

GEOINFO: REPOSITÓRIO DE DADOS ESPACIAIS DA EMBRAPA

Daniela Maciel Pinto
Debora Pignatari Drucker
Júlio César Dalla Mora Esquerdo
Juliana Meireles Fortaleza
Margareth Goncalves Simões
Francisca Rasche
Ricardo de Oliveira Dart
Patrícia Rocha Bello Bertin
Davi de Oliveira Custódio

Empresa Brasileira de Pesquisa Agropecuária

RESUMO

O "GeoInfo - Infraestrutura de Dados Espaciais da Embrapa" é o repositório de dados espaciais da Embrapa, criado a partir das considerações concernentes ao acesso e à ciência aberta, e, em atendimento ao Decreto Lei n. 6.666/08, que instituiu a Infraestrutura Nacional de Dados Espaciais (Inde). A Embrapa é uma instituição de pesquisa aplicada, pública, com foco na geração de soluções para a agropecuária brasileira. Estabelecido como um exemplo bem sucedido de iniciativa, o GeoInfo tem influenciado a Embrapa a promover a abertura de seus dados espaciais gerados no âmbito de suas pesquisas. Este capítulo visa explicitar a experiência da Embrapa na implantação de seu repositório de dados espaciais e os desafios relacionados a um ambiente de produção de dados de pesquisa, o que pressupõe uma mudança de paradigma numa instituição com 42 centros de pesquisa distribuídos pelo país, cada um gerando dados e informações de conteúdo e em contextos bastante diversificados.

Palavras-chave: Dados de pesquisa. Gestão de dados de pesquisa. Agricultura brasileira. Infraestrutura de Dados Espaciais. Agronegócio.

ABSTRACT

The GeoInfo - Embrapa Spatial Data Infrastructure" is a repository that accommodates Embrapa geospatial data, created under the premises of open access and open science, and in compliance with the Decree Law n. 6666/08, which instituted the National Spatial Data Infrastructure (Inde). Embrapa is a public applied research company, focused on generating solutions for Brazilian agriculture and livestock. Established as a successful example of an initiative, GeoInfo has influenced Embrapa to promote the opening of its spatial data generated in the scope of its research. This chapter aims to explain Embrapa's experience in the implementation of its spatial data repository and the challenges related to a production environment of research data, which presupposes a paradigm shift in an institution with 42 research centers distributed across the country, each one generating data and information of quite diversified content and in differentiated contexts.

Keywords: Research data. Research data management. Brazilian agriculture. Geospatial data. Agribusiness.

1 INTRODUÇÃO

A complexidade do processo de geração e de aplicação da geoinformação, que se caracteriza pela onerosidade, pelo caráter decisório e por um alto grau de especialização, fez surgir iniciativas em todo o mundo para estimular o compartilhamento de dados e de informações geoespaciais. Na década de 1990, tornaram-se notáveis os esforços dos países em organizar e em compartilhar a geoinformação por meio da construção das Infraestruturas de Dados Espaciais (IDEs), que, neste texto, serão tratadas como sinônimo de repositório de dados espaciais. Também, neste período, destaca-se a criação do *Open Geospatial Consortium* (OGC), cujo objetivo é promover o estabelecimento de padrões que facilitam o intercâmbio de informações espaciais. Atuando na diplomacia de agregar setores públicos e privados de países que consomem e produzem a geoinformação, o OGC foi e é essencial para o aproveitamento e a consolidação de IDEs nacionais, capazes de congregando dados espaciais de diversas instituições produtoras e consumidoras, disponibilizando-os à sociedade.

Foi diante dessa condição que, no Brasil, em 2008, estabeleceu-se a primeira política pública (PP) relacionada a uma IDE nacional. O Decreto-Lei n. 6.666/08 (Brasil, 2008) instituiu, no país, a Infraestrutura Nacional de Dados Espaciais (INDE) e tornou obrigatória a todos os órgãos públicos (das esferas municipal, estadual e federal) a disponibilização de sua geoinformação.

Engajada nessa política pública e atenta às novas tendências de comunicação científica, a Empresa Brasileira de Pesquisa Agropecuária (Embrapa) reuniu esforços e estabeleceu seu repositório de dados de pesquisa espacial, o GeoInfo. Tais esforços representam o reconhecimento da empresa frente aos movimentos do Acesso Aberto e da Ciência Aberta e à importância do uso e da aplicação das geotecnologias, dos dados e das informações espaciais para se compreender a dinâmica da agricultura no território brasileiro. Com a adoção institucional do GeoInfo, e sua integração à INDE, a Embrapa passou a organizar, a preservar, a qualificar, a gerenciar e a disponibilizar os dados espaciais resultantes de suas pesquisas, provendo-os à sociedade para usos diversos.

Este capítulo visa explicitar a experiência da Embrapa na implantação de seu repositório de dados de pesquisa espacial e os desafios relacionados a um ambiente de produção de dados, o que pressupõe uma mudança de paradigma numa instituição com 42 centros de pesquisa distribuídos pelo país, cada um gerando dados e informações de conteúdo bastante diversificado e em contextos diferenciados.

