



Em direção a uma Ciência do Solo mais aberta

blog.scielo.org/blog/2021/01/08/em-direcao-a-uma-ciencia-do-solo-mais-aberta/

SciELO

January 8, 2021 15:30



Por George Brown, Wilian Demetrio e Alessandro Samuel-Rosa

Os que trabalham com ciências ambientais como a Ciência do Solo sabem que esse *métier* envolve a avaliação de múltiplas variáveis ambientais descritivas e explicativas. O intuito é compreender os complexos fenômenos que ocorrem na natureza, seja trabalhando no campo ou em laboratório, ou ainda com seus derivados, como a micro, meso ou microbiota. Assim, nossas pesquisas geralmente geram enormes planilhas de dados, com resultados de diversas variáveis, muitas delas medidas ao longo do tempo. Estas são frequentemente analisadas usando métodos estatísticos complexos, a fim de encontrar e explorar as relações entre as variáveis e sua significância. Dentre esses métodos, os multivariados se tornaram indispensáveis por serem muito úteis para revelar essas relações de maneira visualmente compreensível.



Imagem: Markus Spiske.

Contudo, uma limitação da quase totalidade dos trabalhos de pesquisa em Ciência do Solo realizados no Brasil é a ausência de acesso aos dados primários usados nas análises estatísticas. Isso impede o reuso dos dados, seja para reproduzir um estudo e verificar suas conclusões ou para realizar novas pesquisas para responder questões que ficaram sem resposta.

Muitas vezes, os dados são apresentados apenas de maneira agregada na forma de tabelas com estatísticas descritivas das variáveis estudadas. Uma das utilidades dos dados agregados é a realização de estudos-síntese baseados em ferramentas meta-analíticas, os quais vêm ganhando espaço na Ciência do Solo. Esses estudos identificam tendências e padrões de comportamento de dados produzidos em muitos trabalhos sobre determinado assunto, especialmente aqueles que apresentam alta variabilidade nos resultados. Porém, com o uso crescente dos métodos estatísticos multivariados, até mesmo essa versão agregada dos dados tem sido omitida nas publicações. Nesses casos, a preferência de autores e editores costuma ser pela apresentação dos dados na forma de figuras e tabelas resultantes das análises multivariadas. Esse tipo de prática prejudica, inclusive, os estudos de meta-análise, o que faz desses trabalhos de pesquisa um ponto final na vida útil dos dados.

Como cientistas, nos preocupamos com a sustentabilidade no avanço do conhecimento. Adotando uma política com foco na disponibilização dos dados primários usados em nossos estudos, aumentaremos a sustentabilidade e utilidade futura de nossas pesquisas e a colaboração entre pesquisadores.

Políticas de dados abertos vêm sendo adotadas por muitos periódicos científicos do exterior, que cada vez mais exigem que dados primários sejam, no mínimo, disponibilizados como material suplementar às publicações, ou que os autores informem onde foram depositados, quer seja em repositórios de dados ou de metadados. Contudo, a prática mais eficiente e apropriada consiste em depositar os dados das pesquisas em repositórios públicos confiáveis especializados para essa finalidade que, além da preservação digital, exigem estruturação e descrição adequada dos metadados, atribuem DOI (*Digital Object Identifier*, identificador de objeto digital) e, portanto, dão maior sustentabilidade e visibilidade aos dados, visto que estes estão indexados em bases de busca *on-line*. O ideal é a organização e disponibilização dos dados de acordo com os princípios FAIR, *Findable, Accessible, Interoperable and Resusable* (encontrável, acessível, interoperável e reutilizável), que se aplicam especialmente ao acesso computadorizado.

A maior parte dos trabalhos de pesquisa em Ciência do Solo que desenvolvemos no Brasil é financiada ou por órgãos públicos ou com recursos públicos, outorgados por órgãos de fomento à pesquisa ou pós-graduação como o CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) e a Capes (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), e as fundações estaduais de amparo à pesquisa (FAPs). Além disso, os salários dos funcionários e empregados públicos, que geram a maior parte dos resultados da ciência no nosso país, também são financiados com recursos públicos. Portanto, precisamos compreender que a disponibilização aberta dos dados produzidos em nossas pesquisas, juntamente com as publicações que elas geraram, é uma questão de sustentabilidade da ciência e responsabilidade para com a sociedade em geral. Logicamente, com a devida

atenção às políticas institucionais de proteção de conhecimento sensível e disponibilização dos dados. Lembrando sempre da regra “tão aberto quanto possível e tão fechado quanto necessário”.

