

Modelo de dados utilizado no repositório brasileiro de dados da pesquisa em ciência do solo (SoilData)

Data model used in the Brazilian soil science research data repository (SoilData)

Vitor Peruzzi de Almeida¹, Aline Mari Huf dos Reis² Taciara Zborowski Horst³,
Alessandro Samuel-Rosa⁴

RESUMO

Com o contínuo avanço da ciência aberta, surgem inúmeros desafios, e um dos mais prementes é a necessidade de padronização na disponibilização de dados compartilhados. O modelo inovador desenvolvido pelo repositório brasileiro de dados da pesquisa sobre o solo, conhecido como SoilData, tem como principal missão preencher essa lacuna crucial. A sua função central é simplificar a organização de dados, tornando-a acessível tanto para os especialistas que já produzem dados quanto para os novatos na área. O SoilData desempenha um papel fundamental na redução da disparidade entre os diversos formatos de dados utilizados na comunidade científica. Este modelo foi cuidadosamente projetado para se adaptar às distintas demandas e profissionais, apresentando-se como planilhas eletrônicas virtuais com cinco páginas ou abas. Após a etapa de organização de dados, torna-se possível o depósito no SoilData, proporcionando a disponibilização dos dados em âmbito internacional. Isso, por sua vez, contribui para o aumento do reconhecimento e valorização dos profissionais envolvidos na pesquisa do solo, exercendo um impacto significativo no avanço da ciência neste campo tão crucial.

PALAVRAS-CHAVE: Ciência aberta; Dados abertos; Modelo de dados.

ABSTRACT

With the continuous advancement of open science, numerous challenges emerge, and one of the most pressing is the need for standardization in the sharing of data. The innovative model developed by the Brazilian soil research data repository, known as SoilData, has as its primary mission to fill this crucial gap. Its central function is to simplify data organization, making it accessible to both experts who already produce data and newcomers in the field. SoilData plays a fundamental role in reducing the disparity among the various data formats used in the scientific community. This model has been carefully designed to adapt to diverse demands and professionals, presenting itself as virtual spreadsheets with five pages or tabs. After the data organization stage, it becomes possible to deposit data into SoilData, facilitating the global availability of data. This, in turn, contributes to increased recognition and appreciation of professionals involved in soil research, exerting a significant impact on the advancement of science in this critical field.

KEYWORDS: Open science; Open data; Data model.

¹Bolsista da Fundação Araucária. Santa Helena, Parana, Brasil. E-mail: vitorperuzi@gmail.com. ID Lattes: 6862089296579302.

²Pesquisadora de Pós-Doutorado. Universidade Federal de Goiás, Goiânia, Goiás, Brasil. E-mail: huf.aline@gmail.com. ID Lattes: 1570834132484121;

³Docente do curso de bacharelado em Agronomia e Engenharia Florestal da Universidade Tecnológica Federal do Paraná, Dois Vizinhos, Paraná, Brasil. E-mail: taciarahtaciatorst@utfpr.edu.br. ID Lattes: 6763043931071514.

⁴Docente do curso de bacharelado em Agronomia da Universidade Tecnológica Federal do Paraná, Santa Helena, Paraná, Brasil. E-mail: alessandrorosa@utfpr.edu.br. ID Lattes: 1609751519717461.

INTRODUÇÃO

Na era atual, em um contexto em que há uma grande quantidade de informações disponíveis e a colaboração é cada vez mais importante, a publicação dos dados de pesquisa em repositórios de dados se torna crucial (SAMUEL-ROSA et al., 2022). Nesse cenário, o repositório SoilData (<https://soildata.mapbiomas.org>) se destaca como um exemplo da ciência do solo brasileira (MAPBIOMAS, 2023a). Atendendo aos padrões FAIR, os dados são "localizáveis", possibilitando que o usuário os encontre de forma eficiente; "acessíveis" de forma aberta, sem a necessidade de criar uma conta ou passar por processos burocráticos; "interoperáveis", utilizando um formato universal de dados para facilitar o reuso; e "reutilizáveis", seguindo padrões de licenças de uso internacionalmente reconhecidos (HENNING et al., 2019).

