



SemeAr DIGITAL: Ações preliminares em gestão de dados

Cássia Isabel Costa Mendes¹ - Embrapa Agricultura Digital, Brasil

Debora Pignatari Drucker² - Embrapa Agricultura Digital, Brasil

Daniel Rodrigo de F. Apolinário³ - Embrapa Agricultura Digital, Brasil

Isaque Vacari⁴ - Embrapa Agricultura Digital, Brasil

Vitor Henrique Vaz Mondo⁵ - Embrapa Agricultura Digital, Brasil

RESUMO CURTO

A digitalização da agricultura intensiva em uso de big data faz emergir o desafio de conciliar a proteção de dados e o avanço tecnológico. O resumo examina ações preliminares de gestão de dados em andamento em um centro de pesquisa desenvolvido pela Embrapa Agricultura Digital e parceiros institucionais. São abordados o projeto piloto Distrito Agro Tecnológico, executado em um município paulista, e o projeto que o sucedeu evoluído para o Centro de Ciência para o Desenvolvimento em Agricultura Digital, em operação em 10 municípios brasileiros. Com abordagem transdisciplinar entre a Agricultura, o Direito e a Ciência da Computação, foram realizadas: consulta ao arcabouço legal e normativo de privacidade e proteção de dados; elaboração de diretrizes de gestão de dados do projeto e aplicação dos princípios FAIR (*Findable, Accessible, Interoperable, Reusable*). Os resultados são: criação de tipologia de dados; análise legal para o tratamento de dados pessoais em projeto de PD&I e plano de gestão de dados submetido à agência de fomento.

Palavras-chave: Agricultura digital. Governança de dados. Direito digital. Proteção de dados.

¹ Advogada. Pós-doutorado em Direito e Doutorado em Desenvolvimento Econômico - cassia.mendes@embrapa.br

² Engenheira Florestal. Doutorado em Ambiente e Sociedade – debora.drucker@embrapa.br

³ Cientista da Computação. Mestrado em Ciência da Computação - daniel.apolinario@embrapa.br

⁴ Tecnólogo em Processamento de Dados. Mestre em Ciência da Computação. isaque.vacari@embrapa.br

⁵ Engenheiro Agrônomo. Doutorado em Produção Vegetal - vitor.mondo@embrapa.br



SemeAr Digital: Ações preliminares em gestão de dados

Introdução

A inserção das tecnologias da informação e comunicação no setor agrícola possibilitou o aumento exponencial na capacidade de geração, transferência, armazenamento e análise de grandes volumes de dados que figuram como insumos essenciais para viabilizar a transformação digital na agricultura.

Segundo Massruhá *et al.* (2021), a agricultura digital é caracterizada pela utilização de tecnologias digitais em todos os elos da cadeia produtiva agrícola, sendo fortemente baseada na análise de big data por meio de ferramentas computacionais para extrair conhecimento relevante para subsidiar o processo decisório na gestão da propriedade rural.

Para fazer frente a esse desafio, a governança de dados contribui para que uma instituição gerencie seus dados como ativos organizacionais relevantes por meio da adoção de processos e políticas para garantir a integridade, a segurança e a proteção legal de seus dados (Mendes, Bertin e Costa, 2023).

Portanto, justifica-se a relevância do tema considerando a necessidade das instituições de ciência e tecnologia (ICTs) abordarem a governança de dados em projetos de pesquisa agrícola, sob a dimensão do direito digital, para conferir segurança jurídica quanto ao caminho adequado para a legitimação do tratamento e compartilhamento de dados em seus projetos de pesquisa.

O trabalho está estruturado em 4 seções incluindo esta introdução, além de apresentar o Centro de Ciência para o Desenvolvimento em Agricultura Digital (CCD-AD/SemeAr Digital), a metodologia utilizada, as análises dos resultados e as considerações finais.

Centro de Ciência para o Desenvolvimento em Agricultura Digital - CCD-AD/SemeAr Digital

As grandes massas de volumes de dados utilizadas na agricultura digital apresentam uma interseção entre os dados coletados em propriedades rurais (denominados dados agrícolas) e os dados pessoais do produtor rural, conforme demonstrado na Figura 1. Isso ocorre porque, por



meio do uso de tecnologias digitais, torna-se possível identificar os agricultores, surgindo o desafio de conciliar a proteção e a privacidade dos dados pessoais, em conformidade com o arcabouço legal vigente, em especial a Lei Geral de Proteção de Dados Pessoais – LGPD (Brasil, 2018), considerando o avanço da grande maioria de pesquisas baseadas em dados (Mendes et al., 2023).

