

Pecha kucha

Plano de Gestão de Dados Geoespaciais da Embrapa: Modelo Adaptado do FioDMP da Fiocruz

*Embrapa's Geospatial Data Management Plan: An Adapted
Model from Fiocruz's FioDMP*

*Plan de Gestión de Datos Geoespaciales de Embrapa:
Modelo Adaptado del FioDMP de Fiocruz*

Viviane Santos de Oliveira Veiga

Doutorado em Ciências/Informação e Comunicação em Saúde

Fundação Oswaldo Cruz

ORCID: <https://orcid.org/0000-0001-8318-7912>

Lattes: <http://lattes.cnpq.br/4983074089687751>

E-mail: viviane.veiga@icict.fiocruz.br

Patricia Corrêa Henning

Doutorado em Ciências/Informação e Comunicação em Saúde

Universidade Federal do Estado do Rio de Janeiro

ORCID: <https://orcid.org/0000-0003-0739-6442>

Lattes: <http://lattes.cnpq.br/0970010723997242>

E-mail: henningpatricia@gmail.com

Márcia Izabel Fugisawa Souza

Doutora em Educação

Embrapa Agricultura Digital

ORCID: <https://orcid.org/0000-0002-6194-9354>

Lattes: <http://lattes.cnpq.br/5202566311237784>

E-mail: marcia.fugisawa@embrapa.br

Tércia Zavaglia Torres

Doutora em Educação

Embrapa Agricultura Digital

ORCID: <https://orcid.org/0000-0001-7091-4201>

Lattes: <http://lattes.cnpq.br/0165420464019947>

E-mail: tercia.torres@embrapa.br

Ivo Pierozzi Junior

Doutor em Ecologia

Embrapa Agricultura Digital

ORCID: <https://orcid.org/0000-0002-4979-1013>

Lattes: <http://lattes.cnpq.br/1666276452783441>

E-mail: ivo.pierozzi@embrapa.br

João dos Santos Vila da Silva

Doutor em Engenharia Agrícola

Embrapa Agricultura Digital

ORCID: <https://orcid.org/0000-0003-3973-9745>

Lattes: <http://lattes.cnpq.br/9815197324138278>

E-mail: joao.vila@embrapa.br

Resumo

Os Planos de Gestão de Dados de Pesquisa (PGDs) adotam frequentemente modelos genéricos que não contemplam plenamente as especificidades dos dados de diferentes áreas ou projetos, o que limita sua efetividade prática. Para enfrentar esse desafio, desenvolveu-se o PGD-GEO, resultado de uma parceria entre a Fiocruz e a Embrapa, com o objetivo de padronizar e aprimorar os processos relacionados à documentação, armazenamento, requisitos legais, preservação e compartilhamento de dados geoespaciais. O modelo, baseado na ferramenta FioDMP da Fiocruz, foi estruturado para ser flexível e adaptável, integrando padrões de metadados específicos da área, formatos amplamente utilizados em geotecnologias, diretrizes para armazenamento, licenciamento e requisitos legais, bem como políticas de compartilhamento. Além disso, ele foi projetado para avaliar a aderência aos princípios FAIR e para integrar-se a outras ferramentas, favorecendo sua incorporação nos fluxos de trabalho dos projetos científicos e tecnológicos da instituição. Com essa iniciativa, busca-se fortalecer a governança dos dados geoespaciais, promover maior eficiência na gestão da informação, e ampliar o impacto científico e tecnológico dos dados produzidos.

Palavras-chave: Plano de gestão de dados; Princípios FAIR; Template de PGD; Dados geoespaciais.

