



Governança de dados na agricultura digital: relato da experiência do C4AI

Cássia I. C. Mendes^{1,4}, Fernando E. Correa², Debora P. Drucker¹, Filipi M. Soares^{2,5}, Antônio M. Saraiva^{2,5}, Alexandre C. D. Delbem^{2,6}, Fernando S. Osório^{2,6}, Cristina G. B. Oliveira³, Juliano Maranhão^{2,4}

¹ Embrapa Agricultura Digital - Av. Dr. André Tosello, 209 - Cidade Universitária
13083-886 - Campinas - SP

² Center for Artificial Intelligence - (C4AI) da Universidade de São Paulo - Av. Prof.
Lúcio Martins Rodrigues, 370 - Butantã, São Paulo - SP, 05508-020

³ Faculdade de Direito de Ribeirão Preto da Universidade de São Paulo (FDRP-USP) -
Av. Bandeirantes, 3900 - Ribeirão Preto - SP, 14040-906

⁴ Faculdade de Direito da Universidade de São Paulo (FD-USP) - Largo São Francisco,
95 - Centro - São Paulo, 01005-010

⁵ Escola Politécnica da Universidade de São Paulo (EP-USP) - Av. Prof. Luciano
Gualberto, Travessa do Politécnico, 380 - Butantã, São Paulo, 05508-010

⁶ Instituto de Ciências Matemáticas e de Computação (ICMC-USP) - Av. Trabalhador
São-Carlense, 400 - São Carlos, 13566-590

{cassia.mendes, debora.drucker}@embrapa.br, acbd@icmc.usp.br, fosorio@icmc.usp.br, {saraiva, cristinagodoy, julianomaranhao, fecorrea, filipisoares}@usp.br

Resumo. A implementação efetiva de sistemas de IA na agricultura digital apresenta desafios organizacionais devido às diversas fontes, formatos e volume de dados processados, para diferentes finalidades. O artigo examina ações de governança de dados empregadas em pesquisas realizadas no Centro de Inteligência Artificial (C4AI) da Universidade de São Paulo (USP). Discorreremos sobre o gerenciamento de conjuntos de dados por meio de uma colaboração transdisciplinar entre Agricultura, Ciência da Computação e Direito. Foram realizadas análises sobre o marco legal de privacidade e proteção de dados, a ISO 27001 e a política de proteção de dados da USP. Os resultados incluem a criação de comitê responsável por desenvolver acordo de uso de dados, política de privacidade e cookies, plano de gerenciamento de dados do projeto, termo de consentimento e ações de treinamento voltadas para a comunidade de pesquisadores de diferentes áreas do conhecimento do C4AI.

Abstract. The practical implementation of AI systems in digital agriculture presents organizational challenges due to the different sources, formats, and volumes of data processed for different purposes. The article examines the data governance actions employed in research at the Center for Artificial

Intelligence (C4AI) at the University of São Paulo (USP). We discuss the management of datasets through a transdisciplinary collaboration between Agriculture, Computer Science, and Law. We carried out analyzes on the legal privacy and data protection framework, including ISO 27001 and data protection policy of the University of São Paulo. The results include the creation of a committee responsible for developing a data use agreement, privacy and cookies policy, project data management plan, consent form, and training actions aimed at the community of researchers from different areas of C4AI knowledge.

1. Introdução

No contexto da sociedade 5.0, caracterizada pela mescla entre o espaço físico (mundo real) e o ciberespaço, alavancando as Tecnologias da Informação e Comunicação (TIC) em todos os setores econômicos, os governos e as instituições são guiados pela ciência, tecnologia e inovação intensificados por meio de uso de big data [Hitachi-Utoko, 2020].

Esse fenômeno também se verifica na agricultura digital colaborativa que utiliza grandes variedades de dados acumulados por meio de diversos sistemas de informação (governamentais, comerciais e *opensources*), com distintas finalidades, tipos de dados e formas de acesso, sendo que, em muitos casos, esses dados são utilizados para treinar ferramentas e algoritmos de inteligência artificial.