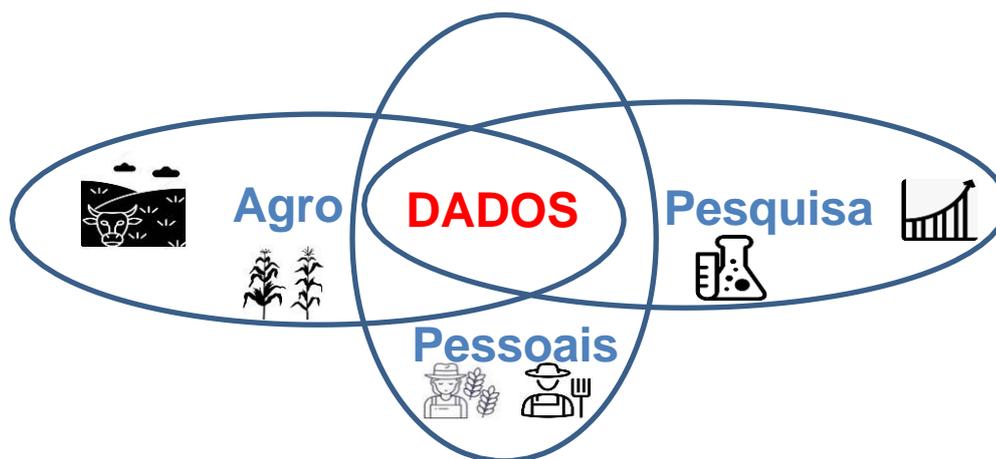


Figura 1 – Interseção entre dados pessoais, agrorados e dados de pesquisa Fonte: elaboração dos autores.

A interseção entre os agrorados coletados em propriedades rurais e os dados pessoais do produtor rural foi evidenciada no CCD-AD/SemeAr, coordenado pela Embrapa Agricultura Digital, intitulado ‘Desenvolvimento e Implementação de Modelo de Distrito Agrodigital como Forma de Desenvolvimento Agropecuário, Econômico, Social e Regional’. O objetivo do Centro é desenvolver e implementar Distritos Agro Tecnológicos (DATs) fomentando a adoção de tecnologias digitais por pequenos e médios produtores rurais. O Centro é desenvolvido em parceria com outros agentes do ecossistema de inovação agropecuária⁶.

O Centro desenvolve tecnologias digitais habilitadoras junto aos produtores rurais, bem como disponibiliza suporte técnico e operação da infraestrutura de conectividade e capacitação. O Centro começou as atividades no início de 2023, sendo que nos anos de 2021

⁶ Instituto de Economia Agrícola (IEA) do Estado de São Paulo, Secretaria de Agricultura e Abastecimento de São Paulo, Instituto Agronômico de Campinas (IAC), Serviço Nacional de Aprendizagem Rural (Senar), Serviço Brasileiro de Apoio às Micro e Pequenas Empresas (Sebrae), Federação da Agricultura e Pecuária do Estado de São Paulo (FAESP) e startups de agtech.



a 2023, duas unidades-pilotos estavam em fase de implementação, ambas no Estado de São Paulo, nos municípios de Caconde e São Miguel Arcanjo (Mondo et al., 2021).

Neste resumo expandido, optou-se por explorar a experiência da unidade-piloto em Caconde, onde foi possível levantar o maior conjunto de dados até o momento, a partir das respostas de 232 produtores rurais a uma consulta *on-line* realizada em 2021.

Metodologia

Nesta pesquisa, utilizou-se uma abordagem exploratória, combinando métodos de: a) análise documental (da legislação aplicável à matéria); b) revisão de literatura; e c) estudo de caso descritivo de um Centro de PD&I da Embrapa, o citado CCD-AD/SemeAr.

Justifica-se a escolha do SemeAr para o estudo de caso pelo fato de que este foi uma das primeiras iniciativas introduzidas após a entrada em vigência da LGPD, em setembro/2020, e que o mesmo previa a coleta de dados, incluindo dados pessoais dos produtores rurais.

O DistritoAgroDIG foi um projeto piloto no que tange à análise de aspectos de direito digital para gestão de dados. Os resultados serão aplicados na iniciativa que o sucedeu, o Centro de Ciência para o Desenvolvimento em Agricultura Digital (CCD-AD/SemeAr Digital). A Tabela 1 relaciona os dez Distritos Agro Tecnológicos (DATs).

