

GeoBR: intercâmbio sintático e semântico de dados espaciais

PAULO DE OLIVEIRA LIMA JÚNIOR¹

GILBERTO CÂMARA²

(recebido em 18/10/2002; aprovado em 26/11/2002)

PALAVRAS-CHAVE

Sistemas de Informação Geográfica – Ontologia – Interoperabilidade

RESUMO

O intercâmbio de dados espaciais é uma questão importante no uso das geotecnologias, impulsionada principalmente pelo alto custo de produção deste tipo de dado e sua necessidade para produção de informações geográficas. É também um desafio, pois ressalta o problema da interoperabilidade entre os sistemas concebidos para armazenar, visualizar e manipular dados espaciais, denominados Sistemas de Informação Geográfica. Cada Sistema de Informação Geográfica (SIG) é desenvolvido independentemente e utiliza um modelo conceitual próprio sem a preocupação de seguir um modelo ou terminologia comum. O resultado é um ambiente heterogêneo onde cada organização tem sua maneira de tratar a informação espacial, vinculada ao SIG por elas adotado. Distorção de dados e perda de informação são problemas que freqüentemente aparecem em uma troca de dados entre organizações que usam SIGs distintos. Este cenário é propício ao desenvolvimento de modelos e formatos de dados que são capazes de representar diferentes tipos de informação geográfica e promover a interoperabilidade entre SIGs distintos. Este artigo descreve o desenvolvimento de um formato para intercâmbio sintático e semântico de dados espaciais entre diferentes SIGs, o qual inclui avanços para interoperabilidade entre sistemas em geral, como o uso do padrão XML e Ontologias. Descreve ainda um software de código aberto para visualização e conversão de arquivos em formatos de diferentes SIGs do mercado e o formato proposto.

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 dado. A falta de modelos conceituais comuns acarreta problemas na troca de dados entre organizações utilizando Sistemas de Informação Geográfica (SIGs) distintos, que incluem distorção de dados, comprometimento de qualidade da informação, perda de definições de atributos e georreferenciamento. Segundo [Hart98], em um ambiente de sistemas heterogê-

E-mail: plima@ciet.com.br

E-mail: gilberto@dpi.inpe.br

neos, a aquisição de dados representa um custo entre 60% e 80% do custo total na implantação de SIGs.

Atualmente, para modelar objetos e fenômenos georreferenciados, cada SIG utiliza um modelo conceitual próprio. Esta diversidade faz com que organizações produtoras de informação georreferenciada sigam regras conceituais vinculadas ao sistema por elas utilizado. O resultado é um ambiente heterogêneo, onde cada organização tem sua maneira de tratar a informação espacial. Assim, o intercâmbio de dados espaciais resalta o problema da interoperabilidade entre os SIGs, sendo esta uma área de pesquisa que cresce a cada dia.

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) vêm adotando soluções *ad-hoc* para o intercâmbio de dados, muitas vezes sem a devida análise crítica da experiência internacional nesta área. O que cria uma lacuna entre Administrações públicas com convênios para troca de dados.

Em SIGs, alcançar a interoperabilidade não é uma tarefa simples, devido à complexidade da informação geográfica envolvida, ocorrendo incompatibilidades em níveis sintático e semântico. Propostas existentes, como a conversão direta entre formatos de exportação próprios dos SIGs mais comuns no mercado Shape (ArcView), MID/MIF (MapInfo), E00 (ArcInfo) e o uso de formatos independentes como SDTS [USGS98], têm maior ênfase no aspecto sintático. A interoperabilidade plena requer não só uma equivalência sintática entre as entidades representadas pelos sistemas, mas inclui também a equivalência de conceitos e significados dessas entidades. Assim surgem propostas como o uso de metadados [FGDC01], modelos genéricos para dados espaciais [Bau+97], [Dav+01] e Ontologias [Fon+00].

Uma abordagem prática do problema da interoperabilidade em SIGs é o uso de formatos de intercâmbio para dados espaciais. Os SIGs oferecem formatos de exportação próprios, e um processo de conversão tem seus prós e contras, principalmente em um ambiente com uma variedade de sistemas diferentes. Qual o formato a adotar? O formato adotado captura o significado da informação bem como as relações entre as entidades descritas? Estas questões são importantes em se tratando do uso de formatos para intercâmbio de dados espaciais.

Do ponto de vista conceitual, uma das melhores alternativas para enfrentar o problema da interoperabilidade é dispor de modelos de dados espaciais de aplicabilidade geral [Bau+97], [Dav+01]. Esses modelos permitem traduzir os dados produzidos por um SIG específico para uma visão genérica da geoinformação e assim formam a base para a proposição de um formato geral para intercâmbio de dados espaciais.

[Câm+96] comentam que, em geral, padrões facilitam o compartilhamento, a integração e a transferência de dados, sendo padrões para linguagens de especificação, transferência de dados, geocodificação ou documentação de formatos. Padrões para SIG vêm sendo definidos em nível internacional, nacional, federal ou industrial, alguns legalmente estabelecidos outros aceitos por consenso por uma considerável parcela da comunidade.

[Fon+00] ressaltam que problemas semânticos irão persistir e impedir a interoperabilidade e que seria interessante conviver com estas diferentes formas de conhecimento sobre a realidade e tentar criar mecanismos para implementar e combinar diferentes visões, ou seja, representar o conhecimento geográfico no computador buscando interoperabilidade pela equivalência semântica dos conceitos entre sistemas distintos. Neste sentido, são propostos trabalhos relacionados a Ontologias e seu uso para interoperabilidade e concepção de SIGs baseados em Ontologias.

Diante do quadro exposto, este artigo descreve a criação de um formato de intercâmbio de dados espaciais independente de sistema, baseado em um modelo genérico para dados espaciais, que possa preservar a qualidade da informação geográfica (topologia e atributos), e a criação de um Dicionário de Termos, baseado em Ontologia, como mecanismo para correspondência de termos no processo de transferência. Este artigo apresenta também a criação de uma ferramenta computacional de fácil utilização para suportar a conversão entre o formato proposto e alguns formatos existentes. Desta forma pretende-se promover a interoperabilidade nos níveis sintático e semântico.

[Thom98] ressalta que mecanismos de transferência de dados entre formatos distintos esbarram nas regras conceituais que devem ser obedecidas para a absorção desses dados geográficos pelos SIGs. Entretanto, uma ferramenta para conversão, gratuita e de fácil utilização que suporte diferentes formatos, é relevante em um cenário onde temos alternativas com alto custo ou opções de conversão entre formatos isolados MIF [Map01] para SHP [ESRI98], SHP para DXF [Auto02], sem interface gráfica e com poucos recursos para manipulação dos dados.

Do ponto de vista prático, adotou-se a tecnologia XML – eXtensible Markup Language [Bra+00] – como sintaxe para descrever os dados espaciais no formato proposto, tornando-os mais acessíveis e semanticamente ricos, fazendo com que o seu significado seja facilmente compartilhado no processo de transferência. Portanto, serão discutidas novas abordagens do problema da interoperabilidade entre sistemas heterogêneos e autônomos através do uso de linguagens baseadas no padrão XML para descrever metadados RDFSchemas [Las+01] e Ontologias DAML+OIL [Con+02].

Este artigo é dividido em cinco seções. Na segunda seção são apresentadas propostas existentes na comunidade para tratar o problema. Essa seção fornece uma base teórica fundamentada na experiência internacional para elaboração e desenvolvimento prático da proposta aqui descrita. A terceira seção aborda o problema da interoperabilidade em SIGs, a descrição e justificativa do formato GeoBR e do dicionário de termos baseado no conceito de Ontologia. A quarta seção descreve a implementação de uma ferramenta para converter arquivos de dados espaciais comuns no mercado e o formato proposto. A conclusão é o assunto da quinta seção, que também apresenta perspectivas de evolução do trabalho.

2. PROPOSTAS PARA INTEROPERABILIDADE

A questão da interoperabilidade entre SIGs vem sendo tratada pela comunidade de geoprocessamento por iniciativas que buscam soluções nos diferentes níveis: sintático e semântico.

Alguns dos trabalhos mais relevantes são comentados a seguir, como o padrão americano para transferência de dados SDTS (*Spatial Data Transfer Standard*) [USGS98], o consórcio OpenGIS [OGC96] e padrões de metadados como proposto pelo FGDC (*Federal Geographic Data Committee*) [FGDC01]. Estas propostas constituem um conjunto de conceitos que servem de base para o trabalho descrito neste artigo.

No Brasil, devemos citar esforços para compartilhamento de dados por propostas de organizações como [BCD01], [CPRM01], e [Geo01] que disponibilizam na Internet dados georreferenciados sobre Minas Gerais.

2.1 Conversão Sintática – Formatos de Exportação

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. Formato, neste caso, é a estrutura lógica usada para armazenar dados de um SIG em arquivos. Para permitir esse tipo de conversão, os SIGs trabalham com duas alternativas:

- Oferecer um formato de exportação ASCII de fácil legibilidade, como o DXF [Auto02] do software Autocad [Auto00], MID/MIF [Map01] do software MapInfo [Map02], E00 [Mori01] do software ArcInfo [ESRI01b] e SPR [INPE00] do software SPRING [INPE01];
- Documentar as estruturas de dados internas, como no caso do formato *Shapefile* SHP [ESRI98] usado pelo software ArcView 8 [ESRI01a].

