



UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA

Relatórios Técnicos
do Departamento de Informática Aplicada
da UNIRIO
n° 00018/2010

Estudo dos Principais Conceitos sobre Integração de Dados Geoespaciais

Priscila Soares
Asterio Tanaka
Fernanda Baião

Departamento de Informática Aplicada

UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
Av. Pasteur, 458, Urca - CEP 22290-240

RIO DE JANEIRO – BRASIL

Estudo dos Principais Conceitos sobre Integração de Dados Geoespaciais

Priscila Soares, Asterio Tanaka, Fernanda Baião

Departamento de Informática Aplicada (DIA) – Universidade Federal do Estado do Rio de Janeiro (UNIRIO)

{priscila.soares, tanaka, fernanda.baiao}@uniriotec.br

Abstract. The geospatial data integration is currently a very important matter, especially in environments where decision making based on the geographical location is needed. This paper presents a research about the different technologies and approaches that exist to support the geospatial data integration. Among the resources and technologies discussed, there are the use of geospatial metadata, ontologies and XML (eXtensible Markup Language), as well as Geographical Data Warehouse and Spatial OLAP. Because of its importance for the integration and dissemination of geospatial data, is also discussed in this paper: the standards organizations OGC (Open Geospatial Consortium) and W3C (World Wide Web Consortium); the standards specific for geospatial metadata; and also what is a Spatial Data Infrastructure (SDI) and the main characteristics of the SDI of Brazil, which received the name of INDE.

Keywords: Data integration; geospatial data; Spatial Data Infrastructure; ontologies; Geographical Data Warehouse.

Resumo. A integração de dados geoespaciais é hoje um assunto de grande importância, principalmente em ambientes em que a tomada de decisão baseada na localização geográfica se faz necessária. Neste trabalho é apresentada uma pesquisa a respeito das diferentes tecnologias e abordagens existentes para auxiliar a integração de dados geoespaciais. Entre os recursos e tecnologias abordados estão o uso de metadados geoespaciais, ontologias e XML (*eXtensible Markup Language*), além de *data warehouse* geográfico e OLAP espacial. Por ser de grande importância para a integração e disseminação de dados geoespaciais, também são abordados neste trabalho as organizações de padronização OGC (*Open Geospatial Consortium*) e W3C (*World Wide Web Consortium*); os padrões específicos de metadados geoespaciais; e também o que é uma Infraestrutura de Dados Espaciais (IDE) e as principais características da IDE do Brasil, que recebeu o nome de INDE.

Palavras-chave: Integração de dados; dados geoespaciais; Infraestrutura de Dados Espaciais; ontologias; Data Warehouse Geográfico.

Sumário

1	Introdução	4
1.1	Objetivo	5
1.2	Organização do trabalho	5
2	Conceitos básicos	5
2.1	Dados Geoespaciais	5
2.1.1	Dados espaciais, geográficos ou geoespaciais – breve esclarecimento	6
2.2	Integração de Dados	6
2.3	Web Semântica	9
2.4	Sistemas de Informação Geográfica	11
3	Recursos e tecnologias	12
3.1	Metadados	12
3.2	XML	13
3.2.1	GML	14
3.2.2	SVG	15
3.2.3	Exemplo de aplicação – GML e SVG	16
3.3	Ontologias	18
3.3.1	Geo-ontologia	20
3.3.2	Geo-campos e geo-objetos	20
3.3.3	Exemplo de aplicação - ontologias	21
3.4	Data Warehouse	24
3.4.1	On-Line Analytical Processing - OLAP	25
3.4.2	Geographical Data Warehouse – GDW	27
3.4.3	Spatial On-Line Analytical Processing - SOLAP	27
3.4.4	Exemplos de projetos – GDW e SOLAP	28
4	Padronização e disseminação	29
4.1	Principais organizações de padronização	29
4.1.1	OGC	29
4.1.2	W3C	30
4.2	Padrões para metadados geoespaciais	30
4.2.1	Content Standard for Digital Geospatial Metadata (CSDGM)	31
4.2.2	Padrão ISO 19115	31
4.3	Infraestrutura de Dados Espaciais	31
4.3.1	INDE	32
4.3.2	Perfil MGB (Metadados Geoespaciais do Brasil)	33
5	Conclusão	34
6	Referências	35

1 Introdução

A busca por informação geoespacial é algo natural do ser humano. Ao saber da ocorrência de um fato ou fenômeno, é natural ouvir a pergunta: “E onde ele ocorreu?”. A busca por essa noção espacial, saber onde algo (seja espaço físico ou fenômeno) está localizado na Terra, e ainda o que existe no espaço em volta deste lugar, vem do princípio da humanidade. Mas o que começou há muito tempo apenas como uma questão de sobrevivência humana, se tornou hoje uma questão estratégica valiosa para empresas e governo. O fato de possuir uma informação geoespacial (também chamada de geográfica) acurada, pode significar para uma empresa vantagem no mercado, e para um governo um meio de planejar suas ações.

Para chegar até essas informações geoespaciais, hoje tão importantes, é necessário ter acesso e conseguir analisar dados geoespaciais que sejam tão acurados quanto à informação que se deseja. O volume de dados geoespaciais existentes hoje é enorme. Esses dados podem ser obtidos utilizando diferentes meios de coleta (receptores GPS, imagens aéreas,...), em diferentes níveis de precisão. Essa grande possibilidade de oferta de dados geoespaciais deveria ser bom para obter as informações buscadas, mas o problema é que esse enorme volume de dados existente está espalhado por diversas fontes de dados, em diferentes formatos. Por exemplo, se uma prefeitura deseja analisar áreas de risco de deslizamentos de terra, ela terá que analisar dados de volume de chuva, relevo e vegetação de uma dada área. Se esses dados foram obtidos por diferentes instituições, sendo armazenados em diferentes banco de dados, com diferentes esquemas, algo precisará ser feito para que a prefeitura consiga obter a informação geoespacial desejada. É em casos como este, em que o acesso a bases de dados distribuídas se faz necessário, é que a integração de dados geoespaciais irá atuar.

A integração de dados, de uma forma geral, visa dar ao usuário uma visão unificada dos dados, mesmo que estes estejam em diferentes bases, permitindo consultas e análises em cima desses dados. A integração dos dados geoespaciais vai buscar isso, tendo que solucionar também problemas específicos de dados geoespaciais como diferenças entre sistemas de coordenadas de referência e diferenças na precisão planimétrica dos dados. Com uma integração adequada, a informação geoespacial poderá ser obtida de forma mais rápida e acurada.

Para auxiliar a integração e disseminação de dados geoespaciais foram tomadas iniciativas em busca do estabelecimento de padrões, como por exemplo a criação do *Open Geospatial Consortium* (OGC), e iniciativas de implantação de Infraestruturas de Dados Espaciais (IDE), como por exemplo a INDE (Infraestrutura Nacional de Dados Espaciais) do Brasil, a NSDI (*National Spatial Data Infrastructure*) dos Estados Unidos e a INSPIRE (*Infrastructure for Spatial Information in the European Community*) da Europa. Além dessas iniciativas, a internet também ajuda muito a disseminar os dados geoespaciais, por isso os padrões estabelecidos pelo *World Wide Web Consortium* (W3C) se fazem também importantes.

Neste trabalho, o que irá ser buscado é quais são os recursos e tecnologias existentes hoje que auxiliam a integração de dados geoespaciais de forma a facilitar a obtenção de informação geoespacial.

1.1 Objetivo

Este trabalho tem como objetivo geral pesquisar as diferentes tecnologias e abordagens utilizadas atualmente na integração e disseminação de dados geoespaciais. Sendo assim foram pesquisados assuntos relativos à *eXtensible Markup Language* (XML) e Ontologias, tecnologias que ganharam mais expressividade devido à iniciativa da Web semântica, e que são importantes aliadas para a integração de dados.

Outro assunto pesquisado diz respeito ao *data warehouse* geográfico (*Geographical data warehouse* - GDW), que é uma base de dados utilizada no suporte à tomada de decisão que precisa armazenar dados já integrados, para que estes possam ser analisados. As iniciativas de padronização e disseminação (OGC, W3C e IDE) citadas na introdução também fazem parte dos assuntos pesquisados neste trabalho.

Com a pesquisa passando pelos tópicos citados acima, buscou-se exemplos de projetos e implementações relativos a integração de dados geoespaciais e possíveis assuntos a serem explorados numa futura dissertação de mestrado.

1.2 Organização do trabalho

O trabalho está organizado em 6 seções, cujos conteúdos podem ser descritos da seguinte forma: na **seção 1** é feita uma introdução ao trabalho, deixando claro seus objetivos; na **seção 2** são apresentados os conceitos considerados fundamentais para entender o trabalho; na **seção 3** são descritos os recursos e tecnologias envolvidos na integração de dados geoespaciais, bem como exemplos de suas aplicações; na **seção 4** são apresentadas as iniciativas de padronização e disseminação de dados geoespaciais; na **seção 5** é apresentada a conclusão; e por fim, na **seção 6** são apresentadas todas as referências utilizadas na pesquisa.

2 Conceitos básicos

Nesta seção são apresentados os conceitos fundamentais necessários para o entendimento deste trabalho. Tais conceitos são: dados geoespaciais; integração de dados; Web Semântica; e Sistemas de Informação Geográfica.

2.1 Dados Geoespaciais

Os dados geoespaciais são aqueles que representam alguma entidade ou fenômeno que esteja associado a uma localização na Terra, ou seja que possuem coordenadas num sistema relativo à Terra. O fato deles possuírem essa componente geoespacial é que os diferencia de dados convencionais.

Segundo o IBGE (Instituto Brasileiro de Geografia e Estatística), os dados geoespaciais podem ser adquiridos de, por exemplo, tecnologias de levantamento, inclusive as associadas a sistemas globais de posicionamento apoiados por satélites, bem como de mapeamento ou de sensoriamento remoto. Para deixar mais claro, exemplos de fontes seriam: receptores GPS (*Global Positioning System*), mapas digitais e imagens aéreas (ou adquiridas por satélites) georreferenciadas.

2.1.1 Dados espaciais, geográficos ou geoespaciais – breve esclarecimento

Antes de prosseguir com o trabalho, vale a pena dar um breve esclarecimento a respeito dos diferentes nomes do que foi chamado até agora de dados geoespaciais. No Plano de Ação para Implantação da INDE [IBGE, 2010] é esclarecido:

- Dados espaciais – são quaisquer tipos de dados que descrevem fenômenos aos quais esteja associada alguma dimensão espacial.
- Dados geográficos ou geoespaciais – são dados espaciais em que a dimensão espacial refere-se ao seu posicionamento na Terra e no seu espaço próximo, num determinado instante ou período de tempo.

Os dados espaciais podem referir-se então a qualquer espaço, não apenas o espaço da Terra; já os adjetivos geográficos e geoespaciais referem-se ambos à superfície e ao espaço próximo a Terra. Como exemplo de espaços não-geográficos é possível citar o espaço do corpo humano, que pode ser captado por instrumentos que geram imagens para diagnósticos, gerando então dados espaciais. Os dados geoespaciais seriam assim uma especialização dos dados espaciais.

Apesar destas explicações, foi possível notar pelos textos de referência deste trabalho que, na prática, o termo espacial vem sendo utilizado com o mesmo significado apresentado para o termo geoespacial. Neste trabalho, os três termos (espacial, geográfico e geoespacial) são então considerados equivalentes, mas a preferência de uso é do termo geoespacial.

2.2 Integração de Dados

A integração de dados é o processo de combinar os dados residentes em diferentes fontes provendo aos usuários uma visão unificada desses dados [LENZERINI, 2002]. ZIEGLER e DITTRICH [2007] afirmam este ser um dos campos de pesquisa mais antigo na área de banco de dados.

A integração pode ocorrer em diferentes níveis, como por exemplo integração ao oferecer uma interface única aos usuários, mantendo os dados fisicamente separados, com a análise dos dados integrados sendo feita pelo próprio usuário; ou ainda uma integração mais complexa, a integração física dos dados, que precisa levar em consideração os aspectos estruturais e semânticos destes. Para maiores esclarecimentos sobre os níveis de integração é recomendada a leitura do texto [ZIEGLER e DITTRICH, 2007] que consta nas referências.

Para facilitar a integração, o uso de linguagens e representações padrões, que já são bem aceitos pela comunidade de pesquisadores e usuários, é muito importante. Quanto mais padronizados a estrutura e o conteúdo dos dados, melhor será para o uso de ambos, homem e máquina.

Dentre os benefícios que a integração de dados pode trazer estão: facilitação do compartilhamento e acesso aos dados, já que para o usuário as bases vão estar unificadas (ainda que as bases não estejam fisicamente unificadas); e diminuição da duplicidade de ações e desperdícios de recursos na produção de dados, já que os dados vão estar sendo compartilhados. A integração vai ainda auxiliar na tomada de decisão rápida e acurada, como apresentado na introdução, o que é de extrema importância a nível estratégico para empresas e governo.

Sobre as dificuldades encontradas para realizar a integração, pode ser dito que são devidas a vários níveis de heterogeneidades dos dados. Segundo SHETH [1999], é possível classificar estas heterogeneidades em:

- de sistemas – causada pelo uso de diferentes sistemas, com arquiteturas e modelos de dados próprios;
- estrutural – causada pelo uso de diferentes esquemas de dados nas diferentes fontes de informação;
- sintática – ocorre quando existem diferenças no formato de representação de um dado;
- semântica – é causada pela diferença no significado dos dados, principalmente em dados que parecem semelhantes.

A partir dessa classificação de heterogeneidades, é possível classificar também a integração em de sistemas, estrutural, sintática ou semântica, onde cada uma destas vai buscar resolver o problema causado pelas heterogeneidades correspondentes. No domínio geográfico, a integração de dados nestas diferentes classificações vem sendo buscada a partir da conversão de formatos entre os diferentes Sistemas de Informação Geográfica, a partir do mapeamento semântico entre bancos de dados geográficos distintos, e a partir do uso de modelos gerais de dados geoespaciais propostos por reconhecidas organizações, como por exemplo o OGC (*Open Geospatial Consortium*) e a ISO (*International Organization for Standardization*) que possui um comitê técnico para tratar das informações geográficas.

Inicialmente, as soluções de integração buscavam solucionar os problemas estruturais e de diferenças entre sistemas e hardware. Hoje o foco está também voltado para semântica, que deve estar explícita nos dados, para que a integração ocorra corretamente. Sendo assim, a integração semântica de dados vem sendo cada vez mais pesquisada.

