

# INTEROPERABILIDADE DE BASES DE DADOS ESPACIAIS HETEROGÊNEAS E DISTRIBUÍDAS

Julia Celia Mercedes Strauch<sup>(1,2)</sup>

e-mail: [julia@cos.ufrj.br](mailto:julia@cos.ufrj.br)

Marta Lima de Queirós Mattoso<sup>(2)</sup>

e-mail: [marta@cos.ufrj.br](mailto:marta@cos.ufrj.br)

Jano Moreira de Souza<sup>(2)</sup>

e-mail: [jano@cos.ufrj.br](mailto:jano@cos.ufrj.br)

(1)EMBRAPA/CNPS - Rua Jardim Botânico, nº 1024, Cep 22460-000

Rio de Janeiro, RJ, Tel: (021) 274 4999, Fax: (021) 274 5291

(2)COPPE/UFRJ - Prog. de Eng. de Sist.e Computação, CP 68511, Cep 21945-970

Rio de Janeiro, RJ, Tel: (021) 590 2552, Fax: (021) 290 6626

## Resumo

O alto custo da aquisição de dados para implantação de Sistemas de Informações Geográficas (SIGs) e a complexidade das análises tem demandado o acesso a dados multidisciplinares armazenados em bases de dados de SIGs localizadas em diversas instituições. No âmbito da tecnologia de Banco de Dados as propostas têm conduzido ao desenvolvimento de um ambiente denominado **Multidatabase (MB)**, que proporciona o compartilhamento de dados entre sistemas. Este trabalho tem por objetivo caracterizar a heterogeneidade entre as bases objetos geográficos, discutir as soluções atuais para promover o compartilhamento de dados entre SIGs, conceituar interoperabilidade entre bases de dados heterogêneas e distribuídas de SIGs e apontar as principais vantagens deste conceito aplicado a SIG.

## Abstract

The high cost of data acquisition to Geographic Information System (GIS) and the analysis's complexity had forced the access to the multidisciplinary data stored into GIS's databases located in any institutions. In the Database technology context the proposes have lead to the development of **Multidabase (MB)** environment, that provide the data sharing among systems. This paper aims to characterise the heterogeneity between geographics objects databases, to discuss the actual solutions to promote data sharing among GIS, to present the interoperability concept among heterogeneous and distributed GIS databases and to point the main advantages of these applied to GIS.

## 1. Introdução

Os Sistemas de Informações Geográficas (SIGs) surgiram da necessidade de di-versas áreas de aplicações com-binarem dados descritivos sobre a superfície terrestre com a sua representação cartográfica, a fim de realizar análises espaciais de acordo com crité-rios definidos pelos usuários. Isto gerou uma diversidade de *hardwares*, *softwares* e sistemas operacionais uti-lizados para o desenvolvimento de SIGs. Esta diversidade au-menta quando se considera que as bases de dados destes sistemas foram proje-tadas inde-pendentemente de forma diferenciada, de acordo com o domínio e finalidade da área de aplicação.

O alto custo da aquisição de dados para a implantação de um SIG e a crescente complexi-dade das análises ambientais, que a cada dia exigem novos parâmetros de diversos domínios de conhecimento não previstos nas questões manifestas e la-tentes, requerem o acesso a dados multidisciplinares arma-zenados em diversas bases de dados localizadas em diferentes instituições.

Do ponto de vista concei-tual dos dados, o compartilha-mento destes é possível, uma vez que nestas bases os objetos geográficos tem em comum o mes-mo espaço geográfico descrito. Contudo, da perspectiva de Ban-co de Dados, essas bases são heterogêneas e distribuídas. A

heterogeneidade destas bases é decorrente: **i)** das áreas modelarem os objetos geográficos de acordo com a semântica necessária a área de aplicação e a finalidade da instituição a qual pertence a informação; e **ii)** dos SIGs possuírem características decorrentes do ambiente de implementação e da adequação da modelagem dos dados ao *software* utilizado para o acesso, gerência e manipulação dos dados, influenciando no formato e estrutura de armazenamento dos dados gráficos e não gráficos dos objetos geográficos. A distribuição é decorrente dessas bases de dados estarem geograficamente localizadas em diferentes instituições [STR95].

