

## **Plataforma Natdata – usando dados de Recursos Naturais do Bioma Pantanal**

Carla Geovana do Nascimento Macário<sup>1</sup>

Alan Massaru Nakai<sup>1</sup>

José Ruy Porto de Carvalho<sup>1</sup>

<sup>1</sup>Embrapa Informática Agropecuária

Av. André Tosello, 209 - Caixa Postal 6041

13083-886 - Campinas - SP, Brasil

{carla.macario, Alan. nakai, jose.ruy}@embrapa.br

**Resumo:** Este trabalho apresenta um exemplo de uso dos dados disponíveis na plataforma Natdata, uma plataforma para integração de informação de Recursos Naturais dos Biomas Brasileiros. O artigo comenta o processo adotado no seu desenvolvimento e um exemplo de uso dos dados do Bioma Pantanal, correlacionando as informações existentes. Com isso, mostramos a importância da plataforma, auxiliando na tomada de decisão considerando a preservação e o uso sustentável dos recursos disponíveis no território brasileiro, em especial no Bioma Pantanal, rico em diversidade.

**Palavras-chave:** banco de dados geográficos, integração de dados, informação de recursos naturais, interoperabilidade.

**Abstract.** This paper presents Natdata, a platform for integration of Information on Natural Resources from Brazilian Biomes. The article comments the process adopted for its development, its features. It also presents an example of use, correlating existing information from Pantanal Biome. With this, we show the importance of the platform, which aims to support decision making regarding the preservation and sustainable use of available resources in Brazil, especially in the Pantanal Biome, with a rich diversity.

**Keywords:** geographic database, database integration, natural resource information, interoperability.

## 1. Introdução

Vivemos num mundo que busca cada vez mais a sustentabilidade, com a preservação dos recursos naturais disponíveis frente às mudanças que o planeta enfrenta. Para que essa sustentabilidade ocorra é necessário que os atuais modelos de produção levem em conta questões relacionadas a diferentes temas: manutenção de áreas ambientais, fauna, flora, mananciais, solo, entre outros. A gestão de recursos naturais envolve diagnósticos em diversas escalas (local, regional e nacional) e os resultados promovem o suporte para os tomadores de decisões.

Tais diagnósticos requerem disponibilidade de informação integrada, o que nos dá condição de fazer uma gestão muito mais eficiente dos recursos naturais do País, permitindo outras simulações e modelagens considerando, por exemplo, a disponibilidade e qualidade de água, manutenção de vegetação em áreas de proteção permanente (APPs) e de áreas de reserva legal (RL), dinâmica de insetos, entre outros. No entanto, os dados de recursos naturais são dados muito heterogêneos e assim, o acesso integrado a eles não é uma atividade simples.

Diversos fatores contribuem para esta indisponibilidade: (i) por serem produzidos por entidades diferentes, estes dados geralmente encontram-se distribuídos geograficamente, normalmente armazenados nas instituições que os produziram; (ii) alguns destes dados ainda não estão organizados em banco de dados, encontrando-se em muitos casos armazenados sob a forma de planilhas, textos e até documentos em papel; (iii) cada instituição adota formato específico, o que dificulta o seu uso em conjunto.

Sob a ótica do uso combinado destes dados, nos deparamos com mais algumas dificuldades, sendo duas delas principais: a manipulação de grandes volumes de dados e a sua heterogeneidade. A primeira delas vem sendo colocada como um dos grandes desafios da ciência moderna e a segunda delas se manifesta de diferentes formas: na representação da informação; no modelo espacial usado para representar o dado; no esquema de dados e no significado semântico do dado. Vencer estas dificuldades tem sido um grande desafio de pesquisa há pelo menos 20 anos.

Em 2009 a Embrapa, Empresa Brasileira de Pesquisa Agropecuária, publicou o relatório “Desafios para uma agricultura sustentável” (Embrapa, 2009), indicando a priorização do inventário dos 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.

