

NATDATA: Integrando Dados de Recursos Naturais dos Biomas Brasileiros

Carla Geovana do Nascimento Macário¹

Eduardo Delgado Assad²

Ivo Pierozzi Jr.³

Sílvio Roberto Medeiros Evangelista⁴

RESUMO: A agricultura nacional exige a intensificação das áreas plantadas aliada à manutenção dos recursos naturais dos biomas brasileiros. Respostas rápidas a questões envolvendo temas como solo, recursos hídricos, biodiversidade e clima nesse caso são essenciais. O Brasil dispõe de um grande acervo de dados sobre estes temas, distribuídos em várias instituições de pesquisa. A heterogeneidade de padrões aliada a essa distribuição dificulta o seu uso combinado. Este trabalho apresenta uma iniciativa que vem sendo desenvolvida pela Empresa Brasileira de Pesquisa Agropecuária - Embrapa, que tem como principal objetivo integrar dados de recursos naturais dos diferentes biomas brasileiros, fornecendo aos usuários um ambiente que permita a consulta rápida e integrada a estes dados.

PALAVRAS-CHAVE: banco de dados geoespaciais, integração de dados geoespaciais, dados de recursos naturais.

NATDATA: Integrating Data of Natural Resources from the Brazilian Biomas

ABSTRACT: The Brazilian agriculture requires the intensification of cultivated areas combined with the maintenance of natural resources from the Brazilian biomes. Rapid answers to questions considering domains such as soil, hydrological resources, biodiversity and climate are very important in this case. Brazil owns a huge data collection considering these themes, distributed in different institutions. This fact, combined to the heterogeneity of standards, makes the use of these data a difficult issue. This work presents an initiative which has been developed under the Brazilian Agricultural Corporation - Embrapa, aiming at the integration of natural resources data, considering the different Brazilian biomes. Doing this we expect to provide an environment that allows our users to get the needed data in a fast and reliable way.

KEYWORDS: geospatial database, geospatial data integration, natural resource data.

1. INTRODUÇÃO E MOTIVAÇÃO

A agricultura moderna é uma atividade que busca o equilíbrio em duas linhas de conduta: a econômica e a ecológica. O Brasil segue essa tendência: a viabilidade de expansão da área de produção e, principalmente, o aumento da produtividade tornaram o setor agrícola brasileiro uma grande oportunidade nacional. Por outro lado a atividade agrícola sofre pressões internacionais generalizadas, referentes a questões como segurança alimentar, sustentabilidade e desmatamento, cada vez mais em evidência face às preocupações mundiais com atitudes éticas e com mudanças climáticas globais.

Na década de 90 surgiu a chamada “agricultura sustentável” como resposta a isso, buscando a construção de um novo modelo de produção com foco no aumento de produtividade, no

¹ Pesquisador A, Embrapa Informática Agropecuária, carla@cnptia.embrapa.br

² Pesquisador A, Embrapa Informática Agropecuária; MCT, assad@cnptia.embrapa.br

³ Pesquisador A, Embrapa Informática Agropecuária, ivo@cnptia.embrapa.br

⁴ Analista A, Embrapa Informática Agropecuária, silvio@cnptia.embrapa.br

melhor uso da terra e na conservação das culturas tradicionais que deram origem e mantêm estes materiais genéticos [10].

Em 2009 a Embrapa publicou o relatório “Desafios para uma agricultura sustentável”[2] priorizando o inventário dos recursos naturais para permitir o aproveitamento sustentável em sistemas de produção de alimentos, fibras, essências e energia, e a adaptação de recursos naturais exógenos a esses sistemas e biomas. Assim vem direcionando esforços de gestão estratégica da Inovação, buscando diagnósticos, monitoramentos, prospecção antecipada de problemas, demandas e necessidades, orientação dos investimentos em conhecimento e organização dos fluxos de geração de tecnologias.

Diagnósticos precisos requerem o acesso a informação atualizada sobre os recursos naturais do país. Durante as últimas três décadas, diversas instituições brasileiras retratando o meio ambiente produziram uma grande quantidade de dados e informação brasileiro, tanto na forma textual (textos e tabelas) quanto na forma não textual (mapas, fotos, imagens). No entanto, essa informação nem sempre está disponível ou chega de forma rápida e simples ao usuário final.

