

GeoInfo - Infraestrutura de Dados Espaciais Abertos para a Pesquisa Agropecuária

GeoInfo - Open Spatial Data Infrastructure for Agricultural Research

GeoInfo - Infraestructura de Datos Espaciales Abiertos para la Investigación Agropecuaria

Debora Pignatari Drucker | debora.drucker@embrapa.br

Embrapa Informática Agropecuária, Campinas, Brasil.

Daniela Maciel Pinto | daniela.maciel@embrapa.br

Embrapa Monitoramento por Satélite, Campinas, Brasil.

Elaine Cristina Cardoso Fidalgo | elaine.fidalgo@embrapa.br

Embrapa Solos, Rio de Janeiro, Brasil.

Davi Oliveira Custódio | davi.custodio@embrapa.br

Embrapa Monitoramento por Satélite, Campinas, Brasil.

Daniel de Castro Victoria | daniel.victoria@embrapa.br

Embrapa Informática Agropecuária, Campinas, Brasil.

Bibiana Teixeira Almeida | bibiana.almeida@embrapa.br

Embrapa Monitoramento por Satélite, Campinas, Brasil.

Margareth Simões | margareth.simoese@embrapa.br

Embrapa Solos, Rio de Janeiro, Brasil.

Claudia Regina De Laia Machado | claudia.delaia@embrapa.br

Embrapa Solos, Rio de Janeiro, Brasil.

Vera Viana dos Santos Brandão | vera.viana@embrapa.br

Embrapa Monitoramento por Satélite, Campinas, Brasil.

Ricardo de Oliveira Dart | ricardo.dart@embrapa.br

Embrapa Solos, Rio de Janeiro, Brasil.

Mário Luiz Diamante Aglio | mario.aglio@embrapa.br

Embrapa Solos, Rio de Janeiro, Brasil.

Francisca Rasche | francisca.rasche@embrapa.br

Embrapa Florestas, Colombo, Brasil.

Maria Regina Capdeville Laforet | regina.laforet@embrapa.br

Embrapa Solos, Rio de Janeiro, Brasil.

Giovana Maranhão Bettiol | giovana.bettiol@embrapa.br

Embrapa Cerrados, Brasília, Brasil.

Sandro Eduardo Marschhausen Pereira | sandro.pereira@embrapa.br

Embrapa Meio Ambiente, Jaguariúna, Brasil.

Ziany Neiva Brandão | ziany.brandao@embrapa.br

Embrapa Algodão, Campina Grande, Brasil.

Marilice Cordeiro Garrastazu | marilice.garrastazu@embrapa.br

Embrapa Florestas, Colombo, Brasil.

José Maria Filippini Alba | jose.filippini@embrapa.br

Embrapa Clima Temperado, Pelotas, Brasil.

Fábia Amorim da Costa | fabia.amorim@embrapa.br

Embrapa Clima Temperado, Pelotas, Brasil.

Renato Cristiano Torres | renato.torres@embrapa.br

Embrapa Agrossilvipastoril, Sinop, Brasil.

Márcia Helena Galina Dompieri | marcia.dompieri@embrapa.br

Embrapa Tabuleiros Costeiros, Aracaju, Brasil.

Sandra Maria Neiva Sampaio | sandra.sampaio@embrapa.br

Embrapa Amazônia Oriental, Belém, Brasil.

Espedito Cezário Martins | espedito.martins@embrapa.br

Embrapa Caprinos e Ovinos, Sobral, Brasil.

Ivo Pierozzi Junior | ivo.pierozzi@embrapa.br

Embrapa Informática Agropecuária, Campinas, Brasil.

Gustavo Bayma-Silva | gustavo.bayma@embrapa.br

Embrapa Meio Ambiente, Jaguariúna, Brasil.

Graziella Galinari | graziella.galinari@embrapa.br

Embrapa Monitoramento por Satélite, Campinas, Brasil.

Flávia Bussaglia Fiorini | flavia.fiorini@embrapa.br

Embrapa Monitoramento por Satélite, Campinas, Brasil.

Celina Maki Takemura | celina.takemura@embrapa.br

Embrapa Monitoramento por Satélite, Campinas, Brasil.

Sergio Aparecido Braga da Cruz | sergio.cruz@embrapa.br

Embrapa Informática Agropecuária, Campinas, Brasil.

Wilson Anderson Holler | wilson.holler@embrapa.br

Embrapa Florestas, Colombo, Brasil.

Leandro Henrique M de Oliveira | leandro.oliveira@embrapa.br

Embrapa, Departamento de Pesquisa e Desenvolvimento, Brasília, Brasil.

Resumo

A geoinformação é essencial para o planejamento e monitoramento das atividades agropecuárias, justificando o emprego de esforços para reuni-la e padronizá-la de acordo com as diretrizes governamentais e possibilitando sua disponibilização à sociedade em geral. O objetivo deste trabalho é apresentar a experiência da Embrapa em construir sua Infraestrutura de Dados Espaciais, denominada GeoInfo, uma iniciativa para organizar, preservar, documentar e ofertar dados geoespaciais abertos produzidos nas pesquisas da empresa, ampliando o potencial de aplicação dessa informação na produção e difusão de conhecimento e inovação. O GeoInfo implementa os padrões da Infraestrutura Nacional de Dados Espaciais e viabiliza a interoperabilidade de dados geoespaciais provenientes de diferentes fontes, inclusive sob o aspecto semântico. Essa iniciativa possibilita a integração das informações geoespaciais produzidas na Embrapa com diversas informações disponíveis em nosso país e permite que a redundância de esforços na obtenção e produção de dados geoespaciais sejam evitados.

Palavras-chave: Geoinformação; Interoperabilidade; Ciência Agropecuária; Geoserviços; Dados de pesquisa; Metadados

Abstract

Geoinformation is essential to plan and monitor agricultural activities, justifying efforts to gather and standardize it according to governmental guidelines and to make it available to decision makers and general public. This work aims on presenting the experience of Embrapa on building its Spatial Data Infrastructure (SDI). This SDI, called GeoInfo, is an initiative to organize, preserve, document and offer geodata produced by the company, in order to increase the application of this information in the production and diffusion of knowledge and innovation. GeoInfo implements the guidelines of the Brazilian National Spatial Data Infrastructure and enables scientific geodata interoperability, encompassing semantics. GeoInfo SDI promotes redundancy of efforts avoidance in obtaining and producing geodata. This initiative enables the integration of geospatial information produced at Embrapa with many other public information sources available worldwide.

Keywords: Geoinformation; Interoperability; Agricultural Science; Geoservices; Scientific Data ; Metadata, Geodata

Resumen

La geoinformación es esencial para la planificación y monitoreo de las actividades agropecuarias, justificando el empleo de esfuerzos para reunirla y estandarizarla de acuerdo con las directrices gubernamentales, posibilitando su disponibilidad para la sociedad en general y los tomadores de decisión. El objetivo de este trabajo es presentar la experiencia de Embrapa en construir su Infraestructura de Datos Espaciales (IDE), denominada GeoInfo, una iniciativa para organizar, preservar, documentar y ofrecer la geoinformación producida por la empresa, para ampliar el potencial de aplicación de esa información en la producción y difusión de conocimiento e innovación. La IDE-GeoInfo implementa los estándares de la Infraestructura Nacional de Datos Espaciales (INDE) y viabiliza la interoperabilidad de datos e informaciones espaciales provenientes de distintas fuentes, incluso bajo el aspecto semántico. El uso del Geoinfo permite que la redundancia de esfuerzos en la obtención y producción de datos geoespaciales sean evitados. Geoinfo es una estructura que posibilita la integración de las informaciones geoespaciales producidas en la Embrapa con diversas informaciones disponibles en todo el mundo.

Palabras Clave: Geoinformación; Interoperabilidad; Ciencias Agropecuarias; Geoservicios; Datos de Investigación; Metadatos.