2 REPOSITÓRIO DE DADOS DE PESQUISA ESPACIAL DA EMBRAPA

2.1 Desenvolvimento e institucionalização do GeoInfo na Embrapa

Para consolidar o “GeoInfo - Infraestrutura de Dados Espaciais da Embrapa” <embrapa.br/geoinfo>, a primeira ação da Embrapa foi aprovar, em 2011, um projeto, intitulado “GeoInfo” para a definição de um modelo de gestão de dados geoespaciais. A partir da articulação de atores internos, relevantes à temática da produção e da gestão de dados espaciais, e do estudo de tecnologias e de ferramentas para a implantação de um repositório de dados espaciais, a Diretoria da empresa decidiu formalizar sua adesão à INDE, em julho de 2013. Em seguida, foi constituído um grupo de trabalho com a função de analisar estrategicamente a estruturação da IDE-Embrapa

e produzir o “Plano de Implantação do nó da Embrapa na INDE”, que, posteriormente, resultou na criação do repositório de dados espaciais, o GeoInfo.

Intitulado de “GT-INDE” (Grupo de Trabalho para implantação do nó da Embrapa na INDE), o grupo atuou de forma integrada ao decreto-lei e ao complexo universo da Embrapa, que engloba, atualmente, 42 centros de pesquisa, espalhados pelo território brasileiro, ora chamados também de “Unidades Descentralizadas (UDs)”.

Com diferentes condições, as Unidades foram estudadas pelo GT-INDE, que levantou informações e requisitos essenciais para indicar o modelo mais adequado para que a Embrapa passasse a integrar a INDE. Nesse processo decisório, considerou-se também as tendências relacionadas ao paradigma da *e-Science*, termo adotado para pesquisas que lidam com grandes volumes de dados ou que fazem uso de métodos computacionais sofisticados e de computação de alto desempenho. Para estabelecer o plano de implantação e para efetivar as indicações feitas pelo GT-INDE junto às UD's da Embrapa, foi elaborado o projeto “Implantação da Infraestrutura de Dados Espaciais da Embrapa (IDE-Embrapa)”, que reuniu um total de 12 unidades-piloto, as quais foram responsáveis pela construção e pela implementação do processo de gestão de dados espaciais da Empresa, operacionalizado por meio do repositório “GeoInfo - Infraestrutura de Dados Espaciais da Embrapa”.

O GeoInfo, portanto, é um repositório destinado à inserção, ao tratamento e à disponibilização de arquivos vetoriais e matriciais, que permite a inclusão de documentos de diferentes extensões (Geotiff, mapas em pdf, tabelas e outros documentos cartográficos). Por seguir os protocolos nacionais e internacionais para o compartilhamento de dados da pesquisa espacial, potencializando o reuso e a ampliação da pesquisa por parceiros e pela sociedade em geral, está associado aos conceitos de: *e-Science*, Ciência Aberta e Acesso Aberto. Foi lançado para a sociedade em 2018 e, antes desse lançamento, passou a contar com um colegiado denominado “Comitê Gestor do GeoInfo (CG-GeoInfo), responsável por acompanhar e por implementar tendências, bem como gerenciar, planejar e executar ações relativas à gestão de dados de pesquisa espacial na Embrapa. Esse comitê articulou, com apoio da Secretaria de Desenvolvimento Institucional (SDI), a adoção corporativa do repositório e interação com as UD's, assistindo-as e orientando-as quanto ao processo de gestão de dados espaciais no GeoInfo.

Com a adoção institucional desse repositório e do processo de gestão de dados espaciais, a Embrapa passou a gerenciar sua produção científica de dados de pesquisa espacial, antes dispersa, a racionalizar os seus esforços na geração de conhecimento e a disponibilizar sua geoinformação, colaborando, assim, com as ações de Pesquisa, Desenvolvimento e Inovação (PD&I) de instituições e de pesquisadores que possuem desafios semelhantes aos da Empresa. Atualmente, o GeoInfo foi designado como ambiente para comprovação do indicador “Transferência Ativa” do Relatório de Desempenho de Gestão das Unidades da Embrapa.

2.2 Arquitetura do repositório

Para atender ao Decreto-Lei nº 6.666/2008 (Brasil, 2008), tendo em vista o novo paradigma da comunicação científica (*e-Science*), foi imprescindível o envolvimento de diversos atores de diferentes áreas para estruturar o processo de gestão de dados de pesquisa espacial (GDPE) da Embrapa e o repositório GeoInfo. Cerca de 70 pessoas, de 12 unidades e uma secretaria de comunicação, além da participação de membros do Instituto Brasileiro de Geografia e Estatística (IBGE), estiveram envolvidos na consolidação desta iniciativa que exigiu, em um primeiro momento, a identificação da arquitetura de *softwares* mais adequada para o estabelecimento do GeoInfo.

A partir das indicações e dos requisitos levantados pelo GT-INDE, optou-se pelo uso dos *softwares open source*, GeoNode e GeoNetwork (Custódio *et al.*, 2018; Drucker *et al.*, 2017). Essa composição segue um modelo arquitetônico organizado em servidores de serviços e servidores de dados (Figura 1), tal como proposto por Café *et al.* (2007).