Diversos fatores contribuem para a indisponibilidade desses dados: (i) por serem produzidos por entidades diferentes, eles geralmente encontram-se distribuídos geograficamente, armazenados nas instituições que os produziram; (ii) alguns deles ainda não encontram-se organizados em banco de dados, estando 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 desses dados, nos deparamos com mais alguns problemas. O maior deles é a chamada heterogeneidade, que 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 na área de banco de dados há pelo menos 20 anos.

A Embrapa, por meio do projeto NATDATA, busca organizar e integrar informações geoespaciais de recursos naturais dos biomas brasileiros, visando 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 a água. O projeto reúne em torno de 150 pesquisadores das mais diversas ciências, além de estabelecer parcerias com universidades e com outros órgãos de pesquisa. Este trabalho descreve o projeto, apresentando alguns resultados já obtidos e os próximos passos a serem desenvolvidos.

2. NATDATA - Visão Geral

Na Embrapa os dados geoespaciais encontram-se em diferentes estágios de organização: enquanto existem sistemas bem sedimentados, como o Agritempo⁵, que gerencia dados de clima, e o Banco de Dados de Solos⁶, que organiza dados de solos do Brasil, temas como coleções biológicas ainda são tratados individualmente. O projeto NATDATA, sigla para Plataforma de Integração de Dados dos Recursos Naturais, tem duração prevista de 3 anos e visa a organização e integração de dados geoespaciais sobre recursos naturais dos biomas brasileiros.

⁵<http://www.agritempo.gov.br/>

⁶<http://www.bdsolos.cnptia.embrapa.br/>

Na proposição do projeto verificou-se a inexistência de um retrato exato da situação desses dados na empresa. Vários fatores contribuem para isso e questões de pesquisa no projeto envolvem a definição do estado da arte destes dados na empresa, sua modelagem e proposta de integração, apoiada por ontologias de domínios e até a proposta e a adoção de um modelo de integração dos diferentes tipos de dados. Ao final do desenvolvimento do projeto pretende-se a implantação de uma plataforma que integre os dados de recursos naturais da Embrapa, ainda muito heterogêneos e provenientes de diferentes domínios, e muitas vezes de outras informações, garantindo a manutenção de sua semântica, e permitindo aos diferentes usuários do sistema o acesso a serviços especializados que viabilizem a consulta e uso adequado desses dados.

Como benefício indireto resultante destas ações de pesquisa, pretende-se o estabelecimento de um conjunto de recomendações que direcionem a integração de dados e a gestão da informação existente na Embrapa sobre os recursos naturais dos biomas brasileiros, facilitando a incorporação de novos bancos de dados temáticos a esta plataforma. Estas recomendações devem incluir questões de propriedade intelectual, interoperabilidade de dados, padrões geográficos e infraestrutura tecnológica, entre outros.

3. EXPERIMENTO INICIAL

O projeto NATDATA vem sendo desenvolvido de maneira incremental. A estratégia adotada para auxiliar na identificação de requisitos para a plataforma de dados foi a realização de um experimento inicial. Nesta etapa buscou-se reunir dados gerados/utilizados pelos projetos Zoneamento da Cana-de-Açúcar e Zoneamento de Riscos Climáticos, já que ambos projetos usam informações de solos, hidrologia, usos da terra e áreas de proteção ambiental, entre outros. Um dos objetivos neste experimento foi permitir a visualização combinada dos diferentes dados envolvidos. Portanto, utilizaram-se as ferramentas HiperEditor e HiperNavegador⁷ desenvolvidos na empresa e que implementam a técnica de visualização de foco+contexto conhecida como árvore hiperbólica.

Para facilitar a inclusão de novos dados, as árvores foram geradas utilizando-se um método semi-automático baseado em *scripts* na linguagem Java e que permite a geração de uma cadeia hierárquica pela nomenclatura dos arquivos de dados. Essa cadeia constitui árvore a ser apresentada ao usuário, onde cada um de seus nós está associado a um dos arquivos de dados. Por exemplo, considerando-se o arquivo *Solos_SP_1.zip*, que contém dados de perfis de solos no estado de São Paulo, é criada uma raiz *Solos* e um nó *SP*. “*1*” simboliza o número de índices dos arquivos daquele nó. No último filho, são criados dois ou mais nós de acordo com o número de arquivos que possuem alguma ligação para esses nós.

