

Rumo à rede de implantação GO FAIR ‘Agro’ Brasil: a experiência de uma organização de PD&I na implantação dos princípios FAIR

Debora Pignatari Drucker¹, Juliana Meireles Fortaleza², Patrícia Rocha Bello Bertin³, Isaque Vacari⁴ e Carla Geovana do Nascimento Macario⁵

1. Introdução

As Ciências Agrárias são multi e interdisciplinares devido à natureza integrativa dos conhecimentos sobre os meios físico, biológico, social e econômico. Além disso, a produção de alimentos, seu comércio e consumo estão intimamente ligados à saúde humana e do ambiente, sendo que a habilidade de conectar conhecimentos sobre esses temas, que muitas vezes são controversos, é essencial para a superação dos problemas complexos da agricultura (HILMIRE, 2016). Considerados a base do método científico (HEIDORN, 2008), dados são essenciais para a análise integrada das diferentes disciplinas que compõem a pesquisa agropecuária, devendo ser tratados como produtos valiosos da atividade de PD&I.

Nesse contexto, a quantidade de dados passíveis de análise vem aumentando exponencialmente, com a transição para uma ciência intensiva em dados em todas as áreas do conhecimento (HEY *et al.*, 2009). No âmbito das Ciências Agrárias, é digno de nota o uso de sensores remotos e proximais para monitorar variáveis de interesse em tempo real, sejam de solo, clima, atmosfera ou de organismos, bem como a produção de grandes volumes de dados com novas tecnologias como as genômicas ou o processamento de linguagem natural. Ao mesmo tempo, dados com características típicas de “cauda longa”, isto é, hete-

1 Doutora em Ambiente e Sociedade, Embrapa Agricultura Digital, debora.drucker@embrapa.br

2 Mestre em Fitotecnia, Secretaria de Desenvolvimento Institucional da Embrapa, juliana.fortaleza@embrapa.br

3 PhD em Gestão da Informação, Secretaria de Desenvolvimento Institucional da Embrapa, patricia.bertin@embrapa.br

4 Mestre em Ciência da Computação, Embrapa Agricultura Digital, isaque.vacari@embrapa.br

5 Doutora em Ciência da Computação, Embrapa Agricultura Digital, carla.macario@embrapa.br

rogêneos, diversos, pouco estruturados, difíceis e custosos para serem obtidos (HEIDORN, 2008; BORGMAN *et al.*, 2016), são acumulados há décadas - registros factuais valiosíssimos sobre os sistemas de produção e patrimônio ambiental, que estão sob o risco de serem perdidos caso não sejam adequadamente tratados e preservados.

Confiabilidade e reprodutibilidade são também pilares do método científico e, para assegurá-las, é fundamental que boas práticas de gestão de dados de pesquisa sejam adotadas. Nesse contexto, os princípios FAIR (*Findable, Accessible, Interoperable, Reusable*) – acrônimo em inglês para dados “localizáveis, acessíveis, interoperáveis e reutilizáveis” (WILKINSON *et al.*, 2016), vêm sendo amplamente adotados como norteadores da gestão de dados de pesquisa e viabilizadores de seu reúso. Aplicações e métricas para adoção concreta dos princípios FAIR têm sido desenvolvidas em todo o mundo, por meio da adoção de padrões, metadados, vocabulários controlados, ontologias e identificadores persistentes que proporcionam significado preciso aos dados e aos demais objetos a eles vinculados (HENNING *et al.*, 2019).

O reconhecimento pela comunidade científica da importância de tornar as práticas de gestão de dados aderentes aos princípios FAIR motivou o surgimento de iniciativas como a GO FAIR, que articula comunidades de prática a partir de um escritório central na Europa e Redes de Implementação (RIs) temáticas e regionais⁶. Uma dessas redes, intitulada *Food Systems*, tem por objetivo apoiar a implementação dos princípios FAIR nas ciências agroalimentares⁷. O Brasil integra a iniciativa com um escritório nacional⁸, ao qual estão associadas diversas RIs - entre elas, a rede GO FAIR Brasil Agro, que contribuirá para a adoção dos princípios FAIR no contexto de instituições produtoras de dados agropecuários (*Go-Change*); a realização de treinamentos e capacitações, em parceria com outras RIs nacionais (*Go-Train*); e a construção colaborativa e implementação de infraestrutura e padrões intercambiáveis (*Go-Build*).

O objetivo deste capítulo é relatar a experiência da Embrapa na incorporação dos princípios FAIR às diretrizes institucionais, aos processos e às práticas de governança e gestão de dados de pesquisa. A narrativa foi construída a partir de estudo de caso com abordagem exploratória, tendo como única unidade de análise a Em-

6 O portal GO-FAIR contém mais informações: <https://www.go-fair.org> Mais informações em: <https://www.rd-alliance.org/>.

7 A Rede de Implementação Food Systems está descrita em <https://www.go-fair.org/implementation-networks/overview/food-systems>.

8 Mais informações sobre o escritório brasileiro podem ser encontradas em: <https://www.go-fair-brasil.org/>.

brapa e os dados coletados por pesquisa documental. Com fundamentação teórica nas conceitualizações da Ciência Aberta, da e-Science e da gestão de dados de pesquisa sob a ótica dos princípios FAIR, espera-se que a análise ora apresentada sirva de base para a construção da rede de implementação GO FAIR Brasil Agro, de modo a beneficiar o sistema nacional de PD&I agropecuário com as lições aprendidas no caso Embrapa.