Os formatos de intercâmbio podem ser formatos documentados nos quais um determinado sistema organiza seus dados (formato de exportação), facilitando a leitura por outro sistema e possibilitando uma conversão sintática. Ou podem ser independentes de sistema, criados com a finalidade de ser usado apenas para transferência.

O processo de transferência de dados geográficos envolve duas fases. A fase de exportação, que consiste em selecionar uma camada do SIG e convertê-lo para um formato de transferência, e a fase de importação, para converter os dados e inseri-los no sistema de destino. Caso o sistema de destino tenha suporte para o formato recebido, uma conversão não é necessária, pois ocorre a importação direta.

Nem sempre o formato de exportação coincide com o formato interno, que por sua vez nem sempre está documentado. Em alguns sistemas, o formato de exportação está descrito por um arquivo ASCII representando um dado geográfico, contendo as informações necessárias para sua decodificação.

Os formatos incluem informações geográficas: posição e formas das entidades, informações sobre atributos – dados não espaciais e em alguns casos informações de apresentação. Geralmente, um arquivo descreve a geometria (pontos, linhas e áreas) das entidades geográficas e outro descreve seus atributos.

Cada sistema possui uma definição para a geometria das entidades representadas. Podemos dividir os formatos em dois tipos: vetoriais e matriciais (*raster*).

Esses formatos intermediários 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, como é o caso das traduções usando-se o formato DXF [Fon+99]. Existem formatos alternativos, independentes de sistema, que procuram evitar a perda de informação, como o SDTS (Spatial Data Transfer Standard), descrito a seguir.

Existem ainda formatos independentes de sistema para intercâmbio de dados como o SDTS - *Spatial Data Transfer Standard* [USGS98]. O SDTS é projetado especificamente como um formato para transferência de dados, não para uso direto por um sistema específico. Para isto, é especificado em partes que procuram abordar o nível conceitual, o lógico e o físico.

Apesar da decisão do governo dos EUA em padronizar o SDTS para todos os órgãos federais americanos, esse formato apresenta diversos problemas. O modelo conceitual do SDTS tem conteúdo semântico limitado, e está fortemente acoplado às definições do sistema Arc/Info-7. O padrão SDTS não contempla os conceitos de modelagem espacial orientada a objetos, não estabelece definições de metadados, não descreve formalmente relacionamentos espaciais e nem tem formas de capturar procedimentos de consulta e análise. Em resumo, o SDTS comporta-se como um formato de intercâmbio tradicional do tipo SHP ou MIF, sem grandes vantagens adicionais.

2.2 OpenGIS

O consórcio OpenGIS define 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 (*Structured Query Language*) [OGC96]. Esta abordagem segue o conceito de API (*Application Programming Interface*), o que fornece uma forma unificada de acesso às funcionalidades de sistemas distintos.

Apesar dos inegáveis avanços, a proposta OpenGIS tem várias limitações. Para começar, 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.

O padrão *OpenGIS* inclui, até o momento, apenas operações topológicas de consulta sobre objetos simples [Ege+91] sem permitir a definição de relacionamentos espaciais para definição de restrições espaciais, como proposto em [Dav+01]. Os problemas de especialização e hierarquia entre classes de objetos também ainda não foram resolvidos.

Adicionalmente, alguns sistemas existentes têm modelos conceituais mais ricos em conteúdo que o OpenGIS, e seu mapeamento para este pode representar sensível perda de informação semântica [Câm+99].

Em seu modelo abstrato, o padrão OpenGIS denomina *Coverage* como uma metáfora bi-dimensional para fenômenos na superfície terrestre. O OpenGIS relaciona diretamente cada tipo de *coverage* com uma geometria particular, como mostra a Figura 1, num relacionamento de especialização. Isto faz com que o conteúdo semântico de uma *coverage* (os tipos de dados geográficos representados) seja confundido com seu con-

teúdo sintático (a estrutura de dados utilizada). Esta relação é incoerente com a definição proposta pelo padrão OpenGIS para *coverages*, que fala em função, domínio e intervalo.

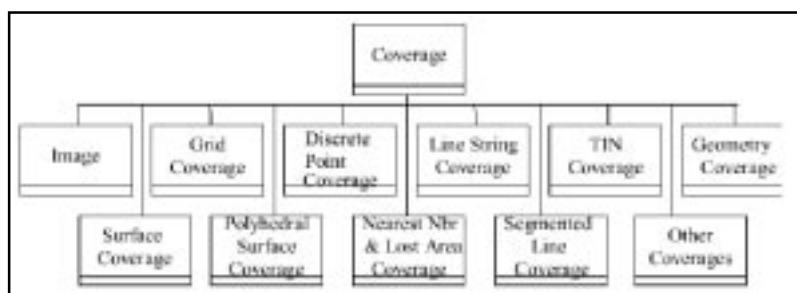


Figura 1 - Subtipos de Coverage FONTE: Adaptada de [OGC96]

Deste modo, o modelo conceitual do OpenGIS não captura a noção que um mesmo geocampo (distribuição de variáveis geográficas contínuas) pode ser representado simultaneamente nos formatos vetorial e matricial.

Finalmente, o uso da linguagem SQL como base para a linguagem de consulta no caso de OpenGIS é questionável. Como mostram [Egen92] e [Câma95], o padrão declarativo do SQL tem diversas limitações para tratar com dados geográficos. O SQL não prevê a existência de uma linguagem de apresentação associada às consultas realizadas, e nem suporta o conceito de que o resultado de consultas retorne objetos e campos, para manipulação posterior.

2.2.1 GML

Seguindo a tendência do uso de padrões para intercâmbio de dados, o OpenGIS usa o padrão XML (eXtensible Markup Language) para definir uma forma de codificar dados geográficos. Para isso especificou a linguagem GML (Geography Markup Language) [OGC02].

O objetivo é oferecer um conjunto de regras com as quais um usuário pode definir sua própria linguagem para descrever seus dados. Para tanto a GML é baseada em Esquema XML (XML Schema) [Fall01]. O Esquema XML define os elementos (*tags*) usados em um documento que descreve os dados. Atualmente a linguagem está em sua versão 2.1.1 e esta inclui Esquemas que contêm os modelos mais básicos de geometria e feições (*features*). Os Esquemas estão publicados nas especificações do OpenGIS [OGC02] e atualmente existem três, a saber:

- *Feature.xsd* – define tipos e elementos concretos e abstratos de acordo com a especificação do OpenGIS.
- *Geometry.xsd* – define a geometria de acordo com a especificação do OpenGIS.
- *Xlinks.xsd* – define formas de ligação entre documentos e elementos dentro de um documento XML.

De posse desses Esquemas, um usuário pode definir seu próprio Esquema para sua aplicação. Mas há algumas exigências a seguir para obter conformidade, descritas em [OGC02]. Desta forma, um desenvolvedor de Esquemas pode criar seus próprios tipos e *tags* derivados da GML e uma aplicação poderá fazer uso dos dados.

Apesar de ser uma iniciativa importante, algumas questões devem ser observadas. A GML 2.1.1 inclui modelos básicos para feições e geometrias. Para representar outros modelos como superfícies, a GML 2.1.1 por si só não é adequada, necessitando ser estendida com a adição de um Esquema de aplicação.

Os Esquemas da GML 2.1.1 sozinhos não são adequados para criar uma instância de documento. Esses devem ser estendidos pela criação de Esquemas de aplicação para domínios específicos, seguindo as regras descritas na especificação; somente assim, pode-se garantir que um Esquema e seus dados serão úteis para um software baseado em GML. Isto exige um investimento na elaboração de Esquemas.

Exemplo:

```
...
<complexType name="AbstractFeatureType" abstract="true">
  <annotation>
    <documentation>
      An abstract feature provides a set of common properties. A concrete
      feature type must derive from this type and specify additional
      properties in an application schema. A feature may optionally
      possess an identifying attribute ('fid').
    </documentation>
  </annotation>
  <sequence>
    <element ref="gml:description" minOccurs="0"/>
    <element ref="gml:name" minOccurs="0"/>
    <element ref="gml:boundedBy" minOccurs="0"/>
    <!-- additional properties must be specified in an application schema -->
  </sequence>
  <attribute name="fid" type="ID" use="optional"/>
</complexType>
...
```

Figura 2 – Fragmento do Arquivo *Feature.xsd*. FONTE: Adaptada de [OGC02]

O fragmento na Figura 2, retirado do Esquema Feature da GML 2.1.1, define um tipo (*AbstractFeatureType*), não é usado para descrever dados espaciais em um arquivo mas é usado na criação de um Esquema de Aplicação de usuário, como a seguir:

```
...
1. <complexType name="hidrografia">
2.   <complexContent>
3.     <extension base="gml:AbstractFeatureType">
4.       <sequence>
5.         <element ref="gml:centerLineOf"/>
6.       </sequence>
7.     </extension>
8.   </complexContent>
9. </complexType>
...
```

Figura 3 – Trecho de um Esquema de Aplicação de Usuário

O trecho ilustrado na Figura 3 descreve a criação de um tipo “hidrografia” que está associado ao tipo *AbstractFeatureType* (linha 3) e conseqüentemente herda todas as suas propriedades definidas no Esquema Feature da GML 2.1.1.

```

...
<element name="rio" type="ex:hidrografia" substitutionGroup="gml:_Feature"/>
...

