

Agricultura 4.0: uma abordagem a partir dos princípios FAIR e tecnologia blockchain para gestão de dados de levantamentos pedológicos

Élton Carneiro Marinho¹, Annatercia Gomes Pinheiro¹, Alessandra Castro Fiorini Bessa², Sérgio Manuel Serra da Cruz^{1,2}, Eber Assis Schmitz¹

¹ Federal University of Rio de Janeiro, Rio de Janeiro RJ, Brasil
elton.marinho@ppgi.ufrj.br
annatercia@ufrj.br
serra@ppgi.ufrj.br
eber@nce.ufrj.br

² Federal Rural University of Rio de Janeiro, Rio de Janeiro RJ, Brasil
alefessa@ufrj.br

Resumo. A vasta quantidade de dados que é gerada diariamente, em todas as áreas do conhecimento, necessita de formas inovadoras de curadoria e extração de conhecimento para que seja usada de modo inteligente, mostrando de onde são extraídas as descobertas com habilidade para transformar a realidade do agronegócio. Atualmente, a agricultura 4.0 está sendo transformada pela *big data*, em especial, dados de qualidade e metadados, são usados pelos agricultores, pesquisadores e formuladores de políticas para auxiliar os processos de decisão. Neste trabalho foi feita uma análise de como os princípios FAIR, os metadados de proveniência e a tecnologia *blockchain* podem auxiliar na proteção de dados. Apresentamos uma proposta de melhoria da plataforma *OpenSoils* que já conta com um aporte para a proveniência de dados e um grande potencial para se tornar um modelo seguro de base para coleta de dados de solo. Este estudo mostra que a adoção da tecnologia *blockchain* nesta plataforma pode contribuir para tornar, tanto estudos pedológicos quanto suas aplicações mais confiáveis, seguras e dinâmicas.

Palavras-chave: e-Science, FAIR, Proveniência, Blockchain, Agricultura 4.0

1 Introdução

Na agricultura, o solo é a base de todas as atividades agrícolas e, provavelmente, o recurso mais crítico, gerando benefícios ambientais, de saúde e socioeconômicos [12]. A quantidade finita de recursos hídricos, de solo e a forma como são utilizados impactam, de certa forma, numa mudança climática que pode ser desastrosa para a vida na Terra como a conhecemos. Logo, compreender os desafios que entremeiam essa temática é de interesse não somente acadêmico, mas das empresas e nações [10].

Atualmente, os modelos agrícolas e sistemas de informação são datacêntricos, tornaram-se ferramentas importantes para avaliar, potencializar e prever o comportamento de diversas variáveis de interesse do agronegócio que afetam diretamente os setores público e privado. No entanto, apesar das recentes pesquisas e das melhorias dos modelos agrícolas, muitos dos atuais modelos são descendentes diretos de investimentos em pesquisa feitos há décadas, e muitos dos principais avanços nas TIC e na ciência de dados da última década ainda não foram totalmente incorporados nas atividades de campo [1].

Além dessa questão, dentro da seara de levantamento de dados pedológicos, a quantidade de *datasets* relacionados aos solos é ainda desconhecida, muitas observações e coletas de dados no campo não são compartilhadas entre pesquisadores, dificultando as tarefas de classificação de solos. A classificação correta dos solos é extremamente importante para, por exemplo, identificar a aptidão e a fertilidade de uma região, entre outras necessidades práticas. Não obstante, é necessário destacar o que os estudos pedológicos manipulam grandes volumes de dados e que, como outras áreas das Ciências, vivenciam uma crise na gestão de dados científicos conforme apontam muitos trabalhos com relação a qualidade e a reprodutibilidade dos dados [4].

A partir de uma grande quantidade de dados bem anotados é possível extrair conhecimentos com habilidade para transformar a realidade social e as economias locais desde que utilizados de modo inteligente. Atualmente, a agricultura 4.0 está sendo transformada por grandes volumes de dados, em especial, usados para ajudar agricultores, pesquisadores e formuladores de políticas a tomar decisões mais inteligentes e informadas sobre os usos dos solos [19].

Pode-se observar que agricultores e tomadores de decisão, de lugares variados do mundo, utilizam uma diversidade de dados pedológicos para decidir como e quando fertilizar, plantar ou colher; pesquisadores reúnem dados de diferentes bases, coletados de inúmeras formas, em seu formato bruto ou manipulado, para extrair conhecimentos cruciais para condução de suas pesquisas; formuladores de políticas precisam de dados baseados em evidências para realizar investimentos e outros grupos podem usá-los para tornar os serviços mais eficientes dentro da cadeia do agronegócio [19].