Introdução

As primeiras décadas do século XXI têm se caracterizado por um crescimento explosivo na capacidade humana para adquirir, armazenar e comunicar dados e informações digitais. No contexto científico, esse novo paradigma denomina-se *e-Science*, ou ciência intensiva em dados¹⁻², e vem provocando transformações na maneira como a informação e o conhecimento são gerados, utilizados e disponibilizados. Neste cenário, os dados passam a ter uma enorme importância, constituindo-se um novo ambiente orientado por dados na comunicação científica³.

Atualmente, um dos ramos no qual a ciência intensiva em dados vem se consolidando é o que trata dos dados espaciais, ou seja, dados que possuem uma localização no espaço geográfico. Neste sentido, a geoinformação, informação geográfica ou informação geoespacial refere-se a todo dado interpretado por especialistas e passível de espacialização, apresentando um vínculo geográfico que permite sua localização de forma imediata e inequívoca⁴. As geotecnologias, conjunto de tecnologias para coleta, processamento, análise e disponibilização de informação geográfica, têm evoluído de forma exponencial nas últimas décadas com o número crescente de satélites e sensores, algoritmos robustos para processamento de imagens e tratamento de dados geoespaciais, máquinas e equipamentos autônomas no campo, entre outras. O processo de popularização da geoinformação tem promovido uma verdadeira revolução, conhecida como a “Era da GeoInformação”, solidificando de vez o conceito de “Sociedade do Conhecimento”⁵.

No âmbito do Brasil, a Constituição Federal garante a todo cidadão o direito de receber dos órgãos públicos informações tanto de interesse particular, como de interesse coletivo ou geral⁵. Nesse sentido, um conjunto de políticas nacionais para regulamentar, orientar e facilitar a disseminação de dados vem sendo instituído na última década. As principais são a Lei de acesso à Informação⁶ que regula o acesso à informação e a Política de Dados Abertos⁷ que estabelece as regras para a disponibilização de informação pelo poder Executivo Federal.

A Infraestrutura Nacional de Dados Espaciais - INDE representa um marco legal para a gestão da informação geoespacial e define um conjunto integrado de tecnologias, políticas, padrões, mecanismos de coordenação, monitoramento e acordos necessários para facilitar e ordenar a geração, o armazenamento, o acesso, o compartilhamento, a disseminação e o uso dos dados geoespaciais⁸. Adicionalmente, a Comissão Nacional de Cartografia (CONCAR), por meio do trabalho dos seus comitês especializados, apresenta os padrões e as normas a serem adotadas na produção dos dados e informações geoespaciais pelos órgãos públicos⁹⁻¹⁰.

A geração e disseminação da geoinformação têm importância estratégica para inúmeros setores, destacando-se o fortalecimento da agricultura brasileira, visto que o Brasil apresenta vasta dimensão territorial e diversidade de sistemas de produção. A enorme quantidade de dados espaciais, informações e conhecimentos sobre a agricultura brasileira, muitas vezes não harmonizados e disponibilizados à sociedade civil e aos tomadores de decisão representa um obstáculo aos órgãos governamentais que planejam o futuro da agropecuária brasileira e desenvolvem políticas públicas para este setor da economia.

O conhecimento da dinâmica espacial da agricultura avançou muito nos últimos anos, e a Empresa Brasileira de Pesquisa Agropecuária, Embrapa, é uma das instituições de pesquisa que investiga essa temática, desenvolvendo pesquisas e produzindo informações em mapeamento, monitoramento, zoneamento e modelagem agroambiental para atender às demandas por informação e subsídios à tomada de decisão e elaboração de políticas públicas. A produção massiva de geoinformação e a necessidade de compartilhá-la, favorecendo novas pesquisas e conhecimentos para o desenvolvimento da agricultura sustentável, tem colocado à Embrapa o desafio de organizar uma infraestrutura de dados espaciais que reúna, padronize e disponibilize a geoinformação de acordo com as diretrizes governamentais.

Em resposta a esse desafio, algumas iniciativas se voltaram à organização, armazenamento e disseminação de dados geoespaciais da Embrapa: Geoportal¹¹, Biblioteca Geoespacial¹², Agritempo¹³ e Banco de Produtos Modis¹⁴; TerraClass¹⁵; GeoPortal¹⁶; Macrozoneamento Ecológico-Econômico Maranhão¹⁷, SOMABRASIL¹⁸, Projeto GeoUpgrade¹⁹ e Portal SIAGEO Amazônia²⁰.

Dada a necessidade de padronizar, em ambiente integrado, as informações espaciais geradas pela Embrapa seguindo os padrões estabelecidos pela INDE e, considerando-se que o acesso aberto aos dados espaciais tem importância estratégica no fortalecimento da pesquisa, desenvolvimento e inovação para a agricultura brasileira, a Embrapa formalizou sua adesão à infraestrutura nacional, e instituiu um grupo para elaborar um plano de trabalho para a implementação do nó da Embrapa na INDE. A estratégia para a implementação baseou-se nas experiências anteriores para a disseminação da geoinformação produzida pela Embrapa, agregada às diretrizes da INDE²¹.

Nesse sentido, o objetivo desse trabalho é apresentar a experiência da Embrapa em construir sua Infraestrutura de Dados Espaciais (IDE), uma iniciativa para organizar, preservar, documentar e ofertar a geoinformação produzida pela empresa. A estratégia para a implantação da IDE-Embrapa, a qual foi intitulada GeoInfo, foi fundamentada em cinco pilares ou componentes⁹: (1) Pessoas envolvidas com o planejamento, a aquisição, o tratamento, a catalogação, a gestão, a análise, a oferta e o uso da geoinformação; (2) Dados propriamente ditos, ou seja, informação geoespacial obtidas por diversas fontes; (3) Tecnologia que viabiliza a infraestrutura computacional que possibilitará a organização, preservação e disseminação da geoinformação; (4) Normas e Padrões que permitem a descoberta, intercâmbio, integração e usabilidade da informação geoespacial e (5) Arcabouço Institucional, ou seja, o contexto de governança e coordenação que garante a implementação e longevidade da infraestrutura, por meio de acordos, articulações e regras. Serão apresentadas as estratégias adotadas em cada um dos cinco pilares para viabilizar a implantação da IDE GeoInfo na Embrapa, com o desafio de atender e envolver nesse processo todas as 53 unidades da empresa, distribuídas no território nacional.

Metodologia

O trabalho teve como base o levantamento realizado com todas as unidades da Embrapa²¹, no qual foram identificados, sensibilizados e capacitados atores envolvidos com a geração e uso da geoinformação produzida pela Embrapa, inclusive gestores, visando conferir maior familiaridade com os temas relacionados à infraestruturas de dados espaciais (IDEs). Os dados levantados permitiram caracterizar o processo de produção, organização, armazenamento e disponibilização da geoinformação da instituição, bem como identificar recursos humanos com habilitação em geoinformação e demanda por capacitação. A análise dos dados levantados evidenciou aspectos frágeis no processo de organização, manutenção e disponibilização da geoinformação produzida na empresa, os quais foram priorizados na elaboração da estratégia para a internalização da INDE na Embrapa, detalhada no plano de trabalho para a implementação do nó da Embrapa na INDE¹.

