

Integrando dados Raster à Plataforma da Embrapa de dados de Recursos Naturais dos Biomas Brasileiros: avaliação do uso da extensão PostGIS Raster

Carla Geovana do N. Macário

Alan Massaru Nakai

Sergio Aparecido B. Cruz¹

Thiago de Siqueira Pereira²

¹Embrapa Informática Agropecuária - CNPTIA

Av. André Toselo, 209 - Caixa Postal 6041

13083-886 - Campinas - SP, Brasil

carla.macario, Alan.Nakai, Sergio.Cruz@embrapa.br

² Faculdade de Tecnologia - Unicamp

R. Paschoal Marmo, 1888 - Jd. Nova Itália

13484-332 - Limeira - SP, Brasil

thiago.pereira@colaborador.embrapa.br

Abstract. Natdata is a platform which aims to integrate data on natural resources from Brazilian Biomes. The platform is under development at Embrapa, the Brazilian Agricultural Research Corporation. Its main goal is the generation of a competitive intelligence for national agriculture, providing support for the actions of environmental zoning, characterization and management of biodiversity, conservation of soil and water. Moreover, it should also allow different users to access specialized services to facilitate query execution and the appropriate use of the integrated data. Some challenges faced during its development were: integration of heterogeneous (big) data, domain independence and data access and retrieval. Much of existing information on Natural Resources are available in raster format. One way to offer such data in the platform is provided by the use of the Raster PostGIS extension available on PostgreSQL. This work evaluates the use of this extension and presents some of the tools developed to maintain the raster database under NatData platform.

Keywords: base de dados geográficos, dados raster, geoprocessamento

1. Introdução

A viabilidade de expansão da área de produção agrícola e, principalmente, o aumento da produtividade tornaram a agricultura brasileira nas últimas décadas uma grande oportunidade nacional. Entretanto, seguindo a tendência mundial, a atividade agrícola sofre pressões internacionais generalizadas, referentes a questões como segurança alimentar, sustentabilidade e desmatamento, cada vez mais em evidência em função das preocupações mundiais com atitudes éticas, mudanças climáticas globais e preservação dos recursos naturais dos biomas brasileiros.

Neste contexto a geotecnologia aparece como disciplina essencial para prover ferramentas que permitam apoiar a tomada de decisão sustentável nos diversos setores e níveis governamentais, dotando a agricultura nacional de inteligência competitiva, fornecendo subsídios para o diagnóstico e ações de gestão, especialmente em áreas estratégicas de zoneamento ambiental, caracterização e gestão da biodiversidade, e da conservação de solo e da água. O Sistema de Monitoramento Agrometeorológico – Agritempo ¹ é um sistema que, a partir de dados geográficos históricos, permite modelagens e simulações para

¹<http://www.agritempo.gov.br/>

culturas considerando as possíveis mudanças globais que terão impacto na sua produtividade. Outras simulações importantes para a sustentabilidade são: recursos hídricos, considerando a disponibilidade e a qualidade de água, vegetação em áreas de proteção permanente, áreas de reserva legal e dinâmica de insetos. Em todos esses casos, assim como no sistema Agritempo, é necessário um conjunto de dados com informações sobre recursos naturais. Infelizmente, os dados existentes ainda são muito heterogêneos e provenientes de diferentes domínios, e muitas vezes de outras instituições, o que dificulta seu uso integrado. Além disso, muitas vezes leva a erros, inconstâncias, atrasos, dificuldade de acesso e imprevisibilidade na geração do conhecimento requerido.

Em 2009 a Embrapa, Empresa Brasileira de Pesquisa Agropecuária, publicou o relatório “Desafios para uma agricultura sustentável” Embrapa (2009), priorizando um inventário de recursos naturais, de maneira a permitir sua avaliação para aproveitamento sustentável em sistemas de produção de alimentos, fibras, essências e energia, e a adaptação de recursos naturais exógenos a esses sistemas e biomas. Portanto, pretendia direcionar esforços de gestão estratégica da Inovação, buscando reduzir erros, inconstâncias e imprevisibilidade na geração do conhecimento requerido, substituindo-os por diagnósticos, monitoramentos, prospecção antecipada de problemas, demandas e necessidades, orientação dos investimentos em conhecimento e organização dos fluxos de geração de tecnologias.