As próximas seções apresentam informações contextuais sobre a Gestão de Dados de Pesquisa (GDP) na Embrapa, explicitam o posicionamento da Empresa no ecossistema global de GDP, descrevem os normativos internos existentes na temática e relatam os resultados alcançados até o momento. Ao final, são descritos os desafios e perspectivas futuras para estabelecer as bases para tornar os dados de pesquisa da Embrapa cada vez mais em conformidade com os princípios FAIR.

2. Um pouco de contexto: a gestão de dados de pesquisa na Embrapa

A Embrapa – instituição pública de pesquisa agropecuária vinculada ao Ministério da Agricultura, Pecuária e Abastecimento (Mapa) – tem a missão de “viabilizar soluções de pesquisa, desenvolvimento e inovação para a sustentabilidade da agricultura, em benefício da sociedade brasileira” (EMBRAPA, 2020, p. 16). Estruturada em 43 centros de pesquisa distribuídos geograficamente em todo o País e com atuação no exterior, a Empresa gera um grande volume de dados nos diversos temas estratégicos da pesquisa agropecuária. Ciente do volume, da velocidade, da variedade e do valor dos dados de pesquisa produzidos no desenvolvimento de suas atividades, a Embrapa tem mobilizado esforços para governar e gerenciar adequadamente esses ativos em todo o seu ciclo de vida, a fim de torná-los localizáveis, acessíveis, interoperáveis e reutilizáveis.

Dentre esses esforços, pode-se destacar a execução, entre 2015 e 2017, do projeto “Governança de Dados e Informação para o Conhecimento na Embrapa: Desenvolvimento de Modelo e Plano de Implantação (GovIE)”, que teve por objetivo conceber, validar e propor um modelo sistêmico para a governança de dados e informação na Empresa. Como resultado do projeto, foram identificadas diversas medidas a serem tomadas para o aprimoramento da governança e gestão de dados de pesquisa na Empresa (Quadro 1), as quais encontram-se em curso de implantação.

Quadro 1. Medidas necessárias para aprimoramento da governança e gestão de dados de pesquisa na Embrapa, associadas aos pilares GO Build, GO Change e GO Train, das Redes de Implantação GO FAIR.

Medidas	Recomendações
Processuais	<ol style="list-style-type: none"> 1. Modelar e implementar os processos corporativos de gestão de dados de pesquisa e de publicação de dados abertos. <i>Change</i> 2. Modernizar o processo de avaliação de desempenho e recompensa dos empregados de forma com o objetivo de favorecer a cultura do compartilhamento e reúso de dados. <i>Change</i> 3. Desenvolver e implementar processos, competências, ferramentas e metodologias que possibilitem a interoperabilidade semântica entre sistemas de informação. <i>Build</i> 4. Possibilitar a agregação dos dados científicos às publicações e aos projetos que os geraram. <i>Build</i> 5. Adotar e operacionalizar o uso de licenças públicas para os ativos digitais. <i>Build</i>
Normativas	<ol style="list-style-type: none"> 6. Elaborar e publicar o Plano de Dados Abertos da Embrapa. <i>Change</i> 7. Elaborar, revisar, atualizar e implementar políticas e normas internas relativas à gestão de dados e informação gerados durante a pesquisa desenvolvida pela Empresa. <i>Change</i> 8. Incluir no Plano Diretor da Embrapa (PDE) diretrizes estratégicas e específicas relacionadas à gestão de dados e informações. <i>C</i> 9. Estabelecer e prover manutenção de um modelo corporativo de dados de pesquisa. <i>Build</i>
Cultura organizacional	<ol style="list-style-type: none"> 10. Garantir a inserção e participação ativa da Embrapa em fóruns e <i>networks</i> nacionais e internacionais nas temáticas de gestão de dados de pesquisa. <i>Change</i> 11. Revalorizar os profissionais da Ciência da Informação da Embrapa. <i>Change</i> 12. Implantar ações de capacitação e comunicação sobre gestão de dados de pesquisa. <i>Train</i> 13. Adotar a prática de elaboração de Planos de Gestão de Dados de Pesquisa. <i>Change</i>
Ferramentas, instrumentos e tecnologia	<ol style="list-style-type: none"> 14. Desenvolver e implantar infraestrutura tecnológica para a gestão de dados de pesquisa, por meio de plataformas consistentes e interoperáveis, para uso corporativo. <i>Build</i> 15. Adotar identificadores persistentes para dados, conjunto de dados e autores. <i>Build</i> 16. Garantir o alinhamento de planos de gestão de dados e arquiteturas da informação agropecuários com desenhos conceituais epistemologicamente sistematizados e globalmente utilizados. <i>Build</i> 17. Construir infraestrutura de dados abertos da Embrapa, interligada ao Portal Brasileiro de Dados Abertos. <i>Build</i> 18. Implantar ferramentas tecnológicas de gestão terminológica e alinhamento conceitual. <i>Build</i>
Estrutura, papéis e responsabilidades	<ol style="list-style-type: none"> 19. Definir instância organizacional responsável pela governança de dados de pesquisa. <i>Change</i>

Fonte: Elaborado pelos autores.

