes do exemplo anterior. Inicialmente, suponha que as classes possuem o mesmo nível de abstração cartográfica, como apresentada na Tabela V.17, abaixo.

Tabela V.17- Exemplo 1 da TG.

Grupo temático: Hidrografia			
C _{mi}	GEO _i	C _{nj}	GEO _j
RioPrin::E ₁	VectorLine	Rio::E ₂	VectorLine
Afluente::E ₁	VectorLine		
		Drenagem::E ₂	VectorLine

Neste exemplo, observa-se que todas as classes possuem a mesma geometria. Assim, a modelagem apresentada na Figura V.11 pode ser utilizada para representar a integração dos esquemas E₁ e E₂. Seja um segundo exemplo, onde o esquema E₂ é representado em uma escala maior, com um nível de abstração cartográfica diferente, no qual a drenagem e os rios são tratados como trechos e apresentados como polígonos (Tabela V.18).

Tabela V.18- Exemplo 2 da TG.

Grupo temático: Hidrografia			
C _{mi}	GEO _i	C _{nj}	GEO _j
RioPrin::E ₁	VectorLine	Rio::E ₂	VectorArea
Afluente::E ₁	VectorLine		
		Drenagem::E ₂	VectorArea

Neste caso, os rios podem apresentar dois níveis de abstração, *vectorLine* ou *vectorArea*. Tendo em vista que a drenagem está no mesmo nível de abstração cartográfica que o Rio, com a geometria *vectorArea*, a modelagem especializa a classe

rio, de acordo com sua geometria, e associa a drenagem ao rio com geometria *vectorArea* (Figura V.12).

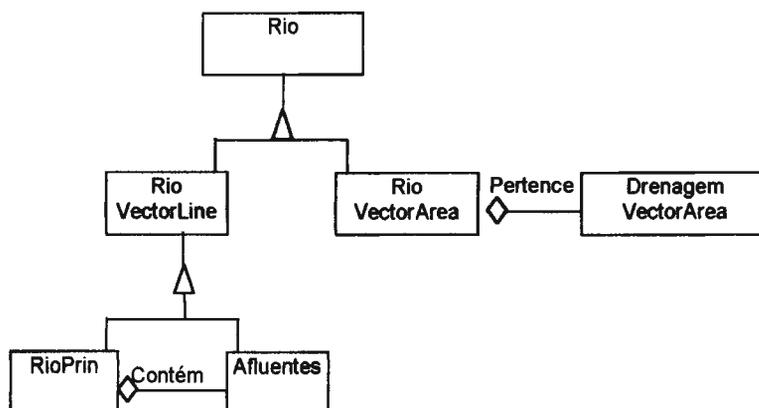


Figura V.12 - Exemplo: modelagem dos relacionamentos considerando a geometria.

- **Identificação das propriedades comuns**

Com o objetivo de encontrar classes com maior proximidade semântica são analisadas as propriedades das classes, de modo a identificar as características temáticas comuns. Desta forma, os pares de classes ($C_{mi}::E_m$; $C_{nj}::E_n$) que possuem proximidade de denominação e de relacionamento devem ser inseridos em uma Tabela Não Normalizada de Propriedades (TNNP), contendo para cada propriedade (P) seus respectivos nomes/termos (T). Nesta tabela (Tabela V.19), as propriedades usadas como chaves deverão ser indicadas com um asterisco.

Tabela V.19- Tabela não normalizada das propriedades.

$C_{mi}::E_m$	$C_{nj}.P_1.T::E_m$...	$C_{nz}.P_x.T::E_m$
$C_{nj}::E_n$			
$C_{mi}.P_1.T::E_n$	X		
.....			
$C_{mx}.P_z.T::E_n$			

Sobre esta tabela é realiza uma pesquisa para encontrar propriedades equivalentes ou implícitas em outras propriedades, formalizada por:

$$\exists T \in P \mid T \text{ é único } \wedge$$

$$C_{mi}.P.T::E_m = C_{nj}.P.T::E_n \vee C_{mi}.P.T::E_m \cong C_{nj}.P.T::E_n$$

Cabe ressaltar que os SIGs, em geral, possuem modelos de dado próprios que oferecem propriedades específicas para cada representação gráfica. A metodologia propõe que estas propriedades sejam ignoradas nesta análise, a menos que façam parte das necessidades dos usuários.

Sobre a TNNP é realizada uma pesquisa comparando-se, para cada par de classes, para o qual foi detectada alguma proximidade semântica nas etapas anteriores, os termos que descrevem as propriedades. Esta comparação consiste em detectar, para cada linha uma coluna com um termo correspondente. Sobre o par de linha e coluna deve ser marcado um "X" que identifica as propriedades comuns ou com alguma proximidade semântica. Esta proximidade encontrada entre as propriedades das classes, deve ser encaminhada aos administradores locais para confirmação. Em função das informações fornecidas pelo administrador local, deve ser aprimorada a modelagem preliminar considerando as propriedades.

Para ilustrar, sejam as classes do exemplo anterior com as propriedades apresentadas abaixo nas Tabelas V.20:

Tabela V.20- Exemplos de TNNP.

RioPrin::E ₁	IdRio*	Nome	Larg-med	Perímetro	Vazão	Profun	Nascente
E ₂ :Rio							
CódigoRio	X						
Denominação		X					
Perímetro				X			
Volume					X		
Profundidade						X	

Afluente::E ₁	IdRio*	Nome	Nascente
Dren::E ₂			
CódigoRio	X		
Denominação		X	
Vazão			

Nesta tabela são analisadas as classes com indicação de proximidade semântica e analisadas as propriedades comuns estabelecendo a seguinte modelagem (Figura V.13):

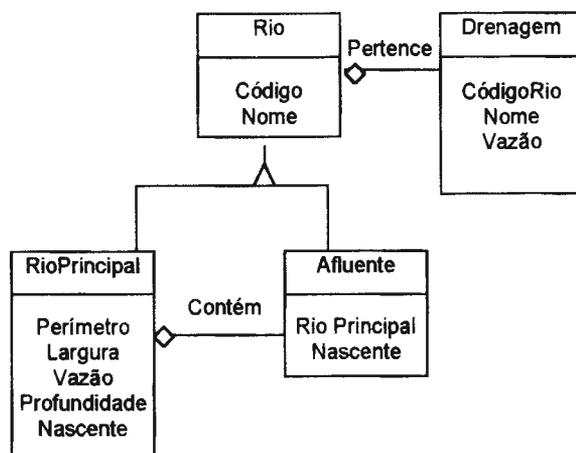


Figura V.13- Exemplo: modelagem dos relacionamentos considerando as propriedades.

- **Análise das descrições das propriedades**

A detecção da proximidade semântica encontrada é refinada através da criação de uma Tabela de Correspondência entre Propriedades (TCP). Nesta tabela, cada propriedade do par de classes com indicação de proximidade semântica é descrita quanto ao nome, tipo e formato, restrição e domínio (Tabela V.21).

Tabela V.21- Tabela com as correspondências entre as propriedades.

E _n :GT _p									
E _m :GT _p :C _{mi}					E _n :GT _p :C _{nj}				
ATRIBUTO	TIPO	FORMATO	RESTRIÇÃO	DOMÍNIO	ATRIBUTO	TIPO	FORMATO	RESTRIÇÃO	DOMÍNIO

Devem ser incluídas as propriedades criadas pelo sistema que façam parte da modelagem e para as quais foram identificadas correspondências na TNNP. Entretanto, cabe ressaltar que os valores fornecidos pelo sistema, em geral, apresentam inconsistências, uma vez que estes são calculados pelo sistema quando comparado com outro obtido por rigorosos processos de medição e introduzido pelo operador. Por exemplo, a propriedade perímetro, na Figura V.13, pode ter sido medida no campo ou calculada pelo sistema.

Nesta análise, os atributos correspondentes são observados quanto ao nome, tipo, formato, domínio e restrição. A análise do domínio é baseada nas características dos valores e na existência de equivalência, estabelecida entre estes valores. Esta análise, algumas vezes será realizada sobre as classes definidas no MDC como *Enumeration*. Esta análise na TCP é descrita pelo seguinte formalismo:

$$G_{PS} : CS (C_{mi}::E_m, C_{nj}::E_n) \Rightarrow \exists \text{ PROPRIEDADE} |$$

$$C_{mi}.ATRIBUTO::E_m \cong C_{nj}.ATRIBUTO::E_n \vee C_{mi}.ATRIBUTO::E_m = C_{nj}.ATRIBUTO::E_n \wedge$$

$$C_{mi}.TIPO::E_m \cong C_{nj}.TIPO::E_n \vee C_{mi}.TIPO::E_m = C_{nj}.TIPO::E_n \wedge$$

$$C_{mi}.FORMATO::E_m \cong C_{nj}.FORMATO::E_n \vee C_{mi}.FORMATO::E_m = C_{nj}.FORMATO::E_n \wedge$$

$$C_{mi}.RESTRICÇÃO::E_m \cong C_{nj}.RESTRICÇÃO::E_n \vee C_{mi}.RESTRICÇÃO::E_m = C_{nj}.RESTRICÇÃO::E_n \wedge$$

$$C_{mi}.DOMÍNIO::E_m \cong C_{nj}.DOMÍNIO::E_n \vee C_{mi}.DOMÍNIO::E_m = C_{nj}.DOMÍNIO::E_n$$

Cabe ressaltar ainda que nesta etapa devem ser observadas as propriedades comuns que são usadas como chaves de acesso, ou que designam códigos para facilitar o acesso quanto as suas definições. De modo geral, as propriedades chaves ou codificadas serão diferentes, devendo ser tratadas na conformação. Para ilustrar, sejam as propriedades ilustradas abaixo na Tabela V.22:

Tabela V.22- Exemplo: Tabelas de correspondência entre as propriedades.

E ₁ :GT _p :RioPrinc					E ₂ :GT _p :Rio				
ATRIBUTO	TIPO	FORMATO	RESTRICÇÃO	DOMÍNIO	NOME	TIPO	FORMATO	RESTRICÇÃO	DOMÍNIO
IdRio	Integer	2			CódigoRio	Integer	5		
Nome	String	25			Denominação	String	40		
Perímetro	Real	6		Km	Perímetro	Real	6		Metro
Profun	Integer	3		Metro	Profundidade	Real	8		Metro
Vazão	Integer	6		m ³ /s	Volume	Real	32		m ³
E ₁ :GT _p :Afluente					E ₂ :GT _p :Drenagem				
ATRIBUTO	TIPO	FORMATO	RESTRICÇÃO	DOMÍNIO	NOME	TIPO	FORMATO	RESTRICÇÃO	DOMÍNIO
IdRio	Integer	2			CódigoRio	Integer	5		
Nome	String	25			Denominação	String	40		

- **Tratamento do relacionamento semântico encontrado entre os esquemas**

A análise do contexto semântico permite descobrir a correspondência de similaridade das classes. Para facilitar a elaboração do esquema global interno, a metodologia propõe duas estratégias.

A **primeira estratégia** consiste em atribuir a cada par de classes, encontrado na etapa de análise, um grau de proximidade semântica (G_{ps}) parcial. Este grau, no futuro, deverá ser utilizado na automação das etapas da *eMultiSIG*, como mecanismo para identificar a proximidade entre as classes.

Ao se tentar definir um critério para o grau de proximidade semântica parcial deve-se levar em consideração a riqueza de heterogeneidade que pode se apresentar em uma análise de esquemas. Por mais estudo e experiência que se tenha é necessário considerar a existência de um caso que não tenha sido contemplado. Desta forma, a solução adotada neste trabalho foi a adoção de um critério flexível, estabelecido heurísticamente, para cada etapa da análise semântica, apresentado na Tabela V.23, que deverá ser testado e avaliado na aplicação da metodologia de forma a verificar a sua adequação.

Tendo em vista que a maior proximidade semântica entre as classes é detectada quando as classes apresentam o mesmo termo para descrevê-las, a metodologia propõe que as classes para as quais é identificado uma proximidade semântica nos termos ($G_{PS:TERMO}$), sejam atribuídos um número real no intervalo de 0 a 4, inclusive.

Tabela V.23- Graus parciais de proximidade semântica.

G_{PS}	Critério
0 a 4	Termo
0 a 2	Relacionamento
0 ou 2	Geometria
0 a 2	Propriedades

O par de classes que possui um $G_{PS:TERMO}$ é analisada quanto ao contexto semântico de seus relacionamentos. A metodologia, então, propõe que seja adicionado um $G_{PS:RELACIONAMENTO}$, designado por um real no intervalo de 0 a 2, inclusive. Este grau, para facilitar, é fornecido pela seguinte expressão:

$$G_{PS:RELACIONAMENTO} = \frac{2 \cdot RC}{NR}$$

Onde:

RC = número de relacionamentos comuns do par de classes que está sendo analisado; e

NR = número total de relacionamentos do par de classes que está sendo analisado.

O par de classes para o qual foi detectada uma proximidade dos termos e uma proximidade dos relacionamentos, a metodologia propõe que seja analisada a proximidade semântica quanto a geometria ($G_{PS:GEOMETRIA}$) e atribuído um valor de 0 ou 2, caso a geometria seja diferente ou igual, respectivamente.

Para o par de classes com proximidade de termos, relacionamentos e geometria, a metodologia propõe a análise das propriedades. Para esta análise, a MultiSIG

atribui um grau de proximidade semântica quanto as propriedades ($G_{PS:PROPRIEDADES}$) que pode variar no intervalo de reais de 0 a 2, inclusive. Este grau é estabelecido em função do número de propriedades comuns, através da seguinte expressão:

Onde:

$$G_{PS:PROPRIEDADES} = \frac{2 \cdot RC}{NR}$$

RC= número de propriedades comuns do par de classes que está sendo analisado;

NR= número total de propriedades do par de classes que está sendo analisado.

O grau de proximidade semântica total ($G_{PS:TOTAL}$) entre o par de classes pertencentes a dois esquemas será dado por:

$$G_{PS: TOTAL} = G_{PS: TERMO} + G_{PS: RELACIONAMENTO} + G_{PS: GEOMETRIA} + G_{PS: PROPRIEDADES}$$

A **segunda estratégia** consiste em função do grau de proximidade semântica separar as classes em três grandes grupos de proximidade semântica, a saber:

- *Grupo de classes idênticas* - encontram-se os pares de classes que apresentam o grau de proximidade igual a 10. Neste caso, as classes apresentam uma exata combinação de suas definições com nenhuma perda da informação;
- *Grupo de classes similares* - encontram-se os pares de classes que apresentam equivalência ou compatibilidade representadas por um grau de proximidade entre 6 e 10, exclusive. Em geral, elas apresentam equivalência semântica, mas não apresentam equivalência estrutural. Isto ocorre quando não há uma exata combinação semântica entre as classes e suas definições nos dois esquemas. Neste grupo, os pares de classes apresentam conflitos de denominação, de abstração usada para modelar ou entre as suas propriedades; e
- *Grupo de classes heterogêneas* - neste grupo se encontram as classes para as quais foram encontrados o grau de proximidade semântica inferior a 6. Para as classes deste grupo é convencional não possuírem nenhuma proximidade semântica entre elas.

As classes idênticas são organizadas em uma tabela, denominada de tabela de classes idênticas (Tabela V.24), com suas respectivas denominações e propriedades. As classes de cada esquema que possuem o relacionamento de identidade são dispostas na tabela, na mesma linha, para facilitar o tratamento.

As classes similares são colocadas em uma tabela, denominada de tabela de classes similares (Tabela V.25), com suas respectivas denominações e propriedades e

um campo destinado a representar a função de mapeamento entre as classes e suas propriedades.

Tabela V.24- Tabela de classes Idênticas.

E _m		E _n	
Classe	Lista (Propriedades)	Classe	Lista (Propriedades)

Os pares de classes deste grupo devem ser disposta na tabela, na mesma linha, para facilitar o tratamento. Esta tabela é encaminhada para os administradores locais e estes identificam as possíveis soluções para os conflitos, de modo a estabelecer a concordância no conjunto compartilhado de termos e definições utilizadas em ambos os domínios.

Tabela V.25- Tabela de classes similares.

E _m			E _n		
Classe	Lista (Propriedades)	Mapeamento	Classe	Lista (Propriedades)	GPS

As classes heterogêneas são descritas em uma outra tabela com todas as suas propriedades (Tabela V.26). Cabe ressaltar que neste grupo devem ser incluídas as classes dos esquemas para os quais foi detectado o relacionamento espacial vazio, apresentado na seção V.8.2.1.

As classes idênticas serão integradas em uma classe virtual, na etapa de integração, com suas propriedades e respectivos mapeamentos de acesso para os esquemas locais. As classes similares serão analisadas na conformação, de modo a resolver os conflitos e estabelecer um mapeamento compreensivo entre as classes. As classes heterogêneas serão tratadas como classes virtuais a serem modeladas no modelo de dados proposto na seção V.6.4.

Tabela V.26- Tabelas de classes heterogêneas.

E _m	
Classes	Lista (Propriedades)

E _n	
Classes	Lista (Propriedades)

V.9.3. Etapa 3: Conformação

Esta etapa tem por objetivo encontrar soluções para os conflitos semânticos e estruturais encontrados nos pares de classes do grupo com indicação de similaridade. Estas soluções devem conformar os objetos geográficos de cada esquema em uma(s) classe(s) virtual(is). Estas classes são definidas através do procedimento de cooperação, que suporta a troca de informações entre os administradores locais.

Após a organização das tabelas, o administrador global deverá estabelecer um fluxo de cooperação entre os administradores locais dos esquemas, de modo a chegarem um consenso sobre as soluções para resolver as heterogeneidades entre as classes com indicação de similaridade. Estas soluções dependerão da negociação entre os administradores locais.

Contudo, a metodologia determina que inicialmente devam ser tratados os conflitos de denominação ao nível de classes e propriedades, a seguir os conflitos estruturais entre as propriedades e, finalmente, os conflitos decorrente da abstração usada na modelagem. Desta forma, a seqüência para tratar as heterogeneidades e ajudar a obtenção de soluções consiste em:

1. Tratar os conflitos de denominação de classes através da padronização dos nomes. A metodologia propõe que as classes empreguem uma denominação genérica, singular, conotativa, prontamente compreendida por todos na federação. Esta denominação é criada a partir do termo selecionado pelos administradores para descrever a classe. No caso de classes idênticas é acordado entre os administradores uma denominação comum a todos os usuários;
2. Tratar os conflitos de denominação das propriedades das classes, de modo análogo ao item anterior;
3. Tratar os conflitos devido a diferença de geometria como uma generalização de classes, de acordo com a geometria;
4. Tratar os conflitos entre as propriedades comuns relacionados as diferenças de tipo, formato, unidades, restrição e domínio, com o objetivo de torná-las compatíveis. Para isto, deve ser determinada funções de mapeamento (F) entre os esquemas que relacionam os objetos de um par de classes com similaridade semântica ($E_m::C_i, E_n::C_j$), a saber:

$$F(E_m::C_i) = E_n::C_j$$

Esta função torna compatível os objetos de uma classe, de um dado esquema, em objetos de outra classe pertencente a outro esquema. Estas funções são criadas para resolver os conflitos relacionados a:

- diferenças de tipo ou de formato - deve ser estabelecido entre os administradores locais um tipo ou um formato comum e a função de transformação para cada esquema local;
 - diferenças de unidades - deve ser estabelecido com os administradores locais uma unidade comum e a função transformação para cada esquema local;
 - diferença de restrição - deve ser estabelecido com os administradores locais uma função de transformação e mapeamento que resolva a diferença de restrição entre os esquemas;
 - diferenças de domínio - deve ser preservada cada propriedade em seu domínio e criada duas propriedades mapeadas para cada esquema local com uma denominação comum, acordada entre os administradores, que o identifique. No caso de se tratar de uma correspondência de 1:1 entre domínios, o mapeamento pode ser implementado através de conversão numérica ou uma tabela de equivalência de valores.
5. Analisar os conflitos de abstração entre as classes e as propriedades, já tratadas nos itens anteriores. Nesta análise devem ser reconhecidas as propriedades que estão presentes no par de classes com indicação de similaridade. Esta análise poderá resultar na seguinte identificação:
- ◆ as propriedades compatibilizadas são comuns;
 - ◆ existência de propriedades implícitas;
 - ◆ nem todas as propriedades são comuns; e
 - ◆ apesar das classes terem a mesma semântica não há propriedades comuns;

Esta análise indica a necessidade das classes serem organizadas, usando as abstrações de generalização/especialização e agregação, a ser aplicada na próxima etapa.

V.9.4. Etapa 4: Integração no esquema global interno

Nesta etapa são definidas como as classes, que se encontram nas tabelas de classes idênticas, similares (conformadas) e heterogêneas, serão modeladas em classes virtuais em um esquema global interno.

Inicialmente, será determinada como as classes similares e conformadas serão representadas nas classes virtuais. Isto é estabelecido através de consenso entre os administradores locais. Todavia, estas soluções poderão ser:

- ❖ Representação do par de classes similares por uma classe virtual, com todas as propriedades comuns compatibilizadas - é criada uma classe virtual contendo a simples união das propriedades do par de classes analisados;
- ❖ Representação do par de classes similares por uma hierarquia de generalização/especialização de classes virtuais - as classes que possuem propriedades comuns, ou as propriedades compatibilizadas nos esquemas são verificadas, de modo a identificar redundâncias que possam ser removidas através da hierarquia de generalização, empregando o mecanismo de herança. Uma das classes no esquema integrado torna-se especialização da outra. Para isto, esta solução analisa a similaridade de denominações adotadas na conformação entre as propriedades. No caso das classes apresentarem interseção de suas propriedades serão criadas duas ou mais classes virtuais. A primeira especialização com as propriedades comuns e as demais com as propriedades que não possuem equivalência estrutural; e
- ❖ Representação do par de classes similares por uma composição de classes virtuais - neste caso é identificado a agregação de classes através do relacionamento de composição, no qual vários objetos formam um objeto de nível semântico maior, onde cada objeto tem funcionalidade própria. Desta forma, esta abstração consiste em criar um objeto composto, formado por agrupamento de objetos. Este relacionamento é materializado nas classes virtuais por propriedade com o tipo *set* ou *list*.

A seguir, todas as classes do par de esquemas, inclusive as que se encontram nas tabelas de classes idênticas e heterogêneas, serão descritas como classes virtuais.

A ~~o~~MultiSIG estende a SAIF CSN com cláusulas abaixo para descrever as classes virtuais do esquema global interno, a saber:

Superclass: Descreve o nome da superclasse virtual, caso seja estabelecida

	uma hierarquia de classes virtuais.
<i>Class:</i>	Descreve o nome da classe virtual, determinado em consenso entre os administradores locais.
<i>Region:</i>	Lista de <i>Region</i> . Define para que regiões geográficas a classe existirá.
<i>Attributes:</i>	Lista de atributos. Define a estrutura da classe.
<i>Geometry:</i>	Se a classe virtual tiver representação gráfica, esta cláusula define como a geometria do objetos será recuperada.
<i>Source schema mapping method:</i>	Descreve para cada região geográfica listada em <i>Region</i> o mapeamento da estrutura da classe para os esquemas locais
<i>Export schema source:</i>	A partir desta cláusula serão definidas as classes dos esquemas de exportação que deram origem a classe virtual, através das seguintes sub-cláusulas:
<i>Origin ClassN:</i>	Nome da classe no esquema de exportação seguido por ":: <i>nome do esquema</i> "
<i>Attributes:</i>	Descreve a lista de atributos da classe no esquema de exportação. No caso de ser uma classe especializada de GGO, isto é, com representação gráfica, será definida a propriedade <i>position</i> . <i>Geometry</i> .
<i>Relationships:</i>	Descreve os relacionamentos que envolvem a classes no esquema de exportação.
<i>Source.DBInfo.dbIdentifier:</i>	Descreve o nome local da classe e sua localização na rede.
<i>Acess method:</i>	Descreve a forma de acesso desta classe origem ao seu esquema local.
<i>Mapping method:</i>	Descreve o mapeamento entre as classes dos esquemas de exportação determinada na conformação.

Nesta descrição, as classes são definidas quanto as classes originais e os respectivos métodos que descrevem as funções de mapeamentos entre os esquemas, definidas na análise espacial e na conformação, e de acesso. A função de mapeamento definida na análise espacial consiste em descrever em que região geográfica a classe e suas propriedades são válidas. A função de acesso tem por objetivo fornecer os

mecanismos para recuperar os dados locais, empregando a sintaxe da linguagem de consulta local, com as devidas operações locais de recuperação dos dados.

Para ilustrar, é assumido que as classes dos esquemas E_1 e E_2 , apresentadas nos exemplos anteriores, pertencem a esquemas que tem como espaço geográfico o Estado do Rio de Janeiro e a Bacia do Rio Paraíba do Sul, respectivamente. Desta forma, a classe RioPrincipal, conformada e integrada nos exemplos anteriores, apresentará a seguinte notação:

Superclass: Rio

Class: RioPrincipal

Region: Estado do Rio de Janeiro, Bacia do Rio Paraíba do Sul

Attributes: Código, Nome, Perímetro, Largura, Vazão, Profundidade, Nascente

Geometry: Linha

Source schema mapping method:

Region = Estado do Rio de Janeiro {

Código	IdRio.RioPrin:: E_1 ; CódigoRio.Rio:: E_2
Nome	Nome.RioPrin:: E_1 ; Denominação.Rio:: E_2
Perímetro	Perímetro.RioPrin:: E_1 ; Perímetro Rio:: E_2
Largura	Largmed.RioPrin:: E_1
Vazão	Vazão.RioPrin:: E_1 ; Volume.Rio:: E_2
Profundidade:	Profun.RioPrin:: E_1 ; Profundidade.Rio:: E_2
Nascente	Nascente.RioPrin:: E_1 ; }

Region = Bacia Hidrográfica do Rio Paraíba do Sul {

Código	CódigoRio.Rio:: E_2
Nome	Denominação.Rio:: E_2
Perímetro	Perímetro Rio:: E_2
Largura	"Informação não disponível para Região"
Vazão	Volpseg.Rio:: E_2
Profundidade:	Profundidade.Rio:: E_2
Nascente	"Informação não disponível para Região"}

Export schema source:

OriginClass1: RioPrin:: E_1

Attributes: IdRio, Nome, Perímetro, Larg_med, Vazão, Profun, Nascente

position.Geometry: VectorLine

Relashionships:

<i>Rel</i>	<i>ObjectA</i>	<i>ObjectB</i>	<i>Attributes</i>	<i>Car</i>
------------	----------------	----------------	-------------------	------------

Contém	RioPrin	Afluente	IDRio	1:n
--------	---------	----------	-------	-----

Source.DBInfo.dbIdentifier: Nome da máquina .nome do arquivo
Access method: comandos para recuperar os dados no esquema local
OriginClass2: Rio::E₂
Attributes: CódigoRio, Denominação, Perímetro, Volume, Profundidade
position.Geometry: VectorLine
Relationships:

<i>Rel</i>	<i>ObjectA</i>	<i>ObjectB</i>	<i>Attributes</i>	<i>Car</i>
Pertence	Drenagem	Rio	CódigoRio	1:1

Source.DBInfo.dbIdentifier: Nome da máquina .nome do arquivo
Access method: comandos para recuperar os dados no esquema local
Mapping method: função para transformar volpseg(cm³/s) em Vazão(m³/s)

Se ocorrer de uma propriedade não estar disponível para uma dada região, em vez do mapeamento para o esquema local, será colocada a seguinte mensagem "Informação não disponível para Região".

V.9.5. Etapa 5: Criação do esquema global externo

Nesta etapa, as classes virtuais definidas na etapa anterior serão transformadas em instâncias do modelo de dados proposto na seção V.6.4 para criar um esquema global externo. Este esquema encapsula através do esquema global interno as localizações dos dados nos esquemas locais, os mapeamentos determinados na etapa de conformação para resolver os conflitos e os métodos de acesso para os esquemas locais.

O modelo de dados do esquema global externo descreve os objetos geográficos a serem disponibilizados na federação em duas hierarquias, *Region* e *Theme*. Para facilitar a definição deste esquema, a  MultiSIG divide esta etapa em duas atividades: a primeira define as instâncias da hierarquia *Region* e a segunda as instâncias da hierarquia de *Theme*.

Estas atividades são realizadas pelo administrador global. A partir das classes virtuais, ele determina a melhor forma de transformá-las em instâncias ou classes especializadas de *Themes* e a forma de associá-las às classes especializadas de *Region*.

V.9.5.1. Definição das instâncias da hierarquia *Regions*

A hierarquia *Region* é definida a partir da análise do contexto espacial entre os esquemas (seção V.9.2.1). Para cada espaço geográfico, detectado na análise espacial, será pesquisado na hierarquia básica proposta para *Region* a existência de classes que definam os espaços geográficos a serem tratados pela federação. Caso não existam, deverá ser criada uma nova subclasse que descreva o espaço geográfico desejado. A seguir, cada subclasse encontrada é instanciada com as regiões resultantes da análise espacial e sua respectiva lista de *Theme* e escala resultante da *MultiSIG*.

A *MultiSIG* propõe, por simplicidade, utilizar a notação para descrever objetos do SAIF (OSN). Esta notação consiste em descrever para cada instância o nome da superclasse *Region*, seguido de um parênteses, dentro do qual será definido o nome da subclasse, seguido novamente de parênteses, dentro do qual são descritos os valores de suas propriedades: *name*, *inverseScale*, *theme*, a saber:

```
region: Region (Subclass(  
    name: String  
    inverseScale: Integer  
    theme: List (Theme)  
    AcessMethod:))
```

Observa-se que o modelo proposto pela *MultiSIG* descreve as instâncias da hierarquia *Region* que definem os espaços geográficos resultantes da integração dos esquemas, de acordo com as escalas em que os dados se encontram nos esquemas locais, conforme observado na notação acima.

O método de acesso tem a finalidade de recuperar os dados no esquema local e é definido da mesma forma que na classe virtual.

Para ilustrar, seja exemplo anterior, para o qual será assumido agora que os esquemas E_1 e E_2 , descrevem o espaço geográfico na escala de 1:100.000, ficando a notação da instância, referente *Region* no esquema global externo, da seguinte forma:

```
region: Region (Estado(  
    name: Rio de Janeiro  
    inverseScale: 100.000  
    theme: List (RioPrincipal)  
    AcessMethod:))
```

Cabe ressaltar que, no caso do espaço geográfico não será tratado pelas subclasses de *Region*, deverá ser criada uma nova subclasse especializada desta, associada à classes especializadas de *Theme*. No caso da existência do espaço geográfico ser tratado por alguma das subclasses existentes na hierarquia *Region*, deverá ser identificada a classe exata e recuperada a notação dos *Themes*, de modo a verificar a existência de temas já tratados pela federação sobre o mesmo espaço geográfico e na mesma escala.

V.9.5.2. Definição das instâncias da hierarquia de *Theme*

As instâncias da classe *Theme* são definidas em função das classes virtuais descritas pela notação proposta pela hierarquia *MultiSIG*. Cada classe virtual será uma instância do modelo de dados proposto para o esquema global externo.

As instâncias da classe *Theme* também serão descritas pela notação para descrever objetos do SAIF (OSN), conforme a seção anterior, a saber:

```
theme: Theme(Subclass(  
    name: String  
    attributes: List(String)  
    geometry: String  
    regions: List (Region)  
    inverseScale: Integer  
    accessMethod:  
    mappingMethod: ))
```

Os métodos de acesso e de mapeamento são os mesmos definidos para as classes virtuais, na atividade de integração.

Assim, após estabelecer a modelagem das classes no modelo de dados do esquema global interno, cada classe criada deverá ser descritas na notação acima proposta para descrever o esquema global externo e estabelecida as funções de acesso para o esquemas locais.

V.9.5.3. Consideração sobre novos esquemas a serem inseridos na federação

As regiões geográficas resultantes da análise do relacionamento espacial devem ser comparadas às classes da hierarquia *Region* do esquema global externo, de forma a

verificar se existe outro esquema já pertencente a federação que trate do mesmo espaço geográfico.

Caso o espaço geográfico não seja contemplado pelas classes da hierarquia *Region*, será criada uma especialização que contemple o espaço geográfico desejado. Na atividade de criação do esquema global externo esta classe será instanciada e associada aos *Themes* do esquema de exportação.

Caso haja uma classe que contemple o espaço geográfico, será verificada a existência de uma instância. Caso não haja a instância que contemple o espaço geográfico desejado, será criada uma nova instância e associada aos *Theme*. Caso o espaço geográfico já exista instanciado em uma das classes da hierarquia *Region*, o esquema federado, expresso no MDC, deve ser recuperado. A seguir, a metodologia deve ser aplicada a parte do esquema federado e do esquema de exportação, de modo a detectar o relacionamento semântico entre os esquemas para integrá-los.

V.9.6. Etapa 6: Validação

Esta etapa tem por objetivo analisar o esquema integrado, a fim de assegurar que os esquemas globais interno e externo apresentem as classes descritas pelos administradores locais no SAIF, que elas estão corretamente disponibilizadas e que o esquema global externo é facilmente compreendido por todos os usuários na federação.

Desta forma, a metodologia propõe três atividades para assegurar as métricas propostas por ÖSZU & VALDURIEZ, (1990) e SCHMITT & SAAKE (1996), de completude, exatidão, minimalidade e compreensão do esquema global.

A primeira atividade consiste em verificar se todas as classes descritas no SAIF se encontram nas tabelas de classe, e se todas as classes descritas nestas tabelas se encontram descritas na notação proposta para as classes virtuais resultante da integração. Esta atividade visa assegurar a métrica de completude.

A segunda atividade consiste em verificar se todas as classes virtuais descritas na notação proposta para o esquema global interno, estão corretamente descritas em relação aos seus esquemas de exportação originais. Esta atividade visa assegurar a métrica de exatidão.

A terceira atividade consiste em verificar se existem classes virtuais redundantes, bem como instâncias das classes do esquema global externo redundantes. Esta atividade objetiva atender a métrica de minimalidade.

Além destas atividades, deve-se avaliar junto aos usuários da federação a compreensão do esquema global externo. Isto é obtido através de observações realizadas pelos usuários da federação, ao realizar consultas sob o esquema global externo.

V.10. Aplicação da metodologia

Para ilustrar a aplicação da metodologia, foram selecionados quatro esquemas de bases de dados geográficas, apresentadas no **Apêndice A**, e comparados, utilizando o **MultiSIG** no **Apêndice B**, a saber:

- Modelagem de dados para o Programa de Desenvolvimento Racional, Recuperação e Gerenciamento Ambiental da Bacia Hidrográfica (Projeto Pró-Guaíba), realizada por GARRAFA (1998) para a Secretaria de Coordenação e Planejamento do Estado do Rio Grande do Sul;
- Modelagem de Dados para Projeto de Qualidade de Águas e Controle de Poluição Hídrica do Ministério e Planejamento e Orçamento, Projeto Estadual de Investimentos para o Rio de Janeiro/Bacia do Rio Paraíba do Sul, realizado no âmbito do Laboratório de Hidrologia da COPPE/UFRJ (LIRA, 1998). Este projeto objetiva o monitoramento e gerenciamento de informações de Recursos Hídricos (Projeto SIG-RHidro);
- Modelagem de dados realizada pela autora para o Estudo do Monitoramento ambiental do Estado do Rio de Janeiro (Projeto **Ambiental**); e
- Modelagem do Projeto Guaíba, aplicado pela autora à Bacia do Rio Paraíba do Sul (Projeto **Paraíba**).

A seleção destes esquemas procurou estudar todos os possíveis casos de heterogeneidades, apresentados na Tabela V.27:

Tabela V.27- Heterogeneidades espaciais e de áreas de aplicação entre os esquemas.

Critério	Esquemas
=Área de Aplicação ∨ =Região Geográfica	Paraíba X RHidro
=Área de Aplicação ∨ ≠Região Geográfica	Paraíba X Guaíba RHidro X Guaíba
≠Área de Aplicação ∨ =Região Geográfica	Ambiental X Paraíba Ambiental X RHidro

Tendo em vista a complexidade e extensão da *MultiSIG*, sua aplicação é apresentada no Apêndice B.

V.11. Comparação com outros trabalhos relacionados na literatura

Aplicar uma metodologia para integração de bases de dados convencionais (SOUZA, 1986, BONJOUR & FALQUET, 1994; SPACCAPIETRA & PARENT, 1994; FRANKHAUSER, 1995) para a integração de bases de objetos geográficos não é suficiente, uma vez que estas metodologias não consideram os aspectos gráficos para a representação dos esquemas, nem a diversidade e riqueza de representação semântica dos dados geográficos armazenados por estas bases de objetos geográficos.

Na literatura são encontradas algumas metodologias para integrar esquemas de bases de objetos geográficos, conforme apresentado na seção IV.2.2. (AGUIAR, 1995; PARENT *et al.*, 1996; YASER, 1997).

AGUIAR (1995) enfatiza os mapeamentos dos modelos de dados locais para um modelo global, entretanto ela não trata as heterogeneidades semânticas relacionadas aos esquemas das bases de dados dos SIGs. A *MultiSIG* tem por objetivo gerar um esquema global para uma federação aberta, onde novas bases de dados poderão participar desta federação. Desta forma, a *MultiSIG* contempla as heterogeneidades semânticas implícitas e explícitas em um esquema de base de dados geográficos. Para isto, propomos três grupos de atividades de análise sobre os esquemas de exportação, de modo a considerar todos os aspectos relacionados à semântica da representação gráfica dos objetos geográficos, à semântica da área de aplicação e à semântica do esquema.

PARENT *et al.* (1996) propõem uma metodologia que tem como um ponto forte a investigação de correspondência e as soluções propostas para a integração. Todavia, esta investigação é realizada sobre os esquemas enriquecidos com informações complementares. A *MultiSIG* para normalizar os esquemas e facilitar as etapas de análise adota um padrão para troca de dados geográficos como o MDC, o qual é estendido com algumas propriedades para suportar a descrição dos metadados sobre os projetos relacionados ao esquema. Este padrão permite descrever os esquemas locais, no nível intencional, e os atributos dos esquemas, no nível extencional. Com isso ele

introduz a vantagem de utilizar um único modelo que descreve todas as características, mesmo que implícitas, do esquema da base de dados do SIG local.

A *MultiSIG* para facilitar a identificação dos conflitos propõe uma taxionomia destes e sobre os esquemas de exportação, expressos no MDC selecionado, são desenvolvidas atividades que geram tabelas, de modo a facilitar a identificação da proximidade semântica entre os esquemas. A *MultiSIG* propõe, para cada tipo de conflito soluções que devem ser acordadas entre os administradores locais.

Cabe ressaltar que uma taxionomia de conflitos, entre esquemas de dados geográficos e o conjunto de soluções, exige um amplo estudo e experiência em análise de esquemas de diferentes áreas de aplicação. Ainda assim, esta taxionomia não esgotará a diversidade de conflitos, devido a riqueza semântica dos dados geográficos. O mesmo acontece com as soluções para os conflitos, uma vez que sempre existirá uma outra forma de acordar entre os administradores locais. Em virtude disto, a *MultiSIG* é uma metodologia que propõe um trabalho cooperativo e interativo entre os administradores locais e o administrador global.

YASER (1997) propõe uma metodologia para compartilhar dados que tem como principal característica considerar a semântica da área de aplicação. Esta consideração é importante, pois esquemas que possuem proximidade de áreas de aplicação, em geral, apresentam uma semântica mais aproximada do que outro que não seja da mesma área de aplicação. Porém, o autor não usa um MDC. Ele constrói uma hierarquia semântica, de acordo com a área de aplicação e disponibiliza a especificação sintática dos objetos geográficos de cada base de dados local para realização de consultas. Isto requer do usuário a compreensão do esquema local.

Assim como YASER(1997), a *MultiSIG* também considera a proximidade semântica entre esquemas que apresentam a mesma área de aplicação. Entretanto, para facilitar a consulta a *MultiSIG* propõe um modelo de dados para o esquema global externo que elimina a necessidade do usuário compreender os esquemas locais. O modelo de dados do esquema global externo no **Sistema MultiSIG** tem a finalidade de encapsular a modelagem dos esquemas locais e os modelos de dados dos SIGs locais e permitir ao usuário localizar os objetos geográficos, através da sua semântica espacial ou temática.

Todas as metodologias encontradas na literatura têm como característica a especificidade para atender a uma federação fechada, formada por comunidades de

informações pré-definidas que tratam do mesmo espaço geográfico. Assim, as metodologias encontradas contemplam características restritas das bases de dados geográficos. A MultiSIG se destaca destas metodologias, devido à amplitude de suas atividades para contemplar todos os aspectos da modelagem, referentes à representação da semântica e ao conceito de espaço geográfico, associado a cada comunidade de informação. Essa necessidade ficou evidente no estudo de caso tratado e é a realidade de diversas empresas que trabalham com os SIG, como por exemplo, o SIG-Solos da Embrapa Solos.

Capítulo VI - *MultiSIG*: Arquitetura proposta para a interoperabilidade semântica

VI.1. Introdução

O presente trabalho tem por objetivo pesquisar e desenvolver um sistema que possibilite a interoperabilidade entre bases de dados de SIGs heterogêneas e distribuídas, sem provocar o mínimo de mudanças nos locais.

A *MultiSIG* é arquitetura proposta para o **Sistema MultiSIG** que oferece um ambiente MBD, sobre o qual os usuários da federação, através de um esquema global externo, poderão realizar consultas sobre os vários esquemas participantes da federação.

Este esquema global externo disponibiliza classes virtuais de *Theme* e *Region*, resultantes da metodologia de integração de esquemas de bases de dados de SIG, apresentada no capítulo anterior. Para isto, a *MultiSIG* oferece uma interface com o usuário que apresenta o esquema global externo, sobre o qual poderão ser formuladas consultas às bases de dados geográficos empregando uma linguagem de consulta gráfica. Esta linguagem, oferece mecanismos para explorar os conteúdos temáticos e espaciais dos objetos geográficos distribuídos sobre a federação.

As consultas são traduzidas para os SIGs locais e distribuídas as bases de dados geográficos, através de métodos para: *i)* acessar as múltiplas formas dos dados armazenados sob diversos ambientes dos SIGs locais; *ii)* recuperar os dados desejados; *iii)* integrar o resultado da consulta; e *iv)* apresentar os resultados, suportando visualização gráfica destes na forma de mapa e/ou tabela.

Como os SIGs apresentam sistemas de gerencia próprios e linguagens de consultas nativas, o **Sistema MultiSIG** para alcançar a interoperabilidade implementa uma arquitetura cliente-servidor orientada a objeto, que suporta a distribuição e recuperação dos dados nas bases locais, de forma segura e eficiente.

Desta forma, este capítulo apresenta a *MultiSIG*, proposta no quarto capítulo, quanto aos seus objetivos, suas características, as funcionalidades pretendidas, seus

componentes, abordagem de implementação e os serviços oferecidos pelo sistema. A seguir, apresenta telas e exemplo de consultas ao esquema global externo, gerado pela aplicação da metodologia no capítulo anterior, apresentada no Apêndice B.

VI.2. Objetivo da *MultiSIG*

A *MultiSIG* tem por objetivo proporcionar interoperabilidade, ao nível físico, entre bases de dados de SIG participantes da federação proposta pelo **Sistema MultiSIG**.

Desta forma, a *MultiSIG* propõe um ambiente MBD que possibilita o compartilhamento de dados provenientes de SIGs heterogêneos e distribuídos sobre a rede global Internet, de forma eficiente e segura. Este compartilhamento é realizado através do conjunto de ferramentas para a realização de consultas sobre um esquema global.

VI.3. Características da *MultiSIG*

A *MultiSIG* pretende atender os objetivos do **Sistema MultiSIG** e do Projeto WHIP2. Ela implementa um ambiente MBD que se utiliza dos recursos de plataformas e aplicativos heterogêneos navegando na rede global Internet, para acessar bases de dados de SIGs heterogêneos.

A escolha desta rede para implementar a federação aberta, proposta pelo **Sistema MultiSIG**, se deu pelo fato desta rede ser bastante democrática, permitindo, qualquer usuário ou comunidade de informação, a um custo relativamente baixo e de forma simples, compartilhar os seus dados e usufruir dos serviços de informação.

A utilização desta rede, nesta arquitetura, promove interoperabilidade entre os SIGs autônomos, heterogêneos e distribuídos que participam da federação, através da troca de mensagens entre os SIG e os componentes da *MultiSIG*. Esta rede também promove a cooperação entre os administradores locais através da comunicação entre administradores locais e o administrador global, proposto pela política de participação, apresentado na seção V.5.

Uma característica proporcionada a *MultiSIG* pela utilização desta rede é uniformidade da interface com o usuário sob, diversas plataformas/sistemas locais. Em

geral cada SIG apresenta uma interface e uma linguagem de consulta própria. A *MultiSIG* oferece uma interface desenvolvida sobre o *browse* desta rede. Esta interface apresenta o esquema global externo. Este esquema contém todos os dados disponíveis na federação e oferece meios para o usuário da federação realizar a consulta, sem precisar conhecer ou compreender a organização original dos dados nos SIG locais.

A *MultiSIG* proporciona interoperabilidade entre os SIGs, preservando as autonomias de projeto, de execução e de comunicação entre os SIG locais participantes da federação proposta pelo **Sistema MultiSIG**.

A autonomia de projeto é assegurada através do esquema global externo apresentado ao usuário. Este esquema é derivado da integração lógica dos esquemas de exportações. A metodologia adotada para derivar este esquema preserva os esquemas locais, uma vez que determina as transformações entre os sistemas para acessar e integrar os resultados, não necessitando alterar as bases de dados locais para torná-las compatíveis. Além disto, os acesso às bases de dados implementados na *MultiSIG* não permite a realização de alterações nas bases de objetos componentes da federação, uma vez que só tem permissão de leitura nas bases de dados locais, garantindo, assim, segurança local.

A autonomia de execução é assegurada pelo mecanismo proposto para tratar o acesso às bases existentes separadamente, isto é, cada processador de consulta local controla a execução das subconsultas às suas base de dados.

A *MultiSIG* proporciona a autonomia de comunicação através do procedimento de cadastro do esquema de exportação e da consulta. Através do procedimento de cadastro, os administradores locais tornam disponíveis as partes dos seus esquemas as quais desejam compartilhar. Através da consulta ao esquema global externo o usuário terá acesso aos dados que se encontram distribuídos na federação. Entretanto, esta autonomia não é transparente para o usuário, uma vez que a consulta é realizada sobre a hierarquia de *Theme* e *Region*, disponíveis na federação. Contudo, o *MultiSIG* possui uma opção (não implementada) para a realização de consulta sobre os metadados, o que tornará possível a consulta a um esquema específico.

Outra característica desta arquitetura é o mecanismo de apoio a cooperação proposto com o objetivo de fornecer suporte as atividades da *MultiSIG* para identificar os esquemas e as classes que apresentam proximidade semântica. Esta cooperação é realizada através de interface específicas e serviços de *e-mails*, oferecidos

pela arquitetura para implementar os procedimentos especificados na política de participação do **Sistema MultiSIG**.

O **Sistema MultiSIG** proporciona uma federação aberta, uma vez que permite que a qualquer momento uma base de dados possa ser incorporada ou retirada da federação. A incorporação é realizada através de interface de cadastro que permite o administrador local descrever seus dados no MDC da federação. A retirada da federação é solicitada pelo administrador local e realizada pelo administrador da federação global, através da retirada dos índices do esquema global externo para as bases de dados locais.

VI.4. Funcionalidades do Sistema MultiSIG

A **MultiSIG** oferece funcionalidades específicas e gerais para os usuários da federação, de acordo com os papéis definidos na política de participação da federação de dados, a saber:

- Funcionalidades para o administrador global:
 - ✓ Criar a federação;
 - ✓ Definir os esquemas participantes da federação;
 - ✓ Criar os usuários de acordo com o seu papel;
 - ✓ Estabelecer comunicação com os administradores locais para aplicar o procedimento de cooperação; e
 - ✓ Cadastrar/Alterar o esquema global externo.
- Funcionalidade para os administradores locais:
 - ✓ Cadastrar o esquema de exportação;
 - ✓ Solicitar a alteração ou remoção do seu esquema de exportação;
 - ✓ Explorar a semântica das classes descritas nos esquemas de exportações; e
 - ✓ Estabelecer comunicação com o administrador global e os demais administradores locais para aplicar o procedimento de cooperação.
- Funcionalidades para os usuários da federação:
 - ✓ Analisar e selecionar classes de *Theme* de interesse;
 - ✓ Analisar e selecionar classes de *Region* de interesse;
 - ✓ Analisar os metadados de um esquema local específico¹;
 - ✓ Submeter consultas temática e/ou espacial;

¹ Especificado, mas não implementado

- ✓ Submeter uma consulta a um esquema local específico;
- ✓ Solicitar dados em um determinado formato²; e
- ✓ Solicitar uma legenda otimizada para a apresentação dos resultados da consulta².

