

Repositório Brasileiro Livre para Dados Abertos do Solo

Manual do Usuário

*Alessandro Samuel-Rosa, Ricardo S. D. Dalmolin, Paulo I. Gubiani, Diego J. Gris, Nicolas
A. Rosin*

25 November 2017

Apresentação

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nome do Chefe do Departamento de Solos

Chefe do Departamento de Solos, UFSM

Prefácio

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Estrutura do documento

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Convenções

Inúmeras siglas são utilizadas ao longo do texto. O significado de cada uma delas, bem como o respectivo endereço na Internet – quando aplicável – são listados a seguir:

- EPSG: European Petroleum Survey Group (<http://www.spatialreference.org/ref/epsg/>)
- **febr**: Repositório Brasileiro Livre para Dados Abertos do Solo (<http://coral.ufsm.br/febr/>)
- INDA: Infraestrutura Nacional de Dados Abertos (<http://wiki.dados.gov.br/>)
- SISB: Sistema de Informação de Solos Brasileiros (<https://www.sisolos.cnptia.embrapa.br/>)

Agradecimentos

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Sobre os autores

Alessandro Samuel-Rosa

Alessandro estudou Ciência do Solo e trabalha na Universidade Federal de Santa Maria (UFSM) desde o final de 2016. Após seus primeiros trabalhos sobre as interações entre o uso da terra e a qualidade do solo na UFSM, em 2012 mudou para o campo da Pedometria. A colaboração com pesquisadores do Centro Nacional de Pesquisas sobre Solos (Embrapa), do International Soil Reference and Information Centre (ISRIC), da Wageningen University and Research Centre (WUR) e da Universidade Federal Rural do Rio de Janeiro (UFRRJ) permitiu-lhe dominar múltiplos métodos e técnicas pedométricas. Seu principal interesse científico é a modelagem espacial do solo com foco em estratégias de amostragem, seleção e calibração de modelos e análise de incerteza. Alessandro publicou recentemente alguns artigos relevantes sobre esses tópicos em revistas internacionais. Alessandro também é o autor de dois pacotes para o R, os quais são amplamente utilizados, cada um com mais de 200 downloads por mês. (Saiba mais aqui)

Ricardo S. D. Dalmolin

Ricardo possui graduação e mestrado em Agronomia pela UFSM e doutorado em Ciências do Solo pela Universidade Federal do Rio Grande do Sul (UFRGS). Atualmente é professor associado da UFSM, atuando no Programa de Pós-Graduação em Ciência do Solo (PPGCS) e no Programa Pós-Graduação em Educação Ambiental. Ricardo é consultor de diversas revistas científicas e líder do Grupo de Pesquisa do CNPq *Gênese, composição e comportamento dos solos do RS*. Ricardo tem experiência na área ambiental e agronômica, com ênfase em gênese, morfologia e classificação do solo, atuando principalmente nos temas ‘levantamento de solos’, ‘classificação de solos’, ‘aptidão de uso das terras’, ‘sensoriamento remoto’ e ‘mapeamento digital de solos’. Ainda foi presidente do Núcleo Regional Sul da Sociedade Brasileira de Ciência do Solo, membro do Conselho Diretor da Sociedade Brasileira de Ciência do Solo, e responsável pelo Museu de Solos do RS. (Saiba mais aqui)

Paulo I. Gubiani

Paulo possui graduação em Agronomia, e mestrado e doutorado em Ciência do Solo pela UFSM, tendo sido foco de sua formação acadêmica o estudo da compactação do solo e seu efeito sobre o desenvolvimento e produtividade das culturas agrícolas. Atualmente Paulo é professor da área de Física do Solo na UFSM,

desenvolvendo estudos, principalmente, sobre os seguintes temas: ‘estrutura do solo’, ‘compactação do solo’, ‘compressibilidade do solo’, ‘relação entre propriedades físicas do solo e crescimento de plantas’ e ‘balanço hídrico’. Paulo também tem grande experiência no uso e desenvolvimento de aplicativos em VBA para planilhas eletrônicas como o Microsoft Office Excel. Recentemente, Paulo publicou importante revisão bibliográfica sobre indicadores físicos de qualidade do solo, numa tentativa de apontar para novos caminhos para a pesquisa em Física do Solo no Brasil. (Saiba mais aqui)

Diego J. Gris

Diego é acadêmico do curso de Agronomia da Universidade Federal de Santa Maria (UFSM). Bolsista de Iniciação Científica no grupo de Pedologia da UFSM. Estudou por um ano (2013) na Cornell University - EUA, com bolsa de graduação sanduíche (CAPES) do programa Ciência sem Fronteiras. (Saiba mais aqui)

Nicolas A. Rosin

Nícolas é acadêmico do curso de Agronomia na Universidade Federal de Santa Maria (UFSM) e bolsista de Iniciação Científica no Laboratório de Pedologia, atuando na área de levantamento e classificação de solos, assim como pedometria. Possui experiência na área da Agronomia com ênfase em Ciência do Solo. (Saiba mais aqui)