Nota-se que as mudanças culturais (*Go Change*), a realização de treinamentos e capacitações (*Go Train*) e a construção e implementação de infraestrutura e padrões intercambiáveis (*Go Build*) permeiam as 19 recomendações do Quadro 1. As medidas processuais contam com recomendações relacionadas com a promoção de mudanças (*Change*: 1 e 2) e com a construção implementação de infraestrutura (*Build*: 3, 4 e 5), assim como as recomendações de medidas normativas número 6, 7 e 8 enquadram-se na categoria *Change*, enquanto a de número 9 é categorizada como *Build*. Já as medidas de Cultura Organizacional contam com uma recomendação de treinamento (*Train*, 12) e as demais são categorizadas como *Change* (10, 11 e 13). Todas as recomendações da medida Ferramentas, instrumentos e tecnologia são de construção e implementação (14, 15, 16, 17 e 18: *Build*), enquanto a medida de Estrutura, papéis e responsabilidades é categorizada como *Change*. No total, nove

recomendações foram categorizadas como de construção e implementação (*Build*), oito como de mudança cultural (*Change*) e uma como treinamento e capacitação (*Train*). Uma das recomendações categorizadas como mudança cultural, a número 10 (“Garantir a inserção e participação ativa da Embrapa em fóruns e networks nacionais e internacionais nas temáticas de gestão de dados de pesquisa”), entendida como uma medida de “Cultura organizacional”, é detalhada na próxima seção.

3. Inserção da Embrapa no ecossistema global de gestão de dados de pesquisa

O fenômeno do Big Data e os novos paradigmas da e-Science e da Ciência Aberta têm promovido uma transformação no fazer científico, com reconfiguração de práticas, regras e comportamentos – sobretudo, na organização e gestão de dados de pesquisa (ALBAGLI *et al.*, 2015). Para melhor compreender e beneficiar-se dessa transformação, a Embrapa tem buscado intercambiar conhecimentos por meio da participação em iniciativas, redes, grupos e fóruns de discussão nacionais e internacionais que tratam sobre GDP. Como parte desses esforços, a Empresa coordenou, entre outubro de 2018 e julho de 2020, o Compromisso 3 do 4º Plano de Ação Nacional em Governo Aberto, conhecido como Compromisso pela Ciência Aberta, que visou “estabelecer mecanismos de governança de dados científicos para o avanço da Ciência Aberta no Brasil” (BRASIL, 2018a, 2018b). A execução do compromisso contou com a parceria de diversos órgãos governamentais e da sociedade civil, entre eles: Ministério da Ciência, Tecnologia e Inovações e Comunicações (MCTI), Fundação Oswaldo Cruz (Fiocruz), Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), Mapa, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), Rede Nacional de Ensino e Pesquisa (RNP), Universidade de Brasília (UnB) e Open Knowledge Brasil (OKBR). Dentre as entregas do compromisso, destacam-se: a elaboração de documentos orientadores para padrões de interoperabilidade e aferição do grau de maturidade de abertura dos dados de pesquisa; o desenvolvimento e implantação de repositórios pilotos institucionais; ações de sensibilização e de capacitação sobre o tema; articulações com editores científicos e agências de fomento; formação de rede interinstitucional de interesse na temática; e o diagnóstico da Ciência Aberta no mundo e no Brasil. A agenda do Governo Aberto mostrou-se um ambiente adequado para o fortalecimento e ampliação de parcerias entre os diversos atores do sistema científico nacional, favorecendo a convergência de ações e evitando esforços duplicados.

Destaca-se ainda a participação ngrupo de trabalho (GT) para implantação de uma Rede de Repositórios de Dados Científicos do Estado de São Paulo, criado

pela Fundação de Apoio à Pesquisa do Estado de São Paulo (Fapesp), contemplando dados e informação das universidades públicas do estado⁹. A Embrapa Informática Agropecuária, uma das unidades de pesquisa da Empresa, integra o GT, focando no compartilhamento de dados e informações da Empresa, os quais podem servir para testes de escalabilidade e de integração de dados agrícolas à rede de repositórios. Além disso, a Empresa compartilha conhecimento técnico e contribui para acelerar as atividades desenvolvidas pelo GT em assuntos como testes de escalabilidade, avaliação de ferramentas ou procedimentos de curadoria de dados, entre outros.

No contexto internacional, a Embrapa integra a iniciativa GODAN (do inglês, *Global Open Data for Agriculture and Nutrition*), que tem por objetivo promover esforços globais para a disponibilização, o acesso e o reúso de dados relevantes em agricultura e nutrição¹⁰. Como parte da rede de mais de 1.110 parceiros, a Empresa em atuado na tradução de materiais instrucionais sobre gestão de dados abertos em agricultura para a língua portuguesa. A Empresa tem contribuído ainda no âmbito da *Research Data Alliance* (RDA¹¹) (“Aliança para Dados de Pesquisa”, tradução livre) - uma iniciativa global lançada em 2013 para incentivo ao compartilhamento aberto e ao reúso de dados de pesquisa. A RDA conta com mais de 11.000 membros na atualidade, envolvendo produtores, usuários e administradores de dados que colaboram para o desenvolvimento de soluções e de boas práticas em tópicos relevantes para a GDP. Os especialistas se reúnem em grupos temáticos, sendo que um deles é o *Interest Group on Agricultural Data* (IGAD), coordenado por representantes da FAO, USDA e Embrapa. Outro grupo internacional com coordenação brasileira é o *Professionalising Data Stewardship*, o qual reúne profissionais de todos os continentes para alcançar uma visão comum quanto à profissionalização da administração de dados de pesquisa. A Rede de Observação de Dados para a Terra (DataOne¹²) é mais uma iniciativa que contribui para a elaboração da estratégia e para a adoção de melhores práticas de gestão de dados na Embrapa, sendo um programa estabelecido para compreender melhor a vida na Terra e o ambiente que a sustenta, conduzido pela comunidade e que fornece acesso a dados em vários repositórios de membros, promovendo as melhores práticas em gestão de dados por meio de recursos e de materiais educacionais. A interface entre a ciência e a política é o foco de atuação do *Intergovernmental Science-Policy Platform on Biodi-*