O aumento exponencial no volume da coleta, transmissão e armazenamento de dados agropecuários exige das organizações a utilização de mecanismos de governança de dados com os objetivos de promover o gerenciamento de dados como um ativo corporativo estratégico, bem como realizar a gestão de riscos associados (tais como privacidade, vazamento ou uso indevido) [Davidson *et al.* 2023].

Os dados podem ser utilizados para o bem social e para melhorar a vida das pessoas. Entretanto, também há o risco de que países, empresas e indivíduos sejam prejudicados caso não se estabeleça uma regulamentação que permita a coleta e o uso ético dos dados, equilibrando a criação de valor econômico e social com o acesso equitativo e a confiança na utilização justa dos dados [World Bank 2021].

Neste contexto, os países têm aprovado leis para regulamentar o tratamento de dados pessoais e não pessoais. A União Europeia, precursora na edição de leis dessa natureza, aprovou o Regulamento Geral de Proteção de Dados (GDPR) para proteção de dados pessoais [União Europeia 2016] e está discutindo a proposta de Regulamento de Dados [União Europeia 2023], em trâmite no Parlamento Europeu, relativo às regras harmonizadas sobre o acesso equitativo aos dados (pessoais e não pessoais) e a sua utilização. No Brasil, destaca-se, nesse sentido, a Lei Geral de Proteção de Dados Pessoais (LGPD) [Brasil 2018], inspirada no GDPR, possuindo muitas semelhanças em diversos momentos. Nota comum naquelas leis é a necessária adoção, por países e organizações, de arcabouços de governança de dados para promover a privacidade, a proteção de dados pessoais, a segurança da informação e a circulação dos dados.

A governança de dados estabelece diretrizes, processos, procedimentos e responsabilidades para que nações e instituições possam lidar e tratar com os dados e as informações, de forma que proteja direitos fundamentais dos indivíduos e organizações afetadas. Conforme sugerido por Davidson *et al.* (2023), a implementação desses processos de governança de recursos de dados pode se concentrar em três níveis: 1) o

macro (setorial, como a agricultura e a indústria); 2) o meso (organizacional e projetos) e 3) o micro (equipes e indivíduos).

Considerando que a regulamentação do tratamento de dados – pessoais e não pessoais – é tema relativamente novo, há lacuna na literatura científica no que diz respeito a trabalhos que abordem o processo de governança de dados nos níveis meso (de organizações públicas de ensino e de PD&I) e micro (equipes e pessoas).

Diante do exposto, o objetivo deste trabalho é discorrer sobre as ações de governança de dados em andamento no âmbito dos projetos de pesquisas desenvolvidas pelo Centro de Inteligência Artificial (C4AI) da Universidade de São Paulo, em visão transdisciplinar entre a Agricultura, Ciência da Computação e Direito, nos níveis meso (numa instituição de ensino, a USP) e micro (na equipe de pesquisadores do C4AI). A contribuição do trabalho é apresentar aspectos teórico-práticos sobre ética e privacidade na gestão de dados no contexto de pesquisas em inteligência artificial na agricultura.

Para tanto, o artigo está estruturado em 6 seções, incluindo esta introdução. Como as ações de governança de dados abrangem quatro etapas, as seções discorrem sobre cada uma delas. A seção seguinte aborda a primeira etapa referente aos aspectos de segurança da informação e à aplicação da ISO 27001. Uma vez utilizadas as ferramentas de segurança da informação em *hardware* e *software* do parque computacional do C4AI, a próxima seção discorre sobre a segunda etapa, ou seja, o cumprimento dos requisitos legais da governança de dados pessoais e não pessoais no âmbito dos projetos. A seção posterior relata a terceira etapa sobre a aplicação dos princípios FAIR (*Findable, Accessible, Interoperable, Reusable*) para a gestão de dados do C4AI. A próxima dispõe sobre a quarta etapa e os resultados obtidos por meio da elaboração de termo e condições de uso de dados, política de privacidade e de cookies, plano de gestão de dados de projeto, termo de consentimento e a proposta de ações de capacitação voltadas para a comunidade de pesquisadores do C4AI. Por último, apresentam-se as conclusões.