Conteúdo

| | |
|---|-----------|
| Apresentação | 3 |
| Prefácio | 5 |
| Estrutura do documento | 5 |
| Convenções | 5 |
| Agradecimentos | 6 |
| Sobre os autores | 7 |
| Alessandro Samuel-Rosa | 7 |
| Ricardo S. D. Dalmolin | 7 |
| Paulo I. Gubiani | 7 |
| Diego J. Gris | 8 |
| Nicolas A. Rosin | 8 |
| 1 Introdução | 15 |
| 1.1 Organização dos Dados | 15 |
| 2 Conjunto de dados | 17 |
| 2.1 Identificação | 18 |
| 2.2 Descrição | 18 |
| 2.3 Versionamento | 19 |
| 2.4 Licença de uso | 20 |
| 2.5 Organização | 20 |
| 2.6 Autoria | 21 |
| 2.7 Contribuições | 22 |
| 2.8 Referências | 22 |
| 2.9 Indexação | 22 |
| 3 Observações no espaço e no tempo | 25 |
| 3.1 Identificação | 25 |
| 3.2 Coordenada temporal | 25 |
| 3.3 Coordenadas espaciais | 25 |
| 3.4 Localização | 27 |
| 3.5 Amostragem | 27 |

| | | |
|----------|---|-----------|
| 4 | Amostras de camadas e horizontes | 29 |
| 4.1 | Identificação | 29 |
| 4.2 | Coordenadas verticais | 29 |
| 4.3 | Variáveis | 30 |
| 5 | Dados sobre os dados (metadados) | 31 |

Lista de Figuras

Lista de Tabelas

| | | |
|-----|--|----|
| 2.1 | Representação da estrutura da tabela <i>dataset</i> | 17 |
| 2.2 | Itens utilizados para a identificação do conjunto de dados. | 18 |
| 2.3 | Itens utilizados para o versionamento do conjunto de dados. | 19 |
| 2.4 | Itens utilizados para a identificação da organização responsável pelo conjunto de dados. . . . | 21 |
| 2.5 | Itens utilizados para a identificação dos autores do conjunto de dados. | 21 |
| 2.6 | Itens utilizados para a identificação dos autores de contribuições para o conjunto de dados. . | 22 |
| 2.7 | Itens utilizados para a indexação do conjunto de dados. | 23 |
| 3.1 | Itens usados para especificar a localização espacial das observações do solo. | 25 |
| 3.2 | Códigos EPSG dos sistemas de referência de coordenadas geográficas para o Brasil. | 26 |
| 3.3 | Códigos EPSG dos sistemas de referência de coordenadas projetadas para o Brasil. | 26 |
| 3.4 | Fontes de coordenadas espaciais. | 27 |

Introdução

Nós acreditamos que o uso sustentável do solo para aplicações ambientais no Brasil exige o conhecimento da distribuição horizontal e vertical do conteúdo de ferro no solo. Por quê? Porque, em alguns locais, os óxidos de ferro podem perfazer até 80% da massa do solo!

Os óxidos de ferro são bem conhecidos por sua forte interação com a matéria orgânica do solo, influenciando assim a quantidade de carbono que o solo consegue armazenar de forma estável. Os óxidos de ferro também são bem conhecidos pela sua forte afinidade com íons fosfato, determinando assim a disponibilidade de fósforo para as plantas. A intrincada relação entre fósforo e matéria orgânica reforça a necessidade do conhecimento da variação tridimensional do teor de ferro do solo em todo o território brasileiro. Tal conhecimento poderia ajudar, por exemplo, na melhoria dos sistemas de classificação do solo, no desenvolvimento de sistemas avançados de recomendação que assegurem o uso mais eficiente de fertilizantes fosfatados, e na construção de políticas públicas de uso e ocupação do solo que respeitem as reais capacidades desse precioso e não-renovável bem natural.

Foi pensando na necessidade de impulsionar o avanço do conhecimento para o uso sustentável do solo brasileiro que decidimos desenvolver um projeto inovador. O objetivo: a construção colaborativa de um repositório de dados de ferro do solo com cobertura nacional que fosse público, gratuito e, sobretudo, fácil de usar e manter. Lançado em dezembro de 2016, o Repositório Brasileiro de Dados de Ferro do Solo (Fe-BR) já conta com a impressionante soma de 232 conjuntos de dados oriundos de todas as vinte e sete unidades federativas do Brasil. Muitos dos conjuntos de dados já foram processados e estão aqui, à sua inteira disposição, esperando para serem usados. Contribua você também!