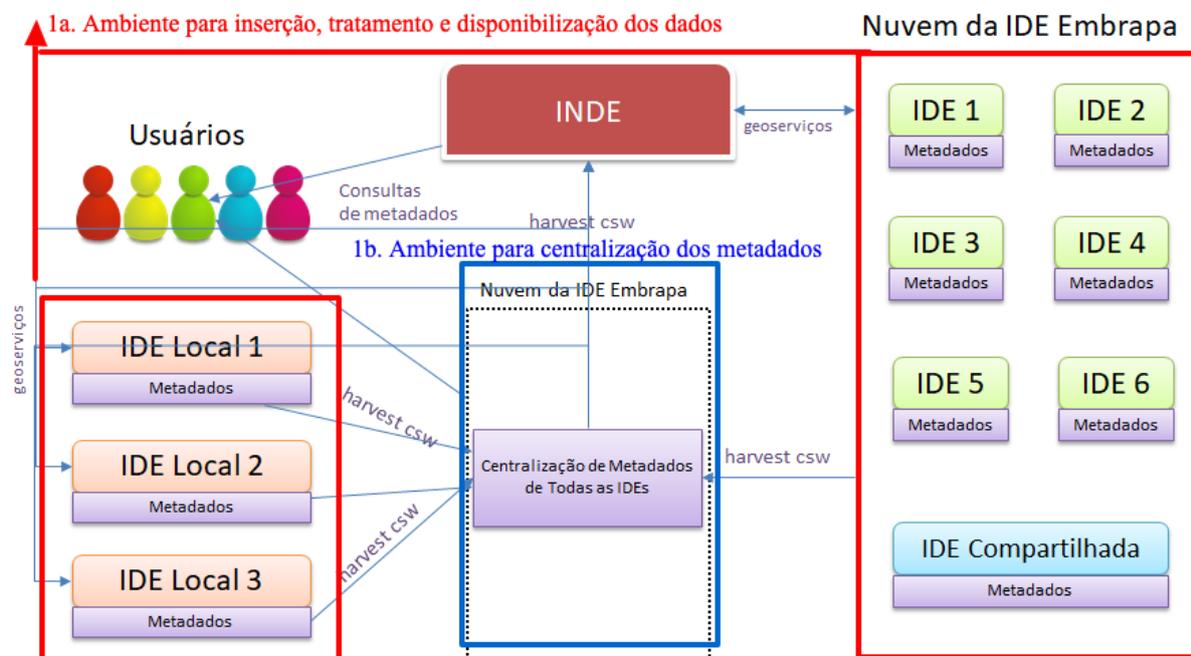


Figura 1. Arquitetura computacional para consolidação das bases de dados e de centralização dos metadados. Fonte: Adaptado pelos autores a partir de Custódio *et al.* (2018).

Por meio dessa estrutura, tem-se um ambiente para inserção e tratamento dos dados (Figura 1a, em vermelho), onde é possível gerenciar permissões de acessos e consolidar meios para a visualização espacial. Esse ambiente é controlado, de forma individualizada⁵⁴, por cada UD produtora de geoinformação da Embrapa, e é operacionalizado por meio do *software* GeoNode⁵⁵ <<http://geonode.org>>. Para configurar a integração entre as bases de dados das diversas UD, tem-se o ambiente de centralização de metadados (Figura 1b, em azul), construído como um agregador a partir do *software* GeoNetwork <<https://geonetwork-opensource.org/>>. Ademais, o GeoInfo é composto por um Sistema Gerenciador de Banco de Dados geoespacial (PostGIS), servidores de mapas (Geoserver) e de metadados (PyCSW), todos embarcados no *software* GeoNode.

Para lidar com a variedade de demandas dos diferentes centros de pesquisa da empresa, foi implementada uma arquitetura computacional flexível, a qual possibilita que cada UD escolha a condição que mais se adapta à sua realidade. Nesse aspecto, foram propostas três modalidades denominadas: Central, Virtual e Local. A modalidade Central é indicada às UD cujo volume de produção de dados geoespaciais não justifica o empreendimento de esforços na manutenção do ambiente computacional ou na administração das bases componentes de uma IDE; já a modalidade Virtual é direcionada para UD cuja produção de dados geoespaciais ocorre a partir de diferentes iniciativas, justificando o empreendimento de esforços para a administração das bases que compõem uma IDE, porém sem a necessidade de empreender esforços na manutenção do ambiente computacional; e a modalidade Local é direcionada para UD com volumosa produção

54 Neste caso, cada UD possui seu próprio repositório, ou IDE.

55 Aplicação Free Open Source Software (FOSS) desenvolvida pela OpenGeo, em parceria com o Banco Mundial.

de dados geoespaciais e que desejam manter infraestrutura própria para catalogação e para publicação dos dados espaciais. As modalidades Central e Virtual fazem uso de uma infraestrutura de nuvem (Figura 1) por meio de um *datacenter* disponibilizado na Embrapa Territorial, cabendo a esta UD a manutenção, a atualização e a evolução das IDEs que compõem o GeoInfo nessas duas modalidades. A gestão das IDEs nessas modalidades ocorre de forma diferenciada, sendo que, na modalidade Central, existe a figura de um único gestor para UD's distintas. Já a modalidade Virtual conta com um gestor do GeoInfo por unidade. Na modalidade Local, há maior facilidade para o acoplamento de *workflows* científicos de análise e de processamento de grandes volumes de dados para a geração de novos conhecimentos a partir dos dados armazenados no repositório. Nesse caso, a responsabilidade pela manutenção da infraestrutura e do software utilizado é da Unidade detentora dos dados (Drucker *et al.*, 2017).

Diferentes customizações foram realizadas nos dois *softwares*, para atender às características e às necessidades da Embrapa. Dentre elas, destaca-se a criação de uma solução para suportar o atributo "Palavras-Chave Embrapa", o qual utiliza uma base de linguagem artificial, especializada no vocabulário de geociências e de geotecnologias, construída para normalizar os dados, possibilitar a indexação dos recursos geoespaciais e otimizar a recuperação dos dados e das informações no repositório. É importante destacar que as customizações implementadas exigem maior esforço e cuidados na infraestrutura, especialmente quando há necessidade de realizar atualizações de versão do GeoInfo.