A convergência da base tecnológica, pela adoção da forma digital na geração e manipulação de conteúdo, a evolução na informática, que propicia processamento mais rápido a custos cada vez menores, e a evolução dos meios de comunicação, que tem permitido a expansão da Internet, são três fatores que favorecem a agricultura digital [30].

Uma das características essenciais na *Open Science* é a reprodutibilidade e reutilização dos dados [34], e isto faz com que a contribuição de uma Plataforma como o *Opensoils* seja de extrema importância no contexto da Agro 4.0 visando à geração de conhecimento, a mobilidade e o aumento da oferta de aplicativos para dispositivos móveis [promovendo] o agronegócio, incluindo também os agricultores familiares, [para que] possam usufruir dos benefícios desta oferta de tecnologia e conhecimento em suas propriedades, propiciando competitividade e melhoria de renda, além do aumento da oferta de alimentos para o Brasil [20].

Este trabalho tem como objetivo avaliar novas tecnologias e propor a atualização tecnológica da plataforma OpenSoils [11][8], que já possui um aporte para a proveniência de dados e um potencial enorme para se tornar um modelo seguro de base de dados para apoiar projetos de coleta de dados de solos no Brasil, como por exemplo o PRONASOLOS [15]. O PRONASOLOS é constituído um grupo de trabalho formado por pesquisadores de várias unidades da Embrapa, do IBGE, da SBCS, da CPRM, da UFRRJ, da UFPI, da UDESC, da UFLA e do Ministério da Agricultura, que formularam um documento que serviu de base para a criação do denominado Programa Nacional de Solos do Brasil (PronaSolos), que deverá fornecer, num prazo de 30 anos, informações adequadas para orientação do uso da terra em todo o Brasil.

Dentro deste contexto, e a partir de revisões bibliográficas sobre os temas de Agricultura Digital, Proveniência de dados digitais, princípios FAIR e da tecnologia Blockchain, com especial atenção na interseção entre eles, apresentamos uma proposta de extensão da plataforma *OpenSoils*, o *Opensoils 2.0*, que trará, além da proveniência dos dados digitais, um arcabouço que contempla os princípios FAIR e esboça uma proposta inicial para elaboração de um Plano de Gestão de Dados. Acredita-se que, no futuro, esta plataforma contribua tornando pesquisas, tanto as voltadas para dados pedológicos quanto as voltadas para aplicações no campo, mais confiáveis, seguras e dinâmicas.

2 FAIR, Proveniência de dados digitais e Blockchain

Devido aos crescentes obstáculos relativos à utilização e reutilização dos crescentes volumes de dados científicos, torna-se necessária a criação de uma infraestrutura global para suportar a abertura dos dados voltados para a publicação, o compartilhamento e a reutilização de dados científicos [14].

Tendo em vista a urgente necessidade de interoperabilidade e a reutilização de dados e metadados, representantes da academia, da indústria, das agências de financiamento e editores uniram-se para projetar um conjunto de regras. O resultado desse esforço foi o FAIR *Guiding Principles* apresentado por Wilkinson et al. [31]. A sigla FAIR é um acrônimo para *Findable* (Localizável), *Accessible* (Acessível), *Interoperable* (Interoperável) e *Reusable* (Reutilizável). São princípios orientadores de alto nível. Os dados científicos e seus metadados devem ser fáceis de serem localizados e compartilhados, tanto para humanos quanto por computadores (*Findable*). Após localizados, os dados precisam ser acessados, possivelmente incluindo autenticação e autorização (*Accessible*). Os dados precisam se integrar a outros dados, precisam interoperar com aplicativos ou fluxos de trabalho para análise, armazenamento e processamento (*Interoperable*). Tudo isso para que os dados possam ser reutilizados. Os metadados e os dados devem ser bem descritos, para que possam ser replicados e/ou combinados em diferentes configurações (*Reusable*).

No contexto da gestão dos dados científicos, os metadados são de extrema importância. Metadados são, de acordo com a NISO (*National Information Standard Organization*), a informação que descreve, explica, localiza, ou possibilita que um recurso informacional seja fácil de recuperar, usar, auditar e gerenciar [25].

