

## Camada de Acesso a Dados Geo-Espaciais na TerraLib 5

Gilberto Ribeiro de Queiroz<sup>1</sup>, Juan P. Garrido<sup>1</sup>, Lúbia Vinhas<sup>1</sup>, Karine Reis Ferreira<sup>1</sup>, Lauro Hara<sup>1</sup>, Marisa Xavier<sup>1</sup>, Emiliano Ferreira Castejon<sup>1</sup>, Gilberto Câmara<sup>1</sup>, Antônio M. V. Monteiro<sup>1</sup>, Ricardo Cartaxo Modesto de Souza<sup>1</sup>

<sup>1</sup>Divisão de Processamento de Imagens – DPI  
Instituto Nacional de Pesquisas Espaciais – INPE

{gribeiro, juan, lubia, karine, miguel, lauro, marisa, castejon, cartaxo}@dpi.inpe.br, gilberto.camara@inpe.br

**Abstract.** *This paper presents the geo-spatial data access module that is being designed for the new generation of TerraLib. Its architecture provides the base building block for robust geographic applications with advanced analytical data processing capabilities. It also enables applications to access and query spatial data over heterogeneous and distributed data sources.*

**Resumo.** *Este trabalho apresenta a camada de acesso a dados geo-espaciais que está sendo projetada para a nova geração da TerraLib. Esta camada possui uma arquitetura que fornece os elementos necessários para construção de aplicativos geográficos com maior robustez, capacidade analítica e de processamento de dados em fontes de dados heterogêneas e distribuídas.*

**Palavras-chave:** *GIS, TerraLib, Software Architecture.*

### 1. Introdução

A TerraLib (Câmara et al., 2008) é um ambiente para pesquisa e desenvolvimento de inovações tecnológicas na área de Geo-Informática. Do ponto de vista de pesquisa, ela serve de suporte à comunidade científica para o desenvolvimento e validação de suas teorias. Ela se tornou a base de diversos sistemas do INPE como TerraView (TerraView, 2009), TerraAmazon (Prodes, 2011), SISMADEN (SISMADEN, 2011) e TerraME (Carneiro, 2006). Outras instituições também têm apostado em seu potencial, fazendo parte da sua rede de colaboradores.

Este projeto tem suas raízes no início dos anos 2000, tendo sido conceitualizado ao redor da criação de uma camada de abstração e acesso a dados espaciais em sistemas de bancos de dados. Na época, as extensões geo-espaciais para SGBDs estavam começando a ser desenvolvidas. Não havia a quantidade de sistemas livres na área de geo-informática disponíveis atualmente. Nem a indústria geo-espacial havia consolidado muitos dos padrões em uso atualmente. O consórcio OGC e a ISO estavam trabalhando nas primeiras versões de especificações para extensões geo-espaciais (OGC, 2009; ISO, 2003). As questões de interoperabilidade utilizando a tecnologia de Web Services neste tipo de sistemas também estavam começando a ser endereçadas. A W3C ainda não havia padronizado a arquitetura SOA.

Por outro lado, ao final dos anos 2000, vários dos padrões ganharam maturidade

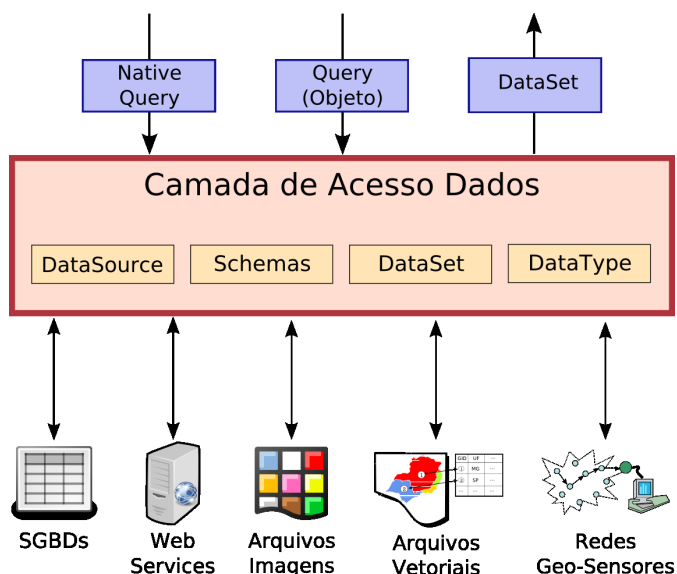
e vários projetos de software livre emergiram quase em todos os campos da computação. Especificamente, na área de bancos de dados, algumas novas tecnologias surgiram para resolver problemas de natureza mais específica, principalmente, para manipular grandes volumes de dados. Sistemas de bancos de dados orientado a coluna (column-oriented DBMS), database engines, No-SQL, a API MapReduce, e SGBDs com suporte a API map/reduce foram desenvolvidos. Conforme apontado por Stonebraker et al. (2008), os sistemas de bancos de dados se tornaram cada vez mais especializados. A idéia de que “one size fits all” não é mais válida.