Abstract

Research Data Management Plans (DMPs) often adopt generic template models that do not fully address the specificities of data from different domains or projects, which limit their practical effectiveness. To address this challenge, the PGD-GEO was developed as the result of a partnership between Fiocruz and Embrapa, with the aim of standardizing and improving processes related to the documentation, storage, legal requirements, preservation, and sharing of geospatial data. Based on Fiocruz's FioDMP tool, the

model was designed to be flexible and adaptable, integrating domain-specific metadata standards, widely used geotechnology data formats, guidelines for storage, licensing, and legal requirements, as well as data-sharing policies. In addition, it was conceived to assess adherence to the FAIR principles and to integrate with other tools, thus facilitating its incorporation into the workflows of scientific and technological projects within the institution. This initiative seeks to strengthen the governance of geospatial data, promote greater efficiency in information management and enhance the scientific and technological impact of the data produced.

Keywords: Data management plan; FAIR principles; DMP template; Geospatial data

Resumen

Los Planes de Gestión de Datos de Investigación (PGD) suelen adoptar modelos genéricos que no contemplan plenamente las especificidades de los datos de diferentes áreas o proyectos, lo que limita su efectividad práctica. Para enfrentar ese desafío, se desarrolló el PGD-GEO, resultado de una asociación entre Fiocruz y Embrapa, con el objetivo de estandarizar y mejorar los procesos relacionados con la documentación, el almacenamiento, los requisitos legales, la preservación y el intercambio de datos geoespaciales. El modelo, basado en la herramienta FioDMP de Fiocruz, fue diseñado para ser flexible y adaptable, integrando estándares de metadatos específicos del área; formatos de datos ampliamente utilizados en geotecnologías, directrices para el almacenamiento, licenciamiento y cumplimiento legal, así como políticas de intercambio. Además, fue concebido para evaluar la adhesión a los principios FAIR y para integrarse con otras herramientas, facilitando su incorporación en los flujos de trabajo de los proyectos científicos y tecnológicos de la institución. Con esa iniciativa, se busca fortalecer la gobernanza de los datos geoespaciales, promover una mayor eficiencia en la gestión de la información, y ampliar el impacto científico y tecnológico de los datos producidos.

Palabras clave: Plan de Gestión de Datos; Principios FAIR; Plantilla de PGD; Datos Geoespaciales

Introdução

A Gestão de Dados de Pesquisa (GDP) é um componente essencial da prática científica contemporânea, apoiando a organização, preservação, acessibilidade e reutilização dos dados ao longo de todo o ciclo de vida da pesquisa. Planos de Gestão de Dados (PGDs) vêm sendo crescentemente exigidos por agências de fomento, periódicos científicos e instituições de pesquisa como instrumentos essenciais da ciência aberta (Torres et al., 2022; Souza et al., 2022).

Apesar dessa evolução, muitos PGDs ainda se apoiam em modelos genéricos que não consideram as especificidades de diferentes domínios científicos. Isso resulta em dificuldades de aplicação e limitações na documentação e na reutilização dos dados.

No caso dos dados geoespaciais, a gestão eficiente apresenta desafios adicionais relacionados ao volume, à diversidade de fontes, à necessidade de precisão cartográfica, à interoperabilidade e à atualização contínua. Imagens de satélite, mapas temáticos, dados de sensoriamento remoto e registros obtidos por sistemas de posicionamento global são exemplos de dados geoespaciais que requerem metadados detalhados, adoção de padrões internacionais (ex. ISO 19115) e protocolos de interoperabilidade definidos pelo Open

Geospatial Consortium (OGC), a fim de assegurar qualidade, integridade e reutilização.

Dante desse cenário, a Embrapa estabeleceu parceria com a Fiocruz para desenvolver um modelo específico de PGD voltado ao domínio geoespacial (PGD-GEO). A colaboração aproveitou a experiência da Fiocruz com a ferramenta FioDMP (Veiga et al., 2019), adaptando-a às demandas da pesquisa agropecuária e ambiental. O PGD-GEO foi concebido para alinhar-se aos princípios FAIR, fortalecer a governança de dados e integrar-se nos fluxos institucionais de pesquisa, contribuindo para práticas mais consistentes de *data stewardship* (curadoria responsável de dados) em ciência aberta.