1.1 Organização dos Dados

Os dados são organizados em tabelas bidimensionais usando planilhas eletrônicas padronizadas construídas a partir de experiências de iniciativas tanto locais como globais. São quatro as tabelas utilizadas, com nomes:

1. *dataset*, para os dados sobre o conjunto de dados,
2. *observacao*, para os dados das observações do solo,
3. *camada*, para os dados das camadas e horizontes, e
4. *metadado*, para os dados sobre os dados.

Uma descrição detalhada do propósito e conteúdo específico de cada uma dessas tabelas é apresentado nas sessões que seguem.

Conjunto de dados

A maximização da disseminação e uso de um conjunto de dados, o apropriado reconhecimento de seus autores ou autoras e instituição responsável, bem como a identificação que alterações e ajustes feitos após a publicação da sua primeira versão, dependem da provisão de dados mínimos sobre aquele conjunto de dados. No **febr** esses dados são inseridos na tabela denominada *dataset*, palavra que na língua inglesa significa ‘conjunto de dados’.

Por conter dados essenciais sobre um conjunto de dados, a tabela *dataset* é sempre a primeira a ser revisada e processada. Quaisquer dados faltantes são solicitados aos autores/responsáveis pelo conjunto de dados. Da mesma forma, sempre que dados incoerentes são identificados, os autores/responsáveis pelo conjunto de dados são consultados iterativamente até que quaisquer dúvidas sejam completamente sanadas. O devido preenchimento da tabela *dataset* é essencial para o posterior processamento dos dados e metadados de um conjunto de dados.

A tabela *dataset* possui cerca de vinte itens estruturados em uma sequência de linhas (Tabela 2.1). Os dados sobre o conjunto de dados são inseridos na segunda coluna, imediatamente ao lado da sua respectiva identificação e definição contidas na primeira coluna. A primeira linha é dedicada à identificação propriamente dita do conjunto de dados no **febr**, a segunda ao título do conjunto de dados, e assim por diante, até as últimas linhas, dedicadas aos termos usados para a descrição de características chave e indexação do conjunto de dados. Cada um dos itens da tabela *dataset* são descritos em detalhe nas próximas seções.

Tabela 2.1: Representação da estrutura da tabela *dataset*.

| item | dado |
|--------------------------------|------|
| <code>dataset_id</code> | <??? |
| <code>dataset_titulo</code> | <??? |
| <code>dataset_descricao</code> | <??? |
| ... | <??? |
| <code>palavras_chave</code> | <??? |
| <code>categoria_vcge</code> | <??? |

2.1 Identificação

Um conjunto de dados no **febr** é identificado por dois itens principais (Tabela 2.2). O primeiro deles, `dataset_id`, é o código identificador único do conjunto de dados, o que significa que não há qualquer `dataset_id` repetido no **febr**.

Tabela 2.2: Itens utilizados para a identificação do conjunto de dados.

| item | dado |
|-----------------------------|--|
| <code>dataset_id</code> | código de identificação do conjunto de dados |
| <code>dataset_titulo</code> | título do conjunto de dados |

Em geral, o código de identificação de um novo conjunto de dados inserido no **febr** é definido pela sua posição (ordem das contribuições, `ctb`) em relação aos demais conjuntos de dados presentes no **febr**. Por exemplo, o código de identificação do primeiro conjunto de dados inserido no **febr** é `dataset_id = ctb0001`, enquanto que do trigésimo terceiro conjunto de dados é `dataset_id = ctb0033`.

No caso dos conjuntos de dados obtidos diretamente do SISB, o código de identificação utilizado no **febr** é o mesmo usado naquele sistema, independente da sua ordem de entrada no **febr**. Por exemplo, `dataset_id = ctb0770`, onde 770 é o código de identificação usado no SISB. Essa estratégia possibilita que, no futuro, e em havendo interesse por parte dos mantenedores do SISB, seja estabelecida comunicação direta do SISB com o **febr**. Assim seria possível atualizar os dados no SISB à medida que ajustes e correções nos dados forem feitos no **febr**.

O segundo identificador de um conjunto de dados é o seu título – `dataset_titulo` –, geralmente relacionado ao título do projeto, tese, dissertação, etc que o gerou. Em geral, sugere-se que seja usado um título específico para o conjunto de dados, o que identifica melhor o seu conteúdo, ao invés do título do trabalho ou projeto que o gerou. Isso é importante para aumentar as chances de o conjunto de dados ser identificado em ferramentas de busca e assim potencializar o seu reuso.

Quanto à formatação do título, procura-se usar a mesma língua da fonte e caracteres em caixa baixa, com exceção da primeira letra do título e dos nomes próprios. Por exemplo, *Conjunto de dados da tese ‘Matéria orgânica e características físicas, químicas, mineralógicas e espectrais de Latossolos de diferentes ambientes’*.

