

Natdata – plataforma de recursos naturais dos biomas brasileiros: informações geoespaciais para sustentabilidade na agricultura

Carla Geovana do Nascimento Macário¹
Eduardo Antônio Speranza¹
João dos Santos Vila da Silva¹

¹Embrapa Informática Agropecuária
Av. André Toselo, 209 - Caixa Postal 6041
13083-886 - Campinas, SP, Brasil
{carla, speranza, jvilla}@cnptia.embrapa.br

Resumo: Este trabalho apresenta o Natdata, uma plataforma para integração de informação de Recursos Naturais dos Biomas Brasileiros. O artigo comenta o processo a ser adotado no seu desenvolvimento para lidar com problemas como a heterogeneidade semântica, de dados e espacial. Em especial destaca-se neste contexto a contribuição do Bioma Pantanal, rico em diversidade de informação, com vários dados já estruturados, podendo servir como um ambiente para validação das atividades sendo desenvolvidas.

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

Abstract: This work presents Natdata, a platform for integration of Natural Resources of Brazilian Biomas data. The article comments the process to be adopted in the platform development, aiming at to deal with problems such as semantic and spatial heterogeneity. In particular stands out the contribution of the Pantanal Bioma, rich in information diversity, having several structured data, which can serve as a validation environment for the activities being developed.

Key Words: geographic database, database integration, natural resource information, interoperability.

1. Introdução

São inúmeros os desafios para uma agricultura sustentável. Um deles é prover um inventário dos recursos naturais, permitindo a avaliação desses recursos para seu aproveitamento sustentável em sistemas de produção de alimentos, fibras, essências e energia, e a adaptação de recursos naturais exógenos a esses sistemas e aos biomas Brasileiros.

A geração e manutenção deste inventário requer esforços de gestão estratégica da inovação, de maneira a 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. Tais medidas exigem o acesso facilitado e integrado a dados organizados sobre os diferentes recursos naturais disponíveis nos biomas brasileiros.

O acesso eficiente a esses dados muitas vezes é dificultado por um conjunto de razões: (i) por serem produzidos por entidades diferentes, estes dados geralmente encontram-se distribuídos geograficamente, normalmente armazenados nas instituições que os produziram; (ii) alguns desses dados ainda não se encontram 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 algumas dificuldades. A maior delas é a chamada heterogeneidade. Ela 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 essas dificuldades tem sido um grande desafio de pesquisa na área de informática, há pelo menos 20 anos.

Em 2009 a Empresa Brasileira de Pesquisa Agropecuária – Embrapa – publicou o relatório “Desafios para uma agricultura sustentável” (Embrapa, 2009). Neste documento, ela indica 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. Neste sentido, deve 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.

A Embrapa iniciou em 2011 o desenvolvimento do Natdata, uma plataforma que deverá integrar dados de Recursos Naturais dos Biomas Brasileiros, visando prover a seus clientes internos e externos um ambiente tecnológico para a gestão da informação

de recursos naturais dos biomas brasileiros. Com isso pretende a geração de inteligência competitiva para a agricultura nacional e fornecer 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.

2. Objetivo

Este trabalho apresenta o Natdata, sua metodologia de desenvolvimento, atividades já realizadas, apresentando um retrato do Bioma Pantanal com relação à informação a ser considerada.

3. Material e Métodos

3.1 A plataforma Natdata

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

Os objetivos específicos do Natdata são: (i) levantar e mapear os dados de recursos naturais da Embrapa com foco em solo, clima, recursos hídricos e componentes da biodiversidade (microrganismos patógenos, bactérias biocontroladoras, artrópodes, vertebrados e plantas); (ii) oferecer um protocolo de recomendações e práticas para o estabelecimento de um processo de integração de banco de dados na Embrapa considerando questões de arquitetura da informação e de propriedade intelectual; (iii) propor instrumentos formais de proteção da propriedade intelectual e de direitos autorais dos dados sobre recursos naturais da Embrapa; (iv) implantar um arcabouço tecnológico em web, com facilidades para consulta, manutenção e evolução; (v) oferecer ferramentas de gestão da informação e do conhecimento relacionadas com reunião, recuperação, organização e disseminação das informações sobre recursos naturais; (vi) implantar a infraestrutura que permita evolução contínua e integração de futuros bancos de dados; e (vii) implantar um portal que permita difusão e transferência de informações.

