

## Investigando a adoção de Princípios FAIR e Proveniência de Dados na Agricultura Digital sob perspectiva das Humanidades Digitais: Um estudo de caso na plataforma OpenSoils

### *Investigating the adoption of FAIR Principles and Data Provenance in Digital Agriculture from the perspective of Digital Humanities: the OpenSoils platform case of study*

Alessandra Castro Fiorini Bessa<sup>1</sup>

Sergio Manuel Serra da Cruz<sup>2</sup>

#### **Resumo**

*Cientistas da informação e profissionais das áreas agrárias atuam em ambientes cada vez mais digitalizados e enfrentam o crescente desafio no gerenciamento de enorme volume de dados em projetos de ciência aberta. Neste trabalho avaliamos como um dos setores da Agricultura Digital poderá se beneficiar de práticas emergentes de gestão de dados comuns às Humanidades Digitais. Verificamos a necessidade de apoiar a reutilização de datasets da área de solos através dos princípios FAIR e da incorporação de metadados de proveniência.*

**Palavras-chave:** *Agricultura Digital. Princípios FAIR. metadados de proveniência.*

**Abstract.** *Information scientists and professionals in the agricultural sciences operate in increasingly digital environments and face the growing challenge of managing huge volumes of data in open science projects. In this paper we evaluate how Digital Agriculture can benefit from emerging data management practices common to the Digital Humanities. We verified the need to support the reuse of agricultural datasets through the FAIR principles and the incorporation of provenance metadata.*

**Keywords:** *Digital Agriculture. FAIR Principles. provenance metadata.*

## **1 Introdução**

Cientistas da informação e profissionais das áreas agrárias atuam em ambientes cada vez mais digitalizados e enfrentam o crescente desafio no gerenciamento de enorme volume de dados em projetos de ciência aberta (ALBAGLI *et al*, 2014). Isso demanda

---

<sup>1</sup> Universidade Federal Rural do Rio de Janeiro – UFRRJ/PPGIHD – [alebessa@ufrj.br](mailto:alebessa@ufrj.br)

<sup>2</sup> Universidade Federal Rural do Rio de Janeiro – UFRRJ/PET-SI – [serra@pet-si.ufrj.br](mailto:serra@pet-si.ufrj.br)

novos olhares que considerem abordagens que ainda não são amplamente internalizadas nas instituições, empreendimentos agrícolas e laboratórios de pesquisas. Pode-se dizer que é “evidente a inclinação de que a noção geral sobre as Humanidades Digitais possui vertentes implícitas na direção dos estudos informacionais tão característicos da Ciência da Informação.” (CASTRO; PIMENTA, 2018), e segundo Russell (2011), um dos grandes objetivos das HD é "criar bases de dados com recursos digitais relevantes para as Humanidades. Isto inclui a captura, estruturação, documentação, preservação e disseminação dos dados."

Apesar do crescente interesse sobre os diferentes recursos de digitalização da agricultura mundial (big data, internet of things, realidade aumentada, robótica, sensores, impressão 3D, integração de sistemas, inteligência artificial, digital twins, blockchain entre outros) e seus impactos na cadeia produtiva do agronegócio ainda existem poucos estudos correlacionando- os com as Humanidades Digitais (HD). Neste trabalho, investigamos as possíveis interseções, em ambientes digitais, entre a área de HD e a área de segurança de solos, investigando como tais áreas poderão se beneficiar dos princípios FAIR (WILKINSON et al, 2016) e metadados de proveniência (BUNEMAN et al., 2000).