A integração semântica de dados pode ser definida como sendo a tarefa de agrupar, combinar e completar os dados de diferentes fontes levando em consideração a semântica dos dados para evitar que dados semanticamente incompatíveis sejam unificados [ZIEGLER e DITTRICH, 2007]. Segundo GOH [1997] as três causas principais da heterogeneidade semântica são:

- conflitos de confusão – ocorrem quando dados parecem ter o mesmo significado, mas são na verdade diferentes;
- conflitos de escala – ocorrem quando diferentes sistemas de referências são utilizados para medir um mesmo valor;
- conflitos de nome – ocorrem quando há valores de elementos que são homônimos ou sinônimos.

Para auxiliar a resolver os conflitos citados podem ser utilizados metadados e ontologias. As ontologias têm um poder maior de explicitar o significado dos dados dentro de um domínio, auxiliando bastante na integração semântica, enquanto que os metadados são semanticamente limitados.

No domínio geográfico, a integração semântica se faz muito importante já que a realidade geográfica facilmente terá uma interpretação diferente por pessoas de diferentes culturas. O “facilmente” se deve ao fato da natureza ser bastante complexa o que leva a diferentes percepções, como demonstrado na Figura 2.1.

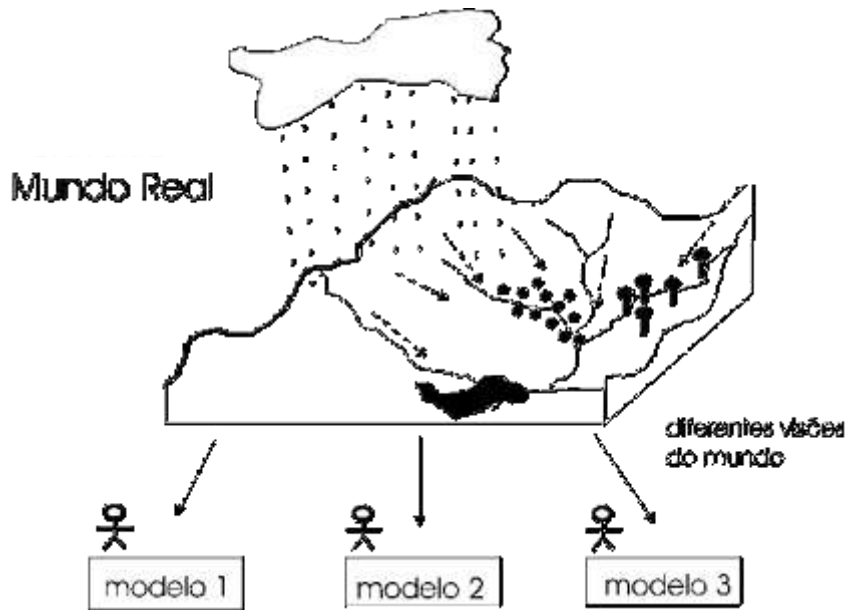


Figura 2.1 – Diferentes modelos de dados advindos de diferentes percepções do mundo real [Adaptado de BARBOSA, 1997].

Um problema para a integração específica de dados geoespaciais que pode ser citado é que existem diferentes sistemas de coordenadas de referência que são utilizados. Como a Terra possui uma superfície muito irregular, são utilizados diferentes superfícies matemáticas (elipsóides) para representá-la e diferentes projeções para que esta representação possa ser visualizada numa folha de papel ou na tela do computador. Apesar de existirem sistemas globais, é muito comum que cada país adote um sistema que possua uma projeção que melhor o represente. A Figura 2.2 abaixo ilustra esse problema, destacando que as diferenças entre os sistemas podem ser resolvidas matematicamente, mas existem ainda outros tipos de diferenças que os dados geoespaciais podem apresentar.

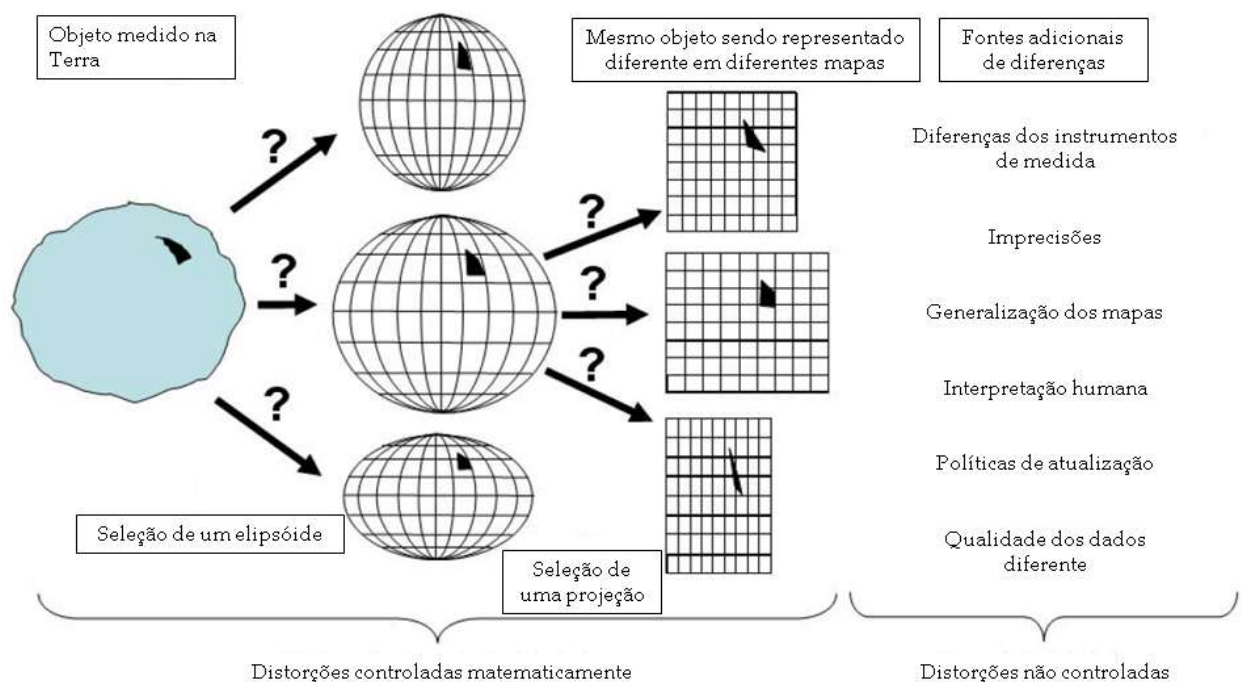


Figura 2.2 – Diferentes sistemas de referência para representação de um objeto medido na Terra [Adaptado de BÉDARD et al., 2007].

As diferenças entre os dados geoespaciais apresentadas na Figura 2.2 que não são totalmente controladas por algoritmos são: as diferenças entre os diversos instrumentos de medida; as imprecisões inerentes dos métodos e ferramentas utilizadas; a generalização dos dados para se adequar às diferentes escalas de representação; as limitações da interpretação humana do fenômeno ou entidade medida; diferentes políticas de atualização dos dados; e qualidade dos dados diferente. Em [BÉDARD et al., 2007] são citados exemplos da atuação das diferenças apresentadas na Figura 2.2 quando a integração de dados geoespaciais se faz necessária. Tais exemplos são: quando se deseja atualizar uma coleção de dados existente, sendo que os dados existentes vieram de fotografias aéreas e as atualizações vieram de levantamento de campo; quando se deseja integrar dados provenientes de mapas de dois municípios adjacentes que podem usar sistemas de referência diferentes; e quando se deseja integrar dados de uma estrada provenientes de um receptor GPS de navegação e de uma imagem aérea.

2.3 Web Semântica

A Web Semântica é uma iniciativa que visa conseguir representar o conteúdo da web de uma forma mais facilmente processável pelas máquinas e usar técnicas inteligentes para aproveitar estas representações [ANTONIOU e VON HARMELEN, 2004]. Com isso, um computador seria capaz de compreender sozinho o significado (semântica) de documentos e dados, se tornando capaz de responder ao usuário com informações mais completas, ao invés de simplesmente apresentar os dados como encontrados. Esta iniciativa foi impulsionada por Tim Berners-Lee e é hoje amplamente disseminada pelo W3C (*World Wide Web Consortium*).

Segundo BERNERS-LEE et al. [2001] a Web Semântica não deve ser vista como uma Web paralela à atual, mas sim como uma extensão desta, onde a informação é fornecida com significado bem definido, permitindo que homem e máquina trabalhem em cooperação. Um informação com significado bem definido propicia o compartilhamento de conhecimento.

A importância da Web Semântica para a integração de dados vem de sua característica de buscar representar os dados de maneira que facilite o processamento e compreensão destes, além de tentar ligar os dados provendo melhor semântica. O objetivo dessa busca é atingido utilizando os recursos da Web Semântica juntamente com os padrões estabelecidos pelo W3C. Entre os recursos da Web Semântica podem ser citados: os metadados explícitos (alcançados com o uso de XML e outras linguagens baseadas nesta), ontologias, lógica e agentes inteligentes. Estes recursos são aplicados para auxiliar a integração de dados tanto convencionais como espaciais.

Em [ANTONIOU e VAN HARMELEN, 2004] é apresentada uma arquitetura em camadas para a Web Semântica, proposta originalmente por Tim Berners-Lee e discutida em trabalhos como [HORROCKS et al., 2005], [GERBER, 2006] e [GERBER et al., 2008]. Essa arquitetura proposta inicialmente representa a Web Semântica em camadas de forma que um agente que atua numa camada deve ter total compatibilidade com as camadas inferiores de desenvolvimento a qual ele se encontra, e ter um entendimento parcial das camadas superiores. Essa arquitetura por pequenos passos facilita a aderência da comunidade de desenvolvedores e usuários fazendo com que o objetivo de atingir totalmente a Web Semântica seja alcançado pouco a pouco. Na Figura 2.3 é apresentada esta arquitetura.

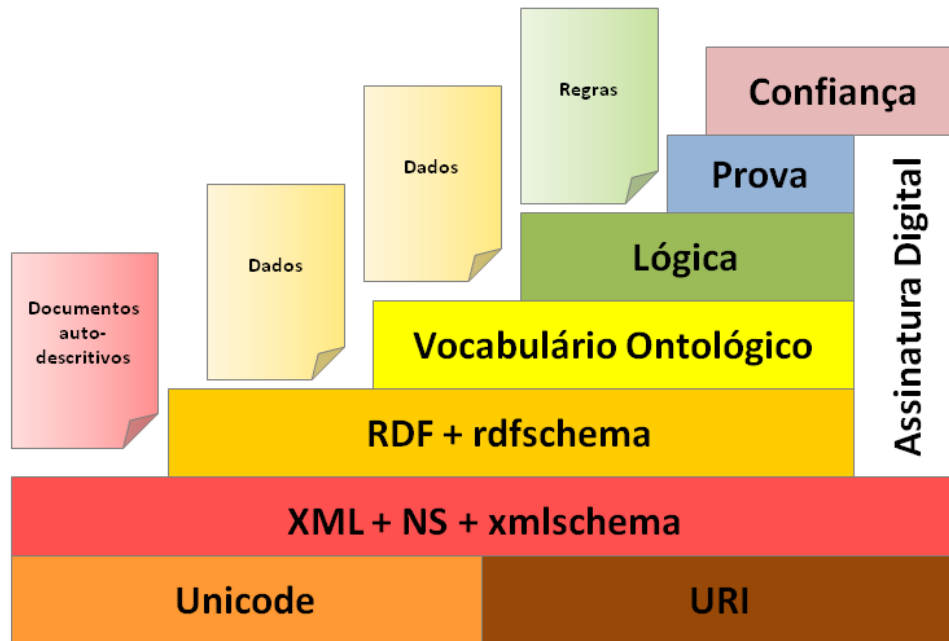


Figura 2.3 – Arquitetura em camadas da Web Semântica. [Adaptado de ANTONIOU e VAN HARMELEN, 2004].

A descrição de cada camada apresentada na figura acima é a que segue (adaptado do site do W3C <www.w3.org>, de [KOIVUNEN e MILLER, 2001] e de [ANTONIOU e VAN HARMELEN, 2004]):

- Unicode e URI – Unicode é um padrão numérico universal, único para cada caractere e independente de plataforma, programa ou linguagem. E URI (Uniform Resource Identifier) é um padrão para identificar um recurso de forma única (o URL - Uniform Resource Locator - é um caso específico de URI). Esta camada irá tratar então da interoperabilidade relativas à codificação de caracteres e identificação de recursos na Web Semântica.

- XML + NS + xmlschema – XML (eXtensible Markup Language) é uma linguagem padrão de representação de dados, independente de plataforma, e que utiliza tags para representar seus elementos, o que é feito segundo uma estrutura hierárquica. Ela será melhor detalhada no tópico 3.2 deste trabalho. Já NS, ou namespaces, permitem utilizar elementos previamente definidos em outras fontes, fazendo referências através de um URI. E o xmlschema é uma linguagem baseada em XML que possui elementos para descrever o esquema de um documento XML.

- RDF + rdfschema – RDF (Resource Description Framework) é um modelo de dados baseado na XML capaz de definir declarações sobre os recursos da Web. E o rdfschema é baseado em RDF, sendo capaz de declarar classes básicas e tipos para descrever termos utilizados no RDF, e que consegue definir vocabulários novos que podem ser referenciados com o uso de um URI. O rdfschema pode ser considerado como uma linguagem primitiva para escrever uma ontologia.

- Vocabulário Ontológico – esta camada trata da definição de termos e dos relacionamentos entre eles fornecendo um vocabulário compartilhado. As ontologias serão tratadas com maior detalhe no tópico 3.3.

- Lógica – a camada de lógica é utilizada para expandir a linguagem da ontologia e permitir a escrita de conhecimento declarativo específico para a aplicação (regras).

Exemplos de linguagens propostas para a descrição dessas regras são a RuleML (Rule Markup Language) e a SWRL (Semantic Web Rule Language).

- Prova - a camada de prova executa as regras e validam as provas junto com a camada superior Confiança.
- Confiança - Validam as provas. A confiança virá do uso de assinaturas digitais e outros tipos de conhecimento baseados em recomendações por agentes confiáveis ou por agências de certificação e classificação e outros consumidores.

2.4 Sistemas de Informação Geográfica

Um sistema de informação geográfica, também conhecido como SIG ou no inglês GIS (*Geographic Information System*), é um sistema capaz de tratar dados geospaciais, permitindo armazenar, manipular, recuperar, analisar e visualizar estes dados. Segundo CÂMARA et al. [2005], a principal diferença de um SIG para um sistema de informação convencional é sua capacidade de armazenar tanto os atributos descritivos como as geometrias dos diferentes tipos de dados geográficos.