Segundo Sayão [27] os metadados podem ser divididos em três categorias conceituais: descritivos¹, estruturais² e administrativos³.

Dessa forma, é fácil perceber como os metadados podem ajudar na gestão dos dados científicos e porque precisam estar bem representados para que possam ser reutilizados pelo próprio pesquisador ou por terceiros. Algumas iniciativas voltadas para Agricultura 4.0, já incluem ferramentas para cadastro e visualização de metadados [9].

O processo de tornar os dados aderentes aos princípios FAIR⁴ é um dos pré-requisitos para gestão de dados científicos [31]. Adicionalmente, ressaltamos que este processo ainda é pouco explorado em dados pedológicos e agrícolas em geral. Porém, considerando que desejamos tornar a Plataforma OS aberta, seguindo os moldes da *Open Science*, deve-se pensar em compartilhamento de dados e metadados - uma característica de repositórios públicos, visando a integrações, visibilidade, aceleração no processo de novas descobertas, democratização do conhecimento, economia de recursos, transparência e progresso, para reuso e reprodutibilidade no domínio agrícola [22].

Os princípios FAIR suportam um conjunto de melhores práticas para compartilhamento dos dados, respeitando restrições éticas, legais ou contratuais; por exemplo, o cumprimento de regulamentos e acessos não autorizados. Adotar estes princípios significa permitir que outros pesquisadores encontrem dados e *datasets* que são disponibilizados dentro de parâmetros para que eles possam integrar-se com outros dados, sendo utilizados por humanos e máquinas e reusados para novas pesquisas [21].

De acordo com as recomendações do *Guidelines on FAIR Data Management in Horizon* [16], a Comissão Europeia, desde o início de 2017, passou a exigir um PGD dos candidatos a financiamentos; garantindo, assim, maior transparência nos seus investimentos, melhor controle na gestão dos dados gerados, facilitando, quando possível, o seu compartilhamento e reuso.

No Brasil, o entendimento sobre esse cenário ainda está em estágio inicial, com a elaboração de propostas de modelos de PGD de pesquisa, baseado nos princípios FAIR. A Embrapa Solos, inclusive, reconheceu essa necessidade há alguns anos e o tema foi debatido em junho de 2018, na Embrapa Solos (Rio de Janeiro-RJ), no workshop 'Gestão de dados de pesquisa: princípios, desafios e oportunidades' [30].

A proveniência é um tipo de metadado que pode ser compreendido como o histórico digital de um dado [7]. É amplamente usada em diversas áreas do conhecimento, em especial na Ciência da Computação, como forma de assegurar o rastreo e originalidade dos dados [29].

¹ Descrevem um recurso com o propósito de descoberta e identificação;

² Documentam como os recursos são compostos por diversos elementos.

³ Fornecem informações que apoiam os processos de gestão do ciclo de vida dos recursos informacionais.

⁴ Este processo também é conhecido como FAIRificação

O metadado de proveniência facilita tanto o rastreamento de um item de dado ou *dataset* quanto a identificação dos processos produtores de dados. Nesse contexto ela também pode assegurar a qualidade dos dados, onde um conjunto de metadados específicos descrevem como o dado se originou, quando, por quem e como foi produzido. Uma boa procedência dos dados pode aumentar a confiança do consumidor de dados [2].

Assim como os bens físicos podem sofrer transformações desde o produtor até o consumidor, o mesmo ocorre com os dados e *datasets*, a qualidade dos dados derivados depende da confiança nos dados e processos que o originaram. Segundo Allemang e Teegarden [2], a confiança nos dados começa com o conhecimento de sua origem. Para reduzir as incertezas, toda informação deve possuir este “certificado de origem”, ou seja, ter sua proveniência registrada. O fato de existir proveniência dos dados permite que haja um conhecimento da cadeia produtiva, identificação de gargalos, localização de um desajuste entre outras informações que podem ser obtidas. A transparência dessas informações gera uma segurança maior do consumidor de dados em relação às suas escolhas [24].