A forma como se organiza e gerencia dados em repositórios desempenha um papel crucial no avanço da pesquisa científica (ANJOS, 2023). É notável como a organização e estruturação dos dados são fundamentais na pesquisa em ciência do solo. Um modelo de dados é essencial para garantir a interoperabilidade dos dados, tornando-os mais acessíveis, compreensíveis e utilizáveis.

O propósito deste estudo foi desenvolver um modelo de planilha eletrônica específico para a organização dos dados de pesquisa em ciência do solo destinados ao depósito no repositório SoilData.

MATERIAL E MÉTODOS

O modelo de dados foi desenvolvido para planilhas eletrônicas, que são amplamente utilizadas por pesquisadores e abrangem um público amplo, podendo ser usadas em qualquer sistema operacional e usando softwares livres. Os padrões adotados foram baseados em experiências de iniciativas locais e globais, construídos com base no conhecimento sobre a estruturação de dados de solo em publicações, relatórios, tabelas de artigos, levantamentos de solos, e outras fontes.

A premissa subjacente é que qualquer observação ou amostra de solo opera em quatro dimensões: coordenadas espaciais horizontais (x e y), temporal (t) e vertical (h). Essa abordagem foi adotada com o objetivo de simplificar a gestão dos dados, abrangendo armazenamento, compartilhamento e reuso eficientes. Durante o desenvolvimento da planilha modelo, cada página foi projetada com características exclusivas, mas mantendo uma interligação eficaz entre as informações em todas elas via definição de chaves primárias e estrangeiras. Isso evita redundâncias desnecessárias e minimiza a lacuna de informações em diferentes partes da planilha. Além disso, o modelo foi concebido para atender às diversas áreas de conhecimento do solo e a diferentes volumes de dados.

RESULTADOS E DISCUSSÃO

A planilha modelo desenvolvida é composta por cinco páginas ou abas, fornecendo uma compreensão abrangente da localização espaço-temporal, bem como das descrições morfo-mecânicas e metodologias. A primeira página, intitulada 'Identificação', armazena informações essenciais para uma compreensão inicial do trabalho, como o título, uma descrição detalhada do projeto, autor, identificador, fonte financiadora e licença de uso (Figura 1). Além disso, essa página inclui a citação dos autores para referências em trabalhos derivados. Essas informações são obrigatórias para o depósito dos dados no repositório SoilData, pois são as mesmas informações que deverão ser inseridas nos metadados do site (King, 2007).

Figura 1 – Demonstrativo da página 'Identificação' na planilha eletrônica desenvolvida pelo SoilData, cujos dados são essenciais para o depósito no repositório.

campo	valor
dados_id	ctb0050
dados_titulo	Projeto RADAMBRASIL. Folha SG.22 Curitiba, parte da folha SG.21 Asunción e folha SG.23 Iguape
dados_autor	Isaias Oenning; José Marcos Moser; Sérgio Hideiti Shimizu; Moacir de Jesus Rauen; Américo Pereira de Carvalho
organizacao_nome	Instituto Brasileiro de Geografia e Estatística (IBGE)
dados_descricao	Dados do solo de levantamento do Projeto RADAMBRASIL cobrindo o norte do Estado do Rio Grande do Sul, quase todo o Estado de Santa Catarina, grande parte do Estado do Paraná e pequena porção do sul do Estado de São Paulo. Inclui observações compiladas de trabalhos prévios realizados na área, totalizando 106 perfis completos e __ pontos de amostras extra. As observações do solo realizadas especificamente para este trabalho ocorreram, quase sempre, em cortes de estradas, sendo em alguns casos utilizadas tradagens, totalizando 64 perfis completos e __ pontos de amostras extras. As descrições do solo seguiram as normas e definições do SNLCS - EMBRAPA e as constantes na Soil Taxonomy (1975). As análises laboratoriais foram realizadas nos laboratórios do SUDESUL e SNLCS, que podem diferir entre si. A classificação dos solos segue as normas do SNLCS. No total, são 170 perfis de solos, 200 pontos de amostras extra (o documento original sugere 201 pontos de amostras extra) e 05 pontos de amostras para avaliação da fertilidade. Os dados dos trabalhos prévios são omitidos, exceto a classificação taxonômica, que foi atualizada.
dados_publicacao	https://www.dropbox.com/s/mkscfu93nhskc8t/OenningEtAl2018.pdf?dl=0
dados_idioma	português
dados_licenca	CC BY 4.0 Atribuição
dados_completude	aproximadamente completos (entre 70 e 95% de cobertura)

Fonte: Elaborado pelos autores (2023).