```

Figura 4 – Definição da tag <rio>

O fragmento ilustrado na Figura 4 mostra a criação de uma *tag* rio do tipo hidrografia, usada para descrever os dados, como mostrado na Figura 5:

```

...
1. <rio>
2.   <gml:description>O rio principal</gml:description>
3.   <gml:name>Rio Grande</gml:name>
4.   <gml:centerLineOf>
5.     <gml:LineString srsName="http://www.opengis.net/gml/srs/epsg.xml#432">
6.       <gml:coord><gml:X>0</gml:X><gml:Y>50</gml:Y></gml:coord>
7.       <gml:coord><gml:X>7</gml:X><gml:Y>60</gml:Y></gml:coord>
8.       <gml:coord><gml:X>1</gml:X><gml:Y>50</gml:Y></gml:coord>
9.     </gml:LineString>
10.   </gml:centerLineOf>
11. </rio>
...

```

Figura 5 – Descrição de Dados de acordo com o Esquema do Usuário

GML 2.1.1 possui pontos, linhas, polígonos e coleções geométricas (MultiPoint, MultiPolygon) definidos por coordenadas cartesianas uni, bi ou tridimensionais associados a eventuais Sistemas de Referência Espacial. Mas as localizações espaciais são definidas apenas por coordenadas cartesianas, coordenadas projetivas não estão previstas. Ainda sobre Sistemas de Referência Espacial, a Seção 4.3.2 *Geometry Elements* da especificação [OGC02] diz que “as coordenadas para uma geometria são definidas em algum Sistema de Referência Espacial (SRS) e toda geometria deve especificar este SRS. GML 2.1.1 não trata dos detalhes da definição de SRS”, assim fica a cargo do usuário criar sua própria forma de tratar o SRS, descaracterizando GML como padrão para intercâmbio.

O Esquema GML GeometrySchema.xsd permite que toda geometria tenha seu SRS associado, o que torna o dado redundante quando representa um conjunto de entidades em um mesmo SRS. Se a associação fosse permitida só a elementos de níveis superiores, como coleções de feições (<featureCollections>), esta redundância seria evitada.

Uma vantagem no uso de XML é a flexibilidade oferecida para criar *tags* que expressam o significado do dado descrito, obtendo-se um documento rico semanticamente. Mas, considere a seguinte situação: Dois usuários de domínios diferentes representam uma determinada entidade, pela GML, como <rio> e <curso_de_agua>. Em uma troca de dados entre os usuários, os Esquemas também devem ser compartilhados, pois só assim uma aplicação poderá saber que <rio> ou <curso_de_agua> são da classe <_Feature> definida pelo Esquema *Feature.xsd* da GML, e então usá-los adequadamente. Desta forma, o problema de acesso aos dados é resolvido. Mas como saber que <rio> é <curso_de_agua> e vice-versa? O aspecto semântico não é considerado de forma efetiva a promover a interoperabilidade. Para amenizar este problema, pode-se acrescentar *tags* que descrevem as entidades e suas relações, ou que identifiquem sinônimos. Mas dentro de um ambiente heterogêneo, o ideal seria a criação de uma forma unificada ou padronizada de realizar isto, definida também nas especificações do OGIS.

A GML tem como base conceitos comuns no domínio dos SIGs, como pontos, linhas e polígonos, e apresenta deficiências de precisão semântica. Como exemplo, temos o caso da definição de linha fechada (*LinearRing*) que não exige na prática que esta seja constituída por mais de um ponto. Ou seja, o fragmento `<LinearRing>
<coordinates>10.0,10.0</coordinates></LinearRing>` é válido, sacrificando o rigor semântico.

Outra questão relevante no uso de GML é a disponibilidade de ferramentas computacionais. Atualmente, uma API em Java para acesso a dados em GML (GML4J) está sendo desenvolvida de forma aberta pela Internet [Sour01]. A sua versão atual é capaz de ler dados e Esquemas GML, desde que estes obedeçam às regras especificadas na Seção 5 da especificação do padrão GML, mas não permite escrever. O software FME (*Feature Manipulation Engine*) [Safe02] suporta a leitura e escrita de dados GML mas não inclui a manipulação de Esquemas de Aplicação.

A validação de dados GML pode ser feita por um *parser* XML. Um programa capaz de ler um arquivo com marcações e identificá-las segundo seu contexto. O *parser* garante a validade de um arquivo de dados segundo os Esquemas GML. Mas não verifica os detalhes exigidos na Seção 5.2 da especificação do padrão GML. Não valida Esquemas de Aplicações de usuários.

Assim, consideramos que os avanços no OpenGIS merecem revisão nas questões levantadas. Parte dessas revisões, como o uso de ontologias para interoperabilidade semântica, uso de um modelo genérico para dados espaciais, definição semântica das geometrias e forma unificada para descrever projeções foram incorporados à proposta GeoBR.

No cenário atual, o uso de XML como padrão para intercâmbio de dados é inquestionável. Mas a simples descrição dos dados por *tags* XML promove a interoperabilidade semântica desde que haja apenas quando houver um software pronto a entender o significado das *tags* e inferir relacionamentos entre entidades produzindo novas informações a partir das que recebeu. O ideal é aproveitar o poder de descrição semântica do padrão XML para estruturar os dados, explicitando relacionamentos entre entidades, de forma que novas informações possam ser inferidas a partir do próprio dado. Portanto, este trabalho propõe o uso de DAML+OIL (*DARPA Agent Markup Language + Ontology Inference Layer*) [Con+02] como forma de descrever ontologias para os dados espaciais e promover a interoperabilidade semântica.

A proposta de vários Esquemas que devem ser estendidos para Esquemas de aplicação é interessante quando se deseja flexibilidade na descrição dos dados, fácil consulta e uso dos dados entre comunidades que adotam o mesmo Esquema. Mas quando se trata de intercâmbio de dados, a criação de Esquemas de aplicação deve prever fatores que tornam esse trabalho mais árduo, como estrutura lógica dos dados e tamanho dos arquivos gerados. Portanto é necessário um investimento na criação de Esquemas de aplicação. Além disso, a existência de diferentes Esquemas de aplicação acarreta em esforço para conversão de dados. Os Esquemas a serem estendidos devem, por sua vez, incluir suporte para todos os tipos de dados espaciais (campos e objetos) a fim de oferecer compatibilidade com dados existentes em outros formatos.

2.3 Uso de Metadados

Metadados descrevem o conteúdo, a condição, o histórico, a localização e o formato do dado. O objetivo do seu uso é ter um mecanismo para identificar qual dado existe, a sua qualidade e como acessá-lo e usá-lo. 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 [FGDC 01].

O padrão FGDC estabelece um conjunto comum de definições para a documentação do dado geográfico, incluindo: identificação, qualidade do dado, organização espacial do dado, referência espacial, informação sobre entidade e atributo, distribuição e referência do metadado. O FGDC também patrocina a criação de uma *clearinghouse* (*National Geospatial Data Clearinghouse*). Trata-se de um site que guia usuários ao melhor dado espacial para seus projetos por meio de pesquisa a metadados disponibilizados no padrão do FGDC por órgãos produtores de dados espaciais.

Como sua ênfase é na disponibilidade da informação, o padrão FGDC não especifica a maneira pela qual a informação está organizada nem o processo de transferência. Com exceção da parte de entidades e atributos, que pode revelar parcela do significado do dado, as demais partes não descrevem a semântica da informação espacial.

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.

Em resumo, a substancial burocracia envolvida em adotar o padrão FGDC não se traduz em benefícios proporcionais. Estes fatos talvez expliquem por que sua adoção ainda está limitada e por que o consórcio OpenGIS propõe seu próprio formato para metadados.

2.4 Ontologias

[Fon+00] ressaltam que problemas semânticos irão persistir e impedir a interoperabilidade, e são claramente os mais difíceis nesta área. Diferentes visões da realidade geográfica sempre existirão em pessoas com culturas diferentes, pois a própria natureza é complexa e leva a percepções distintas. Neste caso seria interessante conviver com essas diferentes formas de conhecimento sobre a realidade e tentar criar mecanismos para implementar e combinar diferentes visões, ou seja, representar o conhecimento geográfico no computador buscando interoperabilidade pela equivalência semântica dos conceitos entre sistemas distintos. Neste sentido, são propostos trabalhos relacionados a Ontologias e seu uso para interoperabilidade e concepção de SIGs baseados em Ontologias [Fon+00], [Câm+00c].

Segundo [Fon+99], no sentido filosófico Ontologia é um sistema específico de categorias que reflete uma visão específica do mundo, e Ontologia como uma ferramenta de engenharia descreve uma certa realidade com um vocabulário específico usando um conjunto de premissas de acordo com o sentido intencional das palavras do vocabulário.

A interoperabilidade plena requer não só uma equivalência sintática entre as entidades representadas pelos sistemas, mas inclui também a equivalência de conceitos e significados dessas entidades. Por exemplo, duas comunidades de informação podem utilizar nomes diferentes para o mesmo conceito (como no caso de “rio” e “curso de água”). Ou ainda, um único conceito para uma comunidade (i.e., “rio”) pode ser expresso com níveis maiores de detalhe por outra (i.e., “rios perenes”, “rios temporários”, “riachos”). para um determinado grupo de pessoas, um rio pode ser chamado de curso d’água e considerado de um tipo específico do SIG adotado. Para outro grupo, a mesma entidade, pode ser chamada de rio e ser de outro tipo próprio do mesmo SIG, ou de outro. Neste sentido, ressalta-se a necessidade que os formatos de intercâmbio de dados disponham de necessidade de um mecanismo que siga suporte o conceito de Ontologias e comunidades de informação geográfica, para que interpretações diferentes de uma mesma realidade geográfica possam ser identificadas e facilmente trocadas.