2. Segurança da informação

Nas ações governança de dados do C4AI, busca-se, na primeira etapa, analisar os pontos relativos à segurança da informação com foco em *hardware* e *software*, ou para boas práticas usadas em áreas estritamente de tecnologia da informação, tais como redes de computadores ou desenvolvimento de sistemas computacionais do C4AI. Cumpre destacar que com o advento das redes sociais e a exposição de informações pessoais e corporativas cada vez mais rápida e frequente, novas metodologias de governança e aplicação/interpretação de leis vêm sendo trabalhadas, por vezes, com alguma sobreposição.

Pode-se citar, como exemplos, tanto a Lei Geral de Proteção de Dados Pessoais (LGPD) como o Marco Civil da Internet. Em ambos casos, ao estudar as boas práticas mencionadas nestas leis, uma parte delas terá a coincidência ou sobreposição de temas relativos à segurança da informação, gestão da informação e ciclo de vida de dados, sendo pontos avaliados por instituições que executam auditorias para verificação de boas práticas e normas como a ISO/IEC 27001 [Lima et al. 2022]. Esta coincidência temática ocorre, principalmente, por serem publicadas mencionadas normas em anos distintos, sendo que o legislador perdeu a oportunidade, em 2018, de harmonizar as normas relevantes para a área da inovação: Legislação protetiva dos direitos da propriedade intelectual; Marco Civil da Internet e Lei Geral de Proteção de Dados.

Deve-se destacar que a obtenção de determinada certificação ou a elaboração de um código de boas práticas não garantem a segurança cibernética da instituição, das pessoas ou dos sistemas. Outro fator importante é a ausência da manutenção de investimentos a médio e a longo prazo para centros de pesquisa e instituições públicas, a fim da renovação periódica das certificações [Fazenda; Fagundes 2015].

Conseqüentemente, são relacionados os pontos que tangem à segurança da informação, à gestão e ao ciclo de vida de dados os quais norteiam as ações de governança de dados do C4AI, com vistas a garantir segurança da informação razoável e compatível com as boas práticas observadas em instituições governamentais ou não.

Primeiro, destacam-se os fundamentos de segurança da informação que guiam as ações de governança de dados do C4AI [Sêmola 2003]:

- **Confidencialidade:** garantia de que a informação será acessada somente pela parte autorizada;
- **Integridade:** manutenção da informação sem alteração não registrada ou com indevida violação. O corrompimento de uma informação pode ocorrer por erro, falha ou intencionalmente;
- **Disponibilidade:** disponibilização da informação para o seu acesso quando solicitada ou necessária.

Outros dois fundamentos, que por vezes não estão diretamente relacionados nas boas práticas; mas que se aplicam ao C4AI, são:

- **Não repúdio:** capacidade do sistema de indicar quem fez a ação, impedindo o usuário de negar essa conduta;
- **Privacidade:** capacidade do sistema de manter anônimo um usuário, impossibilitando o relacionamento entre usuário e suas ações.

Desta forma, a governança de dados deve atender e expor aos usuários e aos produtores de informação do C4AI a necessidade de aderir aos princípios e aos fundamentos de segurança da informação apresentados previamente pelo centro de pesquisa (C4AI).

3. Governança de dados pessoais e não pessoais

Adicionam-se aos fundamentos de segurança da informação, a gestão e o ciclo de vida de dados que integram a governança de dados de uma organização, a qual abrange procedimentos e diretrizes para que as diferentes áreas tratem, de forma padronizada, os dados e as informações corporativas. A gestão de dados é a instrumentalização da governança de dados por intermédio da definição de processos organizacionais para planejar, controlar e monitorar a execução das políticas e diretrizes definidas.

No Brasil, a Lei Geral de Proteção de Dados Pessoais (LGPD), em seu artigo 50, estabelece os requisitos para a elaboração de regras de governança e de boas práticas pelos controladores (públicos ou privados). Em síntese, o programa em privacidade é estruturado em três conjuntos: a) planejamento e iniciação; b) construção e execução; c) monitoramento e avaliação [Mendes, Bertin e Costa 2023].