Metodologia

O desenvolvimento do PGD-GEO seguiu uma abordagem qualitativa, de caráter exploratório e aplicado, baseada em revisão bibliográfica, análise documental e colaboração interinstitucional. O processo foi estruturado em cinco etapas principais:

1. Análise do FioDMP – exame das funcionalidades e estrutura do template genérico da Fiocruz, identificando elementos adaptáveis ao domínio geoespacial;
2. Revisão de diretrizes – levantamento de normativas nacionais e internacionais em gestão de dados geoespaciais, incluindo recomendações da RDA, DataONE, USGS e normas ISO/OGC;
3. Adaptação ao contexto agropecuário – ajustes metodológicos e conceituais para contemplar os requisitos específicos da Embrapa em termos de dados territoriais e ambientais;
4. Avaliação piloto com especialistas – coleta de feedback de pesquisadores e gestores de dados, visando verificar a aderência às práticas institucionais e aos princípios FAIR; e
5. Testes práticos – aplicação piloto do template em projetos selecionados, avaliando usabilidade, completude e potencial de integração em fluxos de trabalho.

Essa abordagem buscou garantir que o PGD-GEO fosse, ao mesmo tempo, tecnicamente robusto, aderente às práticas institucionais e aplicável em cenários reais de pesquisa.

Estrutura do Template PGD-GEO

O template do PGD-GEO é organizado em sete seções que contemplam os principais aspectos da gestão de dados geoespaciais. Cada seção apresenta objetivos claros e elementos obrigatórios, alinhados a padrões nacionais e internacionais (como ISO 19115 para metadados e LGPD para dados pessoais).

Seção A - Informação administrativa

Reúne informações essenciais para a identificação do projeto e do Plano de Gestão de Dados Geoespaciais (PGD-GEO Embrapa), tais como: dados do responsável pela elaboração do PGD, informações do pesquisador principal, título e resumo do projeto, data de início, idioma, agência financiadora e unidades envolvidas. Também contempla a utilização de vocabulários controlados e termos livres, de forma a padronizar a descrição e assegurar a rastreabilidade administrativa e institucional.

Seção B - Descrição dos dados coletados e reutilizados

Detalha os dados utilizados no projeto, considerando tipo, natureza, formato, volume e origem. Inclui ainda: metodologias de coleta ou produção, softwares empregados, estratégias de tratamento e repositórios de origem. Para dados reutilizados, descreve-se o processo de obtenção, restrições de uso, histórico e proveniência, bem como a possibilidade de integração com dados gerados pelo próprio projeto.

Seção C - Documentação e qualidade dos dados

Assegura que os dados sejam devidamente descritos e organizados para uso presente e futuro. Engloba a definição de padrões de metadados (ex.: ISO 19115), elaboração de arquivos README, versionamento de dados, práticas de organização de diretórios e nomeação de arquivos, além de prever medidas de validação, controle de qualidade e requisitos de interoperabilidade entre sistemas.

Seção D - Armazenamento e backup

Define estratégias para garantir a preservação, segurança e recuperação dos dados. Recomenda-se o armazenamento em ao menos dois locais distintos, com definição da frequência de cópias de segurança. Inclui a escolha de plataformas digitais institucionais, mecanismos de controle de acesso e políticas de segurança da informação, em conformidade com normas regulatórias e institucionais.

Seção E - Requisitos legais e éticos

Assegura que o projeto esteja em conformidade com a legislação vigente e com normas éticas aplicáveis. Nesse sentido, a seção contempla a aprovação em comitês de ética (quando pertinente), cadastramento na Plataforma Brasil, atendimento à Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709/2018), diretrizes de confidencialidade, tratamento de dados sensíveis, períodos de embargo por direitos autorais e categorias de sigilo aplicáveis.

Seção F - Compartilhamento de dados

Orienta práticas de transparência, acesso e reuso. Esta seção define previsão temporal para disponibilização dos dados, protocolos e formatos de compartilhamento, licenças de uso (como Creative Commons), políticas de embargo, mecanismos on-line de acesso aberto e registro de identificadores persistentes (a exemplo do DOI), com o objetivo de assegurar a reproduzibilidade, rastreabilidade e conformidade ética e legal.