A solução, inicialmente proposta, consiste basicamente em desenvolver aplicações específicas para converter e migrar dados de um SIG, em um determinado formato, para outro, em outro formato, e vice-versa, proporcionando interoperabilidade entre pares específicos de SIG. Uma outra solução consiste em consolidar alguns formatos de dados mais utilizados como padrões para troca de dados entre SIGs. Entretanto, nos últimos anos, os recentes progressos nas tecnologias de rede, de Sistemas de Gerenciamento de Banco de Dados Distribuídos (SGBDDs), e de comunicação, no âmbito da tecnologia de Banco de Dados, têm conduzido ao desenvolvimento de um ambiente denominado **Multidatabase** (MB), que proporciona o compartilhamento de dados armazenados em SGBDDs pré-definidos e independentes, que gerenciam uma coleção de dados e aplicações específicas [BRI94].

Esta tecnologia estendida a SIG poderá trazer consideráveis benefícios no estabelecimento de mecanismos que permitem a interoperabilidade entre bases de dados georeferenciadas distribuídas entre algumas instituições. A interoperabilidade neste contexto tem por objetivo permitir que dois ou mais SIGs heterogêneos interajam uns com os outros e troquem dados georeferenciados de forma a alcançar resultados unificados.

Assim, este trabalho tem por objetivo caracterizar a heterogeneidade entre as bases objetos geográficos, discutir as soluções atuais para promover o compartilhamento de dados entre SIGs, conceituar interoperabilidade entre bases de dados georeferenciadas heterogêneas e distribuídas de SIGs e apontar as principais vantagens deste conceito aplicado a SIG.

## 2. Heterogeneidade entre bases de dados de SIG

As bases de dados de SIG contêm objetos geográficos que representam fenômenos ligados a natureza ou resultante da atuação do homem sobre o meio. A heterogeneidade entre estas bases de dados é decorrente do fato delas armazenarem diferentes representações lógicas e físicas das conceitualizações a respeito do mundo real. Esta heterogeneidade está presente durante toda a etapa de ciclo de vida de um SIG, a saber: na adoção do modelo conceitual para o seu desenvolvimento; na abstração da aplicação; na especificação dos metadados; no *hardware* utilizado; no formato dos dados; no próprio conteúdo da base e na ferramenta utilizada para o gerenciamento dos dados com suas respectivas linguagens de consulta [AGU95].

Entretanto a heterogeneidade tem sido pesquisada no contexto das diferenças entre os esquemas das bases de dados. Estas diferenças são resultantes da própria tecnologia de Banco de Dados ter capacidade de armazenar uma única representação que captura como cada usuário percebe e concebe a realidade de modo individualizado para sua aplicação. Assim, as diversas bases de dados de SIG heterogêneas e distribuídas foram desenvolvidas independentemente para atender à necessidade de seus usuários. Estas bases não compartilham do mesmo conjunto de dados sobre a realidade que elas tratam, ou se compartilham os mesmos dados não são representados da mesma forma, tornando os seus esquemas incompatíveis (diferentes) ou inconsistentes (faltam informações).

Os objetos geográficos armazenados nas bases de dados de um SIG são descritos por um conjunto de dados gráficos e não gráficos georeferenciados em relação a um sistema de coordenadas. Deste modo, os conflitos decorrentes da heterogeneidade entre os esquemas das bases de dados de SIG podem ser classificados em [STR95]:

- **conflitos entre os dados não gráficos** - são relacionados as diferentes características temáticas utilizadas para descrever os objetos geográficos. Estes conflitos são semelhantes aos conflitos encontrados nas bases de dados convencionais e podem ser classificados em:
  - **conflitos semânticos** - ocorrem quando as conceitualizações sobre a realidade comum é percebida de forma diferente levando as bases de dados a apresentarem divergências semânticas (diferentes significados) relativas as denominações utilizadas e as propriedades que descrevem os objetos geográficos; e
  - **conflitos sintáticos ou estruturais** - ocorrem quando apesar das conceitualizações 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.

- **conflitos entre dados gráficos** - são resultantes da multiplicidade de representações gráficas e diferenças cartográficas, bem como das estruturas de dados utilizadas para armazenar o objeto geográfico. Desta forma os conflitos entre dados gráficos podem ser classificados em:
- **conflito contextual** - ocorre quando a semântica das conceitualizações gráficas são representadas por diferentes representações gráficas (ponto, linha, polígono) empregando diferentes *data*, sistema de projeções, sistema de coordenadas, escalas, semiologia gráfica e etc.; e
- **conflito genérico** - é decorrente das conceitualizações sobre o espaço serem construídas sobre dois modelos conceituais diferentes, de acordo com a forma mais apropriada e eficiente para armazenar e realizar operações necessárias aos usuários, a saber: modelos baseados em níveis (estrutura *raster*) e modelo baseado em objeto espacial (estrutura vetorial). Isto implica em diferenças estruturais e nos operadores geométricos e topológicos proporcionados pelos dois modelos.