Como resposta à essa demanda, em 2010 foi proposto o projeto NatData (MACÁRIO et al., 2012) visando o desenvolvimento de uma plataforma de dados que buscasse prover a Embrapa de um ambiente tecnológico e integrado para a gestão da informação de recursos naturais dos biomas brasileiros, fornecendo subsídios para as ações de diagnóstico e gestão de forma mais precisa. Buscando atingir esse objetivo, no desenvolvimento da plataforma esbarrou-se em questões como heterogeneidade de formatos, dificuldade de integração de diferentes temas e qualidade dos dados. Atualmente boa parte da funcionalidade já está disponível para a Embrapa e em fase de validação, integrando e disponibilizando dados descritivos e vetoriais sobre Solo, Clima, Recursos Hídricos e Biodiversidade. Entretanto, muitos dados pertinentes ao projeto são disponibilizados em formatos raster: dados de clima, dados de projeção climática e imagens de satélite. Faz-se necessário ter meios eficientes de manipulação desses dados, permitindo seu uso conjunto com aqueles disponibilizados na plataforma.

Uma das maneiras de viabilizar a inclusão desses dados na plataforma foi a utilização da extensão PostGIS Raster do PostgreSQL, que possibilita armazenar e analisar os dados raster. Assim, torna-se possível a análise de consultas espaciais, a realização de cálculos como áreas, distâncias e centróides, além de realizar a geração de buffers (zona de influência) e outras operações entre as geometrias. Este trabalho apresenta a avaliação do uso da extensão PostGIS Raster do PostgreSQL na plataforma NatData, comentando o desenvolvimento de ferramentas que auxiliam a manutenção da base de dados raster da plataforma.

2. Metodologia de Trabalho

NatData - Plataforma de Informação de Recursos Naturais dos Biomas Brasileiros, quando proposta deveria funcionar como um sistema web, fornecendo meios para organizar e divulgar informações sobre os recursos naturais de biomas brasileiros, bem como permitir a integração contínua de outros bancos de dados. Para atingir este objetivo, foram consideradas questões como a heterogeneidade, diferente integração de domínio e qualidade dos dados. Além disso, deveria também preservar a semântica de dados, permitindo que os usuários acessem serviços especializados para consulta de dados, análise e usos futuros considerando outras ferramentas. Parente et al. (2013) apresentam uma plataforma semelhante comentando sua importância para a preservação do Bioma Cerrado.

A estratégia para o desenvolvimento do projeto começou com a estruturação dos dados a partir de domínios cuja informação estava mais organizada, como Solo e informações Climáticas. Numa segunda fase, foram trabalhados domínios cujas redes de pesquisa já existiam, como Recursos Hídricos e Biodiversidade, mas nos quais a informação ainda não estava bem organizada. Esperava-se que a experiência de integração resultasse em um conjunto de boas práticas para ser aplicada a outros bancos de dados a serem incorporados.

Como não havia na Embrapa um quadro preciso dos dados disponíveis nos biomas brasileiros, o desenvolvimento da plataforma começou com a aplicação de um questionário junto aos pesquisadores envolvidos no projeto. O questionário abordou questões como o tipo de dados, a autoria, a finalidade e o formato de armazenamento, necessidades de dados, seleção de metadados, interoperabilidade e os objetivos de integração (MACÁRIO et al., 2012). Ao final desse trabalho, com uma imagem real dos dados disponíveis na Embrapa, a equipe passou a modelagem de atividades com foco nas bases de dados individuais, mas também na sua integração.

O desenvolvimento da plataforma iniciou com o resgate/elaboração dos esquemas de dados, como uma forma de identificar os pontos de integração nos diferentes domínios sendo considerados (MACÁRIO et al., 2012). Considerando-se as informações comuns existentes nos esquemas, passou-se ao desenvolvimento de uma ferramenta de pesquisa com base em informação geográfica, considerando especialmente mapas. Nesse ponto também tiveram início discussões sobre padrões de metadados a serem adotados, com foco em questões de interoperabilidade. Um visão detalhada do desenvolvimento e considerações da plataforma pode ser obtida em (MACÁRIO et al., 2014).