2.2 Descrição

A descrição do conjunto de dados, contendo dados básicos para maximizar o seu reuso futuro sem que seja necessário entrar em contato com os seus autores e /ou responsáveis é incluída no item `dataset_descricao`. A descrição do conjunto de dados deve abordar aspectos como:

1. Os motivos para a realização do estudo que levou à geração dos dados,
2. Um resumo dos dados incluídos no conjunto de dados e dos métodos analíticos usados, e
3. Uma descrição do delineamento amostral e a forma de coleta das amostras.

A descrição do delineamento amostral, ou seja, a estratégia utilizada para seleção dos locais de observação e amostragem do solo, fornece os elementos necessários para determinar se – e como – os diferentes conjuntos de dados podem – ou devem – ser utilizados em uma determinada aplicação.

Quando for o caso, a descrição do conjunto de dados deve incluir também:

1. Os motivos para a presença de camadas/horizontes com dados faltantes para uma ou mais variáveis do solo, e

2. Uma descrição das alterações/modificações realizadas quando da publicação de uma nova versão.

Uma descrição exemplar de um conjunto de dados é a que segue:

“Conjunto de dados produzido como parte da Dissertação submetida como requisito parcial para obtenção do grau de Mestre no Curso de Pós-Graduação em Agronomia da Universidade Federal Rural do Rio de Janeiro. Inclui dados do conteúdo de ferro total determinado via extração com solução de ácido sulfúrico para 20 perfis do solo observados no município de Pinheiral, Rio de Janeiro. A seleção dos locais de observação foi feita com base no conhecimento pedogenético da área de estudo, a partir de informações prévias sobre o meio físico e solo, tendo sido selecionados pontos representativos que contemplassem a variação dos fatores de formação do solo. Em cada local foram abertas trincheiras para descrição dos perfis e coleta de amostras dos horizontes para caracterização do solo segundo procedimentos descritos no Manual de Descrição e Coleta de Solo no Campo (Santos et al., 2013). Devido à existência de limitações orçamentárias, apenas amostras de horizontes selecionados foram submetidas à determinação do conteúdo de ferro total. Em alguns casos, apenas os horizontes B diagnósticos foram analisados, em outros o primeiro horizonte A e um ou mais horizontes subsuperficiais.”

2.3 Versionamento

O controle dos passos do processamento de um conjunto de dados é realizado usando dois itens: `dataset_versao` e `publicacao_data` (Tabela 2.3). A versão do conjunto de dados é um dado necessário para identificar os casos em que o conjunto de dados foi modificado por uma razão técnica e/ou científica. Por exemplo, um conjunto de dados contendo inúmeras observações do solo cuja localização não foi georreferenciada (`dataset_versao` = 1.0). Uma vez no **febr**, esse conjunto de dados é modificado consideravelmente de maneira que todas as observações do solo possuam coordenadas espaciais aproximadas, com precisão de cerca de 1000 m (`dataset_versao` = 2.0). Em seguida, modificações menores são feitas, pois verifica-se que, para algumas observações com maior volume de informações disponível, as coordenadas espaciais mais prováveis podem ser inferidas com precisão de cerca de 100 m (`dataset_versao` = 2.1).

Sempre que uma nova versão de um conjunto de dados é preparada, uma descrição sumária das modificações é incluída na descrição do conjunto de dados em `dataset_descricao`. Os responsáveis pelas modificações são então identificados conforme descrito abaixo na Seção 2.7.

Tabela 2.3: Itens utilizados para o versionamento do conjunto de dados.

| item | dado |
|------------------------------|---|
| <code>dataset_versao</code> | versão atual do conjunto de dados |
| <code>publicacao_data</code> | data de publicação da versão atual do conjunto de dados |

A data de publicação do conjunto de dados, `publicacao_data`, não é necessariamente a mesma data de publicação do trabalho que o gerou ou o utilizou pela primeira vez, mas sim a data em que o conjunto de dados foi tornado efetivamente público e disponível para acesso. Quando modificações são feitas no conjunto de dados durante ou após a sua inserção no **febr**, resultando assim em uma nova versão – conforme descrito acima –, então a data de publicação é a data em que essa nova versão do conjunto de dados foi disponibilizada no **febr**.

A data de publicação é apresentada na sequência dia, mês e ano, usando formato numérico `dd-mm-aaaa` ou `dd/mm/aaaa`, ou seja, dois dígitos para o dia e mês, e quatro dígitos para o ano. Por exemplo, `13-06-2001` identifica o dia 13 do mês de junho do ano de 2001. Sempre que possível insere-se os dados de ambos dia, mês e ano. Contudo, dentre os três, o mais importante é aquele relativo ao ano. Quando algum dos dados

não é conhecido, usa-se o símbolo **xx** em substituição, por exemplo, **xx-xx-2001**. Caso a data de publicação da versão atual conjunto de dados seja completamente desconhecida, usa-se a data de publicação do trabalho que o gerou ou utilizou pela primeira vez.