Seção G - Responsabilidades e recursos financeiros para a Gestão de Dados de Pesquisa

Estabelece a alocação de recursos humanos, financeiros e de infraestrutura necessários à gestão dos dados. Define papéis e responsabilidades dentro da equipe, custos associados à manutenção de boas práticas de gestão, e mecanismos para garantir que os dados sejam FAIR (Findable, Accessible, Interoperable, Reusable).

Considerações Finais

A adaptação da ferramenta FioDMP ao contexto da Embrapa resultou no PGD-GEO, um modelo estruturado que fortalece a governança dos dados por meio da documentação, preservação e compartilhamento, em alinhamento com a ciência aberta e os princípios FAIR.

Entre suas contribuições, destacam-se a padronização de práticas, o incentivo à interoperabilidade e o apoio à tomada de decisão baseada em evidências. Contudo, permanecem desafios como a necessidade de customizações contínuas, e a ausência, até o momento, de métricas consolidadas de avaliação da efetividade do modelo.

Os próximos passos incluem:

- Ampliar a aplicação do PGD-GEO em diferentes projetos da Embrapa;
- Desenvolver indicadores de avaliação de completude, aderência ao FAIR e usabilidade;
- Promover capacitações institucionais para disseminação do modelo; e
- Explorar a escalabilidade da proposta para outros domínios além dos dados geoespaciais.

Assim, a iniciativa de desenvolver o PGD-GEO representa um passo importante para consolidar práticas institucionais de gestão de dados na Embrapa, promovendo maior padronização, transparência e alinhamento com os princípios FAIR. Esses avanços contribuem para fortalecer a capacidade da instituição em lidar com grandes volumes de dados geoespaciais, o que é estratégico para a inovação científica e sustentabilidade de pesquisas agropecuárias baseadas em evidências. Nesse processo, a Fiocruz amplia o alcance do FioDMP para além da área da saúde, demonstrando sua aplicabilidade

em outros domínios e consolidando parcerias interinstitucionais em outros campos científicos.

Conflito de Interesses

Os autores declaram não haver conflitos de interesses.

CRediT – Contribuições dos Autores

Viviane Santos de Oliveira Veiga, Patricia Corrêa Henning, Márcia Izabel Fugisawa Souza e Tércia Zavaglia Torres| Concretualização, Escrita – redação original, Supervisão, Recolha de dados, Ivo Pierozzi Junior e João dos Santos Vila da Silva| Escrita – revisão e edição, Supervisão,

Referências

Souza, M. I. F., Visoli, M. C., Torres, T. Z., Falcão, P. R. K., Nhani, A., Jr, Giachetto, P. F., Silva, F. R., Cintra, L. C., Osawa, C. C., Silva, A. R., & Barbosa, L. A. F. (2022). Planos de gestão de dados acionáveis por máquina alinhados aos princípios FAIR para o Laboratório Multiusuário de Bioinformática da Embrapa [Machine-actionable data management plans aligned with the FAIR principles for Embrapa's Multiuser Bioinformatics Laboratory]. Embrapa Agricultura Digital. <https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/1149101/1/Doc183-2022.pdf>

Torres, T. Z., Visoli, M. C., Falcão, P. R. K., Souza, M. I. F., Giachetto, P. F., Nhani, A., Jr, Cintra, L. C., Cunha, L. M. S., & Barbosa, L. A. F. (2022). Gestão de dados de pesquisa no Laboratório Multiusuário de Bioinformática da Embrapa [Research data management at Embrapa's Multiuser Bioinformatics Laboratory]. Athena. <https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1140031/1/LV-Gestao-dados-pesquisa-LMB-2022.pdf>

Veiga, V., Henning, P., Dib, S., Penedo, E., Lima, J. C., Silva, L. O. B., & Pires, L. F. (2019). Plano de gestão de dados FAIR: uma proposta para a Fiocruz [FAIR data management plan: A proposal for Fiocruz]. Liinc em Revista, 15(2), 275–286. <http://revista.ibict.br/liinc/article/view/5030/4349>