9 O portal para acesso aos dados está disponível em: <https://metabuscaador.uspdigital.usp.br/>.

10 Mais detalhes sobre a iniciativa Godan podem ser encontrados em: <https://www.godan.info/>

11 O portal da RDA contém informações sobre membros e grupos de trabalho: <https://www.rd-alliance.org/>.

12 Mais informações em: <https://www.dataone.org/>.

versity and Ecosystem Services (IPBES¹³), que conta com uma força tarefa para dados e conhecimento que, dentre outras ações, propôs a política de dados da plataforma e monitora sua adoção. Em conjunto, as iniciativas mencionadas contribuem para que melhores práticas de GDP sejam disseminadas e adotadas e que, assim, os dados possam ser reusados, conduzindo ao avanço da fronteira do conhecimento e subsidiando a tomada de decisão em diversas esferas. A participação da Embrapa nesses fóruns de discussão possibilita a troca de experiências, a atualização e revisão contínua de estratégias e atividades, como as que se descrevem a seguir.

4. Política de governança de dados, informação e conhecimento da Embrapa

Uma das recomendações GO *Change*, dentre as medidas normativas apresentadas no Quadro 1, é a de no. 7: “Elaborar, revisar, atualizar e implementar políticas e normas internas relativas à gestão de dados e informação gerados durante a pesquisa desenvolvida pela Empresa”. Uma ação fundamental nesse sentido foi a promulgação da Política de Governança de Dados, Informação e Conhecimento da Embrapa1 (Política GDIC) que institui princípios, diretrizes, atribuições e responsabilidades para “fortalecer os mecanismos de geração, organização, tratamento, acesso, preservação, recuperação, divulgação, compartilhamento e reuso dos ativos de informação da Embrapa” (EMBRAPA, 2019, p. 10).

Orientada pelos princípios da Constituição da República Federativa, a Declaração dos Direitos Humanos e os preceitos do movimento da Ciência Aberta, a política estabeleceu 17 princípios para a gestão de dados, informação e conhecimento na Empresa, a saber: (1) Dados, informação e conhecimento como ativos corporativos; (2) Alinhamento estratégico; (3) Desenvolvimento de capacidades e competências; (4) Infraestrutura federada; (5) Análise, inteligência e inovação baseada em dados; (6) Eficiência e economicidade; (7) Conformidade e mitigação de riscos; (8) Interoperabilidade; (9) Licenciamento; (10) Preservação e memória; (11) Privacidade, proteção e confiança; (12) Segurança; (13) Qualidade e integridade; (14) Especificidade epistemológica; (15) Aprendizagem organizacional, continuidade e retenção do conhecimento; (16) Abertura e transparência; (16.1) Acesso Aberto à informação científica; (16.2) Dados Abertos; (17) Monitoramento e responsabilidade na divulgação de informações relevantes.

Apesar de os princípios FAIR não estarem diretamente enunciados nas diretrizes e princípios que compõem a Política GDIC, o documento tem por premissa que “dados e informações bem organizados, documentados, acessíveis e verificados quanto a sua exatidão e validade são mais facilmente compartilháveis e reutilizá-

13 O portal do IPBES contém detalhes sobre a plataforma: <https://ipbes.net/>.

veis”, o que proporciona diversas vantagens à administração (EMBRAPA, 2019). Notadamente, a publicação que originalmente lançou os princípios FAIR (WILKINSON *et al.*, 2016) é uma das referências basilares da Política GDIC, de modo que estes se constituem em elementos transversais a todo o seu conteúdo. A título de exemplo, pode-se destacar a seguinte diretriz: “implantar e sustentar processos que garantam que dados e informações produzidos pela Empresa sejam confiáveis e facilmente recuperáveis, acessíveis, interoperáveis e reutilizáveis” (EMBRAPA, 2019).

Dentre os princípios da Política GDIC, o da ‘Interoperabilidade’ é aquele que melhor denota a necessidade de aplicação dos princípios FAIR. Para que seja contemplado na GDP, esse princípio requer o uso de ferramentas semânticas e padrões de dados e metadados amplamente estabelecidos e difundidos. Esse princípio é fortalecido por meio da diretriz prevista na perspectiva tecnológica que direciona para a inovação e o uso de tecnologias aliadas às tendências internacionais, por meio de serviços, como o compartilhamento e o reúso de dados com amplo atendimento à interoperabilidade.

Os princípios FAIR compreendem, assim, uma referência central para o desenvolvimento do Programa Corporativo de GDP, enunciado no item 8.1, da diretriz estratégica da Política GDIC: “implementar, sustentar e monitorar um Programa Corporativo de Gestão de Dados de Pesquisa e orientar o desenvolvimento de planos de gestão de dados no contexto de projetos de Pesquisa, Desenvolvimento e Inovação” (EMBRAPA, 2019, p. 13).

5. Programa Corporativo de Gestão de Dados de Pesquisa (GDP): Ações em Andamento

5.1 Diagnóstico sobre as práticas de gestão de dados

Ao longo da história da Embrapa, inúmeras práticas de gestão de dados e sistemas de informação foram criados, de acordo com as especificidades das diversas áreas temáticas de atuação dos diferentes centros de pesquisa da Empresa. Assim, uma das ações iniciais do Programa Corporativo de GDP foi realizar um levantamento para diagnosticá-las. Para tal, foi elaborado um questionário eletrônico que buscava retratar pontos importantes relacionados aos dados de pesquisa: caracterização dos dados, coleta e documentação, armazenamento, cópia de segurança, acessibilidade, compartilhamento e reúso, e repositórios de dados de pesquisa. O questionário foi respondido por 854 produtores de dados distribuídos entre 43 unidades de pesquisa com o objetivo de fundamentar e direcionar ações corporativas de aprimoramento da GDP da Empresa, de acordo com as melhores práticas e tendências internacionais de organização e publicação de dados.