O armazenamento dos dados ambientais do local de coleta e amostragem de solo, que auxilia na compreensão do contexto temporal e espacial, é detalhado na página 'Evento' (Figura 2). A coluna "Identificação do Evento" é uma chave primária, sendo de preenchimento obrigatório, uma vez que é essencial para o preenchimento da próxima página. Além disso, é altamente recomendado preencher as colunas de dia, mês e ano, longitude e latitude, a fim de permitir uma localização precisa do evento no espaço (x e y) e no tempo (t).

Figura 2 – Página 'Evento' na planilha eletrônica desenvolvida pelo SoilData, contendo detalhes de armazenamento de dados ambientais.

identificação do evento	página	ordem no doc	evento	mês	evento	ano	descrição da localização	longitude (grau)	latitude (grau)	uso atual da terra
PERFIL-115	464	4	3	8	1980	A 8 km da BR-282 da entrada de Curitibaanos, direção a Santa Cruz do Peri, Município de Curitibaanos (SC). 27°21'S e 50°36'WGr.	-50,60	-27,35	Pastagem natural.	
PERFIL-059	472	12	24	7	1978	No entroncamento da BR280 para São Pascoal. No município de Irineópolis (SC). 26°16'S 50°42'WGr.	-50,70	-26,27	Nenhum.	
PERFIL-159	477	18	6	11	1978	A 12 km de Três de Maio na estrada para Giruá. Município de Independência (RS). 27°53'S e 54°16'WGr.	-54,27	-27,88	Cultura de aveia.	
PERFIL-094	482	22	15	9	1978	A 4 km de Catanduvas em direção a Água Doce. Município de Catanduvas (SC). 27°02'S e 51°37'WGr.	-51,62	-27,03	Nenhum	
PERFIL-089	483	23	#N/D	#N/D	#N/D	Rodovia BR-282, Ponte Serrada - Joaçaba a 10 km de Ponte Serrada. Município de Ponte Serrada (SC). 26°55'S e 50°42'WGr.	-51,93	-26,92	Reflorestamento com Pinus.	

Fonte: Elaborado pelos autores (2023).

A página denominada 'Camada' contém dados das características físicas, biológicas, morfo-mecânicas e químicas das amostras (Figura 3). É essencial que esta seção inclua a "Identificação do Evento", aqui incluída como uma chave estrangeira, que deve corresponder exatamente aos dados da página 'Evento', permitindo a interligação com os mesmos perfis descritos na página anterior. Além disso, a inclusão das profundidades superior e inferior é obrigatória para uma compreensão mais completa da camada, definindo a dimensão vertical dos dados (h).

Figura 3 – Página 'Camada' na planilha eletrônica desenvolvida pelo SoilData, contendo características físicas, biológicas e químicas das amostras com Interligação aos dados da página 'Evento'.