A **Figura 1** esquematiza a plataforma, ilustrando as informações de recursos naturais a serem integradas inicialmente: Solos e Clima, já estruturados em banco de dados e Recursos Hídricos e Biodiversidade – fauna e flora, cujos bancos de dados encontram-se em desenvolvimento. A figura considera também outras informações que podem ser agregadas à plataforma no futuro como Geologia, Uso da Terra e socioeconomia (destacadas em pontilhado). Além disso, apresenta alguns dos serviços que a plataforma pretende disponibilizar a seus usuários.

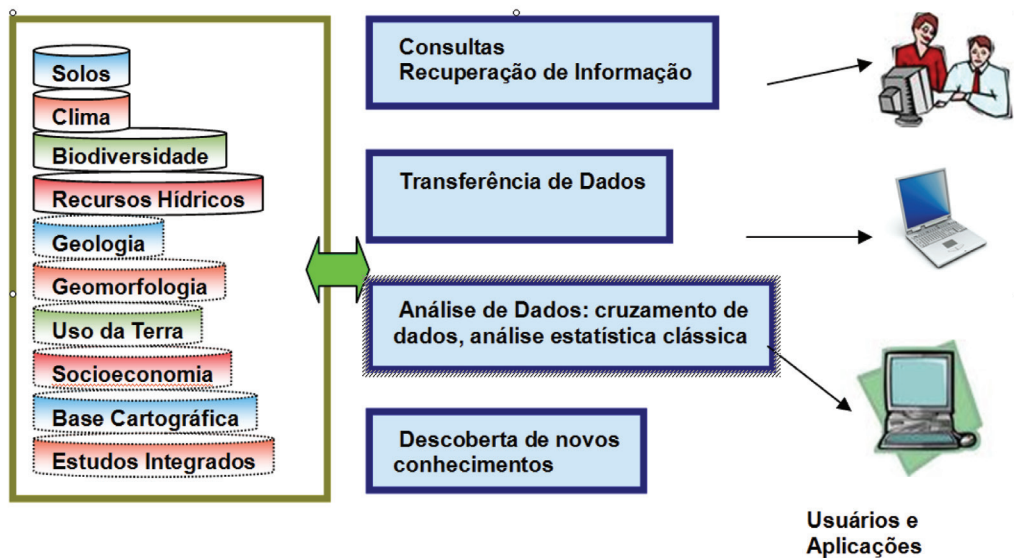


Figura 1. Informações temáticas dos meios físico, biológico e antrópico que podem ser armazenadas no NatDadta e serviços pretende-se fornecer aos usuários.

3.2. Integração de Dados Geográficos –o estado da arte

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.

3.3. Situação na Embrapa

Dados espaciais sobre os recursos naturais brasileiros vêm sendo produzidos ao longo de 30 anos. Mapeamentos sistemáticos do território brasileiro sobre recursos naturais (geologia, geomorfologia, pedologia e vegetação) tiveram suas publicações iniciadas em

1973 no âmbito do Projeto Radam Brasil e ainda estão sendo concluídas pelo IBGE, restando algo em torno de 10% para finalização. Mesmo que esse mapeamento esteja defasado pelo tempo, constitui-se na mais completa base de dados ambientais que o Brasil possui. Ainda com relação à cobertura de dados em nível nacional destacam-se as séries históricas fluviométricas, atualmente sob responsabilidade da Agência Nacional da Água e as séries históricas climatológicas, sob responsabilidade do Instituto Nacional de Meteorologia. Ressalta-se que instituições como INPE, IBAMA, Secretarias Estaduais de Planejamento ou de Meio Ambiente e a própria Embrapa, por meio de sua rede de Unidades descentralizadas por todo o país, geram constantemente dados e informações ambientais nas mais diferentes escalas e abrangências geográficas.

O IBGE, por sua própria incumbência, possui a maior série histórica de dados econômicos e demográficos e, atualmente, já disponibiliza muitos deles via Internet. No entanto, as informações sobre recursos naturais (meios físico e biológico), além de não estarem completas para todo o país, não estão organizadas em um banco de dados operacional. O Ministério do Meio Ambiente, por meio de parcerias, inclusive com várias Unidades da Embrapa, está concluindo o mapeamento da cobertura vegetal (uso e fitofisionomia) do Brasil, ano base 2002, cujos dados estão sendo organizados num banco de dados e parte deles foram disponibilizados à sociedade em geral em 2008. O MAPA possui uma extensa base de dados sobre as cultivares brasileiras.