5.2 Desenvolvimento e Implantação de Repositório de Dados Confiável e Identificadores

Persistentes

O contexto atual requer uma estratégia abrangente para estabelecimento das bases para tornar os dados da Embrapa FAIR, considerando a importância de fortalecer as soluções já existentes na Empresa, como o Sistema de Informação de Experimentos da Embrapa (SIExp¹⁴) e a Infraestrutura de Dados Espaciais da Embrapa – GeoInfo (DRUCKER *et al.*, 2017), bem como a necessidade de acomodar dados para os quais soluções adequadas ainda não haviam sido implementadas. A complexidade e a multidisciplinaridade das Ciências Agrárias demandam soluções tecnológicas que possibilitem a aderência aos dados FAIR e, ao mesmo tempo, que permitam a acomodação de dados de diversas disciplinas e seus modelos de representação e padrões, com vistas a obter um conjunto de descrição central, assim como o tratamento de especificidades via pequenas extensões deste conjunto, para que se viabilize a interoperabilidade dos repositórios de dados científicos de forma geral em diversos níveis. Considerando as melhores práticas mundialmente adotadas, optou-se pela implementação de um repositório de dados de pesquisa confiável como solução para organização, tratamento, preservação e publicação dos dados produzidos pela Embrapa.

O Repositório de Dados da Embrapa, denominado Redape, foi desenvolvido e implantado a partir do software web de código aberto *Dataverse*¹⁵. Adicionalmente, adquiriu-se e viabilizou-se uma infraestrutura computacional própria para o armazenamento dos dados de pesquisa, e designou-se uma equipe responsável pela administração técnica do repositório Redape, hospedado no *Data Center Científico* da Embrapa, sediado na cidade de Campinas, SP, que abriga características essenciais de segurança da informação, tais como: acesso restrito aos computadores, servidores e discos de armazenamento dos dados, bem como defesa contra ataques ao repositório, prevenção do acesso de indivíduos não autorizados a dados restritos, dentre outros.

O Redape suporta o serviço de atribuição de identificadores persistentes, um dos requisitos fundamentais para tornar os produtos de dados disponíveis para a comunidade científica. Um identificador persistente (do inglês *persistent identifier* (PID)) possibilita a identificação unívoca de um conteúdo digital e destina-se a ser uma maneira permanente de identificar e de acessar esse recurso específico. O tipo de PID mais amplamente adotado no meio científico é o DOI (do inglês, *digital object identifier*), o qual gera um link persistente que aponta para o repositório

14 Mais informações em: <http://www.embrapa.br/siexp>.

15 Disponível em: <https://dataverse.org>.

ou para outra localidade digital ao incluir a URL nos metadados. Isso fornece um sistema para a identificação persistente e acionável, bem como para o intercâmbio interoperável. O Redape está sendo validado pela implantação de estudos de caso piloto representativos da pesquisa agropecuária.

5.3 Representação do Conhecimento

De acordo com Meadow *et al.* (2007), a recuperação da informação é um processo de comunicação entre os autores e criadores de registros e os leitores. Esse processo depende de um controle adequado da linguagem (código) entre emissor e receptor e entre os documentos e as requisições dos usuários (JANAITE NETO; FERNEDA, 2016). A construção dos vocabulários controlados visa às atividades de indexação, armazenamento e recuperação da informação, representando conceitos significativos de algum domínio do conhecimento e, se possível, estabelecendo relações entre tipos e até subtipos do domínio (CHANDRASEKARAN *et al.* 1999; CINTRA, 2002; JACOB, 2003; FUJITA, 2004).

De acordo com a Plataforma Lattes do CNPq, as Ciências Agrárias podem ser subdivididas nas seguintes subáreas: Agronomia, Recursos Florestais e Engenharia Florestal, Engenharia Agrícola, Zootecnia, Medicina Veterinária, Recursos Pesqueiros e Engenharia de Pesca, Ciência e Tecnologia de Alimentos (CNPQ, 2021). Essa diversidade de subáreas contribui para o elevado número de termos que podem ser utilizados para indexação, armazenamento e recuperação da informação. O Agrovoc Multilingual Thesaurus, por exemplo, contabiliza 33.388 termos principais e 2.254 termos alternativos, incluindo alimentos, nutrição, agricultura, silvicultura, pesca, nomes científicos e comuns de animais e plantas, meio ambiente, noções biológicas, técnicas de cultivo de plantas, entre outros. O Agrotermos - vocabulário controlado construído por um grupo de trabalho permanente da Embrapa - reuniu aproximadamente 245 mil termos pertinentes ao domínio do conhecimento agropecuário a partir da reunião das terminologias em língua portuguesa encontradas em tesouros agrícolas nacionais e internacionais. A expectativa é de expansão do Agrotermos para um espaço conceitual do conhecimento agropecuário brasileiro e promover uma melhor interoperabilidade entre os sistemas de informação internos e externos.