A arquitetura básica de um SIG é apresentada em [CASANOVA et al., 1996]. A Figura 2.4 abaixo representa tal arquitetura.

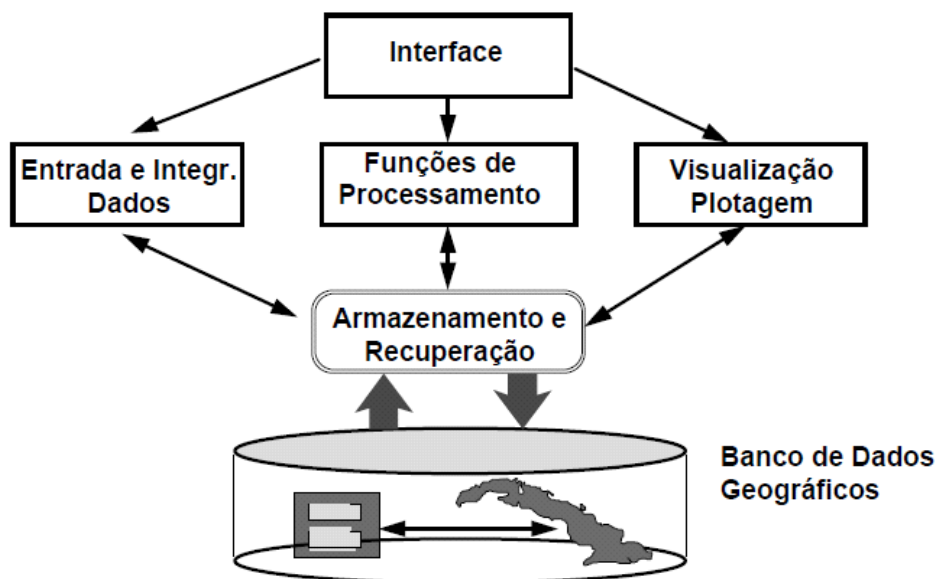


Figura 2.4 – Arquitetura básica de um SIG [CASANOVA et al., 1996].

Na arquitetura apresentada, o nível mais interno é o banco de dados geográficos (ou geoespaciais) e seu sistema gerenciador que irá fornecer suporte para o armazenamento e recuperação dos dados. Já o nível mais externo é a interface com o usuário, que irá prover a forma como o sistema será manipulado. No nível intermediário desta arquitetura estão as funções de entrada, integração, processamento e visualização dos dados geoespaciais. Neste último nível é que o SIG poderá realizar as operações que são específicas de dados espaciais como por exemplo verificar se uma geometria intercepta outra, calcular comprimento ou área, fazer união de geometrias, verificar a posição relativa (ao norte de, ao sul de...), dentre várias outras operações.

Normalmente, para visualizar diferentes tipos de informações geográficas num SIG, é utilizada a sobreposição de camadas (*layers*). Cada camada contém um tema diferente, por exemplo, camada de hidrografia, curvas de nível, ruas, edificações, entre inúmeros outros temas. Como apresentado em [BARROS, 2007], ao sobrepor as camadas dos temas, é possível analisar os dados geográficos de maneira mais simples, já que isso possibilita a combinação de dois ou mais temas para o processamento de uma determinada consulta.

3 Recursos e tecnologias

Nesta seção são apresentados os conceitos dos recursos e tecnologias que podem auxiliar à integração de dados, bem como exemplos de aplicação destes na integração de dados geoespaciais. Tais recursos e tecnologias são: Metadados, XML, ontologias e data warehouse. Esta é a principal seção para responder ao objetivo geral deste trabalho.

3.1 Metadados

Metadados são um conjunto de dados estruturados que servem para descrever características de outros dados sendo capazes de auxiliar na integração sintática de dados e, de forma limitada, na integração semântica. Eles podem descrever o conteúdo, localização, formato, finalidade, qualidade e outras características de um dado, desempenhando um importante papel na identificação, organização, compreensão e recuperação da informação.

Os metadados para dados geoespaciais vão, assim como para dados convencionais, atuar como descritores. Exemplos de metadados específicos para dados geoespaciais são os valores de: projeção, escala, resolução e sistema de coordenadas. Segundo CARNEVALE [2008], os metadados devem oferecer uma base de descritores capaz de documentar qualquer dado geoespacial de modo a permitir seu reconhecimento, sua caracterização e sua eventual recuperação, através de respostas diretas às seguintes questões no mínimo:

- O que? – Título e descrição do conjunto de dados.
- Por que? – Resumo detalhando as razões para a aquisição dos dados e seu uso.
- Quem? – Quem produz, documenta e distribui este dado.
- Quando? – Quando o conjunto de dados foi coletado e qual a frequência de atualização.
- Onde? – Área de abrangência geográfica do dado baseada em latitude e longitude.
- Como? – Como o dado foi construído e como acessá-lo.

Para atuar na integração semântica, os metadados vão depender sobre que elementos eles estão apresentando para descrever o conteúdo do dado, sua finalidade e o domínio. Um metadado rico em informações pode auxiliar a fazer consultas onde seja possível determinar a relevância dos dados para um determinado fim, sem acessar diretamente o dado. Isso pode acarretar numa melhora de performance da consulta.

SHETH [1999] e KASHYAP [1995] apresentam uma classificação de metadados baseada em quanto esses metadados conseguem capturar o conteúdo dos dados. A classificação é a seguinte:

- Metadados independente de conteúdo (content-independent metadata): esse tipo de metadados captura as informações que não dependem do conteúdo dos dados aos quais é associado. Exemplo: última data de modificação de um documento.
- Metadados dependentes do conteúdo (content-dependent metadata): depende do conteúdo dos dados. Exemplo: número de linhas e colunas de uma imagem.
- Metadados que descrevem o conteúdo (content-descriptive metadata): é um caso especial do content-dependent metadata. Ele descreve o conteúdo dos dados, fazendo com que seja possível entender o conteúdo sem precisar acessar os dados diretamente.

Segundo essa classificação, para a integração semântica os metadados mais importantes são os *content-descriptive metadata*, já que este tipo de metadado é que vai conseguir fornecer alguma informação a respeito do significado do dado descrito.

Para facilitar a integração, é importante que os metadados sejam representados de forma independente de plataforma, com uma linguagem compartilhada. Para isso pode ser utilizada a XML (*eXtensible Markup Language*), uma linguagem padrão estabelecida pelo W3C (*World Wide Web Consortium*). Neste trabalho esta linguagem é discutida melhor no tópico 3.2.

Outro ponto importante para facilitar a integração é o uso de padrões para estabelecer os esquemas dos metadados (ou perfis de metadados). Como padrões próprios para dados geoespaciais podem ser citados o padrão ISO 19115, especificado pelo Comitê Técnico 211 da ISO (*International Organization for Standardization*); e o *Content Standard for Digital Geospatial Metadata* (CSDGM), padrão dos Estados Unidos proposto pelo *Federal Geographic Data Committee* (FGDC). Como padrão de metadados para fins gerais pode ser citado o Dublin Core. O padrão ISO 19115 é hoje o padrão para metadados que está sendo mais utilizado, para onde todos os outros estão convergindo, tornando este o padrão de fato para metadados (ele é o padrão aceito pelo OGC – *Open Geospatial Consortium*).

Mais sobre padrões de metadados geoespaciais e também sobre o uso de metadados em uma infraestrutura de dados espaciais é discutido na seção 4 deste trabalho.

3.2 XML

A XML (*eXtensible Markup Language*) é uma linguagem padrão estabelecida pelo W3C (*World Wide Web Consortium*) que é independente de plataforma e apresenta estrutura hierárquica. Essa linguagem é derivada da SGML (*Standard Generalized Markup Language*). Um documento de metadados escrito em XML poderá ser lido e compreendido por diferentes sistemas, e devido a sua estrutura, deixará os metadados escritos de forma explícita, daí sua importância para a integração.

Os elementos num documento XML são descritos através de tags. O elemento mais externo, chamado de elemento raiz ou *root*, é o elemento abaixo do qual todos os outros elementos estarão aninhados. Um exemplo simples de um documento escrito em XML é o que segue:

```
<?xml version="1.0"?>
<livro> <! -- Esse é o elemento raiz-->
  <título> A Semantic Web Primer </título>
  <autor> Antoniou, G.; Van Harmelen, F. </autor>
  <ano> 2004</ano>
</livro>
```

A Figura 3.1 apresenta um gráfico deixando ainda mais claro a estrutura de hierarquia do documento XML.

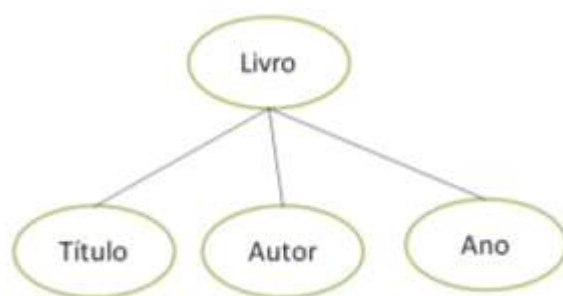


Figura 3.1 – Representação do documento XML dado como exemplo.

ANTONIOU e VAN HARMELEN [2004] estabelecem que para um documento XML ser considerado bem formado é necessário que ele esteja sintaticamente correto, sendo que algumas regras sintáticas são:

- só deve existir um elemento raiz no documento (elemento root);
- cada elemento contém uma tag de abertura e de encerramento;
- as tags não devem se sobrepor, como em `<autor><nome>Lee Hong</author></nome>`, e sim serem fechadas da mais interna para a mais externa, como em `<autor><nome>Lee Hong</nome></author>`;
- os atributos dentro de um elemento devem ter um nome único;
- os elementos e nomes das tags devem ser permissíveis.

A XML serviu como base para várias outras linguagens como por exemplo: RDF, OWL, GML e SVG. Todas essas linguagens citadas podem ser utilizadas no auxílio à integração de dados geoespaciais. As duas últimas (GML e SVG) são apresentadas nos próximos tópicos porque são específicas para trabalhar com dados que tenham geometria, sendo que a GML é própria para manipular dados geográficos e a SVG é capaz de representar geometrias (associadas a dados geográficos ou não).

3.2.1 GML

A *Geography Markup Language* (GML) é uma linguagem estabelecida como padrão de compartilhamento de informações geográficas estabelecido pelo OGC (*Open Geospatial Consortium*). Por ser uma linguagem baseada na XML ela também tem estrutura hierárquica e utiliza tags para identificar seus elementos.

O elemento principal (raiz) de um documento GML é a *feature*, que representa os objetos e fenômenos geográficos existentes no mundo real. Sendo assim, o mundo real poderia então ser representado computacionalmente como um conjunto de *features*. Uma *feature* é definida com um conjunto de propriedades, onde cada propriedade pode ser vista como o trio {nome, tipo, valor} [OGC, 2007].

A GML possui 5 tipos de geometria, todas derivadas de um tipo abstrato de geometria. A geometria “ponto” contém um par de coordenadas, enquanto que a “linha” contém mais pares que são conectados. Uma “caixa” é definida pelos cantos inferior esquerdo e superior direito. “Anel” é uma linha fechada de pelo menos três pontos, onde o ponto inicial e final são o mesmo (então ele tem na verdade quatro pontos). Um “polígono” contém um limite exterior e pode conter buracos, que são um limite interior que marca áreas que são cortadas da área do limite exterior (Figura 3.2).

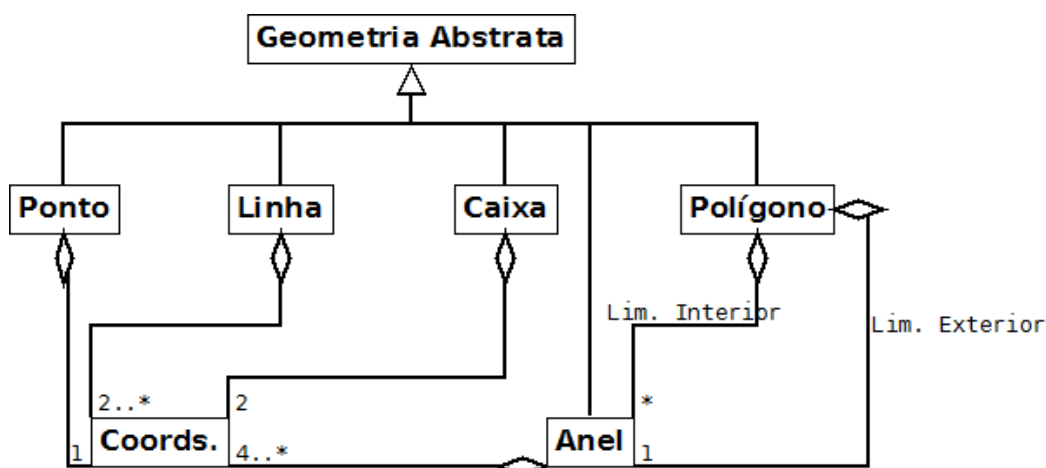


Figura 3.2 – Esquema (em diagrama de classes da UML) das geometrias da GML [Adaptado de MATHIAK et al., 2004].

3.2.2 SVG

SVG (*Scalable Vectorial Graphics*) é uma linguagem, também baseada em XML, que serve para descrever gráficos bidimensionais. Ela é recomendada pela W3C, e possui uma versão mais leve chamada de SVG Tiny que é voltada para dispositivos móveis.

Essa linguagem suporta três tipos de objetos gráficos: vetores, imagem e texto [W3C, 2003]. Uma das vantagens da SVG, é que os gráficos vetoriais gerados por ela podem ser renderizados em diversos tamanhos sem perder a qualidade. Por ser bem mais leve, os dados nessa linguagem são mais rapidamente carregados, propiciando ganho em performance e ampliando as possibilidades de integração de dados que precisam de visualização geométrica.

Abaixo um exemplo de documento SVG extraído da especificação da SVG versão 1.1. Este exemplo irá desenhar 4 retângulos dentro de um retângulo maior. Este último sem preenchimento e de borda azul.

```

<?xml version="1.0" standalone="no"?>
<!DOCTYPE svg PUBLIC "-//W3C//DTD SVG 1.1//EN"
"http://www.w3.org/Graphics/SVG/1.1/DTD/svg11.dtd">
<svg width="5cm" height="4cm" version="1.1"
xmlns="http://www.w3.org/2000/svg">
<desc> Four separate rectangles </desc>
<rect x="0.5cm" y="0.5cm" width="2cm" height="1cm"/>
<rect x="0.5cm" y="2cm" width="1cm" height="1.5cm"/>
<rect x="3cm" y="0.5cm" width="1.5cm" height="2cm"/>
<rect x="3.5cm" y="3cm" width="1cm" height="0.5cm"/>
<!-- Show outline of canvas using 'rect' element -->
<rect x=".01cm" y=".01cm" width="4.98cm" height="3.98cm"
fill="none" stroke="blue" stroke-width=".02cm" />
</svg>
  
```

Ao ler esse documento num visualizador web com suporte a SVG o resultado é o mostrado na Figura 3.3.