Outra iniciativa de banco de dados é a do Sistema de Proteção da Amazônia (SIPAM), cujos dados podem ser úteis ao agronegócio, pois envolvem informações físicas e biológicas da Amazônia Legal. Com a obrigatoriedade de execução do zoneamento ecológico-econômico, alguns Estados como Acre, Rondônia e Mato Grosso geraram uma série de dados ambientais úteis ao agronegócio. Salienta-se que a própria Embrapa já investiu esforço e recursos financeiros com a intenção de estruturar seus dados produzidos em um banco, mas os resultados não se materializaram. Destaca-se o SISGEO na década de 80 e o SISOLOS na década de 90. A iniciativa da Embrapa que tem se consolidado é a do banco de dados de recursos genéticos vegetais (Sibrargen), sob liderança da Embrapa Recursos Genéticos e Biotecnologia, com participação da Embrapa Informática Agropecuária.

Especificamente na área de biodiversidade, tem-se o CRIA¹ – Centro de Referência em Informação Ambiental, uma associação civil, sem fins lucrativos, que atua na área de informação biológica, de interesse industrial e ambiental, e pretende, através de sua atuação, contribuir diretamente para a conservação e utilização racional da biodiversidade no Brasil. O CRIA desenvolveu dois sistemas que atuam como as grandes portas de entrada de informações de biodiversidade do país: (i) *SpeciesLink*², um sistema de informação distribuído para recuperação de dados de acervos de coleções biológicas e de observação em campo; e (ii) SinBiota³, sistema de Informação Ambiental para o Estado de São Paulo desenvolvido para armazenar dados de coletas e inventários realizados pelos projetos vinculados ao Programa Biota/Fapesp, cujo desenvolvimento atualmente é responsabilidade do Instituto de Computação da Unicamp. Uma parte dos dados de biodiversidade da Embrapa encontra-se em algum desses sistemas, já que a empresa não possui o seu próprio banco de dados.

Apesar de todo o esforço envolvido no desenvolvimento dessas diversas bases de dados, não existe no país um banco de dados que reúna informações de recursos naturais,

¹<http://www.cria.org.br/>

²<http://splink.cria.org.br>

³<http://sinbiota.biota.org.br/info/>

permitindo consultas, intercâmbio de dados e análises no ambiente Web.

Além disso, o projeto deve estar alinhado ao INDE, uma ação que vem sendo desenvolvida pelo governo nacional para o estabelecimento de uma Infraestrutura Nacional de Dados Espaciais (INDE). Em janeiro de 2010 foi lançado um documento definindo um conjunto de ações para tornar esta infraestrutura uma realidade, alinhando o país com iniciativas internacionais, como o IDEE na Espanha, SPIRE na Europa e SDI, nos EUA. Dentre os objetivos da INDE estão a indicação de um padrão de metadados para dados geoespaciais e estabelecimento de protocolos e políticas de dados. Portanto, dentro do possível, o projeto pretende interagir com a INDE na definição destes padrões e demais ações, de maneira a tornar os dados da empresa interoperáveis, seguindo recomendações nacionais e mundiais.

3.5. Integração de Dados Espaciais

Na década de 1980, integrar bases de dados significava reunir todos os dados em um único banco de dados. Esta abordagem foi substituída por ambientes abertos, nos quais as informações de cada domínio são integradas em tempo real, no momento de seu uso. A principal vantagem neste caso é preservar os sistemas existentes, adotando soluções que permitam o acesso e o compartilhamento aos dados.

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 (Tryfona, & Sharma, 1996). Peculiaridades de dados espaciais, como visões de espaços e relacionamentos espaciais, são críticas para a representação de informação em um ambiente interoperável, tornando o problema da integração um desafio maior do que quando se manipula dados tradicionais (Strauch et al., 1998).

Ao longo dos anos foram propostas diferentes soluções para o problema da integração de dados tradicionais. Em paralelo, surgiram abordagens que estendem essas soluções para dados espaciais, como forma de prover o compartilhamento efetivo de dados geográficos. Essas 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 desse tipo de informação, a maioria prevê a integração de informação proveniente de um mesmo domínio (Gomez, et al., 2009).