2.4 Licença de uso

A licença de uso do conjunto de dados – especificada usando o item `dataset_licenca` – define como o conjunto de dados pode ser usado por outras pessoas a partir do momento da sua publicação no **febr**. O **febr** aceita apenas conjuntos de dados com licenças de uso não-restritivas, ou seja, que permitem a redistribuição, modificação, adaptação e criação a partir do conjunto de dados submetido, e que não obrigam o licenciamento dos trabalhos derivados sob os mesmos termos. Contudo, o autor/responsável pelo conjunto de dados tem reservado o direito de decidir se os produtos derivados podem ou não ter fins comerciais, desde que previsto pelo órgão financiador. As licenças Creative Commons (<https://br.creativecommons.org/licencas/>) estão entre as mais apropriadas para os conjuntos de dados inseridos no **febr**, segundo as quais todos os novos trabalhos derivados de um conjunto de dados devem atribuir o devido crédito ao autor/responsável pelo conjunto de dados original.

A legislação brasileira ainda não especifica qual deve ser a licença de uso dos conjuntos de dados gerados em projetos conduzidos via financiamento público. Contudo, segundo a INDA, existe consenso entre os diferentes órgãos da administração pública de que uma licença de uso apropriada para esse tipo de conjuntos de dados é a licença Creative Commons CC-BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>). A licença CC-BY 4.0 permite que um conjunto de dados seja distribuído, remixado, adaptado e usado para criar outros produtos, mesmo que para fins comerciais, desde que seja atribuído o devido crédito aos autores/responsáveis pelo conjunto de dados original. Assim, trata-se de uma das licenças mais flexíveis dentre as licenças Creative Commons disponíveis, permitindo a maximização da disseminação e uso dos conjuntos de dados. Assim sendo, a licença CC-BY 4.0 é a licença adotada como padrão no **febr**. Maiores informações sobre as licenças de uso sugeridas para conjuntos de dados públicos podem ser encontradas no fórum de discussão mantido pela INDA no endereço <https://groups.google.com/forum/#!forum/inda-br>.

Para os conjuntos de dados gerados em projetos conduzidos via financiamento privado, onde as partes envolvidas têm interesse em compartilhar os dados desde que não para fins comerciais, sugere-se usar a licença Creative Commons CC-BY-NC 4.0 (<https://creativecommons.org/licenses/by-nc/4.0/>). A licença CC-BY-NC 4.0 possui praticamente os mesmos termos da licença CC-BY 4.0, exceto pelo fato de que a distribuição, remixação, adaptação e derivação de outros produtos não podem ser usados para fins comerciais.

As pessoas interessadas em utilizar os conjuntos de dados disponíveis no **febr** devem prestar atenção, não só à licença de uso especificada no item `dataset_licenca`, como também às determinações da política de uso de dados da organização autora/responsável pelo conjunto de dados. Um exemplo disso são os conjuntos de dados originalmente descarregados do SISB, mantido pela Embrapa, para alimentar o **febr**. Segundo a Embrapa, o emprego dos dados disponíveis em seu SISB para “*finalidades profissionais e/ou comerciais deve ser precedido de estabelecimento de licença de uso específica por parte da Embrapa*”, política essa que se estende à versão do conjunto de dados disponível no **febr**.

2.5 Organização

A organização ou instituição responsável pela geração do conjunto de dados é identificada usando uma série de itens codificados usando o padrão `organizacao_<...>`, onde `<...>` representa um descritor da organização responsável (Tabela 2.4). Fundamentalmente, esses itens armazenam dados básicos sobre os endereços físico e eletrônico da organização.

Tabela 2.4: Itens utilizados para a identificação da organização responsável pelo conjunto de dados.

| item | dado |
|--|---|
| <code>organizacao_nome</code> | nome da organização responsável |
| <code>organizacao_url</code> | endereço da organização responsável na Internet |
| <code>organizacao_pais_id</code> | país da organização responsável |
| <code>organizacao_municipio_id</code> | cidade da organização responsável |
| <code>organizacao_codigo_postal</code> | código postal da organização responsável |
| <code>organizacao_rua_nome</code> | rua da organização responsável |
| <code>organizacao_rua_numero</code> | número da organização responsável em sua rua |

Quando o conjunto de dados é provido por organização outra que não aquela responsável pela sua geração, pode-se identificar a organização que atualmente detém responsabilidade técnica sobre o conjunto de dados, ou seja, a organização provedora do conjunto de dados. Isso é especialmente importante caso modificações consideráveis no conjunto de dados tenham sido feitas por essa organização. O mesmo pode ser aplicado aos conjuntos de dados gerados por duas ou mais organizações. No caso de conjuntos de dados gerados ou sob responsabilidade de organizações como a Embrapa, que possuem várias unidades, identifica-se a unidade da organização responsável pelo conjunto de dados, nunca apenas a organização. Em todo caso, é fundamental garantir que ambos endereços físico e eletrônico sejam válidos e atuais.