Frente a isso a Embrapa propôs e aprovou o projeto NatData -- Plataforma de Recursos Naturais dos Biomas Brasileiros: Integração, Sistematização e Disseminação de Dados e Informações para Sustentabilidade e Competitividade da Agricultura (Macario, 2010). O projeto pretendia o desenvolvimento de uma plataforma que integrasse grandes volumes de informações heterogêneas sobre os recursos naturais dos biomas brasileiros e o foco era a geração de inteligência competitiva para a agricultura nacional, 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. Atualmente boa parte da funcionalidade já está disponível, em fase final de validação, disponibilizando informações diversas sobre os recursos naturais dos Biomas Brasileiros. Dentre os desafios a serem enfrentados estavam a integração de dados heterogêneos, independência de domínios e formas alternativas de acesso e recuperação da informação

## 2. Objetivo

Este trabalho apresenta o Natdata, sua metodologia de desenvolvimento, os resultados obtidos, apresentando um caso de uso considerando dados Bioma Pantanal.

## 3. Material e Métodos

Natdata - Plataforma de Informação de Recursos Naturais dos Biomas Brasileiros é uma plataforma para integração de dados sobre os recursos naturais dos biomas brasileiros. O público-alvo da plataforma é formado por todos aqueles que de alguma forma buscam essa informação, englobando desde gestores de empresas que focam na sustentabilidade do agronegócio brasileiro, até pesquisadores e estudantes que podem usar a informação no desenvolvimento de sua pesquisa.

Os dados fornecidos pela plataforma representam domínios diferentes e isso leva à heterogeneidade em muitos aspectos: linguagem, formato, espacialidade, semântica e uso. Esta heterogeneidade geralmente dificulta a integração de dados. No entanto, dados de recursos naturais têm um ponto em comum: seu atributo geoespacial. O termo geoespacial refere-se a todos os tipos de dados – textuais ou não – sobre objetos e fenômenos do mundo que são associados a características espaciais e referenciam alguma localidade na superfície na Terra. A informação deste tipo de dado é armazenada numa estrutura denominada *espacial*. Um Sistema de Informação Geográfica (SIG) é um sistema que gerencia e recupera informações deste tipo de dado especial (Pires et al., 1993). Estes sistemas são muito utilizados em ambientes de decisão, provendo aos usuários facilidades de combinar as informações de uma determinada região, sejam elas espaciais ou não.

Exemplos desse tipo de dado incluem informação sobre recursos naturais como solos, clima e vegetação, essencial para responder a questões como “Como intensificar a produção de cana-de-açúcar na região de Piracicaba, com a manutenção dos biomas envolvidos?” ou “Qual o padrão climático esperado para uma região de determinado tipo de solo no bioma Cerrado?”. Estas questões são básicas para o estabelecimento de políticas públicas de intensificação agrícola com a conservação ambiental.

Respostas a essas questões envolvem o uso combinado de dados provenientes de diferentes domínios. O acesso integrado às diferentes informações existentes garante agilidade e qualidade na resposta gerada. Esse acesso requer um entendimento único do dado sendo usado/trocado. Ou seja, é necessário que se garanta o significado semântico comum do dado. A garantia da interoperabilidade semântica foi apontada por Sciore et. al. (1994) como sendo a chave para o sucesso na integração de dados, motivadora de várias pesquisas em Banco de Dados, uma das áreas da Tecnologia de Informação. Entretanto, ainda hoje essa integração se mantém como um desafio de pesquisa.

A plataforma NatData deve fornecer todos os dados necessários por meio de um sistema único, usando um formato comum, tornando mais fácil para os usuários a buscá-los. Além disso, a maioria dos sistemas disponíveis oferecem apenas instrumentos de pesquisa tradicionais com base em palavras-chave e melhoradas com filtros. No contexto dos recursos naturais,

descrever os dados desejados usando palavras-chave não é suficiente porque muitos usuários procuram informações considerando questões diferentes. Foi então necessário oferecer ferramentas de pesquisa para recuperar informações considerando recursos geoespaciais, como um bioma, uma região geográfica ou até um determinado local.

#### 4. Metodologia

A plataforma NatData plataforma 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, consideramos questões como a heterogeneidade, diferente integração de domínio e qualidade dos dados.

A plataforma 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. 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 as informações Climáticas. Numa segunda fase, trabalhamos nos 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 da antiga resultaria em um conjunto de boas práticas para ser aplicada a esta última os domínios. 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 (Macario et al., 2012a).