As funcionalidades descritas para os administradores globais e locais são específicas para estes papéis, mas as funcionalidades dos usuários da federação estão também disponíveis para os administradores locais e global.

A *MultiSIG* propõe que o cadastro do esquema exportação seja realizado pelo administrador local, uma vez que esta atividade requer conhecimento sobre o esquema local para estabelecer os mapeamentos deste para o MDC selecionado. Desta forma o **Sistema MultiSIG** oferece interfaces para o administrador local cadastrar os *GenericGeographicObject*, definir os atributos do esquema e o esquema de exportação no MDC, de acordo com a *MultiSIG*.

A autora reconhece a importância da consulta sobre os metadados de um esquema local específico, da solicitação e aquisição de dados com um formato específico e a eficácia da apresentação dos resultados de uma consulta com uma legenda solicitada pelo usuário. Os metadados permitem verificar o conteúdo dos esquemas locais e avaliar a qualidade dos dados locais. A solicitação dos resultados de uma consulta em um formato solicitado pelo usuário, via *file transfer protocol*, por exemplo, supre a necessidade de conversão na aquisição de dados. A especificação da apresentação da consulta facilita, ao usuário da federação, a interpretação do resultado de acordo com a sua semântica. Todavia, tendo em vista não ser este o objetivo em questão, estas funcionalidades, apesar de reconhecidas em uma federação, não serão tratadas no âmbito deste trabalho, o que não impede de futuramente serem implementadas, bem como outras funcionalidades.

VI.5. Linguagem de consulta proposta para o ambiente MULTISIG

A linguagem de consulta em um SIG, também denominada de linguagem de consulta espacial, possui características diferentes das linguagens de consultas convencionais. Ela incorpora técnicas para tratar com propriedades específicas dos objetos geográfico, através de operadores da álgebra georrelacional, para expressar a semântica geográfica. Estas técnicas permitem a linguagem de consulta espacial

² Especificado, mas não implementado

explorar o posicionamento, a geometria e os relacionamentos topológicos, bem como ser integrada a uma interface com o usuário. Outra característica das linguagens de consulta espacial são os mecanismos que elas oferecem para controlar e manipular a apresentação gráfica do resultado das consultas aos objetos geográficos.

No âmbito da tecnologia de Geoprocessamento, não existe linguagem de consulta padrão como a SQL (*Structured Query Language*). Cada SIG possui uma linguagem própria para gerenciar dados geográficos e manipulá-los. Estas linguagens se diferenciam, principalmente, na capacidade semântica, nos graus de liberdade com que elas integram dados gráficos com os não gráficos, e nos algoritmos para a recuperação da informação geográfica.

Uma linguagem de consulta espacial para um ambiente MBD deverá estar acima das diferenças entre os SGBDs das bases de dados geográficos. Ela deverá ser capaz de expressar qualquer consulta às várias bases de dados componentes deste ambiente. Isto exige mais poder que as linguagens de consultas espaciais tradicionais oferecem, uma vez que ela deverá possuir mecanismos para definir o esquema que integra as bases de dados geográficos; para interrogar sobre a estrutura das bases de dados relacionadas, o domínio e a natureza do processo que unirá os dados geográficos, bem como possuir mecanismos para otimizar a execução da consulta.

Segundo BUEHLER & MCKEE^{html} (1996), uma linguagem de consulta para um ambiente MBD, formado por bases de dados de SIG, deve manipular qualquer tipo de objeto geográfico, bem como seus atributos e operações distribuídas, com semântica espaço-temporal sobre eles. Além disto, ela deve apoiar as soluções para os conflitos semânticos e sintáticos dos dados gráficos e não gráficos dos objetos geográficos. Assim, uma linguagem de consulta para uma federação de bases de dados de SIG deverá ter ferramentas para manipulação das componentes gráficas e não gráficas, dos objetos geográficos provenientes de diversos SIGs.

Uma linguagem de consulta espacial para este ambiente requer a combinação da geometria computacional, álgebra de objetos e técnicas de análise espacial. Assim, tendo em vista não ser o objetivo principal deste trabalho e a não disponibilidade de recursos adequados para manipulação de mapas na WEB, a MultiSIG utiliza uma interface gráfica dirigida a eventos, comum a todos os usuários, para realizar consultas sobre o esquema global externo. Sobre esta interface é utilizado um modelo de consultas baseado em OQL (*Object Query Language*) que permite o usuário selecionar o objeto

geográfico desejado. Esta OQL é estendida com a cláusula **Region** para suportar a especificação da consulta sobre o esquema global externo adequadamente, a saber:

```
SELECT <atributos>
FROM <Theme>
WHERE <condição>
BY REGION <Region>
```

A interface de consulta constitui um mecanismo de consulta padrão que contempla as diferenças entre as linguagens de consulta dos SIGs. Ela oferece, no protótipo inicial operadores lógicos pré-definidos para realizar consultas sobre o esquema global. À esta interface poderão ser inseridos, futuramente, operadores que suportem operações espaciais.

Esta consulta ao ser realizada é passada ao processador de consultas, que por sua vez, desmembra as consultas em subconsultas, empregando os mapeamentos estabelecidos na metodologia e as funções de acesso necessárias para a recuperação dos dados nos SIGs locais.

Para formalizar, seja uma consulta global (CG) constituída por uma tupla formada pelo operador (O), o tema (T), a região geográfica (R) os atributos (At_i) e os valores (V_i) selecionados no esquema global:

$$CG \{O, T, R, At_i, V_i\} \mid i \geq 1$$

Esta consulta contém os mapeamentos para os esquemas das bases de dados locais. A consulta global será desmembrada em tantas consultas locais (CL) quantos forem os mapeamentos definidos no esquema global externo para o objeto geográfico (Og_j) selecionado. Este mapeamento é dado por:

$$CG \{O, T, R, At_i, V_i\} \Rightarrow CL \{E_j: Og_j, P\{At_i\}, V_i\} \mid j \geq 1$$

onde:

E_j = identificador dos esquemas onde se encontra o objeto geográfico procurado;

Og_j = denominação do objeto geográfico no esquema local;

$P\{At_i\}$ = conjunto de propriedades do objeto geográfico no esquema local, que representa a propriedade no esquema global externo; e

V_i = conjunto de valores selecionados para selecionar as instâncias no esquema local.

O resultado desta consulta é uma coleção de objetos gráficos e não gráficos heterogêneos que serão tratados pela arquitetura de modo a compor um resultado integrado a ser apresentado na interface com o usuário.

VI.6. Considerações sobre os resultados das consultas no ambiente MultiSIG

Em um ambiente MBD para SIG, os resultados de uma consulta podem apresentar conflitos entre valores, conforme foi visto na seção V.8. Além destes conflitos, pode ocorrer inconsistências devido a diversidade de representações das componentes gráficas e não gráfica, dos objetos geográficos nos esquemas locais, das diferenças de acurácia das bases de dados dos SIGs e, até mesmo, dos parâmetros cartográficos utilizados na representação gráfica dos esquemas.

Os conflitos entre os valores causam erros temáticos e inconsistências entre as representações, erros de acurácia posicional, tais como: superposição inadequada de um conjunto de dados gráficos ou representação geométrica diferente, para o mesmo fenômeno geográfico. Estes erros precisam ser resolvidos antes dos resultados serem apresentados ao usuário.

Todavia, na recuperação dos objetos geográficos é necessário considerar três aspectos de relevante importância, para a recuperação das informações armazenadas sobre um sistema de coordenadas, a saber: geometria, topologia e representação gráfica.

Desta forma, o **Sistema MultiSIG** aplica um conjunto de operações de transformação e generalização para integrar os resultados da consulta e realçar a legibilidade de sua representação gráfica, de forma a ajudar a compreender o fenômeno consultado. Para isto é realizado uma análise extencional dos dados resultantes das consultas, considerando apenas os esquemas utilizados na consulta global. Esta análise avalia os esquemas onde serão realizadas as subconsultas locais, quanto aos:

- parâmetros cartográficos envolvidos nos esquemas;
- a diferença de escala; e
- a diferença de estrutura, matricial e vetorial;

As diferenças entre os parâmetros cartográficos são tratadas por operações de transformação. A diferença entre as estruturas é tratada através da superposição das representações gráficas e posterior conversão para o formato GIF. A diferença de escala é tratada por operações de generalização espacial. Nesta generalização espacial os relacionamentos topológicos originais não devem ser alterados, entretanto, os relacionamentos métricos podem ser alterados, mas não devem causar violar o relacionamento topológico.

Assim, a *MultiSIG* para combinar objetos provenientes dos resultados das subconsultas locais propõe um tratador de resultados que tem a finalidade de:

- combinar a semântica dos objetos correspondentes, de acordo com as suas propriedades temáticas;
- combinar a topologia dos objetos, através da justaposição ou superposição de objetos; e
- combinação geométrica dos objetos, de acordo com as suas localizações;

Cabe ressaltar que a qualidade e integridade dos dados, no ambiente MBD, é mantida pelos fornecedores de dados, enquanto a utilização destes é responsabilidade dos usuários locais. Assim, apesar da *MultiSIG* utilizar este tratador de resultados haverá um relaxamento da acurácia dos dados, principalmente da posicional.

VI.7. Componentes da *MultiSIG*

A *MultiSIG* é definida em três níveis distintos e independentes que se comunicam, via um plano de mensagem orientado a objeto comum (Figura VI.1).

No primeiro nível se encontra a interface com o usuário. Ela tem a função de esconder os detalhes da distribuição dos dados, ou o esquema do modelo de dados locais dos usuários da federação. Neste nível, o usuário interage com a federação, através dos serviços oferecidos.

O segundo nível, denominado de nível de interoperabilidade, proporciona os mecanismos para a distribuição da consulta e comunicação entre os SIG locais e a federação. Nele estão os recursos que permitem o sistema validar e distribuir a consulta às bases de dados locais, bem como tratar os resultados das consultas locais. Assim, este nível utiliza de um servidor HTTP (*Hiper Text Transfer Protocol*), metaesquema, um processador de consultas e um tratador de resultados.

Neste nível, os dados solicitados na consulta são localizados no metaesquema. Usando a informação sobre a distribuição dos dados, a consulta é analisada e descomposta em sub-consultas, de acordo com distribuição dos dados nos SIGs locais, e traduzidas para as respectivas sintaxes das linguagens de consultas destes sistemas.

Estas subconsultas são então submetidas ao terceiro nível, nível local, onde se encontram os servidores de dados dos SIGs, participantes da federação, com seus esquemas locais. Neste nível, cada servidor de dados de SIG executa uma consulta local

e traduz a resposta (relações resultantes) para a sintaxe da linguagem do cliente, e as envia para o segundo nível.

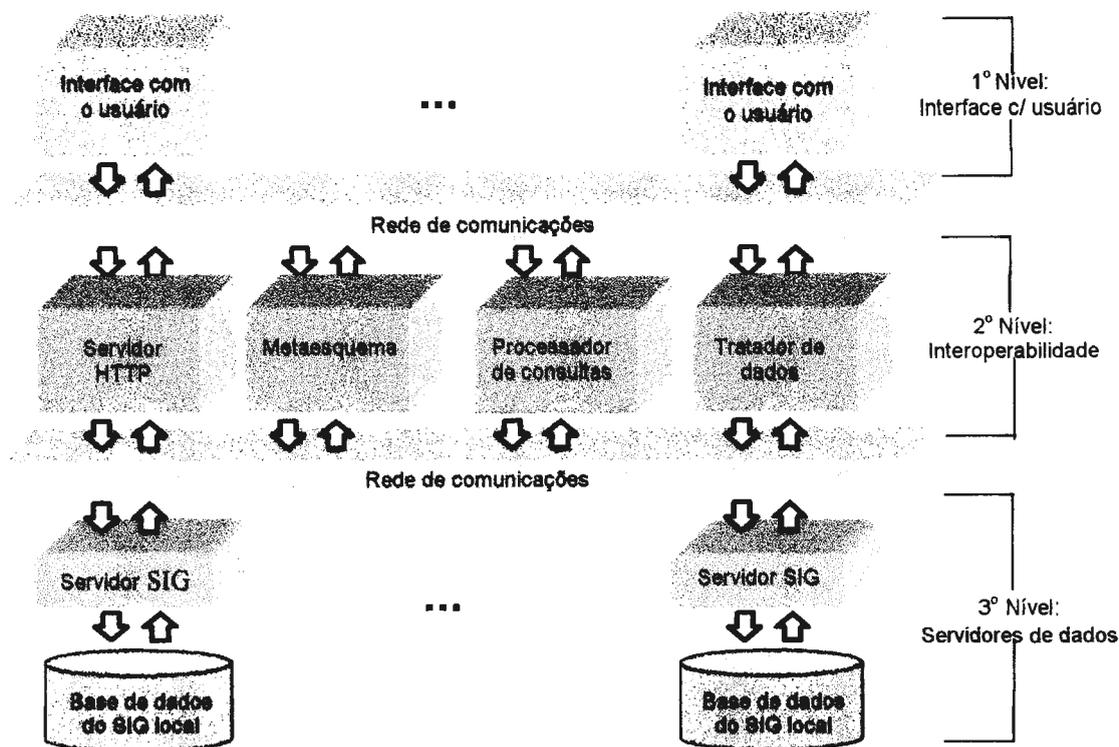


Figura VI.1- Componentes da arquitetura MultiSIG.

No segundo nível, os resultados da consulta são então combinados, de forma a produzir um resultado final a ser apresentado na interface com o usuário. Este resultado poderá ser apresentado na forma de mapa ou de tabela, isto é, a interface com o usuário deverá suportar representação gráfica e *browse* sobre as tuplas encontradas.

VI.7.1. Primeiro nível: interface com o usuário

A principal finalidade deste nível consiste em tornar disponível as funcionalidades propostas pelo Sistema MultiSIG, apresentadas na seção VI.4. Ele capacita os usuários da federação a interagirem com o sistema.

Este nível oferece: *i)* serviços para o administrador local cadastrar seu esquema de exportação na federação; *ii)* serviços para o administrador global definir o esquema global externo; *iii)* serviços para explorar o esquema global externo; *iv)* serviços de consulta; e *v)* serviços para efetuar a comunicação entre os administradores locais e o administrador da federação para estabelecer a cooperação na MultiSIG. Este nível também é utilizado pela arquitetura para apresentar os resultados da consulta.

A consulta será elaborada em uma linguagem gráfica, comum a todos os usuários da federação, que suporta a manipulação textual dos objetos. Assim, através da seleção de predicados temáticos relacionados aos objetos geográficos e/ou da seleção espacial de unidades territoriais, sobre a hierarquia de *Theme* ou de *Region*, respectivamente, é realizado uma consulta. Esta consulta é então submetida para o nível de interoperabilidade, distribuída e mapeada para as sintaxes das linguagens de consultas dos SIG locais.

Assim, a interface gráfica consiste em um conjunto de serviços proporcionados pela federação, a saber:

- de cadastro do esquema de exportação- oferece formulários para os administradores locais cadastrar/alterar os esquemas de exportação na federação, assim como para solicitar a remoção de um esquema da federação;
- de cadastro do esquema global externo - oferece formulários para o administrador da federação criar e alterar o esquema global externo;
- de visualização (*browser*) - oferece facilidades para o usuário explorar e compreender a semântica das classes das hierarquias de *Theme* e *Region*, disponível no esquema global. Este serviço futuramente poderá ser estendido para permitir explorar os esquemas cadastrados na federação, de modo a fornecer suporte para a consulta a um esquema específico;
- de consulta - oferece facilidades para os usuários da federação localizar os dados desejados, selecioná-los e realizar consultas sobre as hierarquias de *Theme* e *Region*. Este sistema, futuramente também poderá ser estendido para suportar consultas aos metadados de um esquema específico, de forma a permitir avaliar a proveniência e a qualidade dos dados;
- de cooperação - oferece um mecanismo que gerencia um fluxo de trabalho entre o administrador global e os administradores locais, de modo a estabelecerem comunicação no processo de integração;
- de aquisição de dados - oferece um conjunto de serviços para trazer os dados de um determinado esquema, ou o resultado de uma consulta aos dados, para um determinado formato ou padrão de dados. Este serviço deverá, ainda, ser especializado, de modo a possibilitar também a conversão de sistemas de representação cartográficos.
- de apresentação dos resultados da consulta - este serviço apresenta os resultados da consulta na forma de relatório ou mapa.

No primeiro protótipo, apresentado neste trabalho, serão especificados e desenvolvidos os serviços de cadastro, de visualização e de consulta, de forma a validar a metodologia.

No futuro, estes serviços poderão ser estendidos conforme as necessidades identificadas na federação. Um exemplo, é a necessidade do usuário em especificar uma legenda para melhor compreender a visualização do resultado da consulta. Este serviço poderá ser oferecido na consulta.

VI.7.2. Segundo nível: de Interoperabilidade

O nível de interoperabilidade tem um papel importante na arquitetura MultiSIG. É o responsável pela coordenação de todas as atividades solicitadas na interface pelo usuário. Ele utiliza técnicas de rede e da arquitetura cliente-servidor para interagir com o usuário e acessar os dados locais.

A principal finalidade deste nível, consiste em gerenciar a execução das requisições dos usuários a um/vários servidor(es) e a seguir aguardar o retorno dos resultados destas requisições para integrar os resultados em uma única representação gráfica e textual, a ser apresentado na interface com o usuário. Este nível também tem por finalidade executar outras tarefas de comunicação solicitadas pelos usuários a federação, na interface com o usuário. Desta forma, este nível é composto por serviços para:

- a) gerenciar o dialogo com o usuário, através da interface;
- b) realizar verificações, de forma a localizar os serviços e dados solicitados pelo usuário;
- c) fragmentar as requisições, de acordo com o serviço solicitado;
- d) no caso de consulta ou aquisição de dados, mapeá-las para as sintaxes das linguagens de consultas dos SIGs locais e submetê-las, através da rede, à camada servidor localizada nos SIG locais; e
- e) integrar os resultados das requisições e elaborar uma apresentação destes.

Para realizar estes serviços a arquitetura, implementada na rede global Internet, será composta por:

- ***Servidor HTTP***

O servidor HTTP, na rede Internet, é o dispositivo que oferece os serviços de suporte à distribuição de informação e serviços de comunicação. Na arquitetura proposta, este servidor é o responsável por executar as tarefas solicitadas pelos clientes e enviar as resposta de volta à interface com o usuário, fornecer serviços de visualização de páginas HTML (*Hyper Text Markup Language*), serviços de comunicação via *e-mail*, e serviços de ftp (*File Transfer Protocol*).

Este servidor estabelece comunicação através do protocolo HTTP, que emprega um conjunto de regras e formatos usados para comunicação entre processos, a fim de realizar uma tarefa. Este protocolo é composto da especificação de uma seqüência de mensagens, que deve ser trocada no processo de comunicação e da especificação do formato dos dados na mensagem.

- ***Metaesquema***

O metaesquema tem a finalidade de armazenar os esquemas de exportações, o esquema global externo e todas as informações resultantes do processo de cooperação para a integração dos esquemas.

Assim, o metaesquema contém a localização das bases de dados participantes da federação, a descrição dos esquemas de exportação, dos atributos dos esquemas de exportação, o esquema global externo e os mapeamentos entre os esquemas locais e global, bem como os mapeamentos da linguagem de consulta global para as linguagens de consulta dos SIGs locais. O metaesquema é, então, utilizado pela  MultiSIG e pela  MultiSIG.

No protótipo inicial, o metaesquema é formadas por um conjunto de classes construtoras, em Java, que armazenam o esquema global externo; e um sistema de arquivo que armazena o esquema de exportação cadastrado na federação proposta.

- ***Processador de consulta***

Este processador é o responsável pela coordenação das atividades relacionadas à execução de uma consulta global nos esquemas locais dos SIG.

A estratégia para recuperação dos objetos geográficos requisitados em uma consulta sobre o esquema global, no ambiente MultiSIG, consiste em transformar uma consulta, expressa na linguagem de consulta gráfica, em um plano de execução, que acesse corretamente todas as bases de dados locais, componentes da federação.

A consulta expressa no esquema global externo é decomposta em subconsultas para os esquemas locais, as quais são então traduzidas para a linguagem de consulta de cada SIG local e submetidas aos servidores de dados locais. O processo de tradução converte as propriedades dos objetos geográficos, no esquema global externo, em atributos correspondentes armazenados pela tabelas gerenciadas pelos esquemas locais.

Assim, o usuário utilizando os formulários propostos pela interface de consulta introduz os objetos selecionados e os predicados da consulta, os quais são transformados em uma *string* de consulta sob o esquema global, que por sua vez, é passada ao processador de consulta global para processá-la, segundo o modelo abaixo:

1. O processador de consulta global executa um método de validação da sintaxe da consulta que avalia:
 - Consistência geográfica - verifica se a *Region* solicitada tem consistência semântica para a realização da consulta;
 - Consistência semântica - verifica se o conteúdo temático da consulta é compatível com a região geográfica dos objetos solicitados; e
 - Consistência sintática - verifica se sintaxe ou os parâmetros de entrada estão corretos. Caso haja um erro, o controle é passado ao tratador de resultados e um erro é relatado ao usuário;
2. O processador de consulta executa um método que pesquisa, no metaesquema, a localização dos dados nos esquemas de exportação. Este método terá como resultado: os esquemas, as tabelas e as propriedades, onde estão localizados os dados solicitados na consulta global;
3. A consulta global, semanticamente correta, é passada para um método que fragmenta a consulta global em subconsultas, de acordo com a localização dos dados encontrados na etapa anterior;
4. Para cada subconsulta são encontrados os mapeamentos que transformam a subconsulta global em uma consulta expressa na linguagem da base de dado local, na qual os dados desejados se encontram;
5. A seguir estas subconsultas são transformadas por métodos tradutores apropriados para cada linguagem de consulta do SIG local; e
6. Esta subconsultas são transmitidas pela rede global e passadas aos servidores de dados locais dos SIGs, segundo a localização dos dados desejados.

Ao final da distribuição das subconsultas aos servidores locais são executados alguns processos para remover resultados intermediários.

No protótipo inicial, o processador de consulta não será implementado. A consulta é realizada através de métodos do esquema global externo que mapeam os elementos para os esquemas locais.

- ***Tratador de resultados***

Podem ocorrer falhas locais na execução das subconsultas, bem como os podem conter componentes gráficas, não gráficas ou ambos, com diferentes formatos, escalas, sistemas de referência e conflitos entre valores.

Desta forma, a *MultiSIG* propõe um tratador de resultados, o qual é responsável por trazer os dados para uma representação que combina/integra dados recuperados de fontes distintas e prepara uma apresentação para a visualização do resultado da consulta. O tratador de resultados realiza um pós-processamento que reformata os dados recuperados e gera os produtos a serem apresentados ao usuário.

O tratador de resultados utiliza os recursos de rede, um SIG acessado internamente, e um conversor de imagens. O SIG utiliza uma interface pública para requisitar e responder as operações de Geoprocessamento. Ao receber os resultados da consulta, o SIG tem a finalidade de integrar os resultados, de modo a compor uma representação gráfica a ser apresentada na interface com o usuário. O conversor de imagens converte a representação para o formato GIF, no caso de mapas. A seguir, o tratador de resultados gera uma HTML, contendo os resultados integrados, que serão apresentados na interface com o usuário. Assim, o tratador de resultados realiza o seguinte processamento:

1. Verifica o tipo de resultado das consultas:
 - Caso de erro - analisar o tipo de erros, o qual pode ser do sistema ou da consulta. Neste caso é elaborado uma HTML, descrevendo o tipo de erro e passada a interface com o usuário; e
 - Caso de resultados - analisar o tipo de resultado. Este pode ser composto por resultados gráficos e não gráficos ou os dois. Neste caso é passada a etapa 2;
2. Analisa os valores dos resultados das consultas locais, de modo a detectar conflitos entre os valores. No caso da existência de conflitos é executado um método para acessar o metaesquema e pesquisar a existência de mapeamentos determinados pela *MultiSIG*, que conformam os resultados das consultas. A seguir é aplicado a função de mapeamento e determinado a existência de conflitos entre os valores;

3. Novamente é verificado a existência de conflitos entre os valores os resultados da consulta, a saber:
 - No caso da existência de conflitos, os valores serão passados ao tratador de resultados que elabora uma HTML, a ser apresentada na interface com o usuário, com todos os valores e os seus respectivos esquemas;
 - No caso da não existência de conflitos e os resultados serem apenas textuais, os valores serão passados ao tratador de resultados que elabora uma HTML, a ser apresentada na interface com o usuário; e
 - No caso da não existência de conflitos entre os valores, se os dados conterem representações gráficas, os resultados são passados para um SIG, de modo a combiná-los em uma representação gráfica, e a seguir, passados ao tratador de resultados que elabora uma HTML, a ser apresentada na interface com o usuário;
4. Elaboração da HTML a ser apresentada na interface com o usuário. Esta elaboração requer a transformação das representações gráficas para o formato GIF.

Cabe, mais uma vez, ressaltar que independente das atividades acima para combinar valores dos dados, haverá o relaxamento de regras de integridade, tal como superposição incompleta de um conjunto de entidades mapeadas. Ou seja, poderá haver perda de acurácia, precisão e/ou significado, bem como a inconsistência entre os dados atualizados com diferente frequência e acurácia.

O tratador de resultados poderá no futuro também interagir com o metaesquema para recuperar funções de mapeamentos dos esquemas locais, tais como: transformações cartográficas, necessárias para tornar compatíveis as representações gráficas, e transformações semânticas, para associar os valores obtidos em um esquema com outros valores obtidos em outros esquemas

VI.7.3. Terceiro nível: Servidores de dados

O terceiro nível contém as plataformas onde estão as bases de dados dos SIGs que participam da federação. Estas plataformas poderão ser heterogêneas e executar sistemas operacionais diferentes.

Este nível possui os serviços para, ao receber uma subconsulta do processador de consulta global, acessar a base de dados local do SIG, resolver a consulta e

transformar o resultado da consulta para o tratador de resultados, de acordo com o seguinte processamento:

1. A subconsulta local é recebida e avaliada em relação a sintaxe da linguagem de consulta local;
2. A subconsulta local é executada;
3. A execução da subconsulta é avaliada, de forma a verificar se a consulta é executada corretamente ou se houve falha;
 - Se a consulta teve sucesso, então os resultados são passados ao tratador de resultados; e
 - Se houve falha é executado um método que solicita o motivo da falha e enviado ao tratador de resultados; e

No processamento local da subconsulta, alguns processos são executados para realizar algumas funções que ajudam a execução da consulta. Assim, ao final do processamento local, um método do servidor é executado para remover resultados intermediários e diretórios usados para a execução da consulta local.

VI.8. Abordagens da Implementação do protótipo MultiSIG

Este trabalho propõe o desenvolvimento de um protótipo, com o objetivo de ajudar a testar a metodologia e tornar o ambiente MultiSIG operacional, a partir dos componentes especificados na seção VI.7. Este protótipo, também deverá fornecer subsídios para o desenvolvimento de outros trabalhos no âmbito do WHIP2.

De modo geral, o ambiente de Geoprocessamento formados pelas comunidades de informação é caracterizado pela diversidade de plataforma e aplicativos de SIG. Desta forma, este trabalho propõe desenvolver a MultiSIG sobre a rede Internet.

Esta rede consiste em um conjunto de redes de computadores interligadas, pelo mundo inteiro, que tem em comum um conjunto de protocolos e serviços. Os usuários conectados na rede Internet podem usufruir de informação e comunicação de alcance mundial. A principal vantagem de utilizar esta rede para implementar a MultiSIG é que ela permite conectividade de redes de tecnologias distintas.

Esta conectividade é alcançada através do protocolo HTTP, que proporciona independência de plataforma e fornece comunicação para transmitir requisições/resultados entre o cliente e o servidor HTTP e vice-versa. Este protocolo é implementado sobre o conjunto de protocolos de comunicação, conhecido como TCP/IP (*Transmission Control Protocol/Internet Protocol*). O protocolo TCP/IP permite a

comunicação entre aplicações em computadores de redes distintas, sem a necessidade de conhecimento da topologia envolvida nesse processo. Além disto, este protocolo oferece flexibilidade de adaptação às tecnologias de redes existentes e futuras. Isto é possível porque o TCP/IP foi concebido de forma independente das tecnologias de rede.

Um outro aspecto importante na escolha da solução para implementar a *MultiSIG* foi a interface com o usuário, a qual pode ser desenvolvida utilizando um *browser* do *World Wide Web*. Este *browser* não só pode ser executado sobre várias plataformas, mas também promove uniformidade na aparência do ambiente *MultiSIG*, além de flexibilidade e elegância através dos recursos de hipertexto.

Para a implementação da interface é utilizado HTML e Java. A HTML é a linguagem para *Web* que cria hipertextos. A linguagem Java, é uma linguagem desenvolvida pela Sun Microsystems. Trata-se de uma linguagem orientada a objeto, adequada a implementação de computação distribuída sobre a rede Internet.

Segundo VCKOVSK (1996), Java permite modelar o dados geográficos como classes virtuais que são transferidas na rede, através de métodos que realçam a funcionalidade de cada sistema.

Sob o ponto de vista de programação, a linguagem Java constitui uma plataforma de desenvolvimento que pode ser utilizada para construir aplicações passíveis de distribuição, segura e eficiente, para toda a comunidade da Internet. A vantagem de usar esta linguagem é ela ser interpretada, podendo ser executada em uma rede heterogênea de equipamentos, ou seja, ela elimina o trabalho de portar aplicações para uma dezena de plataformas, de modo a satisfazer todos os usuários da Internet.

Para a *MultiSIG*, a linguagem Java fornece recursos para a implementação do ambiente MBD, com garantia de autonomia de execução e comunicação, uma vez que oferece recursos de implementar métodos de execução local, além de portabilidade e extensibilidade.

VI.9. Serviços do Sistema MultiSIG

Para alcançar as funcionalidades propostas pelo **Sistema MultiSIG**, a *MultiSIG* é especificada como um conjunto de serviços para realizar aplicações clientes e servidoras que implementam a comunicação e executam os serviços solicitados.

A especificação destes serviços é apresentada, nas seções a seguir, de acordo com os seus componentes, a saber:

VI.9.1. Serviços de suporte a interface com o usuário

A interface com o usuário tem a finalidade de apresentar as funcionalidades dos sistema aos seus usuários. Desta forma, ela é composta pelos seguintes serviços:

- ***Serviço de criação da federação***

Este serviço permite o administrador global dar o *download* no **Sistema MultiSIG** e selecionar uma área na rede onde a federação poderá ser instalada;

- ***Serviço de cadastro das comunidades de informação participantes da federação MultiSIG***

Este serviço permite o administrador global cadastrar as comunidades de informações que participaram da federação;

- ***Serviço de cadastro dos usuários***

Este serviço permite o administrador global cadastrar os usuários da federação de acordo com o seu papel;

- ***Serviço de visualização do esquema global externo***

Este serviço permite ao usuário explorar o esquema global externo, navegando nas hierarquias de *Theme* e *Region*. Este serviço permite conhecer estas hierarquias e as respectivas propriedades de suas classes;

- ***Serviços de visualização do metaesquema***

Este serviço permite ao usuário explorar um esquema de exportação específico, de modo a avaliar o seu conteúdo semântico e qualidade;

- ***Serviço de consulta ao esquema global externo***

Este serviço permite ao usuário selecionar os objetos, as propriedades e os valores no esquema global externo, além dos operadores disponíveis na interface para compor uma consulta de caráter espacial ou temática;

- ***Serviço de consulta a uma base de dados específica***

Este serviço permite ao usuário selecionar um esquema, no metaesquema, e os operadores disponíveis, na interface, para compor uma consulta;

- ***Serviço de cadastro do esquema de exportação***

Este serviço permite ao administrador local cadastrar, ou solicitar a alteração e remoção um esquema na federação. O cadastro consiste em descrever o esquema de exportação no SAIF, estendido de acordo com a *MultiSIG*. A alteração ou remoção consiste em um serviço de e-mail para o administrador da federação providenciar as medidas necessárias a operação solicitada;

- ***Serviço de cadastro do esquema global externo***

Este serviço permite ao administrador da federação cadastrar, alterar ou remover o esquema global externo;

- ***Serviço de apoio a cooperação***

Este é baseado no serviço de correio. Ele permite ao administrador local, trocar *e-mail* com os demais administradores locais, de modo a cooperarem na integração dos esquemas;

- ***Serviço para a seleção gráfica da apresentação dos resultados***

Este serviço permite ao usuário otimizar a ordem e a apresentação dos objetos geográficos, em mapas temáticos, através da seleção de uma legenda de cores e símbolos;

- ***Serviço de aquisição de dados***

Este serviço permite ao usuário solicitar dados em um esquema específico, em um determinado formato ou padrão de dados. Ele requer uma aplicação no servidor HTTP para converter os dados recuperados, em uma consulta, para um formato ou padrão de dados desejado;

- ***Serviço de login***

Apesar de não discutido até o momento, uma arquitetura MDB pressupõe critérios de segurança, que poderão ser implementados através deste serviço. Ele terá a finalidade de verificar a autorização do usuário na federação. Esta autorização será

fornecida pelo administrador da federação. Assim, cada usuário terá um *User id* e permissão admitida pelo administrador global, o qual será analisado pelo processador de consulta;

VI.9.2. Serviços do processador de consulta

Os serviços oferecidos pelo processador de consulta têm por objetivo fragmentar a consulta em subconsultas, aos esquemas locais, a saber:

- ***Serviço de verificação da consulta***

Este serviço tem por objetivo analisar a *string* de consulta, de modo a avaliar a consistência geográfica, semântica e sintática da consulta. Se nesta análise forem encontrados erros, a execução é desviada para o tratador de resultados;

- ***Serviços de localização dos dados no metaesquema***

Este serviço permite localizar no metaesquema os objetos geográficos e as propriedades de interesse solicitados na consulta global. O gerenciador acessa e navega entre os elementos de uma coleção através de um processo iterativo;

- ***Serviço de fragmentação da consulta***

Este serviço permite fragmentar a consulta sobre as bases de dados participantes da federação, de acordo com localização dos dados detectados no serviço anterior. Este serviço é na verdade uma fábrica de subconsultas;

- ***Serviço de mapeamento da consulta global para a subconsulta local***

Este esquema avalia a existência, no metaesquema, de mapeamentos dos elementos do esquema global, solicitado na consulta, para o esquema local. Caso haja mapeamentos, estes serão aplicados;

- ***Serviço de tradução para as linguagens de consultas locais***

Este serviço traduz a subconsulta para a linguagem de consulta usada pelo SIG local, estabelecido de acordo com a fragmentação. Este serviço consiste em pesquisar no metaesquema, para cada esquema local, o conjunto de operações para recuperar os dados localmente. Estas operações são predefinidas e armazenadas no metaesquema.

No metaesquema, para cada esquema, é armazenado um conjunto de operações com os seus qualificadores, e descrição das classes, nas tabelas locais, com suas respectivas propriedades;

- ***Serviço de distribuição das subconsultas aos SIGs locais***

Este serviço tem por objetivo submeter as subconsultas traduzidas aos SIGs locais;

- ***Serviço de limpeza do processador global***

Este serviço remove arquivos e diretórios criados no servidor global para processar a consulta global e distribuí-la;

VI.9.3. Serviços do tratador de resultados

Os resultados são consolidados através da criação de novos tipos de dados, que agregam os resultados em um alto nível de abstração. Desta forma, a MultiSIG oferece serviços que realizam uma análise extencional dos resultados da consulta de modo a integrá-los, a saber:

- ***Serviço de avaliação dos resultados da subconsulta***

Este serviço permite avaliar o resultado da subconsulta, de modo a verificar se houve erros no sistema, erros parciais na execução da consulta, ou conflitos de valores entre os resultados da consulta;

Quando ocorrem erros no sistema, a execução é passada ao serviço gerador de HTML para informar o tipo de erro, a saber:

- ✓ Quando houver erros parciais na execução da consulta, a parte com erro é enviada para o gerador de HTML informar o erro; e
- ✓ Quando houver conflitos entre valores, vai para o serviço de pesquisa no metaesquema;

- ***Serviço de pesquisa no metaesquema***

Este serviço pesquisa no metaesquema se existe funções de mapeamentos entre os valores e chama o serviço de transformação dos valores dos resultados da consulta;

- ***Serviço de transformação dos valores dos resultados da consulta***

Este serviço aplica as funções de mapeamentos entre os valores encontrados no metadesquema;

- ***Serviço gerador de HTML***

Este serviço permite criar uma apresentação em HTML para ser apresentada na interface, de acordo com os tipos de resultados, a saber:

- ✓ Se os resultados das subconsultas forem textuais, será elaborada uma simples HTML ,e enviada para a interface. Neste caso, ocorrem as mensagens de erro na consulta; e
- ✓ Se os resultados das subconsultas envolverem representações gráficas, o gerador de HTML executa o serviço de Geoprocessamento, para compor a apresentação;

- ***Serviço de geoprocessamento***

Este serviço utiliza um SIG para realizar um conjunto de operações para elaborar a representação gráfica, a ser apresentada ao usuário. Este serviço poderá executar as seguintes operações:

- Serviço de generalização espacial - caso haja uma diferença de escala é aplicada uma generalização espacial;
- Serviço de transformação de estrutura - caso haja diferença de estrutura, pode ser aplicado uma operação para conversão de estrutura *raster-vetor* e vice versa;
- Serviço de transformação cartográfica - caso haja diferença de sistemas geodésicos, poderá ser implementado uma função de transformação de *data*; e
- Serviço de transformação de sistema de projeção - caso haja diferença de sistema de projeção será aplicada uma função de transformação de sistema de projeção;

- ***Serviço de limpeza do tratador***

Este serviço remove arquivos e diretórios criados no servidor global para processar a integração dos resultados das subconsultas locais;

VI.9.4. Serviços do Servidor de dados local

O servidor de dados local recebe uma subconsulta do processador de consulta global e a executa no SIG local. Assim, esta camada oferece serviços para acessar a base de dados locais do SIG, resolver a consulta e enviar o seu resultado para o tratador de resultado. Este nível é então composto pelos seguintes serviços:

- ***Serviço de avaliação da subconsulta local***

Este serviço avalia se a *string* de subconsulta está correta, em relação a sintaxe da linguagem de consulta local;

- ***Serviço de execução da subconsulta local***

Este serviço executa seqüência de operações para realizar a subconsulta local;

- ***Serviço de avaliação da subconsulta***

Este serviço avalia a execução da consulta, transforma os resultados da consulta para a federação e os passa para o tratador de resultados, no nível de interoperabilidade. Este serviço remove, ainda, arquivos ou diretórios temporários criados para a execução da consulta local;

- ***Serviço de limpeza local***

Este serviço remove arquivos e diretórios criados no servidor local para processar a subconsulta local;

VI.10. Protótipo MultiSIG

O primeiro protótipo teve por objetivo construir todas as interfaces e realizar a consultas sobre o esquema global externo.

O **Sistema MultiSIG** se apresenta na rede Internet como uma página HTML onde são oferecidos os serviços de criação da federação, cadastro das comunidades de informação, cadastro dos usuários e acesso ao ambiente MultiSIG (Figura VI.2). Para cada um destes serviços há outras páginas HTML que os explicam e executam as operações solicitadas pelo usuário.

O acesso ao ambiente MultiSIG se faz pelo botão da *homepage* do **Sistema MultiSIG**. O **Sistema MultiSIG** é constituído por duas interfaces que oferecem os serviços de suporte ao usuário. A primeira (Figura VI.3), interface principal, oferece

acesso a interface de cadastro do esquema de exportação, serviços de cadastro do esquema global externo, cooperação, visualização e consulta sobre o esquema global externo (Figura VI.5 e VI.6). A segunda oferece serviços de cadastro do esquema de exportação(Figura VI.4).



Figura VI 2- Homepage do Sistema MultiSIG

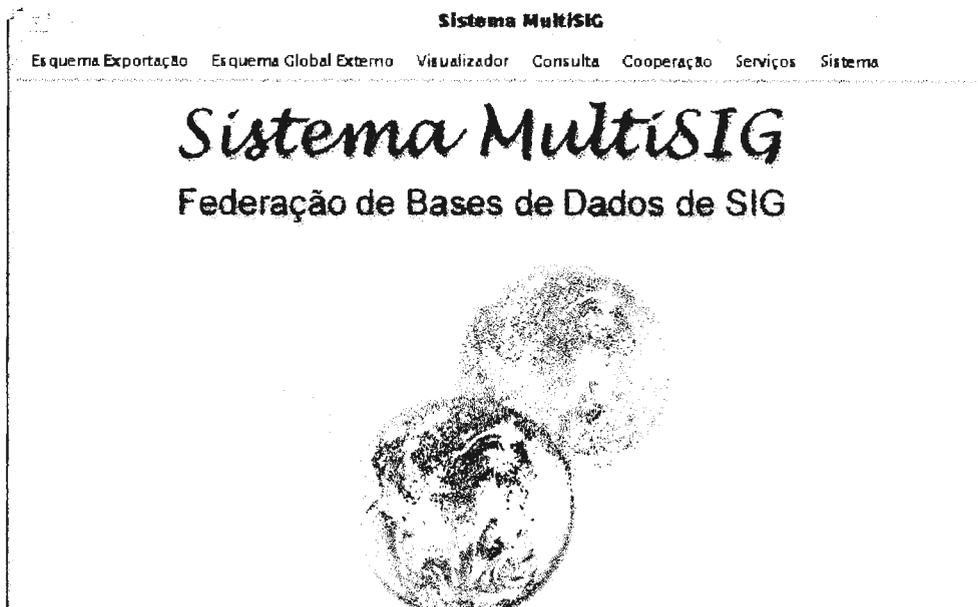


Figura VI.3 - Sistema MultiSIG - interface principal



Figura VI.4 - Sistema MultiSIG: Interface de cadastro do esquema de exportação

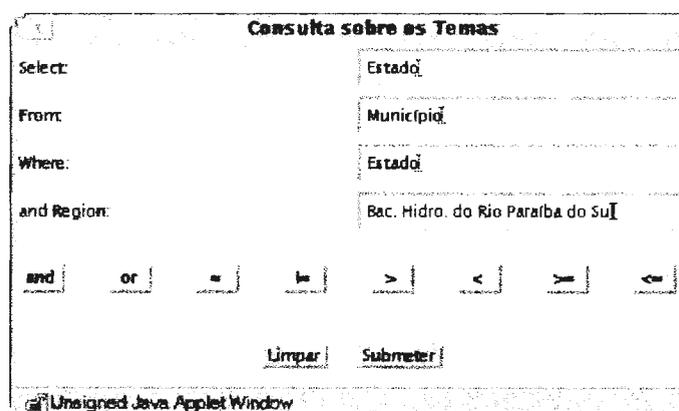


Figura VI.5- Interface de visualização do esquema global externo

Selecione os Temas para a consulta.

Temas	Atributos	Regiões
Município	Estado	Bac. Hidro. do Rio Paraíba do Sul
Bairro	Código	Bac. Hidro. do Rio Paraíba do Sul
Zonas Coleta	Código	Bac. Hidro. do Rio Paraíba do Sul
Resíduo Coletado	Hora Coleta	Bac. Hidro. do Rio Paraíba do Sul
Equipamento Coleta Resíduo	Equipamento	Bac. Hidro. do Rio Paraíba do Sul
Destino Finais Resíduo Sólido	Destino	Bac. Hidro. do Rio Paraíba do Sul
Estado	Código	Bac. Hidro. do Rio Paraíba do Sul
Destino Finais Resíduo Sólido	Código	Bac. Hidro. do Rio Paraíba do Sul
Atos Legais	Código	Bac. Hidro. do Rio Paraíba do Sul
Dados Hidrológicos	Hidrografia	Bac. Hidro. do Rio Paraíba do Sul

Unsigned Java Applet Window

Figura VI.6 - Interface de consulta ao esquema global externo

Capítulo VII - Conclusões e perspectivas futuras

VII.1. Resumo

Os SIGs oferecem ambientes monolíticos, onde as soluções tradicionais para o compartilhamento de dados são caracterizadas por troca de dados entre modelos de dados proprietários. Estas soluções são ineficientes devido à diversidade de estruturas e formatos existentes, além de elas não contemplarem as semânticas dos dados representadas por estas estruturas. Isso faz com que haja a necessidade de criar soluções que promovam a interoperabilidade entre estas bases de dados de SIGs.

O **Sistema MultiSIG** tem por objetivo tornar as bases de dados de SIGs, heterogêneas e distribuídas, em bases interoperáveis e passíveis ao desenvolvimento de novas aplicações. Este sistema promove o compartilhamento de dados, preservando o investimento na informação geográfica armazenada e expandindo as bases de dados das comunidades de informação. Ele representa uma evolução das soluções tradicionais, em Geoprocessamento, para acessar recursos distribuídos com processamento e dados heterogêneos. Essa evolução se caracteriza por um ambiente MBD (Múltiplas Bases de Dados) para comunidades de informação geográfica.

Devido às diferentes formas pelas quais os objetos geográficos podem ser descritos pelos SIGs, a comparação e correlação das informações geográficas de diferentes bases de dados geográficos é extremamente complexa. Assim, este trabalho proporciona, através do **Sistema MultiSIG**, uma federação de dados geográficos aberta, composta por bases de dados de SIGs, heterogêneas e distribuídas. O conceito de federação aberta está associado à participação de novas comunidades de informação ou à saída de comunidades já participantes na federação.

O **Sistema MultiSIG** demandou o desenvolvimento de uma metodologia e de uma arquitetura, denominadas, respectivamente, de **MultiSIG** e **MultiSIG**. A **MultiSIG** coordena as atividades necessárias para capturar a semântica dos esquemas, tratar as diferenças encontradas, resolvê-las e criar um esquema global. A

MultiSIG permite criar uma federação de dados geográficos entre comunidades de informação e oferece serviços para os seus usuários consultarem os dados nas bases dos SIGs componentes desta federação, sem precisar conhecer a localização dos dados, o esquema local e a linguagem de consulta espacial utilizada pelos sistemas locais.

Para criar uma federação aberta, a MultiSIG estabelece a cooperação no processo de integração e a MultiSIG oferece serviços de cadastro, de alteração e remoção dos esquemas de exportação na federação, bem como serviços de apoio à cooperação entre os administradores locais e o administrador global. Os serviços de apoio à cooperação auxiliam os administradores a encontrar um consenso na incorporação dos esquemas de exportação no esquema global.

Desta forma, neste trabalho foram apresentados os conceitos importantes utilizados na construção de uma base de dados geográficos (Capítulo II); foi realizada uma revisão das soluções para permitir a interoperabilidade entre múltiplas bases de dados convencionais, no âmbito da tecnologia de Sistemas de Banco de Dados (Capítulo III); foram feitas considerações sobre as características de um ambiente MBD para SIG, apresentadas algumas soluções encontradas na literatura e proposta a solução MultiSIG (Capítulo IV). Esta solução é composta pela metodologia MultiSIG (Capítulo V) e pela arquitetura MultiSIG (Capítulo VI). A MultiSIG foi ainda aplicada a quatro esquemas apresentados no Apêndice A de modo a validá-la. O seu desenvolvimento é apresentado no Apêndice B.

Dentre os principais resultados deste trabalho, cabe destacar:

- o modelo proposto para o esquema global externo - este é um modelo simples, que fornece uma única percepção do significados dos dados. Ele auxilia o usuário a encontrar os temas e as regiões geográficas de interesse, através das hierarquias de *Theme* e *Region*, respectivamente. Estas hierarquias armazenam dados virtuais na federação;
- a MultiSIG para integração dos esquemas - esta metodologia estabelece uma estratégia para normalizar os esquemas, identificar os relacionamentos entre eles, estabelecer o isomorfismo e a concordância no conjunto compartilhado de termos e definições utilizadas em várias áreas de aplicação, em um esquema global.
- a especificação dos serviços da MultiSIG - esta especificação contempla os serviços, necessários a uma federação aberta, para realizar consultas sobre um

ambiente MBD, considerando as necessidades de apresentação dos dados na forma de tabelas e mapas.