3.6. Interoperabilidade de Dados Espaciais

Interoperabilidade de dados espaciais requer o manuseio apropriado dos diferentes tipos de dados. Uma das abordagens para tratar esta questão é a adoção de padrões bem estabelecidos para a descrição desses 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 (ISO, 2008), o padrão FGDC (FGDC, 1998) e a linguagem de marcação geográfica GML (OGC, 2010). O objetivo desses padrões é prover um conjunto de definições e terminologias comuns para a documentação e a troca de dados. A documentação é geralmente publicada em catálogos de dados espaciais, permitindo que usuários descubram e recuperem os dados (Nogueras-Iso et al., 2005; Maguire & Longley, 2005). 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 (Maguire & Longley, 2005).

Segundo Anzlic (2008), existem três tipos de padrões espaciais: (i) aqueles relacionados a informação geográfica em si, como uso da terra e dicionário de nomes; (ii) os que promovem acesso à informação, permitindo ao usuário identificar, compreender e obter a informação desejada; e (iii) os usados para transporte da informação geográfica, entre produtor e consumidor.

3.7. GeoPortais

Portais são sites na Web que atuam como uma porta ou gateways para uma coleção de recursos de informação, como conjunto de dados, serviços, procedimentos, tutoriais, ferramentas e coleções organizadas de ligações (links) para outros sites. Portanto, um portal é um ambiente na Web que permite a uma organização, ou a usuários e provedores de informação, agregar e compartilhar conteúdo e criar consensos sobre temas (Maguire & Longley, 2005).

GeoPortais, ou portais geoespaciais, organizam e indexam informação geográfica, provendo um conjunto de serviços, como consulta a dados e serviços especializados. Esses portais são considerados elementos chave para a geração de conhecimento e de capacidades geográficas, sendo parte de uma infraestrutura necessária para prover um inventário de recursos naturais e culturais. A maior contribuição desse tipo de portal é a simplificação do acesso à informação geográfica, ajudando e estimulando pessoas que buscam por essa informação (Maguire & Longley, 2005).

Um dos serviços providos por estes portais são catálogos de metadados. Os catálogos são estruturas complexas que permitem que os dados sejam encontrados e recuperados, por meio do acesso às suas descrições via metadados, as chamadas anotações (Nogueras-Iso et al., 2005).

3.8. Metodologia de Desenvolvimento

Uma proposta de trabalho como esta que enfrenta diferentes desafios em diversas áreas, envolvendo informação e conhecimento proveniente de diferentes domínios, prevê a reunião de material e métodos diversos no processo de sua execução. A seguir apresentam-se as etapas de trabalho para o desenvolvimento da plataforma de Informação de Recursos Naturais da Embrapa.

3.8.1. Estado da Arte dos Dados de Recursos Naturais na Embrapa

Não existe um retrato exato da situação dos dados existentes na Embrapa sobre os biomas brasileiros. Alguns fatores contribuem para isso, como a distribuição geográfica da informação e a heterogeneidade de formatos desta informação. Antes de organizar um dado, é necessário ter esse retrato e conhecer a finalidade do uso da informação. Assim, o trabalho de desenvolvimento da plataforma teve início com a elaboração e aplicação de um questionário direcionado à identificação da situação dos dados de recursos naturais. Dentre outras coisas, o questionário abordou o tipo de dado que cada tema manipula, autoria, finalidade e formato de armazenamento. Além disso, questões sobre necessidades para o tratamento deste dado foram consideradas. Por exemplo, quais questões pretendem ser respondidas com esses dados? Qual a integração com demais temas é vislumbrada? Ao final desta etapa, conseguimos obter um retrato real da situação dos dados de recursos naturais de solo, clima, recursos hídricos e coleções biológicas. Este retrato está servindo como insumo para as atividades de modelagem da infraestrutura de integração e dos bancos de dados individuais.

A área de trabalho contemplada neste projeto será o território brasileiro com seus 8,5 milhões de quilômetros quadrados. Serão efetivados acordos multi-institucionais para viabilizar a infraestrutura, apoio logístico e aquisição de materiais necessários ao desenvolvimento do projeto. Serão realizados levantamentos dos dados existentes, tais como mapas, fotos, imagens de satélite, relatórios, bibliografias, etc.