Figura 3.3 – Gráfico gerado pelo documento SVG utilizado como exemplo.

3.2.3 Exemplo de aplicação – GML e SVG

Um exemplo de aplicação da linguagem GML e do SVG para a integração de dados geoespaciais é dado em [MATHIAK et al., 2004] e será apresentado neste tópico. No artigo citado o objetivo é conseguir transformar dados legados do escritório alemão de pesquisa geográfica, em dados com maior capacidade de compartilhamento via internet. Para isso os dados foram transformados para o padrão aberto de troca de dados geográficos GML.

Os dados legados que foram transformados foram gerados num projeto iniciado a mais de 15 anos. Esses dados não apresentavam metadados explícitos, e eram associados a um catálogo de códigos que representavam os tipos de objetos que aquele dado queria representar. Por exemplo, um dado que representasse um objeto rua, não continha o tipo rua, mas sim o código 3101. Além disso, as linhas ou polígonos no modelo de dados original eram representados por pequenos segmentos de linha, sem uma conexão clara entre eles.

Os tipos de elementos disponíveis no catálogo de códigos tiveram então que ser mapeados para elementos correspondentes da GML. Após isso foi feita a transformação dos dados com o auxílio de uma aplicação em Java feita para localizar as informações necessárias dos dados legados e escrever num documento GML. Abaixo o exemplo de um objeto rua mapeado dos arquivos originais (Tabela 3.1) para um arquivo GML.

Tabela 3.1 – Extração dos arquivos originais para o objeto "Bachstr.", que é uma rua [MATHIAK et al., 2004].

Nome do arquivo	Linhas que contém o objeto 86118065
objecttype.txt	86118065, 3101 (significa rua)
name.txt	86118065, Bachstr.
coordinates.txt	86118065, 4437960.070, 5331818.450, 4437967.200, 5331825.410 86118065, 4437952.980, 5331812.550, 4437960.070, 5331818.450

E a extração do documento GML:


```

<Street>
  <gml: name> Bachstr. </gml: name>
  <objectID> 86118065 </objectID>
  <gml: centerLineOf><gml: LineString>
    <gml: coord>
      <gml: X> 4437952.980 </gml: X>
      <gml: Y> 5331812.550 </gml: Y>
    </gml: coord>
    <gml: coord>
      <gml: X> 4437960.070 </gml: X>
      <gml: Y> 5331818.450 </gml: Y>
    </gml: coord>
    <gml: coord>
      <gml: X> 4437967.200 </gml: X>
      <gml: Y> 5331825.410 </gml: Y>
    </gml: coord>
  </gml: LineString></gml: centerLineOf>
  <attributes>
    <status> open for traffic </status>
    <numberOfRoadways denotation="actual number"> 2
  </numberOfRoadways>
    <function> road traffic </function>
    <dedication> communal </dedication>
  </attributes>
</Street>

```

É possível perceber que as coordenadas no documento GML segue uma ordem específica formando a linha, não sendo apenas segmentos de linhas soltos, como no arquivo original. Para melhor visualização desta explicação é apresentada a Figura 3.4, que representa polígonos sendo desenhados como obtido nos dados originais e como obtido nos dados em GML.

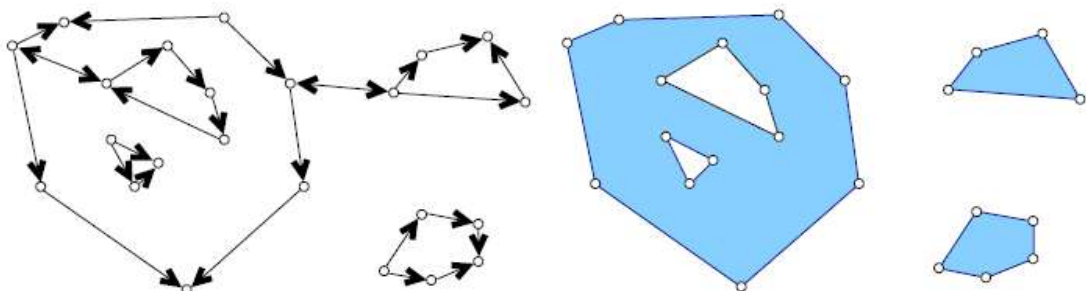


Figura 3.4 – À esquerda, polígonos representados por segmentos de reta (dados originais) e à direita, polígonos derivados da GML [MATHIAK et al., 2004].

Com os dados já em GML, o compartilhamento deles ficou mais fácil. O artigo mostra ainda como visualizar estes dados GML em gráfico de vetores, o que facilitaria a visualização deles pela internet, já que carregar os gráficos de vetores é mais leve que carregar os mapas digitalizados originais. Para isso foi utilizada o XSLT (*eXtensible Style-sheet Language for Transformation*) para mapear os elementos GML para gráficos SVG. O resultado final foram gráficos vetoriais muito parecidos com os mapas originais, já que foram utilizadas nos vetores cores e padrões de preenchimento parecidos com os originais (Figura 3.5).

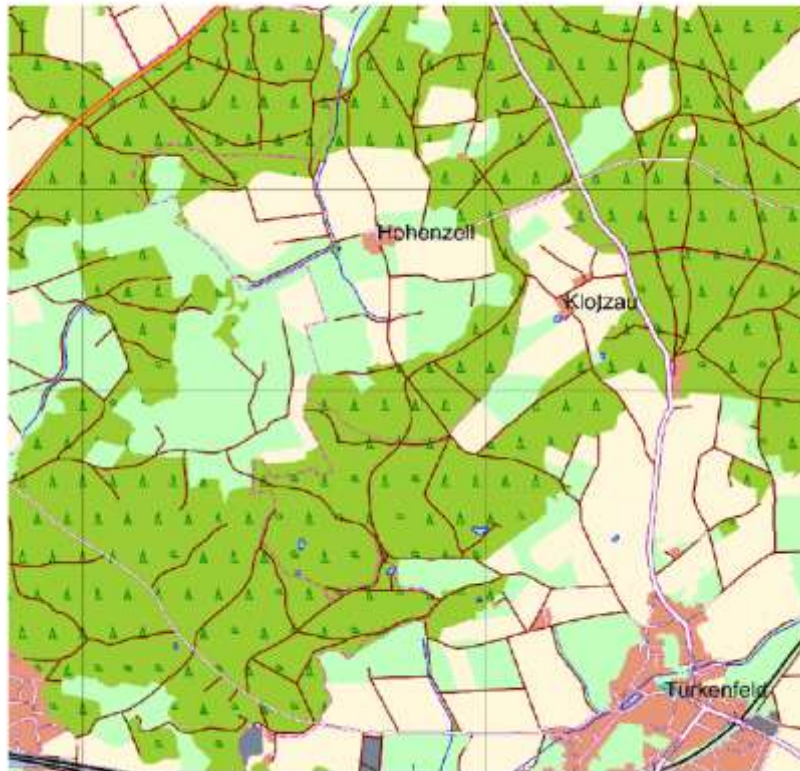


Figura 3.5– Parte de um mapa vetorial gerado [MATHIAK et al., 2004].

Com este exemplo fica claro como as linguagens baseadas em XML podem auxiliar a integração e compartilhamento de dados geoespaciais.

3.3 Ontologias

Ontologia, no contexto da computação, pode ser definido como uma especificação formal e explícita de uma conceituação compartilhada [GRUBER, 1993]. Ou ainda, segundo GUZZARDI [1997], uma ontologia é uma teoria lógica que corresponde ao significado intencional de um vocabulário formal, ou seja, um comprometimento ontológico com uma conceituação particular do mundo.

Na filosofia, a Ontologia (escrita com O maiúsculo) é definida como um ramo da metafísica preocupada com a natureza e as relações entre os seres, estuda a existência do ser [GUZZARDI, 2007]. Essa Ontologia filosófica é única. Já no contexto da computação existem várias ontologias que tentam explicar como um grupo de indivíduos entende um determinado domínio. Nos últimos anos, segundo GUZZARDI [2007], ocorreu uma explosão em trabalhos relacionados a ontologias na ciência da computação.

ção, motivada principalmente pelo crescente interesse na Web Semântica, e no papel chave que as ontologias têm nesta iniciativa.

Na integração de dados, a ontologia será utilizada para a integração semântica. O uso de ontologia nesse tipo de integração se deve ao fato dela ser capaz de gerar um entendimento comum de um domínio e compartilhá-lo no domínio modelado levando em consideração o significado dos termos e seus relacionamentos.

Em [GUARINO, 1998] é proposta uma classificação para as ontologias representada na Figura 3.6 e descritas logo abaixo da figura.

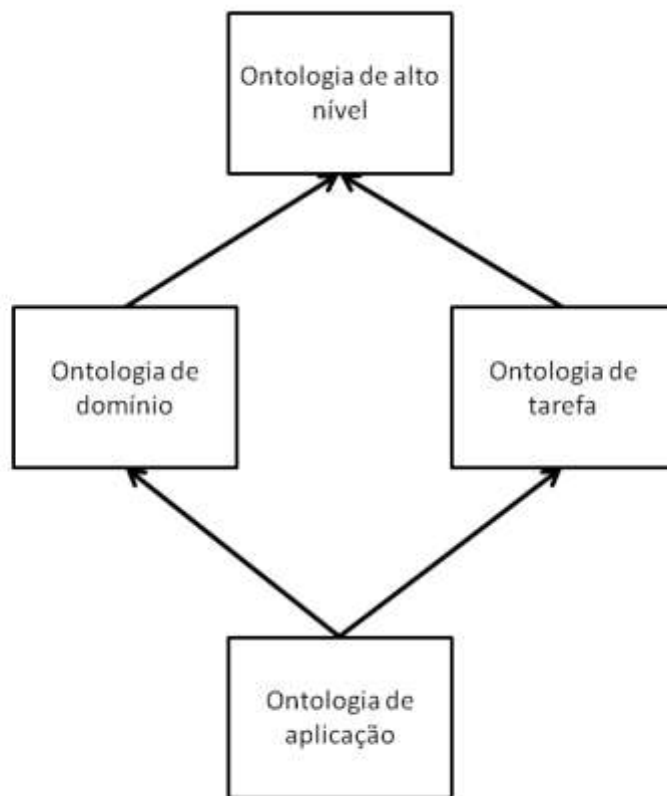


Figura 3.6– Níveis de ontologias [Adaptado de GUARINO, 1998].

- Ontologias de alto nível: descrevem conceitos bem gerais, que são independentes de um domínio particular.
- Ontologias de domínio: descrevem o vocabulário relacionado a um domínio, especializando os termos da ontologia de alto nível.
- Ontologias de tarefas: descrevem uma tarefa ou atividade, especializando os termos da ontologia de alto nível.
- Ontologias de aplicação: descrevem conceitos dependentes tanto do domínio como da tarefa.

Uma ontologia pode ser descrita formalmente de várias formas: através de diagramas de classes da UML por exemplo, ou através de linguagens específicas como a OWL. A OWL (*Web Ontology Language*) é uma linguagem baseada em XML, RDF e RDFschema e foi recomendada pelo W3C. Ela é dividida em 3 tipos diferentes, de acordo com a capacidade de expressividade semântica [W3C]:

- OWL Lite: é a mais simples, com várias restrições. Suporta usuários precisando de uma classificação hierárquica com restrições simples.

- OWL DL: suporta o máximo de expressividade garantindo que tudo será computável em tempo finito.
- OWL Full: é a OWL completa. Suporta usuários que queiram o máximo de expressividade e a liberdade sintática do RDF, mas sem garantias computacionais.

3.3.1 Geo-ontologia

A geo-ontologia é uma ontologia que trata dos conceitos geoespaciais e seus relacionamentos, auxiliando a estabelecer um vocabulário comum para ajudar na integração de dados geoespaciais. Segundo CÂMARA et al. [2005], uma geo-ontologia possui dois tipos básicos de conceitos:

- conceitos físicos, que representam entidades ou fenômenos físicos do mundo real;
- conceitos sociais, que representam entidades ou fenômenos sociais (conceitos estabelecidos pela sociedade).

Esses conceitos podem ainda ser subdivididos de acordo com a identificação de suas fronteiras, ou seja, se esses conceitos representam entidades ou fenômenos com fronteiras bem definidas ou que variam de forma contínua no espaço. Os conceitos físicos que são associados a entidades individualizáveis são chamados de *bona fide* (boa fé em latim). Já os conceitos sociais que são associados a entidades são chamados de *fiat* (fazer em latim). Na tabela 3.2 são apresentados exemplos de cada subdivisão dos conceitos básicos da geo-ontologia descrita por CÂMARA et al. [2005].

Tabela 3.2 – Exemplos de classes de cada tipo de conceitos da geo-ontologia

	Fronteira bem definida	Com variação contínua no espaço
Conceitos físicos	(<i>bona fide</i>) Lagos, ilhas, montanhas, vales	Temperatura, altimetria, declividade, poluição.
Conceitos sociais	(<i>fiat</i>) Lotes, municípios, países	Exclusão social, desenvolvimento humano

Maiores informações sobre os tipos de conceitos até agora apresentados para uma geo-ontologia podem ser encontrados em [SMITH e MARK, 1998].

3.3.2 Geo-campos e geo-objetos

Outra forma de classificar os dados geoespaciais dentro de uma ontologia diz respeito apenas à fronteira e vem da modelagem de dados. Os elementos que possuem fronteiras bem definidas, sendo associados a entidades identificáveis, são classificados como geo-objetos; já os elementos que possuem variação contínua no espaço, são classificados como geo-campos. Na Figura 3.7 é apresentado um exemplo de geo-campo e geo-objeto. O geo-campo dessa figura é representado pela imagem adquirida por um sensor orbital. Essa imagem possui em cada um de seus pontos a associação com uma variável chamada reflectância, que exprime a quantidade de luz refletida pelo solo em cada ponto. A imagem representa então um campo de valores de reflectância. Já o geo-objeto é representado à direita na figura por cada distrito com fronteira bem definida de uma cidade.

Segundo CÂMARA et al. [2005], a diferença essencial entre um geo-campo e um geo-objeto é o papel da fronteira. Num geo-campo, uma pequena partição dele (representada na Figura 3.7 pelo retângulo no centro da imagem orbital) irá conter as mesmas propriedades do geo-campo original. Já uma partição de um geo-objeto não possui as mesmas propriedades que o geo-objeto original.