A	B	C	D	E	F	G	H	I	J
Identificação do evento	Identificação da camada	Descrição morfológica	Profundidade superior (cm)	Profundidade inferior (cm)	Calhau 20-2cm (%)	Cascalho 20-2cm (%)	Carbono (C)	Nitrogênio (N)	Matéria orgânica (MO)
PERFIL-115	A1	bruno-escuro (10YR 3/3); muito argiloso; fraca pequena blocos subangulares e fraca pequena granular; friável, plástico e pegajoso; transição plana e gradual.	0	15	0,00	0,00	3,80	0,23	6,55
PERFIL-115	A3	bruno-escuro (10YR3, 5/3); muito argiloso; fraca, pequena blocos subangulares e fraca pequena granular; friável, plástico e pegajoso; transição gradual.	15	35	0,00	0,00	2,80	0,16	4,83
PERFIL-115	B1	bruno-amarelado-escuro (10YR 4/4); muito argiloso; moderada pequena média blocos subangulares; firme, plástico e pegajoso; transição plana e difusa.	35	65	0,00	0,00	1,80	0,12	3,10
PERFIL-115	B21	bruno-forte (5,5YR 5/6); muito argiloso; moderada média blocos subangulares; firme, ligeiramente plástico e pegajoso, transição plana e difusa.	65	100	0,00	1,30	0,70	0,08	1,21
PERFIL-115	B22	bruno-avermelhado (5YR 4/4); muito argiloso; moderada média blocos subangulares; firme, ligeiramente plástico e pegajoso; transição plana e difusa.	100	150	0,00	0,00	0,80	0,08	1,38
PERFIL-115	B23	vermelho-amarelado (5YR 4/6); muito argiloso; moderada média blocos subangulares; firme, ligeiramente plástico e pegajoso.	150	190	0,00	0,00	0,40	0,06	0,69
PERFIL-059	A1	bruno-avermelhado-escuro (5YR 3/2,5); argila; forte pequena e média granular; muito friável, ligeiramente plástico e ligeiramente pegajoso; transição difusa.	0	30	0,00	0,00	5,40	0,34	9,39
PERFIL-059	A3	bruno-avermelhado-escuro (5YR 3/3); argila; fraca pequena granular; muito friável, ligeiramente plástico e ligeiramente pegajoso; transição gradual.	30	80	0,00	0,00	2,00	0,15	3,45

Fonte: Elaborado pelos autores (2023).

Com o objetivo de promover o máximo de reuso dos dados, foi criada a página "Método", que detalha os métodos de coleta em campo e laboratoriais aos quais as amostras foram submetidas para a obtenção dos dados (Figura 4). O preenchimento de todas as colunas desta página é obrigatório para fornecer o esclarecimento máximo sobre

os métodos utilizados. É essencial destacar que todos os títulos das colunas das páginas 'Evento' e 'Camada' devem estar referenciados na página 'Método'. Nessa seção, deve ser descrito o método de coleta, a descrição e a análise do solo para cada variável, além de indicar sua unidade de medida.

Figura 4 – Página 'Método' na planilha eletrônica desenvolvida pelo SoilData, contendo, descrição dos métodos de coleta, análise e unidades de medida para variáveis do solo contidos nas páginas 'Evento' e 'Camada'.

identificação da página	títulos dos eventos e camadas	unidade de medida	metodo de extração
evento	Classificação do solo	#N/D	Classificação segundo SIBCS (EMBRAPA, 2006)
evento	Longitude (grau)	grau	mapa topográfico na escala 1:500.000
evento	Latitude (grau)	grau	mapa topográfico na escala 1:500.000
camada	Profundidade superior (cm)	cm	A coleta foi realizada nas profundidades de 0-10, 10-20 e 20-30 cm em trincheiras distantes em 100 metros
camada	Profundidade inferior (cm)	cm	A coleta foi realizada nas profundidades de 0-10, 10-20 e 20-30 cm em trincheiras distantes em 100 metros

Fonte: Elaborado pelos autores (2023).

Finalmente, a página 'Histórico' descreve as alterações feitas no arquivo, identificando a natureza dessas mudanças e os responsáveis por elas, conforme evidenciado na Figura 5. Qualquer modificação no documento deve ser obrigatoriamente registrada. Isso permite conhecer e compreender as mudanças ocorridas nos dados ao longo do tempo.

Figura 5 – Detalhes da página 'Histórico' na planilha eletrônica desenvolvida pelo SoilData, contendo o registro de alterações no documento com descrição das mudanças e identificação dos responsáveis.

A	B	C	D
Versão dos dados	Data da versão	Nome ou email do editor	Descrição da mudança
0.0.0	07/07/2020	Alessandro Samuel-Rosa	Entrada no FEBR (dados digitalizados pela ESTATMG)
0.0.1	08/07/2020	Alessandro Samuel-Rosa	Atualiza consistência do conjunto de dados
0.1.0	02/10/2020	Alessandro Samuel-Rosa	Resgata latitude e longitude da descrição de localização
0.2.0	16/05/2021	Alessandro Samuel-Rosa	Melhora identificação das variáveis
1.0.0	09/08/2021	Alessandro Samuel-Rosa	Adiciona colunas para as variáveis camada_id e amostra_id
1.1.0	22/01/2022	Alessandro Samuel-Rosa	define códigos dos campos da tabela com as observações
1.1.1	06/06/2022	Alessandro Samuel-Rosa	define códigos dos campos da tabela camada
1.1.2	01/11/2022	Alessandro Samuel-Rosa	Ajusta códigos dos campos de identificação das tabelas evento e camada

Fonte: Elaborado pelos autores (2023).