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 a construção de ontologias de domínio, como uma forma de identificar os pontos de integração nos diferentes domínios sendo considerados (Macario et al., 2012b). Considerando-se as informações comuns existentes nas ontologias propostas, passou-se ao desenvolvimento de uma ferramenta de pesquisa com base em informação geográfica, considerando especialmente mapas. Neste ponto também teve início discussões sobre padrões de metadados a serem adotados, com foco em questões de interoperabilidade.

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** dá uma visão geral 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.

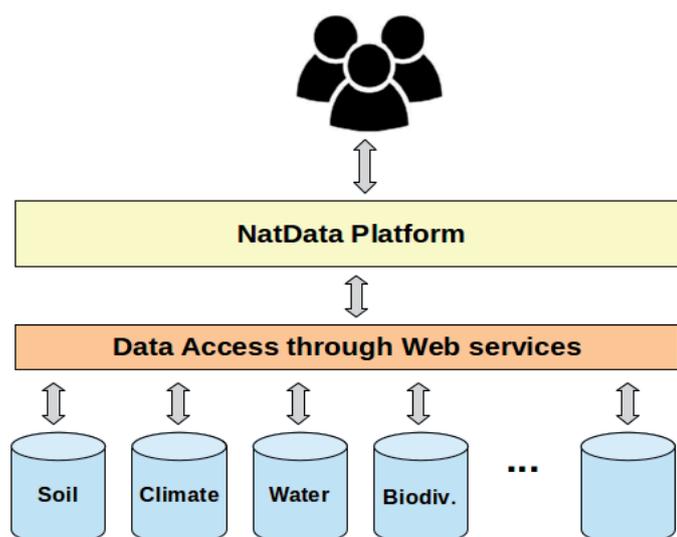


Figura 1. Visão Geral da Estrutura da Plataforma NatData

Nessa abordagem o ponto de integração dos dados da Plataforma em sua maioria é a informação geográfica, uma vez que refere-se a uma localização única na superfície da Terra. Mesmo assim, isso não é suficiente: o usuário precisa obter informações sobre como os dados foram produzidos, por quem e sob quais critérios de qualidade. Assim, é essencial, também armazenar dados de descrição, o que permite a sua compreensão e utilização.

A plataforma NatData foi desenvolvida de maneira a ser independente dos produtores de dados. Isso significa que cada domínio de dados disponibilizados por ela deve ter seu próprio sistema de banco de dados. No início do projeto, já havia dois sistemas: o de solos (DBSolos<sup>1</sup>) e o de clima (Agrimtempo<sup>2</sup>). Por isso foi necessário desenvolver novos sistemas para os domínios de biodiversidade e de recursos hídricos. Entretanto, mesmo para os sistemas prontos houve necessidade de evoluções: o Agrimtempo não foi projetado para exportar seus dados via serviços Web e os dados de solos precisaram ser georreferenciados. Estes sistemas já estão prontos, disponíveis para o público e exportando seus dados adequadamente para a plataforma NatData.

Os sistemas de domínio da água e biodiversidade ainda estão em desenvolvimento. Para a biodiversidade, está sendo adotado a *Biodiversity Data Digitizer* (BDD<sup>3</sup>), um sistema desenvolvido pela Universidade de São Paulo (USP) para a manipulação e publicação de informações sobre a biodiversidade. Para melhorar e manter a qualidade dos dados armazenados, o BDD auxilia os usuários com sugestões de nomes científicos de taxonomias que já estão em bancos de dados oficiais, como o Catálogo da Vida (*Catalogue of Life*). Ele também fornece um ambiente de colaboração, promovendo a interação entre a pesquisa de diversas áreas relacionadas à biodiversidade. Fornecer um sistema para armazenar dados de espécies diferentes não é uma tarefa fácil, porque cada sistema tem características especiais. Decidiu-se considerar apenas as mais importantes, simplificando a tarefa de catalogação de dados. Os recursos fornecidos pelo BDD ajudaram neste trabalho.

No domínio da água ainda não há um consenso sobre o esquema de dados para o banco de dados. Por isso, estamos motivando especialistas a usarem suas planilhas como um primeiro passo para projetar um esquema comum, considerando outros disponíveis em outras instituições

<sup>1</sup> <http://www.sisolos.cnptia.embrapa.br/>

<sup>2</sup> <http://www.agrimtempo.gov.br>

<sup>3</sup> <http://www.agrimtempo.gov.br>

de pesquisa.