2.5 Ferramentas

Apesar das limitações de uma conversão sintática, deve-se reconhecer que a grande maioria dos processos de conversão de dados opera neste nível [Hart98]. Para facilitar esse processo, estão disponíveis ferramentas de conversão, tanto sob forma de programas comerciais quanto disponíveis como *open source* [Free01].

Na área comercial, destaca-se a ferramenta FME - *Feature Manipulation Engine* [Safe02], que além de possibilitar a conversão entre vários formatos comuns no mercado, dispõe de uma interface com o usuário, de fácil uso, que permite manipular e configurar a conversão incluindo funções de transformações topológicas, operações geométricas e sobre atributos e conversões entre projeções. Geralmente as ferramentas disponíveis gratuitamente não passam de simples algoritmos de conversão entre formatos específicos, muitas vezes sem interface gráfica e não dispõem de mecanismos para manipulação do dado. E ferramentas comerciais, como a apresentada, têm alto custo.

3. GEOBR: DESCRIÇÃO GERAL

Na seção anterior, verificamos que para alcançar um alto grau de interoperabilidade na conversão de dados geográficos entre sistemas distintos é preciso atuar nos níveis sintático e semântico. Desta forma, ressalta-se a necessidade de um formato independente e acessível que inclua geometria e atributos das entidades bem como explicita o seu significado, e um mecanismo para manipular e transferir tal significado. O formato GeoBR pretende se diferenciar das propostas descritas neste documento 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 dos dados.
- Suporte à conversão semântica pelo uso de um Dicionário de Termos (Ontologia).
- Suporte para incorporação futura de procedimentos de análise geográfica e relacionamentos entre entidades.
- O formato GeoBR é especificado em XML (eXtensible Markup Language) [Bra+00], o que torna seu conteúdo mais acessível, promovendo o intercâmbio de dados via Internet.
- A proposta do GeoBR é um Esquema único, com elementos pré-definidos, o que faz com que um arquivo GeoBR seja facilmente acessado por uma única interface de programação.
- O esquema GeoBR também define formalmente seus elementos geométricos, assim um elemento <line> com menos de duas coordenadas não é válido.

O formato GeoBR difere dos formatos genéricos de intercâmbio, como DXF [Auto02], SHP [ESRI98], e MIF [Map01], pois enquanto esses 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.

3.1 Uso de XML como padrão de transferência. Por que XML?

O problema da interoperabilidade não é exclusividade para usuários de SIGs, mas é uma questão para todos os sistemas computacionais, e o padrão XML foi projetado com o propósito de resolver este problema.

O padrão XML (eXtensible Markup Language) ou linguagem de marcação extensível usa marcações (*tags*) para codificar o conteúdo de um documento, assim descreve dados, e não sua apresentação, como o exemplo mostrado na Figura 6:

A imagem mostra um exemplo de código XML dentro de um retângulo com uma sombra 3D. O código consiste em uma tag de abertura <pessoa>, seguida por duas tags de atributo <nome>Pedro</nome> e <idade>5</idade>, e uma tag de fechamento </pessoa>. Há pontos de reticências (...) antes e depois do bloco principal.

```
...  
<pessoa>  
    <nome>Pedro</nome>  
    <idade>5</idade>  
</pessoa>  
...
```

Figura 6 - Exemplo de Dados em XML

A forma como os dados descritos serão interpretados e apresentados é função da aplicação que os usará.

Há duas principais vantagens no uso de XML neste trabalho. Além de ser ideal para descrever estruturas de dados hierárquicas e complexas, características comuns em dados espaciais, e permanecer legível, pois o dado é armazenado no formato ASCII (seguindo a sintaxe XML, ao invés de uma forma binária de codificação), também é ideal para tornar o dado mais acessível através de *parsers* e APIs padronizadas. Outra característica importante é que o dado permanece rico semanticamente e seu significado é compartilhado no processo de transferência. Além dessas vantagens, o uso de XML vem aumentando e este é cada vez mais afirmado como um padrão independente de sistema para armazenar e trocar dados.

3.2 Dicionário de Termos (Ontologias)

Comunidades distintas descrevem dados espaciais segundo o vocabulário próprio de cada uma. Esta diferença de nomenclatura entre dados e modelos deve ser considerada no processo de intercâmbio de dados para que não ocorra perda de informação. Uma forma de promover a equivalência de significados dos dados espaciais é utilizar um mecanismo contendo o vocabulário usado para descrever as entidades envolvidas no processo de transferência, seu significado e relacionamentos. Este mecanismo utilizado de forma unificada promove a interoperabilidade semântica de dados espaciais. Portanto, este trabalho propõe o uso de um Dicionário de Termos como mecanismo para definir uma determinada Ontologia.

Para formalizar seus dados de acordo com o vocabulário utilizado por uma dada comunidade, o GeoBR está aberto a inclusão de ontologias através do Dicionário de Termos, indicado por <ontology>. Este elemento indica um arquivo que contém informação semântica sobre os dados descritos no arquivo GeoBR. O trabalho procura manter o Dicionário de Termos e um arquivo GeoBR independentes, para não restringir o acesso aos dados à necessidade de um dicionário, e para que um Dicionário possa ser enriquecido com novos termos sem necessidade de alteração nos dados. Portanto, a seguir será apresentado o uso do Dicionário e em seguida o formato GeoBR. Para explicar a estrutura e o uso de um Dicionário de Termos serão apresentados conceitos e novas abordagens relativas ao uso de Ontologias, linguagens para descrição destas, e um exemplo.

Comunidades de geoinformação possuem, cada uma, um consenso próprio em torno de definições de representação para as entidades de seu domínio. Essas definições também devem ser informadas no processo de intercâmbio, possibilitando o compartilhamento de significado do dado envolvido.

A idéia do uso de um Dicionário de Termos é criar comunidades de geoinformação capazes de explicitar seus conceitos em um arquivo e compartilhar esta informação. A intenção é que o dicionário aqui proposto inclua apenas informações semânticas sobre entidades do mundo real. E que seja estruturado para ser compartilhado independentemente. Assim, em um processo de transferência de dados, consideram-se dois arquivos, um com os dados no formato GeoBR e outro opcional, contendo o significado das entidades descritas pelo primeiro servindo como mecanismo para uma aplicação de equivalência de conceitos com a finalidade de promover a interoperabilidade em nível semântico.

Entretanto, a simples descrição formal das entidades não é suficiente para expressar seu verdadeiro significado. É preciso explicitar também o relacionamento entre as entidades descritas. Em aplicações como cadastro urbano, redes e para procedimentos de Análise Espacial, os relacionamentos são relevantes, como forma de explicitar restrições de integridade espacial do modelo. Neste sentido, Ontologias têm sido usadas para expressar a semântica de dados em várias áreas, tendo destaque no projeto Semantic Web [Ber+02], e em propostas como o uso de RDF (*Resource Description Framework*) Schemas [Las+02] e DAML – DARPA Agent Markup Language [Burk02].

Seguindo a tendência do uso de linguagens de marcação para representação e compartilhamento de conhecimento, o trabalho considera o uso de DAML – DARPA Agent Markup Language para o Dicionário Ontológico.

3.2.1 *Semantic Web*

Semantic Web é uma extensão da Web atual na qual a informação tem um significado bem definido possibilitando computadores e pessoas trabalharem em cooperação [Ber+02]. Neste sentido, a *Semantic Web* é uma visão: a idéia de ter dados na Web definidos e ligados para serem usados por máquinas não apenas para apresentação, mas para automação, integração e reuso entre aplicações. Para isto, os dados devem ser descritos de forma que a máquina entenda seu significado.

3.2.2 *RDF (Resource Description Framework)*

A idéia da *Semantic Web* não é facilmente implementada pois o que acontece na Web é que os dados são originalmente concebidos para consumo humano e leitura pela máquina. Uma solução, já destacada neste trabalho, é o uso de metadados. Visando promover a idéia da *Semantic Web*, o *World Wide Web Consortium* (<http://www.w3c.org>) especificou uma base comum para a definição de metadados, a RDF [Las+02]. Sua principal meta é definir um mecanismo para descrever recursos (entidades) independentemente do domínio de uma aplicação em particular, ou não define, *a priori*, a semântica do domínio de uma aplicação específica.

RDF (*Resource Description Framework*) inclui um modelo abstrato de dados como estrutura conceitual para definição e uso de metadados. Como sintaxe para codificação usa o padrão XML. Como exemplo, adaptado de [Las+02], a seguinte sentença “Paulo Lima é o criador do recurso <http://www.dpi.inpe.br/~lima>” pode ser representada pelo seguinte fragmento RDF:

```
...
<rdf:RDF>
  <rdf:Description about="http://www.dpi.inpe.br/~lima">
    <s:criador>Paulo Lima</s:criador>
  </rdf:Description>
</rdf:RDF>
...
```

3.2.3 DAML (*DARPA Agent Markup Language*)

À medida que o padrão XML e RDF expandiram para atender a iniciativa da Semantic Web, apareceram limitações devido a sua simplicidade e um novo padrão foi proposto, o DAML (*DARPA Agent Markup Language*), [Burk02]. DAML é uma extensão de RDF (*Resource Description Framework*) que inclui a definição de marcações representando propriedades que RDF não tem como de equivalência (*childOf*) e restrição (*disjointWith*). A última versão DAML+OIL (*Ontology Inference Layer*) [Con+02] inclui um conjunto de construções para a criação de Ontologias que permite descrever a informação de forma a ser lida e entendida por máquinas.

