

<https://eventos.utfpr.edu.br/sicite/sicite2020>

Curadoria de dados no Repositório Brasileiro Livre para Dados Abertos do Solo (FEBR)

Data curation in the Free Brazilian Repository for Open Soil Data (FEBR)

RESUMO

Gabriel Panca Santos

gabrielpanca@alunos.utfpr.edu.br

Discente do 2º período do Curso de Agronomia, Universidade Tecnológica Federal do Paraná, Santa Helena, PR, Brasil.

Alessandro Samuel Rosa

alessandrorosa@utfpr.edu.br

Docente do Curso de Agronomia, Universidade Tecnológica Federal do Paraná, Santa Helena, PR, Brasil.

Este artigo apresenta sugestões e soluções para a curadoria de dados no Repositório Brasileiro Livre para Dados Abertos do Solo (FEBR), com destaque para a melhoria da qualidade da descrição dos métodos de análise de solo. Os métodos e procedimentos de curadoria adotados tiveram o intuito de melhorar a consistência documental e a qualidade dos dados. O resgate de informações dos métodos de análise química e física do solo teve prioridade e as devidas informações coletadas foram anexadas às tabelas. Para a maioria dos conjuntos de dados, essas informações foram obtidas diretamente da seção 'Material e Métodos' dos trabalhos originais de levantamento de solos publicados pela Embrapa Solos. Em geral, foram observadas inúmeras inconsistências nos conjuntos de dados. Dentre as principais estão a ausência de descrição dos métodos analíticos no trabalho original, falta de dados, parcialmente completos ou até mesmo esquecimento por parte dos autores são frequentes. Uma estratégia que se mostrou importante foi contatar pesquisadores da área para sanar dúvidas, especialmente no caso de conjuntos de dados com mais de 40 anos desde sua publicação. O artigo também faz o levantamento do padrão de descrição de métodos analíticos em trabalhos que se basearam na edição de 1979 do Manual de Métodos de Análise de Solo da Embrapa. Em geral, a curadoria dos dados mostrou alguns pontos que são necessários para a melhoria documental de futuros projetos. É necessária apresentação dos métodos com mais cautela e completude. Essa apresentação dos métodos é imprescindível desde a captação desses dados para um repositório quanto para um pesquisador que terá como base esse projeto que já fora realizado.

PALAVRAS-CHAVE: Metadados. Padronização. Harmonização.

Recebido: 19 ago. 2020.

Aprovado: 01 out. 2020.

Direito autoral: Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



ABSTRACT

This article presents suggestions and solutions for data curatorship in the Free Brazilian Repository for Open Soil Data (FEBR), with emphasis on improving the quality of the description of soil analysis methods. The methods and curatorial procedures adopted were intended to improve documental consistency and data quality. The recovery of information from chemical and physical soil analysis methods had priority and the appropriate information collected was added to the existing tables. For most data sets, this information was obtained directly from the 'Material and Methods' section of the original soil survey works published by Embrapa Solos. In general, numerous inconsistencies were observed in the data sets. Among the main ones are the lack of



description of the analytical methods in the original work, partial or complete lack of data—forgetfulness by the authors are frequent. A strategy that proved to be important was to contact researchers in the field to solve doubts, especially in the case of data sets with more than 40 years since its publication. The article also surveys the pattern of description of analytical methods in papers that were based on the 1979 edition of Embrapa's Manual of Soil Analysis Methods. In general, the data curation showed some points that are necessary for the documental improvement of future projects. It is necessary to present the methods with more caution and completeness. This presentation of methods is essential both for the capture of this data for a repository as for a researcher who will conduct work based on an already existing project.

KEYWORDS: Metadata. Standardization. Harmonization.

INTRODUÇÃO

Ao longo das últimas décadas, uma quantidade significativa de dados de solo foram produzidos para fins de pesquisa e exploração do território. Entretanto, durante os levantamentos, a quantidade de informações obtidas são elevadas e, na maioria das vezes, os resultados ficaram armazenados em documentos físicos, alguns com mais de 40 anos. O Repositório Brasileiro Livre para Dados Abertos de Solo (FEBR) têm sido fundamental para o agrupamento, armazenamento e compartilhamento dessas informações para a posteridade, maximizando o aproveitamento de recursos alocados na ciência do solo, o que impulsiona a colaboração entre cientistas e fomenta o progresso científico. Deste modo, o cerne do projeto é contribuir para acabar com a fome, promover a agricultura sustentável, deter e reverter a degradação da terra, dentre outros desafios.