A disponibilização aberta dos dados já é exigida por órgãos internacionais de fomento como a União Europeia e os *Research Councils* do Reino Unido, que reforçam a exigência da submissão dos dados como material suplementar por periódicos científicos, conforme comentado anteriormente.

Em um futuro próximo, nossos órgãos de pesquisa também exigirão isso. O CNPq, em parceria com o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), já está desenvolvendo o repositório *Lattes Data* para armazenar dados dos projetos fomentados pelo órgão. No caso da FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo), algumas chamadas já exigem a submissão de um plano de gestão de dados acompanhando a proposta de pesquisa. O programa SciELO também está desenvolvendo um repositório, o SciELO Data, para armazenar os dados vinculados aos manuscritos submetidos aos periódicos da rede SciELO.

Entretanto, seria mais recomendável se, como sociedade científica e como cientistas, adotássemos uma política de dados abertos orientada às especificidades e prioridades de cada área, em vez de esperar uma exigência dos órgãos de fomento ou dos periódicos científicos. Assim, nos adiantamos ao processo e mostramos nossa boa-vontade em relação à sociedade brasileira que, com seus impostos, vem financiando nossos empregos, alunos e pesquisas há décadas.

Num mundo cada vez mais globalizado, no qual grandes sínteses de dados muitas vezes causam bastante alarde, adotar uma ciência mais aberta também aumentará a possibilidade de reuso, aproveitamento e reanálise dos dados de nossas pesquisas e, portanto, sua sustentabilidade, visibilidade e impacto.

O Boletim 45(1) da Sociedade Brasileira de Ciência do Solo¹ apresentou ampla discussão sobre ciência aberta e os principais repositórios públicos nacionais e internacionais de dados abertos do solo. Contudo, desde então, parece não ter havido progresso na abertura dos dados das pesquisas em Ciência do Solo no Brasil. Os principais periódicos da área ainda não exigem que os dados sejam submetidos como material suplementar ou que eles estejam disponíveis em repositórios públicos. Além disso, mestrandos, doutorandos e pós-doutorandos dos principais programas de pós-graduação e instituições de pesquisa da área ainda não precisam elaborar planos de gestão aberta dos dados produzidos.

O desenvolvimento de modelos de organização e armazenamento de dados que facilitam a disponibilização, transferência e reuso com o mínimo esforço humano, vem ocorrendo apenas de maneira pontual e pouco articulada. É fundamental retomarmos essa discussão, visando aumentar a abertura da Ciência do Solo brasileira. Acreditamos que a maneira mais

eficiente de fomentar essa discussão, de modo estruturado e coordenado, seja pela formação de um grupo de trabalho junto à Sociedade Brasileira de Ciência do Solo. Um grupo de trabalho que congregue pesquisadores das diferentes áreas da Ciência do Solo visando viabilizar soluções sustentáveis para a questão dos dados produzidos em nossas pesquisas.

A ciência vem sendo cada vez mais demandada e, ao mesmo tempo, criticada e escrutinada pela sociedade e pelos governos, que buscam soluções para os problemas atuais, ou, às vezes, justificativas para suas ações. Ao adotarmos uma política de ciência aberta, com disponibilização aberta dos dados primários e dos resultados de nossas pesquisas, demonstramos que não temos nada a esconder, e que os dados estão livres para serem usados por quem quiser. Obviamente, isso envolve certos riscos, pois os dados podem também ser usados livremente por grupos de interesse com agendas próprias.