O processo de elaboração da estratégia foi fortemente apoiado na literatura que trata da gestão do ciclo de vida de dados, em especial no trabalho da United States Geological Survey (USGS) – agência do Departamento do Interior dos Estados Unidos – que coordena o Programa Geoespacial desse país, um dos pilares da National Spatial Data Infrastructure (NSDI – Infraestrutura Nacional de Dados Espaciais dos Estados Unidos)²², e no ciclo de vida dos dados proposto pela rede Data Observation Network for Earth (DataONE), ou rede de observação de dados para a Terra, o qual traz o conceito de reuso, com a noção circular de que uma análise leva a resultados que suscitam novas perguntas e o planejamento de uma nova

i Drucker DP, Custódio DO, Fidalgo ECC, Daltio J, Visoli MC. Plano de trabalho: adesão da Embrapa à Infra-estrutura Nacional de Dados Espaciais: INDE. Rio de Janeiro; 2014. Documento institucional não publicado.

pesquisa²³. Todas as etapas do ciclo de vida dos dados são relevantes no contexto de organizar, preservar e disponibilizar dados geoespaciais no âmbito da Embrapa. No contexto da iniciativa GeoInfo, as etapas de validação, descrição, preservação e descoberta são tratadas de forma detalhada.

Com base neste plano, estruturou-se um projeto que contemplou os cinco pilares de uma IDE citados: pessoas, dados, institucional, tecnologia, e normas-padrões. O projeto foi concebido considerando a necessidade de envolver as unidades da Embrapa e compartilhar suas experiências no processo de elaboração, uso e disseminação da geoinformação, ao mesmo tempo considerando a infraestrutura e recursos humanos disponíveis, para criar uma infraestrutura de dados espaciais que atendesse a essas diversas realidades. Para viabilizar a execução do projeto, otimizando recursos, foram selecionadas inicialmente dez unidades para atuar como pilotos no processo de implantação da IDE GeoInfo, com base na lógica de que, a partir delas, as experiências de implantação forneceriam subsídios para a definição e o refinamento de processos para implantação da IDE nas demais unidades da Embrapa. Os critérios para a seleção das unidades foram: disposição em participar da iniciativa e suas diferentes condições em termos de: disponibilidade e uso de infraestrutura; volume de dados; e experiência anterior em criar soluções para a organização, armazenamento e disseminação de dados geoespaciais. Ao longo da execução do projeto, duas novas unidades se inseriram na execução das atividades de criação da infraestrutura necessárias à organização e disponibilização de seus dados espaciais, totalizando doze unidades-piloto.

As metodologias específicas de cada pilar são descritas a seguir:

Pessoas

As Pessoas ou Atores envolvidos com o planejamento, aquisição, tratamento, gerenciamento, análise e uso da geoinformação são um componente chave de uma IDE, que permeia todos os demais. Sendo um dos pilares da INDE, os atores devem ter sua participação e funções examinadas segundo aspectos organizacionais e técnicos⁹. A importância do fator humano em infraestruturas de informação, colaboração e compartilhamento de dados são tópicos que vêm sendo cada vez mais estudados e documentados¹. A primeira tarefa essencial a se realizar na construção de uma IDE consiste na identificação desses atores, considerando a abrangência pretendida⁹. O levantamento realizado com os atores internos forneceu subsídios para a definição dos papéis a serem desempenhados por cada um deles no processo de implantação da IDE GeoInfo. Além disso, foi também considerada a dimensão dos atores externos à instituição Embrapa.

Institucional

O componente institucional compreende as questões de política, legislação e coordenação⁹. Da perspectiva de política, a custódia, o valor e o licenciamento têm papéis importantes²⁴. O fortalecimento institucional, os marcos legais, a captação de recursos e a gestão/coordenação, bem como regras formais e não formais estão contidos nesse arcabouço institucional. No escopo da construção do GeoInfo, no âmbito deste pilar, foi discutido, proposto e estabelecido colaborativamente um processo de publicação de dados geoespaciais abertos pela Embrapa, envolvendo seus geradores, transformadores, gestores e usuários. A definição do processo incluiu aspectos da produção, curadoria, disponibilização e licenciamento da geoinformação pela Embrapa e foram definidas formas de acesso e obtenção dos dados conforme a legislação vigente. Além disso, foi proposta uma estrutura de governança da IDE GeoInfo.

Tecnologia

No contexto do GeoInfo, para viabilizar a internalização do padrões preconizados pela INDE, foi realizada a implantação da infraestrutura física de dados espaciais que subsidia a catalogação dos conjuntos de dados e

que viabiliza a disponibilização dos mesmos. Os procedimentos metodológicos adotados para o desenvolvimento deste pilar envolveu a utilização da arquitetura orientada a serviços (SOA), conforme preconiza a INDE⁹.

Na implantação, os recursos são disponibilizados na forma de serviços, acessados a partir de requisições padronizadas, independente da plataforma ou linguagem de programação utilizada. Dessa forma é possível reduzir os custos de integração de software e compartilhamento de dados, ampliando o acesso dos usuários aos dados e recursos de processamento⁹. No caso do Diretório Brasileiro de Dados Geoespaciais (DBDG), parte integrante da INDE, foi estabelecido que devem ser seguidos pelo menos dois padrões de serviços definidos pela OGC - *Open Geospatial Consortium*²⁵, a saber:

WMS (*Web Map Service*): permite visualização e consulta de mapas georreferenciados

CSW (*Web Catalog Service*): permite publicação e acesso a catálogos de metadados que descrevem dados ou serviços geoespaciais, sendo indispensável para sistemas de buscas de recursos catalogados em uma IDE

Também definiu-se que serviços adicionais deverão seguir, sempre que possível, as especificações OGC correspondentes, como o WFS (*Web Feature Service*), que possibilita interação completa (visualização, consulta, edição, *download*) com dados vetoriais. Adicionalmente deverão ser adotados, preferencialmente, softwares livres ou de código aberto⁹.

Dados

Considerando a diversidade temática das unidades da Embrapa e de dados espaciais produzidos por elas, foram escolhidas as unidades piloto de forma a permitir a identificação dos principais conjuntos de dados espaciais elaborados pela empresa e dos aspectos que poderiam facilitar ou dificultar sua organização e disponibilização. Para permitir a caracterização da produção da geoinformação pela Embrapa, foi selecionada uma grande diversidade dos dados espaciais dessas unidades piloto em termos de tema, formato, escala, abrangência e periodicidade de atualização. Foram levantadas informações sobre o volume, para subsidiar a estimativa de capacidade de armazenamento, e sobre propriedade intelectual, para instruir diretrizes para a política de dados. Por fim, os processos de produção e armazenamento dos dados nas unidades contribuíram para a definição da proposta de processo unificado para a IDE GeoInfo. Ao longo do projeto, o gerenciamento do processo de implantação da infraestrutura nas unidades piloto e acompanhamento dos mecanismos propostos para organizar, armazenar e disponibilizar a geoinformação permitiram atender às demandas surgidas, corrigir rumos e adequar a infraestrutura, suas normas e processos.

Normas e Padrões

O pilar Normas e Padrões permite a descoberta, o intercâmbio, a integração e a usabilidade da informação espacial. Padrões de dados espaciais abrangem sistemas de referência, modelo de dados, dicionários de dados, qualidade de dados, transferência de dados e metadados⁹. Nesse contexto de produção e de especificações diversificadas, a interpretação e o uso adequado dos dados por diferentes usuários demandam a disponibilização de um conjunto de informações sobre esses dados que propicie a compreensão e o entendimento sobre a sua aplicabilidade e forma de utilização¹⁰. Foram realizadas ações visando a internalização e customização do “Perfil de Metadados Geoespaciais do Brasil”¹⁰, de forma a garantir o intercâmbio da informação, bem como atender os serviços web preconizados pela INDE. Dentre o conjunto de metadados essenciais para os processos de descrição e recuperação da informação, consta o campo palavras-chave. A fim de atender um processo de descrição temática adequado foi desenvolvido, no âmbito deste pilar, um vocabulário controlado de geoinformação funcionando como uma ferramenta de auxílio à indexação de termos do domínio geoespacial aplicados à agricultura.

Resultados

A seguir, apresentamos os resultados obtidos no âmbito de cada pilar adotado para a construção da Infraestrutura de Dados Espaciais da Embrapa - GeoInfo.