### 3. Soluções atuais soluções atuais para compartilhar dados entre SIGs

As soluções propostas para promover o compartilhamento de dados entre SIG têm sido eliminar a heterogeneidade entre os diferentes formatos de arquivos proporcionados pelas diversas estruturas de dados utilizadas para representar os dados gráficos e não gráficos dos objetos geográficos [FR094].

A estratégia inicial para promover o compartilhamento de dados entre SIGs consistia no desenvolvimento de aplicações orientadas a produtos, que tinham por finalidade transformar as estruturas de dados de um sistema para outro, através da conversão de formatos de arquivos de dados processados em *batch*. Esta estratégia no âmbito da tecnologia de Banco de Dados não é uma tarefa difícil, entretanto quando aplicada a SIG ela 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. Deste modo, a implementação destas aplicações exigia 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.

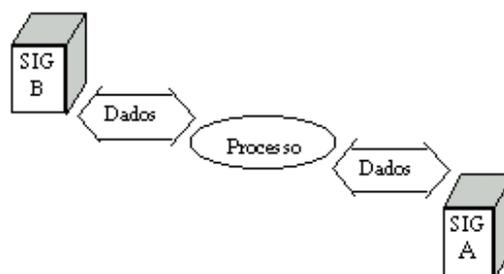


Fig.1 - Conversão de dados

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.

De modo geral, com o amadurecimento dos softwares para SIG disponíveis no mercado, os processos de conversão passaram a ser incorporados aos sistemas através de aplicações que permitem importar/exportar formatos de arquivos de dados que se consolidaram no mercado como padrão, tal como formato DXF (*Drawing eXchange Format*) ou o formato GENERATE do ARC/Info [ARC94]. Entretanto esta solução se torna ineficiente quando há necessidade de trocar dados com um novo tipo de sistema não previsto por estas aplicações.

Deste modo, atualmente uma das soluções propostas pela comunidade usuária dos SIGs é o estabelecimento de um padrão de dados que permita transferência de dados entre os sistemas. Este padrão de dados deve oferecer simplicidade de manuseio, neutralidade em

relação aos sistemas, poder de expressão para qualquer modelo de dados e capacidade de adequação aos novos conceitos da tecnologia. Esta solução representa um denominador comum entre diferentes sistemas e viria reduzir o número de aplicações de conversões de  $n(n-1)$  para  $2n$ , onde  $n$  é o número de formatos de arquivos de dados.

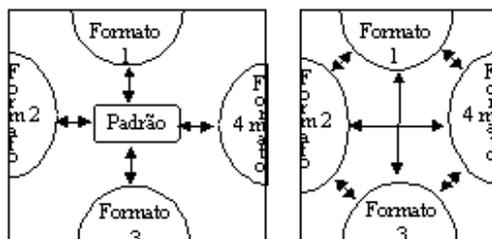


Fig. 2 - Comparação do número de conversores com/sem padrão de dados onde onde significa duas transformações de arquivos

Em alguns países, órgãos com poder normativo estão propondo padrões para o intercâmbio de dados geográficos. Dentre estes destacam-se os padrões americano *Spatial Data Transfer Standard* (STDS) proposto pelo *Federal Geographical Data Committee* (FGDC) [RIB95] e canadense *Spatial Archive Interchange Format* (SAIF) proposto pelo *Survey and Resource Mapping Branch - Ministry of Environment Lands and Parks* (MELP) [SAIFa]. O SDTS é um padrão que fornece formatos para intercâmbio de dados associado a representações dos dados geográficos (vetores, grades, tabelas, etc.) que reflete o modelo conceitual do ARC/Info. O SAIF é um padrão canadense que permite definir estruturas de dados usando um modelo de dados orientado a objeto que independente de *software* e foi projetado para modelar e promover a troca de dados espaço-temporais utilizando ferramentas que estão sendo desenvolvidas, tal como *Feature Manipulation Engine* (FMB) [SAIFb].

O estabelecimento destes padrões requer três considerações. A primeira delas consiste em estabelecer como o padrão pretende atender a tecnologia de heterogeneidade. A segunda é relativa a neutralidade deste padrão que deve ser estabelecida através de um consenso entre os fabricantes e usuários. E a última consiste em estabelecer um padrão de modo que ele possa evoluir com a tecnologia, uma vez que a taxa de inovação da informática torna estes obsoletos mesmo antes de serem completamente estabelecidos.

Uma tendência atual para o compartilhamento de dados entre bases de dados de SIGs é promover a interoperabilidade estendendo as características de um ambiente MB para suportar dados gráficos e não gráficos [GAR,GOD94,LAU94].

