

Intercâmbio de Dados Geográficos: Modelos, Formatos e Conversores

Paulo Lima^{*,1}, Gilberto Câmara^{**,1}, João Argemiro^{**,1}, Antônio Miguel Vieira Monteiro^{**,1}

(1)Área de Geoprocessamento

Divisão de Processamento de Imagens

Instituto Nacional de Pesquisas Espaciais (INPE)

(*)Mestrado, Bolsa CAPES, e-mail: Paulo@nucleo.inpe.br (**) Orientadores

Resumo.

Interoperabilidade de informação geográfica é um problema importante e difícil. Cada Sistema de Informação Geográfica é desenvolvido independentemente com pouca preocupação com teorias ou terminologia comum. Como resultado, é muito difícil para diferentes sistemas compartilhar dados, para usuários treinados em um sistema usar outro, ou para usuários compartilharem procedimentos desenvolvidos em sistemas diferentes. Este cenário leva ao desenvolvimento de modelos de dados e formatos de dados que são capazes de representar diferentes tipos de informação geográfica. Este trabalho descreve o desenvolvimento de um formato novo para intercâmbio de informação espacial entre diferentes SIGs e um software “open source” que tem como objetivo prover um ambiente para visualização e conversão de arquivos de diferentes formatos e o formato proposto.

Palavras Chaves: Sistemas de Informação Geográfica

1 Introdução

Um dos desafios mais importantes no uso das geotecnologias é o intercâmbio de dados espaciais, impulsionado principalmente pelo alto custo de produção deste tipo de dados. A falta de modelos conceituais comuns acarreta problemas na troca de dados entre organizações utilizando SIGs distintos. Segundo [9], em um ambiente de sistemas heterogêneos, a conversão de dados representa um custo entre 60% e 80% do custo total na implantação de SIGs.

No caso brasileiro, o problema do intercâmbio de dados espaciais é agravado pela falta de padrões nacionalmente estabelecidos para intercâmbio de dados geográficos e pela não disponibilidade de ferramentas de baixo custo e fácil utilização para conversão de dados. Em função das deficiências nacionais neste campo, os produtores de dados (públicos e privados) vem adotando soluções “ad-hoc” para o intercâmbio de dados, muitas vezes sem a devida análise crítica da experiência internacional nesta área.

Com base nesta motivação, este trabalho apresenta a proposta de um modelo genérico para dados geográficos e de um formato brasileiro de intercâmbio de dados geográficos baseado neste modelo, denominado GEOBR. Como complemento, pretendemos ainda implementar sob a forma de código “open-source”, uma ferramenta para visualização e conversão entre formatos de dados geográficos comuns no mercado e o formato proposto.

2 Dimensões do Problema da Interoperabilidade Geográfica

2.1 Conversão Sintática

A abordagem mais básica para intercâmbio de dados geográficos é a conversão sintática direta, que procura realizar a tradução dos arquivos de informação geográfica entre diferentes formatos. Para permitir este tipo de conversão, os SIGs trabalham com duas alternativas: (a) oferecer um formato de exportação ASCII de fácil legibilidade, como DXF (Autocad), MID/MIF (MapInfo), E00 (Arc/Info) e SPR (Spring); (b) documentar as estruturas de dados internas, como no caso do SHP (ArcView). Estes formatos não garantem a transferência sem distorção de informações, pois são organizados de acordo com o sistema que os gerou, e quando importados para sistemas conceitualmente diferentes, necessitam de manipulação externa.

2.2 O Padrão OpenGIS

O consórcio OpenGIS pretende definir um modelo de dados genérico e interfaces padronizadas para acesso a bancos de dados geográficos, baseadas em diferentes tecnologias, como XML, COM, Java e SQL [10]. Apesar dos inegáveis avanços, a proposta OpenGIS tem várias limitações, entre elas: a existência de uma API resolve apenas o problema de acesso padronizado a bancos de dados espaciais e não substitui a necessidade da transferência dos dados entre sistemas; Inclui, até o momento, apenas operações topológicas de consulta sobre objetos simples [6] sem permitir a definição de relacionamentos espaciais para definição de restrições espaciais,

como proposto em [4]. O uso da linguagem SQL como base para a linguagem de consulta no caso de OpenGIS é questionável. Como mostram [5] e [1], o padrão declarativo do SQL tem diversas limitações para tratar com dados geográficos.