A nova geração da TerraLib, chamada TerraLib 5, parte deste novo cenário. Ela possui uma arquitetura projetada para atender tanto a demanda de sistemas comerciais, através da adoção de padrões industriais estabelecidos, quanto de aplicações científicas por meio de um ambiente integrado com diversas ferramentas de análise de dados (R, Gnu Octave entre outras). Além disso, dada a natureza multidisciplinar da área de geoinformação, esta nova versão também é projetada para prover a ligação com outros ambientes de programação (Lua, Haskell, Python, PHP, Java) que não apenas a linguagem C++.

Este trabalho tem como objetivo apresentar as camadas da nova geração da TerraLib que lidam com o acesso a dados. A Seção 2 apresenta a camada de acesso a dados da TerraLib 5. A camada de processamento e integração de dados em fontes de dados distribuídas e heterogêneas é apresentada na Seção 3. Por último, são apresentadas algumas considerações sobre o trabalho.

## 2. TerraLib 5 - Camada de Acesso a Dados

O módulo denominado **Camada de Acesso a Dados** (Figura 2) possibilita que as aplicações manipulem dados espaciais em diferentes tipos de fontes de dados, abstraindo desde os tradicionais SGBDs até os Serviços Web da OGC. Este módulo é composto pelas seguintes abstrações: *data source*, *dataset*, *data source catalog*, *dataset type*, *query* e *data source catalog loader*.



**Figura 2. Módulo de Acesso a Dados**

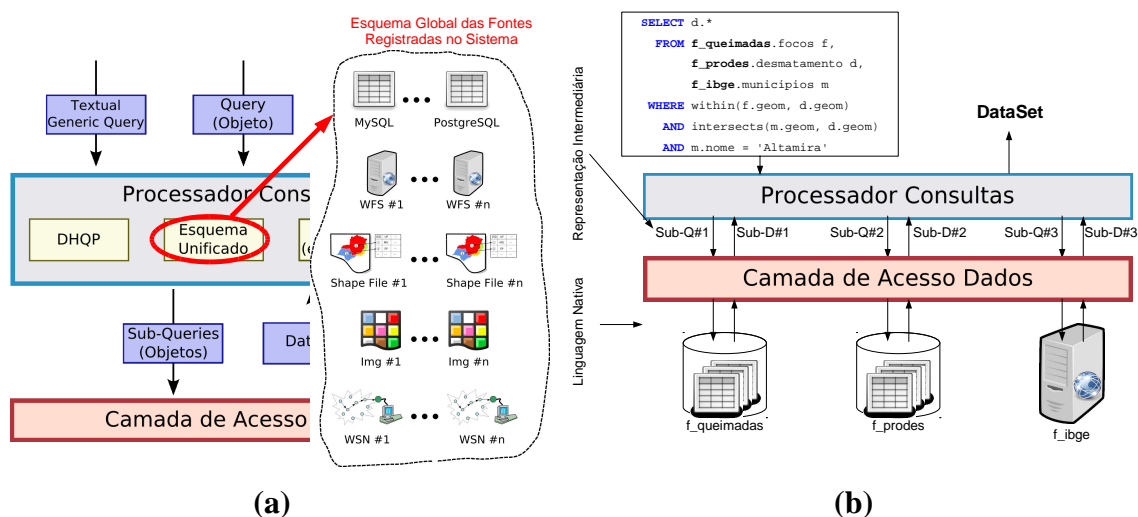
Além de permitir o acesso aos dados, essas abstrações também possibilitam que uma aplicação descubra o que está armazenado em uma fonte de dados e como o dado encontra-se organizado na fonte. A camada de acesso a dados é abstrata, isto é, ela

define apenas a interface que deve ser implementada por *drivers* específicos que entendam dos detalhes de cada tipo de fontes de dados.

### 3. TerraLib 5 - Processamento de Consultas em Fontes de Dados Heterogêneas e Distribuídas

O módulo chamado **Processador de Consultas** (Figura 4), abreviadamente **qp**, é um dos módulos atualmente em construção na TerraLib, projetado no topo da camada de acesso. O qp será capaz de fazer a junção e seleção de dados em diferentes tipos de fontes de dados, o que possibilitará criar aplicações que realizem consultas em fontes de dados heterogêneas e distribuídas. Esta capacidade aumentará o poder de processamento que os pesquisadores terão em mãos quando da realização de experimentos envolvendo dados provenientes de diversas organizações/instituições.