A fim de abordar a questão de integração de dados, inicialmente considerou-se a abordagem de armazéns de dados, os chamados *data warehouse*. No entanto, considerando os resultados dos questionários e também entrevistas realizadas com especialistas, percebeu-se que essa não era a melhor opção. Assim, decidiu-se por um modelo de integração baseado em protocolos, serviços e padrões OGC, superando bancos de dados relacionais, permitindo diferentes usos dos dados e consultas e também facilitando a inclusão futura de outros bancos de dados, para expansão da plataforma. A Figura 1(a) dá uma visão geral da arquitetura da Plataforma NatData. A plataforma oferece ferramentas e serviços OGC para acessar os dados (camada de amarelo), que, por sua vez, utiliza outros serviços (camada laranja) que sabem como obter esses dados nas diferentes bases de dados.

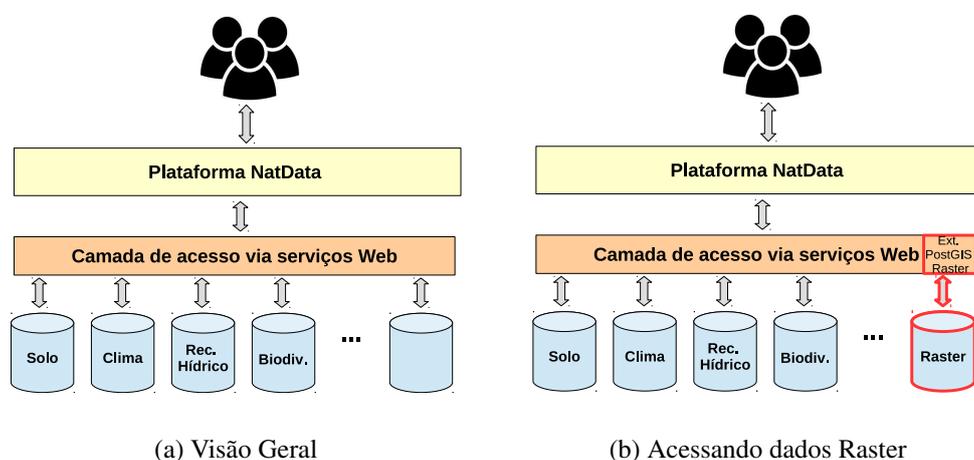


Figure 1: Arquitetura da Plataforma NatData

As bases de dados, por sua característica geográfica, são armazenadas no sistema gerenciador PostgreSQL, usando-se a extensão PostGIS para lidar com características espaciais. Essa extensão provê funcionalidades para estender a linguagem SQL para análise espacial e topológica, permitindo o uso integrado de dados descritivos tradicionais e dados vetoriais. Ao se usar a extensão PostGIS raster, pretende-se que a plataforma manipule também dados raster, permitindo a sua integração com os demais tipos de dados em servidos por ela (Figura 1(b)). Nossa abordagem baseia-se naquela apresentada por Vinhas e Ferreira (2013) com bons resultados.

A Figura 2 demonstra a diferença do uso da extensão PostGIS. No Postgis tradicional inserimos o valor um por um no banco de dados, já na extensão Raster podemos inserir o raster inteiro e dele extrair os valores.