VII.2. Análise das contribuições

O **Sistema MultiSIG**, como solução para compartilhar dados geográficos, expande a base de dados e aplicações das comunidades de informação, tornando possível desenvolver novos tipos de análises, não previstas pelas questões manifestas e latentes, quando da construção das bases de dados dos SIG locais.

O **Sistema MultiSIG** propõe uma solução com quatro camadas de modelos, apresentando como contribuição inovadora o esquema global externo. Este esquema resulta da  MultiSIG e é gerenciado pelo administrador da federação. Este esquema global é desenvolvido sobre um modelo simples, orientado a objeto, que preserva a visão e torna disponível os temas e regiões geográficas. Dentre as vantagens desta solução cabe destacar:

- A facilidade dos usuários do **Sistema MultiSIG** em compreender as informações disponíveis na federação. Os usuários de Geoprocessamento estão acostumados a pensar sobre planos de informações. Um esquema global que não considera este aspecto, levará o usuário da federação a encontrar dificuldades em localizar os dados desejados ou a realizar consultas mais complexas. Esta desconsideração é um problema encontrado nas soluções existentes;
- A transparência na realização das consultas. Os usuários da federação acessam os dados distribuídos como se o usuário estivesse sobre um único sistema;
- A modularidade na descrição dos objetos geográficos, uma vez que, estes são tratados como classes, de acordo com a sua semântica e localização geográfica; e
- A preservação dos esquemas das bases de dados dos SIGs, visto que o modelo de dados do esquema global externo encapsula a resolução dos conflitos e os mapeamentos entre ele e os esquemas locais.

Entretanto, o desenvolvimento do **Sistema MultiSIG** é um trabalho árduo e complexo, devido às diferentes formas pelas quais os dados geográficos podem ser encontrados. No presente trabalho, foi desenvolvida, na íntegra, a  MultiSIG e parte da  MultiSIG.

A *MultiSIG* especifica as atividades necessárias à integração, tratando as diferenças entre modelos de dados de SIG e as diferenças entre as modelagens dos esquemas. A primeira atividade trata da conversão dos esquemas para a extensão proposta para o SAIF, gerando o esquema de exportação. A segunda atividade, cuida das atividades de análise dos esquemas de exportação, propondo soluções para conflitos entre os dados geográficos e integração no esquema global. A aplicação desta última atividade ao estudo de caso contribui com o levantamento dos conflitos encontrados entre os esquemas das bases de dados.

A adoção do SAIF, se mostrou adequada como MDC à primeira atividade para a descrição do esquema de exportação, pois ele possui um metamodelo, que contempla as características espaço temporais dos objetos geográficos, independente do SIG, e possui notações para definir classes e instanciá-las. A notação para definir as classes oferece, ainda, uma cláusula *default*, que permite vincular valores às classes e às suas subclasses. Isto facilita a descrição do conteúdo do esquema, uma vez que, em geral, os planos de informação de uma base de dados de um SIG possuem as mesmas abstrações cartográficas.

O metamodelo do padrão SAIF precisou ser estendido com propriedades específicas, que descrevem a região geográfica e a área de aplicação. Foram, também, refinadas as regras de criação de esquemas no SAIF, uma vez que as originais não representariam adequadamente os esquemas de exportação; pois aumentariam o número de classes necessárias para descrever o esquema de exportação.

A *MultiSIG* propõe soluções para os conflitos, enfatizando a análise dos esquemas de forma a evitar comparar objetos geográficos inconsistentes. Esta análise observa o contexto espacial, o contexto da aplicação e o contexto semântico tratados pelas bases de dados. Na análise do contexto espacial são determinados os esquemas que serão integrados e estabelecida a hierarquia de Regiões a serem disponibilizadas pela federação. Na análise do contexto da aplicação é estabelecida a sequência de pares de esquemas a serem integrados de acordo com a proximidade de áreas de aplicação. E finalmente, na análise semântica são definidas as atividades que estabelecem uma acurada estrutura de correlação entre elementos dos esquemas. Esta última resulta no diagnóstico dos conflitos entre os esquemas que são resolvidos na etapa de conformação dos esquemas, através de técnicas da orientação a objeto aplicadas para compatibilizá-los no esquema global interno.

A necessidades destes três contextos ficou evidente na aplicação da metodologia (Apêndice B). A utilização conjunta destes três conceitos de contextos em uma metodologia para a integração de esquemas de dados geográfico é inovadora. Os trabalhos encontrados na literatura, apresentados na seção IV3.2, em geral, se preocupam apenas com o contexto semântico, exceto YASER (1997) que considera o contexto das áreas de aplicação.

Outra inovação foi a aplicação de um critério de pesos para o grau de proximidade semântica encontrado entre os esquemas. Embora a aplicação deste critério já tenha sido usado na integração de Bases de Dados Convencionais (SOUZA, 1986), ele é inovador em bases de dados geográficos. Este critério de pesos foi estabelecido heurísticamente, para o grau de proximidade semântica parcial encontrado a cada etapa da *MultiSIG*. Este critério foi testado na aplicação da *MultiSIG*, apresentada no Apêndice B, em esquemas de áreas de aplicação e espaço geográficos coincidentes e diferentes. Este critério se mostrou adequado, quando o par de classes analisados apresenta a mesma granularidade de informação. Isto decorre do fato da *MultiSIG* tratar pares de classes com granularidades de informação diferentes como classes diferentes. Cabe ressaltar que os conflitos são diversos e que sempre poderá existir um caso inédito ao qual o critério não se adapta perfeitamente. Assim, se faz necessário sempre reavaliá-lo.

A *MultiSIG* se destaca das propostas encontradas na literatura devido à sua amplitude, detalhamento e conseqüente aplicabilidade. Ela propõe atividades necessárias à integração que contemplam os diferentes aspectos envolvidos na modelagem e construção de uma base de dados geográficos. Assim, embora a *MultiSIG* tenha sido desenvolvida para atender o **Sistema MultiSIG**, ela pode ser adaptada a outras arquiteturas que contemplam múltiplas bases de dados ou a interoperabilidade entre SIGs. Isto é, arquiteturas que empregam modelos de dados diferentes do SAIF estendido e do esquema global externo proposto.

A *MultiSIG* foi proposta sobre a rede Internet, tendo em vista o objetivo de construir uma federação aberta. Esta rede proporciona à arquitetura proposta uniformidade de comunicação, através do protocolo HTTP, e uniformidade de interface entre os usuários da federação.

Esta arquitetura é composta por um servidor HTTP, um metadesenho, um processador de consultas e um tratador de resultados. O servidor HTTP é responsável pela comunicação na rede. O metadesenho armazena os esquemas de exportação, o esquema global externo e todas as informações resultantes do processo de cooperação para a integração dos esquemas. O processador de consultas é responsável pelo processamento e distribuição das consultas. O tratador de resultados é responsável por integrar os resultados da consulta em mapas e tabelas, a serem apresentadas ao usuário.

Este trabalho especifica os serviços para cada componente e implementa um protótipo em Java, com alguns componentes (interface com o usuário, servidor de dados, tratador de resultados) e funcionalidades parciais. Este protótipo objetiva avaliar a funcionalidade e adequação do **Sistema MultiSIG**. Assim, o protótipo desenvolvido permite cadastrar bases de dados na federação, descrevendo os esquemas de exportação no SAIF estendido pela MultiSIG; formular consultas às bases de objetos geográficos componentes da federação; integrar o resultado da consulta e apresentar os resultados, permitindo a visualização gráfica destes.

A construção do protótipo MultiSIG ajudou a validar as atividades da metodologia e os serviços da arquitetura. Este protótipo evidencia os consideráveis benefícios no compartilhamento de dados entre comunidades de informação. Dentre os principais benefícios destacam-se:

- i)* mais eficácia na aquisição e troca de dados geográficos;
- ii)* meios para adquirir informações geográficas, de modo rápido e eficaz, e a um custo relativamente baixo; e
- iii)* redução do custo e do tempo do processo de aquisição de dados geográficos, e por conseguinte, redução do custo e do tempo de implantação de novos SIGs.

As características de programação da linguagem Java, também se mostraram adequadas ao desenvolvimento de uma arquitetura MBD. Ela é uma linguagem de programação orientada a objeto, que permite organizar os códigos dos programas do **Sistema MultiSIG**, clientes e servidores, como unidades lógicas, bem como aproveitar os recursos de encapsulamento, herança e polimorfismo deste paradigma. Além disto, ela possui recursos para a criação de interfaces mais amigáveis. Entretanto suas características principais são a portabilidade e o suporte para a Internet.

Concluindo, o **Sistema MultiSIG**, comparado às propostas analisadas na seção IV.3.3, é um sistema mais abrangente, que contempla tanto a interoperabilidade

semântica quanto a física. A primeira, através da *MultiSIG* que utiliza dois modelos de representação: o SAIF estendido e o modelo de dados proposto para o esquema global externo. A segunda, através da arquitetura, do conjunto de serviços especificados e da linguagem de consulta global adotada. Enquanto as propostas existentes se concentram em um grupo específico de aspectos de interoperabilidade o **Sistema MultiSIG** apoia o ciclo das diversas fases envolvidas no processo de integração. O experimento de validação da integração evidenciou a necessidade e vantagens do ferramental de apoio da *MultiSIG*.

VII.2. Perspectivas futuras

Este primeiro protótipo foi desenvolvido com o objetivo de avaliar a metodologia e possibilitar a realização de consultas sobre o modelo proposto para o esquema global externo. Entretanto, os demais serviços, apresentados no capítulo VI, deverão ser desenvolvidos e estendidos de forma possibilitar:

- a aquisição e o intercâmbio de informações através dos serviços de conversão de formato de dados e de transmissão de dados via *ftp*. Isto introduz novas perspectivas para o **Sistema MultiSIG**, como uma fonte de dados para novos SIGs;
- visualização e consulta a um esquema de exportação específico. Isto permitirá os usuários da federação avaliarem a qualidade dos dados locais;
- gerenciamento da apresentação dos resultados da consulta. Em Geoprocessamento, a apresentação dos dados, de acordo com as necessidades de comunicação da informação, é fundamental para a adequada compreensão dos resultados de uma consulta.
- análises espaciais. Estender o **Sistema MultiSIG** com serviços que oferecem capacidades de análises ou que compartilhem os recursos de análises dos SIG locais introduzirá também uma nova funcionalidade ao ambiente MBD.

Cabe ressaltar que o **Sistema MultiSIG** é a primeira ferramenta do Projeto WHIP (**W**orkbench de **H**eterogeneidade, **I**nteroperabilidade e **P**aralelismo). A esta poderão ser incorporadas novas ferramentas, tendo em vista suportar o estudo e desenvolvimento de novas tecnologias para SIG. Macroscopicamente poderão se desenvolver novos trabalhos nas seguintes direções:

- A *MultiSIG* poderá se utilizar da **lógica fuzzy** (BRAGA, 1998) no processo de análise;
- Aliar aos conceitos de contexto espacial, contexto da aplicação e contexto semântico técnicas de **inteligência artificial** (SULLAIMAN & SOUZA, 1998a e 1998b), visando incorporar conhecimento e facilitar a automação do processo de integração, bem como permitir a realização de *data mining* em bases de dados de SIGs;
- Incorporar ferramentas de **controle de fluxo e cooperação** (PALMA, ANICETO & SOUZA, 1997a 1997b) para promover maior sinergia entre os administradores no processo de integração; e
- Integrar o **Sistema MultiSIG ao gerenciador de objetos GOA++** (MAURO & MATTOSO, 1998) para promover a persistência para os objetos do esquema global interno e externo, de forma a suprir a falta de persistência da linguagem Java.

Uma questão que abre um novo campo de pesquisa é quanto a propagação dos erros temáticos e da acurácia posicional na integração de resultados em um ambiente MBD para SIGs.

Referências Bibliográficas

Foram utilizadas dois tipos de bibliografias para este trabalho que serão apresentadas nas seções a seguir, a saber:

- publicações tradicionais como livros, revistas, Anais, relatórios técnicos, e
- publicações disponíveis somente *on line*, via Internet.

Nestas últimas, para diferenciá-las no texto, os sobrenomes dos autores são seguidas do termo html sobrescrito.

7.1. Referências tradicionais

AALDERS, H., 1996, "Quality metrics for GIS". In: *Proc. of XVII International Symposium on Spatial Data Handling: Advances in GIS Research II*, Ed. M. J. Kraak & M. Molenaar, pp. 277-286.

ABEL, D. J., TAYLOR, K. & KUO, D., 1997, "Integrating modeling Systems for environmental management information Systems", *SIGMOD Record*, v. 26, n. 1 (Mar), pp. 5-10.

ABEL, D.J, 1997, "Spatial Internet Marketplaces: a grand challenge?" In: *Lecture Notes in Computer Science: Advances in Spatial databases*, Ed. Michael Scholl & Agnes Voisard, n. 1262, pp. 3-8, Berlin: Alemanha, Jul.

AGUIAR, C. D. & MEDEIROS & C. B., 1996, "Construção de um modelo básico unificado a partir de sistemas stand-alone". In: *Anais do GIS 96*, pp. 503-515, Curitiba, Mai.

AGUIAR, C. D., 1995, *Integração de Banco de Dados heterogêneos em aplicações de planejamento urbano*. Dissertação de Mestrado, Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, Campinas, SP, Brasil.

AHMED, R, SMEDT, P. D., DU, W., KENT, W., KETABCHI, M. A., LITWIN, W., RAFII A., SHAN, M., 1995, The Pegasus heterogeneous Multidatabase System, *IEEE Computer*, v. 24, n. 12 (Dez), pp. 19-27.

- ARC/INFO, 1991, User's Guide, ARC/INFO Data model, Concepts & Key Terms, ESRI.
- ARMSTRONG, R. & CARR, J., 1994, "The SAIF Toolkit: Opening the data roads between GIS". In: *Symposium of GIS 94*, pp. 333-340, Vancouver, Fev.
- ARONOFF, S., 1989, *Geographic Information Systems: A management perspective*. 1a. ed. Ottawa, WDL Publications.
- BAKKER, M. P. R., 1965, *Cartografia: noções básicas*. DH 21-1.
- BATINI, C., LENZERINI, M. & NAVATHE, S. B., 1986, "A comparative analysis of methodologies for database schema integration", *ACM Computing Survey*, v. 18, n. 4 (Dez), pp. 322-364.
- BERNHARDESEN, T., 1992, *Geographic Information Systems*. 1a. ed. Norway, VIAK IT and Norwegian Mapping Authority.
- BERTINO, E., NEGRI, M. PELAGATTI, G. & SBATTELLA, L., 1994, Applications of object oriented technology to the integration of heterogeneous database systems. In: *Distributed and Parallel Databases*, v. 2, pp. 343-370.
- BLOTT, S. & VCKOVSKI, A., 1995, "Accessing geographical metafiles through a database storage System" In: *Lecture Notes in Computer Science: Advances in Spatial Databases*. Ed. Max Engenhofer & John Herring, pp. 115-126., Portland, Jul.
- BONJOUR, M. & FALQUET, G., 1994, "Concept bases: a support to information system integration". In: *Lecture Notes in Computer Science: Advances Information Systems*. Ed. Tony Wassenan, Gerard Wijers & Spack Brinkemper, pp. 242-245., Utrecht: Holanda.
- BORGES, K. A. V. & FONSECA, F. T., 1996, "Modelagem de dados geográficos em discussão. In: *Anais do GIS 96*, pp. 524-533, Curitiba, Mai.
- BOUGUETTAYA, A., 1995, Using coalitions and services for implementing sharing in large network of databases. In: *Proc. of Second International Symposium on autonomous Decentralised Systems*, , pp. 176-184, Phoenix: Arizona.
- BRAGA, A. L., 1998, *Ferramentas de manipulação nebulosa de dados com aplicação em Sistema de Informação Geográfica*. Dissertação de Mestrado da COPPE/UFRJ, Rio de Janeiro, Rio de Janeiro, Brasil.

- BREITBART, Y. & REYES, T., 1995, "Overview of the ADDS systems". In: *Modern Database Systems: the object model, interoperability and beyond*, Won Kim Editor, ACM Press, New York, pp. 682-701.
- BREITBART, Y., 1990, "Multidatabase Interoperability", *SIGMOD Record*, v. 19, n. 3 (Set), pp. 53-60.
- BREITBART, Y., MOLINA, H. G. & SILBERSCHATZ, A., 1995, "Transaction management in Multidatabase Systems". In: *Modern Database Systems: the object model, interoperability and beyond*, Won Kim Editor, ACM Press, New York, pp. 573-591.
- BREITBART, Y., OLSON, P. L. & THOMPSON, G. R., 1986, "Database integration in a distributed database system", In: *Proc. of International Conference on Data Engineering*, pp. 301-310, California: USA.
- BRIGHT, M. W., HURSON, A. R. & PAKRAD, S., 1994, "Automated resolution of semantic heterogeneity in Multidatabases", *ACM Transactions on Database Systems*, v. 19, n. 2 (Jun), pp. 212-253.
- BRODIE & STONEBRAKER, M., 1995, *Migrating legacy systems: gateways, interfaces and the incremental approach*, 11 ed., Morgan Kaufmanns Publishers Inc., California: USA.
- BUKHERS, O & ELMAGARMID, A.; 1996; Object Oriented Multidatabase Systems, 1a. ed. New Jersey, Prentice Hall,
- BUNEMAN, P, RASCID, L. & ULLMAN J., 1997, "Mediator language- a proposal for standard", *SIGMOD Record*, v. 26, n. 1 (Mar), pp. 39-46.
- CÂMARA, G. & MEDEIROS, J. S., 1996, "Geoprocessamento para projetos ambientais", Relatório INPE, São José dos Campos.
- CASSETTARI, S, 1993, *Introduction to integrated Geo information management*, 1a. ed, London, Capman & Hall.
- CASTELLANOS, M, KUDRASS, T., SALTOR, F. & GARCIA-SOLACO, M., 1994, "Interdatabase existence dependencies: a metaclass approach". In: *Proc. of the Third International Conference on Parallel and Distributed Information Systems*, IEEE Computer Society Press, Austin, pp. 213-216.
- CATTELL, R. G., 1994, "ODMG-93: A standard for object-oriented DBMS". In: *Proc. of the ACM SIGMOD*, Ed. Richards Snodgrass & M. Winslett, v. 23, n. 2 (Jun), pp. 480.

- CHATTERJEE, A. & SEGEV, A., 1991, "Data Manipulation in Heterogeneous Databases", *SIGMOD Record*, v. 20, n. 4 (Dez), pp. 64-68.
- CHEN, A. L. P., 1990, "Database Research at NTHU and ITRI", *SIGMOD Record*, v. 23, n. 3 (Set), pp. 54-59.
- CHOMICKI, J & REVESZ, P., 1997, "Constraint-based interoperability of spatiotemporal databases", In: *Lecture Notes in Computer Science: Advances in Spatial databases*, Ed Michael Scholl & Agnes Voisard, n. 1262, pp. 142-161, Berlin, Jul.
- CHORAFAS, N. D. & STEINMANN, H., 1993, *Solutions for network databases: how to move from heterogeneous structures to federated concepts*, 1a.ed. London, Academic Press.
- CLEMENT, G., LAROUCHE, C., GOUIN, D., MORIN, P. & KUCERA, H., 1997, "OGDI: Toward Interoperability among Geospatial Databases", *SIGMOD Record*, v. 26, n. 3 (Set), pp. 18-23.
- COCKCROFT, S.; 1997, "A taxonomy of data integrity constraints", *Geoinformatica*; Kluwer Academic Publishers, n. 1: 4, Boston, pp. 327 -343.
- COLLET, C., HUHNS, M. N., SHEN, W. M., 1991, "Resource integration using a large knowledge base in Carnot", *IEEE Computer*, v. 24, n.12 (Dez), pp. 55-62.
- CORNELIO, A. & THOMAS, G., 1995, *Geoscope Requirements*, Bellcore International Document.
- DBOUK, M., KVEDARAUSKAS, D. & BOURSIER, P., 1996, "A multi-representation and multi-scale data model for hypermedia GIS applications". In: *Proc. of Geoinformatics'96*, pp. 41-49, Florida.
- DEACON, A., SCHEK, H. J. & WEIKIM, G., 1994, "Semantics-based multilevel transaction management in Federated Systems". In: *Proc. of the X International Conference on Data Engineering*, pp. 452-461, Houston, Fev.
- DEVOGELE, T., TREVISAN, J. & RAYNAL, L., 1996, "Designing and producing a multi-scale database from mono-scale databases". In: *Proc. of Geoinformatics'96*, pp. 14-24, Florida.
- DOGAC, A., DENGI, C., KILLIC, E. *et al.*, 1995, *METU Interoperable Database System*, Technical Report 6-1, Software R&D Center, June.

- DREW, P., KING, R., MCLEOD, D., RUSINKIEWICZ, M. & SILBERSCHATZ, A., 1993, "Report of the Workshop on Semantic Heterogeneity and Interoperation in Multidatabase Systems", *SIGMOD Record*, v. 22, n. 3(Set), pp. 47-56.
- DUPONT, Y. & SPACCAPIETRA, 1996, "Schema integration engineering in cooperative database systems". In: *Proc. of the International Conference Parallel and Distributed Computing Systems*, v. 2, pp. 759-765, Dijon: França, Set.
- ELIASSEN, F. & KARLSEN, R., 1991, "Interoperability and object identity", *SIGMOD Record*, v. 20, n. 4 (Dez), pp. 25-29.
- ELMAGARMID, A. K. & PU, C., 1990, "Introduction to the special issues on heterogeneous databases", *ACM Computing Survey*, v. 22, n. 3 (Set), pp. 175-178.
- ELMAGARMID, A., MANOLA, F., ROSENTHAL, A., MOLINA, H. G., MCLEOD, TEMPLETON, 1990, "Report on the workshop on heterogeneous database systems held at North-western University", *SIGMOD Record*, v. 19, n. 4, pp. 23-31.
- ELSMASRI, R. & NAVATHE, S. B., 1994, "Fundamentals of Database Systems", 2a. ed., California, Benjamin Cummings Publishing Company.
- FELLERS, J. W., CLIFTON, A. & HANDLEY, H., 1995, "Using the Internet to provide support for distributed interactions". In: *Proc. of the Annual Hawaii International Conference on System Sciences*, pp. 52-60.
- FORST, A., KÜHN, E. & BUKHRES, O., 1995, "General Purpose work flow languages", *Distributed and Parallel Databases*, v. 3, n. 2, pp. 187-218.
- FRANKHAUSER, P, MOTZ, R. & HUCK, G., 1995, "SIM Schema integration methodology. IRO/SPEC/D4-4.1-V1.0/PF950210.
- FRIESEN, P. & SONDEHEIM, M., 1994, "Mixing data from diverse sources". In: *GIS'94 Symposium*, pp. 151-156, Vancouver, Fev.
- FROMENT, E. D., 1994, "GIS and data integration: a user's perspective". In: *GIS'94 Symposium*, pp. 157-161, Vancouver, Fev.
- GARCIA-MOLINA, H. & HSU, M., 1995, *Distributed Databases. In: Modern Database Systems the object model, interoperability and beyond*, Won Kim Editor New York, ACM Press, pp. 621-649.
- GARCIA-MOLINA, H., PAPAKONSTANTINOY, Y., QUASS, D., RAJARAMAN, A., SAGIV, Y., ULLMAN, J., WIDON, J., 1995, "The TSIMMIS approach to mediation: data models and languages", In: *Next generation Information Technologies and Systems*, pp. 27-28, Naharia: Israel, Jun.

- GARDELS, K., 1997, "Open GIS and on-line environmental libraries", *SIGMOD Record*, v. 26, n. 1 (Mar), pp. 32-38.
- GARRAFA, I. M.; 1998 , *Análise da adequação de uma hierarquia de classes básicas para modelagem conceitual de SIG, através de um estudo de caso*. Dissertação de Mestrado, Instituto de Informática , UFRGS, Porto Alegre, RS, Brasil.
- GOH, C. H., MADNICK, S. E. & SIEGEL, M. D., 1994, "Context interchange: overcoming the challenges of large-scale interoperable Database Systems in a dynamic environment". In: *Proc. of the III International Conference on Information and knowledge Management*, pp. 1-10, Maryland.
- GOODMAN, J. N., 1994, "Alberta Land related Information System: a federated databases case study". In: *Proc. of Annual Conference Urban and Regional Information System Association*, pp. 421-431, Milwaukee.
- GOTTARD, W., LOCKEMANN, P. C & NEUFEL, A., 1992, "System guided view integration for object-oriented Databases", *IEEE Transaction on Knowledge and Data Engineering*, v.4, n. 1 (Fev), pp. 1-21.
- GÜNTHER, O., 1997, "Environmental Information Systems", *SIGMOD Record*, v. 26, n. 1 (Mar), pp. 3.
- HAMMER, J., 1994, *Resolving semantic Heterogeneity in a federation of autonomous, heterogeneous Database Systems*. Ph.D. Dissertation in Computer Science, University of Southern California, California.
- HEILER, S, 1995, "Semantic interoperability", *ACM Computing Surveys*, v. 27, n. 2 (Jun), pp. 271-273.
- HEIMBGNER, D. & MCLEOD, D., 1985, "A federated architecture for information management", *ACM transaction on office Information Systems*, v. 3, n. 3 (Jul), pp. 253-278.
- HERMELY, A. S., 1992, "Um estudo sobre linguagens de consultas espaciais para GIS", Relatório Técnico, Centro Científico RIO-IBM.
- HURSON, A. R. & BRIGHT, M. W., 1996, "Object-oriented Multidatabase Systems: a solutions for advanced applications", In: *Object-oriented Multidatabase Systems*, Bukhres, O. A. & Elmagarmid Eds., A. K., Prentice Hall, pp. 1- 36.
- HWANG, S. Y., SRIVASTAV, J. & LI, J., 1994, "Transaction recovery in Federated Autonomous Databases". In: *Distributed and Parallel Databases*, Kluwer Academic Publishers, Boston, v. 3, n. 2, pp. 151-182.

- IGRAS, E., 1994, "A framework for query processing in a federated database system: a case study". In: *Proc. of Annual Conference Urban and Regional Information System Association*, pp. 167-178, Milwaukee.
- KARLAPALEM, K., LI, Q. & SHUM, C. D., 1995, "HODFA: An architectural framework for homogenising heterogeneous legacy databases", *SIGMOD Record*, v. 24, n. 1(Mar), pp. 15-20.
- KASHYAP, V. & SHETH, A., 1996, "Semantic and Schematic similarities between database objects: a context-based approach". In: *VLDB Journal*, Eds. P. Apers, H. Scheck, J. Gray & S. Su, pp. 276-304.
- KIM, W. & SEO, J., 1991, "Classifying schematic and data heterogeneity in Multidatabase Systems", *IEEE Computer*, v. 24, n.12 (Dez), pp. 12 -18.
- KIM, W., 1991, "Observations on the ODMG-93 proposal for object-oriented database language", *SIGMOD Record*, v. 23, n. 1 (Mar), pp. 4-9.
- KIM, W., 1995, "Introduction to part 2: Technology for interoperating legacy Databases". In: *Modern Database Systems: the object model, interoperability and beyond*, Won Kim Editor, ACM Press, New York, pp. 510-520.
- KIM, W., CHOI, I., GALA, S. & SCHEEVEL, M., 1995, "On resolving schematic heterogeneity in Multidatabase Systems". In: *Modern Database Systems: the object model, interoperability and beyond*, Won Kim Editor, ACM Press, New York, pp. 521-550.
- KONOPNICKI, D & SHMUELI, O., 1995, "W3QS: A query system for the World-Wide Web". In: *Proc. of the XXI VLDB Conference*, pp. 54-65, Zurich.
- KRAMER, R, NIKOLAI, R, KOSCHEL, A., ROLKER, C., LOCKERMANN, P. & KEITEL, A., 1997, "WWW-UDK: a web-based environmental meta-Information System", *SIGMOD Record*, v. 26, n. 1 (Mar), pp. 16-21.
- KRISHNAMURTH, R. LITWIN, W. & KENT, W., 1991, "Language features for interoperability of Database with schematic discrepancies". In: *Proc. of ACM SIGMOD International Conference on Management of Data*, v. 20, n. 2, pp. 40-49, Denver, Jun.
- KUCERA, H., CHIN, R. & JAMESON, L., 1993, "SAIF - Conceptualisation to realisation". In: *Proc. GIS'93: Eyes on the future, Seventh Annual Symposium*, Vancouver: British Columbia.

- KÜHN, E. & LUDWIG, T., 1988, VIP-MDBS: a logic multidatabase system. In: *Proc. of International Symposium on Database in Parallel and Distributed Systems*, IEEE, , pp. 190-199, Texas: USA.
- KUMAR, K., SINHA, P & BHATT, P. C. P. P., 1997, "Do-GIS : A distributed and object oriented GIS". In: *Advances in GIS Research II: Proc. of VII International Symposium on Spatial data Handling*. Ed. Kraak, Molenaar, Taylor & Francis, pp. 263-275.
- LAURINI, R. & THOMPSON, D., 1992, *Fundamentals of Spatial Information Systems.*, 1a. ed., Londres, Academic Press.
- LAURINI, R., 1994, "Sharing geographic information in distributed databases". In: *Proc. of Annual Conference Urban and Regional Information System Association*, pp. 441-454, Milwaukee.
- LECLERCQ, E., BENSLIMANE, D. & YÉTONGNON, K., 1996, "A distributed object architecture for interoperable GIS", In: *IX International Conference on Parallel and Distributed Computing Systems*, v. 2, pp. 10-11, Dijon: França., Set.
- LECLERCQ, E., BENSLIMANE, D. & YÉTONGNON, K., 1996, "A distributed object architecture for interoperable GIS", In: *IEEE Knowledge and Data Engineering Exchange Workshop*, pp. 73-80, New Port Beach: USA.
- LECLERCQ, E., BENSLIMANE, D. & YÉTONGNON, K., 1997a, "A distributed object oriented model for heterogeneous spatial Databases", In: *X International Conference on Parallel and Distributed Computing Systems*, pp. 625-629, Louisiana: USA.
- LECLERCQ, E., BENSLIMANE, D. & YÉTONGNON, K., 1997b, "Interoperability of GIS: a distributed object architecture based on a canonical and functional models", In: *IX International Association of Science and Technology for Development*, pp. xx-yy, Innsbruck: Austria.
- LIM, E.P, SRIVASTAVA, J. & SHEKHAR S., 1994, "Resolving attribute incompatibility in Database integration and evidential reasoning approach". In: *X International Conference on Data Engineering*, pp. 14-18, Houston, Feb.
- LIRA, F., 1998, "Projeto de Qualidade de Águas e Controle de Poluição do Ministério de Planejamento e Orçamento", Relatório do Projeto PNUD BRA 96/017.

- LISBOA F., J. & IOCHPE, C., 1996, Análise comparativa dos modelos de dados conceituais para Sistemas de Informações Geográficas. Relatório de Pesquisa do Instituto de Informática da UFRGS, RP 266/96.
- LITWIN, W., MARK, L. & ROUSSOPOULOS, N., 1990, "Interoperability of Multiple autonomous databases", *ACM Computing Survey*, v. 22, n. 3 (Set), pp. 265-293.
- MADNICK, S. E., 1995, "From VLDB to VMLDB: dealing with large scale semantic heterogeneity". In: *Proc. of the XXI VLDB Conference*, pp. 11-16, Zurich: Switzerland, Set.
- MALAIKA, S., 1996, "Database management systems and the Internet". In: *Proc. of the XXII VLDB Conference*, pp. 600-601, Mumbai, India, Ago.
- MANOLA, F., 1995, "Interoperability issues in large-scale distributed object systems", *ACM Computing Surveys*, v. 27, n. 2 (Jun), pp. 265-267.
- MAURO, R. C. & MATTOSO, M. L.; "Integração de LPOO e BDOO: uma experiência com JAVA™ e GOA ++", In: *Anais do XIII SBBD*, Maringá: Brasil, pp. 169-184, Out.
- MELO, R. N., 1988, *Banco de Dados Não convencionais: a tecnologia do BD e as novas áreas de aplicação*. VI Escola de Computação, Campinas, R Vieira Gráfica e Editora Ltda.
- MENG, W. & YU, C., 1995, "Query Processing in Multidatabase Systems". In: *Modern Database Systems: the object model, interoperability and beyond*, Won Kim Editor, ACM Press, New York, pp. 551-572.
- MILLINER, S., BOUGUETTAYA, A. & PAPAZOGLU, M., 1995, "A scalable Architecture for autonomous heterogeneous database interactions". In: *Proc. of the XXI VLDB Conference*, Zurich, pp. 515-526.
- MILNE, P, MILTON, S. & SMITH, J., 1993, "Geographical object oriented databases - a case study", *Int. Journal. Geographical Information Systems*, Eds. Taylor e Francis, v. 7, n. 1, pp. 39-55.
- MORELLI, J. C. L, POUBEL, M. B. P. & AQUINO, A. M. F. A, 1994, "Sistema de Informação Geográfica para previsão de safras". In: *GIS Brasil 94*, Ed. Sagres, pp. 33-38, Curitiba: Paraná, Mai.
- NICOLLE C. & CULLOT, N., 1996, "Data models translation in heterogeneous CIS. In: *Proc. of the International Conference Parallel and Distributed Computing Systems*, Dijon: França, v. 2, pp. 605-610, Set.

- OLIVEIRA, R. L. & MAGALHÃES, G. C.; 1993, Metodologias para conversão de esquemas em sistemas de Banco de Dados heterogêneos, Relatório Técnico DCC-17/93; UNICAMP.
- ONSRUD, H. J. & RUSHTON, G., 1995, "Sharing Geographic Information: An introduction". In: Onsrud, H. J. & Rushton, G. (eds), *Sharing Geographic Information*, 1a. ed. New Jersey: USA.
- OSTENSEN, O., 1995, Mapping the future of Geomatics, ISO/TC 211 Bulletin.
- ÖSZU, M. T. & VALDURIEZ, P., 1990, "Principles of Distributed database Systems", 1a. ed., Paris, Prentice Hall International.
- PARENT, C., SPACCAPIETRA, S. & DEVOGELE, T., 1996 "Conflicts in spatial database integration". In: *Proc. of the International Conference Parallel and Distributed Computing Systems*, v. 2, pp. 772-778, Dijon: França, Set.
- PALMA, S.; ANICETO, M. V. & SOUZA, J. M., 1997a, "Collaborative perspective in GIS environment: COOMAN", In: *Proc. of CRIWG'97*, pp. 09-16, Out.
- PALMA, S.; ANICETO, M. V. & SOUZA, J. M., 1997b, "Workflow model and architecture GIS environment", In: *Proc. of II International workshop on CSCW Design*, Bangkok: Tailândia.
- PERINE, L. A., 1995, "In Pursuit of an optimum: a conceptual model for examining public sector policy support of interoperability". In: *Proc. of the Workshop on interoperability and economics of information infrastructure*, Virginia, Jul ou <http://nii.nest.gov/pus/optimum.html>.
- PERIZZO, W., RAJKUMAR, J. & RAM, P., 1991, "HYDRO: a Heterogeneous Distributed Database System". In: *Proc. of ACM SIGMOD Record*, v. 20, n. 2, pp. 32-40.
- PIRES, F.; MEDEIROS, C.B. & SILVA, A. B.; 1993, "Modelling Geographic Information Systems using an Object Oriented Framework", In: *Proc. of the XIII International Conference of the Chilean Computer Society*, Chile, pp. 217-232.
- PIRES, P. F., 1997, *HIMPAR, uma arquitetura para interoperabilidade de objetos distribuídos*. Dissertação de Mestrado em Engenharia de Sistemas e Computação, COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- PITOURA, E., BUKHRES, O., & ELMAGARMID, A., 1995, "Object orientation in Multidatabase systems", *ACM Computing Survey*, v. 27, n. 2(Jun), pp. 141-195.

- QIAN, X., 1993, "Semantic interoperability via intelligent mediation"; In: *Proc. of the III International Workshop on Research issues in Data Engineering: interoperability in Multidatabase Systems*, pp, 228-231, Abr.
- QIAN, X.; 1995, *Correct schema transformations*, In: Technical Report SRI CSL 95-08, Computer Science Laboratory, SRI International, Jul.
- RADEKE, E. & SCHOLL, M. 1994a, "Federation and stepwise reduction of database systems". In: *Proc. of I International Conference on Applications of Database*, pp. 381-399, Sweden.
- RADEKE, E. & SCHOLL, M. 1994b, "Framework for object migration in federated database systems", In: *Proceedings of the III International Conference on Parallel and Distributed Information Systems*, IEEE Computer Society Press, Austin: Texas, pp. 187-194, Set.
- RAMIREZ, M. R., 1994, *Sistemas Gerenciadores de Banco de Dados para Geoprocessamento*. Dissertação de Mestrado da COPPE/UFRJ, Rio de Janeiro, Rio de Janeiro.
- REDDY, M. P., PRASAD, B. E., REDDY, P. G. & GUPTA, A., 1994, "A methodology for integration of heterogeneous databases", *IEEE Transactions on knowledge and Data Engineering*, v. 6, n. 6(Dez), pp. 920-933.
- RIBEIRO, C. H. F. P. & OLIVEIRA, J. P. M., 1995, "Heterogeneity in object with roles conceptual modelling". In: IX. Simpósio Brasileiro de Banco de Dados, Recife, pp. 188-200.
- RIBEIRO, G. P., 1996, *Sistemas de Informação Geográfica: alguns conceitos básicos e aplicações*, IV Encontro do Instituto de Geociências da UFF.
- ROSENTHAL, A. & SELIGNAM, 1994, "Data integration in the large: the challenge of reuse". In: *Proc. of the XX VLDB Conference*, Santiago, Chile, pp. 669-675.
- SAIF, Survey and Resource Mapping Branch, 1995 "Spatial Archive and Interchange format: Formal Definition, release3.2", Ministry of Environment, Lands and Parks, Province of British Columbia, Canada.
- SALTOR, F, CASTELLANOS, M. & GARCIA-SOLACO, G., 1991, "Suitability of data models as canonical models for Federated Databases", *SIGMOD Record*, v. 20, n. 4 (Dez), pp. 44-48.
- SALTOR, F., CAMPDERRICH, B., RODRÍGUEZ, E & RODRÍGUEZ, L. C., 1996 "On schema levels for federated Database Systems". In: *Proc. of the International*

- Conference Parallel and Distributed Computing Systems*, v. 2, pp. 766-771, Dijon: França, Set.
- SCHEUERMANN, P., YU, CLEMENT, MOLINA, H. G., MANOLA, F., MCLEOD, D., ROSENTHAL, A. & TEMPLETON, M., 1990, "Report on the workshop on heterogeneous Database Systems held at North-western University", *SIGMOD Record*, v. 19, n. 4 (Dez), pp. 23-31.
- SCHMITT, I. & SAAKE, G., 1996, "Schema integration and view derivation by resolving intentional and extensional overlapping". In: *Proc. of the International Conference Parallel and Distributed Computing Systems*, v. 2, pp. 751-758, Dijon: França, Set.
- SCIORE, E., SIEGEL, M. & ROSENTHAL, A., 1994, "Using semantic values to facilitate interoperability among heterogeneous Information Systems", *ACM Transaction on Database Systems*, v. 19, n. 2 (Jun), pp. 254-290.
- SHAN, M. C., AHMED, R., DAVIS, J. DU, W. & KENT, W., 1995, "Pegasus : a heterogeneous Information management systems" In: *Modern Database Systems: the object model, interoperability and beyond*, Won Kim Editor, ACM Press, New York, pp. 664-681.
- SHETH, A. & RUSINKIEWICZ, M., 1990, "Management of interdependent data: specifying dependency and consistency requirements". In: *Proc. of the Workshop on the management of replicated data*, Houston, Texas ou http://lstdis.cs.uga.edu/publications/pub_CONSISTENCY.html.
- SHETH, A. P. & LARSON, J. A., 1990, "Federated Database Systems for managing Distributed, heterogeneous and autonomous databases", *ACM Computing Surveys*, v. 22, n. 3 (Set), pp. 183-236.
- SHETH, A. P., 1991, "Semantic issues in Multidatabase Systems", *SIGMOD Record*, v. 20, n. 4(Dez), pp. 5-9.
- SHETH, A., 1995, "Data Semantics: what, where and how", Technical Report TR-CS-95-003, Laboratory Scale Distributed Information System, Lab of Computer Science, University of Georgia.
- SHKLAR, L., CHARLES, C. ST & BEHRENS, C., 1995, Geoscope Detailed Design, Bellcore International Document.
- SILBERSCHATZ, A. & ZDONIK, S., 1997, "Databases Systems: breaking out of the box", *SIGMOD Record*, v. 26, n. 3 (Set), pp. 36-50.

- SMAALEN, J.W.N. Van, 1996, "A hierarchical rule model for Geographic Information abstraction". In: *Proc. of XVII International Symposium on Spatial Data Handling: Advances in GIS Research II*, Ed. M. J. Kraak & M. Molenaar, pp. 215-225.
- SOLACO, M.G., SALTOR, F. & CASTELLANOS, M., 1996, Semantic Heterogeneity in Multidatabase Systems, In: *Object-Oriented Multidatabase Systems: a solution for advanced applications*, Ed. Omran A. Bukhres & Ahmed K. Elmagarmid, New Jersey, pp. 129-202.
- SOUZA, J. M., 1986, *Software tools for Conceptual Schema integration*, Ph.D. Dissertation, University of East Anglia, Local.
- SOUZA, J.M., FERRARI, R., DUARTE, M., E RAMIREZ, M., 1993, "Uma Arquitetura Organizacional para Sistemas de Informação Geográfica Orientados a Objetos". In: *Anais da IV Conferência Latino Americana sobre Sistemas de Informação Geográfica*, São Paulo, pp.
- SPACCAPIETRA, S & PARENT C., 1991, "Conflicts and correspondence assertions in interoperable databases", *SIGMOD Record*, v. 20, n. 4, pp. 49-54.
- SPACCAPIETRA, S. & PARENT, C., 1994, "View integration: a step forward in solving structural conflicts", *IEEE Transactions on Knowledge and Data Engineering*, v.6, n. 2 (Abr), pp. 240-274.
- STRAUCH, J. C. M & MATTOSO, M. L. Q., 1995, *Interoperabilidade entre bases de dados heterogêneas e distribuídas*. In: Relatório Técnico ES-369/96 da COPPE/UFRJ.
- STRAUCH, J. C. M., SOUZA, J. M & MATTOSO, M. L. Q., 1996, Issues in interoperability between GIS databases with heterogeneity and distribution aspects". In: *Proc. of GIS Brasil 96*, Ed. Sagres, Curitiba, pp. 554-566.
- STRAUCH, J. C. M., SOUZA, J. M., & MATTOSO, M. L. Q., 1998c, Estudo do metamodelo Spatial Archive and Interchange Format (SAIF), In: Relatório Técnico da COPPE/UFRJ.
- STRAUCH, J. C. M., SOUZA, J. M., & MATTOSO, M. L. Q., 1998d, Estudo dos modelos de dados dos Sistemas de Informação Geográfica ARC/INFO e SPRING, In: Relatório Técnico da COPPE/UFRJ.
- STRAUCH, J. C. M.; SOUZA, J. M & MATTOSO, M. L. Q., 1998a, A methodology for GIS Databases integration". In: *Proc. of KDEX'98*, (aceito), Taiwan.

- SUBRAHMANIAN, V. S., ADALI, S., BRINK, A., EMERY, R., LU, J., RAPJUT, A., ROGERS, T., ROSS, R., WARD, C., 1995, "Hermes: a heterogeneous, reasoning, and mediators system", University Of Maryland, Maryland ou <http://www.cs.umd.edu/projects/hermes/overview/paper/index.html>.
- SULLAIMAN, A. & SOUZA, J. M., 1998a, "Técnicas de modelagem de dados em projetos de data warehouse", *Developer's Magazine*, Fev.
- SULLAIMAN, A. & SOUZA, J. M., 1998b, "Representação do conhecimento em data warehouse", *Developer's Magazine*, Abr.
- TARI, Z., CHENG, W., YETONGNON, K. & SAVNIK, I., 1996 "Towards cooperative databases: the DOK approach". In: *Proceedings of the International Conference Parallel and Distributed Computing Systems*, v. 2, pp. 595-600, Dijon: França, Set.
- TEMPLETON, M., HENLEY, H., MAROS, E. & BUER, D., 1995, "Intervisio: dealing with complexity of federated database access", *VLDB Journal*, n. 4, pp. 287-317.
- TIMES, V. C. & SALGADO, A. C., 1994, "Uma modelagem orientada a objetos para aplicações geográficas" In: *Anais do IX Simpósio Brasileiro de Banco de Dados*, pp. 293-309, São Carlos: São Paulo, Set.
- TIMPF, S., RAUBAL, M. & KUHN, W., 1997, "Experiences with metadata". In: *Proc. of XVII International Symposium on Spatial Data Handling: Advances in GIS Research II*, Ed. M. J. Kraak & M. Molenaar.
- TSICHRITZIS, D. C. & LOCHOVSKY, F. H., 1982, *Data models*, 1a. ed., New Jersey, Prentice Hall.
- UCHÔA, E. M. A., 1995, *Heros - Um sistema de Banco de Dados heterogêneos*. Dissertação de Mestrado do Departamento de Informática da PUC-RIO.
- VCKOVSKI, A. & BUCHER, F., 1996, "Virtual data sets: smart data for environmental applications. In: *Third International Conference /Workshop on Integrating GIS and Environmental modelling*, Santa Fe: USA ou http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/sf_papers/vckovski_andrej/vbpaper.html
- VENTRONE, V. & HEILER, S., 1991, "Semantic heterogeneity as a result of domain evolution", *SIGMOD Record*, v. 20, n. 4 (Dez), pp. 16-20.
- VERMEER, M. W. W. & APERS, P. M. G., 1996, "The role of integrity constraints in database interoperation". In: *Proc. of the XXII VLDB Conference*, pp. 425-435, Mumbai: India, Ago.

- VIDAL, V. M. P., 1994; "Preservando a semântica de atualizações de visões", Tese de Doutorado da COPPE/UFRJ, Rio de Janeiro, Rio de Janeiro.
- WANG, K. & ZHANG, W., 1996, "Detecting data inconsistency for Multidatabase". In: *Proceeding of the International Conference Parallel and Distributed Computing Systems*, v. 2, pp. 657-663, Dijon: França, Set.
- WEISHAR D. & KERSHBERG, 1991, "Data/knowledge packets as means of supporting semantic heterogeneity in Multidatabase Systems", *SIGMOD Record*, v. 20, n. 4 (Dez), pp. 69-73
- WIEDERHOLD, G., 1992, "Mediators in the architecture of future Information Systems", *IEEE Computer*, v. 25, n.3 (Mar), pp. 38-49.
- WORBOYS, M. F, HEARNshaw, H. M. & MAGUIRE, D.,1990, "Object oriented data modelling for spatial databases", *Geographical Information Systems*, Ed. Taylor & Francis Ltd., v. 4, n.4, pp369-383.
- WORBOYS, M. F., 1994, "Object -oriented approaches to geo-referenced information", *Int. Journal Geographical Information Systems*, Eds. Taylor & Francis, v. 8, n. 4, pp. 385-399.
- WORKBOYS, M. F. & DEEN, S. M., 1991, "Semantic heterogeneity in Distributed Geographic Databases", *SIGMOD Record*, v. 20, n. 4, pp. 30-34.
- YASER,B., 1997, A mechanism for object identification and transfer in a heterogeneous distributed GIS. In: *Advances in GIS Research II: Proc. of VII International Symposium on Spatial Data Handling*. Ed. Kraak, Molenaar, Taylor & Francis, pp. 929-939.
- YU, C., JIA, B., SUN, W. & DAO, S., 1991, "Determining relationships among names in heterogeneous Databases", *SIGMOD Record*, v. 20, n. 4, pp. 79-80.
- ZAJAC, G., 1994, "Integration of GIS applications across multiple platform environments". In *Proc. of Annual Conference Urban and Regional Information System Association*, pp. 140-153, Milwaukee.
- ZHAO, J. L., SEGEV, A. & CHATTERJEE, A., 1995, "A universal relation approach to federated management". In: *Proc. of the 11th Conference on Data Engineering*, pp. 261-270, Taiwan.