2.3 Processo de Gestão de Dados de Pesquisa Espacial (GDPE)

Com a consolidação desta arquitetura, buscou-se estruturar o processo de GDPE, o qual foi fortemente apoiado na literatura que trata da gestão do ciclo de vida de dados científicos. Em especial, baseou-se no trabalho da *United States Geological Survey (USGS)*⁵⁶ (Fauden *et al.*, 2013) e no ciclo de vida dos dados proposto pela rede *Data Observation Network for Earth (DataONE)* (Strausser *et al.*, 2013), que 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 pesquisa. Todas as etapas do ciclo de vida dos dados são relevantes no contexto de organizar, de preservar e de disponibilizar dados geoespaciais no âmbito da Embrapa.

O processo de gestão de dados espaciais da Embrapa (Figura 2) é composto por três etapas principais e envolve atores e perfis distintos. Para sua condução, optou-se pela definição de um gestor do GeoInfo⁵⁷ em cada unidade da Embrapa que implementa e executa as ações planejadas, pelo CG-GeoInfo, relativas à gestão de dados espaciais. Esse agente está encarregado de acompanhar todas as etapas do processo e de atuar, especialmente, nas fases de curadoria e de disponibilização do dado espacial, de modo a garantir a completude e a integridade do processo. Trata-se de uma figura estratégica no contexto da arquitetura e das etapas necessárias para o compartilhamento dos dados, que mantém constante relação com o CG-GeoInfo.

56 Agência do Departamento do Interior dos Estados Unidos que coordena o Programa Geoespacial daquele país e um dos pilares da National Spatial Data Infrastructure (NSDI – Infraestrutura Nacional de Dados Espaciais dos Estados Unidos).

57 Também chamado de administrador local da IDE, ou ponto focal.

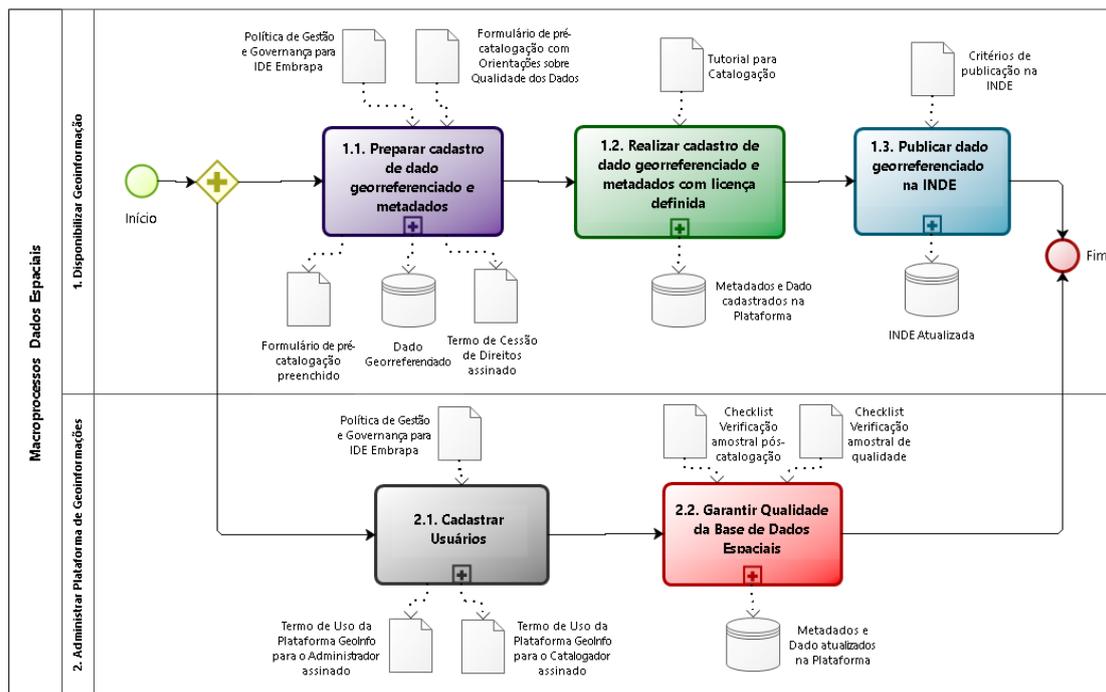


Figura 2. Processo de gestão de dados de pesquisa espacial da Embrapa.
Fonte: Drucker *et al.* 2017.

A primeira etapa inicia-se com a preparação dos dados espaciais pelo autor do dado; nela, é necessário seguir as orientações para nomear o arquivo de acordo com a estrutura “ÁreaGeográfica_Tema_Tempo”, organizar as quatro extensões que compõem o dado espacial vetorial no Geoinfo (.shp, .dbf., .shx e .prj); ajustar o dado para um sistema de referência cartográfica padrão (EPSG 4326); preparar o formulário de pré-catálogo⁵⁸ e indicar questões de propriedade intelectual, de licenciamento⁵⁹ e/ou de direito autoral, caso existam. Em seguida, passa-se para a etapa de inserção e de documentação do dado espacial, descrevendo-se os metadados a partir das informações existentes no formulário de pré-catálogo e as diretrizes para representação do dado espacial no Geoinfo. Essa etapa deve envolver o profissional de informação. Na terceira e última etapa, conduzida pelo gestor do Geoinfo, buscam-se analisar as condições possíveis para a disponibilização ou não do dado espacial para a INDE, tendo em vista as informações inseridas para a documentação do dado no perfil de metadados, em especial quanto à confidencialidade (Drucker *et al.*, 2017).