2.6 Autoria

Os autores e/ou responsáveis pela geração do conjunto de dados são identificados usando os itens `autor_nome` e `autor_email` (Tabela 2.5). Quando o acesso aos autores é impossível ou limitado, identifica-se as pessoas que atualmente detém responsabilidade técnica sobre o conjunto de dados. No caso de trabalhos acadêmicos, como monografias, dissertações e teses, identifica-se tanto o autor como o orientador principal do trabalho acadêmico.

Tabela 2.5: Itens utilizados para a identificação dos autores do conjunto de dados.

| item | dado |
|--------------------------|--------------------------------------|
| <code>autor_nome</code> | nome completo dos autores |
| <code>autor_email</code> | endereço de e-mail atual dos autores |

Sempre que um conjunto de dados tiver dois ou mais autores, insere-se os respectivos dados de identificação separados por ponto e vírgula. Assim como no caso da identificação da organização responsável, é importante garantir que o endereço de e-mail dos autores do conjunto de dados seja válido e atual. Isso é fundamental para permitir o contato com os autores sempre que surgirem dúvidas sobre o conjunto de dados. O conhecimento do nome completo dos autores também é fundamental para permitir que os devidos créditos lhes sejam atribuídos sempre que o conjunto de dados for distribuído, remixado, adaptado ou usado para criar outros produtos (Seção 2.4).

2.7 Contribuições

Os autores de modificações no conjunto de dados são identificados usando três itens similares àqueles usados para identificar os autores ou responsáveis pelo conjunto de dados (Tabela 2.6). No **febr**, modificações constituem contribuições à melhoria do conjunto de dados, sendo sempre mencionadas na descrição do conjunto de dados (Seção 2.2), e geralmente resultando em uma nova versão do do conjunto de dados (Seção 2.3). Como a inserção de um conjunto de dados no **febr** geralmente requer algum tipo de modificação, por menor que seja, os integrantes da equipe do **febr** costuma figurar como autores de contribuições à sua melhoria.

Tabela 2.6: Itens utilizados para a identificação dos autores de contribuições para o conjunto de dados.

| item | dado |
|---------------------------------------|---|
| <code>contribuidor_nome</code> | nome dos autores das contribuições |
| <code>contribuidor_email</code> | endereço de e-mail atual dos autores das contribuições |
| <code>contribuidor_organizacao</code> | organização à qual os autores das contribuições estão afiliados |

A apropriada identificação dos autores das modificações no conjunto de dados é fundamental para permitir o contato com os mesmos sempre que surgirem dúvidas sobre as novas versões do conjunto de dados. Além disso, permite atribuir o devido crédito quando se deseja referenciar uma versão específica do conjunto de dados. Contudo, recomenda-se sempre dar crédito aos autores originais do conjunto de dados, independentemente da sua versão (Seção 2.6).

2.8 Referências

Uma referência permanente e válida a documento ou artigo científico onde a versão original e/ou a versão atual do conjunto de dados foi usado, preferencialmente, pela primeira vez é informada usando o item `dataset_referencia_i`. Sempre que possível usa-se um Digital Object Identifier (DOI, <https://www.doi.org/>), mas uma URL costuma ser a mais usada, especialmente para documentos antigos e trabalhos acadêmicos.

Tantas referências quantas forem julgadas pertinentes podem ser inseridas, numeradas sequencialmente em ordem de importância, usando-se para isso o índice `i` do item `dataset_referencia_i`. No caso de os responsáveis atuais pelo conjunto de dados não serem os seus autores originais, é imprescindível que a referência principal, `dataset_referencia_1`, seja a um trabalho dos autores do conjunto de dados. Já a referência secundária, `dataset_referencia_2`, pode ser a um trabalho dos responsáveis atuais pelo conjunto de dados. No caso de trabalhos acadêmicos, como monografias, dissertações e teses, costuma-se inserir uma referência ao trabalho acadêmico e ao artigo onde o conjunto de dados foi utilizado pela primeira vez. No caso de o conjunto de dados ter passado por modificações consideráveis, uma referência ao trabalho onde a nova versão tenha sido usada pela primeira vez também deve ser inserida. Para adicionar novas referências, basta inserir, na sequência, novas linhas na tabela `dataset`, usando o índice `i` para indicar sua ordem.

2.9 Indexação

Três itens são usados para indexação dos conjuntos de dados e sua otimização para ferramentas de busca (Tabela 2.7). O primeiro deles é a área de especialidade da Agronomia-Ciência do Solo à qual o conjunto de

dados está relacionado. São seis as áreas de especialidade definidas pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES):

- Gênese, Morfologia e Classificação dos Solos
- Física do Solo
- Química do Solo
- Microbiologia e Bioquímica do Solo
- Fertilidade do Solo e Adubação
- Manejo e Conservação do Solo

Tabela 2.7: Itens utilizados para a indexação do conjunto de dados.

| item | dado |
|-------------------|---|
| area_conhecimento | área da especialidade segundo CAPES e CNPq |
| palavras_chave | termos descritores do conjunto de dados |
| categoria_vcge | classificação dentre as atividades de órgãos públicos |

O segundo item consiste de uma lista de termos que descrevem aspectos importantes do conjunto de dados, preferencialmente diferentes daqueles constantes no título, sendo separados por ponto e vírgula. Uma lista de termos bem elaborada aumenta o potencial de descoberta do conjunto de dados por mecanismos de busca.