7.2. Referências bibliográficas *on line*

- BUEHLER, K. & MCKEE, L, 1996, "The OpenGIS guide: Introduction to interoperable Geoprocessing", Ed Open GIS Consortium, OGIS TC 96-001, Wayland, Massachusetts ou <http://www.ogis.org/guide/guide.html>.
- CERES, 1997, "Geomatics technology for CERES", <http://www.regis.berkeley.edu/ceres/architecture.html>.
- GARDELS, K., 1996a, "Open Geodata interoperability services: an object oriented framework for accessing distributed, heterogeneous Geographic Information", http://epoch.cs.berkeley.edu:8000/sequoia/guernewood/ogis/sdh_ogis.html.
- GARDELS, K., 1996b, "Comprehensive data model for distributed, heterogeneous geographic information", http://www.regis.berkeley.edu/gardels/geomodel_def.html.
- GARDELS, K., 1996c, "Sequoia 2000 and Geographic Information: the Guernewood Geoprocessor", http://epoch.cs.berkeley.edu.8000/sequoia/guernewood/sdh_grnwd.html.
- GOODCHILD, M. F., 1996, "Directions in GIS", http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/sf_papers/goodchild_michael/good.html.
- GOODCHILD, M., 1997, "Interoperability of Geographic Information", <http://ra.forest.wisc.edu/mackay/interoperability.html>.
- MARRS, K. A, STEIB, S. A., ABRAMS, C. A. & KAHN, M. G., 1996, "Unifying heterogeneous distributed clinical data in a relational database", http://osler.wustl.edu/~marrs/marrs_scanc2.html.
- MATUSCHAK, B. J., 1996, "GIS is being redefined by current Computing Trends". In: The Electronic Atlas Newsletter, v. 7, n. 9 ou <http://www.electronic-atlas.com/ean79a.html>.
- VODAK, 1996, "VODAK: An object-oriented database for database integration", <http://www.coe.uncc.edu/~whstorey/vodak.html>.

7.3. Bibliografia complementar tradicional

- ABEL, D. J., KILBY, P. J. & DAVIS, J. R., 1994, "The systems integration problem", *International Journal of Geographical Information Systems*, v. 8, n. 1, pp. 1-12.

- BEHRENDT, W., FIDDIAN, N. F. & MADURAPPERNA, A. P., 1997, "Tools to enable interoperation of heterogeneous Databases". In: *XIII International Conference on Data Engineering*, pp. 577, Birmingham: United Kingston, Abr.
- BERGER, B. H. & SMITH, B. P., 1994, "The promise of the distributed computing environments and GIS". In: *GIS'94 Symposium*, pp. 245-250, Vancouver, Fev.
- BRAGA, R.M.M. & TRAVASSOS, G. H., 1996, *Arquiteturas de software: uma abordagem orientada a objetos*, Relatório Técnico COPPE/UFRJ.
- BRIGHT, M. W., HURSON, A. R. & PAKRAD., 1992, "A taxonomy and current issues in Multidatabase Systems", *IEEE Computer*, v 25, n. 5 (Mar), pp. 50-59.
- BROWN, P. & STONEBRAKER, M., 1995, "BigSur: a system for the management of the earth Science data". In: *Proc. of the 21st VLDB Conference*, pp. 720-728, Zurich.
- CALMET, J., JEKUTSCH, S. & SCHÜ, J., 1997, "A generic query-translation framework for a mediator architecture". In: *XIII International Conference on Data Engineering*, pp. 434-443, Birmingham: United Kingston, Abr.
- CÂMARA, G. CASANOVA, M. A., HERMELY, A. S., MAGALHES G. C. & MEDEIROS C. M. B, 1996, "Anatomia de Sistemas de Informação Geográfica". In: *X Escola de Computação, Instituto de Computação, UNICAMP, Campinas*.
- CASTANO, S. & ANTONELLIS, V., 1997, "Semantic dictionary design for Database interoperability". In: *XIII International Conference on Data Engineering*, pp. 43-54, Birmingham: United Kingston, Abr.
- CHATTERJEE, A. & SEGEV, A., 1992, "Resolving data heterogeneity in scientific and statistical databases". In: *Proc. of the Sixth International Working on Scientific and Statistical Database Management*.
- CHEUNG, W & CHENG, HSU, 1996, "The model -Assisted Global Query System for multiple Database in distributed enterprises", *ACM Transactions on Information System*, v. 14, n.4 (Out), pp. 421-470.
- CLEMENTI, E & FELICE, P., 1990, "An Extensible Class Library for Geographic Applications". In: *Proc. of the 2nd International Conference Tools*, Paris.
- CONRAD, S.; EAGLESTONE, B.; HASSELBRING, W.; ROANTREE, M.; SALTOR,F.; SCHÖNHOF, M; STRÄSSLER, M & VERMEER, M; 1997; "Research Issues in Federated Database Systems"; *SIGMOD Record*, v. 26, n. 4(Dez), pp. 54-56.

- DELIS, V. & HADZILACOS, T.; 1997, "On the assessment of generalisation consistency", In: *Lecture Notes in Computer Science: Advances in Spatial databases*, Ed. Michael Scholl & Agnes Voisard, n. 1262, pp. 321-335, Berlin, Jul.
- ENGENHOFER, M. J.; 1997, "Editorial: "Consistency revised"; *Geoinformatica*; Kluwer Academic Publishers, Boston, n. 1: 4; pp. 323 -325.
- FANG, D., HAMMER, J., MCLEOD, D. & SI, A., 1991, "Remote -exchange: an approach to controlled sharing among autonomous, heterogeneous Database Systems", In: *Proc. of IEEE Spring Compcon*, San Francisco, Feb.
- FANG, D.; HAMMER, J. & MCLEOD, 1991, "The identification and resolution of semantic heterogeneity in Multidatabase", In: *Proc. of I International Workshop on Interoperability in Multidatabase*, pp. 7-9, Kyoto, Abr.
- FERRANDINA, F. & FERRAN, G., 1995, "Schema and database evolution in the O2 object Database System". In: *Proc. of the 21st VLDB Conference*, Zurich.
- FERRARI, R., 1997, *Viagem ao SIG: Planejamento estratégico, viabilização, implantação e Gerenciamento de Sistemas de Informação Geográfica*, 1a ed. Curitiba, Editora Sagres.
- FONG, J., 1997, "Converting relational to object-oriented databases", *SIGMOD Record*, v. 26, n. 1 (Mar), pp. 53-58.
- FRANK, A. U. & KUHN, W., 1995, "Specifying Open GIS with functional languages". In: *Proc. of 4th International Symposium*, pp. 184-195, Portland.
- FRANKHAUSER, P; KRACKER, M. & NEUHOLD, E, 1992 "Semantic vs. Structural resemblance of classes", Integrated Publication & Information System Institute, Darmstat, Germany, Nov.
- GANGOPADHYAY, D. & BARSALOU, T., 1991, "On the semantic equivalence of heterogeneous representations in multimodel Multidatabase Systems", *SIGMOD Record*, v. 20, n. 4 (Dez), pp. 35-39.
- GELLER, J., PERL, Y. & NEUHOLD, E., 1991, "Structural schema integration in heterogeneous multidatabase systems using dual model", In: *Proc. of the first International Workshop on interoperability in Multidatabase Systems*, IEEE Computer Society Press, Los Alamitos: California, pp. 200-203.
- GODCHILD, M., HAINING, R. & WISE, S., 1992, "Integrating GIS and spatial data analysis: problems and possibilities". *International Journal Geographical Information Systems*, v. 6, n. 5, pp. 407-423.

- GOODCHILD, M. F., 1992, "Geographical data modelling", *Computers & Geoscience*, v. 18, n. 4 (Dez), pp. 401-408
- GORDON, T. F., KARACAPILIDIS, N. & VOSS, H., 1996, "Zeno: a mediation system for spatial planning, In: *Proc. of the ERCIM Workshop on CSCW and the WEB*, Sant Augustin: Alemanha, Fev.
- HABER, E. M., IOANNIDIS, Y. E. & LIVNY, M, 1995, "Opossum: Desk-top schema management through customisable visualisation". In: *Proc. of the XXI VLDB Conference*, pp. 525-538, Zurich.
- HAMMER, J. & MCLEOD, D., 1993, "An approach to resolving semantic heterogeneity in a federation of autonomous, heterogeneous Database Systems", *International Journal Of Intelligent and Cooperative Information Systems*, v. 2, n. 1, pp. 51-83.
- HAMMER, J., MCLEOD, D. & SI, A., 1993, "Object discovery and unification in Federated Database Systems", In: *Proc. of Workshop on interoperability of Database Systems and Database Applications*, pp. 3-18, Switzerland, Out.
- HAMMER, J.; MCLEOD, D. & SI, A.; 1994, "An intelligent system for identifying and integrating non local objects in federated database Systems", In: *Proc. of the XXVII International Conference on Systems Science*, Computer Society of the IEEE, pp. 389-407, Hawaii, Jan.
- HAO, Q., 1996, "The entity-structure relationship model: towards unified geographic modelling". In: *Proc. of Geoinformatics '96*, pp. 79-88, Florida.
- HULL, R & ZHOU, G., 1996, "A framework for supporting data integration using the materialised and virtual approaches", In: *Proc. of SIGMOD'96*, pp. 481-492, Montreal: Canada.
- ISHIKAWA, H. & KUBOTA, K. 1995, "An autonomous decentralised database System Based on Scripts and Active Objects". In: *II International Symposium on Autonomous Decentralised Systems*, pp. 185, 191, Phoenix.
- JACQUEZ, F., 1994, "How Lee County integrated GIS and legacy technologies to become Florida's most cost-efficient". In: *Proc. of Annual Conference Urban and Regional Information System Association*, pp. 863-872, Milkawkee.
- KALINICHENKO, L., 1991, "An anatomy of the information resource semantic abstraction", *SIGMOD Record*, v. 20, n. 4 (Dez), pp. 21-24.

- KELLEY, W., GALA, S., KIM, W., REYES, T. & GRAHAM, B., 1995, "Schema Architecture of the UniSQL/M Multidatabase Systems". In: *Modern Database Systems: the object model, interoperability and beyond*, Won Kim Editor, ACM Press, New York, pp. 621-649.
- KENT, W., 1991, "The breakdown of the information model in Multidatabase systems", *SIGMOD Record*, v. 20, n. 4 (Dez), pp. 10-15.
- KILBY, P. J. & ABEL, D. J., 1994, "The systems integration problem". *International Journal of Geographical Information Systems*, v. 18, n. 1, pp. 1-12.
- KLIPARCHUK, K. & HOWELL, D., 1994, "Issues in integrating computer systems and applications for mapping purposes in economically emerging countries". In: *GIS'94 Symposium*, pp. 215-221, Vancouver, Fev.
- LEUNG, C. H. C. & KASPI, S., 1994, "A flexible paradigm for semantic integration in Cooperative heterogeneous databases". In: *International Symposium on X Generation Computer Systems*, pp. 141-150, Dez.
- LEUNG, C. H. C. & KASPI, S., 1994, "A flexible paradigm for semantic integration in cooperative heterogeneous database". In: *Int. Symposium on fifth generation Computer Systems*, Workshop on Heterogeneous Cooperative knowledge Bases, pp. 141-150.
- LI, B., 1996, "Strategies for developing network oriented GIS software". In: *Proc. of Geoinformatics'96*, Florida, pp. 162-170.
- LINN, C. & HOWARTH, B., 1994, "A proposed globally distributed Federated Database: a particle performance evaluation". In: *Proc. of the Third International Conference on Parallel and Distributed Information Systems*, IEEE Computer Society Press, Austin, pp. 203-212.
- LITWIN, W., KETABCHI, M. & KRISHNAMURTHY, 1991, "First order normal form for relational databases and Multidatabases", *SIGMOD Record*, v. 20, n. 4 (Dez), pp. 74-76.
- MACFARLANE, A., MCCANN & LIDDELL, H., 1996, "A common data model for meta-data interoperable environments". In: *II International Baltic Workshop on Database and Information Systems*.
- MACFARLANE, A., MCCANN & LIDDELL, H., 1996, "Management tools for distributed interoperable environments", HPCN'96, Brussels, LNCS, Springer, VERIAG.

- MCKEE, L., 1995, "The converging Open GIS drivers", *GIS World*, pp. 76-78.
- MEDEIROS, C. B. & PIRES, F., 1994, "Databases for GIS", *SIGMOD Record*, v. 23, n. 1 (Mar), pp. 107-115.
- MEHROTA S., RASTOGI, R., BREITBART, Y. & SILBERSCHATZ, A., 1992, "The concurrency control problem in Multidatabase: characteristics and solution". In: *Proc. of ACM SIGMOD Record*, v. 21, n. 2, pp. 288-297.
- MELTON, J. & MATTOS, N., 1996, An overview of SQL3: the emerging new generation of the SQL standard - Tutorial.. In: *Proc. of XXII International Conference on Very Large Data Bases*, Mumbai: India, Ago.
- MERZ, U. & KING, R., 1992, "DIRECT: a query facility for multiple Databases". In: *Proc. of ACM SIGMOD Record*, v. 21, n. 2, pp. 2, Jun.
- NEBERT, D., 1996, Information Architecture of a Clearinghouse, WWW Access to Earth Observation/ Geo-referenced Data Workshop, WWW Conference.
- NEIMAT, M. A. & SHAN, M. C., 1991, "Database research at HP Labs", *SIGMOD Record*, v. 20, n. 4 (Dez), pp. 92-97.
- OLIVEIRA, J. L. & MEDEIROS, C. B., 1997, "An environment for modelling and design of Geographic Application", *Geoinformatica*, Kluwer Academic Publishers, Boston, n. 1; pp. 29-58.
- ÖSZU, M. T. & VALDURIEZ, P., 1991, "Distributed Database Systems: where are we now?", *IEEE Computer*, v. 24, n. 8 (Ago), pp. 68-78.
- PAPAKONSTANTINOU, Y., GARCIA-MOLINA, H. & WIDON, J., 1996, "Object fusion in mediator system", In: *Proceedings of XXII VLDB Conference*, Mumbai: India, pp. 413-424 ou disponivel via anonymous ftp: pu/papakonstantinou/1995/fusion.ps de db.stanford.edu.
- PENG, W. & TEMPFLI, K., 1996, "Automated generalisation in a GIS context". In: *Proc. of Geoinformatics '96*, Florida pp. 41-49.
- PIRES, P. F. & MATTOSO, M. L. Q., 1996, Interoperability and heterogeneity issues on the HIMPARG Database Architecture. In: *Anais do XI Congresso de Banco de Dados*.
- QI, L. & TIANJUN, L., 1996, "Global Wide Web: the next generation spatial information system". In: *Proc. of Geoinformatics '96*, Florida, pp. 172-176.

- QIAN, X. & RASCHID, L., 1995, "Query interoperation among oriented and relational databases". In: *Proc. of the International Conference on Data Engineering*, pp. 271-278.
- RAM, S., 1991, "Heterogeneous Distributed Database Systems", *IEEE Computer*, v. 24, n.12 (Dez), pp. 7-10.
- RAMANATHAN, S. & HODGES, J., 1997, "Extraction of object-oriented structures from existing relational databases", *SIGMOD Record*, v. 26, n.1 (Mar), pp. 59-64.
- RAND, R.; 1994, "Meta-tools for enhanced access to global change multiple distributed metadata directories and data collections", In: *Proc. of the GIS/LIS*, pp. 639-646 ou <http://www.sgi.ursus.maine.edu/gisweb/spatdb/gis-lis/gis94080.html>.
- RIBEIRO, C. H. F. P. & OLIVEIRA, J. P. M. DE, 1996 "Multidatabase interoperability through conceptual schema mapping in an object oriented model". In: *Proceedings of the International Conference Parallel and Distributed Computing Systems*, v. 2, pp. 647-652, Dijon: França, Set.
- RIBEIRO, C. H. F. P. & OLIVEIRA, J. P. M., 1995, "Tratamento de heterogeneidade em esquemas conceituais orientados a objetos". In: *Simpósio Brasileiro de Banco de Dados*, Recife, pp. 261-273
- RIBEIRO, G. P. & SOUZA, J. M., 1995, "Padrões de conteúdo para metadados geoespaciais digitais". In: *Anais do XVII Congresso Brasileiro de Cartografia*.
- RIBEIRO, G. P., SOUZA, J. M. & FREITAS, A., 1996, "Digital Metadata: a Brazilian case of Federal Databases". In: *Proc. of I IEEE Metadata Conference*, Maryland, EUA.
- ROSEENTHAL, A. & REINER, D., 1994, "Tools and transformations - Rigorous and otherwise for practical database design", *ACM transactions on Database Systems*, v. 19, n. 2.
- RUSINKIEWICZ, M. & SHETH, A., 1995, "Specification and execution of transactional workflow". In: *Modern Database Systems: the object model, interoperability and beyond*, Won Kim Editor, ACM Press, New York, pp. 592-620.
- RUSINKIEWICZ, M. SHETH, A & KARABATIS, G., 1991, "Specifying interdatabase dependencies in a Multidatabase environment", *IEEE Computer*, v. 24, n. 12 (Dez), pp. 46-51.

- SHELLEY, E.P & JOHMDON, B. D., 1995, "Metadata: Concepts and models". In: *Proc. of the III National Conference on the Management of Geoscience Information and Data*, pp. 1-5, Jul, Adelaide: Australia.
- SHETH, A. & GALA, S. K., 1991, "Attribute relationships: an impediment to automating schema integration". In: *Proc. of the Workshop on heterogeneous Database Systems*, pp. 11-13, Chicago, Dez.
- SHETH, A. & GALA, S. K., NAVATHE, S. B., 1993, "On automatic reasoning for schema integration". *International Journal of Intelligent and Cooperative Information Systems*, v. 2, n.1, pp. 23-50.
- SHETH, A. & KASHYAP, V. 1992, "So far (schematically) yet so near (semantically)". In: *Proc. of IFIP DS-5: Conference on semantics of interoperable database Systems*, pp. 272-301, Lorne: Australia.
- SHETH, A., 1990, "Database research at Bellcore", *SIGMOD Record*, v. 19, n. 3 (Set), pp. 45-52.
- SILBERSCHATZ, A., STONEBRAKER, M. & ULLMAN, J., 1990, "Databases Systems: achievements and opportunities", *SIGMOD Record*, v. 19, n. 4 (Dez), pp. 6-23.
- SILVA, S. da. & MELO, R. N., 1995, "Um modelo de transação para sistemas de gerência de Banco de Dados heterogêneos". In: *Anais do Simpósio Brasileiro de Banco de Dados*, Recife.
- SOLACO, M.G., SALTOR, F. & CASTELLANOS, M., 1995, "A structure based schema integration methodology". In: *Proc. of XI Int. Conference of Data Engineering*, IEEE Computer Society Press, pp. 505-512, Taipei: Taiwan, Mar.
- SOUZA, J.M. & PALMA, S., 1994, "CooMan: A Global Collaborative Project Management System". In: *Proc. of CHEP-94 Conference*, San Francisco, EUA.
- STONEBRAKER, M., AOKI, M. P., DEVINE, R., LITWIN, WITOLD & OLSON, M., 1994, "Mariposa: a new architecture for distributed data". In: *Proc. of X Int. Conf. on Data Engineering*, pp. 54-65, Houston: Texas, Fev.
- STOUT, R. L., 1995, "EDA/SQL". In: *Modern Database Systems: the object model, interoperability and beyond*, Won Kim Editor, ACM Press, New York, pp. 650-663.
- STRAND, E. J., 1993, "How open is your GIS?", *GIS World*, pp. 36-38.

- STRAUCH, J. C. M. & SOUZA, M. J., 1995, "*Uma metodologia para implementação de Sistemas de Informações Geográficas*". In: Relatório Técnico ES-341/95 da COPPE/UFRJ.
- STRAUCH, J. C. M. & SOUZA, M. J., 1998, "Uma Metodologia para Implantação de Sistemas de Informações Geográficas", In: Saúde e Espaço, Eds Najar, A. L. & Marques E. C. Ed. FIOCRUZ, Rio de Janeiro, pp. 109-124.
- STRAUCH, J. C. M., MATTOSO, M. L. Q., 1994, *Sistemas de Informações Geográficas Orientados a Objetos*. In: Relatório Técnico ES-331/94 da COPPE/UFRJ.
- STRAUCH, J. C. M., MATTOSO, M. L. Q., 1997, Aplicação da Orientação a Objeto a Sistemas de Informações Geográficas, In.: *I Workshop de Orientação a Objetos*, COPPE/UFRJ, pp. 11-15.
- STRAUCH, J. C. M., MATTOSO, M. L. Q., 1997, Orientação a Objeto aplicada a GIS, In. *Fator GIS*, n. 20, pp. 58-60.
- STRAUCH, J. C. M., ZIMBRÃO, G., SOUZA, J. M., NETO, M. S., BRAGA, R., 1994, "Critérios de qualidade para groupware", In: *PANEL 94*, Cidade do México.
- STRAUCH, J. C. M., MATTOSO, L. de Q. & SOUZA, M. J., 1996, "Interoperabilidade de bases de dados espaciais heterogêneas e distribuídas, "In: *Anais do I SEGEO: Geoprocessamento: Mitos & Realidade*", pp. 67-75, Rio de Janeiro.
- STRAUCH, J. C. M., MATTOSO, L. de Q. & SOUZA, M. J., 1996, "Interoperabilidade de bases de dados heterogêneas e distribuídas, In: Relatório Técnico ES-369/96 da COPPE/UFRJ.
- SUN, Y., DU, J. & CHEN, Q., 1996 "Research on client/server GIS". In: *Proc. of Geoinformatics '96*, pp. 157-160, Florida.
- TARI, Z. & STOKES, J., 1997, "Designing the Reengineering service for the DOK Federated Database System", In: *XIII International Conference on Data Engineering*, pp. 465-475, Birmingham: United Kingston, Abr.
- THOMAS, G., TOMPSON, G. R., CHUNG, C. W., BARKMEYER, E., CARTER, F, TEMPLETON, M., FOX, S. & HARTMAN, B., 1990, "Heterogeneous distributed databases systems for production use", *ACM Computing Survey*, v. 22, n. 3 (Set), pp. 237-266.

- THOMPSON, G, MOLINA, H. G., GRAY, J., SILBERSCHATZ, A. & STONEBRAKER, M., 1991, "Are Standards the panacea for Heterogeneous Distributed DBMS?", In: *Proc. of ACM SIGMOD Record*, v. 20, n. 2, pp. 446.
- TOMASIC, A. & SIMON, E. 1997, "Improvising access to environmental data using context information", *SIGMOD Record*, v. 26, n. 1(Mar), pp. 11-15
- TRAINA Jr., C.; TRAINA, A. J. M. & BIAJIZ, M; 1994, "O papel da abstração de instanciação em um metamodelo de abstrações para BDOO", In: *Anais do IX Simpósio Brasileiro de Banco de Dados*, pp. 173-187, São Carlos: São Paulo, Set.
- TRESCH, M. & SCHOLL, A, 1994, "A Classification of Multi-database languages". In: *Proc. of the Third International Conference on Parallel and Distributed Information Systems*, Austin: Texas, IEEE Computer Society Press, pp. 195-202, Austin: Texas, Set.
- ULLMAN, J., 1997, "Information integration using logical views", In: *Proceedings of ICDT 97*, Delphi: Grecia, Springer.
- URBAN, S. D. & WU, J., 1991, "Resolving semantic heterogeneity through the explicit representations of data model semantics", *SIGMOD Record*, v. 20, n. 4, pp. 55-58.
- VASSALOS, V. & PAPAKONSTANTINOU, Y., 1997, "Describing and using query capabilities of heterogeneous source", In: *Proceedings of the XXIII VLDB Conference*, Atenas: Grecia.
- VCKOVSKI, A., 1996, "Java as a software system for distributed and interoperable Geoprocessing". In: *Proc. of the International Conference Parallel and Distributed Computing Systems*, v. 2 779-783, Dijon: França, Set.
- VERSO, J., 1997 "Data integration and interrogation". In: *XIII International Conference on Data Engineering*, pp. 588, Birmingham: United Kingston, Abr.
- WALKER, D. R. F. & NEWMAN, I. A., MEDYCKYJ-SCOTT, D. J. & RUGGLES, C. L. N., 1992, "A system for identifying datasets for GIS users". In: *International Journal of Geographical Information Systems*, v 6., n. 6, pp. 511-527.
- WIDOM, J., 1996, "Integrating heterogeneous databases: lazy or eager?", *ACM Computer Survey*, v. 28, n. 714 (Dez), pp. 1-2.
- WIEDERHOLD, G., 1995, "Mediation in Information Systems", *ACM Computing Surveys*, v. 27, n.2 (Jun), pp. 265-267.
- ZIMBRÃO, G, 1997, Interoperabilidade e integração: uma proposta cooperativa, Relatório da COPPE/UFRJ, 26 pg.

7.4. Bibliografia complementar on line

MARKOWITZ, V. M. & RITTER, O., 1995, "There's never time to do it right, but always time to do it over", http://gizmo.lbl.gov/DM_TOLLS/PAPERS/TNT/TNT.html.

NCGIA Workshop on interoperability for GIS: [http://www.spatial.maine.edu/~max/interoperability WS.html](http://www.spatial.maine.edu/~max/interoperability_WS.html).

Open GIS Consortium, 1996a, "The OGC Technical Committee Request Plan & Schedule", Wayland, Massachusetts ou <http://opengis.org/public/request.html>.

Open GIS Consortium, 1996b, "The OGC Technical Committee Technology development process", Wayland, Massachusetts ou <http://opengis.org/public/process.html>.

RADEKE, E. "Extending ODMG for federated database systems", <http://www.cadlab.de/~elke>.

SHKLAR, L, THATTE,S., MARCUS H. & SHETH A., "The inforHanness Information Integration", <Http: 141.142.3.76/SDG/IT94/Proceedings/shklar/shklar.html>.

Apêndice A - Esquemas de bases de dados geográficos utilizados

A.1. Introdução

Este apêndice apresenta os esquemas de dados geográficos utilizados para validar a MMultiSIG, a saber:

- Projeto Pró-Guaíba - Modelagem de dados para o Programa para o Desenvolvimento Racional, Recuperação e Gerenciamento Ambiental da Bacia Hidrográfica (Projeto Pró-Guaíba) realizada por GARRAFA (1998) para a Secretária de Coordenação e Planejamento do Estado do Rio Grande do Sul;
- Projeto SIG-RHidro - Modelagem de Dados para Projeto de Qualidade de Águas e Controle de Poluição Hídrica do Ministério e Planejamento e Orçamento, Projeto Estadual de Investimentos para o Rio de Janeiro/Bacia do Rio Paraíba do Sul, realizado no âmbito do Laboratório de Hidrologia da COPPE/UFRJ (LIRA, 1998);
- Projeto Pró-Paraíba - Tendo em vista testar a MMultiSIG, a modelagem do Projeto Pró Guaíba será aplicada a Bacia Hidrográfica do Rio Paraíba do Sul, criando-se um novo esquema; e
- Projeto SIG-Ambiental - Modelagem de dados realizada pela autora para o Estudo do Monitoramento ambiental do Estado do Rio de Janeiro. Esta modelagem foi construída com a finalidade de fornecer uma base de teste para metodologia.

A.1.1. Projeto Pró-Guaíba

Este projeto objetiva o planejamento e monitoramento estratégico do uso da água e outros recursos naturais da Bacia do Hidrográfica do Rio Guaíba, no Estado do Rio Grande do Sul.

Para ilustrar a metodologia, foi selecionada parte da modelagem original do Projeto Pró-Guaíba, realizada por GARRAFA (1998). A parte selecionada foi dividida em subsistemas para o gerenciamento do resíduo, meio ambiente, hidrografia, infra-

estrutura e divisão territorial, apresentados nas Figuras A.1 e A.2, usando a abordagem Entidade-Relacionamento, com uso da notação IDEF1X, a mesma utilizada por GARRAFA (1998). Esta notação é estendida por pictogramas para representar a geometria de ponto, linha e área das entidades.

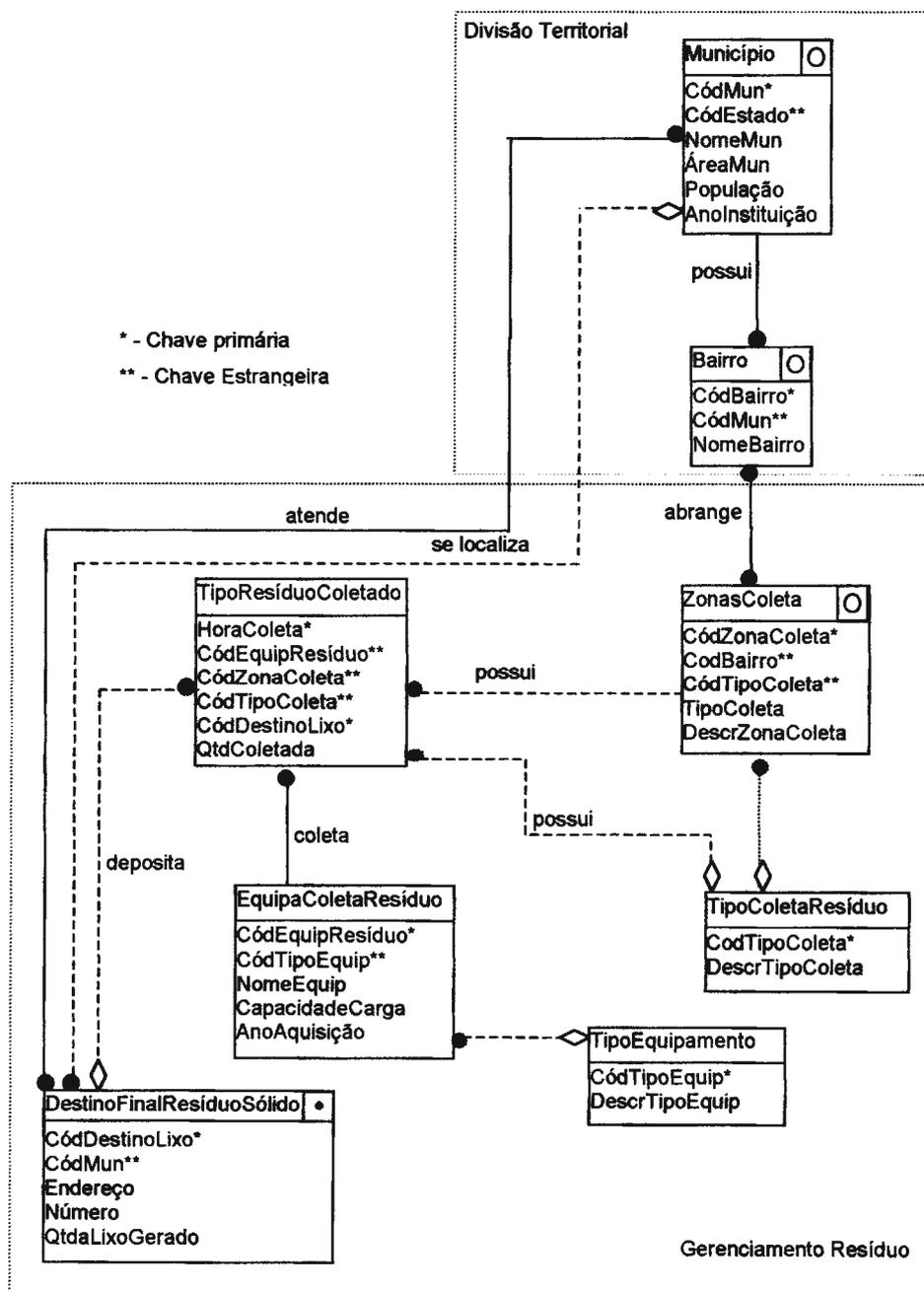


Figura A.1 - Esquema do projeto Pró-Guaíba: Abstração cartográfica em 1:10.000.

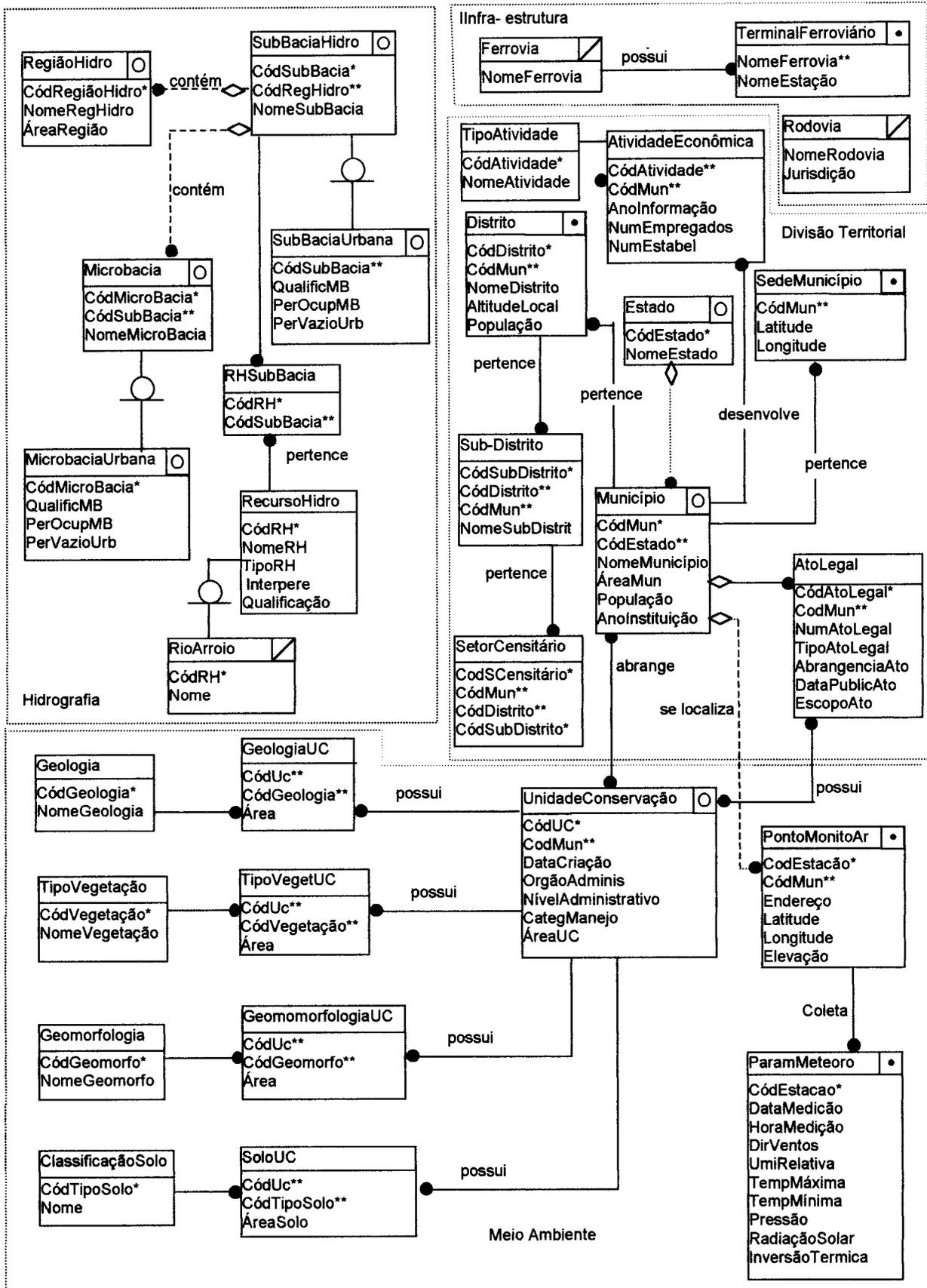


Figura A.2 -Esquema do projeto Pró-Guaíba: Abstração cartográfica em 1:250.000.

A base de dados, para efeito da metodologia, foi adequada ao ARC/Info utilizando as estruturas apresentadas nas Tabelas A.1 e A.2. As estruturas cartográficas para a construção da base de dados foram adquiridas por digitalização das cartas da Diretoria do Serviço Geográfico de Exército (DSG) em 1:50.000. Estas estruturas são generalizadas para 1:250.000 de modo a produzir os mapas temáticos. Para ilustrar os diferentes níveis de abstrações cartográficas usadas em uma base de dados de SIG, foi selecionada uma outra parte do esquema cujas estruturas cartográficas foram adquiridas através de restituição em 1:10.000. Este esquema utiliza para as estruturas cartográficas os *data* SAD-69 e Imbituba no sistema de coordenadas geográficas.

Tabela A.1 - Estruturas físicas usadas para as estruturas do esquema em 1:10.000.

Objetos Geográficos	Estrutura	Representação
Município	Coverage	Polygon
Bairro	Coverage	Polygon
ZonasColeta	Coverage	Polygon
TipoResíduoColetado	Tabela Info	
EquipaColetaResíduo	Tabela Info	
DestinoFinalResíduoSólido	Coverage	Point
TipoEquipamento	Tabela Info	
TipoColetaResíduo	Tabela Info	

Tabela A.2 - Estruturas físicas usadas para as estruturas do esquema em 1:250.000.

Objetos Geográficos	Estrutura	Representação
Estado	Coverage	Polygon
Município	Coverage	Polygon
SedeMunicípio	Coverage	Point
Distrito	Coverage	Point
SubDistrito	Tabela Info	
SetorCensitário	Tabela Info	
UnidadeConservação	Coverage	Polygon
GeologiaUC	Tabela Info	
TipoVegetUC	Tabela Info	
GeomorfologiaUC	Tabela Info	
SoloUC	Tabela Info	
Geologia	Tabela Info	
TipoVegetação	Tabela Info	
Geomorfologia	Tabela Info	
ClassificaçãoSolo	Tabela Info	
RegiãoHidro	Coverage	Polygon
SubBaciaHidro	Coverage	Polygon
SubBaciaUrbana	Coverage	Polygon
Microbacia	Coverage	Polygon
MicroBaciaUrbana	Coverage	Polygon
RHSubBacia	Tabela Info	
RecursoHidro	Tabela Info	
RioArroio	Coverage	Arc
AtoLegal	Tabela Info	
AtividadeEconomic	Tabela Info	

TipoAtividade	Tabela Info	
Ferrovia	Coverage	Polygon
TerminalFerroviário	Coverage	Point
Rodovia	Coverage	Arc
PontoMonitoAr	Coverage	Point
ParamMeteoro	Coverage	Point

A.1.2. Projeto SIG-RHidro

Este projeto objetiva oferecer suporte às atividades de planejamento, monitoramento e gerenciamento de informações de recursos hídricos para a área da Bacia do Rio Paraíba do Sul.

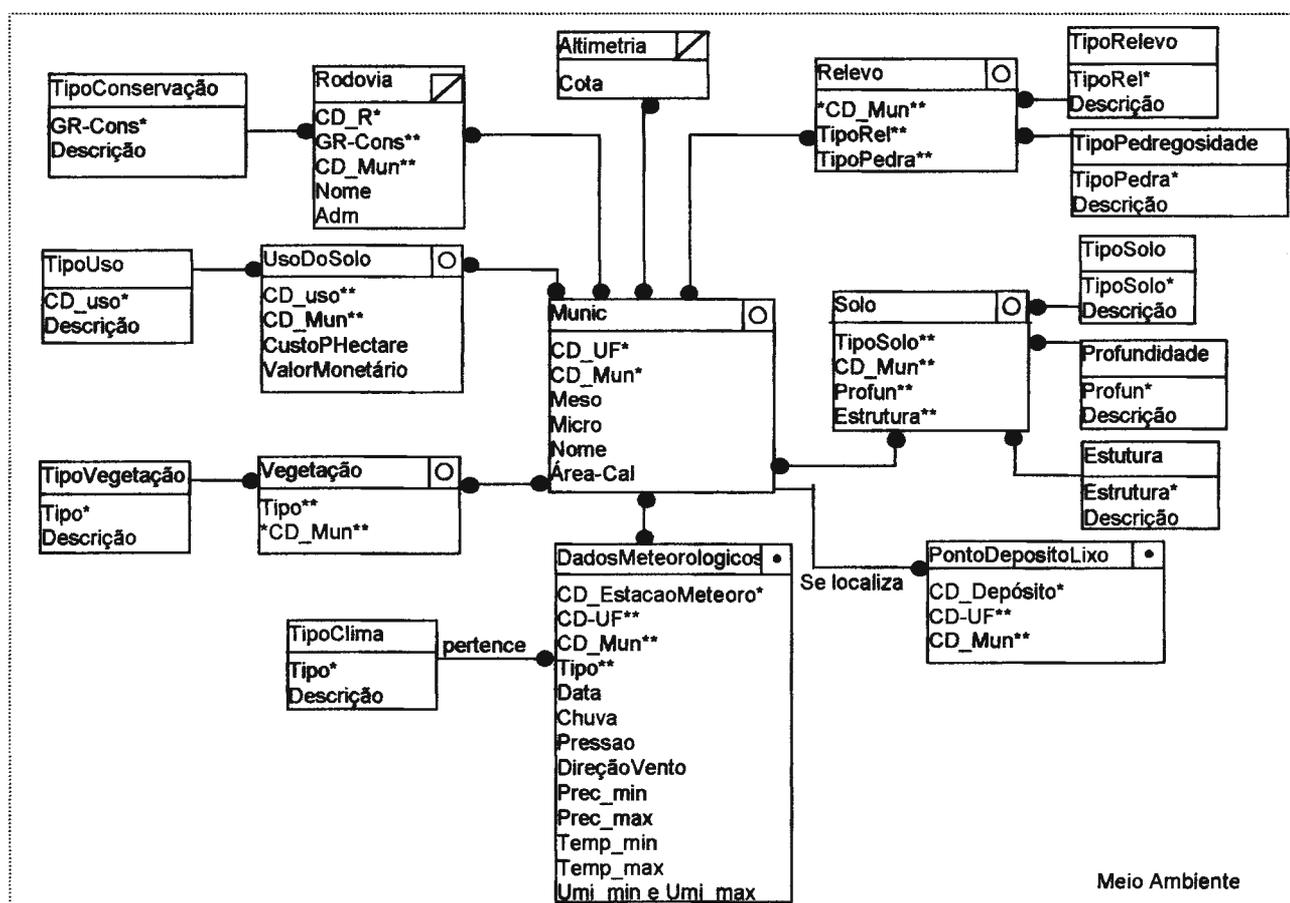


Figura A.3- Esquema do projeto SIG-RHidro: Abstração cartográfica em 1:100.000 (parte 1).

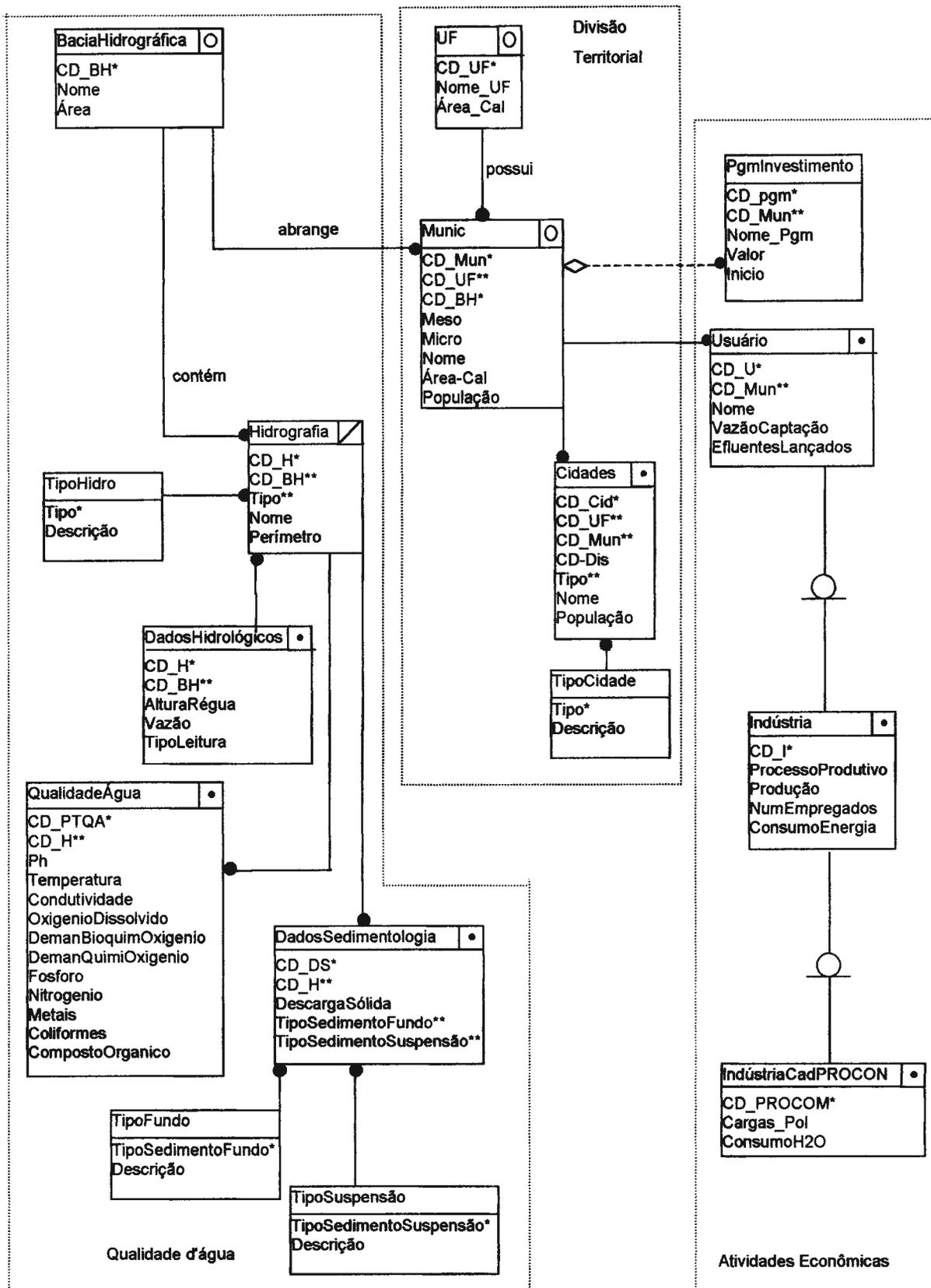


Figura A.4 - Esquema do projeto SIG-RHidro: Abstração cartográfica em 1:100.000 (parte 2).

Tendo em vista as características de implementação deste Projeto, foi criada a modelagem da base de dados já construída. As entidades foram agrupadas nos seguintes

subsistemas: meio ambiente, qualidade d'água, divisão territorial e atividades econômicas, conforme apresentado nas Figuras A.3 e A.4. Para facilitar, foi usada a mesma notação utilizada pelo Projeto Guaíba. Este projeto tem como unidade territorial de estudo os municípios que compõem a Bacia Hidrográfica do Rio Paraíba do Sul, no estado do Rio de Janeiro, São Paulo e Minas Gerais. A construção deste esquema utilizou as cartas do IBGE e DSG em escalas de 1:50.000 e 1:100.000. Os dados em 1:50.000 foram generalizados para 1:100.000, de forma a gerar os mapas temáticos resultantes de análises nesta escala.

Este esquema adota para as estruturas cartográficas o sistema de coordenadas geográficas na no sistema de projeção de Lambert com os *data* SAD-69 e Imbituba.

A base de dados do esquema RHidro, para efeito da metodologia, foi adequada ao ARC/Info utilizando as estruturas apresentadas na Tabela A.3, abaixo.

Tabela A.3 - Estruturas físicas usadas para as estruturas do esquema em 1:250.000.

Objetos Geográficos	Estrutura	Representação
UF	<i>Coverage</i>	<i>Polygon</i>
Munic	<i>Coverage</i>	<i>Polygon</i>
Cidades	<i>Coverage</i>	<i>Point</i>
TipoCidade	<i>Tabela Info</i>	
PgmInvestimento	<i>Tabela Info</i>	
Usuário	<i>Coverage</i>	<i>Point</i>
Indústria	<i>Coverage</i>	<i>Point</i>
IndústriaCadPROCON	<i>Coverage</i>	<i>Point</i>
BaciaHidrográfica	<i>Coverage</i>	<i>Polygon</i>
Hidrografia	<i>Coverage</i>	<i>Arc</i>
TipoHidro	<i>Tabela Info</i>	
Rodovia	<i>Coverage</i>	<i>Arc</i>
TipoConservação	<i>Tabela Info</i>	
UsoDoSolo	<i>Grid</i>	
TipoUso	<i>Tabela Info</i>	
Vegetação	<i>Grid</i>	
TipoVegetação	<i>Tabela Info</i>	
DadosMeteorológicos	<i>Coverage</i>	<i>Point</i>
TipoClima	<i>Tabela Info</i>	
PontoDepositoLixo	<i>Coverage</i>	<i>Point</i>
Solo	<i>Grid</i>	
TipoSolo	<i>Tabela Info</i>	
Profundidade	<i>Tabela Info</i>	
Estrutura	<i>Tabela Info</i>	
Relevo	<i>Grid</i>	
Altimetria	<i>Coverage</i>	<i>Arc</i>
TipoRelevo	<i>Tabela Info</i>	
TipoPedregosidade	<i>Tabela Info</i>	
QualidadeÁgua	<i>Coverage</i>	<i>Point</i>
DadosSedimentologia	<i>Coverage</i>	<i>Point</i>
TipoFundo	<i>Tabela Info</i>	
TipoSuspensão	<i>Tabela Info</i>	
DadosHidrológicos	<i>Coverage</i>	<i>Point</i>

A.1.3. Projeto Pró-Paraíba

Este esquema será idêntico ao esquema do Pró-Guaíba com abstração cartográfica em 1:250.000, considerando a sua aplicação à Bacia do Paraíba do Sul.