#### 4. Interoperabilidade

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 [Per95] define genericamente *interoperabilidade* como a *habilidade de dois ou mais sistemas, sejam computadores, dispositivos de comunicações, base de dados, rede e outras tecnologias de informação; interagirem um com o outro e trocar dados de acordo com um método prescrito a fim de alcançar resultados previsíveis.*

A interoperabilidade é alcançada no ambiente MB por uma arquitetura que suporta o acesso à várias bases de dados, escondendo as heterogeneidades entre os sistemas, esquemas de dados ou aplicações, de forma que os sistemas se comuniquem uns com os outros, proporcionando a troca de dados e expandindo a base de aplicações dos usuários. O objetivo destas arquiteturas é fornecer um ambiente no qual o usuário pode submeter consultas e atualizações acessando os dados de várias bases heterogêneas e distribuídas; denominadas de bases de dados locais; usando definição de dados e uma linguagem de manipulação comum, bem como preservando as autonomias de seus SGBDs.

No âmbito da tecnologia de Banco de Dados são encontradas na literatura diversas propostas de arquiteturas para este ambiente que podem ser classificadas em acoplamento forte e fraco. A primeira delas tem por objetivo desenvolver um ambiente formado por um esquema conceitual global que integra todas as bases de dados com um único método de acesso. A segunda tem por objetivo integrar as bases de dados utilizando uma interface interativa com o usuário para a negociação direta entre as bases de dados componentes da federação ou um programa de aplicação ou uma linguagem de consulta a múltiplas bases de dados, que auxilia a integração dos dados somente no contexto de uma sessão, sem um esquema global.

Assim a tendência em SIGs é substituir as soluções orientadas ao produto por arqui-

teturas cliente-servidores que permitam a interoperação através da eliminação das heterogeneidades entre as bases de dados de SIGs. Para isto é necessário compreender as diferenças entre essas bases de dados projetadas independentemente, capturar a semântica dos dados, os requisitos de consistência, as restrições de integridade e aplicar técnicas desenvolvidas para resolver os conflitos entre os dados gráficos e não gráficos dos objetos geográficos armazenados pelos SIG.

Esta arquitetura deverá ser composta por uma interface que se apresenta como uma única aplicação em diversas plataformas/sistemas e deverá contemplar ainda três requisitos básicos que tratam do gerenciamento de rede em um ambiente MB, a saber:

- os SIGs devem ter **autonomia local**, ou seja eles compartilham os seus dados com os demais SIGs sem comprometer suas restrições de integridade, aplicações específicas ou mecanismos de segurança;
- os SIGs devem **acessar transparentemente** os dados dos demais sistemas, isto é os usuários devem acessar diferentes sistemas como se estivessem acessando a uma aplicação a qual já estão acostumados; e
- o ambiente de interoperabilidade deve possuir mecanismo **de controle de concorrência** eficiente de forma que as consultas realizadas pelos usuários sejam executadas corretamente.

Atualmente, a emergente expansão da *World Wide Web* (WWW) traz novas dimensões tornando conceito de interoperabilidade mais amplo, uma vez que este ambiente deverá utilizar recursos da rede INTERNET permitindo o acesso a dados distribuídos em diferentes ambientes computacionais através da rede global de repositórios de dados [GAR].

## 5. Vantagens de um ambiente MB aplicada a SIG

O desenvolvimento de um ambiente MB que proporciona a interoperabilidade entre bases de dados de SIGs localizadas em diferentes instituições é o grande desafio da década. Este ambiente proporcionará um compartilhamento de dados mais produtivo, trazendo consideráveis benefícios a comunidade de usuários. Dentre os principais benefícios destacam-se:

- mais eficiência no acesso a bases de dados que oferecem uma melhor acuracidade ou estão mais atualizadas;
- expansão das bases de dados e das aplicações dos usuários de SIGs;
- possibilidade de desenvolver novos serviços no acesso a dados multidisciplinares; e
- redução do custo e do tempo do processo de aquisição de dados georeferenciados.

Esta última acarretará na conseqüente redução do custo de desenvolvimento de um SIG, uma vez que o processo de aquisição de dados é a parte mais onerosa para a implantação de um projeto.

## 6. Conclusão

A troca de dados entre bases de dados georeferenciadas é uma necessidade crescente entre as diversas instituições que trabalham com informação georeferenciada. Deste modo estender as pesquisas realizadas para o desenvolvimento de um ambiente MB em Banco de Dados convencionais para a tecnologia de SIGs irá trazer consideráveis benefícios à comunidade de usuários de Geoprocessamento.