O qp tomará como entrada uma consulta expressa na mesma notação suportada pela camada de acesso, analisando a consulta para verificar se mais de uma fonte de dados é referenciada, decompondo-a em sub-consultas se necessário e então traçando os planos de consulta (Figura 4.b). Como não existe homogeneidade entre as capacidades das fontes de dados e, conseqüentemente, das implementações da camada de acesso (*drivers*), mesmo no caso de uma consulta que envolva apenas uma única fonte de dados o qp deverá checar se as operações envolvidas são todas suportadas diretamente pela camada de acesso ou se será necessário criar um plano de execução mais elaborado. Quando possível, nenhum overhead será adicionado à execução de consultas sobre fontes únicas pois a consulta será simplesmente repassada para a camada de acesso. Algum grau de paralelismo poderá ser utilizado quando as sub-consultas forem enviadas à camada de acesso. Após receber os datasets retornados pelas execuções na camada de acesso, o qp poderá realizar alguns refinamentos até finalmente montar o dataset que será entregue à aplicação.



**Figura 3. Módulo de Processamento de Consultas**

Atualmente, o qp está sendo desenvolvido e possui quatro componentes básicos:

1. *global metadata cache manager* (Figura 3.a): integração dos catálogos de cada fonte de dados. As mesmas estruturas da camada de acesso são utilizadas aqui para fornecer uma visão unificada de todos os dados disponíveis nas fontes de dados registradas no sistema.

2. *global data cache manager*: mantém um cache dos datasets mais

frequentemente utilizados nas consultas. Também mantém algumas estruturas de índices temporários para ajudar na execução das consultas que envolvem junções.

3. *query planer*: faz uma análise da consulta, montando a sequência de instruções.

4. *query executor*: coordena a execução das sub-consultas pela camada de acesso, realiza o refinamento dos resultados e devolve o dado final.

Atualmente, este módulo está sendo prototipado utilizando-se a máquina de execuções do SQLite (SQLite, 2011) com o objetivo de validar os itens 1 e 2 acima sem a necessidade de implementação imediata dos itens 3 e 4.

#### 4. Resultados e Conclusões

Os dois módulos descritos neste trabalho são fundamentais para atender às demandas de gerenciamento de dados dos aplicativos geográficos atuais. A nova camada de acesso a dados da TerraLib fornece um sistema de tipos mais flexível e extensível para lidar com dados provenientes de diferentes fontes de dados. Uma vez que novas fontes de dados podem possuir diferentes tipos de dados, como por exemplo, um objeto móvel, é importante que esta camada possibilite a integração destas fontes de dados.

A arquitetura destes módulos contemplam vários aspectos de engenharia que vão além das especificações de interoperabilidade propostas pelo Open Geospatial Consortium (OGC, 2009). A expectativa é que estes módulos ajudem a tornar o trabalho de desenvolvimento de novas aplicações mais fácil e possibilite uma exploração mais eficaz de dados de geo-sensores, dispositivos móveis e bancos de imagens gerados por constelações de satélites de sensoriamento remoto.

#### Referências

- Câmara, G.; Vinhas, L.; Ferreira, K. R.; Queiroz, G. R.; Souza, R. C. M.; Monteiro, A. M. V.; Carvalho, M. T.; Casanova, M. A.; Freitas, U. M. An Open Source GIS Library for Large-scale Environmental and Socio-economic Applications. In. Open Source Approaches to Spatial Data Handling. Berlin, Springer-Verlag, 2008.
- Carneiro, T. Nested-CA: a foundation for multiscale modeling of land use and land change. Doutorado em Computação Aplicada, Junho 2006.
- DPI. Manual do TerraView. Disponível em: <http://www.terralib.org/terraview>. Acesso em: Outubro de 2009.
- ISO. SQL Multimedia and Application Packages - Part 3: Spatial 3rd ed. 2003.
- OGC. OpenGIS® Standards and Related OGC documents. Disponível em: <http://www.opengeospatial.org/standards>. Acesso em: Outubro, 2009.
- Queiroz, G. R.; Ferreira, K. R.; Vinhas, L.; Câmara, G.; Monteiro, A. M. V.; Garrido, J. C. P.; Hara, L.; Xavier, M.; Castejon, E. F.; Souza, R. C. M. TerraLib 5.0: Supporting Data-Intensive GIScience. X Workshop dos Cursos de Computação Aplicada do INPE – WORCAP, 2010.
- SISMADEN. Home page: <http://www.dpi.inpe.br/sismaden>. Acesso em Julho 2011.
- SQLite. Home page: <http://www.sqlite.org/>. Acesso em Julho de 2011.
- Stonebraker, M. Technical perspective - One size fits all: an idea whose time has come and gone. Commun. ACM 51(12): 76 (2008).