O terceiro item é a Categoria do Vocabulário Controlado do Governo Eletrônico (VCGE). O VCGE é um esquema padronizado de assuntos e categorias usado para facilitar a apresentação e identificação dos serviços disponibilizados em uma estrutura de diretórios online. O objetivo do VCGE é ajudar os cidadãos a encontrar informações em catálogos de dados públicos sem a necessidade de conhecer a organização responsável pelo assunto ou categoria.

Observações no espaço e no tempo

Tabela *observacao*. Nessa tabela são inseridas as informações espaciais mais fundamentais sobre as observações do solo.

3.1 Identificação

Identificador exclusivo da observação usado no conjunto de dados de origem (*observacao_id*). Como se trata de um código, não devem ser usados espaços ou caracteres especiais. Espaços são substituídos por um *underscore*, `_`, ou traço, `-`. Caracteres especiais devem ser substituídos pelo caractere correspondente simplificado, ou seja, `á` torna-se `a`, `ç` torna-se `c`, e assim por diante. Exemplo: *observacao_id* = Perfil-01.

3.2 Coordenada temporal

Data da observação (*observacao_date*) no formato dd-mm-aaaa. A data de observação é um dos atributos que, juntamente com as coordenadas espaciais, definem uma observação em um conjunto de dados. Isso significa que a observação repetida, ao longo do tempo, de um mesmo local no espaço, caracteriza a constituição de uma nova observação, portanto, a definição de uma nova *observacao_id*. O valor padrão, para os casos em que a data de observação é desconhecida, é *observacao_date* = xx-xx-xxxx. Quando somente o dia e/ou o mês de observação são desconhecidos, usa-se o primeiro valor da sequência, ou seja, *dd* = 01 e *mm* = 01.

3.3 Coordenadas espaciais

A localização espacial de uma observação do solo é especificada usando coordenadas espaciais, também conhecidas como coordenadas geográficas. Para isso são usados os cinco itens descritos na Tabela 3.1.

Tabela 3.1: Itens usados para especificar a localização espacial das observações do solo.

| item | dado |
|----------------------|--------------------------------------|
| <i>coord_sistema</i> | sistema de referência de coordenadas |

| item | dado |
|----------------|------------------------------------|
| coord_x | coordenada x ou longitude |
| coord_y | coordenada y ou latitude |
| coord_precisao | precisão das coordenadas espaciais |
| coord_fonte | fonte das coordenadas espaciais |

O primeiro item, `coord_sistema`, refere-se ao sistema de referência de coordenadas (SRC) utilizado para o georreferenciamento das observações do solo. A especificação do SRC é fundamental para possibilitar o uso apropriado de dados espaciais em sistemas de informação geográfica (SIG), especialmente para fins de correlação e/ou cruzamento com outros dados espaciais. Tradicionalmente, a especificação do SRC é feita de diversas maneiras, mas a forma mais usual é aquela no formato de texto legível por humanos. Por exemplo, WGS84 / UTM zona 23S, que identifica o datum geodético WGS84 e a zona 23 do hemisfério sul na projeção cartográfica UTM. Esse tipo de identificação não é usado do **febr** pois, não sendo padronizado, dificulta a automatização da manipulação computacional dos conjuntos de dados. Assim, o **febr** adota os códigos padronizados e aceitos internacionalmente do EPSG, os quais são legíveis por humanos e por máquinas. Os códigos dos sistemas de referência de coordenadas geográficas e projetadas mais usados no Brasil são mostrados nas Tabelas 3.2 e 3.3.

Tabela 3.2: Códigos EPSG dos sistemas de referência de coordenadas geográficas para o Brasil.

| Código | Definição |
|-----------|----------------|
| EPSG:4326 | WGS84 |
| EPSG:4225 | Córrego Alegre |
| EPSG:4618 | SAD69 |
| EPSG:4674 | SIRGAS2000 |

Tabela 3.3: Códigos EPSG dos sistemas de referência de coordenadas projetadas para o Brasil.