Os gestores do Centro de Inteligência Artificial (C4AI), para cumprimento à LGPD no âmbito de suas pesquisas, estão implementando ações de governança de dados. O C4AI, vinculado à Universidade de São Paulo (USP), atende às diretrizes estabelecidas pela USP, dentre elas a Portaria 7839/2022 que dispõe sobre a aplicação da LGPD e criou o Comitê Gestor de Governança de Dados e Informações na Universidade. Em seu artigo

6o., a Portaria estabelece que as Unidades ou órgãos da USP deverão internalizar a Política de Proteção de Dados Pessoais, detalhando-a, mediante regulamentação complementar própria e sempre de acordo com as normas gerais da Universidade de São Paulo, tendo em vista as suas respectivas especificidades [Universidade de São Paulo 2022].

Para tanto, a segunda etapa da implementação de ações de governança de dados do C4AI refere-se ao cumprimento dos requisitos legais de gestão de dados pessoais e não pessoais no âmbito de seus projetos. Desse modo, os gestores do C4AI criaram um comitê multidisciplinar, com pesquisadores e professores das áreas de Ciência da Computação, Direito e Agricultura. O comitê ficou responsável pela proposição dos seguintes instrumentos para gestão de dados no âmbito dos projetos de pesquisa em andamento: 1) termo de uso de dados; 2) política de privacidade e de cookies; 3) plano de gestão de dados de projeto; e 4) termo de consentimento. Tais documentos estão detalhados na seção 5 (Resultados).

4. Gestão de dados em órgãos de ensino e pesquisa agrícola

Uma vez implementadas as etapas de segurança da informação e de cumprimento do marco legal, a terceira etapa refere-se à adoção dos princípios FAIR (*Findable, Accessible, Interoperable, Reusable*) para a gestão de dados.

É notório como a transição para a ciência intensiva em dados, em curso nas últimas décadas em todo o mundo, impõe mudanças de processos e de tecnologias de modo a implementar boas práticas de gestão a serem adotadas e a assegurar a confiabilidade e a reprodutibilidade dos resultados. Os princípios FAIR, acrônimo em inglês para dados “localizáveis, acessíveis, interoperáveis e reutilizáveis” e propostos por Wilkinson *et al.* (2016), vêm sendo adotados como norteadores da gestão de dados de pesquisa e viabilizadores de sua reutilização [Drucker *et al.* 2021].

No contexto brasileiro, inúmeras iniciativas em instituições de ensino e pesquisa estão promovendo boas práticas para a gestão de dados científicos e trocando experiências em arenas como a da Parceria para o Governo Aberto [Controladoria-Geral da União 2023]. Notadamente, o Compromisso 3 do 4º Plano de Ação Brasileiro para Governo Aberto – que ficou conhecido como ‘Compromisso pela Ciência Aberta’ – teve como elemento central o pilar dos dados abertos, com o objetivo de “estabelecer mecanismos de governança de dados científicos para o avanço da Ciência Aberta no Brasil” [Fortaleza *et al.* 2022].

A Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) foi a primeira agência de fomento nacional a exigir, desde 2017, planos de gestão de dados (PGD) de pesquisa como item obrigatório em submissões de propostas. Inicialmente, em algumas modalidades, e atualmente, para todos os projetos.

Acompanhando essas iniciativas, o C4AI também adota os princípios FAIR em seus projetos de pesquisa. Destaca-se também que a proposta do C4AI submetida à FAPESP foi acompanhada de um plano de gestão de dados que está sendo implementado pelo Centro de Pesquisa. Nesse sentido, destaca-se que foram disponibilizados serviços para o suporte à pesquisa em nuvem com versionamento de códigos, processamento de modelos e visualização, armazenamento de arquivos e implementação de banco de dados. Outra ação é a proposta de modelo de metadados, que servirá de base para o desenvolvimento

de uma arquitetura de dados, a qual está em desenvolvimento no âmbito do Grupo de Trabalho de Dados, Metadados e Repositórios da Rede GO FAIR Agro Brasil.

5. Resultados

A quarta etapa das ações de governança de dados refere-se aos resultados obtidos: 1) elaboração de documentos; e 2) ações de capacitação.

No que concerne à elaboração de documentos, o já mencionado comitê interno multidisciplinar elaborou os seguintes documentos: a) termo de uso de dados; b) política de privacidade e de cookies; c) plano de gestão de dados de projetos; d) termo de consentimento. Os documentos estão em fase de análise e validação junto à coordenação do C4AI. A Tabela 1 relaciona e descreve, brevemente, os seus conteúdos e as fontes de referência que orientaram a elaboração dos documentos.