2.3 Metadados

A principal proposta de padrão de metadados é do FGDC (*Federal Geographic Data Committee*), comitê que promove a coordenação do uso, troca e disseminação de dados espaciais nos EUA [7]. O grande problema da proposta do FGDC (e do uso de metadados em geral) é a excessiva ênfase em informações que descrevem o processo de produção dos dados. Com relação à sintaxe, o padrão limita-se a indicar qual o formato em que os dados estão disponíveis. No aspecto semântico, suas informações são muito limitadas, pois o FGDC não adota o “modelo padrão” de geoinformação (campos e objetos). Adicionalmente, o padrão do FGDC reflete os compromissos inevitáveis do “projeto de comitê”, pois requer uma excessiva quantidade de informações (de aplicação questionável), com dezenas de formulários.

2.4 Conversão Semântica e Ontologias

O aspecto semântico diz respeito à representação conceitual da informação geográfica presente em cada sistema. Como comunidades com cultura e história diferentes que interpretam distintamente a realidade geográfica e produzem sistemas conceitualmente heterogêneos, a capacidade de transferir dados de um sistema para outro não garante que os dados têm significado para o novo usuário [8]. Nesta visão semântica, a realização plena da interoperabilidade incluem os conceitos: (a) o modelo de dados, como no caso do modelo OMT-G [4]; (b) o dicionário de conceitos, indicado por uma ontologia comum [8]; (c) o dicionário de procedimentos, que contém os diferentes mecanismos de consulta, manipulação e apresentação utilizados para extrair informação dos dados compartilhados [3].

3 O Formato GeoBR

3.1 Descrição Geral

O formato GEOBR pretende se diferenciar das propostas anteriores por incorporar os recentes avanços conceituais em Ciência da Geoinformação, e mais especificamente, pelas seguintes características:

- uso de um modelo de dados orientado-a-objetos que captura as noções de campos e objetos geográficos, e relacionamentos espaciais e hierárquicos entre as classes.
- inclusão opcional de um conjunto mínimo suficiente de informações sobre produção do dados.
- suporte para incorporação futura de ontologias e de procedimentos de análise geográfica.

O formato GEOBR difere dos formatos genéricos de intercâmbio, como DXF, SHP, e E00, pois enquanto estes formatos levam em conta apenas as características do sistema que gerou os dados, o GEOBR preocupa-se em atender a todo o conjunto de tecnologias de geoinformação. O formato GEOBR é especificado em XML (*eXtensible Markup Language*), o que torna seu conteúdo mais acessível, promovendo o intercâmbio de dados via Internet. A tecnologia XML também foi proposta pelo consórcio OpenGIS pela especificação da GML (*Geography Markup Language*) [10]. O GEOBR também se difere da GML pois utiliza um modelo conceitual único e genérico para representar os diferentes tipos de dados espaciais. Deste modo, não requer do usuário a definição de um esquema específico em XML. GeoBR tem definições de diferentes tipos de dados (geo-campos e geo-objetos), enquanto a versão atual do GML tem suporte apenas para geo-objetos simples (*simple feature*). Além disto, GML requer que cada instituição defina seu esquema de dados, o que implica em investimento adicional para conversão de dados. Em tese, quando estiverem disponíveis versões do GML que suportem tipos de dados espaciais mais complexos, poderemos escrever o padrão GeoBR como um esquema XML específico em GML.

Um arquivo GEOBR pode representar uma camada de dados apenas ou um conjunto de camadas de um banco de dados geográfico e é dividido em seções, cada uma identificada por um elemento XML, a saber:

- <ontology>: Dicionário de Ontologias com informação semântica sobre os dados.
- <metadata>: Informações sobre a produção do dado.
- <data_model>: Descrição das entidades reais representadas no formato de acordo com um meta-modelo genérico que inclui os tipos de dados geográficos (uma seção por arquivo, opcional). O modelo inclui: <feature>, dados geográficos individualizáveis, como lotes. <network>, entidades conectadas em topologia arco-nó. <surface>, expressam a variação contínua de uma grandeza quantitativa. <thematic>,

correspondem à variação espacial de uma grandeza qualitativa. <image>, são obtidas por satélites de observação da terra, ou a partir de digitalização de fotos aéreas.