Uma das etapas da construção de um repositório de dados, e também sua manutenção, é a curadoria, onde não consiste somente na captação dos dados documentais, mas também em levar toda a informação coletada para uma warehouse, parecido com um banco de dados, se tornando futuramente um repositório de dados onde os pesquisadores e cientistas da área terão fácil acesso aos dados (SANTOS, 2004). Essa curadoria dos dados tem pleno cuidado com as informações de modo que seja adicionados em lugares seguros, como servidores e outros métodos. Toda essa transposição dos dados e seus subprodutos são meticulosamente cuidados para que as informações não sejam divergentes da realidade. Toda a cadeia de armazenamento dos dados é dimensionada com softwares, de modo que a organização, a conclusão e preservação das informações sejam pensadas antes da captação dos dados nos documentos analisados (CAVANILLAS et al., 2016). Toda essa organização tem um papel imprescindível para qualquer repositório de dados, e o mesmo ocorre no FEBR de modo a validar as informações contidas nessa plataforma a qual disponibiliza os dados coletados.

Toda a parte organizacional realizada nos documentos no sistema, pós tratamento e curadoria dos dados coletados, pode ser descrita como harmonização dos dados. A harmonização é o “processo de reunir dados de diferentes origens, convenções de nomenclatura e colunas e transformá-los em um conjunto de dados coeso” (STEWART, 2017), ou seja, a união de inúmeros dados de diferentes fontes em um só local. A harmonização tem um papel muito

importante para bancos de dados, por exemplo, tendo em sua estrutura fatores que simplifiquem a localização e disponibilização das informações de forma mais dinâmica, maximizando o tempo entre pesquisa e resposta (SANTOS, 2004).

O objetivo deste artigo é apresentar sugestões e soluções para a curadoria de dados no FEBR, com destaque para a melhoria da qualidade da descrição dos métodos de análise de solo.

MATERIAL E MÉTODOS

A curadoria de dados foi realizada em onze conjuntos de dados já publicados no FEBR envolvendo a verificação da consistência documental e melhoria da qualidade das informações. Em todos os conjuntos de dados se procedeu o resgate da descrição dos métodos analíticos para posterior definição de códigos apropriados para as variáveis, armazenados na tabela metadado. A prioridade foi dada a métodos laboratoriais de análise química e física do solo, cujos dados estão armazenados na tabela camada. Para alguns conjuntos de dados também foi realizada a digitalização de dados faltantes nos métodos de análise. Os dados descritivos dos métodos analíticos foram obtidos dos projetos de pesquisas nacionais, realizados por pesquisadores em levantamentos pedológicos e estão disponíveis em meio digital na biblioteca da Embrapa Solos (www.embrapa.br/biblioteca). Os dados descritivos foram encontrados na seção Material e Métodos desses documentos.

Quanto aos métodos analíticos ausentes, inicialmente, não foram adicionadas informações, pois, para a determinação de algumas propriedades de solo possuem mais de um método analítico. Entretanto essa ausência de dados nos métodos analíticos, pós curadoria e processamento dos dados, mudou com objetivo de anexar informações antes ausentes desde que se encaixe nos parâmetros observados futuramente.

RESULTADOS E DISCUSSÃO

Ao todo, foi realizada a verificação da consistência documental e melhoria da qualidade dos onze conjuntos de dados já publicados no FEBR. Durante o resgate das informações descritivas dos métodos analíticos, foram encontradas importantes inconsistências, o que dificultou a captação dos dados referente aos Métodos de Análise da Embrapa Solos de 1979 (Tabela 1). A principal inconsistência foi a ausência parcial ou completa da descrição dos métodos utilizados. Alguns trabalhos apresentam apenas uma descrição parcial dos métodos. Outros apenas citam o nome dos métodos utilizados, mas não apresentam qualquer descrição.

A tabela 2 apresenta os padrões utilizados nas análises de solo, anexando as descrições de métodos analíticos encontradas em diferentes trabalhos de levantamento de solos, com anos diferentes entre eles, que utilizaram a edição de 1979 do Manual de Métodos de Análise de Solo da Embrapa.

Tabela 1 - Verificação de consistência e melhoria de qualidade dos seguintes conjuntos de dados.