Pessoas

A estratégia de implantação da IDE - GeoInfo foi fortemente amparada pelo conceito de construção coletiva pelos diferentes atores envolvidos. Essa lógica direcionou as ações referentes ao pilar “Pessoas” e demais pilares. Houve um forte engajamento da equipe e buscou-se garantir o nivelamento de conhecimentos entre os membros do projeto e suas experiências nas Unidades Piloto. Assim problemas em comum foram identificados e as soluções foram elaboradas conjuntamente. Uma estratégia desenvolvida no projeto foi a estruturação de grupos de trabalhos temáticos. Além disso, foram organizadas capacitações, promovidos encontros presenciais e fóruns virtuais. Os temas tratados nessas interações foram: (1) Processo para publicação dos dados na plataforma GeoInfo; (2) Catalogação de dados na plataforma GeoInfo; (3) Qualidade de dados; (4) Política de dados geo; (5) Termo de Cessão de Direitos; (6) Critérios para publicação de dados na Inde; (7) Estilização de dados para a plataforma GeoInfo e (8) Licenciamento de dados. No total, foram capacitadas 63 pessoas, com a participação de todas as 12 unidades pilotos envolvidas no projeto.

A estratégia de comunicação para o processo de construção do GeoInfo e de adesão da Embrapa à INDE foi estruturada levando em consideração os públicos interno e externo. O público interno são os empregados da Embrapa que atuam de maneira direta ou indireta na produção e utilização da informação espacial e gestores da Empresa. O público externo são os usuários e parceiros que acessam e utilizam a informação geoespacial da Empresa. O planejamento inicial foi voltado para a comunicação interna, em especial para o fortalecimento da cultura de valorização do uso e compartilhamento dos dados e informações geoespaciais, essencial para o alcance dos objetivos do projeto. Paralelamente à comunicação interna, houve a divulgação externa das principais ações relacionadas à adesão da Embrapa à INDE. Foram feitas divulgações jornalísticas voltadas aos dois públicos, buscando dar exposição e clareza sobre todo o processo em desenvolvimento.

A marca Geoinfo foi criada para ser utilizada durante o projeto e, posteriormente, ser apresentada ao público-externo como identificador do serviço disponibilizado pela Embrapa. A equipe de comunicação desenvolveu um logotipo que buscava representar a interação entre a agricultura e os dados geoespaciais. Visando o engajamento do público externo, foi criado um folder para distribuição às 46 Unidades da Embrapa e lideranças corporativas, e uma página webⁱⁱ para a apresentação do projeto ao público interno, a parceiros externos ao projeto e a demais contatos estratégicos. Uma interface para o sistema foi elaborada baseada nas premissas de fácil acessibilidade, clareza visual e concordância com a linguagem visual da Embrapa.

Institucional

O processo para a publicação da geoinformação gerada pela Embrapa foi amplamente debatido. O desenvolvimento do processo foi realizado por uma equipe interdisciplinar, composta por produtores de dados espaciais, cientistas da informação, cientistas da computação, cientistas de dados, profissionais das ciências agrárias, entre outros. Várias versões do processo foram apresentadas e testadas até o desenho final. Esse processo envolveu os principais atores responsáveis pela produção, descrição e publicação dos dados espaciais gerados pela Embrapa, sendo composto por três etapas que simplificam a inserção da informação espacial e de seus metadados: (1) Preparar o Cadastro, (2) Realizar o Cadastro e (3) Publicar o Dado Georreferenciado.

ii Disponível em: <https://www.embrapa.br/geoinfo>

A documentação necessária para subsidiar o processo foi elaborada, incluindo orientações sobre o preparo dos dados georreferenciados, preparo dos metadados para cadastro na plataforma GeoInfo, critérios de qualidade e boas práticas para inserção de dados, e aspectos de propriedade intelectual, incluindo o licenciamento dos dados. A Figura 1 ilustra o Macroprocesso para Publicação da Informação Geoespacial.

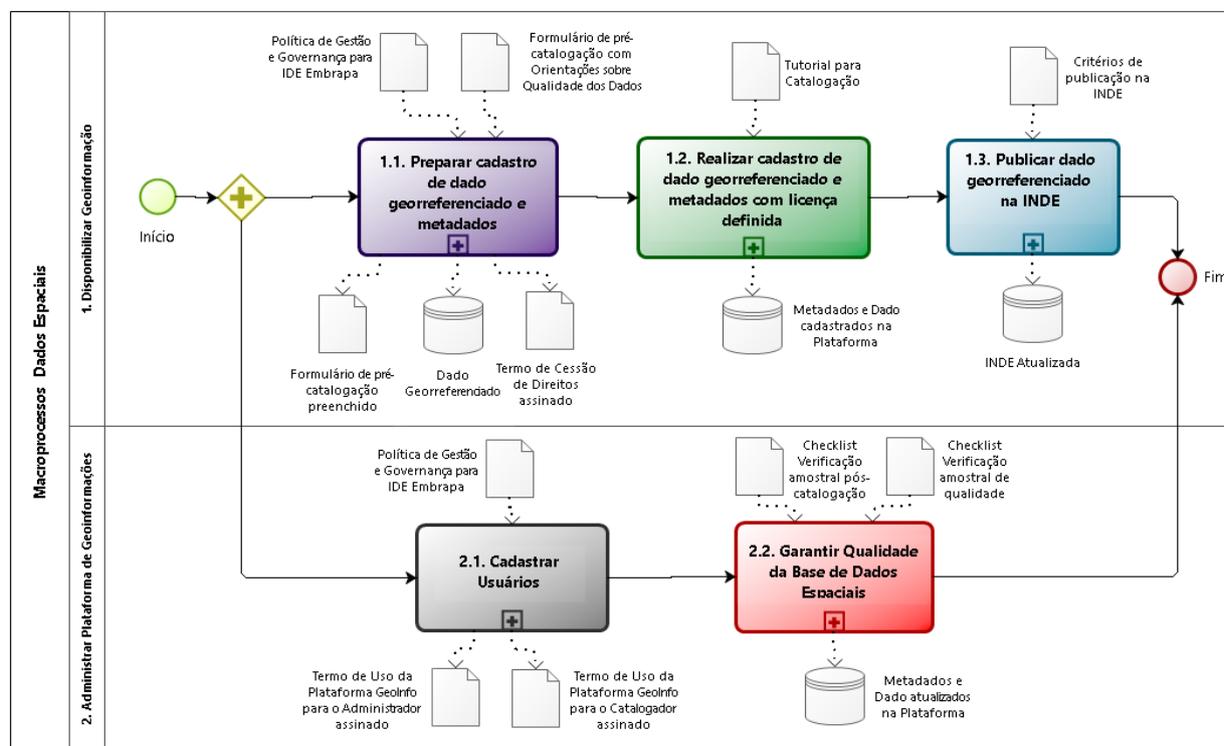


Figura 1 - Macroprocesso para Publicação da Informação Geoespacial

Foi proposta uma estrutura de governança da infraestrutura GeoInfo, composta de um Comitê Gestor, um Grupo de Trabalho para Referenciação da Geoinformação (GT-GeoRefer) e um Grupo de Trabalho para Manutenção da Infraestrutura Computacional (GT-GeoInfra). Foram estabelecidas regras para a composição dos grupos, incluindo formação, tempo de permanência e atribuições dos membros. Foram, ainda, levantados subsídios para a elaboração de uma Política de Gestão e Governança da Infraestrutura de Dados Espaciais Embrapa: GeoInfoⁱⁱⁱ.