Para efeito de organização e visualização dos dados, foi criada uma árvore para cada tema considerado e todas foram associadas a uma árvore maior, identificada como *Recursos Naturais*. A árvore geral foi gerada manualmente, utilizando-se o editor de árvores hiperbólicas. A figura 1 ilustra parte deste experimento, apresentando a árvore geral, que agrega os diversos temas considerados, a árvore para o tema *Solo* e um mapa associado a um nó dessa árvore.

Este experimento permitiu disponibilizar de modo simples uma boa quantidade de dados aos membros e colaboradores envolvidos nos projetos citados, facilitando a consulta aos dados. Fazendo uso deste método automático é possível ainda adicionar ou remover arquivos sempre que necessário, facilitando a manipulação de grandes volumes de dados. Estes resultados devem facilitar a identificação da situação dos dados de recursos naturais da empresa.

⁷<http://repositorio.agrolivre.gov.br/>

4. TRABALHOS RELACIONADOS

O problema de integração de dados espaciais é antigo abordando várias questões. Um dos grandes pontos de pesquisa em bases de dados geográficos é o desafio da interoperabilidade, isto é, a capacidade de acessar remotamente os dados e processá-los em um ambiente aberto, preservando a semântica de cada dado em um dado contexto [14]. Sciore et. al. [12] aponta a garantia da interoperabilidade semântica como sendo a chave para o sucesso na integração de dados, motivadora de várias pesquisas em Banco de Dados.

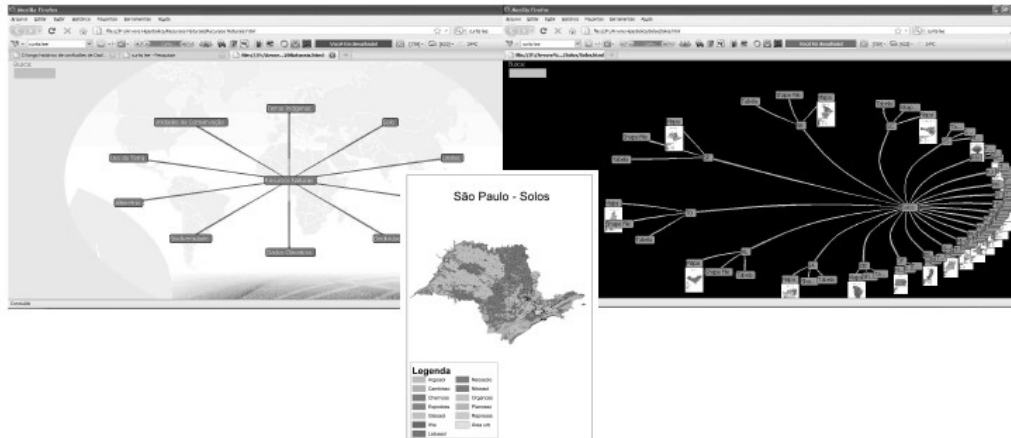


Figura 1. Árvore de Recursos Naturais

Ao longo dos anos foram propostas diferentes soluções para o problema da integração de dados tradicionais. Em paralelo surgiram abordagens que estendem estas soluções para dados espaciais, como forma de prover o compartilhamento efetivo de dados geográficos. Estas abordagens oferecem um ambiente que permite que dois ou mais sistemas heterogêneos interoperem, trocando dados e serviços independentemente de seu formato e estrutura, do esquema definido, das aplicações e da plataforma. Embora existam diferentes propostas para integração deste tipo de informação, a maioria prevê a integração de informação proveniente de um mesmo domínio [13, 14]. Trabalhos mais recentes indicam que o foco da pesquisa se manteve, concentrando-se na proposta de modelos que facilitem a integração [5, 1]. Curiosamente, o uso de ontologias aparece como uma solução metodológica que se baseia em abordagem conceitual e tecnológica e que tem sido desenvolvida e aplicada quando se busca integração de elementos num sistema de alta complexidade [4, 7, 6, 15].

Interoperabilidade de dados espaciais requer o manuseio apropriado dos diferentes tipos de dados. A crescente busca por informação espacial levou ao desenvolvimento de iniciativas para obtenção de metadados específicos, considerando os diferentes formatos, comunidades e agência de dados. Como resultado, tem-se padrões bem estabelecidos para dados geográficos, como o padrão ISO 19115 [8] e o padrão FGDC [3] e têm por objetivo prover um conjunto de definições e terminologias comuns para a documentação e troca de dados. Estas documentações são geralmente publicadas em catálogos de dados espaciais, permitindo que usuários descubram e recuperem os dados [11]. Metadados, e servidores de metadados têm papel crucial na integração de dados, permitindo que usuários integrem dados de múltiplas fontes, organizações e formatos. Em especial, metadados espaciais (ou geográficos) devem incluir informação da fonte de dados (propriedade), data de criação, formato, projeção, escala, resolução e acurácia [9].