A seguir um exemplo do uso do padrão DAML+OIL descrevendo entidades presentes em um arquivo GeoBR:

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:xsd="http://www.w3.org/2000/10/XMLSchema#"
  xmlns:daml="http://www.daml.org/2001/03/daml+oil#"
  xmlns:dex="http://www.daml.org/2001/03/daml+oil-ex#"
  xmlns:exd="http://www.daml.org/2001/03/daml+oil-ex-dt#"
  xmlns="http://www.daml.org/2001/03/daml+oil-ex#"
>
<daml:Ontology rdf:about="">
<daml:versionInfo>Exemplo </daml:versionInfo>
  <rdfs:comment>
    Uma ontologia exemplo.
  </rdfs:comment>
  <daml:imports
    rdf:resource="http://www.daml.org/2001/03/daml+oil#"
  </daml:Ontology>
<daml:Class rdf:ID="Unidade Urbana">
  <rdfs:label>Unidade Urbana</rdfs:label>
  <rdfs:comment>
    Qualquer elemento presente em uma cidade.
  </rdfs:comment>
</daml:Class>
<daml:Class rdf:ID="Lote">
  <rdfs:comment>Unidade Urbana sem construção</rdfs:comment>
  <rdfs:subClassOf rdf:resource="#Unidade Urbana"/>
</daml:Class>
<daml:Class rdf:ID="Edificação">
```

```

<rdfs:subClassOf rdf:resource="#Unidade Urbana"/>
<daml:disjointWith rdf:resource="#Lote"/>
</daml:Class>
<daml: Class rdf:ID="Componente Hidrologico">
  <rdfs:label>Componente Hidrologico</rdfs:label>
  <rdfs:comment>
    Entidade Referente à Hidrologia
  </rdfs:comment>
</daml:Class>
<daml: Class rdf:ID="Rio">
  <rdfs:subClassOf rdf:resource="#Componente Hidrológico">
</daml:Class>
</rdf:RDF>

```

O simples fato de descrever dados usando o padrão XML faz com que um computador seja capaz de, baseado nesses dados, produzir informação nova. Mas apenas pela existência de um software que não é parte do padrão XML. Já um conjunto de afirmações em DAML pode permitir, por si só, a conclusão de outra afirmação DAML.

O dicionário constitui também uma solução facilmente utilizada por ferramentas para tradução semântica. Desta forma, pretende-se fazer com que o dicionário se apresente como um mecanismo para promover a interoperabilidade em nível semântico.

Um importante aspecto da tecnologia XML é a inclusão de *namespaces*. No dicionário de termos, os *namespaces* são usados para indicar domínios diferentes em um mesmo documento. Assim, entidades com mesmo nome e significados diferentes podem co-existir em um mesmo documento.

3.3 Descrição do Formato GeoBR

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 (*tag*) XML, a saber:

<metadata>: Informações sobre a produção do dado (uma secção por arquivo, opcional);

<data_model>: Descrição das entidades reais representadas no formato de acordo com um metamodelo genérico que inclui os tipos de dados geográficos;

<projection>: Descreve uma projeção sob a qual estão as camadas de dados (<layers>) representadas no formato de acordo com um metamodelo genérico que inclui os tipos de dados geográficos;

<layers>: Camadas de informação geográfica;

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.

3.3.1 Metadados

Nesta seção a idéia é resgatar o que há de essencial no FGDC para incluir, de forma simplificada, informações sobre quando, onde, como e quem produziu os dados. Esta seção inclui, até o momento, as seguintes informações:

<originator>: Instituição que compilou os dados originalmente e que detém o *copyright* deles.

<producer>: Instituição que produziu os dados (pode não ser o detentor do *copyright*).

<reference_date>: No formato numérico DD/MM/AAAA, indica a data em que foi realizado o levantamento.

<production_date>: No formato numérico DD/MM/AAAA, indica a data em que os dados foram produzidos.

<description>: Descrição do dado.

<ontology>: Indica o arquivo contendo o Dicionário Ontológico, com informação semântica sobre os dados.

<layers_info>: Informações individuais sobre cada camada, sendo cada uma identificada por <layer> que por sua vez é dividida em: <layer_name> - nome do layer, <resource_map> - descrição do mapa fonte, <digitalization> - informações sobre o processo de digitalização, <platform> - software utilizado para gerar a camada de dados, <data_structure> - tipo de dado vetorial ou matricial.

3.3.2 Modelo Genérico para Dados Geográficos

A comunidade SIG considera duas categorias de dados espaciais: os campos (*fields*), que correspondem a variáveis geográficas contínuas, e os objetos (feições ou *features*), entidades individuais, e independentes de seu entorno, [Couc92]. Apesar de discussões, esta dicotomia é um consenso básico na comunidade de pesquisa em geoinformação, e SIGs mais recentes, como os sistemas ARC-INFO 8, SPRING e a proposta do consórcio OpenGIS implementam esta visão. [Bau+97] e [Dav+01] também adotam esta divisão na proposta de modelos abstratos para representação de aplicações de SIG.

O formato inclui um modelo genérico de dados geográficos, derivado das noções de campos e objetos, para estabelecer uma base comum na qual diferentes sistemas possam buscar equivalência. Cada camada de dados específica será derivada de um tipo definido no elemento <data_model> de acordo com o modelo:

Feições: dados geográficos individualizáveis, como municípios, ocorrências de flora e fauna, lotes. Representados por <feature>. Cada elemento <feature> no GeoBR tem uma ou mais geometria e atributos não espaciais <attributes>, e podem estar associados a uma classe descrita pelo Dicionário Ontológico. Exemplo:

```
...
<data_model>
  <feature type="lote" rdfs:resource="#Lote">
    <attribute name="uso" type="string"/>
```

```

        <attribute name="area" type="string"/>
    </feature>
    <feature type="poste" rdfs:resource="#Poste">
        <attribute name="altura" type="string"/>
        <attribute name="responsável" type="string"/>
    </feature>
    <feature type="cabo" rdfs:resource="#Cabo">
        <attribute name="extensão" type="string"/>
        <attribute name="material" type="string"/>
    </feature>
    <feature type="poste" rdfs:resource="#Poste">
        <attribute name="potencia" type="string"/>
        <attribute name="responsável" type="string"/>
    </feature>

</data_model>
...

```

Redes: entidades conectadas em topologia arco-nó, onde são incluídas informações descritivas através de associação com Feições definidas em <feature> tanto aos nós quanto aos arcos, por exemplo, rede elétrica ou fluvial. Representadas por <network>. Exemplo:

```

...
<network name="rede_elétrica" rdfs:resource="#Rede Elétrica">
    <node_types>poste,transformador</node_types>
    <arc_types>cabo</arc_types>
</network>
...

```

Superfície: expressam a variação contínua de uma grandeza quantitativa, na forma de valores reais. São utilizados para representar valores como altimetria, incidência de poluentes e dados geoquímicos e geofísicos. Representados por <surface> no GeoBR. Exemplo:

```

...
    <surface name="altimetria" rdfs:resource="#Altimetria"/>
...

```

3.3.3 Projeções

Cada dado geográfico transmitido pelo formato GeoBR deve estar necessariamente acompanhado da informação sobre a projeção geográfica utilizada. As projeções utilizadas são definidas pelo elemento <projections> e incluem dados sobre o nome da projeção, modelo da Terra utilizado, hemisfério, latitude e longitude de origem, paralelos padrão, escala, como o exemplo:

```

...

```

```

<projections>
  <projection name="utm01">
    <datum name="SAD-69"/>
    <radius>10000</radius>
    <flat>10000</flat>
    <x_shift>0.0</x_shift>
    <y_shift>0.0</y_shift>
    <z_shift>0.0</z_shift>
  </projection>
</projections>
...

```

Várias projeções podem ser definidas nesta seção e posteriormente utilizadas pelas diferentes camadas de informação na seção <layers>.

3.3.4 Descrição dos Dados

No GeoBR os dados são representados em camadas independentes incluídas na seção <layers>. Cada camada é identificada pelo elemento <layer> que possui respectivamente um nome, uma projeção definida em <projections> e uma classe definida em <data_model>, como no exemplo:

```

<layer name="lotes" projection="utm01" class="lote">

```

Sob o <layer> estão definidas as coordenadas do retângulo envolvente <box> em seguida a geometria e atributos para o caso de Feições ou Redes do modelo genérico.