Projeto	Completo	Parcial	Cita
Flores e Alba (2015)		X	X
Motta et al. (2004)			X
Coelho et al. (2005a)			X
Lima e Rêgo (1999)			X
Coelho et al. (2005b)		X	X
Lima et al. (1999)			X
Santos (1982)		X	X
Witterri et al. (1983)	X		X
Santos et al. (1982b)		X	X
Santos et al. (1982a)	X		X
Gama e Soares (1984)	X		X

Fonte: Autoria própria (2020)

Sem dúvidas um dos maiores problemas encontrados nos trabalhos são as ausências de dados, ao que tangem os métodos analíticos, em especial a forma com que se chegam em determinado resultado, resultados esses que conseguimos nos métodos de análise. Com a ausência dos dados o projeto fica incompleto atrapalhando o resultado final planejado sendo esse disponibilizar os dados para o sistema do FEBR.

Observando-se mais a fundo é notável uma certa disparidade entre os projetos e ausência de forma completa nos métodos de análise. Ao indagar a questão para alguns especialistas no tema, obtive respostas interessantes. Foi relatado que isso é um problema constante, não só no Brasil, mas mundial, às vezes podendo ou não se tratar de uma política interna e editorial. Esses problemas são comuns e cruciais para um projeto, ainda mais para a sua compilação de dados.

Tabela 2 - Descrição de métodos de análise química e física do solo pelo Manual de Métodos da Embrapa Solos de 1979 (VETTORI; PIERANTONI, 1968) segundo três conjuntos de dados publicados no FEBR (GAMA; SOARES, 1984; SANTOS, 1982; WITTERRI et al., 1983)

Nome	Descrição
pH em água (GAMA; SOARES, 1984)	Determinado potenciometricamente na suspensão solo-líquido de 1:2,5 com tempo de contato não inferior a uma hora e agitação da suspensão imediatamente antes da leitura. Suspensão solo-água na proporção 1:1 no caso de horizonte sulfúrico ou material sulfídrico (Solos Tiomórficos).
pH em água (WITTERRI et al., 1983)	Determinado potenciometricamente na suspensão solo-líquido de 1:2,5 com tempo de contato não inferior a uma hora e agitação da suspensão imediatamente antes da leitura.
pH em KCl (GAMA; SOARES, 1984; SANTOS, 1982; WITTERRI et al., 1983)	Determinado potenciometricamente em suspensão salina de KCl 1 N utilizando relação solo-líquido de 1:2,5, com tempo de contato não inferior a uma hora e agitação da suspensão imediatamente antes da leitura.
Carbono orgânico (GAMA; SOARES, 1984; SANTOS, 1982; WITTERRI et al., 1983)	Oxidação da matéria orgânica pelo bicromato de potássio 0,4 N em meio sulfúrico. Quantificação por titulação com sulfato ferroso 0,1 N.
Fósforo assimilável (GAMA; SOARES, 1984; SANTOS, 1982; WITTERRI et al., 1983)	Extração com solução de HCl 0,05 N e H ₂ SO ₄ 0,025 N (North Carolina). Quantificação por colorimetria em presença de ácido ascórbico
Cálcio trocável (GAMA; SOARES, 1984; SANTOS, 1982; WITTERRI et al., 1983)	Extração com solução de KCl N na proporção 1:20. Quantificado com solução de EDTA 0,0125 M.
Cálcio e magnésio trocáveis (GAMA; SOARES, 1984; SANTOS, 1982; WITTERRI et al., 1983)	Extração com solução de KCl N na proporção 1:20. Quantificação com solução de EDTA 0,0125 M.

Autoria própria (2020)

CONCLUSÕES

Após os estudos e levantamento dos dados para a conclusão do projeto, pode ser compreendido dois principais pontos, são eles:

- a. As informações, no que diz respeito aos métodos de análise de solo, quando é feito menção a utilização dos Métodos de Análise de Solos Embrapa 1979, pode ser adotado as informações compreendidas na tabela 2, devido a consistência documental da utilização dos métodos

laboratoriais da Embrapa observados em diferentes projetos com anos e pesquisadores diferentes.

- b. Com relação direta a descrição dos métodos de análise de solo adotado, em cada distinto projeto, muitas vezes sofre uma defasagem na descrição da forma que se chegou em determinado resultado. Muitos trabalhos possuem somente parte das descrições e alguns e realizam apenas a menção sem descrição nenhuma das análises.

Contudo fica evidente, após a captação dos dados de diferentes documentos, a ausência de parte ou de todos os métodos de análise utilizados. É necessário um maior cuidado com esse tipo de informação, ainda mais para sua futura curadoria ou para algum trabalho de pesquisa.