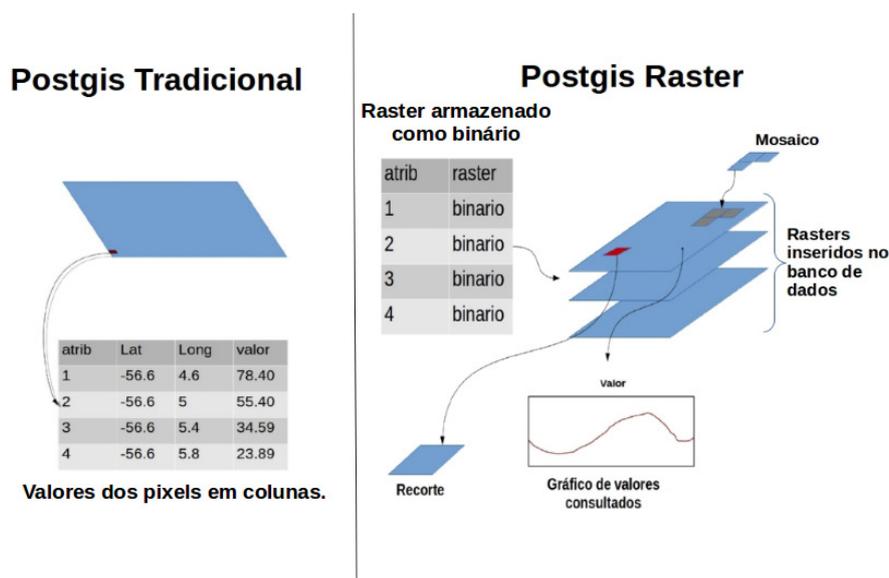


Figure 2: Diferença no uso da extensão PostGIS

Para permitir a manipulação dos dados Raster extensão Raster do PostgreSQL, foi utilizado um computador com processador Intel i7 e 8Gb de memória e sistema operacional Ubuntu 13.04. O banco de dados utilizado é o PostgreSQL junto com a extensão PostGIS 2.1.3 (OSGEO, 2014). Também utilizamos a linguagem de programação Java, com a IDE Eclipse, para criação de ferramentas para inserção e consulta dos dados. Além disso foram utilizados arquivos geotiff e netcdf de modelos de projeção climática para os testes.

3. Resultados e Discussão

A plataforma NatData abriga hoje dados de solos, clima, coleções biológicas, topologia, altimetria e recursos hídricos e permite a consulta a esses dados via um WebGIS ou por meio de palavras-chave. Essa consulta pode ser realizada em um ou mais dos domínios disponíveis. A integração das informações se dá pela sua localização geográfica, ou seja, serão recuperadas informações existentes para cada um dos domínios indicados considerando uma determinada localização (bioma, região geográfica, município, entre outros). Os resultados são disponibilizados na forma de mapas ou de arquivos (tabelas, arquivos textos, entre outros), segundo padrões geográficos definidos pela OGC. A Figura 3(a) ilustra o Menu Principal da NatData, considerando os domínios de dados disponíveis. Os usuários da plataforma têm disponíveis duas formas de realizar suas pesquisas: (i) usando palavras-chave, como em sistemas tradicionais; (ii) considerando características espaciais, com o uso de um mapa e

informando um polígono, regiões geográficas, biomas, ou mesmo uma determinada localidade coordenada. No caso de palavras-chave, o usuário entra em cada um dos domínios desejados e seleciona um ou mais filtros dentre aqueles disponíveis para o domínio. A Figura 3(b) ilustra uma pesquisa considerando palavras-chave para o domínio Clima.

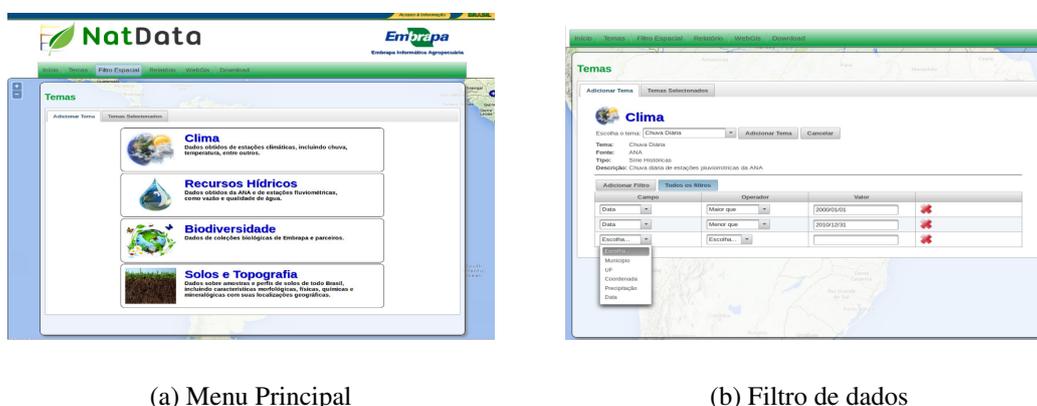


Figure 3: A Plataforma NatData

A consulta espacial é realizada usando um mapa como orientação, no qual o usuário pode selecionar uma área, um bioma, um estado ou mesmo indicar um polígono de interesse. Também estão disponíveis facilidades para união de áreas sobre as quais se deseja obter informações. A Figura 4 ilustra uma consulta espacial. Uma vez selecionada a área de interesse, e os domínios sobre os quais se deseja obter informação, a consulta é processada. O resultado da pesquisa é apresentado usando pontos coloridos, ou ícones, cada um indicando a informação existente para cada domínio em determinada área. O usuário poderá pesquisar as informações em detalhe.