5.4 Plano de Gestão de Dados

O planejamento de GDP é uma das etapas mais importantes no processo de desenvolvimento da pesquisa, pois é nesse momento em que são discutidos como os dados serão tratados durante todo o seu ciclo de vida e como garantir que eles estejam livremente disponíveis - respeitando a privacidade - e passíveis de serem

reutilizados, sob condições e licenças específicas claramente definidas, e que possam ser devidamente citados e referenciados. O Plano de Gestão de Dados (PGD) é, portanto, uma ferramenta fundamental para que boas práticas de gestão de dados sejam aplicadas durante o desenvolvimento da pesquisa até a publicação dos dados. Cientes da importância do PGD, agências de fomento dos Estados Unidos, União Europeia, Holanda, Reino Unido, Austrália, Canadá e Finlândia têm exigido que os projetos de pesquisa venham acompanhados de um PGD alinhado com os princípios FAIR (AVENTURIER, 2017). No Brasil, a Fapesp foi a primeira agência de fomento nacional a anunciar, em 2017, a obrigatoriedade do PGD para as solicitações de financiamento de projetos de pesquisa. As instituições de pesquisa devem inserir em seu processo de desenvolvimento da pesquisa o PGD, não apenas para a obtenção de financiamento em atendimento às agências de fomento, mas também para garantir que os dados sejam devidamente gerenciados. A Embrapa está implantando ações para Gestão de Dados de Pesquisa de seus projetos. Uma delas é exigir um PGD quando da sua submissão no sistema gestor da pesquisa da empresa, tendo o proponente de responder questões como tipo de dados, repositório a ser usado e acesso e compartilhamento. A Embrapa Informática Agropecuária adota essa prática desde 2018. Corporativamente, espera-se que, a partir de 2021, essa prática seja aplicada por todas as unidades de pesquisa, em conformidade com a diretriz estratégica da PGDIC.

6. Considerações Finais

Esse trabalho descreveu esforços que vêm sendo realizados na Embrapa para implantar a gestão de dados de pesquisa fundamentada nos princípios norteadores FAIR e procurou enquadrar as medidas mapeadas de acordo com os pilares da iniciativa GO FAIR. A seção que descreveu a inserção da Embrapa no ecossistema global de GDP demonstrou que há inúmeras frentes de atuação nessa temática, que atende uma das recomendações categorizadas como mudança cultural (*Go-Change*), a número 10: “Garantir a inserção e participação ativa da Embrapa em fóruns e networks nacionais e internacionais nas temáticas de gestão de dados de pesquisa”. Os resultados obtidos até o momento são notórios e têm potencial de ser multiplicados e expandidos nos próximos anos, a partir do estreitamento de relações e laços com instituições parceiras. Essa é uma característica que embasa o pilar *Go-Change*, uma vez que a construção de comunidades fundamenta a iniciativa GO FAIR.

Outra ação de extrema importância para incitar a adoção de práticas aderentes aos princípios FAIR e sustentar a mudança cultural necessária para promover os pilares do movimento GO FAIR foi a promulgação da Política de Governança de Da-

dos, Informação e Conhecimento da Embrapa, aqui descrita. Para incorporar esses princípios no dia a dia da organização, a empresa está implantando o Programa Corporativo de Governança de Dados de Pesquisa, que garantirá os meios, serviços e ferramentas necessários para que os dados produzidos pelos projetos sejam facilmente localizáveis, acessíveis, interoperáveis e reutilizáveis. Dentre as ações que estão em curso para a implantação desse programa corporativo, estão: a realização de um diagnóstico sobre as práticas de gestão de dados; a implantação de um repositório de dados confiável, com a atribuição de identificadores persistentes e que viabiliza a descoberta de dados, até então, desconectados; o desenvolvimento de ações para viabilizar a representação do conhecimento das Ciências Agrárias e o estabelecimento da prática de elaboração de planos de gestão de dados em projetos de pesquisa desenvolvidos pela empresa.

Vale destacar, como desafio e perspectiva futura, a necessidade de mapeamento e descrição formal do processo de gestão de dados de pesquisa, para que se alcance um entendimento mais aprofundado das práticas existentes, de modo que serviços e soluções a serem ofertados seja aderente à cultura organizacional e epistemológica. Os desafios, entretanto, não são elementares, tendo em vista a característica multidisciplinar das Ciências Agrárias e a necessidade de envolver diferentes atores e competências para a consecução de uma modelagem dessa complexidade. Em associação, ações de treinamento e capacitação são de extrema importância para o sucesso da implantação do Programa Corporativo de GDP.

Outro desafio a ser enfrentado é contemplar adequadamente o princípio da interoperabilidade pela adoção de padrões de dados e de metadados - requisito fundamental para permitir que os dados sejam corretamente interpretados e, assim, viabilizar seu reuso. Novamente, considerando a grande diversidade e heterogeneidade dos dados gerados e analisados no contexto das Ciências Agrárias, trata-se de um desafio que requer a participação dos diversos atores das diferentes disciplinas que compõem a pesquisa agropecuária. A atribuição de licenças que esclareçam os termos de uso dos dados é também condição fundamental.

Como demonstrado no caso da Embrapa, a aderência aos princípios FAIR é de fundamental importância para a interoperabilidade tecnológica e semântica de dados no contexto da agricultura. A estratégia aqui apresentada parte da premissa de que os dados são produtos valiosos da atividade de pesquisa, e denota a transição para uma práxis na qual o reuso de dados da pesquisa agropecuária a partir do paradigma dos princípios FAIR é encorajado. Nesse sentido, a construção da GO FAIR Brasil Agro é de fundamental importância para que o trabalho colaborativo beneficie mutuamente as comunidades de prática de gestão de dados afeitas à temática da gestão de dados. Assim, a promoção do reuso de dados de pesquisa agropecuários

contribuirá com a solução de problemas da sociedade brasileira e mundial, considerando a relevância do País no contexto dos sistemas alimentares.