Tabela 1. Instrumentos para gestão de dados em projetos do C4AI

Documento e Destinatário	Descrição	Referências
Termo de uso de dados Pesquisadores do C4AI.	Descreve as ações e as responsabilidades de pesquisadores para tratamento de dados pessoais no âmbito dos projetos.	Lei Geral de Proteção de Dados Pessoais – LGPD
Termo de consentimento para tratamento de dados pessoais Pessoas físicas	Estabelece as hipóteses legais que exigem o consentimento expresso, livre e informado; os direitos do titular de dados pessoais; as responsabilidades dos agentes de tratamento de dados pessoais e os princípios para uso dos dados pessoais, em conformidade com a Lei Geral de Proteção de Dados (LGPD).	[Brasil 2018]
Política de privacidade e de cookies Usuários de websites e aplicativos	Dispõe sobre os cuidados razoáveis que a equipe do C4AI adotou, de acordo com o padrão de mercado e as recomendações da Autoridade Nacional de Proteção de Dados (ANPD) para proteger os dados pessoais.	Guia Orientativo de Cookies e Proteção de Dados Pessoais [ANPD 2022]
Plano de gestão de dados de projeto Agência de fomento à pesquisa	Detalha os dados que serão gerados no projeto, formatos, padrões, compartilhamento, restrições legais e armazenamento de dados.	Gestão de Dados de Projetos [FAPESP 2023]

Fonte: elaboração dos autores.

Adicionalmente, estão previstas ações de capacitação e treinamento para a correta gestão dos dados pelos pesquisadores de diferentes áreas do conhecimento. Assim, serão apresentados os procedimentos para a correta internalização dos documentos, o detalhamento acerca das situações para uso de dados pessoais e tratamento de dados de forma adequada ao âmbito da governança de dados.

6. Conclusões

O uso intensivo de dados na agricultura digital colaborativa é uma realidade inexorável, apresentando desafios às instituições de ensino e de pesquisa concernentes à adoção de melhores práticas de governança de dados.

Este artigo apresentou as ações de governança de dados em andamento no âmbito das pesquisas do C4AI e descreveu as etapas de segurança da informação, de cumprimento do marco legal de proteção de dados pessoais, de adoção dos princípios FAIR e de elaboração de documentos.

Para que ocorra uma efetiva governança de dados, é preciso que organizações estabeleçam diretrizes e procedimentos nos níveis meso (da instituição) e micro (das equipes e dos indivíduos). Como se evidenciou, foram adotadas ações no nível meso (no contexto institucional, com a edição da Portaria USP 7839/2022 que dispõe sobre a aplicação da LGPD e criou o Comitê Gestor de Governança de Dados e Informações), bem como no nível micro (com a proposição do termo de uso de dados, política de privacidade e de cookies, plano de gestão de dados, termo de consentimento e das ações de capacitação).

Neste sentido, ressalta-se que este trabalho não exaure todas as dimensões de um programa de governança em privacidade e em proteção de dados, considerando que as instituições, como é o caso da USP, ainda estão se adequando às políticas e aos processos de coleta, tratamento, processamento de dados pessoais para o atendimento à LGPD.

Com o aumento exponencial na coleta, transmissão e armazenamento de dados agropecuários, a realização eficaz da governança de dados é um desafio que está na pauta das organizações e dos países no contexto da agricultura digital colaborativa, cada vez mais caracterizada pelo trabalho em rede multi-institucional e transdisciplinar.