A.1.4. Projeto SIG-Ambiental

Este projeto foi desenvolvido com objetivo de fornecer suporte ao desenvolvimento da MMultiSIG. A modelagem é apresentada na Figura A.5 no modelo de dados do SPRING. A descrição dos objetos geográficos é complementada pela Tabela A.4.

Este projeto tem como unidade territorial de estudo os municípios do Estado do Rio de Janeiro. As estruturas cartográficas foram adquiridas na escala de 1:50.000. Os mapas temáticos são produzidos em 1:100.000. As estruturas cartográficas da base de dados esta em coordenadas geográficas, nos *data* horizontal e vertical SAD-69 e Imbituba, respectivamente.

Tabela A.4 - Definição dos objetos geográficos no SPRING.

Objetos Geográficos	Definições	Estrutura	Representação
Órgão Executor	Órgão executor do projeto	<i>ObjetoNãoEspacial</i>	
Projeto	Projeto	<i>ObjetoNãoEspacial</i>	
Área de Projeto	Área de Projeto	<i>Geo objeto</i>	<i>Poligono</i>
Município	Limites municipais	<i>Geo objeto</i>	<i>Poligono</i>
Cidades	Localização dos Centros Adm. das Cidades	<i>Geo objeto</i>	<i>ConjuntoDePontos2D</i>
Infra-Estrutura	Infra-estrutura de transporte	<i>Geo objeto</i>	<i>SubdivisãoPlanr</i>
Ferrovia	Ferrovia	<i>Geo-Objeto</i>	<i>Linha</i>
Rodovia	Rodovia	<i>Geo-Objeto</i>	<i>Linha</i>
Indústrias	Indústrias	<i>Geo objeto</i>	<i>ConjuntoPontos2D</i>
BaciasHidrográficas	Limite de bacia Hidrográfica	<i>Geo Objeto</i>	<i>Poligono</i>
Mapa bacias hidrográficas	Bacias hidrográficas com os rios, áreas de projeto e indústrias	<i>Mapa cadastral</i>	<i>SubdivisãoPlanar</i>
Divisão municipal	Municípios com os rios, a infra-estrutura e as cidades	<i>Mapa cadastral</i>	<i>SubdivisãoPlanar</i>
Rios	Hidrografia	<i>Geo objeto</i>	<i>Linha2D</i>
RedeDeDrenagem	Conjunto de rios da região	<i>Mapa cadastral.Red</i>	<i>SubdivisãoPlanar</i>
Vegetação	Tipos de vegetação	<i>Geo-campo Temático</i>	<i>ImagemTemática</i>
Solos	Solo	<i>Geo-campo Temático</i>	<i>ImagemTemática</i>
Floresta	Floresta	<i>Geo-campo Temático</i>	<i>ImagemTemática</i>
Relevo	Relevo	<i>Geo-campo Temático</i>	<i>ImagemTemática</i>

Clima	Clima	<i>Geo-campo Temático</i>	<i>ImagemTemática</i>
Uso da terra	Tipos de uso da terra	<i>Geo-campo Temático</i>	<i>ImagemTemática</i>
Declividade	Declividade	<i>Geo-campo Temático</i>	<i>ImagemTemática</i>
Pontos altimétricos	Pontos cotados da região	<i>Geo-campo Numérico</i>	<i>MapaPontos3d</i>
Altimetria	Curvas de nível	<i>Geo-campo Numérico</i>	<i>ConjuntoIsolinhas</i>
Imagem Landsat	Imagem Landsat	<i>Geo-campo DadoSensor</i>	<i>ImagemTonsDeCinza</i>

Apêndice B - Aplicação da \mathcal{M} MultiSIG

B.1. Introdução

Este apêndice apresenta a aplicação da \mathcal{M} MultiSIG para formar uma federação com os esquemas apresentados no Apêndice A.

As seções a seguir apresentam cada etapa da metodologia proposta, a saber: pré-integração, análise dos esquemas, conformação, integração, definição do esquema externo e validação.

B.2. Pré-Integração

A \mathcal{M} MultiSIG propõe que esta etapa seja dividida em duas atividades, a saber: descrição dos atributos do esquema e descrição dos esquemas de exportação.

B.2.1. Descrição dos atributos dos esquemas

Esta atividade consiste em criar o *GenericGeographicObject* (GGO) para cada esquema, de acordo com as abstrações cartográficas utilizadas, e instanciar os atributos do esquema. Para facilitar esta atividade será realizada para cada projeto.

Nas descrições dos projetos abaixo serão instanciadas os metadados relativos a *SpatialReferencingMM*, *TemporalReferencing*, *GeneralLocationMM* e *SourceMM*. Cabe ressaltar que a \mathcal{M} MultiSIG considera ainda os metadados descritos pelas classes *Quality*, *Lineage*, *ProductDescription* e *UpdateOperation*, que não serão apresentadas devido a extensão do trabalho.

B.2.1.1. Projeto Pró- Guaíba

Tendo em vista que as partes selecionadas do esquema se apresentam em duas

escalas, 1:10.000 e 1:250.000, este esquema será cadastrado na federação de dados proposta pelo Sistema MultiSIG, com duas abstrações cartográficas, a primeira na escala de 1:10.000 e a segunda na escala de 1:250.000.

Este esquema adota o sistema de coordenadas geográficas usando os *data* horizontal e vertical SAD-69 e Imbituba, respectivamente.

```
<Superclass::GeographicObject
  subclass:      GGO::Guaíba.abs1
  subclassing:   "Definidas a seguir"
  attributes:
  metadata: Metadata
  defaults:     position.geometry.qualifier: definite
                position.spatialReferencing: nil
                position.geometry*Coordinate:^Coord2D|^Coord3D
  comments:"Superclasse abstrata que armazena os metadados genéricos para todas as classes
definidas como subclasse de GGO::Guaíba.abs1. As classes em 1:10.000 são traduzidas do Esquema do
Projeto Pró-Guaíba como subclasses de GGO::Guaíba.abs1">
```

```
<Superclass::GeographicObject
  subclass:      GGO::Guaíba.abs2
  subclassing:   "Definidas a seguir"
  attributes:
  metadata:     Metadata
  defaults:     position.geometry.qualifier: definite
                position.spatialReferencing: nil
                position.geometry*Coordinate:^Coord2D|^Coord3D
  comments:"Superclasse abstrata que armazena os metadados genéricos para todas as
classes definidas como subclasse de GGO::Guaíba.abs2. As classes em 1:250.000 traduzidas do Esquema
do Projeto Pró-Guaíba serão definidas como subclasses de GGO::Guaíba.abs2 ">
```

B.2.1.1.A. Instância de Metadados para GGO::Guaíba.abs1

```
GGO::Guaíba.abs1(
  metadata: Metadata (SpatialReferencingMM (
    [coordSystem]: GeodesicCoordinateSystem(
      [coordinateUnits]: {degrees degrees meters}
      [cts]: ConventionalTerrestrialSystem(
        cts: CTSType(
          systemType: UGGI
          systemYear: 67)
      [horizontalReference]: GeodeticHorizontalReference (
        datumAdjustmentSystem: Sistema Geodésico Brasileiro -SAD 69
        [ellipsoid]: GeodeticEllipsoid(
          ellipsoidType: UIGG
          [ellipsoidYear]: 1967
          [semiMajorAxis]: 6378160,0
          [flattening]: 0,00335)
        [verticalReference]: GeodeticVerticalReference(
          [datumAdjustmentSystem]: Sistema Geodésico Brasileiro - Imbituba
          [surface]: seaMeanLevel))
      [hydrographic]: HydrographicReference(
        [tidal]: TidalInfo(
          [reference]: Imbituba
          [tidalHeight]: 0.0
          [remarks]: Datum Vertical Nacional)
```

baseInverseScale:10.000)

TemporalReferencing (
 [type]: local)

GeneralLocationMM(
 [range]: *BoudingBox*(
 minPoint: *Point* (
 [coords]: *Coord2D*(
 c1: XMIN:
 c2: YMIN:))
 maxPoint: *Point* (
 [coords]: *Coord2D*(
 c1: XMAX:
 c2: YMAX:))
 [tiles]: List {*Tile*¹(
 [mapSheetNumber]:
 [mapNumberingSystem]:
 [publicationDate]: *Date* (
 [year]:
 [month]:
 [day]:)
 [mapSeriesName]:
 [mapSheetName]:
 [inverseScale]: 50.000)
 [mapSheetNumber]:
 [mapNumberingSystem]:
 [publicationDate]: *Date* (
 [year]:
 [month]:
 [day]:)
 [mapSeriesName]:
 [mapSheetName]:
 [inverseScale]: 50.000)
 geographicRegion: Vale do Rio Guaíba abrangendo os municípios da Bacia do Rio Guaíba)

SourceMM(
 [terminus]: Set {*Terminus* (
 [terminusType]: producer, consumer, distributor
 [terminusInfo]: *Agency* (
 [contactPerson]: Iria M Garrafa
 [agencyName]: UFRGS
 [address]:
 [telephoneNumber]:
 [email]:
 [country]: Brasil))
 [database]: *DbInfo* (
 [dbmsIdentifier]: ARC/Info
 queryLanguage: AML
 [dbIdentifier]: /
 operationSystem: UNIX
 hardwareType: SUN
 platformName: moscou.cos.ufrj.br)
 [project]: Set {*Project* (
 [projectIdentifier]: Pró-Guaíba
 projectObjective : Planejamento e Monitoramento do Uso D'água
 applicationArea: Recursos naturais

¹ Lista dos cartas que geraram o mapa digital

projectSubsystem: Meio Ambiente, Gerenciamento de Resíduos e Cartografia)
[security]: Security (
[classification]: unclassified)
[remarks]: String))

B.2.1.1.B. Instância de Metadados para GGO::Guaíba.abs2

GGO::Guaíba.abs2(
metadata: Metadata (SpatialReferencingMM (
[coordSystem]: GeodesicCoordinateSystem(
[coordinateUnits]: {degrees degrees meters}
[cts]: ConventionalTerrestrialSystem(
cts: CTSType(
systemType: UGGI
systemYear: 67)
[horizontalReference]: GeodeticHorizontalReference (
datumAdjustmentSystem: Sistema Geodésico Brasileiro - SAD-69
[ellipsoid]: GeodeticEllipsoid(
ellipsoidType: UIGG
[ellipsoidYear]: 1967
[semiMajorAxis]: 6378160,0
[flattening]: 0,00335)
[verticalReference]: GeodeticVerticalReference(
[datumAdjustmentSystem]: Sistema Geodésico Brasileiro - Imbituba
[surface]: seaMeanLevel))
[hydrographic]: HydrographicReference(
[tidal]: TidalInfo(
[reference]: Imbituba
[tidalHeight]: 0.0
[remarks]: Datum Vertical Nacional)
baseInverseScale:250.000)

TemporalReferencing (
[type]: local)

GeneralLocationMM(
[range]: BoundingBox(
minPoint: Point (
[coords]: Coord2D(
c1: XMIN:
c2: YMIN:))
maxPoint: Point (
[coords]: Coord2D (
c1: XMAX:
c2: YMAX:I))
[tiles]: List {Tile ²(
[mapSheetNumber]:
[mapNumberingSystem]:
[publicationDate]: Date (
[year]:
[month]:
[day]:)
[mapSeriesName]:
[mapSheetName]:
[inverseScale]: 50.000)

² Lista dos cartas que geraram o mapa digital

```

[mapSheetNumber]:
[mapNumberingSystem]:
[publicationDate]: Date (
  [year]:
  [month]:
  [day]: )
[mapSeriesName]:
[mapSheetName]:
[inverseScale]: 50.000)

```

map: Graphical Region: Municípios da Bacia do Rio Guaíba, no estado do Rio Grande do Sul)

Source: AML

```

[terminus]: Set { Terminus (
  [terminusType]: producer, consumer, distributor
  [terminusInfo]: Agency (
    [contactPerson]: Iria M Garrafa
    [agencyName]: UFRGS
    [address]:
    [telephoneNumber]:
    [email]:
    [country]: Brasil)
[database]: DbInfo (
  [dbmsIdentifier]: ARC/Info
  [dbIdentifier]: AML
  [dbIdentifier]: /
  [operatingSystem]: UNIX
  [hardwareType]: SUN
  [platformName]: moscou.cos.ufrj.br)
[project]: Set { Project (
  [projectIdentifier]: Pró-Guaíba
  [projectObjective]: Planejamento e Monitoramento do Uso D'água
  [applicationArea]: Recursos naturais
  [projectSubsystem]: Meio Ambiente, Gerenciamento de Resíduos e Cartografia)
[security]: Security (
  [classification]: unclassified)
[remarks]: String)

```

B.2.1.2. Projeto SIG-RHidro

Este esquema será cadastrado na federação de dados proposta pelo Sistema MultiSIG, com uma abstração cartográfica, na escala de 1:250.000. Este esquema adota para sua base cartográfica o sistema de coordenadas da projeção de Lambert, para qual foi usado os *data* horizontal e vertical, SAD-69 e Imbituba, respectivamente.

```

<Superclass::GeographicObject
  subclass: GGO::RHidro.abs1
  subclassing: "Definidas a seguir"
  attributes:
  metadata: Metadata
  defaults:
    position.geometry.qualifier: definite
    position.spatialReferencing: nil
    position.geometry*Coordinate: ^Coord2D | ^Coord3D
  comments: "Superclasse abstrata que armazena os metadados genéricos para todas as classes
  definidas como subclasse de GGO::RHidro.abs1. Todas as classes traduzidas do Esquema do Projeto
  SIG-RHidro serão definidas como subclasses de GGO::RHidro.abs1 ">

```

B.2.1.2.A. Instância de Metadados para GGO::RHidro.abs1

```
GGO::RHidro.abs1(  
  metadata: Metadata (SpatialReferencingMM(  
    [coordSystem]: Projection(  
      [coordinateUnits]: {meters meters meters}  
      [projectionName]: Projeção de Lambert  
      [projectionParameters]: List { ProjectionParameters(  
        [firstStandartParallel]:  
        [secondStandartParallel]:  
        [centralMeridian]:  
        [latitudeProjectionOrigin]:  
        [falseEastingMeters]:  
        [falseNorthingMeters]:)  
      [horizontalReference]: GeodeticHorizontalReference (  
        datumAdjustmentSystem: Sistema Geodésico Brasileiro - SAD 69  
        [ellipsoid]: GeodeticEllipsoid(  
          ellipsoidType: UIGG  
          [ellipsoidYear]: 1967  
          [semiMajorAxis]: 6378160,0  
          [flattening]: 0,00335)  
        [verticalReference]: GeodeticVerticalReference(  
          [datumAdjustmentSystem]: Sistema Geodésico Brasileiro - Imbituba  
          [surface]: seaMeanLevel))  
      [hydrographic]: HydrographicReference(  
        [tidal]: TidalInfo(  
          [reference]: Imbituba  
          [tidalHeight]: 0.0  
          [remarks]: Datum Vertical Nacional)  
        baseInverseScale: 100.000)  
    )  
  )  
  TemporalReferencing (  
    [type]: local)  
  GeneralLocationMM(  
    [range]: BoudingBox(  
      minPoint: Point (  
        [coords]: Coord2D(  
          c1: XMIN:  
          c2: YMIN:))  
      maxPoint: Point (  
        [coords]: Coord2D(  
          c1: XMAX:  
          c2: YMAX:))  
    [tiles]: List { Tile3(  
      [mapSheetNumber]:  
      [mapNumberingSystem]:  
      [publicationDate]: Date (  
        [year]:  
        [month]:  
        [day]: )  
      [mapSeriesName]:  
      [mapSheetName]:  
      [inverseScale]: 50.000)  
      [mapSheetNumber]:  
      [mapNumberingSystem]:  
      [publicationDate]: Date (  
      )  
    )  
  )  
)
```

³ Lista dos cartas que geraram o mapa digital

```

    [year]:
    [month]:
    [day]: )
    [mapSeriesName]:
    [mapSheetName]:
    [inverseScale]: 50.000)}
    [coverage]: Vale do Rio Paraíba do Sul abrangendo os municípios dos estados do Rio de
    Janeiro, São Paulo, Minas Gerais, Espírito Santo)
SourceAM(
  [terminus]: Set {Terminus (
    [terminusType]: producer, consumer, distributor
    [terminusInfo]: Agency(
      [contactPerson]: Flavio Lira
      [agencyName]: COPPE/UFRJ
      [address]:
      [telephoneNumber]:
      [email]:
      [country]: Brasil))
  [database]: DbInfo (
    [dbmsIdentifier]: ARC/Info
    [queryLanguage]: AML
    [dbIdentifier]: /
    [operatingSystem]: UNIX
    [hardwareType]: SUN
    [platformName]: moscou.cos.ufrj.br)
  [project]: Set {Project (
    [projectIdentifier]: RHidro
    [projectObjective]: Planejamento e Gerenciamento dos Recursos Hídricos
    [applicationArea]: Recursos Naturais
    [projectSubsystem]: Meio Ambiente, Qualidade D'Água)
  [security]: Security(
    [classification]: unclassified)
  [remarks]: String))

```

B.2.1.3. Projeto SIG-Ambiental

Este esquema utiliza para a representação gráfica a escala de 1:100.000, no sistema de coordenadas geográficas usando os *data* horizontal e vertical, SAD-69 e Imbituba, respectivamente.

```

<Superclass::GeographicObject
  subclass:      GGO::Ambiental.abs1
  subclassing:   "Definidas a seguir"
  attributes:
  metadata: Metadata
  defaults:      position.geometry.qualifier: definite
                  position.spatialReferencing: nil
                  position.geometry*Coordinate:^Coord2D|^Coord3D
  comments:"Superclasse abstrata que armazena os metadados genéricos para todas as classes
  definidas como subclasse de GGO::Ambiental.abs1. Todas as classes traduzidas do Esquema do Projeto
  SIG-Ambiental serão definidas como subclasses de GGO::Ambiental.abs1 ">

```

B.2.1.3.A. Instância de Metadados para GGO::Ambiental.abs1

```
GGO::Ambiental.abs1(
```

metadata: Metadata (SpatialReferencingMM (
[coordSystem]: GeodesicCoordinateSystem(
[coordinateUnits]: {degrees degrees meters}
[cts]: ConventionalTerrestrialSystem(
cts: CTSType(
systemType: UGGI
systemYear:67)
[horizontalReference]: GeodeticHorizontalReference (
datumAdjustmentSystem: Sistema Geodésico Brasileiro
[ellipsoid]: GeodeticEllipsoid(
ellipsoidType: UIGG
[ellipsoidYear]: 1967
[semiMajorAxis]: 6378160,0
[flattening]: 0,00335)
[verticalReference]: GeodeticVerticalReference(
[datumAdjustmentSystem]: Sistema Geodésico Brasileiro
[surface]: seaMeanLevel))
[hydrographic]: HydrographicReference(
[tidal]: TidalInfo(
[reference]: Imbituba
[tidalHeight]: 0.0
[remarks]: Datum Vertical Nacional)
baseInverseScale:100.000)

TemporalReferencing (
[type]: local)

!GeneralLocationMM(
[range]: BoundingBox(
minPoint: Point (
[coords]: Coord2D(
c1: XMIN:
c2: YMIN:))
maxPoint: Point (
[coords]: Coord2D (
c1: XMAX:
c2: YMAX:))
[tiles]: List { Tile ⁴(
[mapSheetNumber]:
[mapNumberingSystem]:
[publicationDate]: Date (
[year]:
[month]:
[day]:)
[mapSeriesName]:
[mapSheetName]:
[inverseScale]: 50.000)
[mapSheetNumber]:
[mapNumberingSystem]:
[publicationDate]: Date (
[year]:
[month]:
[day]:)
[mapSeriesName]:
[mapSheetName]:
[inverseScale]: 50.000)}
geographicRegion: Municípios do Estado do Rio de Janeiro)

⁴ Lista dos cartas que geraram o mapa digital

```

SourceMAA(
  [terminus]: Set {Terminus (
    [terminusType]: producer, consumer, distributor
    [terminusInfo]: Agency(
      [contactPerson]: Julia Strauch
      [agencyName]: COPPE/UFRJ
      [address]:
      [telephoneNumber]:
      [email]:
      [country]: Brasil))
  [database]: DbInfo (
    [dbmsIdentifier]: ARC/Info
    [dbIdentifier]: AML
    [dbIdentifier]: /
    [operationSystem]: UNIX
    [hardwareType]: SUN
    [platformName]: moscou.cos.ufrj.br)
  [project]: Set {Project (
    [projectIdentifier]: Ambiental
    [projectObjective]: Estudo e Monitoramento do Meio Ambiente
    [applicationArea]: Meio Ambiente
    [projectSubsArea]:
  )
  [security]: Security (
    [classification]: unclassified)
  [remarks]: String))

```

B.2.1.4. Projeto Pró-Paraíba

Tendo em vista este esquema ser igual ao esquema Pró-Guaíba, ele será cadastrado no Sistema MultiSIG com abstrações cartográficas nas escalas de 1:10.000 e 1:250.000. Assim, este esquema adota para sua representação cartográfica, o sistema de coordenadas geográficas usando os *data* horizontal e vertical SAD-69 e Imbituba, respectivamente.

```

<Superclass::GeographicObject
subclass:      GGO::Paraíba.abs1
subclassing:   "Definidas a seguir"
attributes:
metadata:     Metadata
defaults:     position.geometry.qualifier: definite
              position.spatialReferencing: nil
              position.geometry*Coordinate:^Coord2D|^Coord3D
comments:     "Superclasse abstrata que armazena os metadados genéricos para todas as classes
              definidas como subclasse de GGO::Paraíba.abs1. As classes em 1:10.000 são traduzidas do Esquema do
              Projeto Pró-Guaíba como subclasses de GGO::Paraíba.abs1">

```

```

<Superclass::GeographicObject
subclass:      GGO::Paraíba.abs2
subclassing:   "Definidas a seguir"
attributes:
metadata:     Metadata
defaults:     position.geometry.qualifier: definite
              position.spatialReferencing: nil
              position.geometry*Coordinate:^Coord2D|^Coord3D
comments:     "Superclasse abstrata que armazena os metadados genéricos para todas as classes

```

definidas como subclasse de GGO::Paraíba.abs2. As classes em 1:250.000 são traduzidas do Esquema do Projeto PróGuiaba como subclasses de GGO::Paraíba.abs2 ">

B.2.1.4.A. Instância de Metadados para GGO::Paraíba.abs1

GGO::Paraíba.abs1(

```
metadata: Metadata (SpatialReferencingMM (
  [coordSystem]: GeodesicCoordinateSystem(
    [coordinateUnits]: {degrees degrees meters}
    [cts]: ConventionalTerrestrialSystem(
      cts: CTSType(
        systemType: UGGI
        systemYear: 67)
    [horizontalReference]: GeodeticHorizontalReference (
      datumAdjustmentSystem: Sistema Geodésico Brasileiro -SAD 69
      [ellipsoid]: GeodeticEllipsoid(
        ellipsoidType: UIGG
        [ellipsoidYear]: 1967
        [semiMajorAxis]: 6378160,0
        [flattening]: 0,00335)
      [verticalReference]: GeodeticVerticalReference(
        [datumAdjustmentSystem]: Sistema Geodésico Brasileiro - Imbituba
        [surface]: seaMeanLevel))
    [hydrographic]: HydrographicReference(
      [tidal]: TidalInfo(
        [reference]: Imbituba
        [tidalHeight]: 0.0
        [remarks]: Datum Vertical Nacional)
      baseInverseScale:10.000)
TemporalReferencing (
  [type]: local)
GeneralLocationMM(
  [range]: BoudingBox(
    minPoint: Point (
      [coords]: Coord2D(
        c1: XMIN:
        c2: YMIN:))
    maxPoint: Point (
      [coords]: Coord2D (
        c1: XMAX:
        c2: YMAX:))
  [tiles]: List { Tile 5(
    [mapSheetNumber]:
    [mapNumberingSystem]:
    [publicationDate]: Date (
      [year]:
      [month]:
      [day]: )
    [mapSeriesName]:
    [mapSheetName]:
    [inverseScale]: 50.000)
    [mapSheetNumber]:
    [mapNumberingSystem]:
```

⁵ Lista dos cartas que geraram o mapa digital

[reference]: Imbituba
[tidalHeight]: 0.0
[remarks]: Datum Vertical Nacional)
baseInverseScale:10.000)

TemporalReferencing (
[type]: local)

GeneralLocationMM(

[range]: *BoudingBox*(
minPoint: *Point* (
[coords]: *Coord2D*(
c1: XMIN:
c2: YMIN:))
maxPoint: *Point* (
[coords]: *Coord2D*(
c1: XMAX:
c2: YMAX:))
[tiles]: List {*Tile* ⁶(
[mapSheetNumber]:
[mapNumberingSystem]:
[publicationDate]: *Date* (
[year]:
[month]:
[day]:)
[mapSeriesName]:
[mapSheetName]:
[inverseScale]: 50.000)
[mapSheetNumber]:
[mapNumberingSystem]:
[publicationDate]: *Date* (
[year]:
[month]:
[day]:)
[mapSeriesName]:
[mapSheetName]:
[inverseScale]: 50.000)}

geographicRegion Vale do Rio Paraíba do Sul abrangendo os municípios dos estados do Rio de Janeiro, São Paulo, Minas Gerais, Espírito Santo)

SourceMM(

[terminus]: Set { *Terminus* (
[terminusType]: producer, consumer, distributor
[terminusInfo]: *Agency* (
[contactPerson]: Julia Strauch e Iria Garrafa
[agencyName]: UFRGS, COPPE/UFRJ
[address]:
[telephoneNumber]:
[email]:
[country]: Brasil))
[database]: *DbInfo* (
[dbmsIdentifier]: ARC/Info
queryLanguage: AML
[dbIdentifier]: /
operationSystem: UNIX
hardwareType: SUN
platformName: moscou.cos.ufrj.br)

⁶ Lista dos cartas que geraram o mapa digital

[project]: Set {Project (
 [projectIdentifier]: Pró-Paraíba
 projectObjective: Planejamento e Gerenciamento dos recursos hídricos
 projectName: Recursos naturais
 projectSubSystem)
 [security]: Security (
 [classification]: unclassified)
 [remarks]: String))

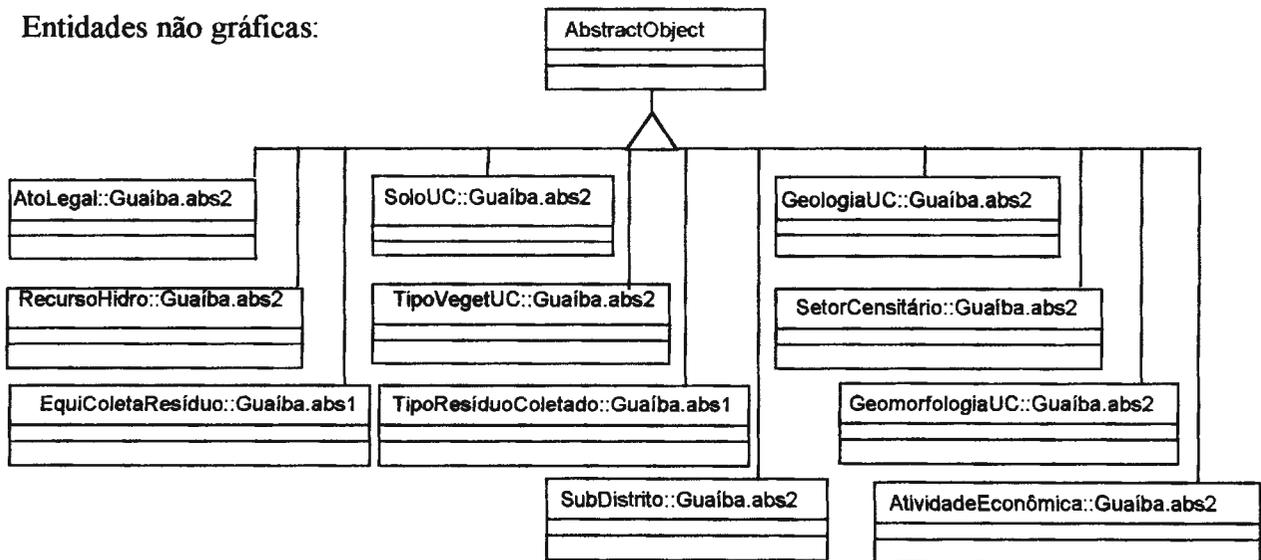
B.2.2. Descrição dos esquemas de exportação

Esta atividade consiste em traduzir os esquemas expressos nos seus esquemas locais para o SAIF, usando as regras de mapeamento proposta pela MultiSIG.

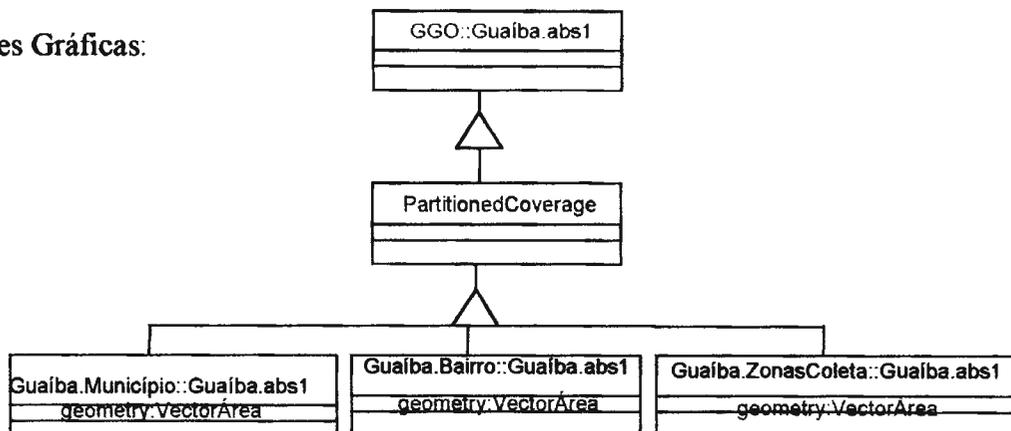
As classes serão apresentadas abaixo para cada projeto usando a notação *Object Modelling Technique*, usada pelo SAIF.

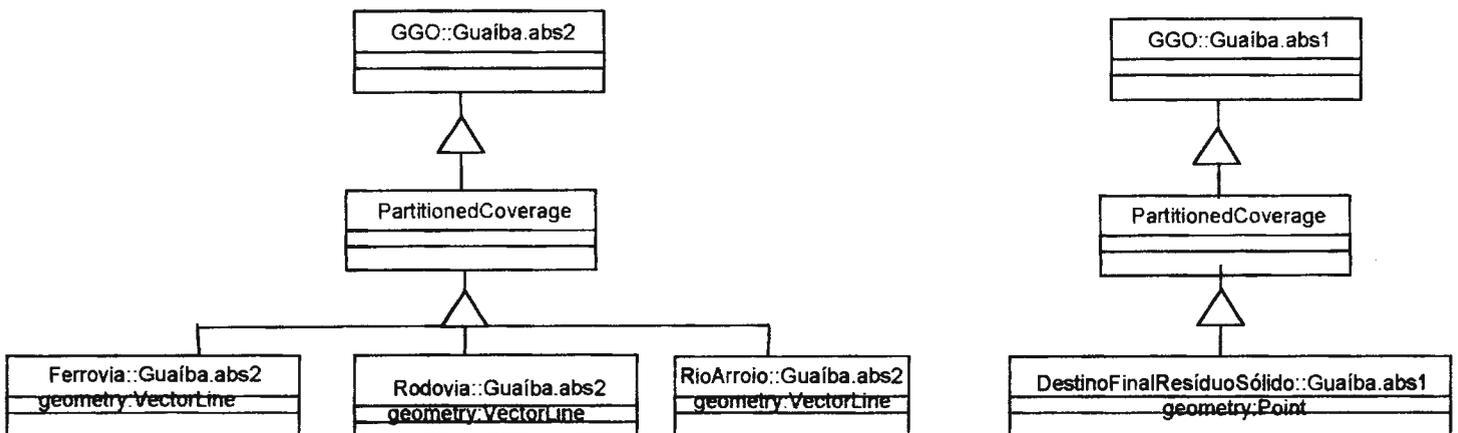
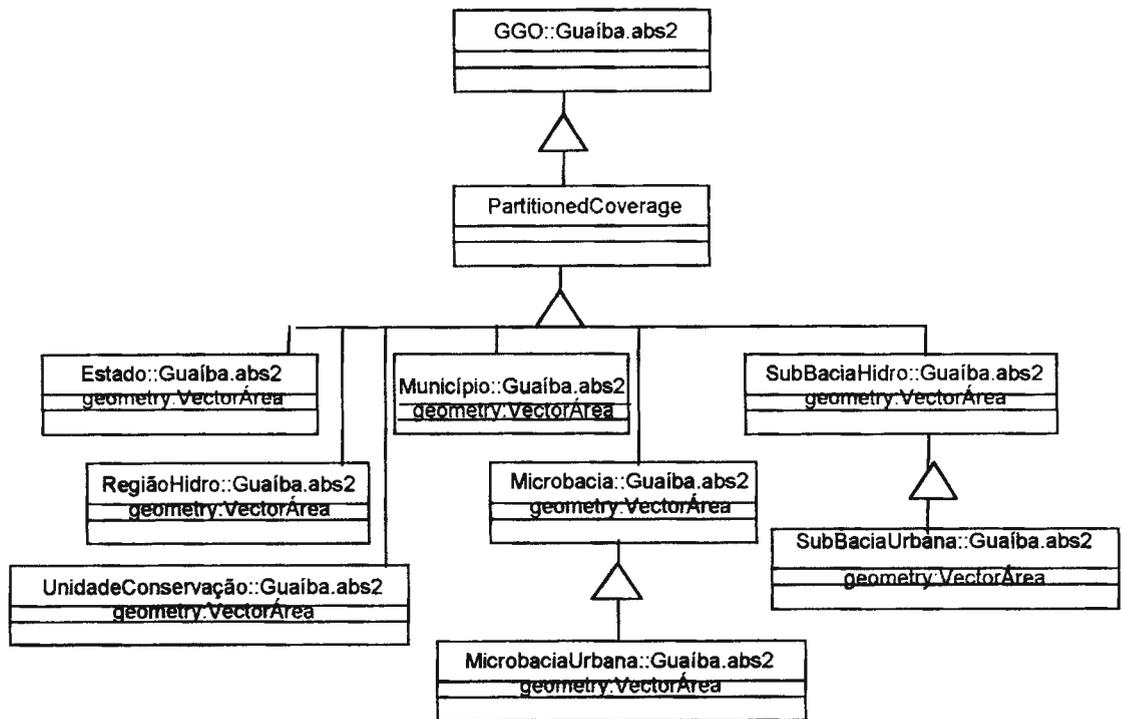
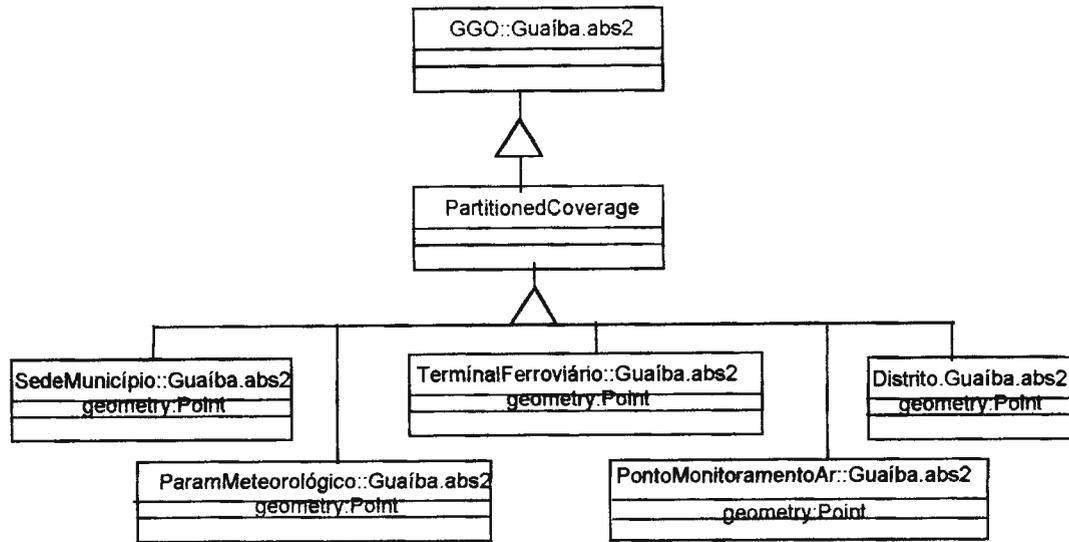
B.2.2.1. Projeto Pró Guaíba

Entidades não gráficas:

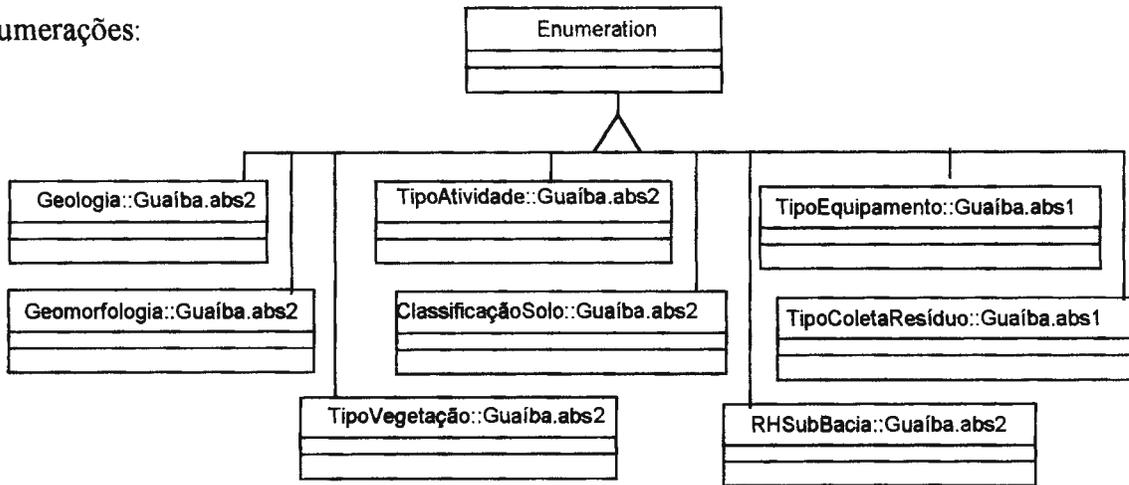


Entidades Gráficas:



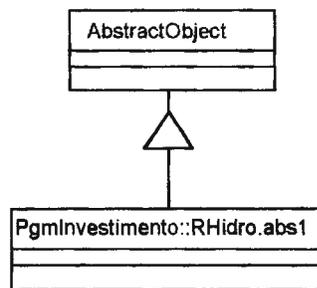


Enumerações:

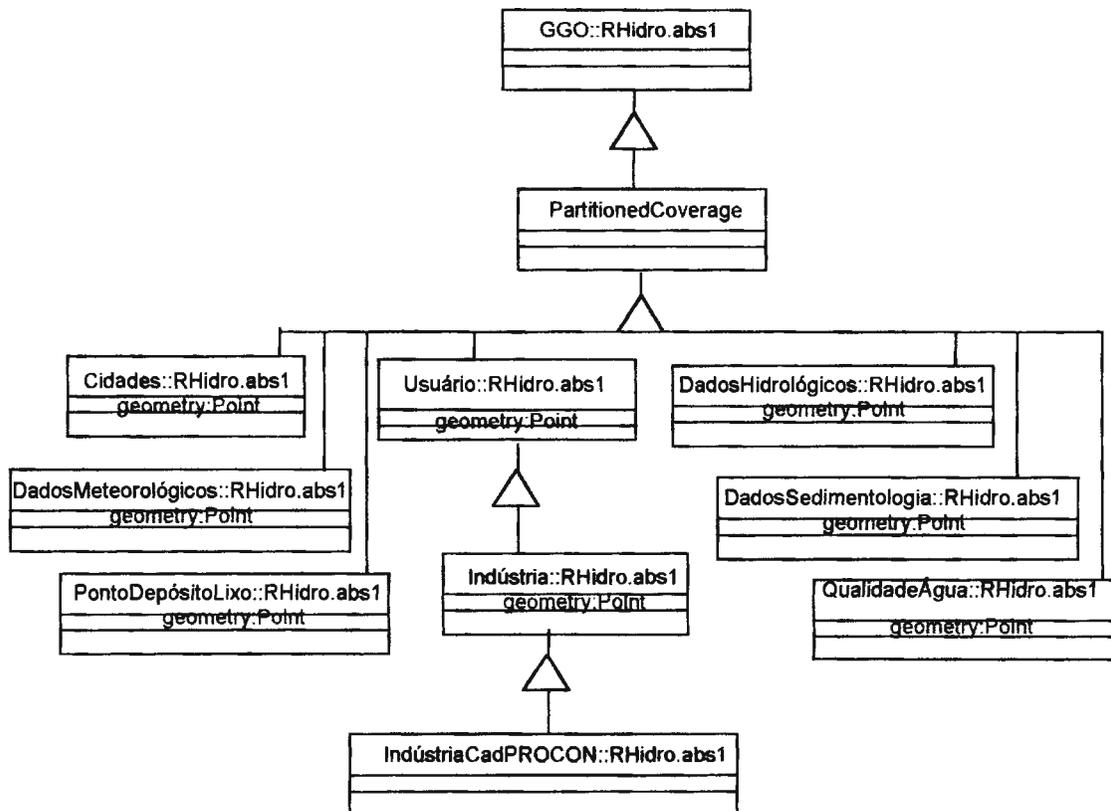


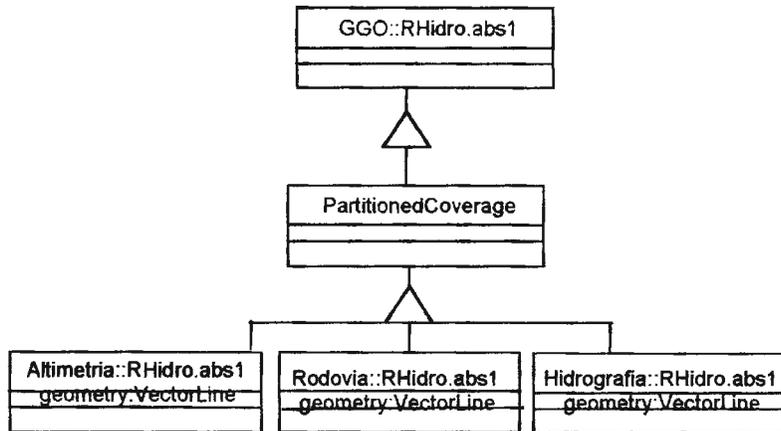
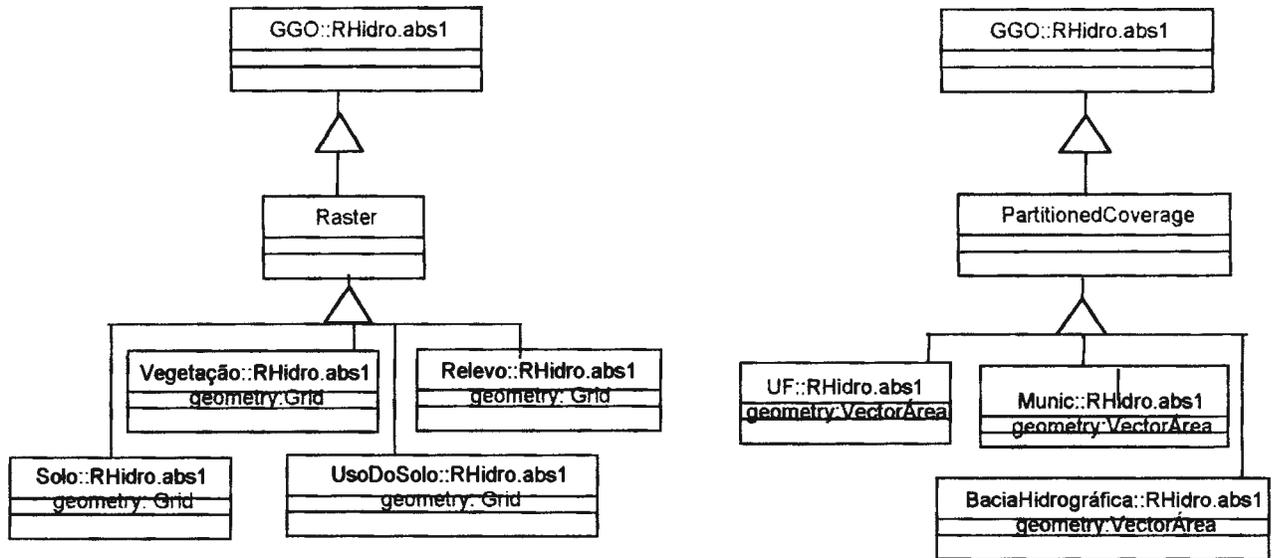
B.2.2.2. Projeto SIG-Rhidro

Entidades não gráficas:

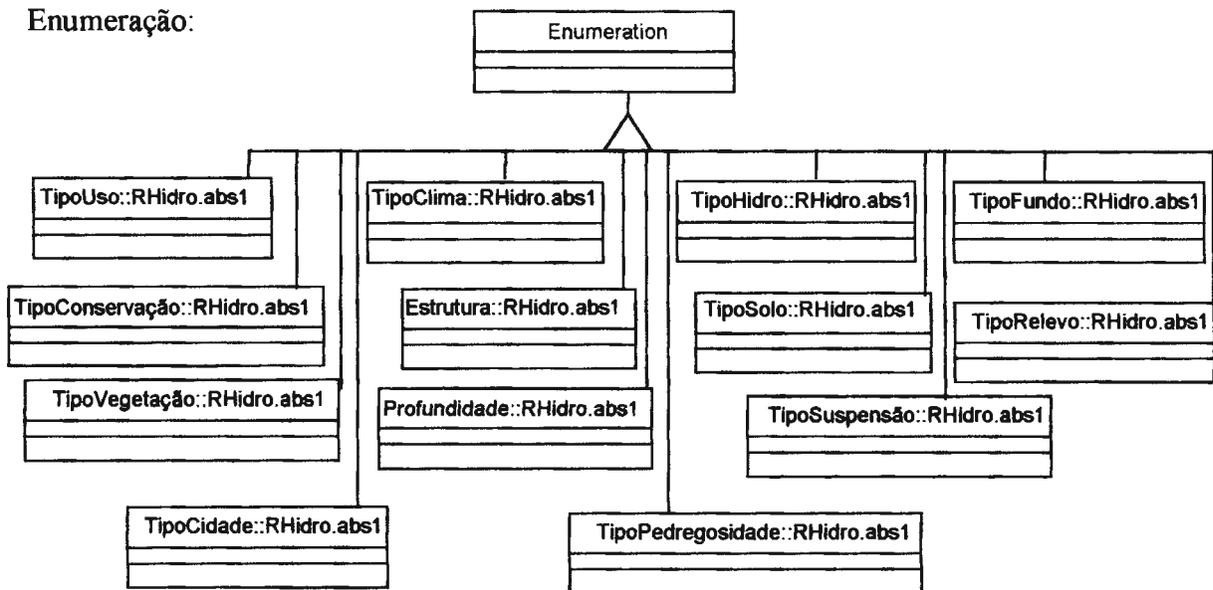


Entidades gráficas:



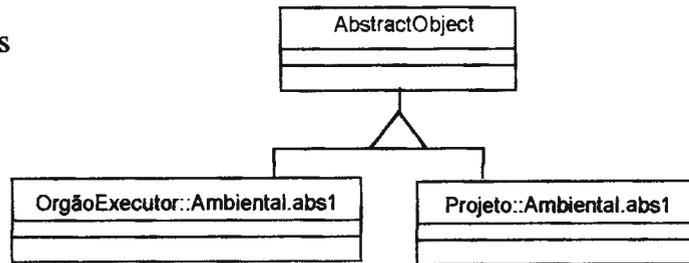


Enumeração:

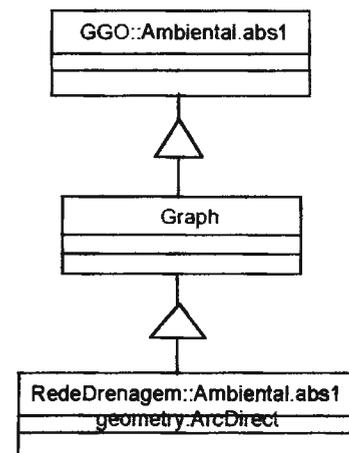
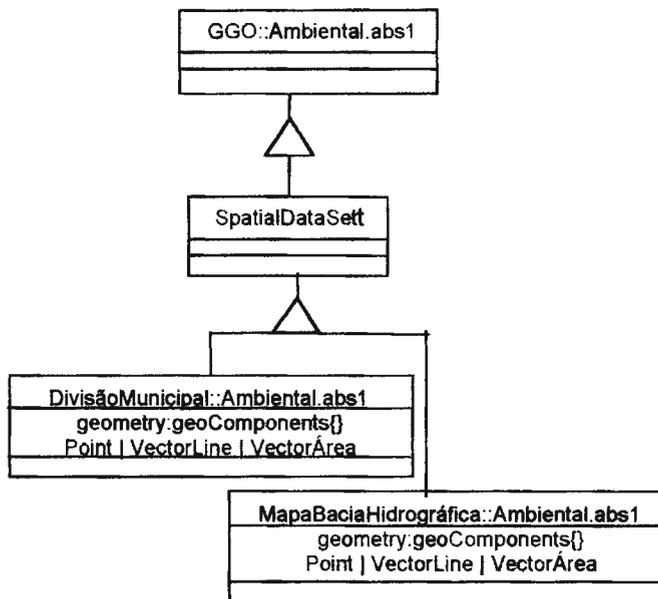
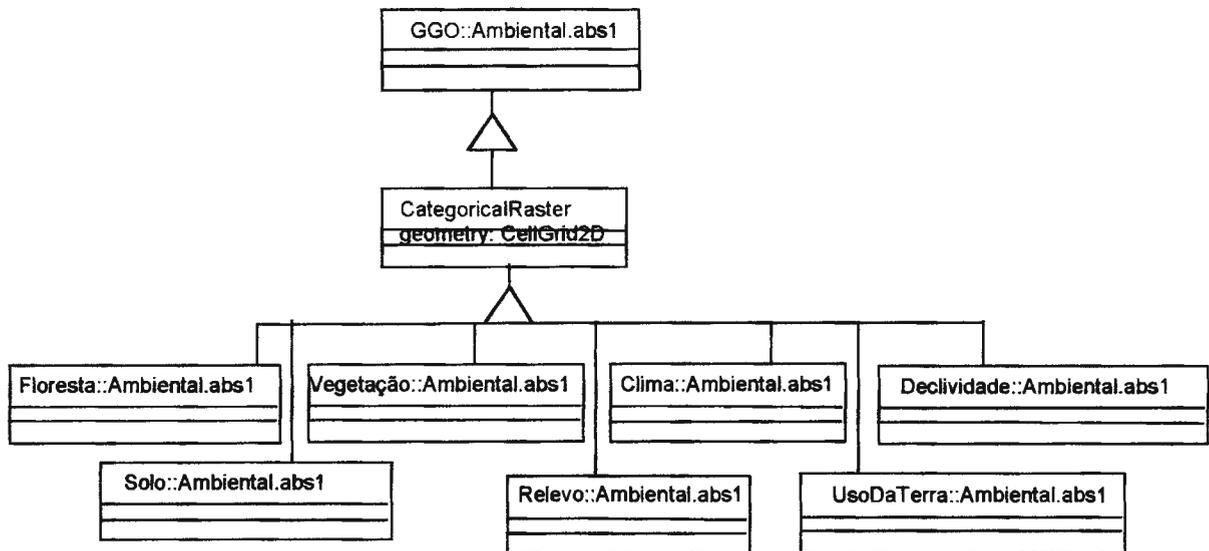


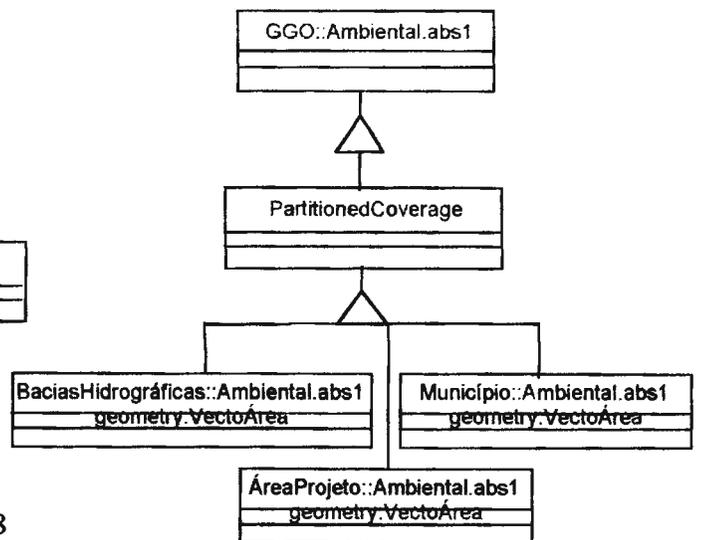
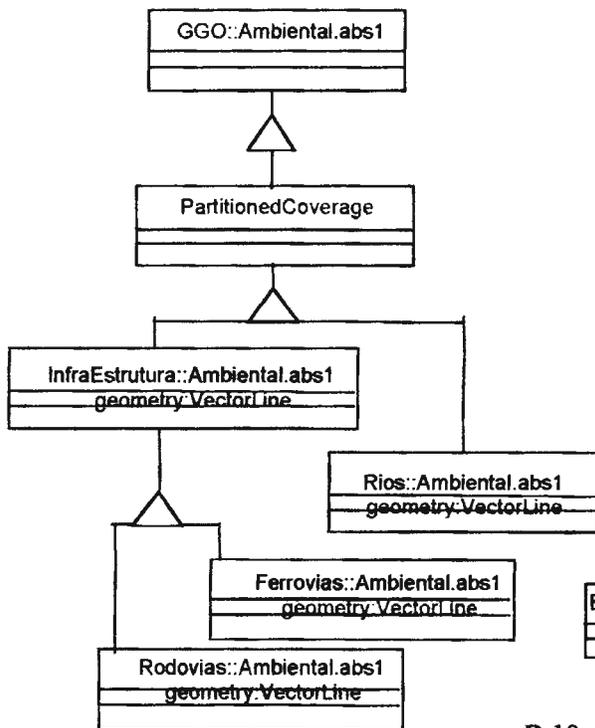
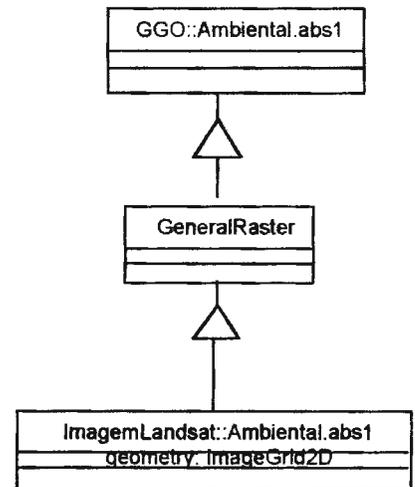
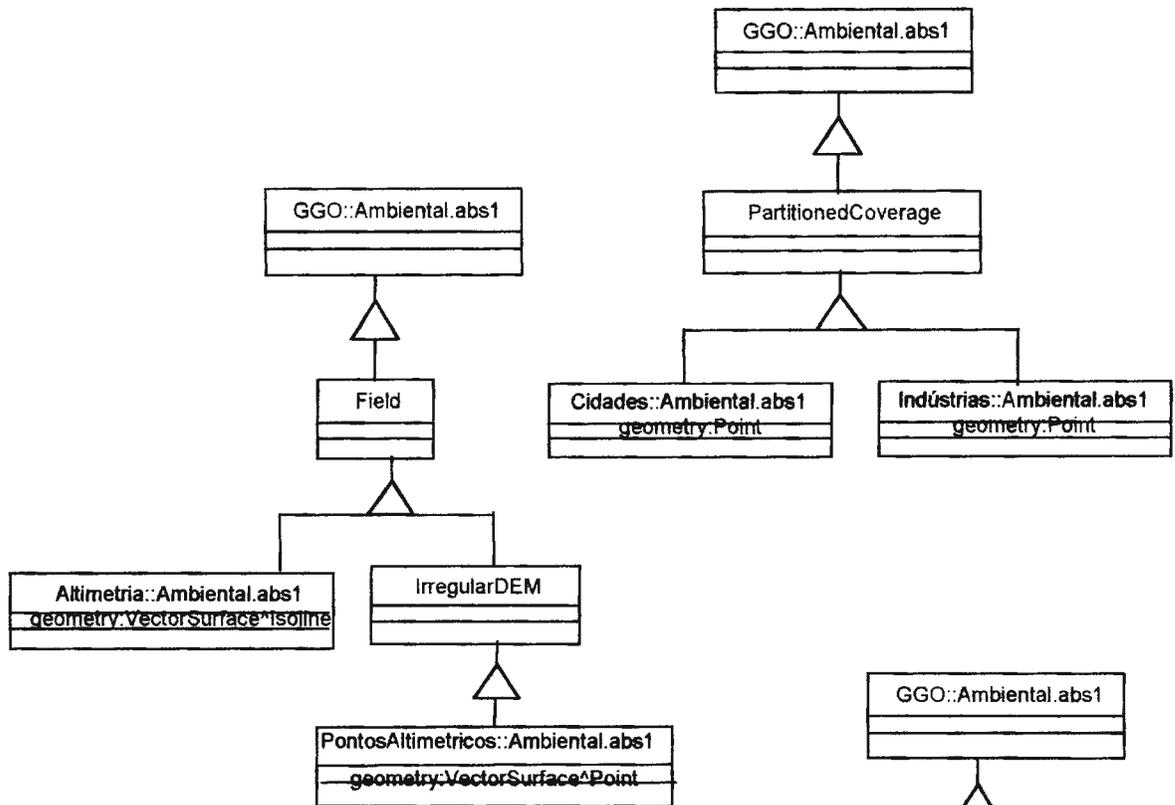
B.2.2.3. Projeto SIG-Ambiental

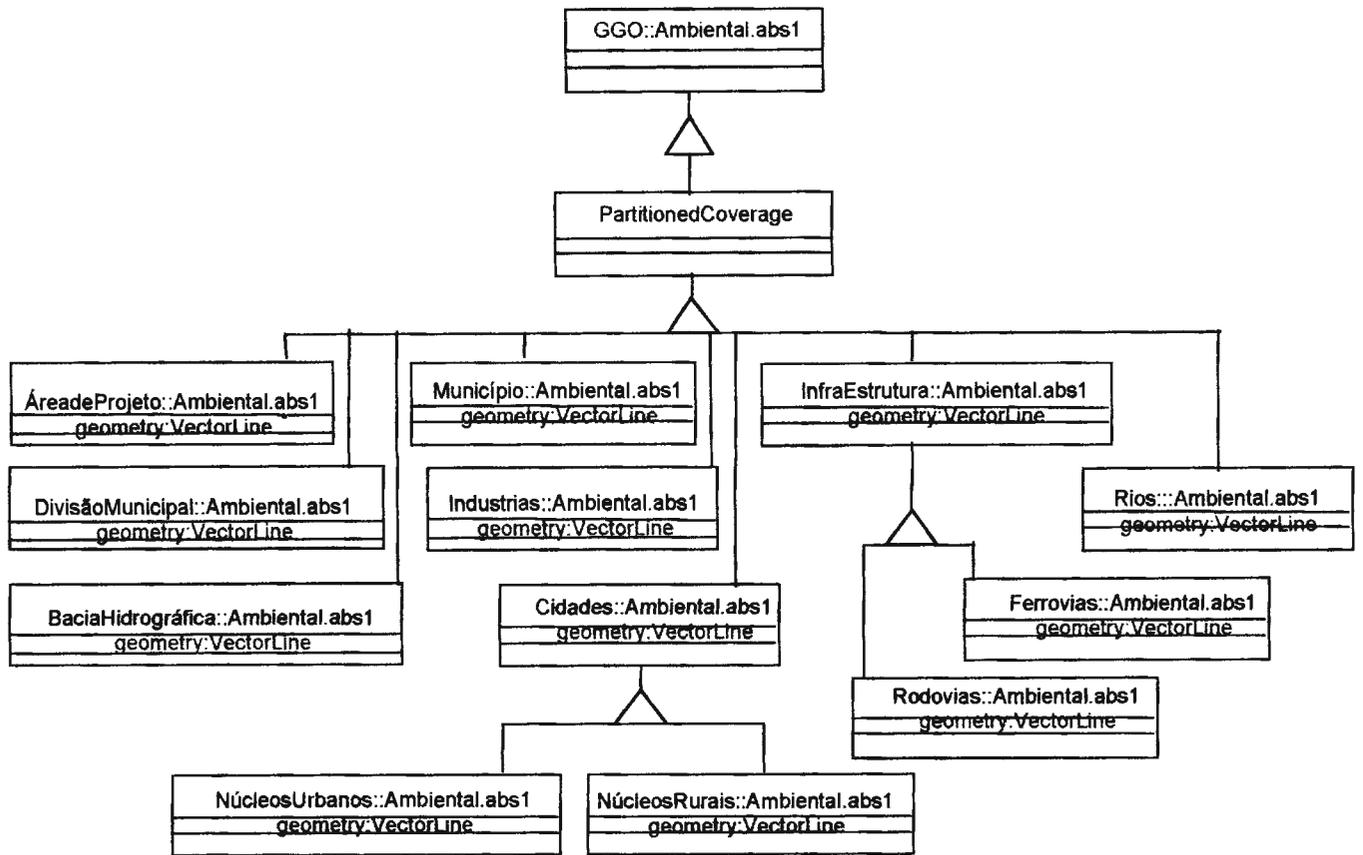
Entidades Não Gráficas



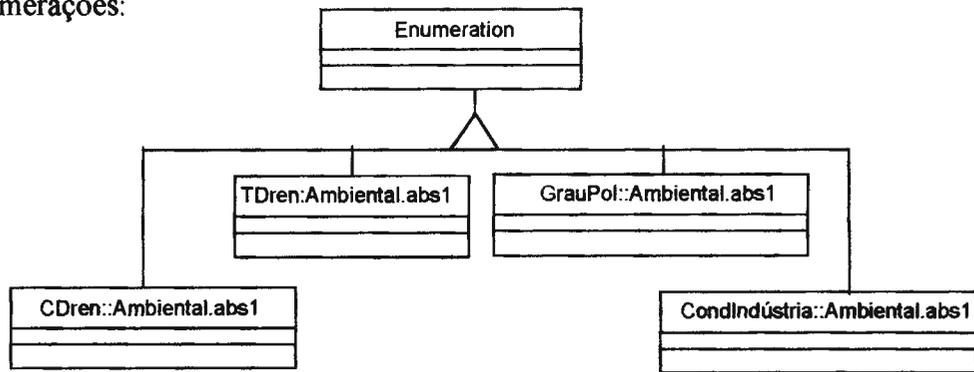
Entidades Gráficas







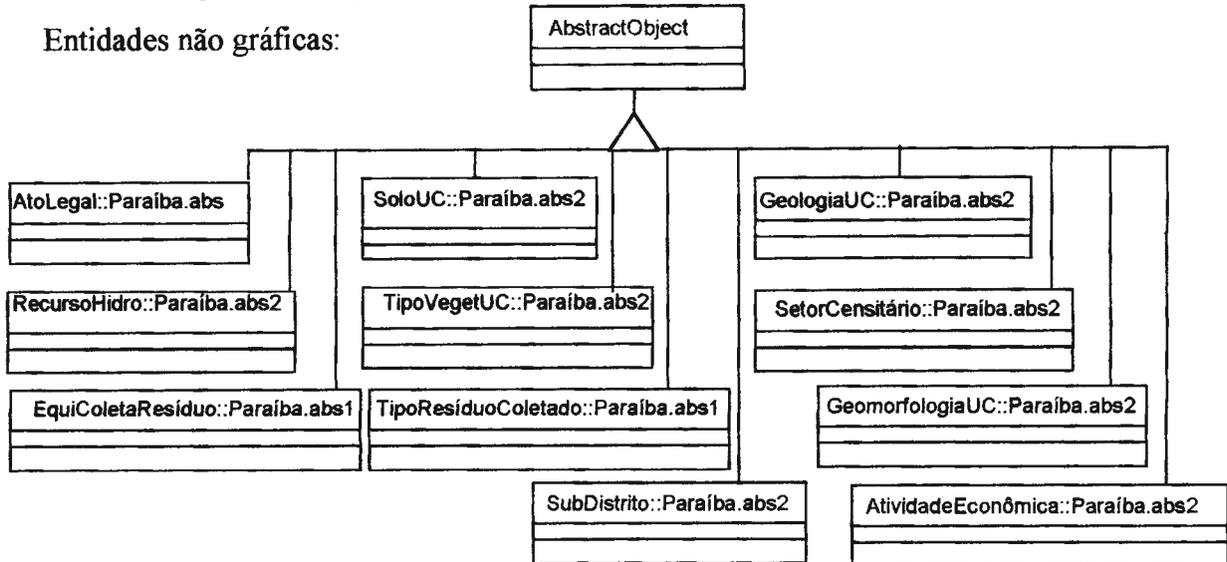
Enumerações:



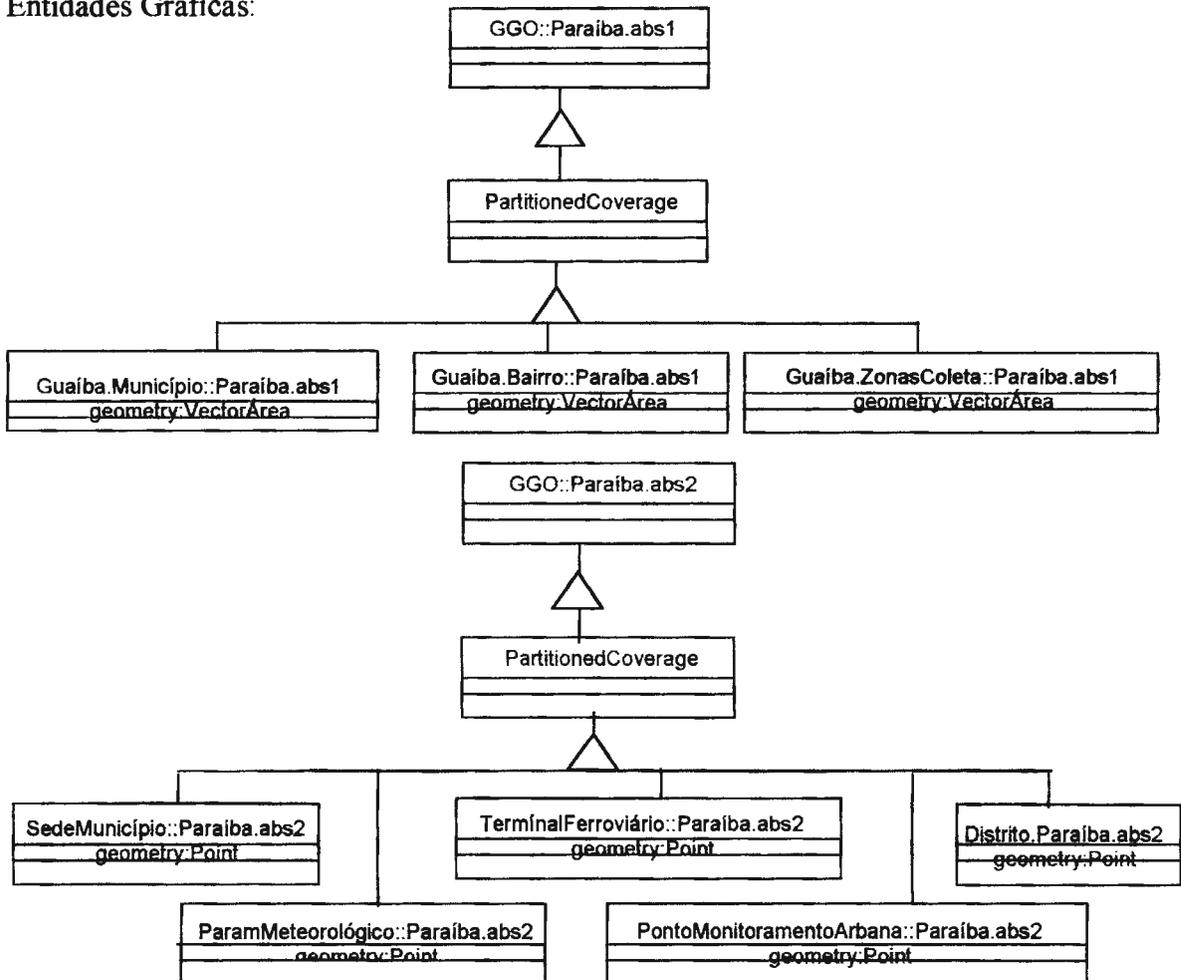
B.2.2.4. Projeto Pró-Paraíba

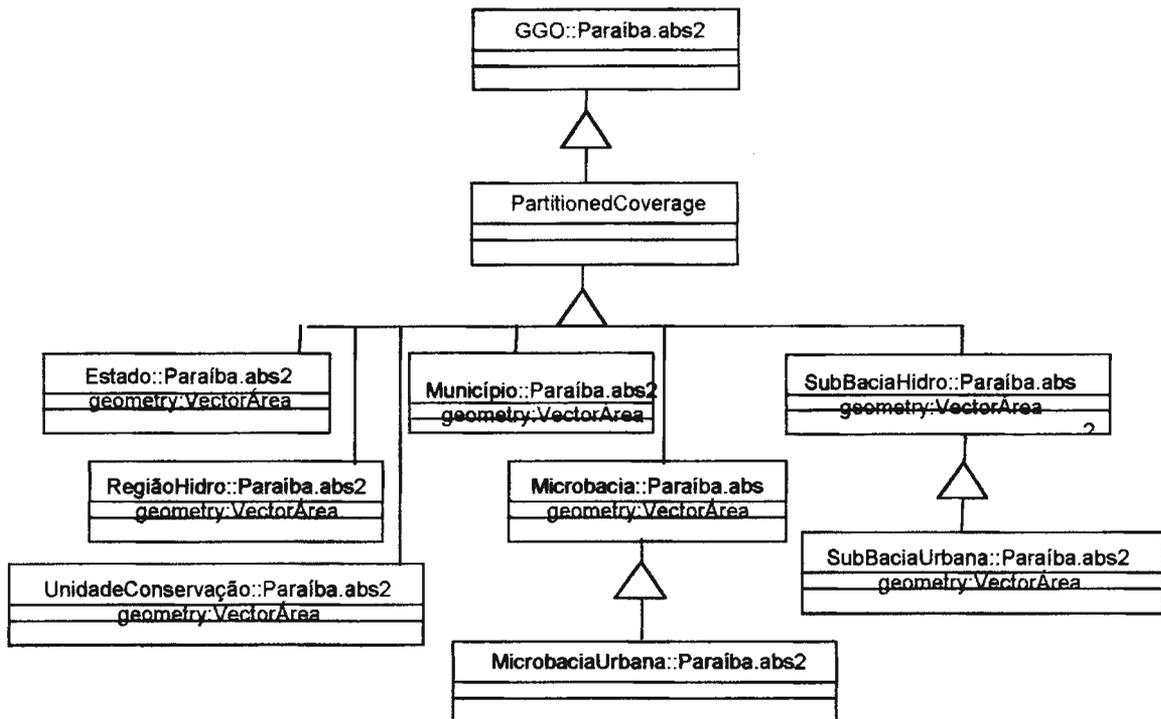
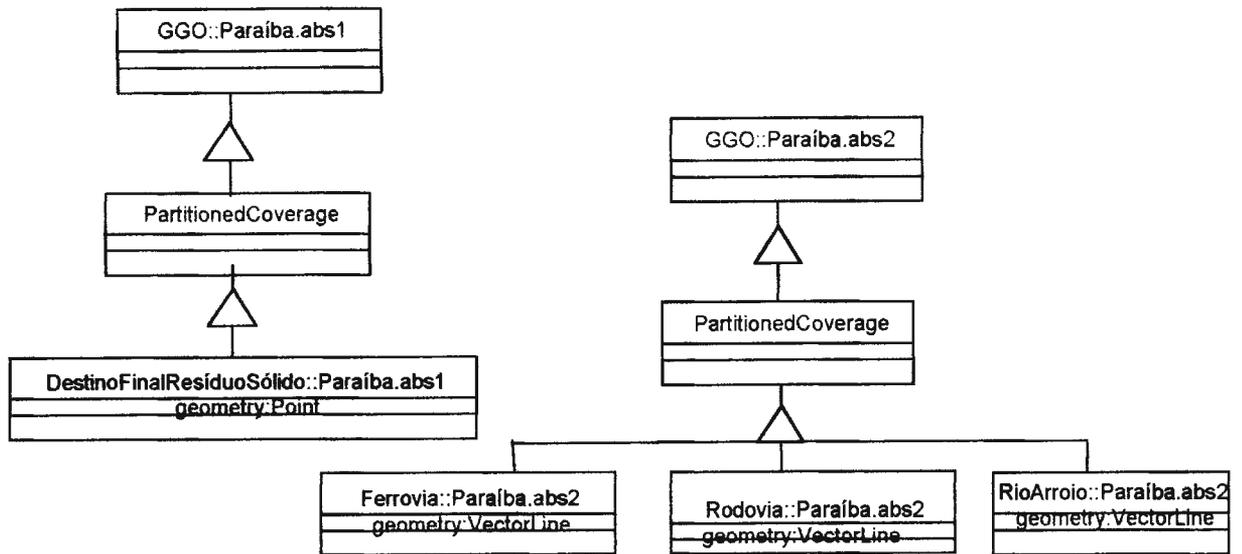
O esquema do Projeto Pró-Paraíba é igual ao projeto do Pró Guaíba, exceto pela localização geográfica, que será a Bacia Hidrográfica do Rio Paraíba do Sul.

Entidades não gráficas:

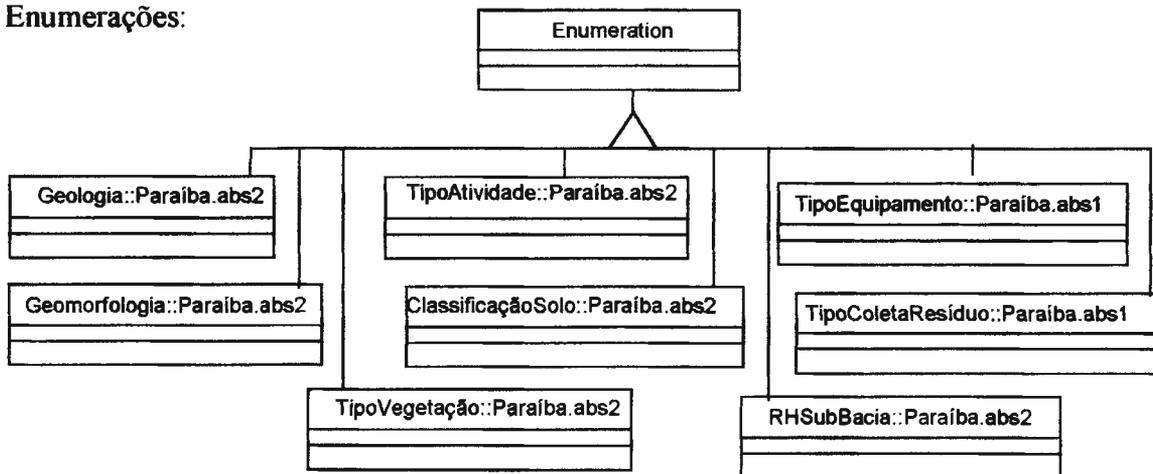


Entidades Gráficas:





Enumerações:



B.3. Análise dos esquemas

A análise dos esquemas é dividida em atividades que comparam os contextos espaciais, da aplicação e semântico.

B.3.1 Análise do contexto espacial dos esquemas

Esta análise é dividida em duas etapas, a saber: análise dos parâmetros espaciais e determinação do relacionamento espacial entre os esquemas.

B.3.1.1. Análise dos parâmetros espaciais

Esta análise compara os atributos do esquema que descrevem os parâmetros cartográficos usados para representar o esquema.

Tabela B.1 - Tabela de Parâmetros Cartográficos (TPC).

ATTRIBUTES	Qualba. Abo1	Qualba. Abo2	RHidro. Abo1	Ambiental. Abo1	Paraíba. Abo1	Paraíba. Abo2
<i>SpatialReferencingMM.coord System</i>	Geodésicas	Geodésicas	Coord. da Projeção	Geodésicas	Geodésicas	Geodésicas
<i>SpatialReferencingMM.ProjectionSystem</i>	Não tem	Não tem	Lambert	Não tem	Não tem	Não tem
<i>SpatialReferencingMM.coord System.HorizontalReference</i>	SAD-69	SAD-69	SAD-69	SAD-69	SAD-69	SAD-69
<i>SpatialReferencingMM.coord System.VerticalReference.</i>	Imbituba	Imbituba	Imbituba	Imbituba	Imbituba	SAD-69
<i>SpatialReferencingMM.base InverseScale</i>	10.000	250.000	100.000	100.000	10.000	250.000

Observa-se na tabela acima que os esquemas possuem os mesmos *data*, entretanto, eles diferem nos sistemas de projeção, sistemas de coordenadas e escala. Isto indica que se os esquemas tratarem do mesmo espaço geográfico, possivelmente, serão necessárias transformações de sistemas de projeção e generalização cartográfica na apresentação dos resultados.

B.3.1.2. Determinação do relacionamento espacial entre os esquemas

Tendo em vista que os espaços geográficos tratados pelos esquemas são descritos pelo seu nome, apenas a propriedade *GeneralLocationMM.Region* será analisada na Tabela de Localização Geográfica.

Para facilitar, os esquemas Guaíba e Paraíba, que apresentam dois níveis de abstrações diferentes, serão tratados em conjunto nesta análise.

Tabela B.2 - Tabela de Localização Geográfica (TLG).

ATTRIBUTES	Guaíba.abs1 e 2	RHidro.abs1	Ambiental.abs1	Paraíba.abs1 e 2
GeneralLocationMM.geographicRegion	Bacia do Guaíba	Bacia do Paraíba do Sul	Estado do Rio de Janeiro	Bacia do Paraíba do Sul

Aplicando-se a operação de interseção espacial aos espaços geográficos acima observa-se que:

$$\text{Guaíba.abs} \cap \text{RHidro.abs1} = 0$$

$$\text{Guaíba.abs} \cap \text{Ambiental.abs1} = 0$$

$$\text{Guaíba.abs} \cap \text{Paraíba.abs1} = 0$$

$$\text{RHidro.abs1} \cap \text{Ambiental.abs1} = \text{Rio de Janeiro}$$

$$\text{RHidro.abs1} \cap \text{Paraíba.abs} = \text{Vale do Paraíba do Sul}$$

$$\text{Ambiental.abs1} \cap \text{Paraíba.abs} = \text{Rio de Janeiro}$$

Esta análise resulta nos seguintes procedimentos de integração dos esquemas de acordo com os relacionamentos espaciais (RE) encontrados, a saber:

- Esquema Guaíba.abs1 e 2
 - RE vazio com todos os outros esquemas: Será verificada na etapa de criação do esquema externo a existência da instância "Bacia do Rio Guaíba" na classe BaciaHidrográfica da hierarquia *Region*. Poderá ocorrer os seguintes casos:
 - Não existe a instância "Bacia do Rio Guaíba": Será criada a instância "Bacia do Rio Guaíba" na classe BaciaHidrográfica, de acordo com a escala, e a ela associada a lista de *Theme* que corresponderá as classes do seu esquema de exportação. Na etapa de definição das instâncias de *Theme* será verificada a existência de instâncias que reflitam as classes dos esquemas de exportação. No caso da existência será aplicada a \mathcal{M} MultiSIG para integrar o esquema de exportação Guaíba.abs1 e 2 a parte do esquema federado que possui relacionamento semântico. Caso contrário, serão criadas novas instâncias de *Theme* com os devidos mapeamentos para os esquemas locais Guaíba.abs1 e abs2;
 - Existe a instância "Bacia do Rio Guaíba" associada a instâncias da classe *Theme*: Será recuperado a parte do esquema federado, expresso no MDC, que descreve esta Região. A seguir será aplicada a \mathcal{M} MultiSIG para integrar

o esquema de exportação Guaíba.abs1 e 2 a parte do esquema federado que possui relacionamento semântico.

- Esquema RHidro.abs1:
 - RE parcial com o esquema Ambiental.abs1: Neste caso a \mathcal{M} MultiSIG propõe que sejam criadas as seguintes instâncias nas classes abaixo da hierarquia *Region*.
 - Estado: "Rio de Janeiro" - será aplicada a \mathcal{M} MultiSIG aos esquemas RHidro.abs1 e Ambiental.abs1 e associados as instâncias de *Themes* resultantes da integração;
 - BaciaHidrográfica: "Bacia do Paraíba do Sul" - Será concatenado as instâncias de *Theme* derivadas do esquema exportação RHidro.abs1
 - RE total com o esquema Paraíba.abs1 e 2- Será aplicada a \mathcal{M} MultiSIG ao esquemas RHidro.abs1 e Paraíba.abs1 e 2
- Esquema Ambiental.abs1
 - RE parcial com o esquema Paraíba.abs1 e 2: Neste caso a \mathcal{M} MultiSIG propõe que sejam criadas as seguintes instâncias nas classes abaixo da hierarquia *Region*.
 - Estado: "Rio de Janeiro" - será aplicada a \mathcal{M} MultiSIG aos esquemas Ambiental.abs1, Paraíba.abs1 e 2 e associado as instâncias de *Theme* resultantes da integração;
 - BaciaHidrográfica: "Bacia do Paraíba do Sul" - Será concatenado as instâncias de *Theme* derivadas do esquema exportação Ambiental.abs1

Esta análise resulta nas seguintes instâncias das classes da hierarquia *Region*:

- BaciaHidrográfica: "Bacia do Rio Paraíba do Sul" com instâncias de *Theme* provenientes da integração dos esquemas RHidro.abs1 e Paraíba.abs1 e 2;
- BaciaHidrográfica: "Bacia do Rio Guaíba" com as instâncias de *Theme* provenientes dos esquemas de exportação Guaíba.abs1 e 2;
- Estado: Rio De Janeiro com as instâncias de *Theme* provenientes da integração dos esquemas (RHidro.abs1, Ambiental.abs1, Paraíba.abs1 e 2).

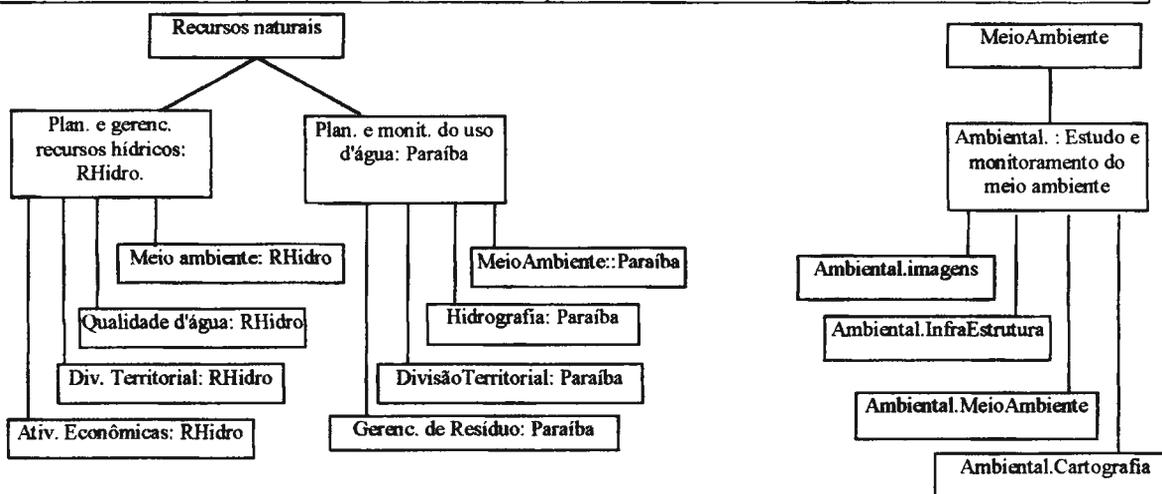
B.3.2. Análise do contexto da aplicação dos esquemas

Esta análise compara os atributos dos esquemas que tratam do mesmo espaço geográfico quanto as áreas de aplicação. Esta análise objetiva estabelecer a ordem para a integração dos esquemas de acordo com a proximidade de área de aplicação.

Desta forma, será realizada a análise relativa aos esquemas RHidro.abs1, Paraíba.abs1 e Ambiental.abs1, ao passo que o esquema do Projeto Guaíba, que não tem em comum o mesmo espaço geográfico, não será analisado nesta fase. Assim, a Tabela do Contexto da aplicação dos esquemas de exportação será:

Tabela B.3 - Tabela do Contexto da Aplicação(TCA).

ATTRIBUTES	Paraíba.abs1e2	RHidro.abs1	Ambiental.abs1
SourceMM.Terminus.Agency.agencyName	URGS e COPPE/UF RJ	COPPE/UF RJ	COPPE/UF RJ
SourceMM.Project.projectIdentifier	Pró-Paraíba	RHidro	Ambiental
SourceMM.Project.objective	Planejamento e monitoramento do uso d'água	Planejamento e gerenciamento dos recursos hídricos	Estudo e monitoramento do meio ambiente
SourceMM.Project.subsystem	Geren. de Residuo Meio Ambiente Hidrografia DivisãoTerritorial	Meio ambiente Qualidade d'água Divisão territorial Atividades economicas	Imagens InfraEstrutura MeioAmbiente Cartografia
SourceMM.Project.applicationArea	Recursos naturais	Recursos naturais	Meio Ambiente



De acordo com a análise da proximidade de área de aplicação e dos seus subsistemas, conforme sugerido pela MultiSIG, será estabelecida a seguinte seqüência para a integração dos esquemas ((RHidro, Paraíba), Ambiental)).

B.3.3. Análise do contexto semântico dos esquemas

Esta etapa analisa os esquemas que tratam do mesmo espaço geográfico e que possuem proximidade de aplicação. Ela tem por finalidade estabelecer o grau de proximidade semântica entre as classes dos esquemas com o objetivo de identificar as similaridades/diferenças entre eles.

As classes dos esquemas que não possuem relacionamento espacial com os outros esquemas, como no caso, Guaíba.abs(1 e 2), serão considerados ao final desta fase como classes heterogêneas.

Desta forma, dando prosseguimento a aplicação da ~~em~~ MultiSIG com o objetivo de testá-la, será analisado o par de esquemas formado por (Paraíba.abs(1 e 2), e RHidro.abs1) de acordo com as atividades abaixo:

B.3.3.1 Análise do contexto semântico dos esquemas Paraíba.abs(1 e 2), e RHidro.abs1

B.3.3.1.1. Análise das denominações

Esta atividade objetiva identificar as similaridades e diferenças entre as classes dos esquemas, em relação a sua semântica no mundo real.

Tabela B.4 - Tabela de Classes (TC).

PARAÍBA			RHIDRO		
CLASSE _i	LISTA(TERMO _{ik})	TIPO	CLASSE _j	LISTA(TERMO _{jk})	TIPO
ESTADO	Estado	GGO	UF	Estado	GGO
MUNICÍPIO	Município	GGO	MUNIC	Município	GGO
DISTRITO	Distrito	GGO			
BAIRRO	Bairro	GGO			
SUBDISTRITO	Subdistrito	AO			
SETOR CENSITÁRIO	Sector censitário	AO			
UNIDADE CONSERVAÇÃO	Unidade de conservação	GGO			
SEDE MUNICÍPIO	Cidades	GGO	CIDADES	Cidades	GGO
			TIPO CIDADE	Tipo de cidade	ENUM
ZONAS COLETA	Zonas de coleta de lixo	GGO			
FERROVIA	Ferrovia	GGO			
TERMINAL FERROVIÁRIO	Terminal ferroviário	GGO			
RODOVIA	Rodovia	GGO	RODOVIA	Rodovia	GGO
			TIPO CONSERVAÇÃO	Class.da cons da rodovia	ENUM
REGIÃO HIDRO	Região hidrográfica	GGO	BACIA HIDROGRÁFICA	Bacia hidrográfica	GGO
SUBBACIA HIDRO	Sub bacia hidrográfica	GGO			
SUBBACIA URBANA	Sub Bacia Urbana	GGO			
RH SUBBACIA	Região da Subbacia hidrográfica	ENUM			
MICROBACIA	Microbacia	GGO			

MICROBACIAÚRBANA	Microbacia urbana	GGO			
RECURSOHIDRO	Class.do Rec. hídrico	AO	TIPOHIDRO	Class.de Hidro	ENUM
RIOARROIO	Hidrografia	GGO	HIDROGRAFIA	Hidrografia	GGO
ATO LEGAL	Ato legal da unidade de conservação e dos municípios	AO			
PONTO MONIT. AR	Dados da estação de monitoramento do ar	GGO			
PARAM. METEORO.	Dados meteorológicos	GGO	DADOS METEOROLÓGICOS	Dados meteorológicos	GGO
			TIPOCLIMA	Tipo de clima	ENUM
GEOLOGIAUC	Geologia	AO			
GEOLOGIA	Class. da geologia	ENUM			
TIPOVEGETUC	Vegetação	AO	VEGETAÇÃO	Tipo de vegetação	GGO
TIPOVEGETAÇÃO	Class. da vegetação	ENUM	TIPOVEGETAÇÃO	Descreve a vegetação	ENUM
GEOMORFOLOGIAUC	Geomorfologia	AO			
GEOMORFOLOGIA	Class. da Geomorfologia	ENUM			
SOLOUC	Solo	AO	SOLO	Tipo solos	GGO
CLASSIFICAÇÃO SOLO	Class. Solo	ENUM	TIPOSOLO	Class.do Solo	ENUM
			PROFUNDIDADE	Intervalo de profundidade	ENUM
			ESTRUTURA	Tipo de estrutura	ENUM
			RELEVO	Class.do relevo	GGO
			TIPORELEVO	Tipo de relevo	ENUM
			TIPOPEDREGOSIDADE	Class.da pedregosidade	ENUM
			ALTIMETRIA	Curvas de níveis	GGO
			USODO SOLO	Uso do solo	GGO
			TIPOUSO	Tipo de uso	ENUM
ATIVIDADE ECONÔMICA	Estabelecimento com Atividade econômica	AO	USUÁRIO	Estabelecimento com atividade econômica	GGO
TIPO ATIVIDADE	Tipo de atividade	ENUM			
			INDÚSTRIA	Indústria	GGO
			INDÚSTRIA CAD. PROCON	Estabelecimento de indústria cadastrada no PROCON	GGO
TIPO RESÍDUO COLETADO	Tipo de resíduo sólido	AO			
EQUIP. COLETA RESÍDUO	Equipamento de coleta de resíduo sólido	AO			
DESTINO FINAL RESÍDUO SÓLIDO	Destino final do resíduo sólido	GGO	PONTO DEPÓSITO LIXO	Ponto representativo de depósito de lixo	GGO
TIPO COLETA RESÍDUO	Tipo de coleta	ENUM			
TIPO EQUIPAMENTO	Tipo de equipamento	ENUM			
			DADOS SEDIMENTOLOGIA	Dados sedimentologia	GGO
			TIPO FUNDO	Tipo de fundo do rio	ENUM
			TIPO SUSPENSÃO	Tipo de suspensão	ENUM
			QUALIDADE ÁGUA	Qualidade d'água	GGO
			PGM INVESTIMENTO	Pgm de investimento	AO
			DADOS HIDROLÓGICOS	Dados hidrológicos	GGO

Esta análise resulta na distribuição das classes dos esquemas nos seguintes grupos temáticos:

DIVISÕES DO ESPAÇO GEOGRÁFICO	
PARAÍBA	RHIDRO
Estado::Paraíba.abs2	UF::RHidro.abs1
Município::Paraíba.abs2	Munic::RHidro.abs1
Distrito::Paraíba.abs2	
SedeMunicípio::Paraíba.abs2	Cidades::RHidro.abs1

	DadosHidrologicos::RHidro.abs1 DadosSedimentologia::RHidro.abs1 TipoFundo::RHidro.abs1 TipoSuspensão::RHidro.abs1 QualidadeÁgua::RHidro.abs1
--	--

GERENCIAMENTO DE RESÍDUO	
PARAÍBA	RHIDRO
TipoResíduoColetado::Paraíba.abs1 EquipaColetaResíduo::Paraíba.abs1 DestinoFinalResíduoSólido::Paraíba.abs1 TipoColetaResíduo::Paraíba.abs1 TipoEquipamento::Paraíba.abs1	PontoDepósitoLixo::RHidro.abs1

B.3.3.1.2. Análise dos relacionamentos semânticos

Esta etapa consiste em analisar para cada classe os relacionamentos, dentro do grupo temático e em relação aos outros grupos temáticos.

Tabela B.5 - Tabela de Relacionamentos (TR).