No Brasil, as HD e o Agronegócio têm interesses sociais, ambientais e econômicos múltiplos, apresentam diferentes graus de maturidade e poucos estudos que busquem compreender as interfaces em comum. Percebemos que transformação digital tanto nas humanidades quanto no agronegócio não é mais uma opção, é uma necessidade imprescindível para a ser trilhada pelos diversos atores da sociedade brasileira. No cenário internacional, começam a surgir iniciativas tais como o RDA<sup>3</sup>, CODATA<sup>4</sup> e o CGIAR<sup>5</sup> que ressaltam a importância da gestão de grandes volumes de dados

---

<sup>3</sup> Research Data Alliance – iniciativa que visa construir uma infraestrutura social e técnica para permitir o compartilhamento aberto e a reutilização de dados. Disponível em: <https://www.rd-alliance.org/about-rda/>

<sup>4</sup> Comitee on Data – iniciativa que promove a Ciência Aberta e maior disponibilidade e usabilidade dos dados para todas as áreas de pesquisa. Disponível em: <https://codata.org/about-codata/our-mission/>

<sup>5</sup> Programas de pesquisas em sistemas agroalimentares. Disponível em: <https://www.cgiar.org/research/research-portfolio/>

científicos que se utilizam de proveniência de dados e princípios FAIR tanto nos contextos das HD (ALLEA, 2020) quanto da Agricultura Digital (AD) (KLERSKX et al, 2019). Este manuscrito é um resumo que apresenta os primeiros esforços interdisciplinares desenvolvidos em nosso grupo de pesquisa para investigar as correlações entre HD e agronegócio (com foco na AD), considerando a adoção de Princípios FAIR e Proveniência de Dados para o enriquecimento dos Repositórios Digitais<sup>6</sup> (RD) de solos da plataforma OpenSoils (CRUZ et al, 2018).

## 2 Metodologia

O método de investigação adotado neste trabalho foi o exploratório, de abordagem de análise de conteúdos cujo objeto foi o estudo de caso aplicado ao RD da plataforma OpenSoils (CRUZ et al, 2018). O levantamento sistemático de dados na literatura (KITCHENHAM; CHARTERS, 2007) envolveu os domínios da HD e AD, realizado a partir de literatura da área e documentação disponível em repositórios e acervos públicos. Neste trabalho, incorporamos essa percepção de Moreira et al. (2019), o autor apregoa que sob a ótica das pesquisas centradas em grandes volumes de dados digitais, a gestão sistemática dos RD é importante não só para perpetuar pesquisas e seus resultados como também assegurar sua interoperabilidade de dados promovendo o intercâmbio e disseminação dos seus metadados. Nesta pesquisa aplicamos essa percepção ao contexto do RD da plataforma, um repositório voltado para a área de solos que já conta com algum aporte para a proveniência de dados.

## 3 Princípios Fair e Proveniência de Dados

Os princípios FAIR são elementos norteadores de alto nível para aprimorar a gestão de dados e que ainda são pouco explorados tanto em HD quanto em AD. Existe farta literatura sobre o tema. Resumidamente temos que os princípios referem-se a: i) localização de dados (*Findable*) por humanos e aos computadores; ii) acessibilidade de dados (*Accessible*), sob autenticação e autorização; iii) interoperabilidade de dados

<sup>6</sup> Os repositórios digitais (RDs) são bases de dados que reúnem de maneira organizada a produção científica de uma instituição ou área temática. Disponível em: <https://bit.ly/36epmDS>

(*Interoperable*) com aplicativos ou fluxos de trabalho para análise, armazenamento e processamento; iv) tudo isso possibilitando a reutilização dos dados (*Reusable*), graças a presença de metadados que descrevem os *datasets* bem descritos, para que possam ser replicados e/ou combinados em diferentes configurações. Já a Proveniência de Dados, do ponto de vista da computação, é definida como metadados que descrevem entidades, atividades e pessoas envolvidas na produção de um dado ou coisa, que podem ser usadas para formar avaliações sobre sua qualidade ou confiabilidade de dados e processos e segundo Sayão (2010), é a “história do objeto [digital] armazenado; pode incluir informações sobre sua fonte ou origem, sua cadeia de custódia; registra também as ações de preservação sofridas pelo objeto e seus efeitos, por exemplo: as migrações efetuadas”. Neste trabalho, nosso referencial teórico de proveniência é o modelo da PROV, utilizado pelo W3C (W3C, 2013).