Utilizando as cinco páginas de planilha eletrônica, é possível criar conjuntos de dados facilmente interpretáveis para reuso pela comunidade científica, impulsionando ainda mais as pesquisas na área de solos no Brasil.

CONSIDERAÇÕES FINAIS

A estrutura cuidadosamente elaborada do modelo de dados do repositório SoilData beneficia tanto especialistas já atuantes nesse campo quanto aqueles que estão começando sua jornada na área. Após extensos estudos sobre o modelo de dados, foi desenvolvida a primeira versão que consiste em cinco tabelas: Identificação, Evento, Camada, Método e Histórico. Este modelo já está em uso por pesquisadores no Brasil para o depósito de dados no SoilData. Reconhecemos, no entanto, que pode ser necessário fazer ajustes para atender a certos tipos de dados e planejamos atualizá-lo periodicamente para melhor atender às demandas da comunidade científica brasileira.

Agradecimentos

Agradecemos à Universidade Tecnológica Federal do Paraná (UTFPR) pelo espaço de aprendizado, à Fundação Araucária, Instituto Arapyaú e CNPq pelo financiamento do projeto, bem como às bolsas de iniciação científica da UTFPR e ao suporte oferecido pela FUNTEF, FUNAPE e IAMAP. Também expressamos nossa gratidão aos colaboradores do Laboratório de Pedometria e aos parceiros externos.

Conflito de interesse

Não há conflito de interesse

REFERÊNCIAS

ANJOS, M. A. **Repositório de Dados do Solo Brasileiro**: Concepção e Implementação. Disponível em: https://drive.google.com/file/d/1Rdrl-g8A6TNTq2EObsXGLwBG-4lhqE8j/view?usp=sharing_eil_se_dm&ts=65098453&usp=embed_facebook>. Acesso em: 19 set. 2023.

HENNING, Patricia Corrêa et al. GO FAIR e os princípios FAIR: o que representam para a expansão dos dados de pesquisa no âmbito da Ciência Aberta. **Em Questão**, [s. l.], p. 389–412, 2019. Disponível em: <https://www.seer.ufrgs.br/index.php/EmQuestao/article/view/84753>. Acesso em: 21 out. 2023.

KING, Gary. An Introduction to the Dataverse Network as an Infrastructure for Data Sharing. *Sociological Methods & Research*, [s. l.], v. 36, n. 2, p. 173–199, 2007. Disponível em: <http://journals.sagepub.com/doi/10.1177/0049124107306660>. Acesso em: 21 out. 2023.

MAPBIOMAS. "Projeto MapBiomias – Coleção 8 da Série Anual de Mapas de Cobertura e Uso de Solo do Brasil. MapBiomias Data, 2023a. Disponível em: <<http://mapbiomas.org>>.

XIII Seminário de Extensão e Inovação
XXVIII Seminário de Iniciação Científica e Tecnológica da UTFPR

Ciência e Tecnologia na era da Inteligência Artificial: Desdobramentos no Ensino Pesquisa e Extensão
20 a 23 de novembro de 2023 - *Campus Ponta Grossa, PR*



SAMUEL-ROSA, Alessandro et al. CURADORIA E GESTÃO DE DADOS E METADADOS NO REPOSITÓRIO DE DADOS DO SOLO BRASILEIRO (FEBR). OPEN SCIENCE RESEARCH VII, [s. l.], v. 7, n. 1, p. 1288–1300, 2022. Disponível em: <http://www.editoracientifica.com.br/artigos/curadoriae-gestao-de-dados-e-metadados-no-repositorio-de-dados-do-solo-brasileirofebr>. Acesso em: 21 out. 2023.