- <projection>: descreve uma projeção sob a qual estão as camadas de dados (<layers>) representadas no formato de acordo com um meta-modelo genérico que inclui os tipos de dados geográficos.
- <layers>: camadas de informação geográfica. Incluem as seguintes geometrias <line>: coordenadas vetoriais em 2D. <node>: correspondem as entidades de topologia arco-nó. <polygon> coordenadas de entidades do tipo região. <point>: localizações individuais (X,Y). <sample>: localizações individuais do tipo (X,Y,Z). <contour>: linhas associadas a regiões de mesma cota. <grid>: grade regular. <table>: atributos descritivos dos dados geográficos.
- <relationships>: descreve o relacionamento entre as entidades descritas pelo arquivo.

A validação de um arquivo GeoBR é feita por um esquema XML, onde estão definidos os tipos e elementos que podem ser usados para a descrição de dados no formato GeoBR.

4 Conversor de Dados GeoBR

Para viabilizar o uso do formato GEOBR pela comunidade brasileira, estamos desenvolvendo uma ferramenta para auxiliar na tarefa de conversão de dados, baseada em um ambiente gráfico que permite ler e mostrar na tela múltiplos arquivos de dados geográficos, (geometria e atributos) nos seguintes formatos: GEOBR, SHP, MIF/MID, E00 e SPR. Realiza a conversão entre os formatos suportados através de uma interface que permitirá ajustar parâmetros necessários para a tradução de um determinado formato para outro. A construção do software está sendo feita em C++, com o uso da biblioteca de classes TerraLib [2] e do ambiente de interface multi-plataforma Qt [11]. A figura 1 mostra a interface do conversor.

Referencias

- [1] Câmara, G. (1995). *Modelos, Linguagens e Arquiteturas para Bancos de Dados Geográficos*. Ph.D., Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, SP. <www.dpi.inpe.br/teses/gilberto>.
- [2] Câmara, G., R. Souza, B. Pedrosa, L. Vinhas, A. Monteiro, J. Paiva, M. Carvalho and M. Gattass (2000). *TerraLib: Technology in Support of GIS Innovation*. II Workshop Brasileiro de Geoinformática, GeoInfo2000, São Paulo.
- [3] Câmara, G. M., A.M.V.; Paiva, J.A.C; Souza, R.C.M. (2000). *Action-Driven Ontologies of the Geographical Space*. GIScience 2000, Savannah, GA, AAG. <<http://www.giscience.org>>.
- [4] Davis, C., K. Borges and A. Laender (in press). "OMT-G: An Object-Oriented Data Model for Geographic Applications." *GeoInformatica*.
- [5] Egenhofer, M. (1992). "Why not SQL!" *Int. Journal of Geographical Information Systems* 6(2): 71-85.
- [6] Egenhofer, M. and R. Franzosa (1991). "Point-Set Topological Spatial Relations." *International Journal of Geographical Information Systems* 5(2): 161-174.
- [7] FGDC (2001). Content Standard for Digital Geospatial Metadata WorkBook. Reston, VA, Federal Geographic Data Committee.
- [8] Fonseca, F., M. Egenhofer and K. Borges (2000). *Ontologias e Interoperabilidade Semântica entre SIGs*. II Workshop Brasileiro em Geoinformática - GeoInfo2000, São Paulo. <<http://www.tecgraf.puc-rio.br/geoinfo2000>>.
- [9] Hartman, R. (1998). *GIS Data Conversion - Strategies, Techniques and Management*. N.Y., Onword Press.
- [10] OGC, Ed. (1996). *The OpenGIS® Guide - Introduction to Interoperable Geoprocessing and the OpenGIS Specification*. MA, Open GIS Consortium, Inc.
- [11] Troll_Tech (2001). *Qt: The Official Documentation*. Oslo, Noyes Publications.