```

...
<layers>
  <layer name="lotes" projection="utm01" class="lote">
    <box>
      <x1>12778</x1>
      <y1>6737</y1>
      <x2>81882</x2>
      <y2>83878</y2>
    </box>
    <feature id="1">
      <polygon>
        <xy>20,30</xy>
        <xy>40,50</xy>
        <xy>30,60</xy>
        <xy>20,30</xy>
      </polygon>
      <attributes>comercial, 240m2</attributes>
    </feature>
  </layer>
</layers>

```

```

</feature>
  </layer>
<layer name="cemig" projection="utm01" class="rede eletrica">
  <box>
    <x1>12778</x1>
    <y1>6737</y1>
    <x2>81882</x2>
    <y2>83878</y2>
  </box>
  <node id="1" x="100" y="50">
    <attributes type="Poste">14m,Cemig</attributes>
  <attributes type="Transformador">1000,Cemig</attributes>
  </node>
  <node id="2" x="50" y="40">
    <attributes type="Poste">14m,Cemig</attributes>
  </node>
  <arc id="1" o_node="1" d_node="2">
    <line>
      <xy>100,50</xy>
      <xy>60,60</xy>
      <xy>70,65</xy>
      <xy>50,40</xy>
    </line>
    <attributes type="cabo">15m,cobre</attributes>
  </arc>
</layer>
</layers>

```

...

O exemplo mostra duas camadas de informação representadas no GeoBR, uma camada com lotes e seus atributos e outra com uma rede elétrica e seus elementos.

3.3.4.1 Geometria

O formato inclui as primitivas básicas presentes na maioria dos SIGs de mercado [ESRI01a], [INPE01], [Map02]: pontos, linhas e polígonos além de coleções (PointSet, LineSet e PolygonSet) para entidades representadas por mais instâncias de uma primitiva. A geometria e os respectivos elementos do GeoBR são apresentados a seguir:

Pontos 2D: localizações individuais, associadas a descrições e atributos. Representados por <point> no arquivo GeoBR. Compostos por apenas um par de coordenadas (x,y).

Coleção de Pontos: conjunto de coordenadas utilizadas para representar feições compostas por mais de um ponto. Representados por `<pointSet>`, contendo pelo menos dois ou mais elementos `<point>`.

Linhas: coordenadas vetoriais em 2D, usualmente associadas a dados com topologia arco-nó ou arco-nó-polígono. Podem estar associadas a dados dos tipos Feições ou Redes. Representadas por `<line>`. Compostos por pelo menos dois pares de coordenada (x,y).

Coleção de Linhas: conjunto de coordenadas utilizadas para representar feições compostas de mais de uma linha. Representadas por `<lineSet>`, contendo pelo menos dois ou mais elementos `<line>`.

Polígonos: coordenadas de entidades do tipo região, que usualmente delimitam feições individuais. São usados para dados do tipo Feição. Representados por `<polygon>`. Compostos por pelo menos três coordenadas (x,y).

Coleção de Polígonos: coordenadas para feições com mais de uma região. Representados pelo elemento `<polygonSet>`. Neste caso, o elemento `<polygonSet>` contém dois ou mais elementos `<polygon>`. Por exemplo, o Estado do Pará é constituído de vários polígonos, representando seus limites.

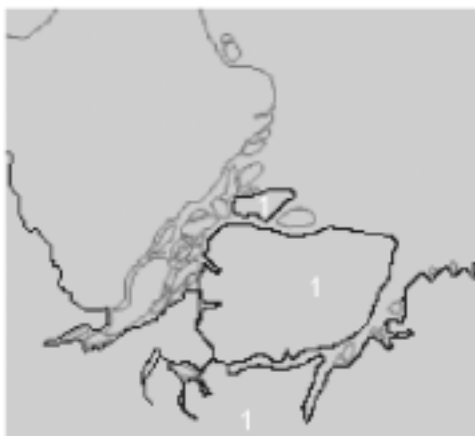


Figura 7 - Polígonos Representando a mesma Entidade

A Figura 7 mostra o Estado do Pará, e seus limites destacados em linhas vermelhas.

Nó: correspondem aos nós em entidades de topologia arco-nó, e que devem ser identificadas individualmente. São usadas em dados do tipo Rede. Representados no GeoBR por `<node>`.

Arco: correspondem aos arcos em entidades de topologia arco-nó, e que devem ser identificadas individualmente. São usadas em dados do tipo Rede. Representado por `<arc>`.

Amostras 3D: localizações individuais, associadas a amostras de uma grandeza quantitativa, do tipo (X,Y,Z). Representadas por `<sample>`. Possuem apenas uma coordenada (x,y) e um valor associado.

Ex:`<sample value="200" x="1233" y="4564">`

Isolinhas: linhas associadas a regiões de mesma cota, usadas para transmitir informações 3D na forma de contornos. Representadas por `<countour>`. Ex:

```
<countour value="100">
<xy>231,324</xy>
<xy>241,300</xy>
<xy>238,466</xy>
</countour>
```

Grade: grade regular consistindo de valor com espaçamento regular nas direções horizontal e vertical, onde os valores podem ser inteiros ou reais (dados de MNT). Representada por <grid>. Ex:

```
<grid rows="3" cols="4" resX="30" resY="30" x="100" y="100">
  <row>23,12,54,94</row>
  <row>98,88,56,44</row>
  <row>27,11,48,28</row>
</grid>
```

O formato não prevê primitivas para o intercâmbio de imagens, pois sua transmissão em arquivos ASCII pode ser inviável devido ao volume de dados. Sugerimos o uso do formato GeoTIFF [Rit+00], por já estar padronizado *de facto*.

3.3.4.2 Atributos

Os dados não espaciais das entidades descritas pelo GeoBR são representados como atributos dos tipos Feição ou Rede do modelo genérico. Os nomes dos atributos são definidos na seção <data_model>. No exemplo abaixo, a entidade “lote”, do tipo Feição <feature> tem dois atributos, “uso” e “área”. O modelo de Redes informa quais Feições definidas por <feature> podem estar associadas aos seus nós e arcos.