## 5. Resultados Parciais 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 da informação 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 2** mostra a tela inicial do sistema, considerando os domínios de dados disponíveis.



**Figura 2.** Tela inicial da Plataforma NatData

Os usuários da plataforma tem 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-chaves, o usuário entra em cada um dos domínios desejados e seleciona um ou mais filtros dentre aqueles disponíveis para o domínio. A **Figura 3** ilustra uma pesquisa considerando palavras-chave para o domínio Clima.

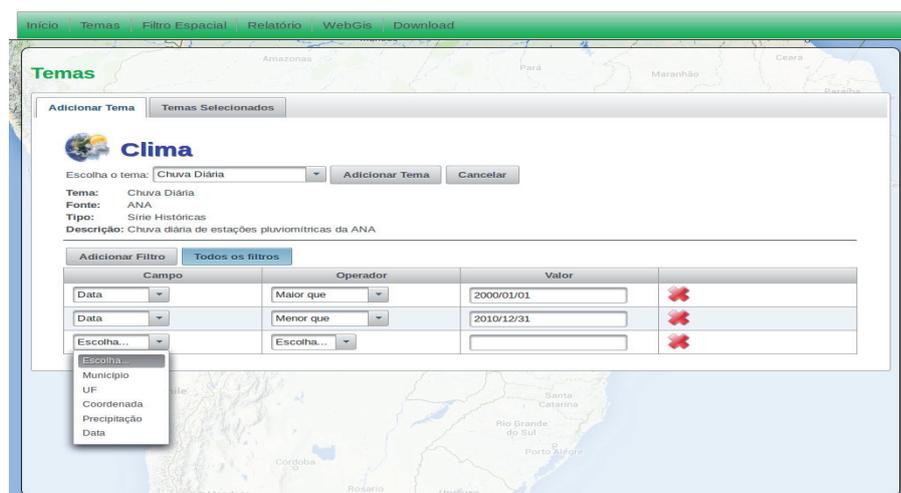


Figura 3. Consulta baseada em palavras-chave

A Figura 4 ilustra uma consulta espacial. Nela, usando um mapa como orientação, 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.

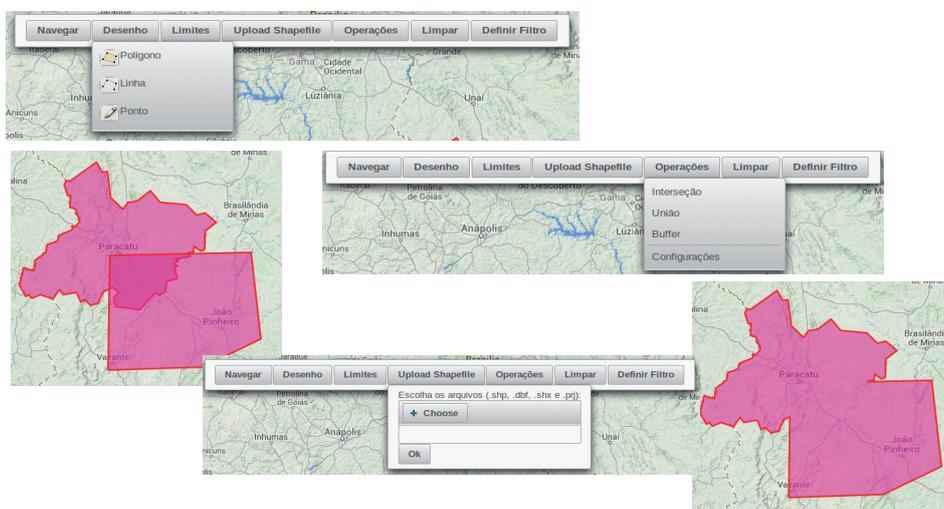


Figura 4. Consulta espacial

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. A Figura 5 ilustra o resultado para a área informada.