Diferentes documentos foram produzidos consoante ao processo de GDPE, como instrumentos orientadores para cada uma das etapas e atores previstos nas ações de inserção, de tratamento e de disponibilização dos dados. Esses documentos foram organizados e disponibilizados de forma a atingir todas as UD, por meio da intranet corporativa da Embrapa. O canal possibilita acesso a esses documentos e a outras informações, tais como a citação de dados espaciais e de tutoriais para geração de produtos derivados dos dados armazenados/disponibilizados no Geoinfo.

Para tornar esse processo uma ação rotineira às UD geradoras de dados espaciais, consolidou-se o Geoinfo na estrutura programática da PD&I da Embrapa, reconhecendo-o como

58 Instrumento para garantir a documentação do dado antes de ser inserido no Geoinfo, pelo autor do dado.

59 Para os dados espaciais inseridos no Geoinfo adota-se, como padrão não condicionante e obrigatório, a licença Creative Commons NonComercial (CC BY NC).

ambiente para comprovação dos resultados de projetos e de ações de pesquisa, quando esses se referem a bases de dados espaciais. Essa inserção ratifica o GeoInfo como ambiente estratégico para a empresa posicionar-se frente ao Decreto Lei 6.666/08 e aos movimentos Acesso e Ciência Aberta, além de possibilitar o gerenciamento seguro dos dados espaciais (Drucker *et al.*, 2017; Pinto *et al.*, 2017).

2.4 Perfil de Metadados

Em um repositório institucional, é imprescindível a definição e a estruturação do perfil de metadados que deverá abranger toda a coleção a ser disponibilizada, de modo a assegurar que todos os metadados necessários para representação de um recurso sejam utilizados e estejam devidamente preenchidos (Pavão *et al.*, 2015). Para a representação e a interoperabilidade dos dados espaciais, utilizou-se o padrão ISO 19115:2003 (International Organization for Standardization, 2003), o qual propõe um núcleo comum de oito atributos para a descrição dos produtos geoespaciais. Essa norma apresenta, ainda, outros 354 atributos que podem ser utilizados e aplicados a um perfil de metadados, de acordo com as necessidades institucionais. Além dela, foram adotadas as orientações da INDE, descritas no Perfil de Metadados Espaciais do Brasil (Perfil MGB; Comissão Nacional de Cartografia, 2009).

Buscando a consolidação de um perfil de metadados para os dados espaciais da Embrapa, realizou-se a análise da norma ISO e do Perfil MGB. Essa atividade foi realizada por meio de um grupo de trabalho denominado Grupo de Trabalho de Referência da Geoinformação (GT-Geo-Refer), criado durante a vigência do projeto IDE-Embrapa e considerou, ainda, a apresentação dos dados quando de sua exportação para outros ambientes. Para isso, respeitaram-se os padrões de descrição para a interoperabilidade semântica entre as bases de dados e as diretrizes estabelecidas pelos princípios FAIR (*Findability, Accessibility, Interoperability and Reusability*) relativas à gestão de dados científicos (Wilkinson *et al.*, 2016), o que precisa ser intensificado em oportunidades futuras de revisão da arquitetura, do processo e do perfil de metadados do GeoInfo. Ao final, conforme descrito em Pinto *et al.* (2017), foram identificadas a necessidade de criação de novos atributos e de alteração na nomenclatura de alguns itens. Atualmente, o perfil de metadados do GeoInfo contém 33 elementos para descrição dos recursos geoespaciais temáticos, dos quais 17 são campos obrigatórios (Figura 3).

Metadados : Uso e Ocupação do Vale do Paraíba do Sul (SP) no ano de 1985.

[Visitar Camada](#)

Identificação

| | |
|-----------------------|--|
| Título | Uso e Ocupação do Vale do Paraíba do Sul (SP) no ano de 1985. |
| Resumo | Representação temática das categorias de uso e ocupação do Vale do Paraíba do Sul, na perspectiva da cobertura vegetal e atividade mineradora. A classificação do tipo não supervisionada foi praticada pelo uso de imagem Landsat 5, composição colorida da banda visível (R,G, B), com auxílio do software ArcGis 10.3 (ESRI). Distribuição espacial da cobertura vegetal, no vale do paraíba do sul, para o ano de 1985. A região compreende 34 municípios. |
| Licença | Public Domain (PD) |
| Creation Data | Mai 26, 2017, 9:21 a.m. |
| Tipo | Vector Data |
| Palavras-chave | eucalipto mapa de uso e cobertura sensoriamento remoto |
| Ano de criação | Jan. 1, 2015, meia-noite |
| Categoria | Geoscientific Information |
| Regiões | São Paulo |
| Publicado | Sim |
| Caracterizado | Não |

Proprietário

| | |
|--------------------|--|
| Nome | Embrapa Monitoramento por Satélite (embrapa) |
| email | embrapa@embrapa.br |
| Cargo | Nenhum |
| Empresa | Embrapa Monitoramento por Satélite |
| Localização | Av. Soldado Passarinho, nº 303, Fazenda Jardim Chapadão CEP: 13070-115 Campinas - SP São Paulo BRA |
| Voz | (19) 3211-6200 |
| Fax | (19) 3211-6222 |