Segundo o Consórcio W3C, a proveniência pode ser descrita em termos de entidades, atividades e pessoas envolvidas na produção de um dado, podendo ser usada para formar avaliações sobre sua qualidade ou confiabilidade. A proveniência é a informação sobre entidades, atividades e pessoas envolvidas na produção de um dado ou coisa, que pode ser usada para formar avaliações sobre sua qualidade, confiabilidade [35]. A família de documentos PROV define um modelo, serializações correspondentes e outras definições de suporte para permitir o intercâmbio de informações de proveniência em ambientes heterogêneos, como a Web. O padrão PROV permite representar e trocar informações de proveniência usando formatos amplamente disponíveis, como RDF e XML. Além disso, fornece definições para acessar informações de proveniência, validá-las e mapear no padrão de metadados Dublin Core [36]. Dublin Core é um esquema de metadados que visa descrever objetos digitais.

A agricultura 4.0, promove muitas melhorias no campo que um impacto direto na cadeia produtiva e no próprio consumidor final [23]. Neste contexto, os agricultores e tomadores de decisão, principais atores no processo, tem um papel primordial, o processo de tomada de decisão apoiado em sistemas datacêtricos podem ajudar a definir os melhores usos do solo para atender suas demandas. Um produtor consciente e mais exigente, pode direcionar a melhor forma de utilização do solo, para culturas mais saudáveis e uma agricultura mais sustentável.

Conhecer e organizar os dados pedológicos ajuda com a melhor perspectiva de sua utilização. Ao sumarizar as necessidades deste contexto, torna-se possível gerar novos artefatos que agreguem valor a esta cadeia produtiva, justificando a necessidade de proveniência, de rastreabilidade e de transparência de dados.

Na web, com a facilidade que se tem de copiar e transformar dados, está se tornando cada vez mais difícil determinar as origens de um determinado dado [7]. Seguindo os parâmetros de segurança e privacidade, considerando que mecanismos de controle são criados para permissão de acesso de usuários (níveis) enquanto a privacidade controla quais dados se pode acessar. “Você não pode ter privacidade sem segurança, mas pode

ter segurança sem privacidade.” [6]. Devido a crescente importância contida nos dados atualmente, estes devem ser protegidos sob uma tecnologia que minimize ou mesmo impeça qualquer tipo de alteração indevida. Para isso, além dos princípios FAIR e da proveniência de dados, é necessário um protocolo de confiança, uma tecnologia que forneça a proteção necessária para os *datasets*.

A tecnologia blockchain é aplicada para *datasets* descentralizados e auto reguláveis. Através dela os dados podem ser gerenciados e organizados de uma maneira transformadora: aberta, permanente, verificada e compartilhada, sem a necessidade de uma autoridade central [13]. Formada por bancos de dados compartilhados, distribuídos e tolerantes a falhas e sem nenhuma autoridade certificadora onde todos os participantes da rede podem compartilhar. Nenhuma entidade controla ou intermedia a *blockchain* [28].

Esta tecnologia pode ser compreendida como um livro razão cuja manutenção é feita pela cooperação e interação dos nós em uma rede. Neste livro são guardadas todas as transações ocorridas. Estas transações não podem ser alteradas ou excluídas, garantindo um armazenamento imutável dos dados [3].

A unidade básica de dados de uma rede *blockchain* é o bloco, sendo este uma estrutura de dados responsável por armazenar informações sobre um conjunto de transações [3]. Uma transação é uma unidade de informação dentro de um bloco e pode representar qualquer coisa: dinheiro, ativos financeiros, músicas, propriedades, dados em geral etc.

Cada bloco possui uma identificação única e contém a identificação do bloco anterior, perfazendo assim uma cadeia de blocos (“*Block Chain*”). O primeiro bloco, também chamado de bloco gênese não possui identificação para o bloco anterior. Outra propriedade de um bloco é o *timestamp*, esta propriedade torna mais difícil para um usuário de má-fé manipular a rede *blockchain*. O *timestamp* guarda informações sobre data e hora de criação do bloco. Estes conceitos permitem que as transações possam ser rastreadas de maneira histórica [3].

Cada bloco pode ser dividido em duas partes, o cabeçalho e o corpo do bloco. O cabeçalho possui informações de controle enquanto o corpo do bloco possui as informações sobre as transações [33].

Os dados a serem incluídos nas transações podem perfeitamente atender às metas descritas nos princípios FAIR para reprodutibilidade. Os metadados de proveniência descritos podem ser definidos para esta tecnologia, assim como um protocolo aberto, permitindo autenticação e autorização quando necessário.

As informações existentes em uma *blockchain* podem ser públicas e estar disponíveis para o escrutínio de qualquer usuário ou tomador de decisão que tenha interesse, lembrando que estes registros são protegidos por potentes criptografias [5].