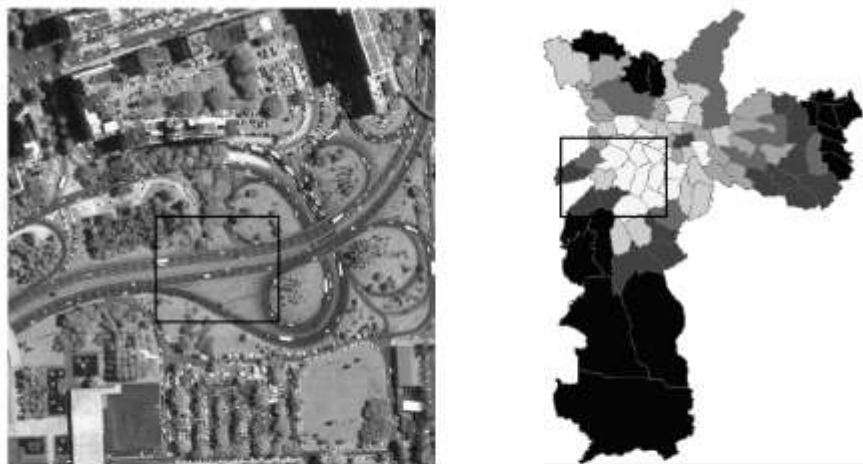


Figura 3.7– Exemplo de geo-campo e conjunto de geo-objetos [CÂMARA et al., 2005].

3.3.3 Exemplo de aplicação - ontologias

Para exemplificar como as ontologias podem ser aplicadas para auxiliar a integração de dados geoespaciais, são apresentados neste tópico os SIGs baseados em ontologias. Além disso, também é apresentada uma curiosidade que é como uma imagem aérea ou adquirida por sensor orbital pode ser classificada dentro de uma geo-ontologia.

A respeito dos Sistemas de Informação Geográfica baseados em ontologias, FONSECA et al. [2002] afirma que a próxima geração de sistemas de informação deve poder resolver a heterogeneidade semântica para fazer uso do grande volume de informações disponíveis com a chegada da internet e computação distribuída. Sendo assim ele propõe o uso de ontologias no desenvolvimento e também no uso de SIGs.

O uso de ontologias no desenvolvimento e uso de sistemas de informações (SI) é apresentado em [GUARINO, 1998]. Nesse artigo é afirmado que o desenvolvimento de um SI baseado em ontologias pode ser feito a partir de uma biblioteca de ontologias preexistente, onde o conteúdo semântico expresso pelas ontologias selecionadas seria transformado num componente do sistema. FONSECA et al. [1999, 2000, 2002] estende essa forma de desenvolvimento para tratar de geo-ontologias, propondo os Sistemas de Informação Geográfica baseados em ontologias (SIG-O) ou em inglês *Ontology-Driven Geographic Information Systems* (ODGIS). Num SIG-O as ontologias seriam mapeadas para componentes de software através de técnicas de orientação a objeto.

Um SIG baseado em ontologias é desenvolvido a partir dos seguintes componentes: um servidor de ontologias, onde ficarão armazenadas as ontologias especificadas formalmente; um editor de ontologias, que vai permitir aos especialistas criar e editar ontologias; um folhador de ontologias, que vai fornecer ao usuário final metadados referentes às informações disponíveis; um tradutor de ontologias, que vai traduzir as ontologias para classes através da orientação a objeto; um servidor de classes, onde ficarão armazenadas as classes derivadas de ontologias; e um folhador de classes,

que vai permitir procurar por classes que servirão como base para aplicações SIG. Um esquema do uso de ontologias no desenvolvimento de um SIG-O é apresentado na Figura 3.8. Este tipo de sistema vai facilitar a integração semântica já que utiliza classes que foram baseadas em ontologias.

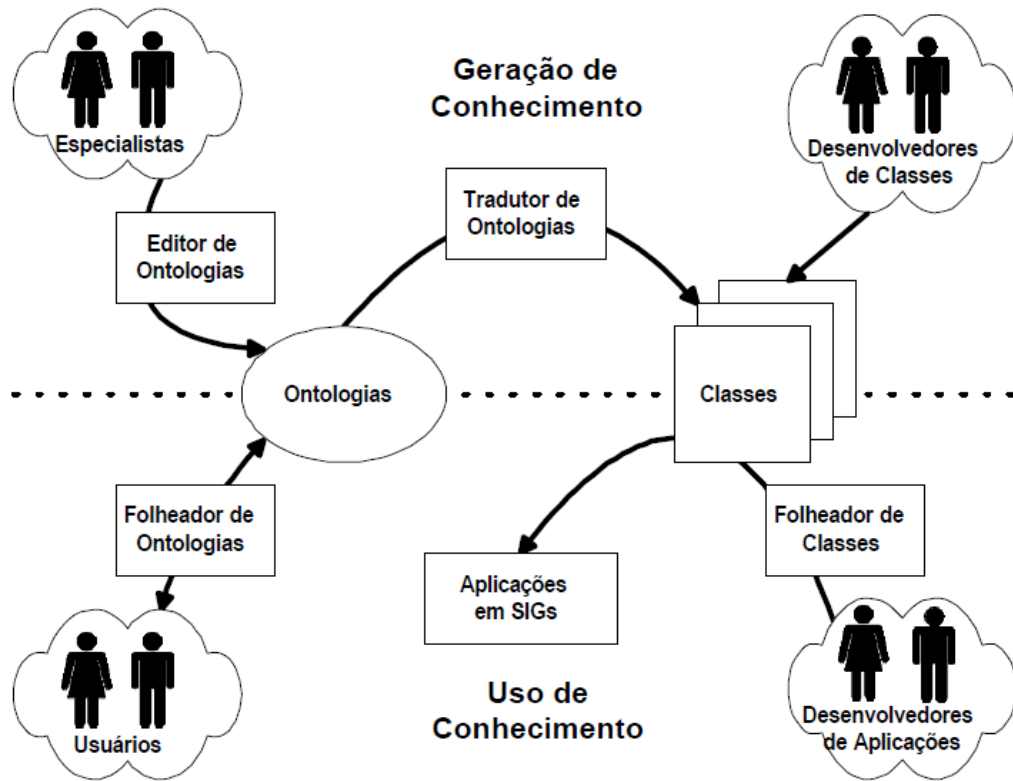


Figura 3.8 – Esquema do desenvolvimento de um SIG-O [FONSECA et al., 2000].

O outro exemplo de uso de ontologias é na verdade uma curiosidade e diz respeito a como uma imagem adquirida por sensor orbital pode ser representada dentro de uma geo-ontologia. Em [CÂMARA et al., 2001] é apresentada uma ontologia própria para imagens já que estas podem ser classificadas tanto em geo-campos quanto em geo-objetos, não se encaixando então em nenhuma classificação única de uma geo-ontologia.

A classificação como geo-campo vem do fato que uma imagem pode ser considerada como um conjunto de valores de luz refletida associado a cada ponto da imagem, ou ainda, como um sub-conjunto de um Modelo Digital de Terreno (MDT). E a classificação como geo-objeto vem do fato que uma imagem pode ser considerada como um conjunto de objetos que podem ser extraídos através de procedimentos de análise. Para maiores explicações da classificação da imagem na perspectiva como conjunto de objetos, é recomendado o texto [BITTNER e WINTER, 1999].

Ambas as classificações da imagem possuem as suas vantagens e nenhuma pode ser excluída. Sendo assim, CÂMARA et al. [2001] propõe uma ontologia para imagens com múltiplos níveis considerando que imagens de sensoriamento remoto são, ontologicamente, instrumentos para capturar a dinâmica da superfície terrestre. Essa ontologia seria capaz de englobar múltiplas perspectivas para uma mesma imagem.

A ontologia de múltiplos níveis das imagens seria composta por três ontologias distintas mas que se relacionam (Figura 3.9):

- Ontologia física: descreve os processos físicos da criação da imagem, ou seja, descreve o conhecimento relacionado a captura pelo sensor da luz refletida pela superfície da Terra.
- Ontologia estrutural: descreve as estruturas geométricas e descritivas que podem ser extraídas ou detectadas na imagem através de técnicas de processamento.
- Ontologia de métodos: consiste de um conjunto de algoritmos e estruturas de dados que representam conhecimento reutilizável em forma de técnicas de processamento de imagem que podem ser utilizadas para transformar a imagem do nível físico para o nível estrutural.

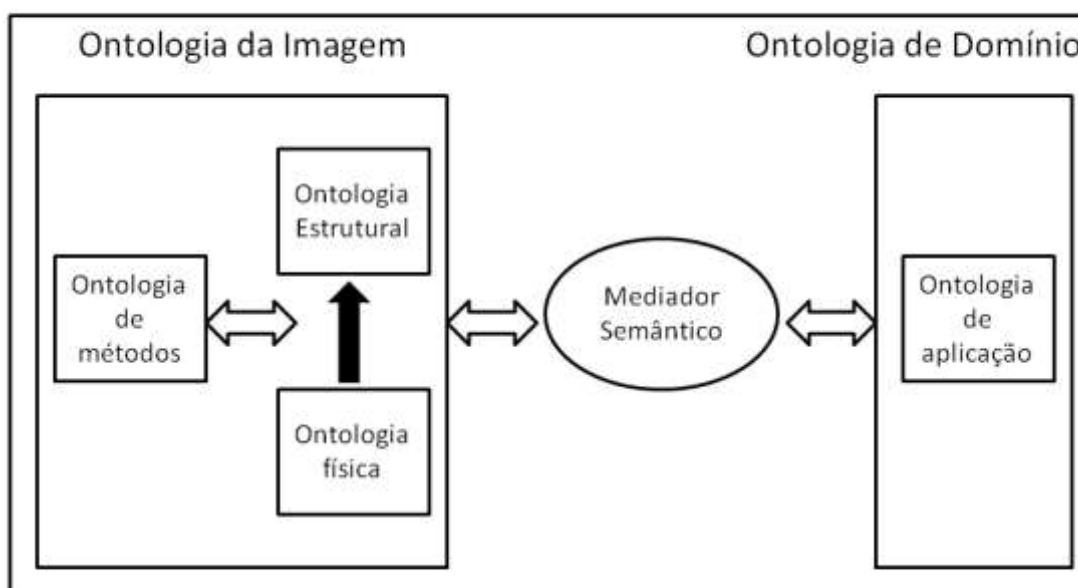


Figura 3.9 – Representação da ontologia de múltiplos níveis da imagem [Adaptado de CÂMARA et al., 2001].

A ontologia proposta para as imagens vai conseguir lidar tanto com geo-campos quanto com geo-objetos, sendo independente de qualquer outra ontologia de domínio que um usuário esteja utilizando. Para fazer o relacionamento entre a ontologia das imagens e a ontologia de domínio utilizada, é proposto um mediador semântico. Esse mediador seria capaz de identificar que algoritmo de processamento seria necessário utilizar para extrair as estruturas desejadas da imagem; e seria capaz também de fazer um mapeamento entre os conceitos da ontologia de domínio e as estruturas extraídas da imagem.

Ainda falando de ontologias com aplicação em dados geoespaciais, vale a pena citar o texto [KOLAS et al., 2005] onde são apresentados cinco tipos de ontologias que poderiam dar suporte à chamada Web Semântica Geoespacial. A primeira seria uma ontologia de base que fornecesse o principal vocabulário que as outras ontologias irão referenciar, sendo utilizada pela base de conhecimento e pelos serviços da Web Semântica; a segunda é uma ontologia de filtros geoespaciais, que também é utilizada pela base de conhecimento e pelos serviços web; a terceira é uma ontologia de domínio, que representa os dados mapeados a partir da ontologia geoespacial de base para a ontologia de um grupo específico de usuários; a quarta é a ontologia de feições, que define dados existentes na especificação da Web Feature Service do OGC; e a quinta é

a ontologia de serviços geoespaciais, que liga a ontologia de base e a de filtros com o OWL-S (OWL para serviços) para criar descrições de serviços de feições da Web Semântica. A Figura 3.10 mostra a arquitetura proposta dessas ontologias no suporte a uma base de conhecimento.



Figura 3.10 – Arquitetura de ontologias geoespaciais no suporte a uma base de conhecimento [Adaptado de KOLAS et al., 2005].

3.4 Data Warehouse

O *data warehouse* (DW) é um exemplo de onde a integração de dados se faz muito importante, já que ele lida com dados adquiridos em diversas fontes diferentes. Segundo INMON [2005], um *data warehouse* pode ser definido como uma coleção de dados orientada ao assunto, integrada, variante no tempo e não volátil, que provê suporte a tomada de decisão gerencial. INMON [2005] afirma ainda que não tem sentido trazer dados do ambiente operacional para o *data warehouse* sem integrá-los. Se os dados chegarem até o DW sem serem integrados, eles se tornam inúteis, pois não vão poder ser utilizados para a tomada de decisão acurada.

Um *data warehouse* irá então armazenar dados adquiridos de diferentes fontes, integrando-os a fim de permitir sua posterior análise. Por ser variante no tempo e não volátil, o *data warehouse* permite fazer análises históricas sobre os dados. Essas duas características são responsáveis também pelo DW tender a ser muito mais volumoso que um banco de dados operacional.

Segundo CHAUDHURI e DAYAL [1997], para facilitar as análises complexas e visualização, os dados num *data warehouse* são tipicamente modelados com o modelo multidimensional. Nesse modelo, as medidas numéricas do negócio (ou métricas) são armazenadas em tabelas chamadas fato, que estão ligadas a outras tabelas, as chamadas tabelas dimensão, que armazenam as descrições auxiliares do fato. As dimensões possuem uma série de atributos em diferentes níveis de detalhes, que estão dispostos segundo uma hierarquia. Esses níveis da hierarquia é que vão atribuir a granularidade do fato, isto é, definir se o fato vai ser descrito com maior ou menor detalhe.

Para popular as tabelas fato e dimensão, é utilizado um processo chamado de *Extract/Transform/Load* (ETL). Nesse processo os dados são extraídos das bases operacionais, transformados e então carregados no *data warehouse*. Em [KIMBALL e CASERTA, 2004], o processo de ETL é expandido, sendo apresentado como tendo 4 fases (o processo de transformação é dividido em duas fases):

- Extração – Nessa fase, os dados provenientes dos sistemas fontes são salvos no disco com o mínimo de re-estruturação, fazendo com que a extração seja o mais simples e rápida possível.
- Limpeza – Essa fase é relacionada a qualidade do dado. Nela é checado se os dados extraídos possuem valores válidos, consistentes, não duplicados e se atendem às regras do negócio.
- Conformação – É nessa fase em que a integração total entre dados provenientes de diferentes fontes é feita.
- Entrega – Nessa fase os dados já estão estruturados de acordo com o esquema do data warehouse (usualmente o esquema estrela), prontos para serem armazenados no DW e analisados.