Buscar-se-á mapear e recuperar dados e informações geradas nas diversas Unidades da empresa e nas instituições federais e estaduais, além daqueles gerados no âmbito de programas ou projetos como Radam Brasil, Sipam, Prodes, Probio, Macrozoneamentos ou Zoneamentos Ecológico-econômicos de alguns Estados da Federação. Os dados serão organizados em forma de quadros, tabelas, mapas, figuras e gráficos, estendendo-se até o início da carga no banco de dados. Os dados espacializados dos diferentes temas selecionados serão trabalhados, preferencialmente, nas escalas de 1:1.000.000 e 1:250.000, não descartando porém, as diversas escalas existentes, inclusive até aquelas pontuais ou locais.

3.8.2. Plataforma de integração dos dados

Será implantada uma plataforma de integração de dados de diferentes temas de recursos naturais. Esta plataforma deve ser independente dos domínios envolvidos e também de fácil evolução, de maneira a acomodar novos bancos de dados que venham a ser integrados. Da mesma forma, deve considerar questões como distribuição geográfica, necessidades/requisitos dos diversos temas e uso de software livre.

3.8.3 Elaboração de Bancos de Dados Individuais de Recursos Naturais

Neste projeto pretende a manutenção de sistemas já existentes – solos e clima – e projeto e implantação de outros, como é o caso do tema coleções biológicas da Embrapa. No caso de sistemas existentes, é possível serem necessários ajustes para facilitar a sua integração aos demais bancos. A obtenção das informações temáticas deverá obedecer ao seguinte procedimento metodológico: reunião do material existente nos diferentes locais, sistematização dos dados, conversão dos dados analógicos em digitais, migração para os formatos selecionados, geração de conversores para dados digitais legados, inserção dos dados no BDA, elaboração da documentação final e treinamento dos usuários colaboradores.

A sistematização dos dados envolve a leitura crítica, atualização, padronização e conversão das informações de cada tema obtido para o(s) formato(s) especificado(s) no BDA, mediante ajustes necessários efetuados manualmente ou digitalmente.

A inserção dos dados nos bancos de dados ambientais consiste no armazenamento das informações de cada temática em meio digital, podendo ocorrer, quando se tratar de dados geográfico, por digitalização via mesa digitalizadora, scannerização e vetorização ou digitalização diretamente da tela do microcomputador, sobre um dado pré-existente. Quando se tratar de dados alfanuméricos poderá ser feita por transferência de arquivos ou digitação.

3.8.4. Infraestrutura de Integração dos Dados de Recursos Naturais

Por razões técnicas e operacionais, está sendo priorizada a integração de dados referentes aos temas solos, clima, recursos hídricos e biodiversidade. Entretanto, sendo objetivos deste projeto a definição e a validação de um processo de integração de dados, bem como a implementação de uma plataforma de apoio a este processo, entendemos que outros temas de interesse da Embrapa como uso das terras, cobertura vegetal, unidades

de conservação e dados socioeconômicos, poderão ser facilmente integrados ao ambiente desenvolvido.

Cada um dos temas a serem trabalhados no projeto encontra-se em estágio diferente quanto à organização de seus dados. Portanto, a abordagem a ser utilizada pretende a integração gradual dos diversos dados de recursos naturais. Dentre as vantagens decorrentes, tem-se a validação da arquitetura quanto a questões de adaptação e extensibilidade.

3.8.5. Portal de Informação Sobre Recursos Naturais

A estratégia a ser adotada para disseminação do conhecimento produzido é a implantação de um portal de informações. O portal deve reunir não só o conhecimento, mas também prover serviços e mecanismos de busca inteligentes, fornecendo meios aos potenciais usuários do sistema de obter a informação desejada. Pretende-se que este Portal não apenas reúna a informação dos diferentes temas, mas funcione como elemento de comunicação entre as redes de pesquisa estabelecidas, oferecendo diferentes canais de informação, mídias sociais, acesso aos sistemas individualmente, acesso e exportação de dados.

Exemplos de serviços especializados incluem a geração e o uso de mapas, catálogos de dados, expansão de consultas baseadas em ontologias e sistemas especialistas.

A estratégia de ação para desenvolvimento do projeto prevê o início do trabalho de estruturação de temas cujas informações já estão mais organizadas, como solo e clima, e de outros cujas redes de pesquisa existem e buscam esta organização, como recursos hídricos e biodiversidade. A experiência de integração desses quatro temas deve resultar num conjunto de boas práticas a serem utilizadas na adição de informação de outros temas envolvendo recursos naturais.