Informações da Camada

Imagem de Identificação



| | |
|----------------------------|----------------|
| Resolução Espacial | --- |
| Sistema de Projeção | EPSG:4326 |
| Extensão x0 | -46.3654810857 |
| Extensão x1 | -44.1588781007 |
| Extensão y0 | -23.6342933503 |
| Extensão y1 | -22.4003490513 |

Características

| | |
|-----------------------------------|--|
| Frequência de manutenção | Dados São Atualizados Quando Necessário |
| Finalidade | Projeto "Análise da distribuição geoespacial e de aspectos ambientais da eucaliptocultura na bacia do Rio Paraíba do Sul - GEOVALE. Código SEG: 02.12.10.002.00.00. |
| Idioma | Português |
| Extensão temporal | Jan. 1, 1985, meia-noite - Jan. 1, 1985, meia-noite |
| Qualidade dos dados | O trabalho foi realizado a partir da análise das Imagens Landsat 5, sensores TM, 1985. Mais informações sobre a metodologia utilizada podem ser obtidas na publicação: RONQUIM, C. C.; SILVA, R. F. B.; FIGUEIREDO, E. B. de; BORDONA, R. O. Carbon sequestration associated to the land-use and land-cover changes in the forestry sector in Southern Brazil. Proceedings of SPIE - International Society for Optical Engineering, v. 9998, p. 99981T-1- 99981T-14, 2016. Disponível em: < https://ainfo.cnptia.embrapa.br/digital/bitstream/item/149463/1/Paper-Carbono-Ronquim-SPIE-16.pdf >. Acesso em: 27 jun. 2017. |
| Informações complementares | No information provided |

Pontos de Contato

| | |
|--------------------|---|
| Nome | Área de Transferência de Tecnologias (transferencia.tecnologias) |
| email | cnpm.geoinfo@embrapa.br |
| Cargo | Nenhum |
| Empresa | Embrapa Monitoramento por Satélite |
| Localização | Av. Soldado Passarinho, nº 303, Fazenda Jardim Chapadão 13070-115 Campinas - SP BRA |

Figura 3. Exemplo de dado espacial catalogado a partir das orientações definidas pelo GeoRefer.
Fonte: GeoInfo. Endereço para visualização do dado espacial no GeoInfo: < http://geoinfo.cnpm.embrapa.br/layers/geonode%3Auso_vale_1985 >.

Destaca-se que a discriminação dos autores/produtores de dados não estava prevista como atributo no Perfil MGB e no núcleo da ISO 19115:2003, tendo sido implementada, no GeoInfo, por meio da seção CI_RoleCode, da ISO 19115:2003. Essa seção possui 11 possíveis perfis associados, dentre os quais, a autoria. Esse atributo reconhece a produção do dado, pelo autor, garantindo-lhe o

direito autoral⁶⁰ sobre o dado. Num ambiente de pesquisa, a descrição individualizada dos responsáveis pela geração do dado espacial é fundamental para o reconhecimento dos pesquisadores, haja vista as novas tendências de comunicação científica baseadas no dado de pesquisa.

É importante ressaltar o entendimento de que processo de gestão dos dados espaciais da Embrapa, em especial no que tange à descrição e à representação, necessita ser periodicamente revisitado, buscando analisar a necessidade de ajustes para que se possa aproveitar ao máximo a organização desses insumos, que são resultados das pesquisas realizadas pela Embrapa, de modo a orientar a empresa sobre sua gestão de PD&I.

3 DESAFIOS

A implementação e o desenvolvimento de um repositório de dados foi um dos desafios vencidos pela Embrapa. Para que o repositório seja efetivo e cumpra a sua função, é preciso vencer outras questões, que estão relacionados com a cultura do compartilhamento e do reuso dos dados, com a sustentabilidade do repositório, com a privacidade dos dados, com a identificação de atores no processo de gestão de dados espaciais, com a propriedade intelectual e com a comunicação científica.

A Embrapa, que é uma empresa pública, possui uma função social, portanto, a discussão sobre o compartilhamento de dados deve se dar apenas quando questões éticas e legais estejam envolvidas, ou seja, aqueles dados que não ferem a privacidade, a soberania nacional, o interesse comercial, e que tiverem sido gerados a partir de pesquisas financiadas com recursos públicos, são bens públicos e devem ser compartilhados. Cabe à Embrapa acompanhar as discussões e os esforços relacionados ao paradigma da *e-Science* e aos movimentos “Governo Aberto” e “Ciência Aberta”, como forma de democratizar o acesso aos dados, às informações e ao conhecimento gerados pela Empresa. Além disso, é preciso que os pesquisadores da Embrapa também assumam o papel de reutilizadores de dados, não só daqueles gerados em suas pesquisas, bem como dados externos. Para isso, é necessário que eles sejam capacitados para desenvolver habilidades em pesquisa intensiva em dados.

A Lei nº 13.709, que dispõe sobre o tratamento de dados pessoais (Brasil, 2018) exige a anonimização dos dados, para seu uso e sua disponibilização. Mesmo aplicando esse cuidado aos dados espaciais disponibilizados no GeolInfo, é necessário buscar assegurar, continuamente, o equilíbrio entre o compartilhamento e a privacidade dos dados, tendo em vista as possibilidades de obter novos resultados a partir do cruzamento de diferentes bases de dados. Por isso, tendo em vista o contexto das instituições de pesquisa pública brasileira, devem-se identificar e construir estratégias para viabilizar a manutenção, a evolução e a sustentabilidade do repositório GeolInfo, o que é um desafio, especialmente num cenário de contingenciamento de recursos. Investimentos financeiros são importantes, principalmente para suprir a necessidade de manutenção de *hardware* e de *software*, que demandam constante substituição e atualização. As mudanças tecnológicas acontecem rapidamente, exigindo acompanhar a evolução das ferramentas e adotar tecnologias mais eficazes para garantir a relevância e o aprimoramento contínuo do repositório.