Tecnologia

O arcabouço tecnológico do GeoInfo foi desenhado de forma a atender os diferentes perfis das 46 unidades da Embrapa e sua estrutura distribuída pelo país, considerando as necessidades de consumidores, produtores eventuais e grandes produtores de geoinformação. Para isso, foram definidas três arquiteturas de solução: (1) Infraestrutura Central: direcionada para Unidades cujo volume de produção de dados geoespaciais não justifica o empreendimento de esforços na manutenção do ambiente computacional ou na administração de usuários da plataforma; (2) Infraestrutura Virtual: direcionada para Unidades cuja produção de dados geoespaciais ocorre a partir de diferentes iniciativas, justificando o empreendimento de esforços para a administração de usuários, porém sem a necessidade de empreender esforços na manutenção do ambiente computacional; e (3) Infraestrutura Própria: direcionada para Unidades com

iii De Laia CR, Drucker, DP, Pinto, DM, Fidalgo, ECC, Simões M, Laforet MRC, Almeida BT, Custodio DO, Criscuolo C, Vitória DC. Subsídios para uma Política de Gestão e Governança da Infraestrutura de Dados Espaciais da Embrapa: GeoInfo. Rio de Janeiro; 2017. Documento institucional não publicado.

volumosa produção de dados geoespaciais e que desejam manter infraestrutura própria para catalogação e publicação dos dados espaciais. Tanto em (1) como em (2), um ambiente virtual, integrado à infraestrutura central, é disponibilizado para que as Unidades façam a catalogação e publicação de seus dados espaciais, sendo que em (2) a administração de usuários é realizada pela unidade detentora dos dados. Já a solução (3) facilita o acoplamento de workflows científicos de análise para a geração de novos conhecimentos a partir dos dados armazenados no repositório. A responsabilidade pela manutenção da infraestrutura e do software utilizado é da Unidade detentora dos dados. Cabe a equipe de administração da Estrutura Central auxiliar na implantação e integração das infraestruturas próprias com a infraestrutura central da Embrapa.

Em todas as soluções definidas, a responsabilidade pela gestão dos dados é da Unidade detentora. A equipe de administração da estrutura central presta auxílio no processo de implantação do catálogo e das estruturas virtuais e locais, e supervisão para promover melhoria da solução implantada. A supervisão consiste também na gestão das atualizações de software e aquisições de novos itens de hardware, bem como o suporte aos usuários. A integração de todos os metadados e serviços, cadastrados nas diferentes infraestruturas, é realizada por um agregador de metadados, através do processo de harvest. Esse agregador também é responsável pela ligação do GeoInfo com a INDE. Uma representação gráfica das diferentes modalidades e como se dá sua integração é apresentada na Figura 2.

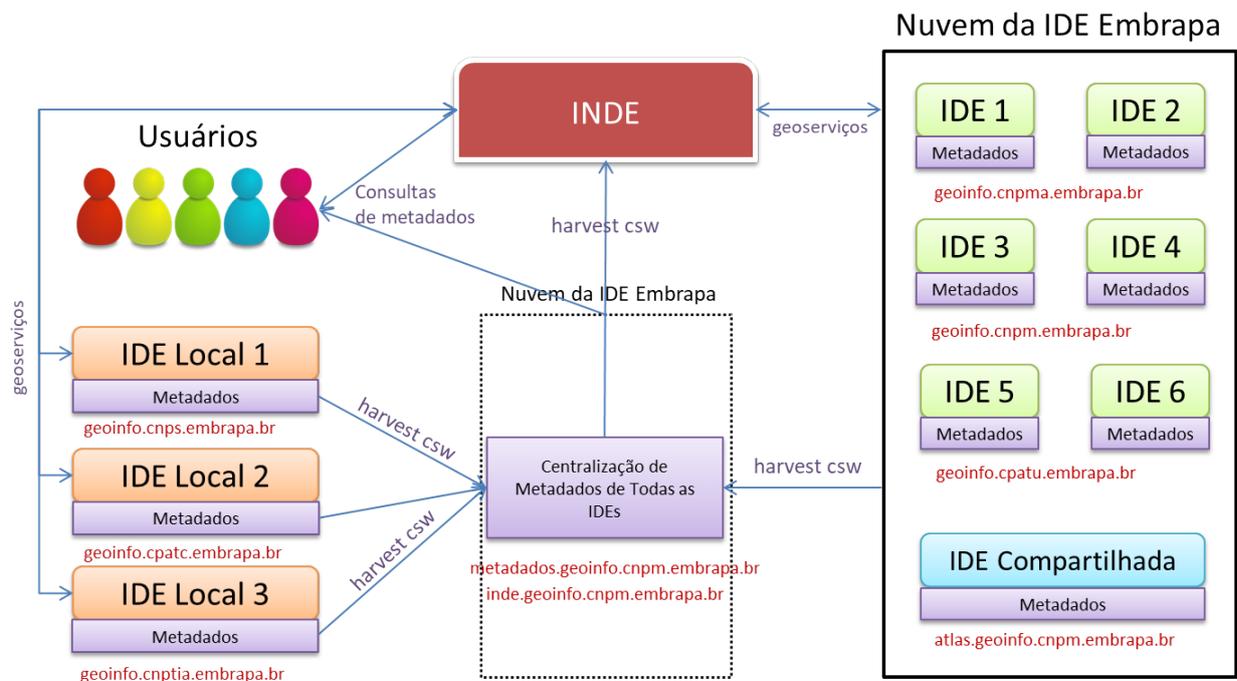


Figura 2 - Arquitetura da IDE-GeoInfo detalhando as relações entre as estruturas compartilhada e virtuais, hospedadas em uma nuvem privada, as estruturas locais, o agregador de metadados e a ligação com a INDE.

A implantação da estrutura da IDE-GeoInfo foi realizada utilizando softwares livres, que implementam os padrões OGC. Para tal foi utilizada a plataforma GeoNode^{iv}, versão 2.4, que integra um banco de dados geoespacial (PostGis) a um servidor de mapas (GeoServer) e a um catálogo de metadados (PyCSW), todos orquestrados por um Sistema de Gerenciamento de Conteúdo (Content Management System - CMS em inglês) em ambiente Web. Essa plataforma tem a vantagem de prover um ambiente amigável para a catalogação dos dados, orquestrando todas as ações necessárias para a publicação dos dados, metadados

iv Disponível em: <http://geonode.org/>

e geosserviços por uma simples interface web. Para o agregador de metadados foi utilizado o catálogo de metadados GeoNetwork^v, versão 3.0.4.

Ao todo, das 12 Unidades que participaram da implantação do GeoInfo, três optaram por utilizar infraestrutura própria: Embrapa Agrossilvipastoril, Embrapa Informática Agropecuária e Embrapa Solos. Oito unidades optaram por utilizar a estrutura virtualizada, onde um servidor dedicado foi posto à disposição da unidade: Embrapa Algodão, Embrapa Amazônia Oriental, Embrapa Clima Temperado, Embrapa Florestas, Embrapa Meio Ambiente, Embrapa Monitoramento por Satélite, Embrapa Pecuária Sudeste e Embrapa Tabuleiros Costeiros. Por fim, apenas uma Unidade, Embrapa Caprinos e Ovinos, optou pela infraestrutura central, por meio do cadastramento de seus dados em um repositório comum. As infraestruturas central e virtual, o agregador de metadados^{vi}, além de toda a administração do sistema ficou a cargo da Embrapa Monitoramento por Satélite. Estabeleceu-se um ambiente de desenvolvimento e testes a fim de realizar correções e melhorias necessárias na plataforma de forma continuada. Tais alterações são replicadas às estruturas central, virtual e própria utilizando a plataforma de gerenciamento de código GitLab^{vii}, instalado paralelamente a infraestrutura central. Além do sistema controlador de versão Git, que faz parte do GitLab, esse também possibilita a criação de documentação através de um WIKI e o registro de problemas (issues). Dessa forma, essa plataforma foi utilizada tanto para a documentação técnica (instalação, configuração, atualização) quanto o registro de problemas e necessidades de melhorias (issue track).