4. Resultados Parciais e Discussão

O desenvolvimento do projeto teve início com a aplicação de questionários buscando elaborar um retrato da situação dos dados de recursos naturais na Embrapa. Dentre outras coisas o questionário pretendeu um levantamento das formas utilizadas para organização dos dados. Foram aplicados questionários juntos a todos os membros do projeto, distribuídos em quatro temas: clima, solos, recursos hídricos e biodiversidade. Apesar de solos e climas serem bases já estruturadas, a aplicação do questionário permitiu identificar as possibilidades de integração existentes.

Como esperado, grande parte dos dados encontram-se estruturados sob forma de planilhas, e arquivos texto e em alguns casos sistemas como o SpeciesLink. Se por um lado a estruturação usando abordagens simples como planilhas eletrônicas auxilia na proposta de um banco de dados comum, pois de certa forma propõe uma forma de estruturar os dados, por outro lado leva a diferentes visões dos dados, o que pode dificultar um esquema de dados comum que atenda a todos os seus possíveis usuários. A adoção de sistemas prontos leva à necessidade de implementação de mecanismos que permitam a interoperabilidade de dados entre sistemas semelhantes.

Em especial com relação ao Bioma Pantanal tem-se dados relacionados à unidade de pesquisa Embrapa Pantanal. São eles: (i) a informação de recursos hídricos está totalmente organizada sob a forma de planilhas excell, contendo informação descritiva numérica, gráficos, mapas e fotos; (ii) em torno de 80% da coleção entomológica daquela unidade está organizada principalmente em arquivos texto, armazenando informação descritiva numérica, fotografias e mapas; (iii) 90% da coleção de peixes encontra-se organizada usando principalmente o sistema Systat, mas também adota planilhas excell, arquivos

texto e banco de dados FoxPro; (iv) 100% da coleção de vertebrados está organizada, sendo 95% sob a forma de planilhas excell e 5% em fotografias; e (v) 70% da coleção de herbário está organizada, adotando arquivos texto, fotografias e arquivos Clipper;

Com base nas respostas recebidas, num passo seguinte do desenvolvimento do projeto, passou-se à elaboração de diagramas que retratassem os dados que estão presentes em cada um dos temas, tentando-se identificar as possibilidades de integração entre eles. A **Figura 2** ilustra uma parte desses diagramas, descrevendo as informações relacionadas à Espécie e destacando em amarelo os dados que foram identificados como comuns aos temas. Isso permite direcionar os trabalhos de integração da plataforma, bem como auxilia na estruturação dos novos bancos de dados.

Paralelo a este trabalho iniciou-se a avaliação de melhor estrutura para implementação da plataforma para integração dos dados. Neste exercício pretendem-se duas coisas: avaliar a melhor arquitetura para acomodar os dados temáticos e a tecnologia para implementar esta arquitetura. Apparentemente uma decisão simples, ela deve considerar vários fatores como heterogeneidade dos dados, heterogeneidade espacial e heterogeneidade semântica. Esta avaliação vem sendo feita usando dados de solos e de clima, considerando o uso de serviços web para a importação e exportação dos dados, bem como a adoção de serviços geográficos padrões que facilitam a geração de mapas e troca de dados.

Finalmente, a implementação dos novos bancos também vem sendo trabalhada mediante entrevistas com especialistas de cada domínio. Um esquema preliminar para o banco de dados de Biodiversidade e de Recursos Hídricos está sendo elaborado para apresentação no workshop do projeto que ocorrerá em agosto.