| Zona e hemisfério | Córrego Alegre | SAD69 | WGS84 | SIRGAS2000 |
|-------------------|----------------|-------|-------|------------|
| 18N | - | 29168 | 32618 | 31972 |
| 18S | - | 29188 | 32718 | 31978 |
| 19N | - | 29169 | 32619 | 31973 |
| 19S | - | 29189 | 32719 | 31979 |
| 20N | - | 29170 | 32620 | 31974 |
| 20S | - | 29190 | 32720 | 31980 |
| 21S | 22521 | 29191 | 32721 | 31981 |
| 22S | 22522 | 29192 | 32722 | 31982 |
| 23S | 22523 | 29193 | 32723 | 31983 |
| 24S | 22524 | 29194 | 32724 | 31984 |
| 25S | 22525 | 29195 | 32725 | 31985 |

Os valores das duas coordenadas espaciais, x e y, são especificadas usando os itens `coord_x` e `coord_y`, respectivamente. Aqui, x é o mesmo que longitude e y é o mesmo que latitude. A diferença fundamental entre as duas nomenclaturas é que, enquanto longitude/latitude refere-se a coordenadas geográficas, x/y refere-se a coordenadas projetadas. No caso das coordenadas geográficas, usa-se o formato decimal, sendo a unidade de medida o grau (°). Já para as coordenadas projetadas, a unidade de medida é o metro (m). Em ambos os casos, utiliza-se a vírgula como separador decimal, sendo esse o padrão da língua portuguesa.

Os dois

Precisão (`coord_precisao`) com que as coordenadas espaciais foram determinadas (em metros).

`coord_fonte`

Tabela 3.4: Fontes de coordenadas espaciais.

| Código | Definição |
|--------|--------------------------------|
| GPS | aparelho GPS |
| MAPA | mapa analógico ou digital |
| WEB | serviço web como o Google Maps |

3.4 Localização

Identificação do município (`municipio_id`), estado (`estado_id`) – ou unidade federativa – e país (`pais_id`) onde a observação foi realizada. Como o **febr** trata apenas de conjuntos de dados produzidos no Brasil, usa-se o código ISO 3166-1 alpha-2 do Brasil, ou seja, **BR**. No caso do estado – ou unidade federativa –, usa-se a abreviação da respectiva unidade federativa (UF). O nome do município é escrito por extenso, conforme encontrado na fonte.

3.5 Amostragem

Tipo de amostragem (`amostra_tipo`). Opções: SIMPLES ou COMPOSTA.

Número de amostras (`amostra_quanti`). Um (1) quando o valor da variável anterior é SIMPLES ou mais para COMPOSTA. Expressar o valor com números inteiros.

Área amostral (m^2) (`amostra_area`). Por exemplo, a área de uma trincheira costuma ser de 1 m^2 .

Amostras de camadas e horizontes

Tabela *camada*. Nesta tabela são inseridos os dados das camadas amostradas e onde o conteúdo de ferro tenha sido determinado.

4.1 Identificação

observacao_id. Identificador exclusivo da observação usado no conjunto de dados de origem. Refere-se à tabela *observacao*.

camada_numero Número da camada ou horizonte na observação, atribuído consecutivamente de cima para baixo. Há alguma confusão com essa variável, por vezes sendo atribuído valor numérico sequencial para todo o conjunto de dados. É preciso ficar claro que a numeração refere-se à ordem das camadas dentro do perfil à que pertencem, ou seja, a primeira camada amostrada de um perfil receberá, sempre, o número inteiro 1. O valor padrão, para o caso do perfil observado possuir apenas uma camada, é **camada_number = 1**.

camada_nome Nome da camada ou horizonte. Exemplo: A, B, C etc. Conjuntos de dados que são produto de trabalhos edafológicos raramente terão informações sobre o nome da camada amostrada. Caso essa informação esteja presente, deve-se verificar com o responsável pelo conjunto de dados se a informação está correta ou foi criada apenas para preencher a coluna da planilha eletrônica. O valor padrão, para o caso do perfil observado ser oriundo de estudo edafológico, é **camada_nome = NA**.

amostra_codigo Código laboratorial da amostra. Usado para identificar as repetições de laboratório. Alguns grupos de pesquisa destinam suas amostras para análise em laboratórios especializados, os quais identificam as amostras com códigos específicos, que chamamos de **sample_code**. Em outros casos, as análises são feitas pelo laboratório do próprio grupo de pesquisa, o qual não necessariamente atribui um código de identificação único às amostras. Mas esses laboratórios podem realizar as análises com duas ou mais repetições, sendo então o número da repetição de uma determinada amostra identificado como **sample_code**. O valor padrão, para o caso em que não há código de identificação único e/ou a amostra não foi analisada com repetições, é **amostra_codigo = 1**.

4.2 Coordenadas verticais

Profundidade do limite superior, **profund_sup**, e inferior, **profund_inf**, da camada ou horizonte observados.

4.3 Variáveis

`fe_<extração>_<determinação>`

O valor padrão, para o caso em que o conteúdo de ferro não está disponível, é `fe_<extração>_<determinação> = NA`.

Dados sobre os dados (metadados)

Tabela *metadado*

Bibliografia