Em [CHAUDHURI e DAYAL, 1997] é apresentada uma arquitetura básica de um *data warehouse* (Figura 3.11):

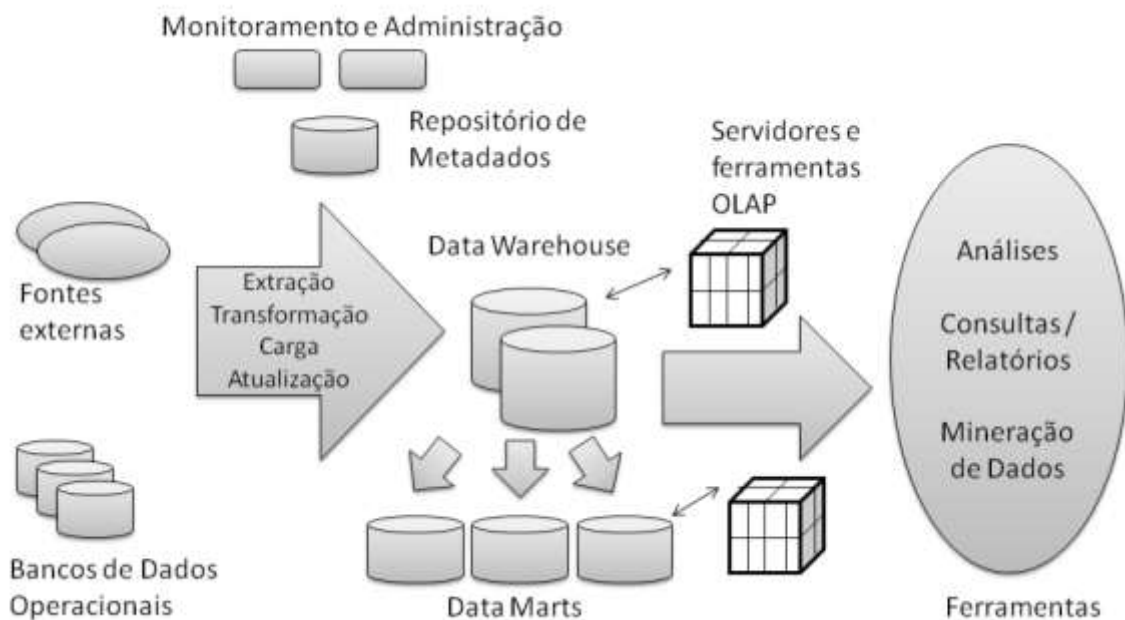


Figura 3.11 – Arquitetura de básica de um *data warehouse* [Adaptado de CHAUDHURI e DAYAL, 1997].

A arquitetura da Figura 3.11 representa as ferramentas para extrair os dados de múltiplas fontes; para limpar, transformar e integrar esses dados; para carregá-los dentro do *data warehouse*; e para atualizar o *data warehouse* periodicamente. Representa também os *data marts*, que são bases de dados que apresentam visões de processos de negócio ou de departamentos de uma empresa. As ferramentas OLAP apresentam as visões multidimensionais dos dados, permitindo fazer análises para diferentes finalidades. E esta arquitetura apresenta ainda um repositório de metadados para armazenar e gerenciar os metadados, além de ferramentas para monitorar e administrar todo o sistema do *data warehouse*.

3.4.1 On-Line Analytical Processing - OLAP

OLAP pode ser definido como um conjunto de tecnologias que permitem dar suporte a tomada de decisão através de consultas, análises e cálculos sobre dados armazenados num *data warehouse*. Um sistema OLAP possibilita realizar consultas que fazem

cruzamentos sobre dados agregados extraídos do *data warehouse* em diferentes níveis de detalhe e sob diferentes perspectivas.

Segundo CHAUDHURI e DAYAL [1997] e PAIM [2003], operadores típicos de um sistema OLAP incluem (Figura 3.12):

- Roll-up – este operador oferece uma visão dos dados num nível mais elevado da hierarquia dimensional; uma visão mais agregada.
- Drill-down – oferece uma visão mais refinada, com um nível de detalhe maior.
- Slice and dice – estes operadores permitem, fixando-se um valor de uma dimensão, analisar os dados por uma camada particular ou fatia do cubo (slice), ou por um sub-cubo dos dados (dice).
- Rotate ou pivot – operador que permite re-orientar a visão dos dados modificando a posição das dimensões no eixo do cubo dimensional, oferecendo uma nova configuração para análise.

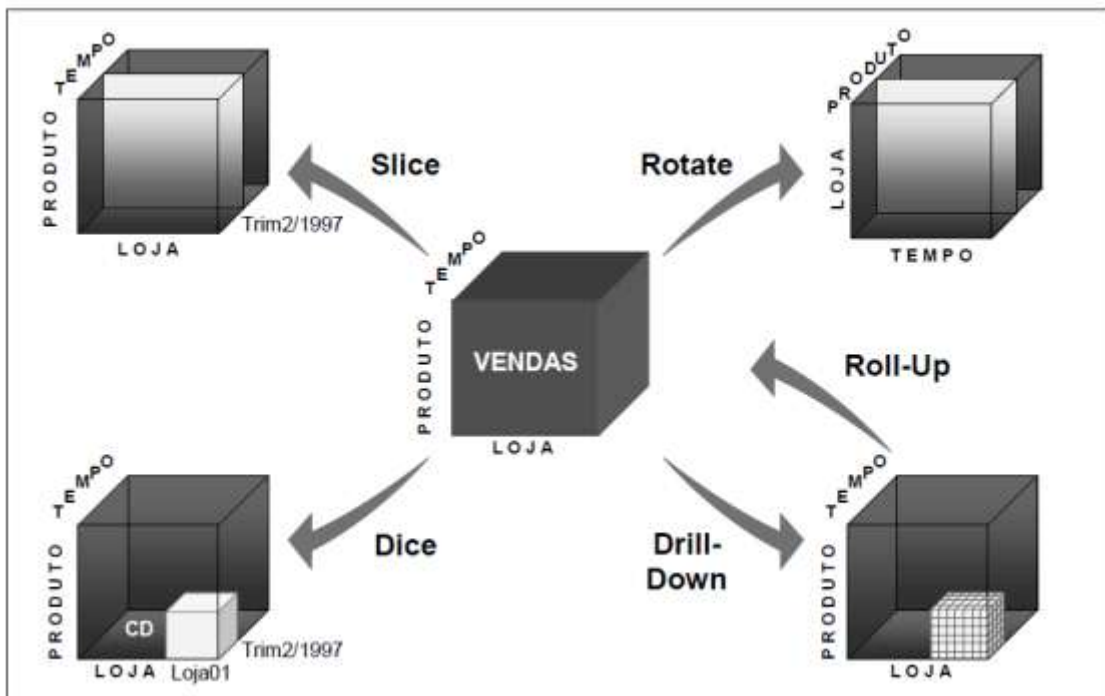


Figura 3.12 – Operadores OLAP [PAIM, 2003].

Como apresentado em [PINHO, 2008], um dos principais componentes de um sistema OLAP é o servidor OLAP. Este servidor fica entre o *data warehouse* e o cliente, armazenando os dados para serem analisados. Os servidores OLAP podem ser:

- Relational OLAP (ROLAP) – armazena os dados em banco de dados relacionais.
- Multidimensional OLAP (MOLAP) – armazena os dados em estruturas otimizadas para acesso multidimensional.
- Hybrid OLAP (HOLAP) – armazena alguns dados em formato multidimensional e outros em relacional.

3.4.2 Geographical Data Warehouse – GDW

Um GDW é um *data warehouse* estendido para que possa manipular e armazenar dados geográficos. Ele mantém as características tradicionais de um *data warehouse* e é capaz de manipular tanto dados geoespaciais quanto dados não espaciais que irão se complementar no auxílio à tomada de decisão [AHMED, 2008]. Para ele então, não apenas a integração de dados convencionais (não espaciais) é importante; a integração de dados geoespaciais também o é.

Segundo AHMED [2008], um GDW pode conter três tipos diferentes de dimensões geoespaciais (Figura 3.13): dimensão não geométrica, dimensão geométrica e dimensão híbrida.

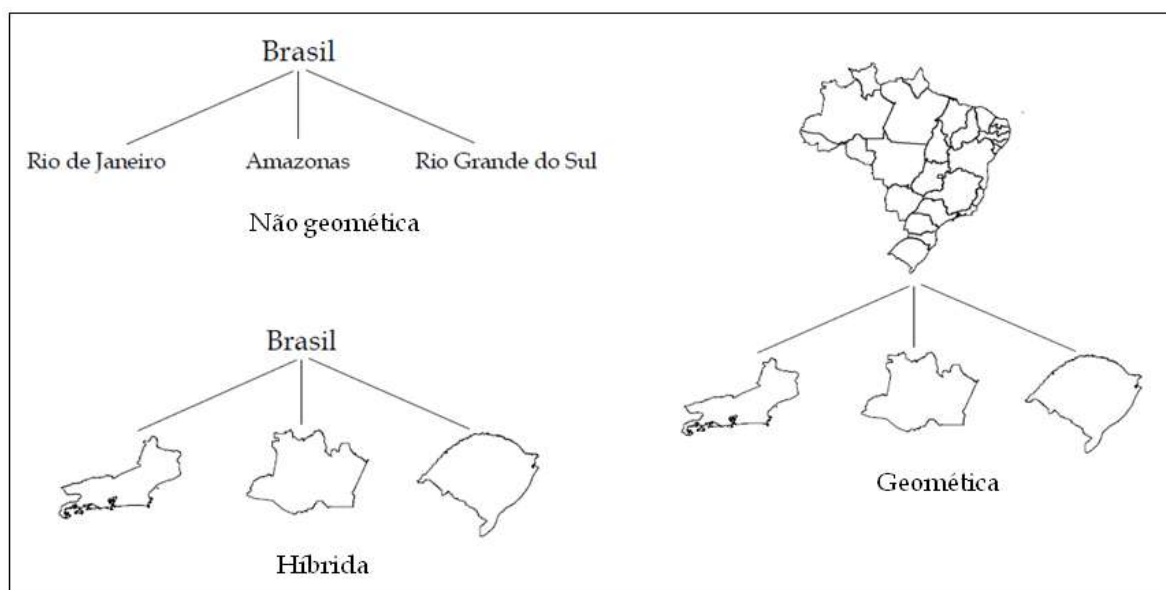


Figura 3.13 – Representação dos tipos de dimensões espaciais num GDW [Adaptado de AHMED, 2008].

Num sistema de GDW, a componente espacial da arquitetura do *data warehouse* deve consistir em três elementos [AHMED, 2008]: uma ferramenta para agregar os dados geoespaciais; um banco de dados que suporte dados espaciais; e ferramentas de análise e representação de dados geoespaciais sob a forma de mapas, gráficos ou tabelas (ferramentas SOLAP). Segundo AHMED [2008], o fator de atração mais importante num GDW é a capacidade que ele tem de fornecer dados que podem ser consultados e relacionados a áreas geográficas.

3.4.3 Spatial On-Line Analytical Processing - SOLAP

Spatial OLAP pode ser definido como um sistema OLAP tradicional ao qual foi integrado uma plataforma visual, construída especialmente para dar suporte rápido e fácil a análise espaço-temporal e exploração de dados seguindo uma abordagem multidimensional, composta de visualizadores cartográficos, bem como de visualizadores em forma de tabelas e diagramas [BÉDARD et al., 2007; AHMED, 2008]. Segundo AHMED [2008], as principais características desejáveis num sistema SOLAP são:

- Possuir uma interface flexível que suporta diferentes formatos de visualização dos dados, exibindo-os de forma cartográfica (mapas) ou não cartográfica (tabelas, diagramas).
- Todas as operações de navegação, bem como as funções de análise espaço-temporal, devem estar disponíveis em todas as formas de exibição dos dados.
- Deve fornecer meios de calcular medidas (ou métricas) a partir de medidas existentes.
- Deve dar suporte à análise espacial em diferentes dimensões do mesmo cubo de dados, além de fornecer suporte a múltiplas representações geométricas e múltiplas fontes de dados.

A Figura 3.14 apresenta a estrutura de um ambiente SOLAP que atende às características citadas acima. Essa estrutura é composta por:

- Ferramentas para extração, integração, limpeza e carga dos dados oriundos de fontes convencionais e de bancos de dados geográficos;
- Data Warehouse Geográfico (Geographical DW);
- Mecanismos para processamento analítico-multidimensional e geográfico e;
- Aplicações com interface gráfica, sincronizando mapas, tabelas e gráficos para a visualização e análise dos dados.

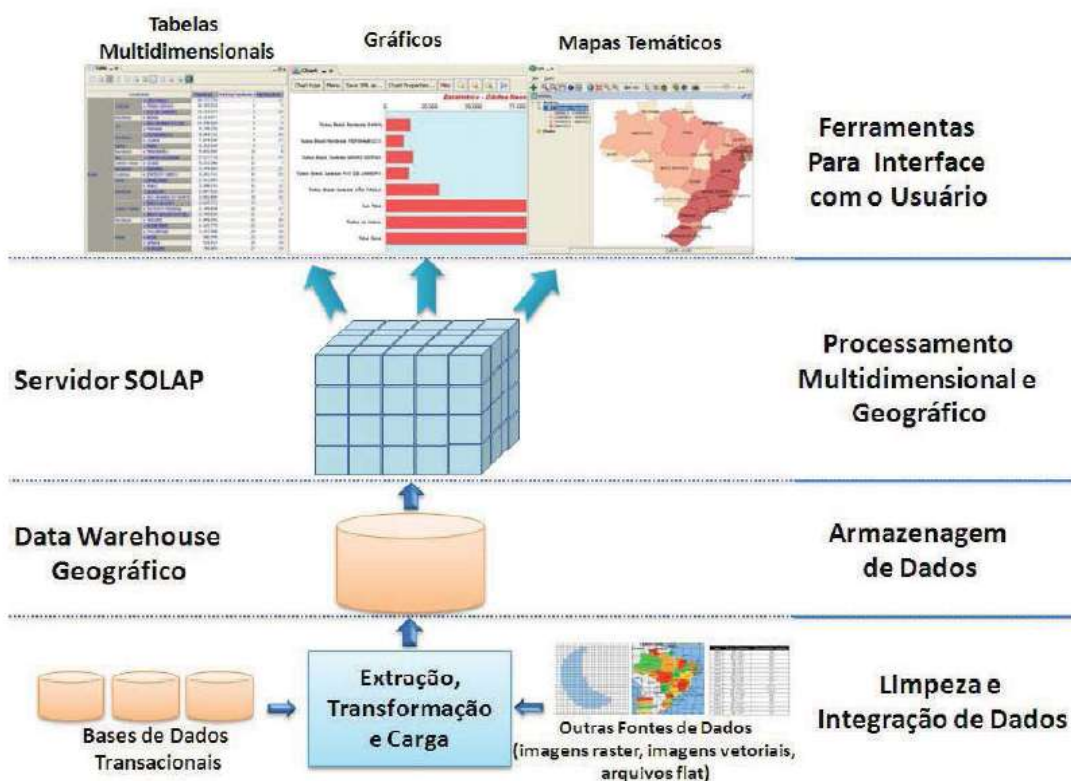


Figura 3.14 – Estrutura de um Ambiente SOLAP [SILVA, 2008].