Dados

Os dados disponibilizados pelas unidades piloto apresentaram diversidade temática, de escala e de abrangência. Há unidades temáticas de pesquisa que atuam em todo o território brasileiro e que apresentaram dados espaciais relacionados ao seu tema principal, embora não se limitem a ele. Esses dados foram monitoramento de uso e cobertura das terras (Embrapa Monitoramento por Satélite), estudos da dinâmica espacial da agricultura (Embrapa Informática), mapeamento de solos (Embrapa Solos) e de uso sustentável do território pela agricultura (Embrapa Meio Ambiente). Nesses casos, as escalas e abrangência dos dados variaram de escala local, 1:1.000, a nacional, 1:5.000.000. Unidades de pesquisa de produtos em geral apresentaram dados espaciais associados aos locais de ocorrência desses produtos, sendo compostos por temas que caracterizam tais produtos, como produção, produtividade ou condições físicas e climáticas regionais para o desenvolvimento das culturas (Embrapa Algodão). Unidades de pesquisa conhecidas como ecorregionais apresentaram dados de escala e abrangência compatível com sua área de atuação. Nestes casos, os temas eram os de importância à realidade ecorregional, por exemplo, a dinâmica do uso e cobertura das terras em áreas desflorestadas (Embrapa Amazônia Oriental); caracterização dos solos, de atividades extrativistas e ocorrência de doenças (Embrapa Tabuleiros Costeiros); caracterização do meio físico e zoneamentos edafoclimáticos para culturas de expressão regional (Embrapa Clima Temperado). Algumas unidades apresentaram dados espaciais de suas fazendas ou campos experimentais (Embrapa Florestas, Embrapa Caprinos e Ovinos e Embrapa Clima Temperado), ou ainda de áreas experimentais, criadas para fins de pesquisa (Embrapa Pecuária Sudeste)²⁶. As informações mais comuns a respeito dessas áreas são sua localização, descrição dos sistemas de uso das terras e caracterização física, como a identificação das diversas unidades de solos, visando a sua utilização em pesquisas nas áreas experimentais. A Figura 3 ilustra um exemplo de conjunto de dados geoespaciais catalogados na plataforma GeoInfo, com uma visualização simplificada dos respectivos metadados.

v Disponível em: <http://geonetwork-opensource.org/>

vi Disponível em: <http://metadados.geoinfo.cnpm.embrapa.br>

vii Disponível em: <http://gitlab.com>



Figura 3 Mapa de solos do estado do Rio de Janeiro, em escala 1:250.000, e visualização simplificada dos respectivos metadados.

Os metadados completos podem ser visualizados pela interface web ao acionar o botão “Metadados da Camada”, bem como podem ser consumidos por serviço web ou por download do arquivo XML (eXtensible Markup Language), nos formatos ISO, FGDC, eBRIM, Dublin Core, DIF ou Atom.

Normas e Padrões

O processo de elaboração do formulário de descrição do dado espacial contou com um grupo de trabalho composto por 11 pessoas, representantes das Unidades Piloto, as quais foram encarregadas de analisar e implementar ações voltadas para a descrição do dado espacial²⁷. Foram utilizados como referências: Norma ISO 19115:2003²⁸, Perfil MGB²⁹ e Documentação do software GeoNode³⁰. Para a descrição dos dados espaciais foi estruturado um formulário de pré-catalogação, para preenchimento pelo autor, composto por 15 campos, dos quais 12 são de preenchimento obrigatório. Produziu-se, ainda, uma tabela comparativa entre o perfil MGB e o perfil padrão do GeoNode, que evidenciou as interfaces e necessidades de ajustes pela Embrapa. Foram criados novos atributos: Ano de criação do Conjunto de Dados Geográficos (CDG), Lista de Autores, e realizada alteração de nomenclatura em outros atributos para facilitar o preenchimento pelos usuários. Um manual para descrição do dado espacial na Embrapa foi produzido²⁷.

Os principais tesouros usados na Embrapa para a indexação e recuperação de documentos não contemplavam adequadamente termos sobre a área da geoinformação, em particular no idioma português. Assim, buscou-se operacionalizar aspectos inerentes à gestão de dados de pesquisa, perpassando por entendimentos e a aplicabilidade de conceitos e ferramentais de gestão da informação e do conhecimento³¹. O vocabulário controlado foi compilado a partir de textos científicos produzidos por pesquisadores e analistas da Embrapa e publicados em eventos científicos importantes na área de geotecnologias aplicadas

à agricultura. As etapas consistiram primeiramente na limpeza do corpus, onde os arquivos originais, em diversos formatos, foram convertidos para txt (UTF-8). Em seguida foi realizada limpeza no conteúdo desses arquivos, que consistiu na remoção de agradecimentos, links, autoria, filiação, notas de título e números de página. Foram mantidas as referências (bibliográficas) porque os títulos normalmente contêm termos técnicos relevantes. O resultado foi um corpus com 1.118 páginas de texto e 586.945 palavras. A etapa seguinte foi a extração dos candidatos a termo e envolveu as etapas de (1) detecção do idioma, (2) remoção de sentenças em inglês e (3) extração das listas de candidatos a termo. A primeira lista conteve 7.839 candidatos a termo, que foram limpos de forma manual pela equipe até chegar à lista final, com 799 candidatos a termos. A equipe criou, manualmente, os relacionamentos entre hiperônimos e hipônimos para todos os 799 termos, e gerou uma lista tripla. Esse processo pautou-se em proposta de metodologia de organização e engenharia do conhecimento³². A lista de termos resultante foi disponibilizada na ferramenta GeoNode, para validação por parte da equipe do projeto. Além dos termos do lista disponível, o usuário pode, em um campo de termos livres, incluir outras palavras que representem a informação ali catalogada e que não esteja contemplada na lista. A análise futura desses termos inseridos poderá retroalimentar o vocabulário desenvolvido.

Discussão

A estratégia adotada para viabilizar a implantação do GeoInfo, a Infraestrutura de Dados Espaciais da Embrapa, foi bem sucedida, considerando que foi estabelecida e implementada uma iniciativa para organizar, preservar, descrever, qualificar e ofertar a geoinformação gerada pela Embrapa de forma a ampliar o potencial de aplicação dessa informação na produção e difusão de conhecimento e inovação. A plataforma criada está em consonância com os padrões preconizados pela INDE e viabiliza a interoperabilidade de dados heterogêneos provenientes de diferentes fontes, inclusive sob o aspecto semântico. A lógica de estabelecer um processo assegura a perpetuação da iniciativa ao longo do tempo, com novos dados sendo gerados, descritos e publicados. Através da plataforma Geoinfo, o retrabalho e redundâncias na produção de dados espaciais é evitado, possibilitando sua integração a outras informações disponíveis em nosso país através de uma infraestrutura única.

Alguns fatores foram fundamentais para o sucesso da iniciativa, com destaque para o fato dos resultados terem sido obtidos a partir do engajamento de diferentes atores relevantes ao processo de produção de dados espaciais no contexto dessa instituição de pesquisa agropecuária, constituindo uma equipe interdisciplinar e uma construção coletiva, o que tem se mostrado um fator de sucesso em outras iniciativas⁹. Além disso, estratégia de iniciar a construção da IDE a partir de uma amostra representativa do universo de unidades da Embrapa foi fundamental para permitir maior proximidade e interação entre os participantes, bem como para viabilizar a validação da solução digital, a qual foi concebida para ser escalável. Ações de comunicação e capacitação foram fundamentais para aproximar os envolvidos, tanto no contexto do público interno de uma instituição formada por uma rede de centros de pesquisa, quanto considerando o público externo e suas demandas variadas. A construção coletiva do processo para a publicação de dados geoespaciais a partir da experiência das unidades piloto requereu a elaboração de várias versões do processo até o desenho final, contendo a versão mais simplificada possível. Uma série de documentos de apoio foram elaborados, englobando aspectos de qualidade de dados e metadados, propriedade intelectual, licenciamento, política e descrição de metadados. Sob o aspecto tecnológico, a adoção de ferramentas de código aberto foi consonante com o propósito de publicar dados abertos e a flexibilidade da arquitetura implantada forneceu, ao mesmo tempo, autonomia para as unidades com diferentes realidades e possibilidade de agregação do acervo de dados espaciais em uma infraestrutura única, facilitando a recuperação da informação pelo usuário. Dados espaciais relevantes para a pesquisa agropecuária e para a sociedade foram catalogados e disponibilizados

ao público sob o formato aberto, interoperáveis e consumíveis por máquinas e humanos. Padrões de metadados foram analisados de forma a melhor descrever o dado espacial. Uma terminologia para dados geográficos foi elaborada para representar adequadamente a geoinformação.