Como cientistas, cuidamos dos dados, e fazemos uso criterioso deles, em nossas publicações. Portanto, é necessário construir, como sociedade científica, concomitantemente com os possíveis usuários de nossos dados, diretrizes de uso responsável que envolveriam, por exemplo, citação das fontes e adoção de critérios para a contextualização, interpretação e uso dos dados abertos. Essas diretrizes poderiam ser disponibilizadas junto com os arquivos dos dados, em repositórios de dados confiáveis. Claramente, existem muitos benefícios mas também possíveis armadilhas a serem evitadas no processo de abertura dos dados de nossas pesquisas, e esses temas merecem maior discussão, visando a melhor compreensão do processo de abertura científica ao qual estamos nos expondo e ao qual precisamos nos preparar.

Nota

1. Os Desafios da Open Science para a Ciência do Solo. *Boletim Informativo da Sociedade Brasileira de Ciência do Solo* [online]. 2019, vol. 45, no.1, ISSN: 1981-979X [viewed 8 January 2020]. Available from: <https://www.sbcs.org.br/wp-content/uploads/2019/06/Boletim-SBCS-Volume-45-N%C3%BAmero-1.pdf>

Referências

CNPq e IBICT lançam Lattes Data [online]. Governo Federal – Governo do Brasil. 2020 [viewed 8 January 2020]. Available from: <https://www.gov.br/cgu/pt-br/governo-aberto/noticias/2020/2/cnpq-e-ibict-lancam-lattes-data>

Lançamento do repositório SciELO Data [online]. SciELO.org. 2020 [viewed 8 January 2020]. Available from: <https://mailchi.mp/scielo/scielo-data-pt>

Os Desafios da Open Science para a Ciência do Solo. *Boletim Informativo da Sociedade Brasileira de Ciência do Solo* [online]. 2019, vol. 45, no.1, ISSN: 1981-979X [viewed 8 January 2020]. Available from: <https://www.sbcs.org.br/wp-content/uploads/2019/06/Boletim-SBCS-Volume-45-N%C3%BAmero-1.pdf>

WILKINSON, M.D., et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* [online]. 2016, vol. 3, no 1, 160018 [viewed 8 January 2020].
<http://doi.org/10.1038/sdata.2016.18>. Available from:
<https://www.nature.com/articles/sdata201618>

Links externos

Gestão de dados – FAPESP: <https://fapesp.br/gestaodedados>

Repositório Brasileiro Livre para Dados Abertos do Solo: <https://www.pedometria.org/febr/>

SciELO Data: <https://data.scielo.org/>

Sociedade Brasileira de Ciência do Solo (SBCS): <https://www.sbc.org.br/>

Sobre George Brown

George Brown é pesquisador na Embrapa Florestas e professor da pós-graduação em Ciência do Solo na Universidade Federal do Paraná. É formado em Ciências Naturais na Universidade de Wisconsin, e com doutorado em Ecologia na Universidade de Paris. Atua na ecologia do solo, usando os animais como bioindicadores da qualidade ambiental, e é um dos poucos especialistas em minhocas no Brasil. Editou vários livros e publicou mais de 100 trabalhos em periódicos científicos, interagindo com centenas de profissionais no Brasil e no exterior.



Sobre Wilian Demetrio

Wilian Demetrio é pós-doutorando no Instituto Nacional de Pesquisas Especiais (Inpe). É formado em Agronomia pela Uniguauçu, com mestrado e doutorado em Ciência do Solo pela Universidade Federal do Paraná. Sua área de estudo e interesse é ecologia do solo, com o uso de invertebrados, com ênfase em minhocas, como indicadores da qualidade do solo.



Sobre Alessandro Samuel-Rosa

Alessandro Samuel-Rosa é professor de classificação e mapeamento do solo, estatística aplicada e probabilidade, e geoprocessamento e sensoriamento remoto na Universidade Tecnológica Federal do Paraná. Seus interesses de pesquisa incluem otimização de configurações amostrais, modelos (geo)estatísticos e aprendizado de máquina, e padronização e harmonização de dados. Ele é o líder do Laboratório de Pedometria, que desenvolve os pacotes *pedometrics*, *spsann* e *febr* para R, e mantém o Repositório Brasileiro Livre para Dados Abertos do Solo.