3 Trabalhos relacionados

No setor agrícola e na agricultura 4.0, os cientistas de dados continuam se surpreendendo com o fato de que quantidades maciças de dados podem ser aliados na solução de problemas supostamente intratáveis em Pedologia. Nos últimos anos, projetos de dados abertos e compartilhados em larga escala e algumas pesquisas forneceram, nesse sentido, exemplos positivos de como planejar e conduzir iniciativas globais de compartilhamento de dados.

O sucesso na execução da primeira venda de grãos usando blockchain serviu de inspiração para mostrar o uso potencial desta tecnologia na cadeia de abastecimento agrícola. Nas iniciativas de blockchain podem ser usadas para resolver problemas práticos da vida real, entre elas a de apoio ao pequeno agricultor [18].

Allemand e Teegarden [2] elaboraram um documento que descreve as prioridades para a criação de um ecossistema de dados global eficaz para a agricultura, desde o engajamento das partes interessadas até o fornecimento, o compartilhamento e a colaboração com os dados. Este documento foi encomendado pela Syngenta com a assistência da GODAN⁵ para catalisar o consenso sobre os desafios e princípios que devem ser abordados na construção de um ecossistema global de dados para a agricultura; uma experiência baseada na iniciativa Open PHACTS⁶, cujo objetivo era disponibilizar dados científicos que podem beneficiar a indústria farmacêutica em geral.

Além da iniciativa da GODAN para a criação de um ecossistema de dados global, outras iniciativas que valem ser citadas são: O CABI e o ODI. O CABI é uma organização internacional, intergovernamental e sem fins lucrativos que fornece informações e aplica conhecimentos científicos para resolver problemas na agricultura e no meio ambiente. O ODI (*Open Data Institute*), que trabalha com o CABI e o GODAN para incentivar boas práticas de gerenciamento de dados agrícolas. O ODI busca explorar, aprimorar o acesso e o compartilhamento de dados nos programas agrícolas em diversas regiões do planeta e tornar as pesquisas financiadas abertamente acessíveis. Para tanto, nesta iniciativa, procura-se entender os desafios e tornar os dados FAIR.

Dentro desse contexto, para se entender como a comunidade científica brasileira está aderindo a esses princípios, a iniciativa GO FAIR Brasil-Saúde, como a primeira rede de implementação ativa em operação e que atua em todos os domínios da saúde, serve de modelo para outras áreas como agricultura, por exemplo. Esta iniciativa demonstra o grande interesse em implementar os princípios do FAIR em suas práticas de gerenciamento de dados de pesquisa. [26]

⁵ GODAN (Global Open Data): uma iniciativa que busca apoiar os esforços globais para tornar os dados agrícolas e nutricionalmente relevantes disponíveis, acessíveis e utilizáveis para uso irrestrito em todo o mundo.

⁶ Projeto patrocinado, cinco anos, pela *Innovative Medicines Initiative* na *European Commission*,

Segundo Sales [26] é necessário considerar a inclusão do Brasil nos fóruns internacionais de tratamento padronizado de dados de pesquisa, começando por áreas estratégicas, como a agricultura e energia que contribui para moldar políticas de comportamento, financiamento e pesquisa no país.

Tendo em vista as três principais dimensões práticas da e-Science, isto é, o compartilhamento de recursos computacionais, o acesso distribuído a grandes conjuntos de dados e o uso de plataformas digitais para colaboração e comunicação [32], fica evidente também o interesse das comunidades brasileiras na implementação dos princípios do FAIR. Demonstrando, dessa forma, uma preocupação vigente em prol de uma boa gestão de dados de futuras pesquisas.

4 Opensoils 2.0

A plataforma *OpenSoils* foi concebida no intuito de contribuir com as políticas brasileiras de proteção e mapeamento de solos, projetando e estabelecendo as bases para um esforço de longo prazo. O OpenSoils é baseado nos fundamentos de *open science* e *e-science* para a área de segurança de solos, sendo uma infraestrutura aberta, elástica, multiusuário, multicamada e orientada para armazenar dados primários, secundários de solos e sua proveniência. Acreditamos que esta plataforma irá posicionar o Brasil como um dos principais atores mundiais no que tange a pesquisa e inovação nesta área [8].