```
...
<data_model>
  <feature type="lote" rdfs:resource="#Lote">
    <attribute name="uso" type="string"/>
    <attribute name="area" type="string"/>
  </feature>
  ...
</data_model>
...
```

O valor dos atributos de uma feição é descrito, na seção <layer>, de acordo com a sequência definida em <data_model>, podendo uma feição ter mais de um atributo separado por vírgulas. Exemplo: <attributes>comercial, 240m2</attributes>

3.3.5 Procedimentos de Análise

A grande vantagem do uso de SIGs para lidar com dados geográficos é a possibilidade de produzir novas informações baseadas nas existentes. Para isto

os SIGs disponibilizam procedimentos como Álgebra de Mapas e Análise Espacial.

Em sua primeira versão o formato GeoBR não inclui mecanismos para abordar tais procedimentos, mas é concebido para permitir esta extensão. Uma opção é a inclusão de uma linguagem de análise espacial, que descreva formalmente operadores no modelo. A sequência resultante da aplicação dos operadores é incluída no formato disponibilizando um histórico de transformações a que foi submetido o dado antes de ser descrito.

3.4 GeoBR x GML: Uma Comparação

A tecnologia XML também foi proposta pelo consórcio OpenGIS pela especificação da GML (*Geography Markup Language*), que é um conjunto de regras que são estendidas para codificar informação geográfica em XML [OGC 96].

A GML é composta de Esquemas XML divididos em três arquivos *feature.xsd*, *geometry.xsd* e *xlinks.xsd* que definem respectivamente o modelo abstrato, a geometria e a ligação entre documentos ou partes de documentos. Esses arquivos devem ser estendidos para criar outros Esquemas XML para aplicações de domínios específicos. Assim, cada instituição pode criar sua própria forma de armazenar seus dados. Supõe-se que grandes instituições irão criar seus próprios Esquemas GML ou adotar o Esquema de terceiros e então este será usado por seus softwares promovendo a interoperabilidade.

O intercâmbio de dados entre instituições que tenham Esquemas GML diferentes requer a conversão entre Esquemas que pode ser feita por XSLT [Clar99] (regras para transformar arquivos XML) sendo considerada trivial, desde que os Esquemas sejam similares, ou outra linguagem.

As principais diferenças entre o GeoBR e a GML são:

- GeoBR 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. Ele é um Esquema XML. Instâncias de documentos XML são criadas a partir do Esquema GeoBR.
- GML requer que cada instituição defina seu Esquema de dados, o que implica em investimento adicional para conversão de dados.
- GeoBR tem definições de diferentes tipos de dados (geocampos e geo-objetos), enquanto a versão atual do GML tem suporte apenas para geo-objetos simples (*simple feature*).
- GeoBR inclui uma forma unificada para intercâmbio em nível semântico.

Em tese, quando estiverem disponíveis versões de GML que suportem tipos de dados espaciais mais complexos, poderemos escrever o padrão GeoBR como um esquema XML específico em GML.

4.CONVERSOR-TERRATRANSLATOR

Como parte deste trabalho, está sendo desenvolvido um software para materializar a proposta teórica (GeoBR) descrita neste artigo. O software é denominado TerraTranslator.

4.1 Descrição Geral

O propósito do TerraTranslator é auxiliar a tarefa de conversão de dados entre formatos provenientes de diferentes sistemas. A pretensão é suportar a leitura e escrita nos formatos mais comuns, a saber: Shape Files (ArcView – ESRI), MapInfo Interchange File (MapInfo), E00 (ArcInfo – ESRI), ASCII Spring (SPRING – INPE), e o formato proposto GeoBR. Pretendemos ainda suportar dados no Padrão GML – *Geography Markup Language* (OpenGIS). Neste caso, elementos descritos nos Esquemas do Padrão GML, e não Esquemas de Aplicação de usuários. Para cumprir seu propósito, o TerraTranslator é baseado em um ambiente gráfico com as seguintes funcionalidades:

Visualização – ler e mostrar na tela múltiplos arquivos de dados geográficos (geometria e atributos), inicialmente nos seguintes formatos: SHP (ArcView - ESRI), MIF (MapInfo) e dados no formato proposto. Disponibilizará também recursos para controlar a visualização, como zoom e manipulação de janelas e tabela de atributos.

Conversão – conversão entre os formatos suportados através de uma interface que permitirá ajustar parâmetros necessários para a tradução de um determinado formato para outro.

O TerraTranslator é concebido como um software *open source*, portanto considera-se sua documentação, o próprio código que estará devidamente comentado e disponível via Internet. A seguir é descrito o desenvolvimento do sistema em duas etapas: Projeto da Interface e Implementação do tradutor.

4.2 Implementação do Tradutor

A construção do software segue a metodologia de Orientação a Objetos [Mey97], [Booc94], como linguagem de programação, será usada C++, [Stro00].

A implementação do tradutor conta ainda com a adoção de bibliotecas específicas, a saber:

TerraLib [Câm+00b] para suporte a dados espaciais, Qt [Trol01] para interface com o usuário e SAX (*Simple API for XML*) [Megg01] como *parser* XML. Outra API para acesso a dados XML é DOM (Document Object Model), adequada para muitas aplicações, seu uso requer maiores recursos do sistema, especialmente se o documento é grande, e em geral dados geográficos geram arquivos grandes. Neste caso, SAX é mais adequada.

Atualmente os formatos suportados pelo tradutor são: Shape Files (ArcView – ESRI), MapInfo Interchange Files (MapInfo) e GeoBR - feições. Os dados são mostrados em documentos que possuem uma área para visualização da

geometria e uma tabela para os atributos relativos a esta. Vários arquivos podem ser abertos ao mesmo tempo, como mostra a Figura 8.

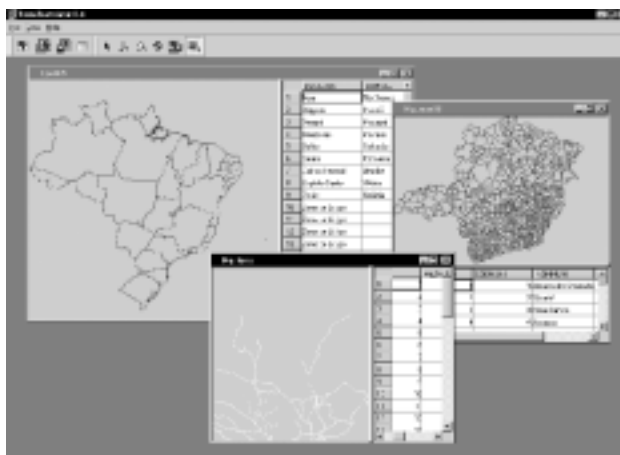


Figura 8 - Dados Provenientes de Arquivos do Tipo Shape Files (shp) e MapInfo Interchange Files (mif)

Para exportar dados no formato GeoBR é apresentada uma caixa de diálogo onde se podem ajustar os parâmetros para a criação do arquivo, a saber, Metadados (*Metadata*), Projeções (*Projections*) Atributos (*Attributes*), Informação Semântica (*Dictionary*) e Modelo de Dados (*Data Model*).

O usuário pode escolher dentre os atributos das feições importadas, apenas aqueles que achar necessário exportar no formato GeoBR. Em versões posteriores será incluída uma forma de mudar o nome dos atributos de acordo com o dicionário ontológico.

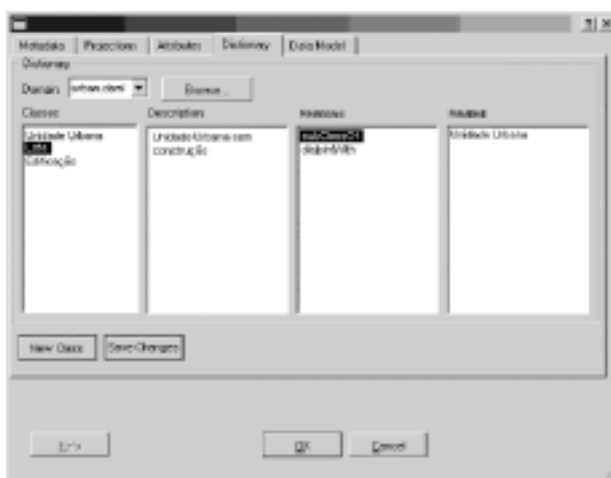


Figura 9 - Dicionário de Termos

O usuário pode abrir um arquivo DAML contendo o dicionário de termos e verificar as entidades e seus relacionamentos, bem como criar novas entidades e incluir no dicionário. Após abrir um dicionário, o usuário poderá utilizar as classes definidas nele para descrever seus dados. A interface de manipulação do dicionário é ilustrada na Figura 9.

5. CONCLUSÕES E PERSPECTIVAS

O avanço da tecnologia em Computação inclui soluções aplicáveis ao problema da interoperabilidade entre SIGs, que vêm aprimorar ou suprir as iniciativas existentes. O estudo mostra que para acontecer um alto grau de interoperabilidade é preciso atuar nos níveis sintático e semântico. Em propostas existentes que oferecem uma forma, independente ou não de sistema, para intercâmbio de dados espaciais, o aspecto semântico é restrito, exemplo – SDTS [USGS98], em outros casos não é relevante, exemplo – formatos de exportação. E o aspecto sintático, nestes casos, não incorporou os recentes avanços como a afirmação do padrão XML.

Soluções que consideram os aspectos mencionados, como é o caso do padrão GML2.1.1 [OGC02], ainda dependem de revisões para a inclusão de novos tipos de dados espaciais. Versões futuras prometem a incorporação de superfícies (*coverages*) como TIN, superfícies polédricas, curvas, topologia, bibliotecas de sistemas de coordenadas, dados temporais (*Temporal Features*), endereços postais e a criação de metadados. Mas na versão atual, esses dados dependem da criação de um Esquema de aplicação. A questão da não disponibilidade, até o momento, de uma API padronizada completa para lidar com dados em GML também é relevante.

A proposta do formato GeoBR apresenta um conjunto de inovações com relação aos formatos atualmente existentes, a saber:

- Modelo de Dados que procura fornecer equivalência para dados existentes nos diferentes SIGs de mercado.
- Intercâmbio Semântico de dados de forma padronizada pelo uso de DAML+OIL. A existência de um arquivo DAML contendo a descrição de termos enriquece o processo de transferência dos dados, agregando valor semântico aos dados transferidos.
- Uso de XML como padrão de transferência tornando o dado mais acessível.

Considerando a substancial complexidade do problema de interoperabilidade de dados geográficos, procurou-se elaborar uma proposta que incluísse todas as informações relevantes, com um mínimo de redundância. Espera-se com este trabalho oferecer uma forma simples de realizar o intercâmbio de informações espaciais nos níveis sintático e semântico.

Apesar de esperar que o GeoBR possa usar elementos da GML, algumas características relevantes estão incluídas no GeoBR:

- GeoBR é definido por um Esquema XML único, o que o torna mais unificado e poupa o trabalho de investir na criação de Esquemas de usuário.

- GeoBR prevê uma forma unificada para descrever projeções.
- GeoBR associa as projeções a camadas de dados espaciais (*layers*) evitando o processo de repetir a projeção para todas as geometrias presentes em uma camada, evitando o problema da redundância de dados.
- GeoBR está associado a um arquivo contendo a descrição dos termos utilizados como nomenclatura do seu conteúdo. A existência de dois arquivos, separando dados de seu significado, contribui para que esses possam ser tratados independentemente. O arquivo contendo o significado pode ser estendido com novas informações sem necessidade de alteração dos dados. A cada intercâmbio com usuários de outros domínios, o dicionário deve ficar mais rico, incorporando novos termos e significado.