Como desafios e perspectivas futuras, a modelagem de processos de gestão de dados foi iniciada na Empresa e envolve inúmeros atores de diferentes competências e áreas de atuação, trazendo complexidade para seu desenho e melhoria. Além disso, ações de treinamento e educação são de extrema importância para o sucesso da implantação do Programa Corporativo de GDP. Outro desafio a ser enfrentado é contemplar adequadamente a interoperabilidade pela adoção de padrões de dados e de metadados, requisito fundamental para permitir que os dados sejam corretamente interpretados e, assim, viabilizar seu reuso. Novamente, considerando a grande diversidade e heterogeneidade dos dados gerados e analisados no contexto das Ciências Agrárias, trata-se de um desafio que requer a participação dos diversos atores das diferentes disciplinas que compõem a pesquisa agropecuária. Além disso, a atribuição de licenças de uso para os dados é também pré-requisito para que os tipos de reuso permitidos sejam conhecidos.

Por fim, a elaboração de estratégias de monitoramento do Programa Corporativo de GDP com vistas à aderência aos princípios FAIR, permitindo a incorporação de melhorias, é uma perspectiva fundamental para assegurar seu sucesso. Uma referência de base é o trabalho realizado no âmbito do Marco 9 do Compromisso pela Ciência Aberta, intitulado “Proposição de conjunto de indicadores para aferição da maturidade em Ciência Aberta”. Apesar de os princípios FAIR não necessariamente implicarem em abertura de dados, o conjunto de indicadores para aferição do grau de maturidade de abertura de dados científicos fornece critérios objetivos que podem ser utilizados também para mensuração de sucesso nos Eixos Governança, Cultura Organizacional, Gestão de Dados Científicos e Infraestrutura Tecnológica (Fortaleza *et al.* 2020). Indicadores mais específicos poderão ser desenvolvidos, tais como: métricas de catalogação de dados de pesquisa em repositórios institucionais; quantitativo de acesso aos recursos disponibilizados; descrição e implantação de processos; adoção de licenças de uso; e estabelecimento de mecanismos de recompensa ao compartilhamento e reuso de dados.

As ações aqui apresentadas sinalizam alguns passos de um caminho para promover a transição para a lógica de que dados são produtos valiosos das atividades de pesquisa e na qual o reuso de dados da pesquisa agropecuária a partir do paradigma dos princípios FAIR é encorajado. Nesse sentido, a construção da GO FAIR Brasil Agro é de fundamental importância para que o trabalho colaborativo beneficie mutuamente as comunidades de prática de gestão de dados afeitas à temática da gestão de dados. Assim, a promoção do reuso de dados de pesquisa agropecuários contribuirá com a solução de problemas da sociedade brasileira e mundial, consi-

derando a relevância do país no contexto dos sistemas alimentares.

7. Referências

- ALBAGLI, Sarita; MACIEL, Maria Lucia; ABDO, Alexandre Hannud (org.)
Ciência aberta, questões abertas. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015.
- AVENTURIER, Pascal. Plano de Gestão de Dados: uma introdução. *In:*
 AVENTURIER, Pascal. **A publicação científica:** blog do Pascal Aventurier
 sobre as publicações científicas e os dados de pesquisa. Avignon, 17 maio 2017.
 Disponível em: <https://publicient.hypotheses.org/1660>. Acesso em: 30 out.
 2020.
- BORGMAN, Christine L. *et al.* data management in the long tail: science,
 software, and service. **International Journal of Digital Curation**, Edinburg, v.
 1, n. 1, p. 128- 148, Oct. 2016. DOI 10.2218/ijdc.v1i1i.428.
- BRASIL. Controladoria Geral da União. Inovação e governo aberto na ciência
 - monitoramento e execução: compromisso 3. Estabelecer mecanismos de
 governança de dados científicos para o avanço da ciência aberta no Brasil.
 2018a. Disponível em: [https://www.gov.br/cgu/pt-br/governo-aberto/a-ogp/
 planos-de-acao/40-plano-de-acao-brasileiro/compromisso-3-docs/inovacao-
 e-governo-aberto-na-ciencia-monitoramento-e-execucao](https://www.gov.br/cgu/pt-br/governo-aberto/a-ogp/planos-de-acao/40-plano-de-acao-brasileiro/compromisso-3-docs/inovacao-e-governo-aberto-na-ciencia-monitoramento-e-execucao). Acesso em: 3 mar.
 2021.
- BRASIL. Ministério da Transparência e Controladoria-Geral da União. Secretaria
 de Transparência e Prevenção da Corrupção. Diretoria de Transparência e
 Controle Social. Coordenação-Geral de Governo Aberto e Transparência. **4º
 Plano de Ação Nacional em Governo Aberto.** Brasília, DF, 2018b. Disponível
 em: [http://governoaberto.cgu.gov.br/esta-aberta-consulta-publica-do-40-
 plano-de-acao-nacional-para-governo-aberto/40-plano-de-acao-nacional_
 portugues.pdf](http://governoaberto.cgu.gov.br/esta-aberta-consulta-publica-do-40-plano-de-acao-nacional-para-governo-aberto/40-plano-de-acao-nacional_portugues.pdf). Acesso em: 6 out. 2020.
- CHANDRASEKARAN, B.; JOSEPHSON, John R.; BENJAMINS, V. Richard.
 What are ontologies, and why do we need them? **IEEE Intelligent Systems**, v.
 14, n. 1, p. 20-26, Feb. 1999. DOI 10.1109/5254.747902.
- CINTRA, Anna Maria Marques; TALAMO, Maria de Fatima Gonçalves Moreira;
 LARA, Marilda Ginez Lopes de; KOBASHI, Nair Yumiko (Org.). **Para
 entender as linguagens documentárias.** 2. ed. São Paulo: Polis, 2002.
- CNPQ. **Diretório dos grupos de pesquisa no Brasil:** áreas do conhecimento
 –Ciências Agrárias. Brasília, DF, [2021]. Disponível em: [http://lattes.cnpq.br/
 web/dgp/ciencias-agrarias](http://lattes.cnpq.br/web/dgp/ciencias-agrarias). Acesso em: 3 mar. 2021.
- DRUCKER, Debora Pignatari *et al.* GeoInfo - infraestrutura de dados espaciais
 abertos para a pesquisa agropecuária. **RECIIS: Revista Eletrônica de**