#### **4 Correlações entre Humanidades Digitais e Agricultura Digital**

Nossa investigação demanda uma conjunção de saberes e apresenta vários desafios, um deles é reconhecer o gigantismo e a complexidade dessas áreas; o outro é detectar o delineamento das interseções, compreendê-las e, por fim investigá-las. Portanto, é necessário definir um recorte. Neste trabalho investigaremos como um dos mais principais recursos não renováveis da cadeia de produção do agronegócio (que tem imensos impactos no dia a dia das sociedades) pode se beneficiar de abordagens da HD. No caso específico, iniciamos nossos estudos sobre o recurso solo cujos dados digitalizados estão armazenados em um RD.

Segundo Alves (2016) há uma forte ligação entre a investigação em Humanidades e a incorporação de métodos e ferramentas das Tecnologias Digitais. O termo ‘Humanidades Digitais’ foi cunhado internacionalmente há pouco mais de uma década. Na AD, ainda existem poucos trabalhos evidenciando sua correlação com as HD mesmo quando se considera o grande volume de dados e a extrema necessidade de gestão de dados produzidos. Atualmente, a AD, em especial no que diz respeito à

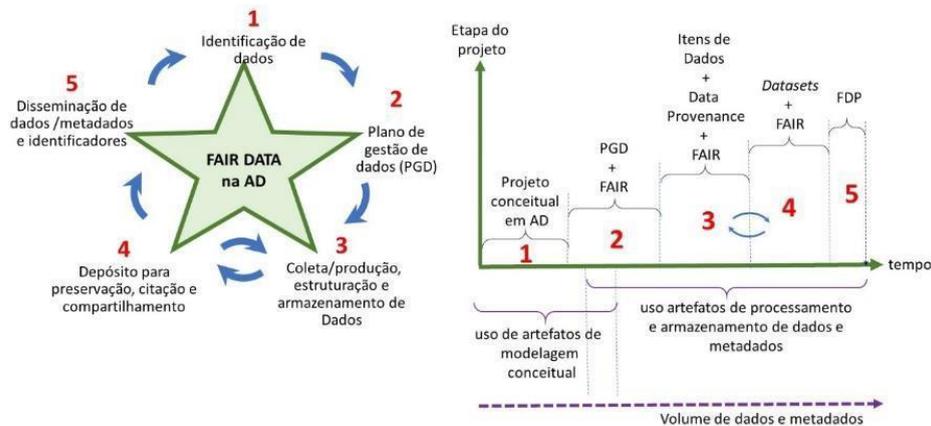
sustentabilidade e segurança de solos são temas centrais da Agenda 2030 da ONU<sup>7</sup>. A área de solos é central nas cadeias agroalimentares e industriais das sociedades modernas. Está sendo transformada por grandes volumes de dados coletados por sensores, experimentos de campo e imagens de satélites e mapas, em especial, ajudando consumidores, agricultores, profissionais da área, pesquisadores e formuladores de políticas a tomar decisões mais inteligentes e racionais sobre usos, recomendações e mapeamentos dos solos nas áreas urbanas e rurais. (L'HÉNAFF; SMITH, 2018 apud MARINHO; PINHEIRO; BESSA, 2020).

Defendemos que as HD podem propiciar novas possibilidades na investigação cruzada com a AD devido a sua conexão com práticas arquivísticas, bibliométricas, humanísticas e computacionais no que diz respeito ao trato dos ativos informacionais. As correlações podem ser ilustradas a seguir (Figura 1). Se iniciam com a elaboração e identificação dos projetos, definição de suas regras, atores e seus papéis ao longo do ciclo de vida do projeto de geração dos dados, em geral não há grandes esforços de processamento de informações, os maiores esforços estão nas áreas de modelagens e devem levar em conta a definição e a identificação dos atributos de dados (semanticamente bem definidos), seus tipos, formatos e estruturas e datasets a serem produzidos.