Considerando a busca por financiamento em editais públicos, ressalta-se um outro desafio, no qual propostas para a manutenção e para a atualização de um repositório já existente dificilmente são consideradas competitivas. Dessa forma, torna-se oportuno identificar soluções

60 No processo de gestão de dados espaciais por meio do GeolInfo, é proposto um termo de cessão de direitos autorais para o dado espacial.

criativas que possam, também, configurar parcerias com a iniciativa privada, reservando-se os aspectos inerentes à característica democrática promovida por meio do GeolInfo. Os recursos humanos também são importantes para a continuidade do repositório, sendo necessária uma equipe dedicada e que tenha acesso frequente a cursos e a treinamentos para o desenvolvimento de habilidades profissionais. Sendo assim, a sustentabilidade do repositório requer o comprometimento institucional.

A identificação de atores considerados essenciais ao processo de gestão de dados espaciais também é algo desafiador e requer uma tramitação articulada, pautada na comunicação direta. O processo é dependente de ações relacionadas à cultura organizacional e, por isso, requer estratégias de comunicação interna eficientes e campanhas para fornecer o máximo de informação o quanto possível.

Uma tendência em comunicação científica que pode servir como incentivo aos pesquisadores para compartilhar os dados é a atribuição de identificadores persistentes aos conjuntos de dados disponibilizados, os quais facilitam a citação e a extração de métricas de acesso aos dados, que, atualmente, passam a ser citáveis e passíveis de atribuição de crédito. No caso do repositório GeolInfo, passar a atribuir identificadores persistentes, como, por exemplo, a adoção do *Digital Object Identifier* (DOI) para bases de dados, é uma das metas para o futuro próximo.

4 CONSIDERAÇÕES FINAIS

A componente espacial é essencial para a integração de dados e para a compreensão da agropecuária nacional. Na Embrapa, diversos conjuntos de dados geoespaciais são gerados e analisados no contexto de suas pesquisas, os quais agora podem ser compartilhados e incorporados a outros estudos da própria empresa ou de outras instituições que possuem desafios semelhantes ou complementares.

Por meio do repositório GeolInfo, dados gerados tanto por observações ou por levantamentos *in loco*, quanto por processamento de dados de sensoriamento remoto ou por sensores de campo, podem ser acessados, compreendidos, integrados e reutilizados, viabilizando novas análises e a geração de novo conhecimento para a agropecuária e para a sociedade brasileira. A estratégia adotada para viabilizar a implantação do repositório GeolInfo foi bem sucedida, considerando-a como 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 na difusão de conhecimento e de inovação. Discussões acerca do dado de pesquisa como produção científica, passível de citação, a exemplo do projeto *Making Data Count*⁶¹ (MAKE..., 2020), proporcionam à Embrapa a oportunidade de analisar a possibilidade de inclusão de um novo formato de publicação para seus dados, a fim de valorizá-los no contexto de produção, estimulando seus pesquisadores a compartilharem seus dados brutos de forma estruturada no repositório GeolInfo. Uma vez organizadas e documentadas, as bases de dados também tornam-se aptas a serem compartilhadas de maneira adequada para a sociedade.

O GeolInfo 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 inserção do processo na programação de PD&I da Embrapa assegura a perpetuação da iniciativa ao longo do tempo, com novos dados sendo gerados, descritos e publicados. Por in-

61 Iniciativa que busca coletar e padronizar métricas sobre a utilização dos dados de pesquisa, por meio de visualizações, downloads e citações.

termédio do repositório Geoinfo, o retrabalho e as redundâncias na produção de dados espaciais são evitados, possibilitando sua integração a outras informações disponíveis em nosso país por meio de uma infraestrutura única.

Nesse aspecto, a Embrapa posiciona-se como importante apoiadora das ações de Acesso Aberto e Ciência Aberta, tornando transparente seus produtos de pesquisa, expressos pelos dados espaciais, o que traz à Empresa a oportunidade de inserir, em seu contexto de gestão da PD&I, a análise sobre os resultados de pesquisa ou, propriamente, os dados como resultantes de seus investimentos, contextualizando-os ao consumo pela sociedade. Assim, faz-se necessário investir em mecanismos para monitorar o acesso e o impacto do uso dos dados espaciais disponibilizados.

5 REFERÊNCIAS

BRASIL. Decreto n. 6.666, de 27 de novembro de 2008. Institui, no âmbito do Poder Executivo federal, a Infraestrutura Nacional de Dados Espaciais - INDE, e dá outras providências. **Diário Oficial [da] República Federativa do Brasil**, Poder Executivo, Brasília, DF, 28 nov. 2008. p. 57.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado. **Diário Oficial [da] República Federativa do Brasil**, Poder Executivo, Brasília, DF, 15 ago. 2018.

CAFÉ, L. M. A.; FACHIN, G. R. B. Provedores de dados, provedores de serviços e periódicos em Ciência da Informação, Biblioteconomia e áreas afins. **Encontros Bibli** (Online), v. 1. sem, p. 59-76, 2007. Disponível em: <https://doi.org/10.5007/1518-2924.2007v12nesp1p59>. Acesso em: 10 nov. 2022.