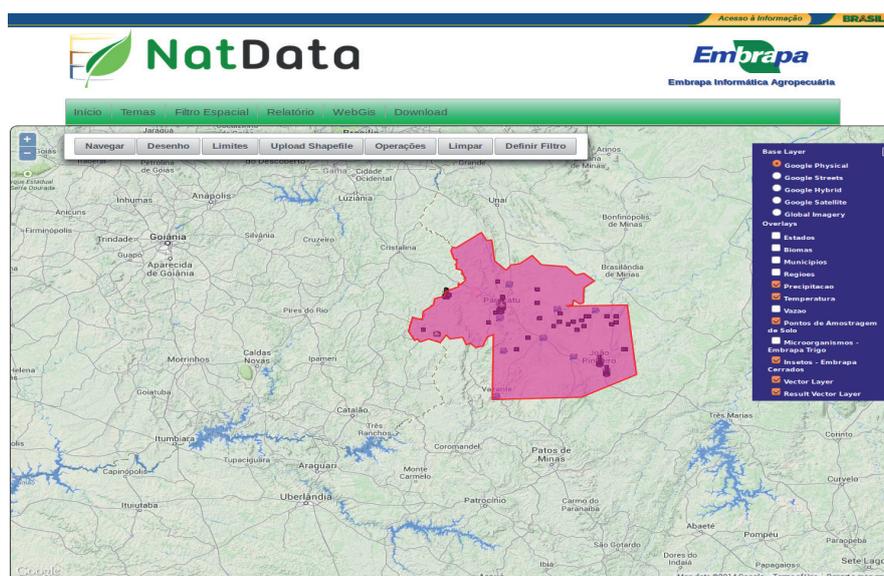


Figura 5. Pontos de informação recuperados para uma dada área de interesse

O usuário poderá pesquisar as informações em detalhe. Na Figura 6 o usuário acessa informações sobre insetos e Figura 7 apresenta as informações chuva existente para um ponto selecionado.