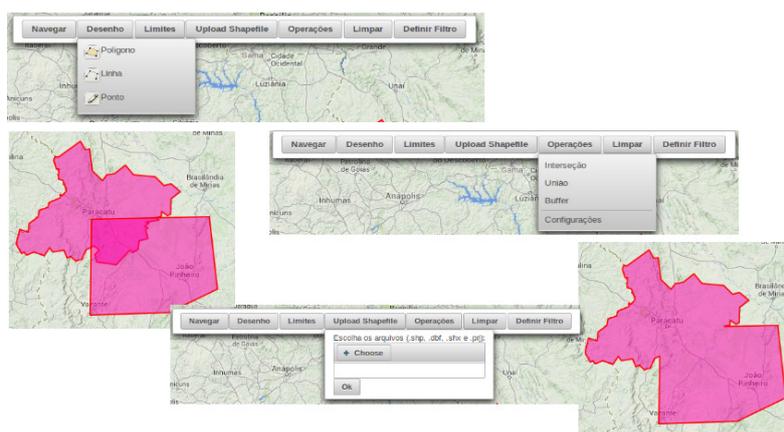


Figure 4: Consulta usando operadores espaciais

A Figura 5(a) apresenta o resultado para a área selecionada no exemplo anterior. Na Figura 5(b) o usuário acessa detalhes sobre as informações chuva existente para um ponto selecionado.

A execução de consultas integradas dos dados vetoriais e raster, permitiria, por exemplo, visualizar os dados de biodiversidade num mapa de solos ou de relevo. Com isso, seria facilmente possível uma identificação visual de algumas relações existentes. A inserção de dados raster na plataforma NatData é realizada por meio da ferramenta raster2pgsql que faz

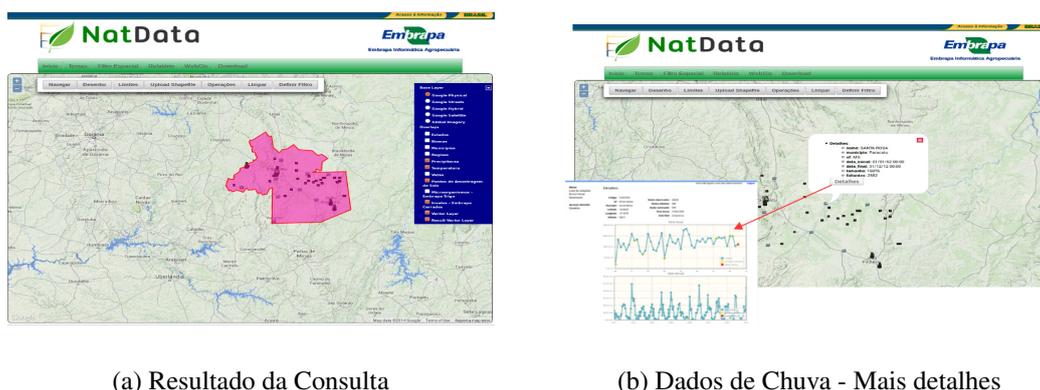


Figure 5: Resultado da Consulta Espacial

parte da extensão PostGIS. A linha de comando da Figura 6, mostra um exemplo de inserção de um arquivo geotiff.

```
$> raster2pgsql -c -s 4326 -I -F nomedoraster.tiff nome_schema.nome_da_tabela |
psql -d nome_do_banco
```

Figure 6: Código para Criação da Tabela Raster

Neste exemplo, a ferramenta é utilizada para gerar uma nova tabela (nome_schema.nome_da_tabela) a partir dos dados da imagem nomedoraster.tiff, utilizando a projeção EPSG 4326. A nova tabela é criada no banco de dados nome_do_banco.