COMISSÃO NACIONAL DE CARTOGRAFIA (Brasil). Comitê de Estruturação de Metadados Geoespaciais. Comissão Nacional de Cartografia. **Perfil de metadados geoespaciais do Brasil: perfil MGB**. Rio de Janeiro: CEMG-Concar, 2009. 194 p.

CUSTODIO, D. de O.; PINTO, D. M.; VICTORIA, D. de C.; DRUCKER, D. P.; FIDALGO, E. C. C.; DOMPIERI, M. H. G.; SIMÕES, M. O uso de ferramentas open source para compor uma IDE (Infraestrutura de Dados Espaciais) na Embrapa. In: SIMPÓSIO BRASILEIRO DE INFRAESTRUTURA DE DADOS ESPACIAIS, 1., 2018, Rio de Janeiro. **Anais...** Rio de Janeiro: IBGE, 2018. p. 47-48. Disponível em: <https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1100849/1/4988.pdf>. Acesso em: 29 abr. 2020.

DRUCKER, D. P.; PINTO, D. M.; FIDALGO, E. C. C.; CUSTÓDIO, D. de O.; VICTORIA, D. de C.; ALMEIDA, B. T. de; SIMÕES, M.; MACHADO, C. R. de L.; SANTOS, V. V. dos; DART, R. de O.; AGLIO, M. L. D.; RASCHE, F.; LAFORET, M. R. C.; BETTIOL, G. M.; PEREIRA, S. E. M.; BRANDÃO, Z. N.; GARRASTAZU, M. C.; FILIPPINI ALBA, J. M.; COSTA, F. A. da; TORRES, R. C.; DOMPIERI, M. H. G.; SAMPAIO, S. M. N.; MARTINS, E. C.; PIEROZZI JUNIOR, I.; SILVA, G. B. S. da; GALINARI, G.; FIORINI, F. B.; TAKEMURA, C. M.; CRUZ, S. A. B. da; HOLLER, W. A.; OLIVEIRA, L. H. M. de. GeoInfo - infraestrutura de dados espaciais abertos para a pesquisa agropecuária. **Reciis: Revista Eletrônica de Comunicação, Informação & Inovação em Saúde**, Rio de Janeiro, v. 11, p. 1-17, 2017. DOI: <https://doi.org/10.29397/reciis.v11i0.1356>. Disponível em: <https://www.recis.icict.fiocruz.br/index.php/reciis/article/view/1356>. Acesso em: 10 nov. 2022

FAUNDEEN, J.L., BURLEY, T.E.; CARLINO, J.A.; GOVONI, D.L.; HENKEL, H.S.; HOLL, S.L.; HUTCHISON, V.B.; MARTÍN, E.; MONTGOMERY, E.T.; LADINO, C.C.; TESSLER, S.; ZOLLY, L.S. **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>. Acesso em: 10 nov. 2022.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). **Geographic Information - Metadata**: ISO 19115:2003. London, England; 2003. 152 p.

MAKE DATA COUNT. 2020. Disponível em: <https://makedatacount.org>. Acesso em: 10 nov. 2022.

PAVÃO, C. G.; COSTA, J. S. B.; FERREIRA, M. K.; HOROWITZ, Z. Metadados e repositórios institucionais: uma relação indissociável para a qualidade da recuperação e visibilidade da informação. In: CONFERÊNCIA LUSO-BRASILEIRA SOBRE ACESSO ABERTO, 6., 2015. **Anais...** Salvador: UFBA; 2015.

PINTO, D. M.; BRANDÃO, S. V. V. dos; DRUCKER, D. P.; DOMPIERI, M. H. G.; RASCHE, F.; HOLLER, W. A.; BETTIOL, G. M.; CUSTODIO, D. de O.; VICTORIA, D. de C.; GONÇALVES, L. de M. P. B. Descrição dos dados da pesquisa geoespacial: a experiência da Embrapa. **Reciis: Revista Eletrônica de Comunicação, Informação & Inovação em Saúde**, Rio de Janeiro, v. 11, p. 1-16, 2017. DOI: [10.29397/reciis.v11i0.1418](https://doi.org/10.29397/reciis.v11i0.1418).

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. [**Proceedings...**]. California: University of California, 2011. p. 126–131.

WILKINSON, M.D.; DUMONTIER, M.; AALBERSBERG, I.J.; APPLETON, G.; AXTON, M.; BAAK, A.; BLOMBERG, N.; BOITEN, J.W.; SANTOS, L.B.S.; BOURNE, P.E.; BOUWMAN, J.; BROOKES, A.J.; CLARK, T.; CROSAS, M.; DILLO, I.; DUMON, O.; EDMUNDS, S.; EVELO, C.T.; FINKERS, R.; GONZALEZ-BELTRAN, A.; GRAY, A.J.G.; GROTH, P.; GOBLE, C.; GRETHE, J.S.; HERINGA, J.; HOEN, P.A.C.; HOOFT, R.; KUHN, T.; KOK, R.; KOK, J.; LUSHER, S.J.; MARTONE, M.E.; MONS, A.; PACKER, A.L.; PERSSON, B.; ROCCA-SERRA, P.; ROOS, M.; SCHAIK, R.; SANSONE, S.A.; SCHULTES, E.; SENGSTAG, T.; SLATER, T.; STRAWN, G.; SWERTZ, M.A.; 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]. **Sci. Data**, v. 3, p.160018. DOI: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18), 2016.