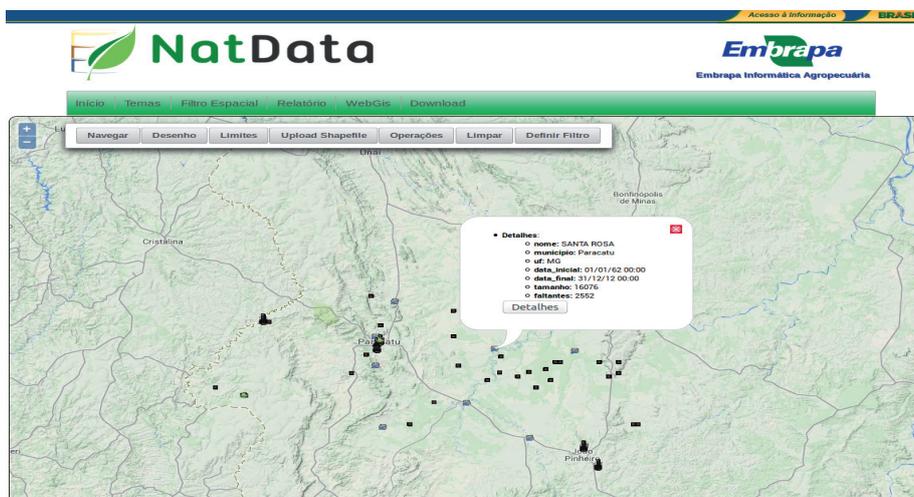


Figura 6. Detalhes de informação de um ponto recuperado - metadados

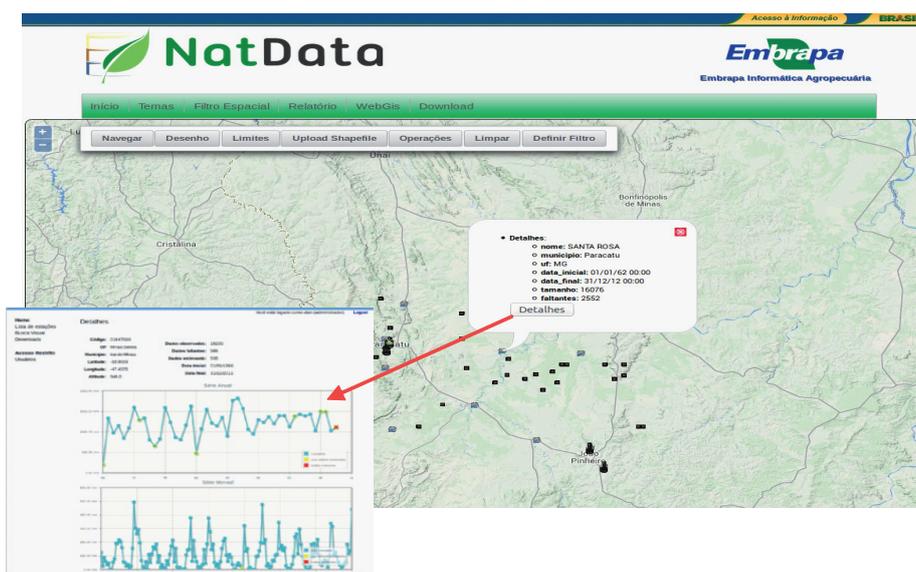


Figura 7. Detalhes de informação de um ponto recuperado na forma de um gráfico

#### 4. Dados sobre o Bioma Pantanal

Em relação ao Bioma Pantanal a plataforma abriga dados de Herbários, Peixes e Clima. Estes dados ainda estão sendo trabalhados para sua apresentação em mapas e para seu acesso via consulta espacial. O projeto NatData viabilizou ações de revisão de qualidade dos dados, devendo estar disponíveis para acesso em breve.

Tendo todas essas informações, o usuário poderá pesquisar as informações em detalhe para o Bioma Pantanal, por exemplo correlacionando dados de clima e de biodiversidade. Com isso, é possível acompanhar a evolução de ocorrência de uma espécie relacionando-a às mudanças e variações de clima.

Além disso, as informações podem ser exportadas em vários formatos – mapas, *shapefiles*, planilhas, texto – para uso em outros sistemas. O acesso facilitado aos dados integrados, permite inúmeras variações e usos, contribuindo para a gestão dos recursos naturais disponíveis nos Biomas Brasileiros.

#### 4. Conclusão

Este trabalho apresentou o Natdata, uma plataforma para integração de dados de recursos naturais dos Biomas Brasileiros. Dentre outras coisas, com a plataforma pretende-se a geração de inteligência competitiva para a agricultura nacional, fornecendo subsídios para as ações de diagnóstico e gestão de forma mais precisa, especialmente nas áreas estratégicas de zoneamento ambiental, caracterização e manejo da biodiversidade, conservação do solo e da água.

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.

Trabalhos futuros envolvem melhoria na informação geográfica e a disponibilização de um conjunto de serviços que permitam a execução de análise e simulações, baseados em *workflows* científicos.

## 5. Agradecimentos

A equipe do projeto agradece ao Fundo Embrapa-Monsanto.

## 6. Bibliografia

Embrapa. **Desafios para uma agricultura sustentável**. Embrapa, Assessoria de Comunicação Social. – Brasília, DF : Embrapa - 79 p. 2009

Macário, C. G. N, et al. **Plataforma de Recursos Naturais dos Biomas Brasileiros: Integração, Sistematização e Disseminação de Dados e Informações para Sustentabilidade e Competitividade da Agricultura**, Projeto Embrapa código 02.10.04.002.00.00, 2010.

Macário, C. G. N.; Fonseca, C. E. L.; Pierozzi Junior, I.; Torres, T. Z. **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, 8., 2012, Brasília, DF. Informing decisions on biodiversity and natural resources conservation: book of abstracts. Brasília, DF: UnB, 2012. 1 CD-ROM.

Macário, C. G. Do N.; Speranza, E. A.; Silva, J. dos S. V. da. **Natdata - plataforma de recursos naturais dos biomas brasileiros: informações geoespaciais para sustentabilidade na agricultura**. In: SIMPÓSIO DE GEOTECNOLOGIAS NO PANTANAL, 4., 2012, Bonito, MS. Anais... Brasília, DF: Embrapa, 2012. p. 859-869. 1 CD-ROM. Geopantanal 2012. Sciore, E.; Siegel, M.; Rosenthal, A. Using semantic values to facilitate interoperability among heterogeneous information systems. *ACM Trans. Database Syst.*, 19(2):254–290. 1994.

Pires, M. F.; Medeiros, C. M. B.; Silva, A. B. **Modelling geographic information systems using an object oriented framework**. Technical Report DCC-93-13, DCC/Unicamp, Campinas, Brazil. 1993.