O *OpenSoils* é uma plataforma distribuída que não se limita ao universo acadêmico-científico, dedicada para a execução de ciência aberta na área de Segurança de Solos do Brasil. Dentre as funcionalidades que apresenta, a coleta *online* de dados heterogêneos de solos diretamente no campo com conectividade com a base de dados e a plataforma na nuvem traz uma solução *Web* para pesquisar e compartilhar dados curados abertos com semântica explícita, análise de grandes volumes de dados e geração de mapas de solo.

Considerando características essenciais destes princípios na plataforma *OpenSoils*, citamos o aprimoramento da capacidade das máquinas de encontrar e usar automaticamente os dados pedológicos, além de apoiar sua reutilização pelos pesquisadores. A utilização e aplicação destas diretrizes, favorece o compartilhamento de informações em plataformas com acesso aberto; repositórios integrados; *datasets* e quaisquer outros conjuntos de dados e/ou informação que se queira FAIRificar [21].

A partir da avaliação dos temas descritos anteriormente, FAIR, Proveniência de dados digitais e da tecnologia blockchain, propomos uma extensão do *Opensoils*, o *OpenSoils 2.0*.

Dentro desse cenário de gestão de dados científicos, tem sido observada a crescente necessidade de se trabalhar com dados capturados de forma metódica, que demandam uma gestão segura para que possam ser reutilizados em pesquisas futuras [27]. Assim, surgiu a proposta da elaboração de um Plano de Gestão de Dados (PGD) para a plataforma *OpenSoils 2.0*, trata-se um documento formal e dinâmico, contendo a

descrição detalhada dos dados de pesquisa, contemplando todo o ciclo de vida de pesquisa em Ciência de Solos.

No caso da plataforma *OpenSoils 2.0*, sobre os objetivos de tornar dados FAIR, pode-se dizer que é necessário adicionar descritores de dados, identificadores e *links* de longa duração [31]. Para torná-los acessíveis, é preciso definir quem poderá acessar e como será este acesso. Caso os dados pedológicos não possam ser totalmente abertos, é necessário permitir o acesso através de um repositório com características de acesso limitado. Para haver interoperabilidade, pode-se usar padrões, metadados e formatos comuns de legibilidade. E, para que sejam reusados, a elaboração de um PGD ajude na compreensão dos dados pedológicos e permissões apropriadas que determinem como os dados e metadados poderão ser reusados em novas pesquisas ao longo tempo [16].

A facilidade com que se pode copiar e transformar dados na Web atualmente tornou cada vez mais difícil determinar as origens de um determinado dado [7]. Os dados devem ser protegidos sob uma tecnologia que minimize ou mesmo impeça qualquer tipo de alteração indevida. Para isso, além dos princípios FAIR e da proveniência de dados, é necessário um protocolo de confiança, uma tecnologia que forneça a proteção necessária para os *datasets* da plataforma *OpenSoils 2.0*.

Por não possuir uma entidade centralizadora é mais resistente a ataques maliciosos. Os dados gravados em suas transações são duráveis, confiáveis, imutáveis e fornecem transparência ao sistema. Estes fatores mostram que a *blockchain* é uma tecnologia confiável e segura que pode ser incorporada a plataforma *OpenSoils 2.0*, assegurando proteção adicional aos seus dados, o que alinha esta tecnologia com os princípios FAIR [17] e os metadados de proveniência.

5 Considerações Finais

Este trabalho demonstrou, através de uma revisão bibliográfica, que existe a viabilidade de aplicação dos temas propostos à plataforma *Opensoils*. A implementação destes conceitos deu origem à extensão desta plataforma, a que demos o nome de *OpenSoils 2.0*. Foi demonstrada a potencialidade destes conceitos, tecnologia *blockchain* alinhada com a aplicação dos princípios FAIR voltada para o domínio das Ciências dos Solos, destacando a relevância da sua incorporação na extensão da plataforma *OpenSoils*.

Essa abordagem oferece excelentes recursos para facilitar a gestão de grandes volumes de dados pedológicos, mitigando as dificuldades que envolvem a reprodutibilidade das pesquisas dessa área e estimulando o reuso e compartilhamento seguro de *datasets*. Também foi destacada a necessidade da aplicação de princípios FAIR e da adoção de tecnologias seguras relacionadas à gestão de dados científicos. No entanto, reconhecemos que ainda são pouco estudadas na literatura da área de Ciência de Solos.