Tabela 1 – Municípios que receberão Distritos Agrotecnológicos (DATs)

Região	Município	Estado
Sudeste	Alto Alegre	São Paulo
	Caconde*	
	Lagoinha (SP)	
	São Miguel Arcanjo*	
	Jacupiranga	
	Ingaí	Minas Gerais
Sul	Vacaria	Rio Grande do Sul
Norte	Breves (PA)	Pará
Nordeste	Boa Vista do Tupim (BA)	Bahia
Centro Oeste	Guia Lopes da Laguna (MS)	Mato Grosso do Sul

Fonte: Fapesp (2023)

* Municípios que participam desde o projeto piloto



Análise dos Resultados

No projeto piloto, no município de Caconde, a primeira etapa foi caracterizar os produtores rurais e mapear as condições para viabilização da agricultura digital. Para tanto, foi realizada uma consulta *on-line* aos agricultores locais, cujo formulário com as questões aplicadas e as respostas estão disponíveis em Mondo *et al.* (2022). Com as respostas, foi possível categorizar os conjuntos de dados (dados pessoais, dados agrícolas e dados de pesquisa) levantados a partir das respostas de 232 produtores rurais. Após essa categorização pelos pesquisadores, foi elaborada uma tipologia de dados, conforme Quadro 1, relacionando a legislação aplicável a cada categoria.

Quadro 1. Tipologia de dados: categoria, subtipos, definição e marco legal aplicável aos dados coletados no projeto SemeAr

Categorias	Subtipos	Definição	Legislação aplicável
Dados Pessoais	Dados pessoais	Relacionados à pessoa natural identificada ou identificável.	·Lei Geral de Proteção de Dados Pessoais.
	Dados pessoais sensíveis	Dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado genético ou biométrico, quando vinculado a uma pessoa natural.	
Agrodados	Dados da propriedade rural	Relacionados às operações e gerenciamento da propriedade rural.	·Lei de propriedade industrial;
	Dados agronômicos	Relacionados à produção de plantas, planejamento da produção, dados de solo, clima, entre outros.	·Lei de acesso à informação;
	Dados de implementos agrícolas	Obtidos por meio de controladores de sistemas de sensores da máquina, sistemas criptografados para evitar reengenharia reversa, manutenção.	·Lei de proteção ao programa de computador; ·Lei de direito autoral;



Dados de Pesquisa, Desenvolvimento e Inovação	De insumos agrícolas	Sobre composição de fertilizantes, rações, produtos fito farmacêuticos.	<ul style="list-style-type: none"> ·Lei de proteção de cultivares ·Marco civil da internet; ·Normas de bioinsumos (fertilizantes); ·Normas de bioética; ·Legislação ambiental; ·Lei Geral de Proteção de Dados Pessoais.
	Dados de pesquisa	Registros factuais (pontuações numéricas, registros textuais, imagens e sons) produzidos ou utilizados como fontes primárias para a pesquisa científica e tecnológica e que são necessários para validação dos seus resultados.	<ul style="list-style-type: none"> ·Lei de propriedade industrial; ·Lei de acesso à informação; ·Lei de proteção ao programa de computador; ·Lei de Direito Autoral; ·Lei de proteção de cultivares;
	Dados obtidos de terceiros	Obtidos de parceiros institucionais de projetos de PD&I, de provedores de serviços/produtos/insumos agrícolas, entre outros.	<ul style="list-style-type: none"> ·Marco Civil da Internet; ·Lei Geral de Proteção de Dados Pessoais.

Fonte: Mendes, Bertin e Costa (2023)

A tipologia apresentada demonstra a integração de conjuntos de dados – pessoais, dados agrícolas e de PD&I – no contexto da pesquisa agrícola em órgãos públicos.

O que se depreende da tipologia de dados apresentada no Quadro 1 é que, enquanto o tratamento de dados pessoais é regulamentado pela LGPD, os dados não pessoais, sejam eles dados agrícolas ou dados de pesquisa, dispõem de proteção jurídica em situações específicas e de acordo com requisitos legais.



No que tange ao tratamento de dados pessoais, a aprovação da LGPD exigiu dos formuladores de políticas públicas um processo de calibração de instrumentos e adequação das práticas de governança de dados do Governo Federal. O artigo 50 da LGPD apresenta diretrizes para a implantação do programa de governança em privacidade a ser observado pela administração pública, quais sejam: as condições de organização para tratamento de dados pessoais, o os procedimentos, incluindo reclamações de titulares, as normas de segurança, os padrões técnicos, as obrigações dos envolvidos, as ações educativas, os mecanismos internos de supervisão e mitigação de riscos (Brasil, 2018).