DIVISÕES DO ESPAÇO GEOGRÁFICO									
PARAÍBA					RHIDRO				
REL _{mi}	C _{me}	C _{mf}	CAR _{mi}	AT. DO REL _{mi}	REL _{nj}	C _{nk}	C _{nl}	CAR _{nj}	AT. DO REL _{nj}
CONTEM	Estado	Município	1:n	CódEstado	POSSUI I	UF	Munic	1:n	CD_UF
PERTENCE	Sede Município	Município	1:1	CódMun	POSSUI I	Munic	Cidades	1:n	CD_UF CD_Mun
					POSSUI I	Cidades	TipoCidade	1:1	tipo
POSSUI	Município	AtoLegal	1:n	CódMun					
ABRANGE	Município	Unidade Conservação	N:m	CódMun CódUc					
PERTENCE	Distrito	Município	N:1	CódMun					
POSSUI	Município	Bairro	1:n	CódMun					
ABRANGE	Bairro	ZonasColeta	N:m	CódBairro CódZonas Coleta					
PERTENCE	SubDistrito	Distrito	N:1	CódDistrito CódMun					
PERTENCE	Setor Censitário	SubDistrito	N:1	CódSubDistrit o CódDistrito CódMun					
POSSUI	Unidade Conservação	AtoLegal	N:m	CódMun					
					POSSUI	Munic	Altimetria	n:m	
					POSSUI	Munic	Relevo	n:n	CD_Mun
					POSSUI	Munic	Solo	n:n	CD_Mun
					POSSUI	Munic	PontoDepo- sito Lixo	1:n	CD_Mun
					POSSUI	Munic	Dados Meteorológi- cos	1:n	CD_Mun
					POSSUI	Munic	Vegetação	N:m	CD_Mun
					POSSUI	Munic	UsoDoSolo	N:n	CD_Mun
					POSSUI	Munic	Rodovia	N:m	CD_Mun
					POSSUI	Munic	Pgm Investimento	N:m	CD_Mun
DESENVOL- VE	Município	Atividade Econômica	1:n	CódMun	POSSUI	Munic	Usuário	1:n	CD_Mun
POSSUI	Atividade	Tipo	1:n	CódAtividad					

	Econômica	Atividade		e					
					CONTEM	Bacia Hidrográfica	Munic	N:m	
ATENDE SE LOCALIZA	Destino Final Resíduo Sólido	Município	1:n 1:1	CódMun					
SE LOCALIZA	Ponto Monitor Ar	Município	1:1	CódMun					
POSSUI	Zonas Coleta	Tipo Resíduo Coletado	N:m	CódTipo Coleta					
POSSUI	Unidade Conservação	Geologia Uc	1:n	CódUc					
POSSUI	Unidade Conservação	Solo UC	1:n	CódUc					
POSSUI	Unidade Conservação	Geomorfologia Uc	1:n	CódUc					
POSSUI	Unidade Conservação	Tipo VegetUc	1:n	CódUc					
ÁREA DE ESTUDO DA HIDROGRAFIA									
PARAÍBA					RHIDRO				
REL _{mi}	C _{me}	C _{mf}	CAR _{mi}	AT. DO REL _{mi}	REL _{nj}	C _{nk}	C _{nl}	CAR _{nj}	AT. DO REL _{nj}
CONTEM	Região Hidro	SubBacia Hidro	1:n	CódReg Hidro					
SE ESPECIALIZA	SubBacia Hidro	SubBacia Urbana	1:n	CódSub Bacia					
CONTEM	SubBacia Hidro	Microbacia	1:n	CódSub Bacia					
SE ESPECIALIZA	Microbacia	Microbacia Urbana	1:n	Cód Microbacia					
POSSUI	SubBacia Hidro	RHSub Bacia	1:n	CódSub Bacia					
PERTENCE	Recurso Hidro	RHSub Bacia	1:1	CódRH					
SE ESPECIALIZA	Recurso Hidro	Rio Arroio	1:n	CódRH					
					ABRANG E	Bacia Hidrográfica	Hidrografia	1:n	CD_BH
					POSSUI	Hidrografia	Tipo Hidro	1:1	Tipo
					ABRANG E	Bacia Hidrográfica	Munic	N:m	CD_BH
					POSSUI	Hidrografia	Dados Hidrológicos	1:n	CD_H
					POSSUI	Hidrografia	Qualidade Água	1:n	CD_H
					POSSUI	Hidrografia	Dados Sedimentologia	1:n	CD_H
					POSSUI	Dados Sedimentologia	Tipo Fundo	1:1	Tipo Sedimento Fundo
					POSSUI	Dados Sedimentologia	Tipo Suspensão	1:1	Tipo Sedimento Suspensão
INFRA-ESTRUTURA									
PARAÍBA					RHIDRO				
REL _{mi}	C _{me}	C _{mf}	CAR _{mi}	AT. DO REL _{mi}	REL _{nj}	C _{nk}	C _{nl}	CAR _{nj}	AT. DO REL _{nj}
POSSUI	Ferrovia	Terminal Ferroviário	1:n	Nome Ferrovia					
					POSSUI	Munic	Rodovia	n:m	
					TEM	Rodovia	Tipo Conservação	1:1	Gr. Cons
DADOS DO MEIO AMBIENTE									
PARAÍBA					RHIDRO				
REL _{mi}	C _{me}	C _{mf}	CAR _{mi}	AT. DO REL _{mi}	REL _{nj}	C _{nk}	C _{nl}	CAR _{nj}	AT. DO REL _{nj}
POSSUI	Unidade Conservação	Geologia UC	1:n	CódUc					
POSSUI	Geologia UC	Geologia	1:1	CódGeologia					
POSSUI	Unidade Conservação	Tipo VegetUc	1:n	CódUc					
POSSUI	Tipo VegetUc	Tipo Vegetação	1:1	Cód Vegetação					
POSSUI	Unidade Conservação	Geomorfologia UC	1:n	CódUc					
POSSUI	Geomorfologia	Geomorfologia	1:1	Cód					

	gia UC	gia		Geomorfo					
POSSUI	Unidade Conservação	SoloUC	1:n	CódUc					
POSSUI	SoloUC	Classificação Solo	1:1	CódTipoSolo					
					POSSUI	Munic	Vegetação	1:n	CD_Mun
					TEM	Vegetação	Tipo Vegetação	1:1	Tipo
					POSSUI	Munic	UsoDoSolo	1:n	CD_Mun
					TEM	UsoDoSolo	TipoUso	1:1	CD_uso
					POSSUI	Munic	Solo	1:n	CD_Mun
					TEM	Solo	TipoSolo	1:1	TipoSolo
					TEM	Solo	Profundidade	1:1	Profun
					TEM	Solo	Estrutura	1:1	Estrutura
					POSSUI	Munic	Relevo	1:n	CD_Mun
					TEM	Relevo	TipoRelevo	1:1	TipoRel
					TEM	Relevo	Tipo Pedregosidade	1:1	TipoPedra
						Munic	Altimetria	n:m	
DADOS DE ATIVIDADES ECONÔMICAS									
PARAIBA					RHIDRO				
REL _{mi}	C _{me}	C _{mf}	CAR _{mi}	AT. DO REL _{mi}	REL _{nj}	C _{nk}	C _{nl}	CAR _{nj}	AT. DO REL _{nj}
					POSSUI	Munic	Pgm Investimento	n:m	CD_Mun
DESENVOLVE	Município	Atividade Econômica	1:n	CódMun	CONTEM	Munic	Usuário	1:n	CD_Mun
POSSUI	Atividade Econômica	Tipo Atividade	1:1	CódAtividade	SE ESPECIALIZA	Usuário	Indústria	1:n	CD_U
					SE ESPECIALIZA	Indústria	IndústriaCad PROCON	1:n	CD_I
DADOS DE CAMPO									
PARAIBA					RHIDRO				
REL _{mi}	C _{me}	C _{mf}	CAR _{mi}	AT. DO REL _{mi}	REL _{nj}	C _{nk}	C _{nl}	CAR _{nj}	AT. DO REL _{nj}
SE LOCALIZA	PontoMonitor Ar	Município	1:1	CódMun	POSSUI	Munic	Dados Meteorológicos	1:n	CD_Mun
					PERTENCE	Dados Meteorológicos	TipoClima	1:1	CD_Estação Meteorolo
COLETA	PontoMonitor Ar	Param Meteorológico	1:n	CódEstação					
					POSSUI	Hidrografia	Dados Hidrológicos	1:n	CD_H
					POSSUI	Hidrografia	Qualidade d'água	1:n	CD_H
					POSSUI	Hidrografia	Dados Sedimentologia	1:n	CD_H
GERENCIAMENTO DE RESÍDUO									
PARAIBA					RHIDRO				
REL _{mi}	C _{me}	C _{mf}	CAR _{mi}	AT. DO REL _{mi}	REL _{nj}	C _{nk}	C _{nl}	CAR _{nj}	AT. DO REL _{nj}
POSSUI	ZonasColeta	TipoResiduo Coletado	1:n	CódTipo Coleta					
POSSUI	ZonasColeta	TipoColeta Residuo	1:n	CódZona Coleta					
POSSUI	TipoResiduo Coletado	TipoColeta Residuo	1:1	CódTipo Coleta					
	TipoResiduo Coletado	EquipaColeta Residuo	1:1	CódEqui Residuo					
	EquipaColetaResiduo	TipoEquipamento	1:1	CódTipo Equip					
ATENDE	DestinoFinal ResiduoSólido	Município	N:m	CódMun					
SE LOCALIZA	DestinoFinal ResiduoSólido	Município	1:n	Cód DestinoLixo	POSSUI	Munic	PontoDepositoLixo	1:n	CD_Mun
DEPOSITA	TipoResiduo Coletado	DestinoFinal ResiduoSólido	N:m						

Uso CD_uso.UsoDoSolo::RHidro.abs1
 Município CD_Mun.UsoDoSolo::RHidro.abs1
 CustoPorHectare CustoPHectare.UsoDoSolo::RHidro.abs1
 ValorMonetário ValorMonetário.UsoDoSolo::RHidro.abs1
 Region = Bacia Hidrográfica do Rio Guaíba {
 Uso "Informação não disponível para Região"
 Município "Informação não disponível para Região"
 CustoPorHectare "Informação não disponível para Região"
 ValorMonetário "Informação não disponível para Região"}

Export schema source:

OriginClass1: UsoDoSolo::RHidro.abs1
 Attributes: CD_uso, CD_Mun, CustoPHectare, ValorMonetario
 position.Geometry: VectorÁrea
 Relationships:

Rel	ObjectA	ObjectB	Attributes	Car
Possui	Município	UsoDoSolo		1:n

 Source.DBInfo.dbIdentifier:
 Acess method:

Superclass:

Class: Vegetação
 Region:
 Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba
 Attributes:
 Tipo, Município
 Geometry: Área
 Source schema mapping method:
 Region = Bacia Hidrográfica do Rio Paraíba do Sul {
 Tipo Tipo.Vegetação::RHidro.abs1
 Município CD_Mun.Vegetação::RHidro.abs1
 Region = Bacia Hidrográfica do Rio Guaíba {
 Tipo "Informação não disponível para Região"
 Município "Informação não disponível para Região"}

Export schema source:

OriginClass1: Vegetação::RHidro.abs1
 Attributes: Tipo, CD_Mun
 position.Geometry: VectorLine
 Relationships:

Rel	ObjectA	ObjectB	Attributes	Car
Possui	Município	Vegetação		1:n

 Source.DBInfo.dbIdentifier:
 Acess method:

Superclass:

Class: Relevo
 Region:
 Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba
 Attributes:
 Tipo, Pedregosidade, Município
 Geometry: Área
 Source schema mapping method:
 Region = Bacia Hidrográfica do Rio Paraíba do Sul {
 Tipo TipoRel.Relevo::RHidro.abs1
 Pedregosidade TipoPedra.Relevo::RHidro.abs1
 Município CD_Mun.Relevo::RHidro.abs1
 Region = Bacia Hidrográfica do Rio Guaíba {
 Tipo "Informação não disponível para Região"
 Pedregosidade "Informação não disponível para Região"
 Município "Informação não disponível para Região"}

Export schema source:

OriginClass1: Relevo::RHidro.abs1
 Attributes: TipoRel, TipoPedra, CD_Mun
 position.Geometry: VectorLine
 Relationships:

<i>Rel</i>	<i>ObjectA</i>	<i>ObjectB</i>	<i>Attributes</i>	<i>Car</i>
Possui	Município	Relevo		1:n

Source.DBInfo.dbIdentifier:
Acess method:

Superclass:

Class: Solo

Region:

Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba

Attributes:

Tipo, Profundidade, Estrutura, Município

Geometry: Área

Source schema mapping method:

```
Region = Bacia Hidrográfica do Rio Paraíba do Sul {
  Tipo                TipoSolol.Solo::RHidro.abs1
  Profundidade        TipoPedra.Solo::RHidro.abs1
  Estrutura            Estrutura.Solo::RHidro.abs1
  Município            CD_Mun.Solo::RHidro.abs1}
```

```
Region = Bacia Hidrográfica do Rio Guaíba {
  Tipo                "Informação não disponível para Região"
  Profundidade        "Informação não disponível para Região"
  Estrutura            "Informação não disponível para Região"
  Município            "Informação não disponível para Região"}
```

Export schema source:

OriginClass1: Solo::RHidro.abs1

*Attributes: TipoSolol, Profun, estrutura, CD_Mun
position.Geometry: VectorLine*

Relationships:

<i>Rel</i>	<i>ObjectA</i>	<i>ObjectB</i>	<i>Attributes</i>	<i>Car</i>
Possui	Município	Relevo		1:n

Source.DBInfo.dbIdentifier:

Acess method:

B.6. Criação do esquema externo

Esta etapa é dividida nas seguintes atividades:

B.6.1. Definição das instâncias da hierarquia *Regions*

Nesta etapa são descritas as instâncias da hierarquia *Region* que definem os espaços geográficos resultantes da integração dos esquemas de acordo com as escalas.

Tendo em vista o objetivo desta aplicação da  MultiSIG testá-la, não serão descritos os métodos de acesso para cada instância das subclasses de *Region*.

```
region: Region (BaciaHidrográfica (
  name: Do Rio Paraíba do Sul                /* Esquemas: Paraíba.abs1
  inverseScale: 1:10.000
  theme: List (Município, Bairro ZonasColeta, TipoResiduoColetado, EquipaColetaResiduo,
  DestinoFinalResiduoSólido))
```

```
region: Region (BaciaHidrográfica(
  name: Do Rio Guaíba                        /* Esquemas: Guaíba.abs1
  inverseScale: 1:10.000
  theme: List (Município, Bairro ZonasColeta, TipoResiduoColetado, EquipaColetaResiduo,
  DestinoFinalResiduoSólido))
```

region: *Region (BaciaHidrográfica)*
 name: Do RioParaíba do Sul /* *Esquemas: Paraíba.abs2 U RHidro.abs1*
 inverseScale: 1:250.000
 theme: *List (Estado, Município250, Cidades, SedeMunicípio, UnidadeConservação, Distrito, SubDistrito, SetorCensitário, AtoLegal, BaciaHidrográfica, SubBacia, SubBaciaUrbana, Microbacia, MicrobaciaUrbana, Hidrografia, Ferrovia, TerminalFerroviário, Rodovia, GeologiaUC, GeomorfologiaUC, VegetaçãoUC, SoloUC, AtividadeEconômica, PontoMonitoAr, DadosMeteorologicos, PgmInvestimento, Usuário, Indústria, IndústriaPROCON, DadosHidrológicos, DadosSedimentologia, QualidadeÁgua, Altimetria, UsoDoSolo, Vegetação, Relevo, Solo))*

region: *Region (BaciaHidrográfica)*
 name: Rio Grande do Sul /* *Esquema: Guaíba.abs2*
 inverseScale: 1:250.000
 theme: *List(Estado, Município250, SedeMunicípio, UnidadeConservação, Distrito, SubDistrito, SetorCensitário, AtoLegal, BaciaHidrográfica, SubBacia, SubBaciaUrbana, Microbacia, MicrobaciaUrbana, Hidrografia, Ferrovia, TerminalFerroviário, Rodovia, GeologiaUC, GeomorfologiaUC, VegetaçãoUC, SoloUC, AtividadeEconômica, PontoMonitoAr, DadosMeteorológicos))*

B.6.2. Definição dos Temas

theme: *Theme(*
 name: Município
 attributes: Código, Nome, Estado, População, Área, AnoInstituição,
 geometry: Área
 regions: *List (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)*
 inverseScale: 1:10.000)

theme: *Theme(*
 name: Bairro
 attributes: Código, Nome, Município
 geometry: Área
 regions: *List (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)*
 inverseScale: 1:10.000)

theme: *Theme(*
 name: ZonasColeta
 attributes: Código, Bairro, TipoColeta, Descrição
 geometry: Área
 regions: *List (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)*
 inverseScale: 1:10.000)

theme: *Theme(*
 name: ResíduoColetado
 attributes: HoraColeta, Equipamento, ZonaColeta, TipoColeta, DestinoLixo, Quantidade
 regions: *List (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)*
 inverseScale: 1:10.000)

theme: *Theme(*
 name: EquipamentoColetaResíduo
 attributes: Equipamento, Tipo, Capacidade, AnoAquisição
 regions: *List (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)*
 inverseScale: 1:10.000)

theme: *Theme(*
 name: DestinoFinalResíduoSólido
 attributes: Destino, Município, Endereço, Número, LixoGerado
 geometry: Ponto
 regions: *List (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)*

inverseScale: 1:10.000)

theme: *Theme*(

name: Estado

attributes: Código, Nome, Área

geometry: Área

regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)

inverseScale: 1:250.000)

theme: *Theme*(

name: Município250

attributes: Código, Nome, Estado, Mesoregião, Microregião, Área, População, AnoInstituição

geometry: Área

regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)

inverseScale: 1:250.000)

theme: *Theme*(

name: Cidades

attributes: Código, Nome, Estado, Município, Distrito, População, Tipo

geometry: Ponto

regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul)

inverseScale: 1:250.000)

theme: *Theme*(

name: SedeMunicípio

attributes: Código, Nome, Estado, Município, Distrito, População, Tipo, Latitude, Longitude

geometry: Ponto

regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)

inverseScale: 1:250.000)

theme: *Theme*(

name: UnidadeConservação

attributes: Código, Município, Criação, Administração, NívelAdministrativo, Manejo, Área

geometry: Área

regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)

inverseScale: 1:250.000)

theme: *Theme*(

name: Distrito

attributes: Código, Nome, Município, Altitude, População

geometry: Ponto

regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)

inverseScale: 1:250.000)

theme: *Theme*(

name: SubDistrito

attributes: Código, Nome, Distrito, Município

regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)

inverseScale: 1:250.000)

theme: *Theme*(

name: SetorCentitário

attributes: Código, Município, Distrito, SubDistrito

regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)

inverseScale: 1:250.000)

theme: *Theme*(

name: AtoLegal
attributes: Código, Município, Número, Tipo, DataPublicação, Escopo
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)
inverseScale: 1:250.000)

theme: *Theme*(
name: BaciaHidrográfica
attributes: Código, Nome, Área
geometry: Área
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)
inverseScale: 1:250.000)

theme: *Theme*(
name: SubBacia
attributes: Código, Nome, BaciaHidrográfica
geometry: Área
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)
inverseScale: 1:250.000)

theme: *Theme*(
name: SubBaciaUrbana
attributes: Código, CódigoSB, NomeSB, BaciaHidrográfica, Qualificação, Ocupação, OcupaçãoUrbana
geometry: Área
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)
inverseScale: 1:250.000)

theme: *Theme*(
name: Microbacia
attributes: Código, Nome, SubBacia
geometry: Área
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)
inverseScale: 1:250.000)

theme: *Theme*(
name: MicrobaciaUrbana
attributes: Código, CódigoMB, NomeMB, SubBacia, Qualificação, Ocupação, OcupaçãoUrbana
geometry: Área
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)
inverseScale: 1:250.000)

theme: *Theme*(
name: Ferrovia
attributes: Nome
geometry: Linha
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)
inverseScale: 1:250.000)

theme: *Theme*(
name: TerminalFerroviário
attributes: Ferrovia, Estação
geometry: Ponto
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)
inverseScale: 1:250.000)

theme: *Theme*(
name: Rodovia
attributes: Código, Nome, Juridição, GrauConservação
geometry: Linha
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)
inverseScale: 1:250.000)

theme: *Theme*(
name: GeologiaUc
attributes: UnidadeConservação, Geologia, Área
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)
inverseScale: 1:250.000)

theme: *Theme*(
name: VegetaçãoUC
attributes: UnidadeConservação, Vegetação, Área
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)
inverseScale: 1:250.000)

theme: *Theme*(
name: GeomorfologiaUc
attributes: UnidadeConservação, Geomorfologia, Área
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)
inverseScale: 1:250.000)

theme: *Theme*(
name: SoloUC
attributes: UnidadeConservação, Solo, Área
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)
inverseScale: 1:250.000)

theme: *Theme*(
name: Hidrografia
attributes: Código, Nome, Perímetro, Tipo
geometry: Linha
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)
inverseScale: 1:250.000)

theme: *Theme*(
name: AtividadeEconômica
attributes: Atividade, Município, Ano, NumEmpregados, NumEstabelecimento
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)
inverseScale: 1:250.000)

theme: *Theme*(
name: PontoMonitoramentoAr
attributes: Estação, Município, Endereço, Latitude, Longitude, Elevação
geometry: Ponto
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)
inverseScale: 1:250.000)

theme: *Theme*(
name: DadosMeteorológicos
attributes: Estado, Município, CódigoEstação, Clima, DataMedição, HoraMedição,
DireçãoVentos, UmidadeRelativa, UmidadeMáx, UmidadeMín, TemperaturaMáx,
TemperaturaMín, Pressão, RadiaçãoSolar, InversãoTérmica
geometry: Ponto
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba)
inverseScale: 1:250.000)

theme: *Theme*(
name: PgmInvestimento
attributes: Programa, Nome, Município, Valor, Início
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul)
inverseScale: 1:250.000)

theme: *Theme*(

name: Usuário
attributes: Código, Nome, Município, VazãoCaptação, EfluentesLançados
geometry: Ponto
regions: List (Bacia Hidrográfica do Rio Paraíba do Sul)
inverseScale: 1:250.000)

theme: Theme(
name: Indústria
attributes: Código, CódigoU, NomeU, Município, VazãoCaptação, EfluentesLançados,
ProcessoProdutivo, NumEmpregados, ConsumoEnergia
geometry: Ponto
regions: List (Bacia Hidrográfica do Rio Paraíba do Sul)
inverseScale: 1:250.000)

theme: Theme(
name: IndústriaCadPROCON
attributes: Código, CódigoI, CódigoU, NomeU, Município, VazãoCaptação, EfluentesLançados,
ProcessoProdutivo, NumEmpregados, ConsumoEnergia, Poluentes, ConsumoÁgua
geometry: Ponto
regions: List (Bacia Hidrográfica do Rio Paraíba do Sul)
inverseScale: 1:250.000)

theme: Theme(
name: DadosHidrológicos
attributes: Hidrografia, BaciaHidrográfica, AlturaRegua, Vazão, TipoLeitura
geometry: Ponto
regions: List (Bacia Hidrográfica do Rio Paraíba do Sul)
inverseScale: 1:250.000)

theme: Theme(
name: DadosSedimentologia
attributes: Código, Hidrografia, DescaragaSólida, TipoSedimentoFundo,
TipoSedimentoSuspensão
geometry: Ponto
regions: List (Bacia Hidrográfica do Rio Paraíba do Sul)
inverseScale: 1:250.000)

theme: Theme(
name: QualidadeÁgua
attributes: Ponto, Hidrografia, PH, Temperatura, Condutividade, OxigênioDissolvido,
DemanBioquimOxigênio, DemanQuimiOxigênio, Fosforo, Nitrogênio, Metais,
Coliformes, CompostoOrganico
geometry: Ponto
regions: List (Bacia Hidrográfica do Rio Paraíba do Sul)
inverseScale: 1:250.000)

theme: Theme(
name: Altimetria
attributes: Cota
geometry: Linha
regions: List (Bacia Hidrográfica do Rio Paraíba do Sul)
inverseScale: 1:250.000)

theme: Theme(
name: UsoDoSolo
attributes: Uso, CustoPorHectare, ValorMonetário,
geometry: Área
regions: List (Bacia Hidrográfica do Rio Paraíba do Sul)
inverseScale: 1:250.000)

theme: Theme(
name:

name: Vegetação
 attributes: Tipo, Município
 geometry: Área
 regions: List (Bacia Hidrográfica do Rio Paraíba do Sul)
 inverseScale: 1:250.000)

theme: Theme(
 name: Relevo
 attributes: Tipo, Município, Pedregosidade
 geometry: Área
 regions: List (Bacia Hidrográfica do Rio Paraíba do Sul)
 inverseScale: 1:250.000)

theme: Theme(
 name: Solo
 attributes: Tipo, Profundidade, Estrutura, Município
 geometry: Área
 regions: List (Bacia Hidrográfica do Rio Paraíba do Sul)
 inverseScale: 1:250.000)

B.7. Integração do Esquema Ambiental.abs12 ao esquema resultante da integração de (Paraíba.abs1 e 2, RHidro.abs1)

Na análise realizada na seção B.3.2 foi observado que o esquema Ambiental, apesar de tratar do mesmo espaço geográfico, possui área de aplicação diferente dos esquemas Paraíba.abs(1 e 2) e RHidro.abs1. Deste modo, foi realizado primeiro a integração destes esquemas, e ao esquema resultante desta integração será aplicada a  MultiSIG, conforme as seções a seguir.

Para facilitar, o esquema resultante da integração dos esquemas Paraíba.abs(1 e 2) e RHidro.abs1, será denominado de **Integração.abs**, visto que a abstração cartográfica resultante da integração está em 1:250.000.

B.7.1. Análise do contexto semântico dos esquemas (Integração.abs, Ambiental.abs1)

B.7.1.1. Análise das denominações

Esta atividade objetiva identificar as similaridades e diferenças entre as classes dos esquemas, em relação a sua semântica no mundo real.

Tabela B.14 - Tabela de Classes (TC).

INTEGRAÇÃO.ABS			AMBIENTAL.ABS1		
CLASSE _i	LISTA(TERMO _{ik})	TIPO	CLASSE _j	LISTA(TERMO _{jk})	TIPO
			ÓRGÃOEXECUTOR	Órgão executor do projeto	AO
			ÁREAPROJETO	Projeto	GGO

			PROJETO	Área de Projeto	AO
ESTADO	Estado	GGO			
MUNICÍPIO250	Limites municipais	GGO	MUNICÍPIO	Limites municipais	GGO
			DIVISÃO MUNICIPAL	Limites municipais com cidade e infra-estrutura	GGO
CIDADES	Todas as cidades	GGO	CIDADES	Local. do centro adm das cidades	
SEDEMUNICÍPIO	Sede dos municípios	GGO			GGO
UNIDADECONSERVAÇÃO	Unidade de conservação	GGO			
DISTRITO	Distrito	GGO			
SUBDISTRITO	Subdistrito	AO			
SETORCENSITÁRIO	Sector censitário	AO			
ATOLEGAL	Ato legal da unidade de conservação e dos municípios	AO			
BACIAHIDROGRÁFICA	Bacia hidrográfica	GGO	BACIAHIDROGRÁFICA	Limite de bacia hidrográfica	GGO
			MAPAS BACIAS HIDROGRÁFICAS	Bacias hidrográficas com os rios, Áreas de projeto e indústrias	GGO
SUBBACIA	Sub bacia hidrográfica	GGO			
SUBBACIAÚRBANA	Sub bacia urbana	GGO			
MICROBACIA	Microbacia	GGO			
MICROBACIAÚRBANA	Microbacia urbana	GGO			
HIDROGRAFIA	Hidrografia	GGO	REDEDEDRENAGEM	Conjunto de rios da região	GGO
			RIOS	Hidrografia	GGO
			INFRA-ESTRUTURA	Infra-estrutura de transporte	GGO
TERMINALFERROVIÁRIO	Terminal ferroviário	GGO			
FERROVIA	Ferrovias	GGO	FERROVIA	Ferrovias	GGO
RODOVIA	Rodovias	GGO	RODOVIA	Rodovias	GGO
GEOLOGIAUC	Geologia da unidade de conservação	AO			
VEGETAÇÃOUC	Vegetação da unidade de conservação	AO			
GEOMORFOLOGIAUC	Geomorfologia da unidade de conservação	AO			
SOLOUC	Solo da unidade de conservação	AO			
ATIVIDADEECONÔMICA	Estabelecimento com atividade econômica	AO			
PONTO MONITORAMENTO AR	Dados da estação de monit. do ar	GGO			
DADOSMETEOROLÓGICOS	Dados meteorológicos	GGO			
PGM INVESTIMENTO	Pgm de investimento	AO			
USUÁRIO	Estabelecimento com atividade econômica	GGO			
INDÚSTRIA	Tipo de estab. Econômico	GGO	INDÚSTRIA	Indústrias	GGO
INDÚSTRIACADPROCON	Indústria cadastrada no ProCON	GGO			
DADOSHIDROLÓGICOS	Dados hidrológicos	GGO			
DADOSSEDIMENTOLOGIA	Dados de sedimentologia	GGO			
QUALIDADEÁGUA	Dados de qualidade d'água	GGO			
ALTIMETRIA	Curvas de nível	GGO	ALTIMETRIA	Curvas de nível	GGO
USODO SOLO	Class. uso do solo	GGO	USODATERRA	Tipo de uso da terra	GGO
VEGETAÇÃO	Class. vegetação	GGO	VEGETAÇÃO	Tipo de vegetação	GGO
RELEVO	Class. relevo	GGO	RELEVO	Tipo de relevo	GGO
SOLO	Class. solo	GGO	SOLO	Tipo de solo	GGO

			FLORESTAS	Tipo de floresta	GGO
			CLIMA	Tipo de clima	GGO
			DECLIVIDADE	Declividade	GGO
			PONTOSALTIMÉTRICOS	Pontos cotados da região	GGO
			IMAGEM LANDSAT	Imagem do satélite Landsat	GGO

Esta análise resulta na distribuição das classes dos esquemas nos seguintes grupos temáticos:

DIVISÕES DO ESPAÇO GEOGRÁFICO	
INTEGRAÇÃO.ABS	AMBIENTAL.ABS
Estado::Integração.abs Município250::Integração.abs Cidades::Integração.abs SedeMunicípio::Integração.abs UnidadeConservação::Integração.abs Distrito::Integração.abs Subdistrito::Integração.abs SetorCensitário::Integração.abs AtoLegal::Integração.abs	Município::Ambiental.abs1 DivisãoMunicipal::Ambiental.abs1 Cidades::Ambiental.abs1

ÁREA DE ESTUDO DA HIDROGRAFIA	
INTEGRAÇÃO	AMBIENTAL
BaciaHidrográfica::Integração.abs SubBaciaIntegração.abs SubBaciaUrbana::Integração.abs Microbacia::Integração.abs MicroBaciaUrbana::Integração.abs Hidrografia::Integração.abs	BaciaHidrográfica::Ambiental.abs1 MapasBaciasHidrográficas::Ambiental.abs1 RedeDeDrenagem::Ambiental.abs1 Rio::Ambiental.abs1

INFRA-ESTRUTURA	
INTEGRAÇÃO	AMBIENTAL
Ferrovia::Integração.abs TerminalFerroviário::Integração.abs Rodovia::Integração.abs	InfraEstrutura::Ambiental.abs1 Ferrovia::Ambiental.abs1 Rodovia::Ambiental.abs1

DADOS DO MEIO AMBIENTE	
INTEGRAÇÃO	AMBIENTAL
GeologiaUC::Integração.abs VegetaçãoUC::Integração.abs GeomorfologiaUC::Integração.abs SoloUC::Integração.abs Altimetria::Integração.abs UsoDoSolo::Integração.abs Vegetação::Integração.abs Relevo::Integração.abs Solo::Integração.abs	Altimetria::Ambiental.abs1 PontosAltimétricos::Ambiental.abs1 UsoDaTerra::Ambiental.abs1 Vegetação::Ambiental.abs1 Relevo::Ambiental.abs1 Solos::Ambiental.abs1 Florestas::Ambiental.abs1 Declividade::Ambiental.abs1 Clima::Ambiental.abs1 ImagemLandsat::Ambiental.abs1

DADOS DE ATIVIDADES ECONÔMICAS	
INTEGRAÇÃO	AMBIENTAL
AtividadeEconômica::Integração.abs Usuário::Integração.abs Indústria::Integração.abs IndústriaCadPROCON::Integração.abs	Indústria::Ambiental.abs1

DADOS DE CAMPO	
INTEGRAÇÃO	AMBIENTAL
PontoMonitoramentoAr::Integração.abs DadosMeteorológicos::Integração.abs DadosHidrológicos::Integração.abs DadosSedimentologia::Integração.abs QualidadeÁgua::Integração.abs	

GERENCIAMENTO DE PROJETO	
INTEGRAÇÃO	AMBIENTAL
PgmInvestimento::Integração.abs	OrgãoExecutor::Ambiental.abs1 ÁreaProjeto::Ambiental.abs1 Projeto::Ambiental.abs1

B.7.1.2. Análise dos relacionamentos semânticos

Esta etapa consiste em analisar para cada classe os relacionamentos, dentro do grupo temático e em relação aos outros grupos temáticos. De maneira a abreviar, esta análise será restringida as classes que possuem proximidade semântica dentro de cada grupo temático.

Tabela B.15 - Tabela de Relacionamentos (TR).

DIVISÕES DO ESPAÇO GEOGRÁFICO									
INTEGRAÇÃO					AMBIENTAL				
REL _{mi}	C _{me}	C _{mf}	CAR _{mi}	AT. DO REL _{mi}	REL _{nj}	C _{nk}	C _{nl}	CAR _{nj}	AT. DO REL _{nj}
PERTENCE	Município250	Estado			CONTÉM	Divisão Municipal	Município	1:n	
					CONTÉM	Divisão Municipal	Infra Estrutura	1:n	
					CONTÉM	Divisão Municipal	Cidades	1:n	
					CONTÉM	Divisão Municipal	Rios	1:n	
PERTENCE	Município250	Estado	1:1	estado					
PERTENCE	Cidades	Estado	1:1	estado					
PERTENCE	Cidades	Município250	1:1	município	PERTENCE	Cidades	Município	1:1	Códigomum
PERTENCE	Cidades	Distrito	1:1	distrito					
PERTENCE	Sede Município	Estado	1:1	estado					
PERTENCE	Sede Município	Município250	1:1	município					
PERTENCE	Sede Município	Distrito	1:1	distrito					

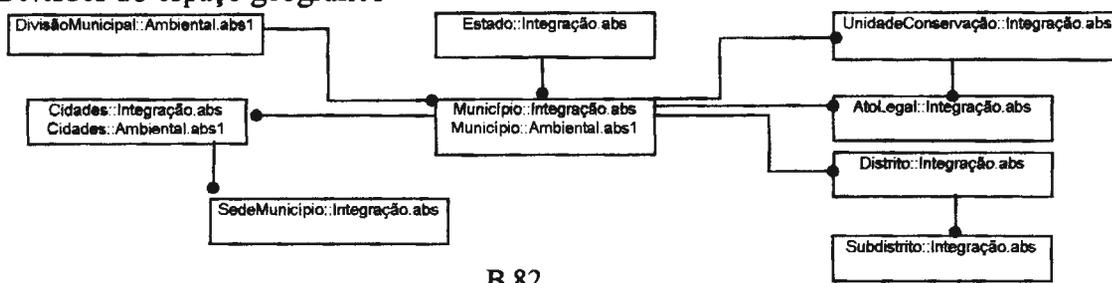
PERTENCE	Unidade Conservação	Município250	1:n	município					
PERTENCE	Distrito	Município250	1:1	município					
PERTENCE	Subdistrito	Distrito	1:1	distrito					
PERTENCE	Subdistrito	Município250	1:1	município					
PERTENCE	Setor Censitário	Município250		município					
PERTENCE	Setor Censitário	Distrito		distrito					
PERTENCE	Setor Censitário	SubDistrito		subdistrito					
DESENVOLVE	Município250	Atividade Econômica	1:n	município					
PERTENCE	AtoLegal	Município250		município					
PERTENCE	Pgm Investimento	Município250	1:n	município					
ÁREA DE ESTUDO DA HIDROGRAFIA									
INTEGRAÇÃO					AMBIENTAL				
REL _{mi}	C _{me}	C _{mf}	CAR _{mi}	AT. DO REL _{mi}	REL _{nj}	C _{nk}	C _{nj}	CAR _{nj}	AT. DO REL _{nj}
					CONTÉM	MapaBacia Hidrográfica	Rios	n:m	
					CONTÉM	MapaBacia Hidrográfica	AreaProjeto	n:m	
					CONTÉM	MapaBacia Hidrográfica	Bacia Hidrográfica	n:m	
					CONTÉM	MapaBacia Hidrográfica	Indústrias	1:n	
					POSSUI	Rede Drenagem	Rios	1:n	
					POSSUI	Rios	TipoDren	1:1	
					POSSUI	Rios	Classdren	1:1	
PERTENCE	SubBacia Hidrográfica	Bacia Hidrográfica	1:n	Bacia Hidrográfica					
SE ESPECIALIZA	SubBacia Hidrográfica	SubBacia Urbana	1:n	CódigoSB					
PERTENCE	SubBacia Urbana	Bacia Hidrográfica		Bacia Hidrográfica					
CONTÉM	Microbacia	SubBacia	1:n	SubBacia					
SE ESPECIALIZA	Microbacia	Microbacia Urbana	1:n	Cód Microbacia					
POSSUI	SubBacia Hidrográfica	RHSub Bacia	1:n	CódSub Bacia					
INFRA-ESTRUTURA									
INTEGRAÇÃO					AMBIENTAL				
REL _{mi}	C _{me}	C _{mf}	CAR _{mi}	AT. DO REL _{mi}	REL _{nj}	C _{nk}	C _{nj}	CAR _{nj}	AT. DO REL _{nj}
					CONTÉM	Divisão Municipal	Infra Estrutura		
					SE ESPECIALIZA	Infra Estrutura	Ferrovia		
					SE ESPECIALIZA	Infra Estrutura	Rodovia		
PERTENCE	Terminal Ferroviário	Ferrovia	1:n	Ferrovia					
DADOS DO MEIO AMBIENTE									
INTEGRAÇÃO					AMBIENTAL				
REL _{mi}	C _{me}	C _{mf}	CAR _{mi}	AT. DO REL _{mi}	REL _{nj}	C _{nk}	C _{nj}	CAR _{nj}	AT. DO REL _{nj}
PERTENCE	GeologiaUC	Unidade Conservação	1:n	CódUc					
POSSUI	GeologiaUC	Geologia	1:1	Cód Geologia					
PERTENCE	TipoVegetUC	Unidade Conservação	1:n	CódUc					
POSSUI	TipoVegetUC	TipoVegetação	1:1	Cód Vegetação					
PERTENCE	Geomorfologia UC	Unidade Conservação	1:n	CódUc					
POSSUI	Geomorfologia UC	Geomorfologia	1:1	Cód Geomorfo					
PERTENCE	SoloUC	Unidade Conservação	1:n	CódUc					

POSSUI	SoloUC	Classificação Solo	1:1	Cód TipoSolo					
					POSSUI	ÁreaProjeto	Solos		
					TEM	Solos	TipoSolo	1:1	Tipo
					POSSUI	ÁreaProjeto	Vegetação		
					TEM	Vegetação	TipoVegetação	1:1	Tipo
					POSSUI	ÁreaProjeto	Florestas		
					TEM	Florestas	TipoFloresta	1:1	Tipo
					POSSUI	ÁreaProjeto	UsoDaTerra		
					TEM	UsoDaTerra	TipoUso	1:1	Tipo
					POSSUI	ÁreaProjeto	Relevo		
					TEM	Relevo	TipoRelevo	1:1	Tipo
					POSSUI	ÁreaProjeto	Clima		
					TEM	Clima	TipoClima	1:1	Tipo
POSSUI	Município250	Solo							
POSSUI	Município250	UsoDoSolo							
POSSUI	Município250	Vegetação							
POSSUI	Município250	Relevo							
POSSUI	Município250	Altimetria							
DADOS DE ATIVIDADES ECONOMICAS									
INTEGRAÇÃO					AMBIENTAL				
REL _{mi}	C _{me}	C _{mf}	CAR _{mi}	AT. DO REL _{mi}	REL _{ni}	C _{nk}	C _{nl}	CAR _{ni}	AT. DO REL _{ni}
					CONTÉM	MapaBacia Hidrográfica	Indústrias	1:n	
					POSSUI	Indústrias	TpoInd	1:n	tipo
					POSSUI	Indústrias	Poluição	1:n	Graupol
PERTENCE	Pgm Investimento	Município250							
GERENCIAMENTO DE PROJETO									
INTEGRAÇÃO					AMBIENTAL				
REL _{mi}	C _{me}	C _{mf}	CAR _{mi}	AT. DO REL _{mi}	REL _{ni}	C _{nk}	C _{nl}	CAR _{ni}	AT. DO REL _{ni}
					EXECUTA	Orgão Executor	Projeto	n:m	Códigoor
					POSSUI	Projeto	ÁreaProjeto	1:1	Códigopr
					PERTENCE	Projeto	Bacia Hidrográfica	1:1	Códigobh
					CONTÉM	ÁreaProjeto	Florestas	1:n	
						ÁreaProjeto	Solos	1:n	Códigopr
						ÁreaProjeto	Vegetação	1:n	Códigopr
						ÁreaProjeto	Relevo	1:n	Códigopr
						ÁreaProjeto	Clima	1:n	Códigopr
						ÁreaProjeto	UsoDaTerra	1:n	Códigopr
						ÁreaProjeto	Altimetria	1:n	Códigopr
						ÁreaProjeto	Declividade		
						ÁreaProjeto	Pontos Altimétricos		
						ÁreaProjeto	Imagem Landsat		

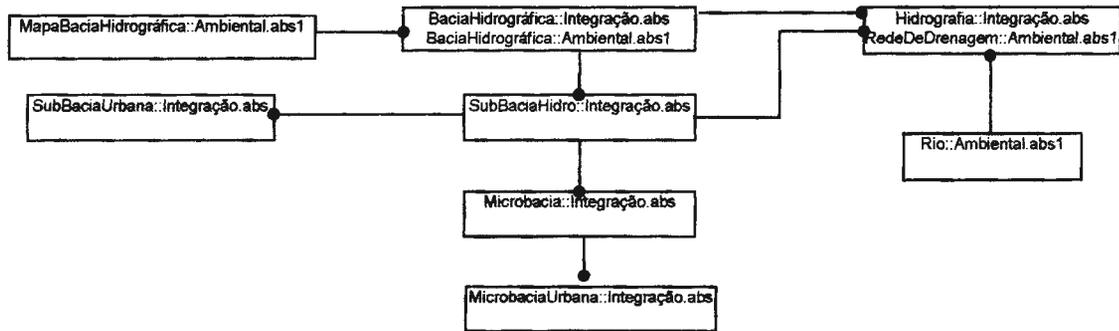
Tendo em vista que o grupo temático DadosDeCampo possui classes resultantes da integração dos esquemas Paraíba e RHidro, as classes destes grupos serão consideradas heterogêneas em relação ao esquema Ambiental.abs1.

Esta análise resulta na modelagem preliminar da integração do par de esquemas, apresentada nas figuras esquemáticas abaixo, onde —●— representa que existe algum relacionamento:

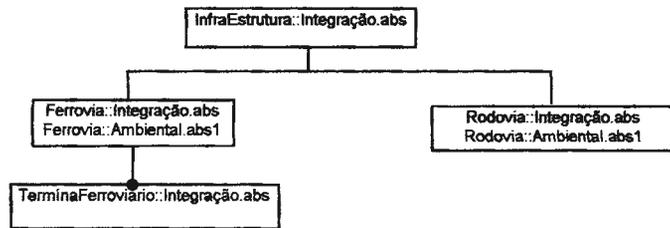
• Divisões do espaço geográfico



- Hidrografia

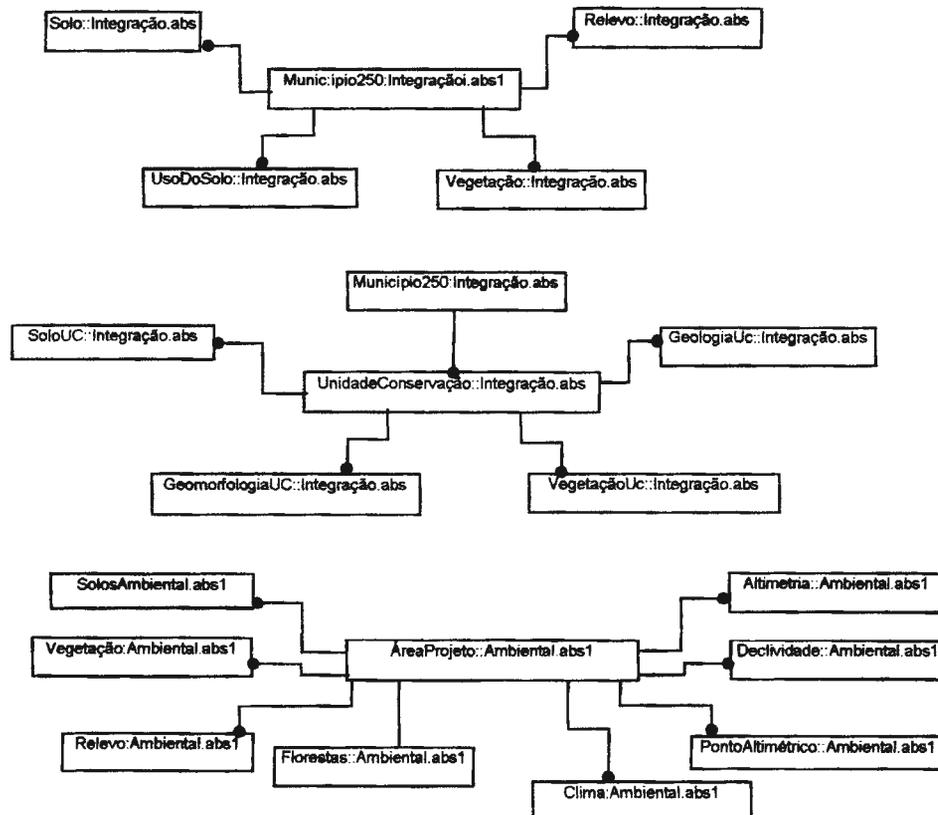


- Infra-estrutura



- Meio ambiente

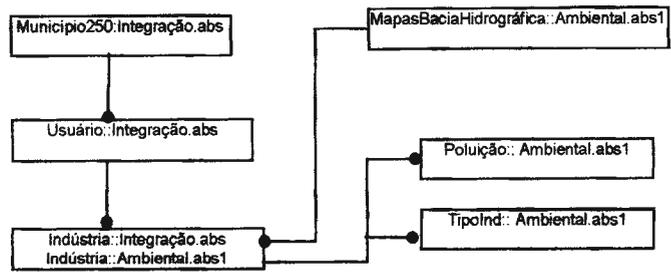
Neste grupo temático observa-se que a granularidade da informação temática é diferente nos dois esquemas.



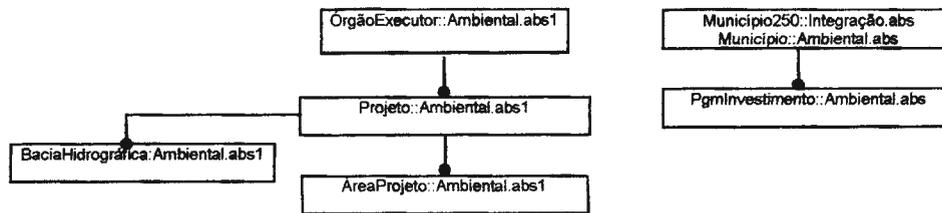
Assim, as classes Solos::Ambiental.abs1, Vegetação::Ambiental.abs1, Florestas::Ambiental.abs1, Relevo::Ambiental.abs1, Clima::Ambiental.abs1, Declividade::Ambiental.abs1, PontoAltimétrico::Ambiental.abs1 e

Altimetria::Ambienta.abs1, que possuem granularidade de informação diferente, não serão integradas, apesar de possuírem semântica aproximada.

- **Dados de atividades econômicas**



- Gerenciamento de projeto



B.7.1.3. Análise das geometrias das classes

Esta análise objetiva a identificação das geometrias utilizadas para representar os objetos geográficos que possuem representação espacial.

Tabela B.16 - Tabela de Geometrias (TG).

INTEGRAÇÃO		AMBIENTAL	
C _{int}	GEO _{int}	C _{amb}	GEO _{amb}
Estado	VectorÁrea		
Município250	VectorÁrea	Município	VectorÁrea
		DivisãoMunicipal	Point VectorLine VectorÁrea
Distrito	Point		
UnidadeConservação	VectorÁrea		
Cidades	Point	Cidades	Point
SedeMunicípio	Point		
		InfraEstrutura	VectorLine
Ferrovia	VectorLine	Ferrovia	VectorLine
Rodovia	VectorLine	Rodovia	VectorLine
TerminalFerroviário	Point		
BaciaHidrográfica	VectorÁrea	BaciaHidrográfica	VectorÁrea
		MapasBacias Hidrográficas	Point VectorLine VectorÁrea
SubBacia	VectorÁrea		
SubBaciaUrbana	VectorÁrea		
Microbacia	VectorÁrea		
MicroBaciaUrbana	VectorÁrea		
Hidrografia	VectorLine	Rios	VectorLine
		RedeDrenagem	VectorLine

Usuário	Point		
Indústria	Point	Indústria	Point
IndústriaCadaProcon	Point		
PontoMonitoraAr	Point		
DadosHidrológicos	Point		
DadosMeteorológicos	Point		
QualidadeAgua	Point		
DadosSedimentologia	Point		
UsoSolo	VectorÁrea		
Vegetação	VectorÁrea		
Relevo	VectorÁrea		
Solo	VectorÁrea		
Altimetria	VectorLine	Altimetria	Isoline
		UsoDaTerra	CellGrid2D
		vegetação	CellGrid2D
		Relevo	CellGrid2D
		Solo	CellGrid2D
		Floresta	CellGrid2D
		Clima	CellGrid2D
		Declividade	CellGrid2D
		PontosAltimétricos	DEMPoint
		ImagemLandsat	ImageGrid2D

B.7.1.4. Identificação das propriedades comuns

Esta análise tem por objetivo identificar as propriedades comuns entre as classes de cada grupo temático para as quais foi verificada alguma proximidade semântica nas atividades anteriores.

De modo a ilustrar a atribuição dos graus de proximidade semântica, estes são calculados e apresentados a seguir com as Tabelas Não Normalizadas das Propriedades.