Um modelo de Plano de Gestão de dados também é proposto, pois através da elaboração, manutenção e atualização deste documento pode-se acompanhar o ciclo de

vida dos dados. Neste PGD estão contemplados os princípios que padronizam a localização, acessibilidade, interoperabilidade e reutilização dos dados.

A aderência dos dados de solo aos princípios FAIR potencializa a possibilidade de que esses dados sejam confiáveis para reuso tanto no ambiente acadêmico quanto no ambiente empresarial.

Os conceitos descritos agregam valor aos estudos pedológicos já que os princípios FAIR organizam as informações de forma a serem amplamente reproduzíveis; a tecnologia *blockchain* é aderente a estes princípios; há previsão da preservação da propriedade (autoria) e o controle sobre seus dados; há acesso instantâneo ao banco de dados completo, já que cada um pode ser um nó na *blockchain*; há garantia de que os dados anexados à *blockchain* sejam imutáveis e invioláveis; há participação ativa nas decisões de governança por parte dos produtores de dados.

Dentro do contexto da Pedologia ainda há muito que precisa ser feito. Uma limitação deste trabalho é que ainda é necessário ampliar e detalhar melhor as informações que permeiam o universo de classificação de solos para que essa proposta seja bem contemplada numa adaptação futura à plataforma. A elaboração de um PGD que acompanhe o ciclo de vida desses dados dentro das limitações éticas e de acordo com as atualizações vindas da literatura que envolvem esse domínio de solos é o próximo passo a ser tomado.

Do ponto de vista de expandir o uso da proposta para uma aplicação, verificamos que o que foi apresentado no trabalho é perfeitamente incorporável ao arcabouço ferramental e ao propósito da plataforma *OpenSoils*. Desta forma, a implementação dessa proposta se tornaria uma contribuição a essa infraestrutura eletrônica, no contexto do aumento da confiabilidade e da proteção dos dados, além de ampliar sua abordagem para uma conjuntura de dados compartilhados, fornecendo conjuntos de dados pedológicos de maior qualidade e maior confiabilidade nos processos decorrentes da Agricultura Digital.

Agradecimentos: Os autores agradecem à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de mestrado.

Referências

1. Agropensa: Summary for Policymakers. In: Intergovernmental Panel on Climate Change (ed.) Climate Change 2013 - The Physical Science Basis. pp. 1–30 Cambridge University Press, Cambridge (2014). <https://doi.org/10.1017/CBO9781107415324.004>.
2. Allemang, D., Bobbin, T.: A Global Data Ecosystem for Agriculture and Food. 24, 2, 23 (2016).
3. Alves, P.H. et al.: Desmistificando Blockchain: Conceitos e Aplicações. *Comput. e Soc.* August, 1–24 (2018).
4. Baker, M.: 1,500 scientists lift the lid on reproducibility. *Nature*. 533, 7604,