Dentre os fatores de sucesso na implantação de uma IDE, conforme a experiência de iniciativas internacionais que largaram na frente nesse esforço, está a mudança institucional quanto à cultura de documentação dos dados geoespaciais, através de adoção de padrão de metadados⁹. Podemos afirmar que esse esforço foi realizado através da IDE GeoInfo, trazendo o desafio de perpetuar esse alcance nas próximas etapas. Além disso, os autores apontam um fator que deve ser revisitado continuamente para o sucesso da iniciativa, o da conscientização da gestão e do corpo técnico, de forma crescente e permanente, quanto à necessidade de conhecer e compor seus catálogos de dados e informações para integração e análise objetivando a tomada de decisão e o avanço do conhecimento. Consideramos que esse é um fator que deve ser revisitado continuamente para o sucesso da iniciativa. O fator seguinte é a implementação e incorporação gradativa (sistemática e permanente) de catálogos de dados e metadados, segundo padrão internacional, das bases geoespaciais existentes; o que será o próximo passo da presente iniciativa, ao incorporar outras unidades da instituição que ainda não participam do GeoInfo. Por fim, o fator de implantação de mecanismos de busca e acesso, segundo padrão de dados e metadados, que propicie a usabilidade dos dados em nível nacional, regional e internacional, foi alcançado plenamente ao se adotar o ferramental de código aberto descrito no presente trabalho.

Considerações Finais

A integração consistente de dados espaciais oriundos de diversas fontes requer conhecimento de conceitos, normas e especificações inerentes aos dados e às aplicações a que se destinam⁹. No âmbito da Embrapa, a estratégia de adesão à INDE se deu pela implantação de uma Infraestrutura de Dados Espaciais, a IDE GeoInfo, internalizando um processo para o armazenamento, a organização, a preservação, a recuperação e a disseminação da geoinformação.

Neste trabalho, apresentamos o complexo esforço de construção do GeoInfo, a Infraestrutura de Dados Espaciais da Embrapa. O repositório de dados geoespaciais implantado permite que o usuário possa encontrar e navegar pelos dados e respectivos metadados relacionados com a agricultura no Brasil, permitindo a integração consistente de dados oriundos de diversas fontes pela adoção de padrões de interoperabilidade, inclusive sob o aspecto semântico. Foram empreendidos esforços para subsidiar a representação e a descrição da geoinformação pelos metadados. A experiência de implantação nas unidades piloto forneceu subsídios para a definição de processos para implantação nas demais unidades da empresa. Foi definido um processo para a gestão do ciclo de vida dos dados geoespaciais gerados pela Embrapa, de forma a perpetuar o fornecimento de serviços e informações para a Embrapa e para a sociedade, em consonância com a legislação vigente.

Para o usuário externo, é possível interagir com os dados por meio de consultas e de algumas ferramentas da plataforma, e explorar os metadados em uma página web de interface amigável, bem como por download do arquivo XML ou consumo por geoserviço, ou seja, consumível por máquina. Sob o ponto de vista interno à instituição, o processo de publicação de dados e metadados geoespaciais foi interdisciplinar e envolveu a atividade de profissionais de ciências agrárias, ciência da informação, ciência de dados, geografia, entre outros. Esse processo levou ao entrosamento da equipe e conhecimento de todas as etapas de publicação do dado pelos envolvidos.

Assim, a iniciativa da IDE GeoInfo segue as diretrizes estabelecidas pela iniciativa FAIR para a gestão de dados científicos (Findability, Accessibility, Interoperability, and Reusability)³³, considerando a necessidade de promover o reuso desses dados. A IDE GeoInfo está inserida em um contexto no qual a exposição de

dados pelo compartilhamento em repositórios leva ao aumento de citações dos artigos originais aos que os mesmos estão associados³⁴ e está em consonância com a crescente tendência à valorização de todos os produtos da pesquisa científica, em especial os dados^{33,35}. Acompanhar e participar do processo de ganho de importância de repositórios institucionais de dados científicos³⁶ no contexto acadêmico é um desafio para a evolução futura da presente iniciativa, incluindo a certificação desse repositório de dados científicos de acordo com critérios mundialmente adotados^{37,38}.

O componente espacial é essencial para a integração de dados da agropecuária. Com isso, diversos conjuntos de dados geoespaciais são gerados e analisados no contexto das pesquisas conduzidas pela Embrapa. Com a implantação da Infraestrutura de Dados Espaciais da Embrapa - GeoInfo, dados gerados tanto por observações ou levantamentos de campo, quanto por processamento de dados de sensoriamento remoto ou sensores de campo podem ser acessados, compreendidos, integrados e reusados, viabilizando novas possibilidades de análise e a geração de novo conhecimento relevante para a agropecuária e a sociedade brasileira.

Esses processos beneficiam a Embrapa para a adequação à legislação sobre gerenciamento de dados geoespaciais vigentes no Brasil. Ao mesmo tempo, a experiência adquirida permitirá que a empresa contribua com o aprimoramento de processos no âmbito da INDE, e os aprendizados aqui relatados podem ser úteis para o estabelecimento de iniciativas semelhantes em outras instituições.

Referências

1. Borgman CL. Scholarship in the digital age: information, infrastructure and internet. Cambridge: MIT Press; 2007.
2. Gray J. Jim Gray on eScience: a transformed scientific method. In: Hey T, Tansley S, Tolle K (Ed.). The fourth paradigm: data-intensive scientific discovery. Redmond: Microsoft Research; 2009.
3. Sayão LF, Sales LF. Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores. Rio de Janeiro: Comissão Nacional de Energia Nuclear; 2015.
4. Huisman O, By RA (Ed.). Principles of geographic information systems: an introductory textbook. The Netherlands: ITC; 2009. Disponível em: www.itc.nl/library/papers_2009/general/PrinciplesGIS.pdf
5. Simões MG, Oliveira SEM, Ferraz RPD, Santos HG, Manzatto CV. Democratização da informação de solos do Brasil: Geoportal e banco de dados de solos com acesso via web. Cadernos de Ciência & Tecnologia [Internet]. 2015 Jan./Ago. [citado em 19 set. 2017]; 32(1/2):55-69. Disponível em: <https://seer.sct.embrapa.br/index.php/cct/article/view/23295>
6. Brasil. Lei n. 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei no 8.112, de 11 de dezembro de 1990; revoga a Lei no 11.111, de 5 de maio de 2005, e dispositivos da Lei no 8.159, de 8 de janeiro de 1991; e dá outras providências. Diário Oficial da União, Brasília, DF, 18 nov. 2011. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm
7. Brasil. Decreto n. 8.777, de 11 de maio de 2016. Institui a Política de Dados Abertos do Poder Executivo federal. Diário Oficial da União, Brasília, DF, 12 maio 2016. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/d8777.htm
8. Brasil. Decreto n. 6.666, de 27 de novembro de 2008. Institui, no âmbito do Poder Executivo federal, a Infra-Estrutura Nacional de Dados Espaciais - INDE, e dá outras providências. Diário Oficial da União, Brasília, DF, 28 nov. 2008. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2008/Decreto/D6666.htm
9. Brasil. Ministério do Planejamento, Orçamento e Gestão. Comissão Nacional de Cartografia. Plano de ação para implantação da Infraestrutura Nacional de Dados Espaciais. Rio de Janeiro: Concar; 2010.
10. Brasil. Ministério do Planejamento, Orçamento e Gestão. Comissão Nacional de Cartografia. Perfil de metadados geoespaciais do Brasil: perfil MGB: versão homologada. Rio de Janeiro: Concar; 2011.