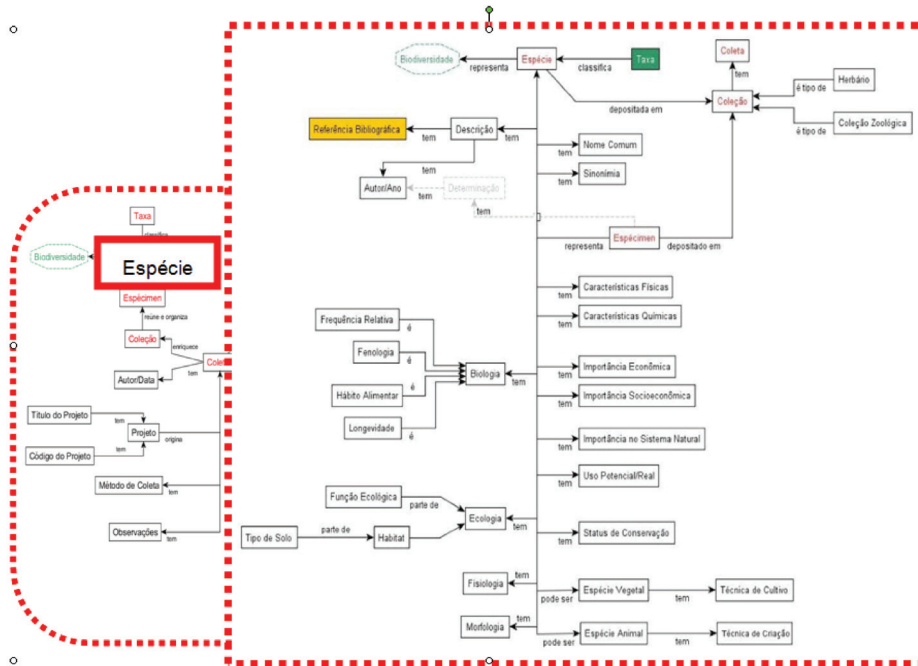


Figura 2. Parte de um diagrama de integração de dados de recursos naturais - Espécie.

Em especial o Bioma Pantanal tem grande importância no projeto, dado que para este Bioma há uma grande disponibilidade de Bancos de Dados de Recursos naturais, como Uso e Cobertura da Terra, áreas de preservação, relevo entre outros. Assim, pretende-se usar estes bancos de informação como base não apenas para estruturar novos temas, como também para validar a integração de dados considerando uma grande heterogeneidade de informação.

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, destaca-se um retrato da situação dos dados de recursos naturais na Embrapa montado a partir de respostas a um questionário aplicado junto aos pesquisadores envolvidos. Nesse retrato identifica-se que o Bioma Pantanal, além de apresentar uma diversidade de informação temática, possui grande parte desses dados já estruturados em meios digitais, podendo servir como um ambiente para validação das atividades sendo desenvolvidas no projeto.

5. Agradecimentos

A equipe do projeto agradece ao Fundo Embrapa-Monsanto.

6. Bibliografia

- Anzlic **The spatial information council**. Anzlic. Disponível em: <<http://www.anzlic.org.au>>. Acesso em março/2012.
- Buccella, A., Cechich, A. Towards integration of geographic information systems. **Electron. Notes Theor. Comput. Sci.**, 168:45–59. 2007.
- Embrapa **desafios para uma agricultura sustentável**. Embrapa, Assessoria de Comunicação Social. – Brasília, DF : Embrapa - 79 p. 2009
- FGDC . FGDC-STD-001-1998. **Content Standard for Digital Geospatial Metadata**. Washington, D.C. 1998.
- Gomez, L., Haesevoets, s., Kuijpers, B. ; Vaisman, A. A. Spatial aggregation: data model and implementation. **Information Systems**, 34:551–576. 2009.
- ISO **ISO 19115:2003 Geographic information – Metadata**. ISO. Disponível em: <<http://www.iso.org/iso/home.htm>>. Acesso em março/2012.
- Maguire, D.J.; Longley, P.A. The emergence of geoportals and their role in spatial data infrastructures. **Computers, Environment and Urban Systems**, 29 (2005) 3–14.
- Nogueras-Iso, J.; Zarazaga-Soria, F.; Bejar, R., Lvarez, P.; Muro-Med, P. OGC catalog services: a key element for the development of spatial data infrastructure. **Computers & Geosciences**, 31:199–209. 2005.
- 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
- 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.
- Spaccapietra, S., Cullot, N., Parent, C., Vangenot, C. On spatial ontologies. In **Brazilian Symposium on GeoInformatics - GEOINFO**, Campos do Jordão. 2004.
- Strauch, J., Souza, J., Mattoso, M. A methodology for gis database integration. In **IEEE Knowledge and Data Engineering Exchange Workshop (KDEX'98)**, 1998. p 151–159., Taipei, Taiwan., IEEE CS Press.
- Tryfona, N.; Sharma, J. **Advanced Information Systems Engineering**, volume 1080/1996, chapter On information modeling to support interoperable spatial databases, 1996. p. 210–221. Springer Berlin / Heidelberg.