A Embrapa, buscando sincronizar o respaldo legal ao tratamento e à proteção de dados com a continuidade do desenvolvimento de suas pesquisas, vem adotando medidas práticas para implantação do seu programa de governança em privacidade de dados.

Pode ser adotada a hipótese legal de realização de estudos por órgão de pesquisa, conforme o artigo 7º, inciso IV, da LGPD, garantindo, quando possível, a anonimização, quando presentes os requisitos: a) o controlador ou operador ser um órgão de pesquisa (como é o caso da Embrapa); b) os dados pessoais serem utilizados dentro do órgão para a finalidade estrita prevista na pesquisa; c) o órgão de pesquisa garantir que os dados pessoais não serão revelados em caso de divulgação dos resultados; d) o órgão assumir a responsabilidade pela segurança da informação e se comprometer a não transferir os dados a terceiros em circunstância alguma..

Como explicam Mendes *et al.* (2023), especificamente no escopo do pré-projeto piloto em Caconde CCD-AD/SemeAr, com base no Artigo 7º, inciso IV, da LGPD, o qual prevê a hipótese legal de tratamento de dados pessoais para a realização de estudos por órgão de pesquisa, garantindo a anonimização, foram tomadas as seguintes medidas para atendimento à legislação:

- a) apenas a Embrapa Agricultura Digital, como coordenadora do projeto, teve acesso aos dados brutos coletados (principalmente dados pessoais dos 232 agricultores);
- b) foram utilizados mecanismos para anonimização de dados, por meio da supressão de alguns campos de dados (tais como nome e e-mail do produtor rural);
- c) somente os dados agregados do projeto, após eliminados os dados pessoais, foram compartilhados com os parceiros institucionais para análise).

Tais medidas buscam atender os requisitos legais para o adequado enquadramento na hipótese legal de tratamento de dados pessoais por órgão de pesquisa.



Ações em Andamento

No âmbito CCD-AD/SemeAr Digital, a próxima etapa prevê a aplicação da tipologia de dados (conforme Quadro 1) para todos os conjuntos de dados a serem levantamentos nos demais municípios que receberão Distritos Agrotecnológicos (DATs), como relacionado na Tabela 1, com análise, caso a caso, de qual é a hipótese legal para tratamento dos dados.

As atividades de gestão de dados do Centro têm como referência o plano elaborado na ocasião de submissão da proposta à Fapesp e seguiu as "Diretrizes para Planos de Gestão de Dados (PGD) para Propostas de Centros da Fapesp⁷". Trata-se de um documento com seis itens, sendo: (i) Um sumário descritivo dos dados digitais a serem produzidos e gerenciados no Centro;

(ii) Gerenciamento dos objetos digitais armazenados – formatos, metadados e interoperabilidade;

(iii) Compartilhamento interno e externo; (iv) Gerenciamento dos dados – responsabilidades, preservação e segurança; (v) Restrições éticas, de confidencialidade, legais e outras e (vi) Repositórios. O plano foi elaborado conjuntamente pelas instituições participantes do Centro e é um instrumento para consulta e atualização se necessário.

A aderência aos princípios FAIR (*Findable, Accessible, Interoperable and Reusable*, do inglês Encontráveis, Acessíveis, Interoperáveis e Reusáveis), propostos por Wilkinson *et al.* (2016), embasam o plano de gestão de dados do Centro. Em especial, os dados são acompanhados por metadados que os descrevem e estão sendo adotados padrões de interoperabilidade para facilitar sua integração e análise, com priorização de formatos abertos.

Para viabilizar o gerenciamento de dados do Centro é necessária a utilização de tecnologias digitais que atuam em todo o ciclo de vida dos dados, desde a sua coleta até sua disponibilização para o usuário final do dado. Estas tecnologias englobam tanto infraestrutura computacional quanto ferramentas que são dotadas com funcionalidades que ajudam a garantir a aplicação dos princípios FAIR e de dispositivos legais relacionados aos dados.