---

<sup>7</sup> Agenda 2030 com os 17 Objetivos de Desenvolvimento Sustentável - Disponível em: <http://www.agenda2030.com.br/>

**Figura 1: Representação conceitual (estrela) com as etapas de um projeto AD e correlação com princípios FAIR e proveniência de dados e diagrama que representa as etapas ao eixo do tempo.**



Fonte: Os autores, 2020.

A segunda etapa do ciclo se alia a necessidade de definição de planos de gestão dos grandes volumes de dados (PGD) para os projetos de AD. Destacamos que ainda existem poucos trabalhos na literatura que explicitamente correlacionam PGD com AD e papéis e atores. Os PGDs na AD podem ser elaborados por ferramentas digitais e devem incluir as metodologias da criação/coleta de dados, modelos de dados, documentação e metadados; aspectos legais e éticos, políticas de armazenamento durante a pesquisa, além de indicações de como os dados serão preservados e compartilhados e descrição das responsabilidades e proprietários dos dados. A partir deste ponto do ciclo a processamento das tarefas tende a ser massivo, a adoção dos princípios FAIR é mais evidente e passa a desempenhar um papel central para o sucesso do projeto.

A terceira etapa do ciclo se aproxima da operacionalização da coleta de itens de dados e execução de tarefas computacionais, além da estruturação e armazenamento dos *datasets*. Essa etapa é central, há um ciclo interno de produção de itens de dados com os *datasets*. Ela assegura a confiabilidade e a acessibilidade dos dados ao longo do tempo. A partir deste ponto a proveniência de dados, do tipo retrospectiva, passa a desempenhar um papel crucial, visto que sua coleta e armazenamento junto aos

metadados e *datasets* são mecanismos complementares que visam assegurar os aspectos de qualidade dos dados, reprodutibilidade dos processos geradores de dados em AD.

A quarta etapa está intimamente relacionada com a preservação e compartilhamento dos *datasets* ao longo do tempo. Inevitavelmente, a etapa suscita novas situações que estão além das questões tecnológicas anteriores, levantam muitos pontos relacionados que se alinham com o sucesso do projeto que vão desde a propriedade intelectual, aspectos legais de licenças de uso/compartilhamento até pontos ligados à privacidade e autoria (citação) dos *datasets*. Por fim, a quinta etapa trata dos requisitos necessários a assegurar a curadoria e eficiência da disponibilidade dos *datasets* ao longo do tempo, destacamos que aos *datasets* devem ser atribuídos os *persistent identifiers* (PID) além de etiquetas de metadados para assegurar a acessibilidade automática por máquinas através de RDs do tipo *Fair Data Points* (FDP), ampliando a visibilidade desejada dos dados e do projeto propriamente dito, além de aprimorar a preservação dos ativos digitais presentes nos *datasets*.

## 5 Considerações finais

Este trabalho apresentou um primeiro estudo interdisciplinar que visou a investigar aspectos comuns entre HD e AD e avaliar onde elas se aproximam, considerando a adoção de princípios FAIR e Proveniência de Dados tendo como caso de uso a plataforma OpenSoils ([www.opensoils.org](http://www.opensoils.org)). Ressaltamos que este texto não é um trabalho exaustivo e se encontra em andamento.

Dentro desse contexto e do recorte proposto, analisamos neste trabalho apenas o RD da plataforma, percebemos que ele pode evoluir e se beneficiar mais explicitamente dos Princípios FAIR e da Proveniência ao longo de suas próximas fases de desenvolvimento, contribuindo para tornar pesquisas voltadas para segurança de solos e aplicações no campo, mais confiáveis, seguras e dinâmicas e cada vez mais alinhadas com a fundamentação de Ciência Aberta e dos Dados Abertos. Adicionalmente, apresentamos uma primeira versão do ciclo de vida contendo etapas que explicitam correlações entre Princípios FAIR e Proveniência de Dados comuns às áreas de HD e AD.