- Divisões do espaço geográfico

MUNICÍPIO250	Código	Nome	Estado	Mesoregião	Microregião	Área	População	Anoinstituição
MUNICÍPIO								
Códigos			X					
Nomemun		X						
Códigomun	X							
Áreaof						X		
Perímetro								

$$GPS=GT + GR + GG + Gpi = 4 + 2.1/8 + 2 + 2.4/9 = 7,14$$

CIDADES	Código	Nome	Estado	Município	Distrito	População	Tipo
CIDADES							
Nome		X					
Código	X						
Códigomun				X			
Área							
permetro							

$$GPS=GT + GR + GG + Gpi = 4 + 2.1/5 + 2 + 2.3/9 = 7,15$$

- Área de estudo da hidrografia

BACIAHIDROGRÁFICA	Código	Nome	Área
REDEDRENAGEM			
Nome		X	
CódigoBh	X		
Área			X
Perímetro			

$$GPS=GT + GR + GG + Gpi = 4 + 2. 1/3 + 2+ 2.3/4 = 8,0$$

HIDROGRAFIA	Código	Nome	Perímetro	Tipo
RIOS				
Nome		X		
C Dren				
T dren				X

$$GPS=GT + GR + GG + Gpi = 4 + 2.1/3 + 2 + 2.2/4 = 7,67$$

- Infra estrutura

FERROVIA	Código	Nome
FERROVIA		
Bitola		
Nome		X
Código	X	
Origem		
Destino		

$$GPS=GT + GR + GG + Gpi = 4+2.0/2 + 2 + 2.2/5 = 6,8$$

RODOVIA	Código	Nome	Jurisdição	GrauConservação
RODOVIA				
Pistas				
Nome		X		
Código	X			
Origem				
Destino				

$$GPS=GT + GR + GG + Gpi = 4+ 2.0/1 +2+2.2/7 = 6,57$$

- Dados de atividades econômicas

INDÚSTRIAS	Código	CódigoU	NomeU	Município	Vazão	Efluentes	Processo	Num	Consumo
INDÚSTRIA					Captação	Lançados	Produtivo	Empregados	Energia
Nome			X						
Códigoor		X							
Tipo									
Graupol									

$$GPS=GT + GR + GG + Gpi = 4+ 2.0/1 +2+2.2/11 = 6,18$$

B.7.1.5. Análise das descrições das propriedades

Esta atividade tem por finalidade comparar as propriedades identificadas nas tabelas anteriores que possuem alguma proximidade. Esta comparação avalia a

definição das propriedades de modo a identificar a necessidade de mapeamentos para conformá-las.

Na tabela abaixo, as propriedades de códigos de unidades territoriais do par de esquemas são os códigos normalizados pelo IBGE, bem como as Áreas também são as fornecidas pelo IBGE.

Tabela B.17 - Tabela de Correspondência entre Propriedades (TCP).

DIVISÕES DO ESPAÇO GEOGRÁFICO									
MUNICÍPIO250::INTEGRAÇÃO.ABS					MUNICÍPIO::AMBIENTAL.ABS1				
ATRIBUTO	TIPO	FORMATO	RESTRIÇÃO	DOMÍNIO	ATRIBUTO	TIPO	FORMATO	RESTRIÇÃO	DOMÍNIO
Código					Códigomun	Integer	5		
Nome					Nomemun	String	30		
Estado					Códigos	Integer	2		
Mesoregião									
Microregião									
Área									
População									
AnoInstituição									
					Áreaof	Real	8,3		
					Perímetro	real	8,3		
CIDADES::INTEGRAÇÃO.ABS2					CIDADE::AMBIENTAL.ABS1				
ATRIBUTO	TIPO	FORMATO	RESTRIÇÃO	DOMÍNIO	ATRIBUTO	TIPO	FORMATO	RESTRIÇÃO	DOMÍNIO
Código					Código				
Nome					Nome				
Estado									
Município					Códigomun				
Distrito									
População									
Tipo									
					Área				
					Perímetro				
ÁREA DE ESTUDO DA HIDROGRAFIA									
BACIAHIDROGRÁFICA::INTEGRAÇÃO.ABS					BACIAHIDROGRÁFICA::AMBIENTAL.ABS1				
ATRIBUTO	TIPO	FORMATO	RESTRIÇÃO	DOMÍNIO	ATRIBUTO	TIPO	FORMATO	RESTRIÇÃO	DOMÍNIO
Código					Códigobh				
Nome					Nome				
Área					Área				
					Perímetro				
HIDROGRAFIA::INTEGRAÇÃO.ABS					RIO::AMBIENTAL.ABS1				
ATRIBUTO	TIPO	FORMATO	RESTRIÇÃO	DOMÍNIO	ATRIBUTO	TIPO	FORMATO	RESTRIÇÃO	DOMÍNIO
Código									
Nome					Nome				
Perímetro									
Tipo									
					Cdren				
					Tdren				
INFRA ESTRUTURA									
FERROVIA::INTEGRAÇÃO.ABS					FERROVIA::AMBIENTAL.ABS1				
ATRIBUTO	TIPO	FORMATO	RESTRIÇÃO	DOMÍNIO	ATRIBUTO	TIPO	FORMATO	RESTRIÇÃO	DOMÍNIO
Nome					Nome				
Linha					Código				
					Origem				
					Destino				
					Bitola				
RODOVIA::INTEGRAÇÃO.ABS2					RODOVIA::AMBIENTAL.ABS1				
ATRIBUTO	TIPO	FORMATO	RESTRIÇÃO	DOMÍNIO	ATRIBUTO	TIPO	FORMATO	RESTRIÇÃO	DOMÍNIO
Código					Código				
Nome					Nome				
Jurisdição									
GrauConservação									
					Origem				
					Destino				
					Pistas				
DADOS DE ATIVIDADES ECONÔMICAS									

INDÚSTRIAS:: INTEGRAÇÃO.ABS					INDÚSTRIAS:: AMBIENTAL.ABS1				
ATRIBUTO	TIPO	FORMATO	RESTRIÇÃO	DOMÍNIO	ATRIBUTO	TIPO	FORMATO	RESTRIÇÃO	DOMÍNIO
Código					Códigooor				
CódigoU					Nome				
NomeU					Tipo				
					Grupol				
Município									
Vazão captação									
Efluentes Lançados									
Processo Produtivo									
Num Empregados									
Consumo Energia									
Poluentes									
Consumo Água									

B.7.1.6. Tratamento do relacionamento semântico encontrados

Esta etapa consiste em separar as classes do par de esquemas nos grupos de proximidade semântica abaixo:

- *Grupo de classes idênticas*

Não existem pares de classes similares entre os esquemas Integração.abs1 e Ambiental.abs1.

Tabela B.18 - Tabela de Classes Idênticas.

INTEGRAÇÃO.ABS		AMBIENTAL.ABS1		
CLASSES	LISTA(PROPRIEDADES)	CLASSES	LISTA(PROPRIEDADES)	GPS

- *Grupo de classes similares*

Tabela B.19 - Tabela de Classes Similares.

INTEGRAÇÃO.ABS			AMBIENTAL.ABS1		
CLASSES	PROPRIEDADES	MAPEAMENTOS	CLASSE	PROPRIEDADES	GPS
Município	Código Estado Nome Mesoregião Microregião Área População AnoInstituição		Munic	Códigomun Códigos Nomemun Áreaof Perimetro	7,52

Cidade	Código Nome Estado Município Distrito População Tipo		Cidades	Código Nome Códigomun Área Perímetro	6,88
BaciaHidrográfica	Código Nome Área		BaciaHidrográfica	Nome Códigobh Área Perímetro	8
Hidrografia	Código Nome Perímetro Tipo		Rios	Nome Cdren Tdren	7
Rodovia	Código Nome Jurisdição GrauConservação		Rodovia	Nome Código Origem Destino Pistas	
Ferrovia	Nome Linha		Ferrovia	Nome Código Origem Destino Bitola	
Indústrias	Código CódigoU NomeU Município VazãoCaptação EfluentesLançados ProcessoProdutivo NumEmpregados ConsumoEnergia Poluentes ConsumoÁgua		Indústrias	Nome Códigooor Tipo Graupol	

- *Grupo de classes heterogêneas*

Tabela B.20 - Tabela de Classes Heterogêneas do esquema Ambiental.

AMBIENTAL.ABS1	
CLASSES	PROPRIEDADES
Órgão Executor	Nomeor, Códigoor, endereço, vinculação
Projeto	Códigobh, Códigopr, nomepr, Códigooor
ÁreaProjeto	Códigobh, Códigopr, Códigooor, Área, Perímetro
Solos	Tipo, profund, pors, permea, textura
Florestas	Tipo, altura, densidade
Declividade	Grau
Vegetação	Tipo
Relevo	Tipo
Clima	Tipo, preci, temp, umi
UsoDaTerra	Tipo, adap
Altimetria	Altitude
PontoAltimétrico	Cota
RedeDrenagem	Número, microregião, escala, ano
DivisãoMunicipal	Número, microregião, escala, ano
MapasBaciasHidrográficas	Número, Códigobh
ImagemLandsat	

Tabela B.21 - Tabela de Classes Heterogêneas do esquema Integração.

INTEGRAÇÃO.ABS	
CLASSES	PROPRIEDADES
Estado	Código, Nome, Área
SedeMunicípio	Código, Nome, Estado, Município, Distrito, População, Tipo, Latitude, Longitude
UnidadeConservação	Código, Município, Criação, Administração, NivelAdministrativo, Manejo, Área
Distrito	Código, Nome, Município, Altitude, População
SubDistrito	Código, Nome, Distrito, Município
SetorCentitário	Código, Município, Distrito, SubDistrito
AtoLegal	Código, Município, Número, Tipo, DataPublicação, Escopo
SubBacia	Código, Nome, BaciaHidrográfica
SubBaciaUrbana	Código, CódigoSB, NomeSB, BaciaHidrográfica, Qualificação, Ocupação, OcupaçãoUrbana
Microbacia	Código, Nome, SubBacia
MicrobaciaUrbana	Código, CódigoMB, NomeMB, SubBacia, Qualificação, Ocupação, OcupaçãoUrbana
TerminalFerroviário	Ferrovia, Estação
GeologiaUc	UnidadeConservação, Geologia, Área
VegetaçãoUC	UnidadeConservação, Vegetação, Área
GeomorfologiaUc	UnidadeConservação, Geomorfologia, Área
SoloUC	UnidadeConservação, Solo, Área
AtividadeEconômica	Atividade, Município, Ano, NumEmpregados, NumEstabelecimento
PontoMonitoramentoAr	Estação, Município, Endereço, Latitude, Longitude, Elevação
DadosMeteorológicos	Estado, Município, CódigoEstação, Clima, DataMedição, HoraMedição, DireçãoVentos, UmidadeRelativa, UmidadeMáx, UmidadeMín, TemperaturaMáx, TemperaturaMín, Pressão, RadiaçãoSolar, InversãoTérmica
PgmInvestimento	Programa, Nome, Município, Valor, Início
Usuário	Código, Nome, Município, VazãoCaptação, EfluentesLançados
IndústriaCadPROCON	Código, CódigoI, CódigoU, NomeU, Município, VazãoCaptação, EfluentesLançados, ProcessoProdutivo, NumEmpregados, ConsumoEnergia, Poluentes, ConsumoAgua
DadosHidrológicos	Hidrografia, BaciaHidrográfica, AlturaRegua, Vazão, TipoLeitura
DadosSedimentologia	Código, Hidrografia, DescargaSólida, TipoSedimentoFundo, TipoSedimentoSuspensão
QualidadeÁgua	Ponto, Hidrografia, PH, Temperatura, Condutividade, OxigênioDissolvido, DemanBioquimOxigênio, DemanQuimiOxigênio, Fosforo, Nitrogênio, Metais, Coliformes, CompostoOrganico
Altimetria	Cota
UsoDoSolo	Uso, CustoPorHectare, ValorMonetário
Vegetação	Tipo, Município
Relevo	Tipo, Município, Pedregosidade
Solo	Tipo, Profundidade, Estrutura, Município

B.7.2. Conformação

Nesta etapa são tratados os conflitos para as classes que apresentam similaridade e definido os mapeamentos entre elas. Esta atividade é realizada apenas nos pares de classes que se encontram no grupo de classes similares dos esquemas Integração.abs(Paraíba.abs2, Rhidro.abs1) e Ambiental.abs1.

Tabela B.22 - Tabela de Classes Similares.

INTEGRAÇÃO.ABS2			AMBIENTAL.ABS1	
CLASSES	PROPRIEDADES	MAPEAMENTOS	CLASSE	PROPRIEDADES
Município	Código Estado Nome Mesoregião Microregião Área População AnoInstituição		Munic	Códigomun Códigos Nomemun Áreaof Perímetro

Cidade	Código Nome Estado Município Distrito População Tipo		Cidades	Código Nome Códigomun Área Perímetro
BaciaHidrográfica	Código Nome Área		BaciaHidrográfica	Nome Códigobh Área Perímetro
Hidrografia	Código Nome Perímetro Tipo		Rio	Nome Cdren tdren
Rodovia	Código Nome Jurisdição GrauConservação			Nome Código Origem Destino Pistas
Ferrovia	Nome Linha			Nome Código Origem Destino Bitola
Indústrias	Código CódigoU NomeU Município VazãoCaptação EfluentesLançados ProcessoProdutivo NumEmpregados ConsumoEnergia Poluentes ConsumoÁgua		Indústrias	Nome Códigooor Tipo Graupol

B.7.3. Integração

Esta etapa tem por objetivo criar o esquema global resultante da integração dos esquemas (Integração.abs(Paraíba.abs1e2, Rhidro.abs1), Ambiental.abs1). Elas são expressas usando a notação proposta pela *MultiSIG*.

As classes que possuem proximidade semântica e foram conformadas entre os esquemas são re-escritas as suas classes virtuais na notação proposta pela *MultiSIG* e as classes heterogêneas são também expressas na mesma notação criando novas classes virtuais para o esquema global, a saber:

Superclass:

Class: ÓrgãoExecutor

Region:

Estado do Rio de Janeiro

Attributes:

Nome, Código, Endereço, Vinculação

Source schema mapping method:

Region = Estado do Rio de Janeiro {

Nome Nome.ÓrgãoExecutor::Paraiba.abs1
 Código Códigoor.ÓrgãoExecutor::Paraiba.abs1
 Endereço Endereço.ÓrgãoExecutor::Paraiba.abs1
 Vinculação Vinculação.ÓrgãoExecutor::Paraiba.abs1 }

Export schema source:

*OriginClass1:*ÓrgãoExecutor::Ambiental.abs1
Attributes: Nomeor, Códigoor,Endereço, Vinculação
Relationships:

<i>Rel</i>	<i>ObjectA</i>	<i>ObjectB</i>	<i>Attributes</i>	<i>Car</i>
Executa	ÓrgaoExecutor	Projeto	Códigoor	1:n

Source.DBInfo.dbIdentifier:
Acess method:

Superclass:

Class: Projeto

Region:

Estado do Rio de Janeiro

Attributes: Projeto, Nome, BaciaHidrográfica, ÓrgãoExecutor

Source schema mapping method:

Region = Estado do Rio de Janeiro {
 Projeto Códigoopr.Projeto::Ambiental.abs1
 Nome Nomepr.Projeto::Ambiental.abs1
 BaciaHidrográfica Códigobh.Projeto::Ambiental.abs1
 ÓrgãoExecutor Código.Projeto::Ambiental.abs1 }

Export schema source:

*OriginClass1:*Projeto.Ambiental.abs1
Attributes: Códigoopr, Nomepr, Códigobh, Códigoor
Relationships:

<i>Rel</i>	<i>ObjectA</i>	<i>ObjectB</i>	<i>Attributes</i>	<i>Car</i>
Executa	ÓrgaoExecutor	Projeto	Códigoor	1:n
Possui	Projeto	ÁreaProjeto	Códigopr	1:n
Pertence	Projeto	BaciaHidrográfica	Códigobh	

Source.DBInfo.dbIdentifier:
Acess method:

Superclass:

Class: ÁreaProjeto

Region:

Estado do Rio de Janeiro

*Attributes:*Projeto, BaciaHidrográfica, ÓrgãoExecutor, Área, Perímetro

Geometry: Área

Source schema mapping method:

Region = Estado do Rio de Janeiro {
 Projeto Códigoopr.ÁreaProjeto::Ambiental.abs1
 BaciaHidrográfica Códigobh.ÁreaProjeto::Ambiental.abs1
 ÓrgãoExecutor Código.ÁreaProjeto::Ambiental.abs1
 Área Área.ÁreaProjeto::Ambiental.abs1
 Perímetro Perímetro.ÁreaProjeto::Ambiental.abs1 }

Export schema source:

*OriginClass1:*ÁreaProjeto.Ambiental.abs1
Attributes: Códigoopr, Códigobh, Códigoor, Área, Perímetro
Relationships:

<i>Rel</i>	<i>ObjectA</i>	<i>ObjectB</i>	<i>Attributes</i>	<i>Car</i>
Executa	MapaBacias Hidrográfica	Projeto	Códigoor	1:n
Possui	Projeto	ÁreaProjeto	Códigopr	1:n
Possui	ÁreaProjeto	Solos, Vegetação, Relevo, Florestas, Declividade, Clima, UsoDaTera,		

Source.DBInfo.dbIdentifier:
Acess method:

Superclass:

Class: Municipio

Region:

Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaiíba, Estado do Rio de Janeiro

Attributes:

Código, Estado, Nome, Mesoregião, Microregião, Área, Perímetro, População, AnoInstituição, Perímetro

Geometry: Área

Source schema mapping method:

```
Region = Bacia Hidrográfica do Rio Paraíba do Sul {
  Código          CódMun.Município::Paraíba.abs1;
  Nome            Nome.Município::Paraíba.abs1
  Estado          CódEstado.Município::Paraíba.abs1
  Área            Área.Município::Paraíba.abs1
  Perímetro       "Informação não disponível para Região"
  População       População.Município::Paraíba.abs1
  AnoInstituição AnoInst.Município::Paraíba.abs1}
```

```
Region = Bacia Hidrográfica do Rio Guaiíba {
  Código          CódMun.Município::Guaíba.abs1
  Nome            NomeMunicípio.Município::Guaíba.abs1
  Estado          CódEstado.Município::Guaíba.abs1
  Área            ÁreaMun.Município::Guaíba.abs1
  Perímetro       "Informação não disponível para Região"
  População       população.Município::Guaíba.abs1
  AnoInstituição AnoInst.Município::Guaíba.abs1}
```

```
Region = Estado do Rio de Janeiro{
  Codigo          Códigomun.Município::Ambiental.abs1
  Nome            Nomemun.Município::Ambiental.abs1
  Estado          Códigos.Município::Ambiental.abs1
  Área            Áreaof.Município::Ambiental.abs1
  Perímetro       Perímetro.Município::Ambiental.abs1
  População       "Informação não disponível para Região"
  AnoInstituição "Informação não disponível para Região"}
```

Export schema source:

OriginClass1: Município::Paraíba.abs1

Attributes: CódMun, CódEstado, NomeMun, ÁreaMun, População, AnoInstuição,
position.Geometry: VectorÁrea

Relationships:

Rel	ObjectA	ObjectB	Attributes	Car
Possui	Município	Bairro	CódMun	1:n
Atende	Município	DestinoFinalResíduoSólido	CódMun	1:n

Se localiza

Source.DBInfo.dbIdentifier:

Acess method: ARC; AP; map m1; mapex OriginClass1; reselect OriginClass1 poly attributes

=condição;

Polygonshades OriginClass1; List OriginClass1 poly; map end; quit, Display1040,

PostScript

OriginClass2: Município::Guaíba.abs1

Attributes: CódMun, CódEstado, NomeMun, ÁreaMun, População, AnoInstuição

position.Geometry: VectorÁrea

Relationships:

Rel	ObjectA	ObjectB	Attributes	Car
Possui	Município	Bairro	CódMun	1:n
Atende	Município	DestinoFinalResíduoSólido	CódMun	1:n

Se localiza

Source.DBInfo.dbIdentifier:

Acess method: ARC; AP; map m1; mapex OriginClass2; reselect OriginClass2 poly attributes

=condição; Polygonshades OriginClass1; List OriginClass poly; map end; quit, Display1040,

PostScript

OriginClass3: Município::Ambiental.abs1

Attributes: Código, Códigos, Nomemun, Áreaof, Perímetro

position.Geometry: VectorÁrea

Relationships:

Rel	ObjectA	ObjectB	Attributes	Car
Contém	Divisão Municipal	Município		1:n
Pertence	Cidades	Município	Códigomun	1:n

Source.DBInfo.dbIdentifier:

Acess method:

Superclass:

Class: Cidades

Region:

Bacia Hidrográfica do Rio Paraíba do Sul, Estado do Rio de Janeiro

Attributes

Código, Nome, Estado, Município, Distrito, População, Tipo, Área, Perímetro

Geometry: Ponto

Source schema mapping method:

Region = Bacia Hidrográfica do Rio Paraíba do Sul {
Código CD_Cid.Cidades::RHidro.abs1
Nome Nome.Cidades::RHidro.abs1
Estado CD_UF.Cidades::RHidro.abs1
Município CD_Mun.Cidades::RHidro.abs1
Distrito CD_Dis.Cidades::RHidro.abs1
População População.Cidades::RHidro.abs1
Tipo Tipo.Cidades::RHidro.abs1
Área "Informação não disponível para Região"
Perímetro "Informação não disponível para Região"}
Region = Estado do Rio de Janeiro{

Código Código.Cidades::Ambiental.abs1
Nome Nome.Cidades::Ambiental.abs1
Estado "Informação não disponível para Região"
Município Códigomun.Cidades::Ambiental.abs1
Distrito "Informação não disponível para Região"
População "Informação não disponível para Região"
Área Área.Cidades::Ambiental.abs1
Perímetro Perímetro.Cidades::Ambiental.abs1}

Export schema source:

OriginClass1: Cidades::RHidro.abs1

Attributes: CD_UF, CD_Mun, CD_Dis, CD_Ci, Nome População, Tipo
position.Geometry: Point

Relationships:

Rel	ObjectA	ObjectB	Attributes	Car
Possui	Munic	Cidades	CD_Mun	1:n
Possui	Cidades	TipoCidade	Tipo	1:1

Source.DBInfo.dbIdentifier:

Acess method:

OriginClas2: Cidades::Ambiental.abs1

Attributes: Código, Nome, Códigomun, Área, Perímetro
position.Geometry: Point

Relationships:

Rel	ObjectA	ObjectB	Attributes	Car
Contém	Divisão Municipal	Cidades	Código	1:n
Pertence	Cidades	Município	Códigomun	1:n

Source.DBInfo.dbIdentifier:

Acess method:

Superclass:

Class: BaciaHidrográfica

Region:

Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guafba, Estado do Rio de Janeiro

Attributes:

Código, Nome, Área, Perímetro

Geometry: Área

Source schema mapping method

Region = Bacia Hidrográfica do Rio Paraíba do Sul {
Código CódRegHidro.RegiãoHidro::Paraíba.abs2; CD_BH.BaciaHidro::RHidro.abs1
Nome NomeRegHidroRegiãoHidro::Paraíba.abs2; Nome.BaciaHidro::RHidro.abs1
Área ÁreaRegião.RegiãoHidro::Paraíba.abs2; Área.BaciaHidro::RHidro.abs1
Perímetro "Informação não disponível para Região"
Número "Informação não disponível para Região"
Escala "Informação não disponível para Região"
Microrregião "Informação não disponível para Região"
Ano "Informação não disponível para Região"}
Region = Bacia Hidrográfica do Rio Guafba {

Código CódRegHidro.RegiãoHidro::Guaíba.abs2
 Nome NomeRegHidroRegiãoHidro::Guaíba.abs2
 Área ÁreaRegião.RegiãoHidro::Guaíba.abs2
 Perímetro "Informação não disponível para Região"}
 Region = Estado do Rio de Janeiro {
 Código Códigobh.BaciaHidrográfica::Ambiental.abs1
 Nome Nome.BaciaHidrográfica::Ambiental.abs1
 Área Área.BaciaHidrográfica::Ambiental.abs1
 Perímetro Perímetro.BaciaHidrográfica::Ambiental.abs1 }

Export schema source:

OriginClass1: RegiãoHidro::Paraíba.abs2
Attributes: NomeRegHidro, CódRegHidro, ÁreaRegião
position.Geometry: VectorÁrea
Relationships:

Rel	ObjectA	ObjectB	Attributes	Car
Contém	RegiãoHidro	SubBacia	CódRegiãoHidro	1:n

Source.DBInfo.dbIdentifier
Acess method:

OriginClass2: BaciaHidro::RHidro.abs1
Attributes: CD_BH, Nome, Área
position.Geometry: VectorÁrea
Relationships:

Rel	ObjectA	ObjectB	Attributes	Car
Abrange	BaciaHidro	Município	CD_BH	
Abrange	BaciaHidro	Hidrografia	CD_BH	

Source.DBInfo.dbIdentifier
Acess method: ARC; AP; mapex originClassN ;reselect originClass geometry condiçãoARC; A,

mapex

OriginClass3: RegiãoHidro::Guaíba.abs2
Attributes: NomeRegHidro, CódRegHidro, ÁreaRegião
position.Geometry: VectorÁrea
Relationships:

Rel	ObjectA	ObjectB	Attributes	Car
Contém	RegiãoHidro	SubBacia	CódRegiãoHidro	1:n

Source.DBInfo.dbIdentifier
Acess method:

OriginClass4: BaciaHidrográfica::Ambiental.abs1
Attributes: Códigobh, Nome, Área, Perímetro
position.Geometry: VectorÁrea
Relationships:

Rel	ObjectA	ObjectB	Attributes	Car
Contém	MapaBacia Hidrográfica	BaciaHidrográfica	Códigobh	1:n

Source.DBInfo.dbIdentifier
Acess method:

Superclass:

Class: Hidrografia

Region:

Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba

Attributes:

Código, Nome, Perímetro, Tipo, Drenagem

Geometry: Linha

Source schema mapping method:

Region = Bacia Hidrográfica do Rio Paraíba do Sul {
 Código CódRh.RioArroio::Paraíba.abs2; CD_H.Hidrografia::RHidro.abs1
 Nome Nome.RioArroio::Paraíba.abs2; Nome.Hidrografia::RHidro.abs1
 Perímetro Perímetro.Hidrografia::RHidro.abs1
 Tipo Tipo.Hidrografia::RHidro.abs1
 Drenagem "Informação não disponível para Região"}
Region = Bacia Hidrográfica do Rio Guaíba {
 Código CódRh.RioArroio::Guaíba.abs2
 Nome Nome.RioArroio::Guaíba.abs2
 Perímetro "Informação não disponível para Região"
 Tipo Λ Região não possui esta informação"
 Drenagem "Informação não disponível para Região"}
Region = Estado do Rio de Janeiro {

Código "Informação não disponível para Região"
 Nome Nome.Rios::Ambiental.abs1
 Perimetro "Informação não disponível para Região"
 Tipo Cdren.Rios::Ambiental.abs1
 Drenagem Tdren.Rios::Ambiental.abs1 }

Export schema source:

OriginClass1: RioArroio::Paraiba.abs2

Attributes: CódRh, Nome
position.Geometry: VectorLine

Relationships:

Rel	ObjectA	ObjectB	Attributes	Car
Se especializa	RecursoHidrico	RioArroio		

Source.DBInfo.dbIdentifier

Acess method: ARC; AP; mapex originClassN ;reselect originClass geometry condiçãoARC; AP,

mapex

OriginClass2: HidrografiaRHidro.abs1

Attributes: CD_H, Nome, Perimetro, Tipo
position.Geometry: VectorLine

Relationships:

Rel	ObjectA	ObjectB	Attributes	Car
Possui	Hidrografia	TipoHidro		CD_H
Possui	Hidrografia	QualidadeÁgua		
Possui	Hidrografia	DadosSedimentologia		
Possui	Hidrografia	DadosHidrológicos		

Source.DBInfo.dbIdentifier

Acess method:

OriginClass3: RioArroio::Guaiba.abs2

Attributes: CódRh, Nome
position.Geometry: VectorLine

Relationships:

Rel	ObjectA	ObjectB	Attributes	Car
Se especializa	RecursoHidrico	RioArroio		

Source.DBInfo.dbIdentifier

Acess method: ARC; AP; mapex originClassN ;reselect originClass geometry condiçãoARC; AP,

mapex

OriginClas4: Rios::Ambiental.abs1

Attributes: Nome, Cdren, Tdren
position.Geometry: VectorÁrea

Relationships:

Rel	ObjectA	ObjectB	Attributes	Car
Contém	MapaBacia Hidrográfica	BaciaHidrográfica	Códigobh	1:n

Source.DBInfo.dbIdentifier:

Acess method:

Superclass:

Class: Rodovia

Region:

Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba, Estado do Rio de Janeiro

Attributes:

Código, Nome, Jurisdição, GrauConservação, Origem, Destino, Pistas

Geometry: Linha

Source schema mapping method:

Region = Bacia Hidrográfica do Rio Paraíba do Sul {

Código CD_R.Rodovia::RHidro.abs1
 Nome NomeRodovia.Rodovia::Paraiba.abs2; Nome.Rodovia::RHidro.abs1
 Jurisdição Jurisdição.Rodovia::Paraiba.abs2; adm.Rodovia::RHidro.abs1
 GrauConservação GR_cons.Rodovia::RHidro.abs1
 Origem "Informação não disponível para Região"
 Destino "Informação não disponível para Região"
 Pistas "Informação não disponível para Região" }

Region = Bacia Hidrográfica do Rio Guaíba {

Código "Informação não disponível para Região"
 Nome NomeRodovia.Rodovia::Guaiba.abs2;
 Jurisdição Jurisdição.Rodovia::Guaiba.abs2;
 GrauConservação "Informação não disponível para Região"

```

Origem          "Informação não disponível para Região"
Destino         "Informação não disponível para Região"
Pistas         "Informação não disponível para Região"}
Region = Estado do Rio de Janeiro{
Código         Código.Rodovia::Ambiental.abs1
Nome           Nome.Rodovia::Ambiental.abs1
Jurisdição     "Informação não disponível para Região"
GrauConservação "Informação não disponível para Região"
Origem         Origem.Rodovia::Ambiental.abs1
Destino        Destino.Rodovia::Ambiental.abs1
Pistas         Pistas.Rodovia::Ambiental.abs1}

```

Export schema source:

```

OriginClass1: Rodovia::Paraíba.abs2
Attributes: NomeRodovia, Jurisdição
           position.Geometry:VectorLine
Relationships:
  Rel      ObjectA      ObjectB      Attributes      Car
  Contém  Município    Rodovia      N:m

Source.DBInfo.dbIdentifier
Access method:

OriginClass2: Rodovia::RHidro.abs1
Attributes: CD_R, Nome , Adm, GR_cons,
           position.Geometry:VectorLine
Relationships:
  Rel      ObjectA      ObjectB      Attributes      Car
  Contém  Município    Rodovia      N:m

Source.DBInfo.dbIdentifier
Access method:

OriginClass3: Rodovia::Guaíba.abs2
Attributes: NomeRodovia, Jurisdição
           position.Geometry:VectorLine
Relationships:
  Rel      ObjectA      ObjectB      Attributes      Car

Source.DBInfo.dbIdentifier
Access method:

OriginClas4:Rodovia::Ambiental.abs1
Attributes: Nome, Código, Origem, Destino, Pistas
           position.Geometry:VectorÁrea
Relationships:
  Rel      ObjectA      ObjectB      Attributes      Car
  Se      InfraEstrutura Rodovia      l:n
  especializa

Source.DBInfo.dbIdentifier:
Access method:

```

Superclass:

Class: Ferrovia

Region:

Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba, Estado do Rio de Janeiro

Attributes:

Nome

Geometry: Linha

Source schema mapping method:

```

Region = Bacia Hidrográfica do Rio Paraíba do Sul {
Nome      NomeFerrovia.Ferrovia::Paraíba.abs2
Código    "Informação não disponível para Região"
Origem    "Informação não disponível para Região"
Destino   "Informação não disponível para Região"
Bitola    "Informação não disponível para Região"}
Region = Bacia Hidrográfica do Rio Guaíba {
Nome      NomeFerrovia.Ferrovia::Guaíba.abs2
Código    "Informação não disponível para Região"
Origem    "Informação não disponível para Região"
Destino   "Informação não disponível para Região"
Bitola    "Informação não disponível para Região"}

```

Region: Estado do Rio de Janeiro {
 Nome Nome.Ferrovia::Ambienta.abs1
 Código Código.Ferrovia::Ambienta.abs1
 Origem Origem.Ferrovia::Ambienta.abs1
 Destino Destino.Ferrovia::Ambienta.abs1
 Bitola Bitola.Ferrovia::Ambienta.abs1 }

Export schema source:

OriginClass1: Ferrovia::Paraiba.abs2

Attributes: NomeFerrovia
 position.Geometry:VectorLine

Relationships:

Rel	ObjectA	ObjectB	Attributes	Car
Possui	Ferrovia	TerminalFerroviário		

Source.DBInfo.dbIdentifier

Access method:

OriginClass2: Ferrovia::Guaiba.abs2

Attributes: NomeFerrovia
 position.Geometry:VectorLine

Relationships:

Rel	ObjectA	ObjectB	Attributes	Car
Possui	Ferrovia	TerminalFerroviário		

Source.DBInfo.dbIdentifier

Access method:

OriginClass4:Ferrovia::Ambiental.abs1

Attributes: Nome, Código, Origem, Destino, Pistas
 position.Geometry:VectorArea

Relationships:

Rel	ObjectA	ObjectB	Attributes	Car
Se especializa	InfraEstrutura	Ferrovia		1:n

Source.DBInfo.dbIdentifier:

Access method:

Superclass:

Class: Solos

Region:

Estado do Rio de Janeiro

Attributes: Tipo, Profundidade, Porosidade, Permeabilidade, Textura

Geometry: Área

Source schema mapping method:

Region = Estado do Rio de Janeiro {
 Tipo Tipo.Solos::Ambiental.abs1
 Profundidade Profund.Solos::Ambiental.abs1
 Porosidade Pors.Solos::Ambiental.abs1
 Permeabilidade Permea.Solos::Ambiental.abs1
 Textura Text.ÁreaProjeto::Ambiental.abs1 }

Export schema source:

OriginClass1: Solos.Ambiental.abs1

Attributes: Tipo, Profund, Pors, Permea, Text

Relationships:

Rel	ObjectA	ObjectB	Attributes	Car
Possui	ÁreaProjeto	Solos		

Source.DBInfo.dbIdentifier:

Access method:

Superclass:

Class: Florestas

Region:

Estado do Rio de Janeiro

Attributes: Tipo, Altura, Densidade

Geometry: Área

Source schema mapping method:

Region = Estado do Rio de Janeiro {
 Tipo Tipo.Florestas::Ambiental.abs1

Altura Altura.Florestas::Ambiental.abs1
Densidade Densidade.Florestas::Ambiental.abs1 }

Export schema source:

OriginClass1: Florestas.Ambiental.abs1

Attributes: Tipo, Altura, Densidade

Relationships:

<i>Rel</i>	<i>ObjectA</i>	<i>ObjectB</i>	<i>Attributes</i>	<i>Car</i>
Possui	ÁreaProjeto	Florestas		

Source.DBInfo.dbIdentifier:

Acess method:

Superclass:

Class: Declividade

Region:

Estado do Rio de Janeiro

Attributes: Grau

Geometry: Área

Source schema mapping method:

Region = Estado do Rio de Janeiro {
Grau Grau.Declividade::Ambiental.abs1 }

Export schema source:

OriginClass1: Declividade.Ambiental.abs1

Attributes: Grau

Relationships:

<i>Rel</i>	<i>ObjectA</i>	<i>ObjectB</i>	<i>Attributes</i>	<i>Car</i>
Possui	ÁreaProjeto	Declividade		

Source.DBInfo.dbIdentifier:

Acess method:

Superclass:

Class: Vegetação

Region:

Estado do Rio de Janeiro

Attributes: Tipo

Geometry: Área

Source schema mapping method:

Region = Estado do Rio de Janeiro {
Tipo Tipo.Vegetação::Ambiental.abs1 }

Export schema source:

OriginClass1: Vegetação.Ambiental.abs1

Attributes: Tipo

Relationships:

<i>Rel</i>	<i>ObjectA</i>	<i>ObjectB</i>	<i>Attributes</i>	<i>Car</i>
Possui	ÁreaProjeto	Vegetação		

Source.DBInfo.dbIdentifier:

Acess method:

Superclass:

Class: Relevô

Region:

Estado do Rio de Janeiro

Attributes: Tipo

Geometry: Área

Source schema mapping method:

Region = Estado do Rio de Janeiro {
Tipo Tipo.Relevô::Ambiental.abs1 }

Export schema source:

OriginClass1: Relevô.Ambiental.abs1

Attributes: Tipo

Relationships:

<i>Rel</i>	<i>ObjectA</i>	<i>ObjectB</i>	<i>Attributes</i>	<i>Car</i>
Possui	ÁreaProjeto	Relevô		

Source.DBInfo.dbIdentifier:

Acess method:

Superclass:

Class: Clima

Region:

Estado do Rio de Janeiro

Attributes: Tipo, Precipitação, Umidade, Temperatura

Geometry: Área

Source schema mapping method:

```
Region = Estado do Rio de Janeiro {  
  Tipo                Tipo.Clima::Ambiental.abs1  
  Precipitação        Preci::Ambiental.abs1  
  Umidade              Umi.Clima::Ambiental.abas1  
  Temperatura          Temp.Clima::Ambiental.abs1 }
```

Export schema source:

OriginClass1: Clima.Ambiental.abs1

Attributes: Tipo, Preci, Umi, Temp

Relationships:

Rel	ObjectA	ObjectB	Attributes	Car
Possui	ÁreaProjeto	Clima		

Source.DBInfo.dbIdentifier:

Acess method:

Superclass:

Class: UsoDaTerra

Region:

Estado do Rio de Janeiro

Geometry: Área

Attributes: Tipo, Adaptação

Source schema mapping method:

```
Region = Estado do Rio de Janeiro {  
  Tipo                Tipo.UsoDaTerra::Ambiental.abs1  
  Adaptação           Adap.UsoDaTerra::Ambiental.abs1 }
```

Export schema source:

OriginClass1: UsoDaTerra.Ambiental.abs1

Attributes: Tipo, Adap

Relationships:

Rel	ObjectA	ObjectB	Attributes	Car
Possui	ÁreaProjeto	UsoDaTerra		

Source.DBInfo.dbIdentifier:

Acess method:

Superclass:

Class: Altimetria

Region:

Estado do Rio de Janeiro

Attributes: Altitude

Geometry: Linha

Source schema mapping method:

```
Region = Estado do Rio de Janeiro {  
  Altitude            Altitude.Altimetria::Ambiental.abs1 }
```

Export schema source:

OriginClass1: Altimetria.Ambiental.abs1

Attributes: Altitude

Relationships:

Rel	ObjectA	ObjectB	Attributes	Car
Possui	ÁreaProjeto	Altimetria		

Source.DBInfo.dbIdentifier:

Acess method:

Superclass:

Class: PontoAltimétrico

Region:

Estado do Rio de Janeiro

Attributes: Cota

Geometry: Ponto

Source schema mapping method:

Region = Estado do Rio de Janeiro {
 Cota Cota.PontoAltimétrico::Ambiental.abs1}

Export schema source:

OriginClass1: PontoAltimétrico.Ambiental.abs1

Attributes: Cota

Relationships:

Rel	ObjectA	ObjectB	Attributes	Car
Possui	ÁreaProjeto	PontoAltimétrico		

Source.DBInfo.dbIdentifier:

Access method:

Superclass:

Class: RedeDrenagem

Region:

Estado do Rio de Janeiro

Attributes: Número, Microregião, Escala, Ano

Geometry: Linhas

Source schema mapping method:

Region = Estado do Rio de Janeiro {

Número Número.ReduDrenagem::Ambiental.abs1

Microregião Microregião.ReduDrenagem::Ambiental.abs1

Escala Escala.ReduDrenagem::Ambiental.abs1

Ano Ano.ReduDrenagem::Ambiental.abs1 }

Export schema source:

OriginClass1: RedeDrenagem.Ambiental.abs1

Attributes: Número, Microregião, Escala, Ano

Relationships:

Rel	ObjectA	ObjectB	Attributes	Car
Contém	RedeDrenagem	Rios		

Source.DBInfo.dbIdentifier:

Access method:

Superclass

Class: DivisãoMunicipal

Region:

Estado do Rio de Janeiro

Attributes: Número, Microregião, Escala, Ano

Geometry: Áreas, Pontos, Linhas

Source schema mapping method:

Region = Estado do Rio de Janeiro {

Número Número.DivisãoMunicipal::Ambiental.abs1

Microregião Microregião.DivisãoMunicipal::Ambiental.abs1

Escala Escala.DivisãoMunicipal::Ambiental.abs1

Ano Ano.DivisãoMunicipal::Ambiental.abs1 }

Export schema source:

OriginClass1: DivisãoMunicipal.Ambiental.abs1

Attributes: Número, Microregião, Escala, Ano

Relationships:

Rel	ObjectA	ObjectB	Attributes	Car
Contém	Divisão Municipal	Município, Rios, InfraEstrutura, Cidades		

Source.DBInfo.dbIdentifier:

Access method:

Superclass

Class: MapasBaciaHidrográficas

Region:

Estado do Rio de Janeiro

Attributes: Número, BaciaHidrográfica

Geometry: Áreas, Pontos, Linhas

Source schema mapping method:

Region = Estado do Rio de Janeiro {

Número Número.MapasBaciaHidrográficas::Ambiental.abs1

BaciaHidrográfica Códigobh.MapasBaciaHidrográficas::Ambiental.abs1 }

Export schema source:

OriginClass1: MapasBaciaHidrográficas.Ambiental.abs1

Attributes: Número, Códigobh
Relationships:

<i>Rel</i>	<i>ObjectA</i>	<i>ObjectB</i>	<i>Attributes</i>	<i>Car</i>
Contém	MapasBacia Hidrográficas	Rios, Indústrias		

Source.DBInfo.dbIdentifier:
Acess method:

Superclass

Class: ImagemLandsat
Region:
 Estado do Rio de Janeiro
Source schema mapping method:
Region = Estado do Rio de Janeiro {}
Export schema source:
OriginClass1: ImagemLandsat.Ambiental.abs1
Source.DBInfo.dbIdentifier:
Acess method:

B.7.4. Criação do esquema externo

Esta etapa é dividida nas seguintes atividades:

B.7.4.1. Definição das instâncias da hierarquia *Regions*

Nesta etapa são descritas as instâncias da hierarquia *Region* que definem os espaços geográficos resultantes da integração dos esquemas de acordo com as escalas.

Tendo em vista que as classes similares já faziam parte do esquema global externo e suas denominações não foram alteradas, será criada apenas mais uma instância para a hierarquia *Region*, a saber:

```

region: Region (Estado(
    name: Rio de Janeiro                               /* Esquema: Ambiental.abs1
    inverseScale: 1:250.000
    theme: List(ÓrgãoExecutor, Projeto, ÁreaProjeto, Município250, Cidades, InfraEstrutura,
                Ferrovia, Rodovia, Indústrias, BaciasHidrográficas, MapasBaciasHidrográficas,
                DivisãoMunicipal, Rios, RedeDrenagem Vegetação, Solos, Florestas, Relevo, Clima,
                UsoDaTerra, Declividade, PontosAltimétricos, Altimetria, ImagemLandsat))

```

B.7.4.2. Definição dos Temas

As classes virtuais definidas na seção B.5 serão expressas no esquema global, conforme a notação abaixo:

```

theme: Theme(
    name: ÓrgãoExecutor
    attributes: Nome, Código, Endereço, Vinculação
    geometry: Não tem
    regions: List (Estado do Rio de Janeiro)
    inverseScale: 1:250.000

theme: Theme(
    name: Projeto
    attributes: Projeto, Nome, BaciaHidrográfica, ÓrgãoExecutor
    geometry: Não tem

```

regions: *List* (Estado do Rio de Janeiro)
inverseScale: 1:250.000)

theme: *Theme*(
name: ÁreaProjeto
attributes: Projeto, BaciaHidrográfica, ÓrgãoExecutor,Área, Perímetro
geometry: Não tem
regions: *List* (Estado do Rio de Janeiro)
inverseScale: 1:250.000)

theme: *Theme*(
name: Município 250
attributes: Código, Estado , Nome, Mesoregião, Microregião, Área, Perímetro, População, AnoInstituição,
geometry: Área
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba,
Estado do Rio de Janeiro)
inverseScale: 1:250.000)

theme: *Theme*(
name: Cidades
attributes: Código, Nome, Estado, Município, Distrito, População, Tipo, Área, Perímetro
geometry: Point
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba,
Estado do Rio de Janeiro)
inverseScale: 1:250.000)

theme: *Theme*(
name: BaciaHidrográfica
attributes: Código, Nome, Área, Perímetro
geometry: Área
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba,
Estado do Rio de Janeiro)
inverseScale: 1:250.000)

theme: *Theme*(
name: Hidrografia
attributes: Código, Nome, Perímetro, Tipo, Drenagem
geometry: Área
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba, estado
do Rio de Janeiro)
inverseScale: 1:250.000)

theme: *Theme*(
name: Ferrovias
attributes: Nome
geometry: Linha
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba,
Estado do Rio de Janeiro)
inverseScale: 1:250.000)

theme: *Theme*(
name: Rodovia
attributes: Código, Nome, Jurisdição, GrauConservação, Origem, Destino, Pistas
geometry: Linha
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba,
Estado do Rio de Janeiro)
inverseScale: 1:250.000)

theme: *Theme*(

name: Solos
attributes: Tipo, Profundidade, Porosidade, Permeabilidade, Textura
geometry: Área
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba, Estado do Rio de Janeiro)
inverseScale: 1:250.000)

theme: *Theme*(
name: Florestas
attributes: Tipo, Altura, Densidade
geometry: Área
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba, Estado do Rio de Janeiro)
inverseScale: 1:250.000)

theme: *Theme*(
name: Declividade
attributes: Grau
geometry: Área
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba, Estado do Rio de Janeiro)
inverseScale: 1:250.000)

theme: *Theme*(
name: Vegetação
attributes: Tipo
geometry: Área
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba, Estado do Rio de Janeiro)
inverseScale: 1:250.000)

theme: *Theme*(
name: Relevo
attributes: Tipo
geometry: Área
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba, Estado do Rio de Janeiro)
inverseScale: 1:250.000)

theme: *Theme*(
name: Clima
attributes: Tipo, Precipitação, Umidade, Temperatura
geometry: Área
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba, Estado do Rio de Janeiro)
inverseScale: 1:250.000)

theme: *Theme*(
name: UsoDaTerra
attributes: Tipo, Adaptação
geometry: Área
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba, Estado do Rio de Janeiro)
inverseScale: 1:250.000)

thcmc: *Theme*(
name: Altimetria
attributes: Altitude
geometry: Linha

regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba, Estado do Rio de Janeiro)
inverseScale: 1:250.000)

theme: *Theme*(
name: PontoAltimétrico
attributes: Cota
geometry: Ponto
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba, Estado do Rio de Janeiro)
inverseScale: 1:250.000)

theme: *Theme*(
name: RedeDrenagem
attributes: Número, Microregião, Escala, Ano
geometry: Linha
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba, Estado do Rio de Janeiro)

theme: *Theme*(
name: DivisãoMunicipal
attributes: Número, Microregião, Escala, Ano
geometry: Linha, Polígono, Pontos
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba, Estado do Rio de Janeiro)
inverseScale: 1:250.000)

theme: *Theme*(
name: MapasBaciasHidrográficas
attributes: Número, Microregião, Escala, Ano
geometry: Linha, Polígonos, Pontos
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba, Estado do Rio de Janeiro)
inverseScale: 1:250.000)

theme: *Theme*(
name: ImagemLandsat
attributes: Não tem
geometry: Não tem
regions: *List* (Bacia Hidrográfica do Rio Paraíba do Sul, Bacia Hidrográfica do Rio Guaíba, Estado do Rio de Janeiro)
inverseScale: 1:250.000)
inverseScale: 1:250.000)

B.8. Validação

Esta etapa consiste em verificar as seguintes métricas:

- ◆ **Completeness** - é verificado que todas as classes descritas no SAIF se encontram nas tabelas de classes, e se todas as classes descritas nesta tabela se encontram descritas na notação proposta para as classes virtuais resultante da integração;
- ◆ **Exactness** - é verificado que todas as classes virtuais descritas na notação proposta estão corretamente descritas em relação aos seus esquemas de exportações originais;

- ◆ Minimalidade - é verificado que não existem classes virtuais redundantes, bem como instâncias das classes do esquema externo redundantes; e
- ◆ Compreensão - esta métrica será avaliada no ambiente MultiSIG quando for realizadas consultas sob o esquema externo.