7. Referências

- Autoridade Nacional de Proteção de Dados - ANPD (2022). Guia Orientativo Cookies e proteção de dados pessoais. Disponível em: <https://www.gov.br/anpd/pt-br/documentos-e-publicacoes/guia-orientativo-cookies-e-protecao-de-dados-pessoais.pdf>. Acesso em: 05 jul. 2023.
- Brasil. Lei no 13.709, de 14 de agosto de 2018. (2018) Lei Geral de Proteção de Dados Pessoais. Disponível em: https://www.planalto.gov.br/ccivil_03/ato2015-2018/2018/lei/L13709compilado.htm. Acesso em: 12 jul. 2023.
- Controladoria-Geral da União. (2023). Governo Aberto. Disponível em: <https://www.gov.br/cgu/pt-br/governo-aberto/a-ogp>. Acesso em: 17 jul. 2023.
- Davidson, E., Wessel, L., Winter, J. S. and Winter, S. (2023). Future directions for scholarship on data governance, digital innovation, and grand challenges. *Information and Organization*. Volume 33, Issue 1, March, 100454. DOI: <https://doi.org/10.1016/j.infoandorg.2023.100454>
- Drucker, D. P.; Cruz, S. M. S. da; Saraiva, A. M.; Fortaleza, J. M.; Bertin, P. R. B.; Simao, V. P. M.; Telles, M. A.; Silva, A. R. da; Santos, P. S. S.; Macário, C. G. do N. (2021). Implantação da Rede Temática GO-FAIR Agro Brasil: primeiros passos. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 13., Bagé. Anais [...]. Bagé: Unipampa, 2021. p. 164-171. 2177-9724

- Fapesp. Gestão de Dados. (2023). Disponível em: <https://fapesp.br/gestaodedados>. Acesso em: 05 jul. 2023.
- Fazenda, R.; Fagundes, L. (2015). Análise dos desafios para estabelecer e manter sistema de gestão de segurança da informação no cenário brasileiro. In: Anais do XI Simpósio Brasileiro de Sistemas de Informação. SBC, p. 307-314.
- Fortaleza, J. M.; Bertin, P. R. B.; Drucker, D. P. (2022). O compromisso pela ciência aberta na Open Government Partnership: avanços na governança de dados científicos no Brasil. In: Príncipe, E.; Rode, S. de M. (org.). Comunicação científica aberta. Rio de Janeiro: Instituto Brasileiro de Informação em Ciência, p. 189-206. (Coleção PPGCI 50 anos).
- Hitachi-Utokyo Laboratory. (2020). Society 5.0 A People-centric Supersmart Society. Tokyo: Spring Open-University of Tokyo. Disponível em: https://library.oapen.org/bitstream/handle/20.500.12657/41719/2020_Book_Society50.pdf?sequence=1. Acesso em: 05 jul. 2023.
- Lima, P. R. S.; Ferreira, L. M. M.; De Albuquerque Peixoto, A. L. V. (2022). Gestão da segurança da informação: análise de políticas de defesa cibernética e estratégias para a proteção de dados e informações da administração pública brasileira. P2P E INOVAÇÃO, v. 9, n. 1, p. 206-221.
- Mendes, C.I.C.; Bertin, P. R. B.; Costa, M. M. (2023) Programa de Governança em Privacidade e Proteção de Dados Pessoais na Administração Pública Federal. *Administração de Empresas em Revista*. Vol.2, n.32|e-6367| p.506-543 |Abril/Junho. Disponível em: <http://revista.unicuritiba.edu.br/index.php/admrevista/article/view/6367>. Acesso em: 12 jun. 2023.
- Sêmola, M. Gestão da segurança da informação: uma visão executiva. Rio de Janeiro: Ed. 2003.
- União Europeia. (2016). Regulation (EU) 2016/679 (General Data Protection Regulation). Official Journal of the European Union. Disponível em: <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN>. Acesso em: 24 maio 2023.
- União Europeia (2023). Regulamento Dados 2022/0047. Proposta de Regulamento do Parlamento Europeu e do Conselho relativo a regras harmonizadas sobre o acesso equitativo aos dados e sua utilização. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:52022PC0068>. Acesso em: 31 maio 2023.
- Universidade de São Paulo. (2022). Portaria GR no. 7839, de 21 de novembro de 2022. Disponível em: <https://leginf.usp.br/?portaria=portaria-gr-no-7839-de-21-de-novembro-de-2022>. Acesso em: 12 jun. 2023.
- Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018. <https://doi.org/10.1038/sdata.2016.18>.
- World Bank (2021). World Development Report 2021: Data for Better Lives. 2021. DOI: <https://doi.org/10.1596/978-1-4648-1600-0>.