Este ambiente possibilitará interoperabilidade entre bases de dados de SIG através da integração de informações multidisciplinares das diversas bases de objetos geográficos heterogêneas distribuídas. Isto deverá expandir a base de dados dos usuários de Geoprocessamento, permitindo pensarem sobre que tipos de dados eles querem compartilhar e realizar novos tipos de análises dos dados.

O Projeto HIP (**H**eterogeneidade, **I**nteroperabilidade e **P**aralelismo do Programa de Engenharia de Sistemas e Computação da COPPE/UFRJ tem como projeto de pesquisa o desenvolvimento de uma arquitetura que possibilite a interoperabilidade entre bases de objetos geográficos heterogêneas e distribuídas. Esta arquitetura deverá proporcionar um modelo de dados que descreva os objetos geográficos utilizando técnicas para compatibilizar as referências geográficas, a estrutura semântica dos objetos geográficos heterogêneos, a qualidade dos dados e outras questões relativas aos metadados. Esta arquitetura deverá se beneficiar do paradigma orientado a objetos na adoção do modelo de dados, bem como no

acesso às bases de dados georeferenciadas. Ela deverá ainda utilizar algum formalismo para definir os metadados de forma a auxiliar os usuários a encontrarem os objetos geográficos no ambiente distribuído [RIB96].

O desenvolvimento deste ambiente é um primeiro passo ao desenvolvimento da tecnologia de bibliotecas geográficas digitais, que nada mais são do que uma hiperbase de dados geográficos descentralizada. O desenvolvimento destas bibliotecas permitiram um eficiente acesso a mapas e imagens através do uso de sofisticados *browsing*, catálogo de dados e avançadas técnicas de pesquisa.

**Agradecimentos:** Ao CNPQ pelo suporte a esta pesquisa.

## Referências bibliográficas

[AGU95] Aguiar, C. D.; Integração de Sistemas de Banco de Dados Heterogêneos em Aplicações de Planejamento Urbano; Dissertação de Mestrado apresentada ao Instituto de Matemática e Ciência da Computação da UNICAMP, 1995

[ARC94] ARC/INFO: Data Management: Concepts, data models, database design and storage, v. 7.1; ESRI; 1994

[BRI94] Bright, M. W.; Hurson, A. R. & Pakzad.; A taxonomy and current issues in Multidatabase Systems; In: Multidatabase Systems: An advanced solution for global information sharing; IEEE Computer Society Press; 1994

[FRO94] Froment, E. D.; Gis and data integration: a user's perspective; GIS'94 Symposium, Vancouver, 1994, pg 157-161

[GAR95] Garcia-Molina, H. & Hsu, M.; Distributed Databases; In: Modern Database Systems the object model, interoperability and beyond; Won Kim Editor; ACM Press; New York; 1995

[GAR] Gardels, K. The open gis approach to distributed geodata and geoprocessing; [http://..cgia.uscb.edu/conf/SANTA\\_FE\\_CR\\_ROM/sf\\_papers/gardels\\_ken/ogismodl.html](http://..cgia.uscb.edu/conf/SANTA_FE_CR_ROM/sf_papers/gardels_ken/ogismodl.html)

[GOD94] Goodman, J. N.; Alberta land related Information System, a federated databases case study, Proceedings of Annual Conference URISA, 1994

[LAU94] Laurini, R.; Sharing geographic information in distributed databases; Proceedings of Annual Conference URISA, 1994

[Per95] Perine, L. A.; In Pursuit of an optimum: a conceptual model for examining public sector policy support of interoperability; Proceedings of the Workshop on interoperability and economics of information infrastructure, Virginia, 1995

[RIB96] Ribeiro, G.P. & Souza, J. M.; Digital Geospatial Metadata: an Brazilian Case of Federal Databases; first IEEE Metadata Conference; Maryland, USA; 1996

[SAIFa] SAIF - Profile 1: British Columbia Specification and Guidelines for Geomatics; <http://www.env.gov.bc.ca/~srmb/stk/toc.html>

[SAIFb] SAIF/FMEBC: Spatial Archive and Interchange Format - Feature Manipulation Engine; [http://www.env.gov.bc.ca/~srmb/fmbc/SAIF\\_FMEBC.html](http://www.env.gov.bc.ca/~srmb/fmbc/SAIF_FMEBC.html)

[STR95] Strauch, J. C. M & Mattoso, M. L. Q.; Interoperabilidade entre bases de dados heterogêneas e distribuídas, Relatório Técnico, COPPE, 1995