3.4.4 Exemplos de projetos – GDW e SOLAP

O número de projetos de pesquisa relacionados a GDW e SOLAP é grande. Diferentes abordagens vem sendo adotadas na busca de integrar uma componente espacial em

data warehouse e em sistemas OLAP. Há projetos que visam a integração das tecnologias de OLAP e SIG estendendo o conceito de cubo de dados, como o Map Cube [SHEKHAR et al. 2001]; outros que visam conectar aplicações OLAP e GIS desenvolvidas separadamente, como o GOAL [KOUBA et al., 2000], o GOLAPA [FIDALGO et al., 2001] e o SIGOLAP [FERREIRA, 2002]; e há ainda outros projetos que visam o desenvolvimento de aplicações OLAP espaciais, como o GeoOLAP [MANHÃES et al., 2008]. Além desses projetos, existem pesquisas específicas para propor um *framework* para projetar esquemas multidimensionais geográficos, como o GeoDWFrame [FIDALGO et al, 2004a e 2004b], e também pesquisas sobre linguagens de consulta próprias para tratar de dados espaciais numa estrutura multidimensional, como a GeoMDQL [SILVA, 2008].

4 Padronização e disseminação

Nesta seção são apresentadas as principais organizações de padronização que fornecem suporte à integração e disseminação de dados geoespaciais e os principais padrões para metadados geoespaciais. Além disso, também é apresentado o que são as Infraestruturas de Dados Espaciais (IDE) e as principais características da IDE do Brasil.

4.1 Principais organizações de padronização

A padronização na maneira em que os dados são gerados e armazenados facilita o compartilhamento dos dados, pois como eles estão representados num padrão livre de plataforma, diferentes sistemas serão capazes de entendê-los sem precisar fazer nenhum tipo de transformação. O estabelecimento desses padrões são então de extrema importância para a integração de dados.

Os padrões aceitos internacionalmente são estabelecidos por organizações de grande reconhecimento. Nos tópicos a seguir são apresentadas as duas principais organizações no contexto deste trabalho, que são o OGC (*Open Geospatial Consortium*), que estabelece os padrões relacionados a dados geoespaciais, e o W3C (*World Wide Web Consortium*), que estabelece os padrões relacionados a Web.

4.1.1 OGC

O *Open Geospatial Consortium* é uma organização sem fins lucrativos criada em 1994, visando promover o desenvolvimento e uso de padrões para a geração, exploração e intercâmbio de dados geoespaciais. Ela é formada por membros do meio acadêmico, empresas que atuam com dados geoespaciais e órgãos governamentais de diversos países.

Os padrões criados pela OGC são chamados de padrões abertos (*Open standards*) pois possuem as seguintes características [OGC]:

- São criados de forma aberta, não-proprietária, com a participação internacional de qualquer companhia, agência ou organização.
- Têm direito de distribuição livre de qualquer imposto.
- Suas especificações têm acesso livre e são neutras de qualquer tecnologia.

O conjunto das especificações do OGC chama-se OpenGIS. Essas especificações vêm como resultados de diferentes grupos de pesquisas como por exemplo:

- 3 DIM (3D Information Management)
- Catalog
- Coordinate Reference System
- Coverages
- Data Preservation
- Data Quality
- Geo Rights Management
- Geography Markup Language (GML)
- Geosemantics
- Metadata (mantém correspondência com a ISO /TC 211 Metadata Standard)
- Security
- Web Feature Service

4.1.2 W3C

O *World Wide Web Consortium* é uma comunidade internacional, liderada por Tim Berners-Lee, onde membros de organizações, especialistas próprios e o público em geral trabalham juntos para desenvolver padrões web, visando garantir o crescimento e enriquecimento da experiência na web. Os princípios que guiam o trabalho do W3C são [W3C]:

- Web para todos, que visa dar acesso a Web a qualquer pessoa, seja qual for seu hardware, software, infraestrutura de rede, idioma, cultura, localização geográfica, ou habilidade física ou mental. Envolve iniciativas de acessibilidade e internacionalização.
- Web em tudo, que visa prover acesso em qualquer lugar, a qualquer hora utilizando qualquer dispositivo. Envolve iniciativas de web móvel e web voltada a outros dispositivos.

A Web Semântica, apresentada no tópico 2.3, aparece como uma visão do W3C, sendo apresentada como web de dados conectados na busca te trazer informações mais completas.

4.2 Padrões para metadados geoespaciais

Os padrões para metadados geoespaciais são importantes para a integração de dados pois definem um esquema padrão para os metadados, o que elimina as heterogeneidades estruturais. Eles também possuem em suas especificações os metametadados, que descrevem cada um dos elementos de metadados, para que não haja diferenças de interpretação sobre que valores determinado elemento deve receber.

Tais padrões são estabelecidos por organizações como o OGC, que os define e compartilha com a comunidade promovendo a disseminação. Segundo o IBGE [2009], os padrões de metadados geoespaciais estão conceituados e estruturados em seções com funções específicas de:

- Identificar o produtor e a responsabilidade técnica de produção;
- Padronizar a terminologia utilizada;
- Garantir o compartilhamento e a transferência de dados;
- Viabilizar a integração de informações;
- Possibilitar o controle de qualidade;
- Garantir os requisitos mínimos de disponibilização.

A seguir são apresentados os dois principais padrões para metadados geoespaciais.

4.2.1 Content Standard for Digital Geospatial Metadata (CSDGM)

O CSDGM é o atual padrão de metadados dos Estados Unidos [FGDC]. Ele foi proposto pelo FGDC (*Federal Geographic Data Committee*) que é um comitê interinstitucional norte-americano que promove a coordenação do uso e disseminação de dados geoespaciais. Sua primeira versão foi publicada em 1994 e em 1998 foi feita uma revisão. Apesar desse ainda ser um padrão bastante utilizado, desde março deste ano (2010) o FGDC começou a migrar para o padrão de metadados ISO 19115.

Este padrão faz parte da Infraestrutura Nacional de Dados Espaciais dos EUA (*National Spatial Data Infrastructure - NSDI*). Ele estabelece um conjunto comum de terminologia e definições para a documentação do dado geográfico. Como exemplos de elementos definidos, é possível citar: identificação, qualidade do dado, referência espacial e referência do metadado.

4.2.2 Padrão ISO 19115

Padrão especificado pelo Comitê Técnico 211 (TC 211 - *Geographic information/Geomatics*) da ISO (*International Organization for Standardization*), este padrão se tornou o mais aceito pela comunidade produtora de dados espaciais, passando a ser adotado como padrão para a infraestrutura de dados espaciais de vários países. Ele foi publicado em 2003 e teve uma revisão em 2006.

Esse padrão é muito amplo possuindo cerca de 300 elementos, dos quais 8 são obrigatórios. Um número tão grande de elementos permite utilizar esse padrão nos mais diferentes domínios, e os elementos obrigatórios garantem a interoperabilidade. Esse padrão pode também ser utilizado como base para estabelecimento de perfis de metadados, ou seja, ele pode ser customizado.

Ele é o padrão aceito pelo Open Geospatial Consortium – OGC.

4.3 Infraestrutura de Dados Espaciais

As Infraestruturas de Dados Espaciais (IDE) – ou em inglês *Spatial Data Infrastructure* (SDI) – facilitam a integração na medida em que elas promovem o compartilhamento e disseminação dos dados geoespaciais, bem como de seus metadados. Elas ainda especificam que padrões devem ser adotados pelos produtores de dados.

Segundo VAN HARMELEN [2008], para obter total interoperabilidade uma infraestrutura de dados espaciais deve conseguir resolver os problemas de integração nos diferentes níveis: nível físico, sintático e semântico. E para atingir a integração semân-

tica é necessário expressar não somente os dados mas também o modelo dos dados em linguagens formais e de preferência padronizadas.

Segundo a GSDI (*Global Spatial Data Infrastructure Association*) [2004], uma IDE deve ser mais do que um conjunto de dados ou base de dados; uma IDE deve garantir suficiente documentação (metadados), um meio de descobrir, visualizar e avaliar os dados (catálogos e mapas), e alguns métodos para fornecer acesso aos dados geográficos. Além disso, fornecer serviços adicionais ou software para suportar aplicações de dados. Para a descoberta ou recuperação de dados, as ontologias farão um papel de grande importância, fornecendo o significado pretendido dos dados.

Com o crescente interesse na utilização dos dados geoespaciais, cada vez mais países estão buscando estabelecer as suas infraestruturas. Como citado na introdução, no Brasil está sendo implementada a INDE (Infraestrutura Nacional de Dados Espaciais); na Europa foi implantada a INSPIRE (*Infrastructure for Spatial Information in Europe*); nos Estados Unidos foi criada a NSDI (*National Spatial Data Infrastructure*); e estes são só alguns exemplos.

4.3.1 INDE

A Infraestrutura Nacional de Dados Espaciais (INDE) é o nome da IDE brasileira que vem sendo especificada desde 2008 e começou a ser implantada no início de 2010, sendo gerenciada pelo IBGE (Fundação Instituto Brasileiro de Geografia e Estatística). Ela foi criada para integrar todos os dados geoespaciais existentes nas diversas instituições do governo brasileiro facilitando o compartilhamento e disseminação dos dados. Esse compartilhamento foi estabelecido como sendo obrigatório para todos os órgãos do governo federal.

A INDE é uma infraestrutura orientada a serviços o que significa, segundo o Plano de Ação da INDE, que ela pode ser entendida como a confluência entre diversos provedores de dados geográficos, cada qual fornecendo acesso a dados através de serviços web específicos, que podem ser encontrados através de mensagens XML. Para escolher quais dados e quais serviços preenchem suas necessidades, o usuário realiza buscas através de um repositório de metadados sobre informações e geosserviços disponíveis (Figura 4.1). O repositório de metadados da INDE pode ser acessado através do site: <www.metadados.inde.gov.br>, onde é possível buscar e fazer *download* dos metadados em XML. Os metadados estarem em XML e seguindo um perfil baseado no padrão ISO 19115 (perfil MGB) facilita o compartilhamento desses metadados que poderão ser utilizados para escolher que dados são adequados para determinado uso.

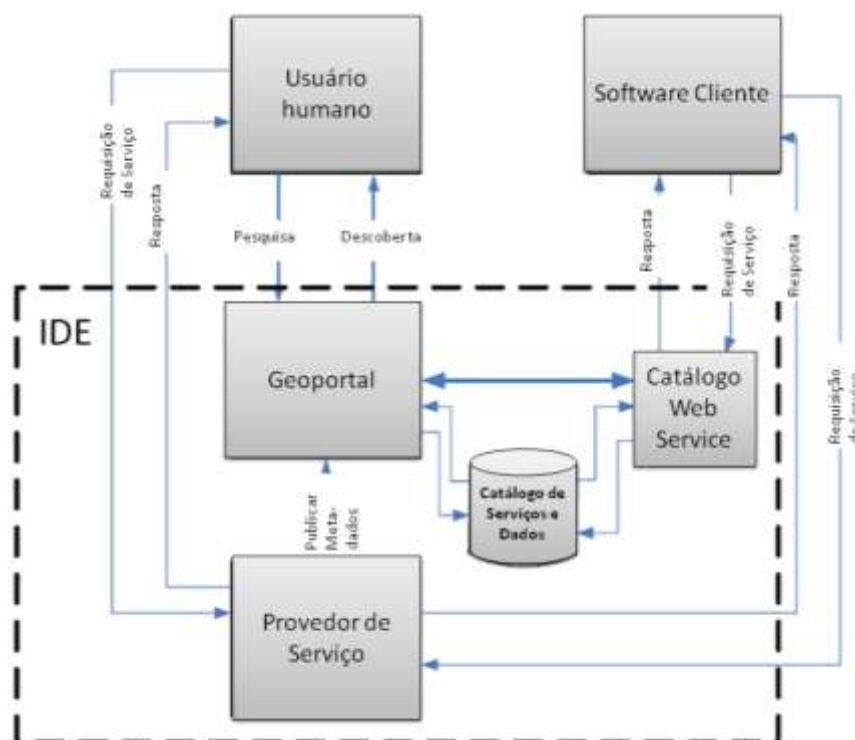


Figura 4.1 – Infraestrutura de Dados Espaciais orientada a Serviços [Adaptado de DAVIS JR. e ALVES, 2006].

Outros serviços como visualização de mapas, download de arquivos vetoriais e buscas através de nomes ou áreas aproximadas também são disponíveis via internet no portal da inde <www.inde.gov.br>.

4.3.2 Perfil MGB (Metadados Geoespaciais do Brasil)

Segundo IBGE [2009] um perfil de metadados estabelece um conjunto básico de elementos que retratam as características dos produtos geoespaciais de uma determinada comunidade e garante sua identificação, avaliação e utilização consistente. Esse conjunto básico é proposto como o núcleo comum a todos os tipos de produtos geoespaciais, sendo que os produtos de mapeamento especial, cadastral e temático requerem maior detalhamento dos itens de algumas seções dos metadados para retratar suas peculiaridades.

O perfil MGB é baseado na norma ISO 19115, cujo esquema padrão estabelecido foi customizado pelo IBGE para se adequar aos dados geoespaciais do Brasil. Antes da elaboração do perfil MGB, foram analisados os perfis já estabelecidos em vários outros países, todos também baseados na norma ISO 19115, como por exemplo os perfis:

- MIG – Metadados de Informação Geográfica (Portugal) (*);
- NEM – Núcleo Espanhol de Metadados (Espanha) (*);
- NAP – North American Profile (EUA/Canadá);
- LAMP – Latin American Metadata Profile (proposto para América Latina).