AGRADECIMENTOS

Este trabalho foi financiado pela Fundação Araucária de Apoio ao Desenvolvimento Científico e Tecnológico do Paraná na forma de bolsa de iniciação científica (Edital UTFPR/PROPPG – 02/2019). Os autores são gratos à Marcos Alexandre dos Anjos (UTFPR), Maiara Pusch (UNICAMP) e Taciara Horst-Heinen (UFSM) pelos comentários em uma versão preliminar do artigo. Os autores também são gratos à Luciana Sampaio (Embrapa Solos), João Herbert (Embrapa Milho e Sorgo), Wenceslau Geraldes Teixeira (Embrapa Solos), pelas informações prestadas sobre os conjuntos de dados.

REFERÊNCIAS

CAVANILLAS, J. M. et al. (EDS.). **New horizons for a data-driven economy: a roadmap for usage and exploitation of big data in Europe**. Cham: Springer International Publishing AG, 2016.

COELHO, M. R. et al. **Levantamento pedológico de uma área-piloto relacionada ao Projeto BiosBrasil (Conservation and Sustainable Management of Below-Ground Biodiversity: Phase I), Município de Benjamin Constant (AM): janela 6**. Rio de Janeiro: Embrapa Solos, 2005a.

COELHO, M. R. et al. **Solos das áreas-piloto do Projeto BiosBrasil (Conservation and Sustainable Management of Below-Ground Biodiversity: phase I), Município de Benjamin Constant, Estado do Amazonas**. Rio de Janeiro: [s.n.].

FLORES, C. A.; ALBA, J. M. F. **Zoneamento edáfico de culturas para o município de Santa Maria - RS, visando o ordenamento territorial**. Brasília, DF: Embrapa, 2015.

GAMA, J. R. N. F.; SOARES, A. F. **Levantamento de reconhecimento dos solos e avaliação da aptidão agrícola das terras de uma área de colonização no Município de Urucará, Estado do Amazonas**. Rio de Janeiro: EMBRAPA-SNLCS, 1984.

LIMA, A. A. C. et al. **Levantamento de reconhecimento de média intensidade dos solos e avaliação da aptidão agrícola das terras do Pólo Juruá-Solimões, Amazonas.** Rio de Janeiro: Embrapa Solos, 1999.

LIMA, A. A. C.; RÊGO, R. S. **Levantamento de reconhecimento de alta intensidade dos solos e avaliação da aptidão agrícola das terras da área de Novo Paraíso, Roraima.** Rio de Janeiro: Embrapa Solos, 1999.

MOTTA, P. E. F. D.; BARUQUI, A. M.; SANTOS, H. G. D. **Levantamento de reconhecimento de média intensidade dos solos da região do Alto Paranaíba, Minas Gerais.** Rio de Janeiro, RJ: Embrapa Solos, 2004. Rio de Janeiro: Embrapa Solos, 2004.

SANTOS, H. G. DOS et al. **Levantamento de reconhecimento de média intensidade dos solos e avaliação da aptidão agrícola das terras de uma área sob influência dos rios Araguari, Falsino e Tartarugal Grande-Território Federal do Amapá.** Rio de Janeiro: EMBRAPA-SNLCS, 1982a.

SANTOS, R. D. DOS et al. **Levantamento de reconhecimento de baixa intensidade dos solos e avaliação da aptidão agrícola das terras do projeto de colonização Apiaú-Território Federal de Roraima.** Rio de Janeiro: [s.n.].

SANTOS, R. D. D. **Levantamento de reconhecimento de baixa intensidade dos solos e avaliação da aptidão agrícola das terras de área piloto no Município de Barreirinha - Estado do Amazonas.** Rio de Janeiro: EMBRAPA-SNLCS, 1982.

SANTOS, V. Projecto e implementação de sistemas de Data Warehousing. 2004.

STEWART, L. **Dados 101: O que é Harmonização de Dados?** Disponível em: <<https://datorama.com/blog/2017/04/14/data-101-what-is-data-harmonization/>>. Acesso em: 4 ago. 2020.

VETTORI, L.; PIERANTONI, H. **Análise granulométrica: novo método para determinar a fração argila.** Rio de Janeiro: Equipe de Pedologia e Fertilidade do Solo, 1968.

WITTERRI, K. P. et al. **Levantamento de reconhecimento de média intensidade dos solos e avaliação da aptidão agrícola das terras de 21.000 hectares no Município de Tefé, Amazonas.** Rio de Janeiro: EMBRAPA-SNLCS, 1983.