5. CONCLUSÕES E TRABALHOS FUTUROS

A agricultura brasileira caminha na busca o aproveitamento sustentável dos recursos naturais presentes nos biomas do país. Neste sentido o acesso rápido e eficiente a informação livre de erros é essencial. A Embrapa, assim como outras instituições, dispõe de um grande volume de informações resultantes das pesquisas que realiza. Entretanto esses dados encontram-se em estágios de organização diferentes, dificultando o seu acesso, bem como sua disponibilização aos consumidores deste dado.

Este trabalho descreveu o projeto NATDATA, que visa a organização e disponibilização de dados geoespaciais de recursos naturais numa plataforma única e integrada. Um experimento realizado inicialmente no desenvolvimento do projeto foi apresentado, considerando os resultados obtidos. Atualmente está sendo aplicado um questionário junto aos pesquisadores produtores de dados visando a obtenção de um retrato do dado sendo armazenado e os formatos de armazenamento adotados. Com isso espera-se ter uma visão clara das possibilidades de integração e atividades a serem desenvolvidas, servindo como guia no processo de integração dos dados.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Agustina Buccella and Alejandra Cechich. Towards integration of geographic information systems. *Electr.. Notes Theor. Comput. Sci.*, 168:45–59, 2007.
- [2] Embrapa. *Desafios para uma agricultura sustentável..*, Brasília, 2009. Embrapa, ACS.
- [3] FGDC. *FGDC Geospatial Metadata*, 1998.
- [4] F. Fonseca, C. Davis, and G. C^ amara. Bridging ontologies and conceptual schema in geographical information integration. *Geoinformatica*, 7(4):307–321, 2003.
- [5] L. Gomez, S. Haesevoets, B. Kuijpers, and A. A. Vaisman. Spatial aggregation: data model and implementation. *Information Systems*, 34:551–576, 2009.
- [6] F. Hakimpour. Using ontologies to resolve semantic heterogeneity for integrating spatial data base schemata. PhD thesis, Zurich University, Switzerland, 2003.
- [7] F. Hakimpour and A. Geppert. Global schema generation using formal ontologies. In Springer, editor, *ER'02:21st International Conference on Conceptual Modeling.*, v.2503 of *Lecture Notes in Computer Science*, p. 307–321, 2002.
- [8] ISO. *ISO 19115:2003 Geographic information – Metadata*. ISO, 2008. Disp. Em: <<http://www.iso.org/iso/home.htm>>.
- [9] D.J. Maguire and P.A. Longley. The emergence of geoportals and their role in spatial data infrastructures. *Computers, Environment and Urban Systems*, 29:3–14, 2005.
- [10] C. Marcatto. *Agricultura sustentável: conceitos e princípios*. Technical report, Rede Ambiente, 2002.
- [11] J. Nogueras-Iso, F.J. Zarazaga-Soria, R. B´ jar, P.J. Alvarez, and P.R. Muro-Med. OGC catalog services: a key element for the development of spatial data infrastructure. *Computers & Geosciences*, 31:199–209, 2005.
- [12] E. Sciore, M. Siegel, and A. Rosenthal. Using semantic values to facilitate interoperability among heterogeneous information systems. *ACM Trans. Database Syst.*, 19(2):254–290, 1994.
- [13] J.C.M. Strauch, J.M. Souza, and M.L.Q. Mattoso. A methodology for gis database integration. In *IEEE Knowledge and Data Engineering Exchange Workshop (KDEX'98)*, pages 151–159., Taipei, Taiwan., November 1998. IEEE CS Press.
- [14] Nectaria Tryfona and Jayant Sharma. *Advanced Information Systems Engineering*, volume 1080/1996, chapter On information modeling to support interoperable spatial databases, pages 210–221. Springer Berlin / Heidelberg, 1996.
- [15] Z. Zhang. *Ontology query languages for the semantic web: a performance evaluation*. Master's thesis, University of Georgia, Athens, 158pp. 2005., 2005.