(*) Ambos perfis (MIG e NEM) dentro da iniciativa europeia INSPIRE.

O perfil MGB, importante para a integração de dados no Brasil, possui um perfil resumizado, que representam os elementos mínimos que devem ter os metadados geo-

espaciais produzidos. Esse perfil sumarizado possui 23 elementos dos quais 11 são de preenchimento obrigatório. Abaixo a tabela 4.1 que representa estes elementos.

Tabela 4.1: Elementos do Perfil MGB sumarizado. [IBGE 2009]

Entidades e elementos do núcleo de metadados do Perfil MGB sumarizado			
Entidade/Elemento	Obrigatoriedade	Entidade/Elemento	Obrigatoriedade
1. Título	Obrigatório	12. Tipo de Representação Espacial	Opcional
2. Data	Obrigatório	13. Sistema de Referência	Obrigatório
3. Responsável	Obrigatório	14. Linhagem	Opcional
4. Extensão Geográfica	Condicional	15. Acesso Online	Opcional
5. Idioma	Obrigatório	16. Identificador Metadados	Opcional
6. Código de caracteres do CDG	Condicional	17. Nome Padrão de Metadados	Opcional
7. Categoria Temática	Obrigatório	18. Versão da Norma de Metadados	Opcional
8. Resolução Espacial	Opcional	19. Idioma dos Metadados	Condicional
9. Resumo	Obrigatório	20. Código de Caracteres dos Metadados	Condicional
10. Formato de Distribuição	Obrigatório	21. Responsável pelos Metadados	Obrigatório
11. Extensão Temporal e Altimétrica	Opcional	22. Data dos Metadados	Obrigatório
		23. Status	Obrigatório

5 Conclusão

Com este trabalho foi possível perceber que a integração de dados geoespaciais é um assunto de grande importância e que vem sendo cada vez mais pesquisado meios de poder utilizar de forma eficiente esses dados geoespaciais no suporte à tomada de decisão. Foi possível perceber também que a integração de dados, sejam eles geoespaciais ou convencionais, não se dá apenas transformando o formato do arquivo de armazenamento de dados de um sistema para outro. É preciso levar em consideração questões sintáticas e a semântica dos dados. Isso é claro, ao se tratar de uma integração mais complexa, a física. Como visto, nem sempre esse tipo de integração é necessária.

Na pesquisa a respeito dessa integração mais complexa, chegou-se a assuntos como Geographical Data Warehouse e Spatial On-Line Analytical Processing. Foi visto que muita pesquisa vem sendo desenvolvida nessa área, com projetos em universidades e lançamentos de ferramentas SOLAP proprietárias e abertas. O que não parece ainda definido de uma maneira amplamente aceita é como projetar esse data warehouse estendido para manipular dados geográficos.

Dentre todos os assuntos abordados neste trabalho, foi percebido que há também muito interesse em pesquisar o uso de ontologias para a integração de dados geoespaciais e a integração de dados em diferentes sistemas de coordenadas.

6 Referências

- AHMED, T. O. **Spatial On-line Analytical Processing (SOLAP): Overview and Current Trends**. In: International Conference on Advanced Computer Theory and Engineering (ICACTE '08). 2008.
- ANTONIOU, G.; VAN HARMELEN, F. **A Semantic Web Primer**. The MIT Press, Cambridge, Massachusetts. 2004.
- BARBOSA, C. C. F. **Álgebra de mapas e suas aplicações em sensoriamento remoto e geoprocessamento**. Dissertação (Mestrado). Instituto Nacional de Pesquisas Sociais. São José dos Campos, São Paulo. 1997.
- BARROS, D. M. V. **Especificação de testes funcionais para o processador de consultas da arquitetura GOLAPA**. Monografia (Graduação). Universidade Federal de Pernambuco. 2007.
- BÉDARD, Y.; RIVEST, S.; PROULX, M.-J. **Spatial On-Line Analytical Processing (SOLAP): Concepts, Architectures and Solutions from a Geomatics Engineering Perspective**. In: WREMBEL, R.; KONCILIA, C. (Eds.). *Data Warehouses and OLAP: Concepts, Architectures and Solutions*, Chap. 13, IRM Press (Idea Group), London, UK, p. 298-319. 2007.
- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. **The Semantic Web**. In: *Scientific American*, volume 284. 2001.
- BITTNER, T.; WINTER, S. **On ontology in Image Analysis**. In: AGOURIS, P.; STEFANIDIS, A. (Eds.), *Integrated Spatial Databases: Digital Images and GIS*. Springer-Verlag: Berlin. p. 168-191. 1999.
- CÂMARA, G.; EGENHOFER, M.; FONSECA, F.; MONTEIRO, M. **What's in an Image?** In: COSIT'01 - Conference on Spatial Information Theory, 2001. *Lectures Notes in Computer Science*, vol. 2205, p. 474-488. 2001.
- CÂMARA, G.; CASANOVA, M.; DAVIS, C.; VINHAS, L.; QUEIROZ, G.R. **Banco de Dados - Livro online do INPE**. 2005. Disponível em: < <http://www.dpi.inpe.br/livros/bdados/> >.
- CARNEVALE, M.T. **Infra-estrutura para informações espaciais**. 2008. 111 f. Dissertação (Mestrado) - Universidade do Estado do Rio de Janeiro, Faculdade de Engenharia, Rio de Janeiro, 2008.
- CASANOVA, M.A.; CÂMARA, G.; HEMERLY, A.S.; MAGALHÃES, G.C.; MEDEIROS, C.M.B. **Anatomia de Sistemas de Informação Geográfica**. 1996. Disponível em: < <http://www.dpi.inpe.br/geopro/livros/anatomia.pdf> >
- CHAUDHURI, S.; DAYAL, U. **An Overview of Data Warehousing and OLAP Technology**. In: *ACM SIGMOD Record*, v. 26 n. 1, p. 65-74. 1997.
- DAVIS JR, C.A.; ALVES, L.L. **Infra-Estruturas de Dados Espaciais: Potencial para Uso Local**. In: *Informática Pública*, vol. 8, p 65-80. 2006.
- FERREIRA, A. C. F. **Um modelo para suporte à integração de análises multidimensionais e espaciais**. Dissertação (Mestrado). Universidade Federal do Rio de Janeiro. 2002.
- FIDALGO, R. N.; TIMES, V. C.; SOUZA, F. F. **GOLAPA: Uma Arquitetura Aberta e Extensível para Integração entre SIG e OLAP**. In: *Simpósio Brasileiro de Geoinformática - GEOINFO*. 2001.

- FIDALGO, R. N.; TIMES, V. C.; SILVA, J.; SOUZA, F. F. **Geodwframe: A framework for guiding the design of geographical dimensional schemas**. In: Proceedings of the Data Warehousing and Knowledge Discovery. Zaragoza, Spain, p. 26-37. 2004a.
- FIDALGO, R. N.; TIMES, V. C.; SILVA, J.; SOUZA, F. F.; SALGADO, A. C. **Providing Multidimensional and Geographical Integration Based on a GDW and Metamodels**. In: Simpósio Brasileiro de Banco de Dados (SBBD). 2004b.
- FONSECA, F.T.; EGENHOFER, M. J. **Sistemas de Informação Geográficos Baseados em Ontologias**. In: Informática Pública 1 (2): 47-65. 1999.
- FONSECA, F. T.; EGENHOFER, M. J.; BORGES, K. **Ontologias e Interoperabilidade Semântica entre SIGs**. In: II Workshop Brasileiro em Geoinformática - GeoInfo2000, SãoPaulo. 2000.
- FONSECA, F. T.; EGENHOFER, M. J.; AGOURIS, P.; CÂMARA, G. **Using Ontologies for Integrated Geographic Information Systems**. Transactions in GIS, vol. 6, no. 3, p. 231 – 257. 2002.
- GERBER, A. **Towards a comprehensive functional layered architecture for the semantic web**. Tese (Doutorado). Universidade da África do Sul. 2006.
- GERBER, A.; VAN DER MERWE, A; BARNARD, A. **A Functional Semantic Web Architecture**. In: 5th European semantic web conference on The semantic web: research and applications. 2008.
- GOH, C. H. **Representing and Reasoning about Semantic Conflicts in Heterogeneous Information Sources**. Phd Thesis, MIT, 1997.
- GRUBER, T. **Toward Principles for the Design of Ontologies Used for Knowledge Sharing**. International Journal Human-Computer Studies, Vol. 43, Issues 5-6, p.907-928. 1993.
- GSDI. **Developing Spatial Data Infrastructures: the SDI Cookbook**. V.2, 2004. Disponível em: < <http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf> >
- GUARINO, N. **Understanding, Building, and Using Ontologies: A Commentary to "Using Explicit Ontologies in KBS Development"**, by van Heijst, Schreiber, and Wielinga. International Journal of Human and Computer Studies, Vol. 46,p.293-310. 1997.
- GUARINO, N. **Formal Ontology and Information Systems**. In: FOIS'98 - Formal Ontology and Information Systems 98. 1998.
- GUIZZARDI, G. **On Ontology, ontologies, Conceptualizations, Modeling Languages, and (Meta)Models**. In: Frontiers in Artificial Intelligence and Applications, Databases and Information Systems IV, IOS Press, Amsterdam. 2007.
- HORROCKS, I.; PARSIA, B.; PATEL-SCHNEIDER, P. F.; HENDLER, J. **Semantic Web Architecture: Stack or Two Towers?** In: PPSWR - Workshop on Principles and Practice of Semantic Web Reasoning. 2005.
- IBGE. **INDE - Infraestrutura Nacional de Dados Espaciais**. < <http://www.inde.gov.br/> >.
- IBGE. **Perfil de Metadados Geoespaciais do Brasil**. Brasil, 2009. Disponível em: < http://www.concar.ibge.gov.br/arquivo/Perfil_MGB_Final_v1_homologado.pdf >

IBGE. **Plano de ação para implantação da Infra-Estrutura Nacional de Dados Espaciais.** Brasil, 2010. Disponível em: <<http://www.concar.ibge.gov.br/arquivo/PlanoDeAcaoINDE.pdf>>.

INMON, W.H. **Building the Data Warehouse.** 4th Edition. Wiley Publishing, Inc. Indianapolis, IN. 2005.

KASHYAP, V.; SHAH, K.; SHETH, A. **Metadata for building the MultiMedia Patch Quilt.** In: Jajodia S, Subrahmanian V S (eds) Multimedia Database Systems: Issues and Research Directions. Springer-Verlag: 297-319. 1995.

KIMBALL, R.; CASERTA, J. **The data warehouse ETL toolkit : practical techniques for extracting, cleaning, conforming, and delivering data.** Wiley Publishing, Inc. Indianapolis, IN. 2004.

KOIVUNEN, M. R.; MILLER, E. **W3C Semantic Web Activity.** In: Semantic Web Kick-off Seminar in Finland. 2001.

KOLAS, D.; HEBELER, J.; DEAN, M. **Geospatial semantic web: Architecture of ontologies.** In: RODRIGUEZ, M. A.; CRUZ, I. F.; EGENHOFER, M. J.; LEVASHKIN, S. (Eds.). 1st International Conference on GeoSpatial Semantics, pages 183–194, Cidade do México, México. 2005.

KOUBA, Z; MATOUSEK, K.; MIKSOVSKÝ, P. **On data warehouse and gis integration.** In: Proceedings of the 11th International Conference on Database and Expert Systems Applications. Springer-Verlag, London, UK, p. 604–613. 2000.

LENZERINI, M. **Data Integration: A Theoretical Perspective.** Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems – PODS, pages 243-246. 2002.

MANHÃES, R. S.; GONZÁLEZ, S. M.; COLONESE, G.; CARVALHO, R. A.; TANAKA, A. K. **GeoOlap: An Integrated Approach for Decision Support .** In: Second IFIP TC8 International Conference on Research and Practical Issues of Enterprise Information Systems (CONFENIS 2008), Research and Practical Issues of Enterprise Information Systems. Springer-Verlag. Pequim, China. 2008.

MATHIAK, B.; KUPFER, A.; NEUMANN, K. **Using XML languages for modeling and Web-visualization of geographical legacy data.** In: VI Brazilian Symposium on GeoInformatics (GeoInfo 2004), Campos do Jordão (SP). 2004.

OGC. **Open Geospatial Consortium.** <<http://www.opengeospatial.org/>>.

OGC. **OpenGIS Geography Markup Language (GML) Encoding Standard.** Versão 3.2.1. 2007.

PAIM, F. R. S. **Uma Metodologia para Definição de Requisitos em Sistemas Data Warehouse.** Dissertação (Mestrado). Universidade Federal de Pernambuco. 2003.

PINHO, F. R. **Estendendo o Servidor OLAP Mondrian com UDF Envolvendo Operadores Espaciais.** Monografia (Graduação). Universidade Federal de Pernambuco. 2008.

SHEKHAR, S.; LU, C. T.; CHAWLA, S.; VATSAVAI, R. R. **Map Cube: A Visualization Tool for Spatial Data Warehouses.** In: MILLER, H. and HAN, J. (Eds). Geographic data mining and knowledge discovery. 2001.

SHETH, A. **Changing Focus on Interoperability in Information Systems: from System, Syntax, Structure to Semantics.** In: GOODCHILD, M.; EGENHOFER, M.;

FEGEAS, R.; KOTTMAN, C. (Eds.) *Interoperating Geographic Information Systems*. p. 5-29, Kluwer Academic Publishers, Norwell, Ma. 1999.

SILVA, J. **GeoMDQL: Uma linguagem de consulta geográfica e multidimensional**. Tese (Doutorado). Universidade Federal de Pernambuco. 2008.

SMITH, B.; MARK, D. **Ontology and Geographic Kinds**. In: *International Symposium on Spatial Data Handling*. Vancouver, Canada. p. 308-320. 1998.

VAN HARMELEN, F. **Semantic Web Technologies as the Foundation for the Information Infrastructure**. In: VAN OOSTEROM, P.; ZLATANOVA, S. (Eds) *Creating spatial information infrastructures: towards the spatial Semantic Web*. p. 38-51. 2008.

W3C. **World Wide Web Consortium**. < <http://www.w3.org/> >.

W3C. **Scalable Vector Graphics (SVG) 1.1 Specification**. Versão 1.1. 2003.

ZIEGLER, P., DITTRICH, K. R. **Data Integration – Problems, Approaches, and Perspectives**. In: KROGSTIE, J.; OPDAHL, A. L.; BRINKKEMPER, S., (Eds), *Conceptual Modelling in Information Systems Engineering*, pages 39-58. Springer, Berlin. 2007.