Portanto, o planejamento da gestão de dados de um projeto precisa contemplar o uso de recursos computacionais como hardware, softwares de virtualização, sistemas especializados de storage, softwares específicos para armazenamento, processamento, visualização de dados, dentre outros. Há uma diversidade de ferramentas relacionadas ao

7

<https://fapesp.br/14974/diretrizes-para-planos-de-gestao-de-dados-pgd-para-propostas-de-centros>



gerenciamento de dados e a escolha de um conjunto de ferramentas pode variar de acordo com as necessidades e restrições do projeto.

O CCD-AD/SemeAr Digital adota como premissa a utilização de ferramentas open-source para compor as soluções a serem desenvolvidas. O Centro foi concebido com a participação de várias instituições com responsabilidades distintas relacionadas a coleta e uso de dados. No entanto, para o Centro alcançar seus objetivos, os dados coletados e/ou gerados pelas instituições precisam ser armazenados e disponibilizados. Por isso, essa próxima etapa do projeto consiste na elaboração de uma proposta para construção de uma infraestrutura (hardware e software) para armazenamento, controle de acesso, processamento e disponibilização de dados para consultas e integrações.

Considerações Finais

O relato do CCD-AD/ SemeAr demonstra como a análise da tipologia dos dados – pessoais, dados agrícola e de PD&I – contribui para gerar definir as estratégias jurídicas a serem adotadas na gestão de dados do projeto, fornecendo subsídios para análises em iniciativas semelhantes no âmbito da pesquisa agropecuária.

No âmbito da pesquisa pública agrícola há a necessidade de conciliação de uma estratégia para adequação às regras da LGPD, principalmente quanto às hipóteses legais que autorizam o tratamento de dados para fins de pesquisa por órgãos públicos de pesquisa.

Para instituições de pesquisa agrícola baseadas em ciência, como a Embrapa, cujos dados são insumos essenciais para seus projetos de PD&I, é imprescindível coordenar o avanço científico com a proteção de dados para conferir a segurança jurídica e o tratamento adequado aos dados obtidos e produzidos e/ou custodiados. Para tanto, faz-se necessário identificar e distinguir os diferentes conjuntos de dados que integram ou resultam da sua atividade-fim.

Nesse escopo, as tecnologias para a gestão dos dados são essenciais para que o CCD- AD/SemeAr Digital possa alcançar os resultados esperados. Portanto, a equipe do Centro está empenhada na construção de uma plataforma de dados baseada em soluções de código aberto, usando a infraestrutura computacional da Embrapa Agricultura Digital. Por meio desta plataforma será possível coletar, armazenar, processar e disponibilizar uma grande quantidade de dados de diferentes fontes e formatos que será gerada/produzida durante a execução do projeto.



Referências

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. **Lei Geral de Proteção de Dados Pessoais**. 2018. Disponível em: Acesso em: 07 mar. 2023.

FAPESP. **Centro da Fapesp sediado na Embrapa inicia ações de inclusão digital em dez municípios brasileiros**. Disponível em: <https://agencia.fapesp.br/centro-da-fapesp-sediado-na-embrapa-inicia-acoes-de-inclusao-digital-em-dez-municipios-brasileiros/50214>. **Agência Fapesp**. Acesso em: 08 fev. 2024.

MASSRUHÁ, S. M. F. S.; LEITE, M. A. de A.; MENDES, C. I. C. **Agricultura digital: tendências, oportunidades e desafios**. In: GIACOBBO, D. G.; FROTA, L. M. (coord.). O Papel do Agronegócio Brasileiro nas Novas Relações Econômicas Mundiais. Rio de Janeiro: Editora Synergia, 2021.

MENDES, C. I. C.; MARANHÃO, J. S. de A.; BERTIN, P. R. B.; MONDO, V. H. V.; PIRES, F. C. **Governança de dados para a pesquisa agrícola: segurança jurídica e autorregulação**. **Revista Cadernos de Ciência & Tecnologia**, Brasília, volume no. 40, edição no. 27209, 2023a. Disponível em: <https://seer.sct.embrapa.br/index.php/cct/article/view/27209>. Acesso em: 07 mar. 2024.

MONDO, V. ; LUCHIARI JR, A. ; BOLFE, E. L. ; SILVA, V. J. ; FREDO, C. E. . Projeto SemeAr: mapeamento das condições para viabilização da agricultura digital no município de Caconde, Estado de São Paulo. Textos para Discussão - Instituto de Economia Aplicada de SP, São Paulo, p. 1 - 25, 01 fev. 2022.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. **The FAIR Guiding Principles for scientific data management and stewardship**. *Sci. Data* 3, 160018. <https://doi.org/10.1038/sdata.2016.18>