Destacamos que a plataforma, não originalmente concebida segundo os princípios da HD, poderá ser refatorada e enriquecida para se alinhar ao ciclo descrito na Figura 1, em especial ao que diz respeito à adaptação de RD e aumento da acessibilidade, reusabilidade e interoperabilidade dos dados. Como trabalhos futuros pretendemos aprofundar esses estudos e delinear o processo de abertura de dados para torná-los mais facilmente localizáveis através da *fairificação* do repositório da plataforma OpenSoils e disponibilização através de um FDP.

## Referências

ALBAGLI, S., APPEL, A. L., MACIEL, M. L. *E-science*, ciência aberta e o regime de informação em ciência e tecnologia. Disponível em:

<https://ridi.ibict.br/bitstream/123456789/854/1/124-540-1-PB.pdf>

ALLEA (2020). Sustainable and FAIR Data Sharing in the Humanities: Recommendations of the ALLEA Working Group E-Humanities. Edited by Natalie Harrower, Maciej Maryl, Beat Immenhauser, and Timea Biro. Berlin. DOI: 10.7486/DRI.tq582c863

ALVES, D. As Humanidades Digitais como uma comunidade de práticas dentro do formalismo acadêmico: dos exemplos internacionais ao caso português. Disponível em: <https://journals.openedition.org/lerhistoria/2496>

BUNEMAN P., KHANNA S., TAN WC. (2000) Data Provenance: Some Basic Issues. Disponível em: [https://doi.org/10.1007/3-540-44450-5\\_6](https://doi.org/10.1007/3-540-44450-5_6)

CASTRO, R. M. de, PIMENTA, R. M. Novas práticas informacionais frente às humanidades digitais: a construção de acervos digitais. *Inf. Inf.*, Londrina, v. 23, n. 3, p. 523 – 543, set./dez. 2018.

CRUZ, S. M. et al. Proveniência de dados na agricultura (2018). Disponível em: <https://bit.ly/3o2ixuV>

HENNING, P. C. et al. GO FAIR e os princípios FAIR: o que representam para a expansão dos dados de pesquisa no âmbito da Ciência Aberta. Disponível em: DOI: 10.19132/1808-5245252.389-412

KITCHENHAM, B., CHARTERS, S. Guidelines for performing Systematic Literature Reviews in Software Engineering. EBSE Technical Report EBSE-2007-01. Disponível em: <https://bit.ly/3o3EGsR>

KLERKX, L., JAKKU, E., LABARTHE, P. A review of social science on digital agriculture, smart farming and agriculture 4.0: New contributions and a future research agenda. Disponível em: <https://doi.org/10.1016/j.njas.2019.100315>

MARINHO, E.C.; PINHEIRO, BESSA et al., 2020. Proteção de Dados: Investigações sobre gerenciamento de dados pedológicos usando os princípios FAIR e tecnologias blockchain. Disponível em: <https://bit.ly/3q9IdYD>

MOREIRA, J. L. R.; BONINO, L.; PIRES, L. F.; VAN SINDEREN, M.; HENNING, P. Repositórios para dados localizáveis, acessíveis, interoperáveis e reutilizáveis (FAIR): adaptando um repositório de dados para se comportar como um FAIR Data Point. *Liinc Em Revista*, 15(2), 2019. Disponível em: <https://bit.ly/2JfwMhh>

RBHD, Rio de Janeiro, v. 1, n. 1, Dossiê Temático 3, p. 72-80, jan./jun., 2021

RUSSELL, I. G. ¿Qué son las Humanidades Digitales? Revista Digital Universitaria 1 de julio de 2011 Vol.12, No.7. Disponível em: <https://bit.ly/3mgEGFg>

SAYÃO, L. F. “Uma outra face dos metadados: informações para a gestão da preservação digital,” Encontros Bibli: Rev. Eletrônica Bibliotecon. e Ciência da Informação, pp. 1–31, 2010.

WILKINSON, M., DUMONTIER, M., AALBERSBERG, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>