11. GeoPortal Digital. Rio de Janeiro: Embrapa Solos, 2008. Citado em 19 set 2017. Disponível em: <http://mapoteca.cnps.embrapa.br>
12. Biblioteca Geoespacial. São Carlos: Embrapa Meio Ambiente, 2014. Citado em 19 set 2017. Disponível em: <http://geo.cnpma.embrapa.br/>
13. Agritempo: sistema de monitoramento agrometeorológico. Campinas: Embrapa Informática Agropecuária, 2014. Citado em 19 set 2017. Disponível em: <http://www.agritempo.gov.br/agritempo/index.jsp>
14. Banco de produtos Modis. Campinas: Embrapa Informática Agropecuária, 2013. Citado em 19 set 2017. Disponível em: <https://www.modis.cnptia.embrapa.br/geonetwork/srv/pt/main.home>
15. Almeida CA, Coutinho AC, Esquerdo JCDM, Adami M, Venturieri A, Diniz CG, Dessay N, Durieux L, Gomes AR. High spatial resolution land use and land cover mapping of the Brazilian Legal Amazon in 2008 using Landsat-5/TM and MODIS data. *Acta Amazonica*, 2016. Citado em 19 set, 2017, 46(3)291-302. Disponível em: <https://acta.inpa.gov.br/fasciculos/46-3/PDF/AA-2015-0550.pdf>
16. GeoPortal Digital. Sete Lagoas: Embrapa Milho e Sorgo, 2014. Citado em 19 set 2017. Disponível em: <http://geoportal.cnpms.embrapa.br/>
17. Macrozoneamento Ecológico-Econômico do Estado do Maranhão [Internet]. Campinas: Embrapa Monitoramento por Satélite, 2014. Citado em 19 set 2017. Disponível em: <https://www.cnpm.embrapa.br/projetos/macrozee/>
18. SomaBrasil. Campinas: Embrapa Monitoramento por Satélite, 2014. Citado em 19 set 2017. Disponível em: <http://mapas.cnpm.embrapa.br/somabrasil/webgis.html>
19. Bayma-Silva G, Nogueira SF, Drucker DP, Siqueira GR, Fernandes RM, Custodio DO, Alvarenga SVA. Plataforma web para sistemas de informação geoespacial (SIG): aplicações no projeto Geodegrade. In: *Simpósio Brasileiro de Sensoriamento Remoto*; São José dos Campos: INPE; 2015.
20. Mengatto Junior, E. A.; Silva, J. dos S. V. da; Oliveira, R. C. de Ferramenta técnica como suporte a gestão do território. In: *Simpósio de Geotecnologias no Pantanal São José dos Campos*: INPE; Brasília, DF: Embrapa, 2016. p. 856-864.
21. Drucker DP, Custodio DO, Fidalgo ECC, Daltio J, Visoli MC. Preservação e organização da geoinformação em instituições: o caso da construção da infraestrutura de dados espaciais da Embrapa. In: *Simpósio Brasileiro de Sensoriamento Remoto*; São José dos Campos: INPE; 2015.
22. Faundeen JL, Burley TE, Carlino JA, Govoni DL, Henkel HS, Holl SL, Hutchison VB, Martín E, Montgomery ET, Ladino CC, Tessler S, Zolly LS. The United States Geological Survey Science Data Lifecycle Model. Virginia: U. S. Department of the Interior; U.S. Geological Survey; 2013. Disponível em: <https://pubs.usgs.gov/of/2013/1265/pdf/of2013-1265.pdf>
23. Strasser C, Cook R, Michener W, Budden A, Koskela R. DataONE: promoting data stewardship through best practices. In: *Proceedings of the Environmental Information Management Conference*; California: University of California; 2011. p. 126–131. Disponível em: <https://eim.ecoinformatics.org/eim2011/eim-proceedings-2011>
24. Warnest M. A collaboration model for national spatial data infrastructure in federated countries (dissertação). Melbourne: University of Melbourne; 2005.
25. Open Geospatial Consortium. Citado em 19 set. 2017. Disponível em: <http://www.opengeospatial.org/>
26. Bettiol GM, Moura L, Nogueira SF, Drucker DP, Oliveira PPA, Quartaroli CF, Rodrigues CAG, Silva GBS. Web platform for geospatial information: applications for geopecus project. In: *Simpósio Internacional sobre Gases de Efeito Estufa na Agropecuária*; Campo Grande: Embrapa Gado de Corte; 2016.
27. Pinto DM, Brandão VVS, Drucker DP, Dompieri MHG, Rasche F, Holler WA, Bettiol GM, Custódio DO, Victoria DC, Gonçalves LMPB. Descrição dos dados da pesquisa geoespacial: a experiência da Embrapa. *Ref. Eletron Comun Inf Saúde*. 2017 nov. [citado em 30 nov 2017]; 11(sup):1-17. Disponível em: <<https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/1418>>
28. International Organization for Standardization (ISO). *Geographic Information – Metadata*: ISO 19115:2003. London, England; 2003. 152 p.
29. Comitê de Estruturação de Metadados Geoespaciais. *Perfil de metadados geoespaciais do Brasil: perfil MGB*. Rio de Janeiro: CEMG-Concar; 2009.

30. GeoNode. Open Source Geospatial Content Management System. Citado em 19 set. 2017. Disponível em: <http://geonode.org/>
31. Pierozzi Junior I, Souza M IF, Torres TZ, Oliveira LHM, Queiros LR. Gestão da informação e do conhecimento. In: Massruhá SMFS, Leite MAA, Luchiari Junior A, Romani LAS (Ed.). Tecnologias da informação e comunicação e suas relações com a agricultura. Brasília, DF: Embrapa; 2014. p. 235-258.
32. Di Felippo A, Aluísio SM, Oliveira LHM, Almeida GMB. OntoMethodus: a methodology to build domain-specific ontologies and its use in a system to support the generation of terminographic products. In: Workshop em Tecnologia da Informação e da Linguagem, Espírito Santo: Universidade Federal do Espírito Santo; 2008.
33. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, Santos LBS, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, Lei J, Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons, B. The FAIR Guiding Principles for scientific data management and stewardship (Internet). 2016. Sci. Data 3:160018 doi: 10.1038/sdata.2016.18
34. Piwowar HA, Day RS, Fridsma DB. Sharing detailed research data is associated with increased citation rate. PloS one. 2007 jan. Citado em 15 set 2017;2(3). Disponível em: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000308>
35. Kratz JE, Strasser, C. Making data count. Sci Data (internet), 2015. Acesso em 15 set 2017; 2(150039). Doi: 10.1038/sdata.2015.39.
36. Bjork K, Isaak D, Vyhnanek K. The changing roles of repositories: where we are and where we are headed. 2013. Citado em 15 set 2017. Disponível em: <https://research.libraries.wsu.edu:8443/xmlui/handle/2376/4357>
37. DDC: because good research needs good data. Trustworthy repositories. Edinburgh, 2017. Citado em 15 set 2017. Disponível em: <http://www.dcc.ac.uk/resources/repository-audit-and-assessment/trustworthy-repositories>
38. Husen SE, Wilde ZG, Waard A, Cousijn H. Recommended versus certified repositories: mind the gap. Data Science Journal, 2017. Citado em 15 set 2017;16. Disponível em: <https://datascience.codata.org/articles/10.5334/dsj-2017-042/>