- 452–454 (2016). <https://doi.org/10.1038/533452a>.
5. Bolfe, É.L. et al.: Futuro da Agricultura Brasileira. Embrapa. 212 (2018).
 6. Brian Dean: Privacy vs. Cybersecurity | Secureworks, <https://www.secureworks.com/blog/privacy-vs-security>, last accessed 2020/08/19.
 7. Buneman, P. et al.: Data Provenance: Some Basic Issues. Lect. Notes Comput. Sci. Found. Softw. Technol. Theor. Comput. Sci. 1974, 87–93 (2000). https://doi.org/10.1007/3-540-44450-5_6.
 8. Ceddia, M., Cruz, P.: OpenSoils : e-Science em Segurança de Solos. In: Tical. (2018).
 9. Corrêa, B. et al.: Uma ferramenta para cadastro e visualização de metadados em Agricultura de Precisão utilizando GeoNetwork. May, 39–41 (2011).
 10. Cruz, S. et al.: Desenvolvendo Sistemas Agrícolas de Próxima Geração: Um Estudo em Ciência de Solos. 135–144 (2020). <https://doi.org/10.5753/wcama.2019.6428>.
 11. Cruz, S.M.S. da et al.: OpenSoils : Uma Plataforma de Apoio à Ciência do Solo. November, (2019).
 12. Da Cruz, S.M.S. et al.: Towards an e-infrastructure for Open Science in Soils Security. 2, Figure 1, (2020). <https://doi.org/10.5753/bresci.2018.3273>.
 13. Delft University of Technology et al.: Tecnologia Blockchain: uma visão geral. 01 Introdução. Harv. Bus. Rev. 6, 2, 1–4 (2017). <https://doi.org/10.6084/m9.figshare.5607778.v1>.
 14. Dutch Techcentre for Life Sciences: Jointly designing a Data FAIRPORT - Dutch Techcentre for Life Sciences, <https://www.dtls.nl/2014/01/20/jointly-designing-data-fairport/>, last accessed 2020/05/20.
 15. EMBRAPA: Estratégias pronasolos parcerias. Doc. Solos. 183, 93 (2016).
 16. European Commission: Guidelines on Fair Data Management in Horizon 2020. December, 6 (2016).
 17. Jacobsen, A. et al.: FAIR Principles: Interpretations and Implementation Considerations. Data Intell. 10–29 (2019). https://doi.org/10.1162/dint_r_00024.
 18. Kamilaris, A. et al.: The rise of blockchain technology in agriculture. ICT Updat. (2018).
 19. L'Hénaff, P., Smith, F.: Creating FAIR and open agricultural data ecosystems – The ODI, <https://theodi.org/article/creating-fair-and-open-agricultural-data-ecosystems>, last accessed 2020/05/20.
 20. Maria, S. et al.: AGRO 4 . 0 – RUMO À AGRICULTURA DIGITAL. Presented at the .
 21. Martínez-Lavanchy, P.M., Hüser, F.J., Buss, M.C.H., Andersen, J.J., Begtrup, J.W.: (8) Research Data Management (RDM) - FAIR Principles - YouTube, <https://www.youtube.com/watch?v=OvEHYCSmzCA>, last accessed 2020/05/20.
 22. Medeiros, C.B. et al.: Academia Brasileira de Ciências - O MUNDO A PARTIR DO CORONAVÍRUS, Ed. 20 | DESVENDANDO A OPEN SCIENCE: BENEFÍCIOS, GARGALOS E DESAFIOS CORPO | Facebook, <https://www.facebook.com/abciencias/videos/2783241185239956/>, last accessed 2020/08/19.

23. Prodemge: Agricultura 4.0, www.prodemge.gov.br, (2018).
24. Ribeiro, J.G. et al.: Agricultura 4.0: Desafios À Produção De Alimentos E Inovações Tecnológicas. Simpósio Eng. Produção - Sienpro. 1–7 (2018).
25. Riley, J. (NISO): Understanding Metadata - What Is Metadata? (2017).
26. Sales, L. et al.: GO FAIR Brazil: A Challenge for Brazilian Data Science. *Data Intell.* 238–245 (2019). https://doi.org/10.1162/dint_a_00046.
27. Sayão, L.F.: Uma outra face dos metadados: informações para a gestão da preservação digital. *Encontros Bibli Rev. Eletrônica Bibliotecon. e Ciência da Informação.* 1–31 (2010). <https://doi.org/10.5007/1518-2924.2010v15n30p1>.
28. Shetty, S. et al.: Data provenance assurance in the cloud using blockchain. May 2017, (2020). <https://doi.org/10.1117/12.2266994>.
29. Silva, F.C. da: Tratamento e preenchimento de falhas de séries de dados meteorológicos utilizando workflows científicos paralelos em ambientes de GPU. (2014).
30. Takahashi, T.: Livro verde-sociedade da informação no Brasil. (2000).
31. Wilkinson, M.D. et al.: Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data.* 3, 1–9 (2016). <https://doi.org/10.1038/sdata.2016.18>.
32. Wouters, P.: What is the matter with e-science: thinking aloud about informatisation in knowledge creation. *The Pantaneto Forum.* July 2006, January 2006, (2006).
33. Zheng, Z. et al.: An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends. *Proc. - 2017 IEEE 6th Int. Congr. Big Data, BigData Congr. 2017.* October, 557–564 (2017). <https://doi.org/10.1109/BigDataCongress.2017.85>.
34. 2.4 - Investigação Reprodutível e Análise de Dados - Manual de Formação em Ciência Aberta, https://foster.gitbook.io/manual-de-formacao-em-ciencia-aberta/02introducaoacienciaaberta/04investigacao_reprodutivel_e_analise_de_dados, last accessed 2020/08/19.
35. PROV-Overview, <https://www.w3.org/TR/prov-overview/>, last accessed 2020/06/20.
36. World Wide Web Consortium (W3C), <https://www.w3.org/>, last accessed 2020/03/02.