No código fonte da Figura 7, apresentamos um exemplo de uma função Java que demonstra a utilização da API PostGIS Raster. Neste exemplo, a função recebe como parâmetros as coordenadas de um ponto da imagem e o código da projeção e retorna o valor do pixel. No SQL da linha 2, utilizou-se a função ST_Value() que retorna o valor de um pixel da coluna raster (rast) de acordo com sua coordenada geográfica. A função ST_GeomFromtext() é utilizada para gerar um objeto geométrico a partir das coordenadas textuais e de seu código de projeção. Este objeto geométrico representa o ponto que é utilizado na função ST_Value(). Na linha 3 utilizamos uma classe de conexão ao banco de dados (omitida para simplificar o código) para estabelecer a conexão com o banco de dados. As linhas 5 e 6 realizam a consulta ao banco, e da linha 7 a 11 são impressos os resultados desta consulta.

4. Conclusões

Este trabalho apresentou a plataforma Natdata, uma plataforma para integração de dados de recursos naturais dos Biomas Brasileiros. Considerando os resultados já obtidos, percebe-se a importância de uma plataforma que integre os diferentes tipos de dados num único local. Além disso, a facilidade de exportação dos dados em diferentes formatos, aumenta a usabilidade, permitindo sua adoção em diversos ambientes e setores decisórios que buscam a sustentabilidade.

Como os dados servidos pela plataforma são essencialmente vetoriais, fez-se necessária a adoção de mecanismos que permitissem também manipulação e integração de dados Raster. Neste trabalho foi avaliado o uso da extensão PostGIS Raster do ambiente PostgreSQL como opção de solução. O uso do PostGIS Raster possibilitará a criação de ferramentas para automatizar o armazenamento e as consultas aos dados raster do projeto NatData, proporcionando mais agilidade e confiabilidade do que a solução anterior, que consistia

```

2.     String sql = "SELECT rid, ST_Value(rast, ponto) As blpval FROM
nome_da_tabela, st_geomfromtext('POINT('+longitude+" "+latitude+"',
"+projecao+") As ponto ";
3.     Connection con = Conexao.getConnection();
4.     try {
5.         PreparedStatement preparador = con.prepareStatement(sql);
6.         ResultSet resultado = preparador.executeQuery();
7.         System.out.println("rid | blpval |");
8.         while (resultado.next()){
9.             System.out.print(resultado.getInt("rid"));
10.            System.out.println("|"+resultado.getDouble("blpval")
+"|");
11.        }
12.        preparador.close();
13.        con.close();
14.    }catch (SQLException el) {
15.        el.printStackTrace();
16.        System.out.println("erro");
17.    }

```

Figure 7: Exemplo de Uso da API PostGIS Raster

em armazenar os dados como arquivos e manipulá-los por meio de ferramentas externas à plataforma.

5. Agradecimentos

A equipe do projeto agradece ao Fundo Embrapa-Monsanto.

References

EMBRAPA. *Desafios para uma agricultura sustentável*. [S.l.]: Embrapa. Assessoria de Comunicação Social., 2009.

MACÁRIO, C. G. et al. Platform for natural resources of brazilian biomes: integration, systematization and dissemination of data and information for competitiveness and information for competitiveness and sustainability of agriculture. In: *International Conference on Ecological Informatics: Informing decisions on biodiversity and natural resources conservation*. Brasilia-DF: [s.n.], 2012. v. 6.

MACÁRIO, C. G. et al. Natdata: a platform to integrate geospatial data from natural resources of the brazilian biomes. In: *2014 IEEE 10th International Conference on eScience*. [S.l.: s.n.], 2014. p. 256–262.

OSGEO. *PostGIS 2.1 Manual*. <<http://postgis.net/stuff/postgis-2.1.pdf>>, 2014. Acesso em: Out/2014.

PARENTE, L. L. et al. Uma plataforma de pesquisa para o cerrado: Disseminando e promovendo dados e conhecimento científico sobre um bioma ameaçado. In: *XVI Simpósio Brasileiro de Sensoriamento Remoto - SBSR*. [S.l.: s.n.], 2013.

VINHAS, L.; FERREIRA, K. R. Explorando a extensão wkt raster do postgis para armazenamento e manipulação de imagens de sensoriamento remoto. In: *XVI Simpósio Brasileiro de Sensoriamento Remoto - SBSR*. [S.l.: s.n.], 2013.