Alguns aspectos da primeira versão do GeoBR devem ser destacados e considerados limitações ou questões de discussão para uma próxima versão:

- O GeoBR é orientado a camadas de informação. Apesar da tendência atual em trabalhar com feições individuais e independentes, os principais SIGs usam o conceito de camadas de informação ou *layers*. Consideramos esta uma questão aberta à discussão e possível mudança no futuro.
- O GeoBR não inclui uma linguagem de consulta a dados no formato. Seu objetivo principal é o intercâmbio de dados.
- O GeoBR não prevê o intercâmbio de imagens.

Algumas perspectivas relevantes são:

- Revisão da estrutura do formato GeoBR quanto a aspectos técnicos e operacionais do padrão XML, visando promover questões como tamanho do arquivo e acesso aos dados.
- Possibilidade de relacionamento do Esquema GeoBR com a GML 3.0.
- Possibilidade de uso de elementos da linguagem X3D [ISO02], para representar superfícies e dados em 3 dimensões.
- Reformulação do Dicionário de Termos, de acordo com avanços do uso de Ontologias para intercâmbio de dados e novas versões do padrão DAML+OIL.
- Possível descrição de todo o formato pelo padrão DAML.
- Descrição de operadores espaciais. Considera-se a inclusão de operadores básicos definidos formalmente para relações espaciais, como: toca, em, cruza, sobrepõe e disjunto, [Egen92]. Possivelmente identificados pelas *tags* <touch>, <in>, <cross>, <overlap>, <disjoint>, <adjacent_to>, <coincide>, <contain>, <near>. Explicitar este tipo de relacionamento é importante para aplicações que necessitam saber a topologia das entidades geométricas que representam a reali-

dade (ponto, linha e polígono). Por exemplo: uma linha que cruza um polígono pode ser um rio que cruza um município.

- Para versões posteriores do formato relacionamentos como generalização e agregação, conforme [Dav+01], e restrições espaciais entre as classes também devem ser verificadas.

Espera-se que este trabalho venha preencher a atual lacuna de propostas para intercâmbio da geoinformação no Brasil.

KEYWORDS

Geographic Information Systems — Ontology — Interoperability

ABSTRACT

The exchange of spatial data is an important issue in the use of the GIS technology, impelled mainly by the high cost of production of this data type and its need for production of geographical information. It is also a challenge, because it stands out the problem of interoperability among the systems conceived to store, visualize and manipulate spatial data, denominated Geographic Information Systems (GIS). Each GIS software is developed independently and it uses an own conceptual model without the concern in following a common model or common terminology. The result is a heterogeneous environment where each organization has its way to treat the spatial information, linked to the GIS by them adopted. Data distortion and loss of information are problems that frequently appear in a change of data among organizations that use different GIS. This scenery is favorable to the development of models and formats of data that are capable to represent different types of geographical information and to promote the interoperability among different GIS softwares. This article describes the development of a format for syntactic and semantic exchange of spatial data, based on a proposed model and advances in XML technology and Ontology. Describes an open source software for visualization and conversion of files in formats of different GIS softwares of the market and the proposed format.

REFERÊNCIAS BIBLIOGRÁFICAS

- [Auto00]. Autodesk, *AutoCAD2000*. Computer Aided Design Software. San Rafael, CA. 2000.
- [Auto02]. Autodesk, *DXF Reference Guide* [online] <<http://www.autodesk.com/techpubs/autocad/dxf/dxf2002.pdf>> Mar. 2002.
- [Bau+97]. BAUZER, MEDEIROS C., PIRES, F. and OLIVEIRA, J. *An Environment for Modeling and Design of Geographic Applications*. *GeoInformatica* v. 1, n. 1, p. 29-58. Apr. 1997.
- [BCD01]. BCDAM - *Sistema de Bases Compartilhadas de Dados Sobre a Amazônia*. [online] <<http://www.bcdam.gov.br/>>. Mar. 2001.
- [Ber+02]. BERNERS-LEE, T., HANDLER, J., LASSILA, O. *The Future of the Web* [online] <<http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html>>. Mar. 2002.

- [Burk02]. BURKE, M. *The DARPA Agent Markup Language*. [online] <<http://www.daml.org/>>. Mar. 2002.
- [Booch94]. BOOCH G., *Object-Oriented Analysis and Design – with applications*. California, Benjamin/Cummings Publishing Company, Inc., 1994
- [Bra+00]. BRAY T., PAOLI, J., SPERBERG-MACQUEEN, C., MALER, M., Maler, *Extensible Markup Language 1.0 (Second Edition) World Wide Web Recommendation* [online] <<http://www.w3.org/TR/2000/REC-xml-20001006>>. Oct. 2000.
- [Câma95]. CÂMARA, G. *Modelos, Linguagens e Arquiteturas para Bancos de Dados Geográficos*. Ph.D. São José dos Campos, SP, Instituto Nacional de Pesquisas Espaciais (INPE) 1995.
- [Câm+96]. CÂMARA, CASANOVA, M., HEMERLY, A. MAGALHÃES, G. e MEDEIROS. C. *Anatomia de Sistemas de Informação Geográfica*. Curitiba, SAGRES, 1996. 192 p.
- [Câm+99]. CÂMARA, G., THOMÉ, R. G., Thomé, FREITAS, U. and MONTEIRO, A. *Interoperability In Practice: Problems in Semantic Conversion from Current Technology to OpenGIS*. 2nd International Conference on Interoperable Operating Systems, Zurich, 1999.
- [Câm+00b]. CÂMARA, G., SOUZA, R., PEDROSA, B., VINHAS, L., MONTEIRO, A., PAIVA, J., CARVALHO, M. and GATTASS, M.. TerraLib: Technology in Support of GIS Innovation. In: Workshop Brasileiro de Geoinformática (GeoInfo2000), 2., São Paulo, 2000. *Anais*. São José dos Campos: INPE, 2000b. p. 126 - 133.
- [Câm+00c]. CÂMARA, G. M., A.M.V.; PAIVA, J.A.C; SOUZA, R.C.M. *Action-Driven Ontologies of the Geographical Space*. GIScience 2000, Savannah, GA, AAG 2000c.
- [Clar99]. CLARK, J. *XSL Transformations (Version 1.0) World Wide Web Recommendation* [online] <<http://www.w3.org/TR/xslt>>. Nov. 1999.
- [CPRM01]. Companhia de Pesquisa de Recursos Minerais (CPRM). *Serviço Geológico do Brasil*. [online] <<http://www.cprm.gov.br>>. Mar. 2001.
- [Con+02]. CONOLLY, D., HARMELEN, F., HORROCKS, I., MCGUINNES, D., PATEL-SCHNEIDER, P., STEIN, L. *DAML+OIL (March 2001) Reference Description*. W3C Note 18 December 2001. [online] <<http://www.w3.org/TR/daml+oil-reference>>. Mar. 2002.
- [Couc92]. COUCLELIS, H. People Manipulate Objects (but Cultivate Fields): Beyond the Raster-Vector Debate In: *GIS theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, 65—77. 1992.
- [Dav+01]. BORGES K, DAVIS., C. A., LAENDER A. H. F. *OMT-G: An Object-Oriented Data Model for Geographic Applications*. *GeoInformatica*, v. 5, n. 3, p. 221-260, Sep. 2001.
- [Egen92]. EGENHOFER, M. Why not SQL! *International Journal of Geographical Information Systems*, v. 6, n. 2, p. 71-85, 1992.
- [Ege+91]. EGENHOFER, M. and FRANZOSA. R. *Point-Set Topological Spatial Relations*. *International Journal of Geographical Information Systems* v. 5, n. 2, p. 161-174, 1991.
- [ESRI98]. Environmental System Research Institute Inc. (ESRI). *Shapefile Technical Description*. Redland CA, 1998.
- [ESRI01a]. Environmental System Research Institute Inc. (ESRI). *ArcView 8.2 GIS Software*. Redland CA, 2001a.

[ESRI01b]. Environmental System Research Institute Inc. (ESRI). *ArcInfo*. GIS Software. Redland CA, 2001b.

[Fall01]. FALLSIDE, D., *XML Schema Part 0: Primer W3C Recommendation*, 2 May 2001 [online] <<http://www.w3.org/TR/2001/REC-xmlschema-0-20010502>>. Jun. 2001.

[FGDC01]. Federal Geographic Data Committee (FGDC). *Content Standard for Digital Geospatial Metadata WorkBook*. Reston, VA. [online] <<http://www.fgdc.gov/metadata/constan.html>>. Mar. 2001.

[Fon+99]. FONSECA, F., EGENHOFER, M.. *Ontology-Driven Geographic Information Systems* In: MEDEIROS C. B. (Ed.), *7th ACM Symposium on Advances in Geographic Information Systems*, Kansas City, MO, p. 14-19. November 1999.

[Fon+00]. FONSECA, F., EGENHOFER, M. and BORGES. K. Ontologias e Interoperabilidade Semântica entre SIGs. In: Workshop Brasileiro em Geoinformática (GeoInfo2000), 2. ed. São Paulo, 2000. *Anais*. São José dos Campos: INPE, 2000. p. 45 - 52.

[Free01]. FreeGIS Consortium. *FreeGIS Software and Free Geo-Data: Conversion Tools*. [online] <<http://www.freegis.org/>>. Mar. 2001.

[Geo01]. GeoMinas. *Programa Integrado de Uso da Tecnologia de Geoprocessamento pelos Órgãos de Estado de Minas Gerais*. [online] <<http://geominas.mg.gov.br/>>. Mar. 2001.

[Hart98]. HARTMAN, R. *GIS Data Conversion - Strategies, Techniques and Management*. New York, Onword Press. 1998.

[INPE01]. Instituto Nacional de Pesquisas Espaciais (INPE). *SPRING 3.5: Sistema de Processamento de Informações Georreferenciadas*. São José dos Campos, 2001.

{INPE00}. Instituto Nacional de Pesquisas Espaciais (INPE). *Formato de Intercâmbio de Dados Geográficos - ASCII - SPRING* [online] <http://www.dpi.inpe.br/spring/portugues/exemplos/ascii_format.htm>. Mar. 2000.

[ISO02]. ISO/IEC - International Organization for Standardization and International Eletrotechnical Commission. *Information technology — Computer graphics and image processing — eXtensible 3D (X3D)*. [online] <<http://www.web3d.org/TaskGroups/x3d/specification-milestone4/part01/index.html>>. Jul. 2002.

[Las+01]. LASSILA, O., SWICK, R., *Resource Description Framework (RDF) Model and Syntax Specification - World Wide Web Recommendation* [online] <<http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>>. Mar. 2001.

[Map01]. MapInfo Co., *MapInfo Professional User's Guide - Appendix J: MapInfo Interchange Format*. [online] <www.mapinfo.com/community/free/library/interchange_file.pdf>. Mar. 2001.

[Map02]. MapInfo Co., *MapInfo Professional*. GIS Software. Troy, New York, 2002.

[Megg01]. MEGGINSON, D. *SAX - Simple API for XML* [online] <<http://www.saxproject.org/>>. Nov. 2001.

[Meye97]. MEYER, B. *Object-Oriented Software Construction*. New Jersey, Prentice Hall. 1997.

[Mori01]. MORISSETE, D. *Analysis of Arc Export File Format*. *Arc/Info (Rev. 6.1.1)*. [online] <http://pages.infinit.net/danmo/e00/docs/v7_e00_cover.html>. Apr. 2001.