- Comunicação, Informação & Inovação em Saúde, Rio de Janeiro, v. 11, p. 1-17, 2017. Suplemento. Disponível em: <https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1083246/1/GeoInfo.pdf>. Acesso em: 1 mar. 2021.
- EMBRAPA. Política de Governança de Dados, Informação e Conhecimento da Embrapa. **Boletim de Comunicação Administrativa**, Brasília, DF, ano 45, n. 16, p. 1-19, 5 abril. 2019. Disponível em: <https://www.embrapa.br/politica-de-governanca-de-dados-informacao-e-conhecimento>. Acesso em: 3 mar. 2021.
- EMBRAPA. Regimento das Secretarias da Embrapa. **Boletim de Comunicação Administrativa**, [S.l.], ano 44, n. 8, p. 1-26, 1 fev. 2018. Disponível em: <https://www.embrapa.br/documents/10180/1546282/Regimento+das+Secretarias+da+Embrapa/d629c401-d2e6-fd8d-5154-cbbaa1e3313>. Acesso em: 30 out. 2020.
- EMBRAPA. VII **Plano Diretor da Embrapa 2020–2030**. Brasília, DF: Embrapa, 2020. 31 p. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/217274/1/VII-PDE-2020.pdf>. Acesso em: 3 mar. 2021.
- FORTALEZA, J.M.; BERTIN, P. R. B.; DRUCKER, D.P.; ASSIS, T.B.; COSTA, M.P. Conjunto de indicadores para aferição do grau de maturidade de abertura dos dados científicos. Brasília, DF: Embrapa, CNPq, OKBR, Ibict, MCTI, 2020. 14 p.
- FUJITA, Mariângela Spotti Lopes. A leitura documentária na perspectiva de suas variáveis: leitor-texto-contexto. **DataGramZero: Revista de Ciência da Informação**, Rio de Janeiro, v. 5, n. 4, ago. 2004. <https://www.brapci.inf.br/index.php/article/download/7646>. Acesso em: 30 out. 2020.
- HEIDORN, P. Bryan. Shedding light on the dark data in the long tail of science. **Library Trends**, Baltimore, v. 57 n. 2, p. 280-299, Fall 2008. DOI 10.1353/lib.o.0036.
- HENNING, Patrícia Corrêa *et al.* Desmistificando os princípios fair: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados FAIR. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, Paraíba, v. 14, n. 3, p. 175-192, 2019. DOI 10.22478/ufpb.1981-0695.2019v14n3.46969.
- HEY, Tony; TANSLEY, Stewart; TOLLE, Kristin, (ed.). **The fourth paradigm: data-intensive scientific discovery**. Redmond: Microsoft Research, 2009. Disponível em: https://www.microsoft.com/en-us/research/wp-content/uploads/2009/10/Fourth_Paradigm.pdf. Acesso em 1 set 2020.
- HILIMIRE, Kathleen. Theory and practice of an interdisciplinary food systems curriculum. **NACTA Journal**, Rupert, v. 60, n. 2, p. 227-233, June 2016. DOI 10.2307/nactajournal.60.2.227.
- JACOB, Ellin K. Ontologies and the semantic web. **Bulletin of the American Society for Information Science and Technology**, Washington, DC, v. 29, n. 4,

p. 19-22, Apr./May 2003. DOI 10.1002/bult.283.

JANAITE NETO, Jorge; FERNEDA, Edberto. Ontologia como recurso de padronização terminológica. **Informação em Pauta**, Fortaleza, v. 1, n. 1, p. 30-45, jan./jun. 2016. DOI 10.32810/2525-3468.ip.v1i1.2016.2967.

MEADOW, Charles T.; BOYCE, Bert R.; KRAFT, Donald H.; BARRY, Carol. **Text information retrieval system**. 3rd ed. Amsterdam: Elsevier, 2007. Disponível em: https://diglibrary.weebly.com/uploads/1/8/5/1/18511482/text_info_retrieval_system.pdf Acesso em: 30 out. 2020.

WILKINSON, Mark D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, London, v. 3, p. 1-9, 2016. DOI 10.1038/sdata.2016.18.

► Como citar com o DOI individual

DRUCKER, Debora Pignatari; FORTALEZA, Juliana Meireles; BERTIN, Patrícia Rocha Bello; VACARI, Isaque; MACARIO, Carla Geovana do Nascimento. Rumo à rede de implantação GO FAIR 'Agro' Brasil: a experiência de uma organização de PD&I na implantação dos princípios FAIR. *In*: SALES, Luana Farias; VEIGA, Viviane dos Santos; HENNING, Patrícia; SAYÃO, Luís Fernando (org.). **Princípios FAIR aplicados à gestão de dados de pesquisa**. Rio de Janeiro: Ibict, 2021. p. 47-62. DOI: 10.22477/9786589167242.cap4