

**Empresa Brasileira de Pesquisa Agropecuária
Embrapa Agricultura Digital
Ministério da Agricultura, Pecuária e Abastecimento**

DOCUMENTOS 183

Planos de Gestão de Dados Acionáveis por Máquina Alinhados aos Princípios FAIR para o Laboratório Multiusuário de Bioinformática da Embrapa

*Márcia Izabel Fugisawa Souza
Marcos Cezar Visoli
Tércia Zavaglia Torres
Paula Regina Kuser Falcão
Antonio Nhani Junior
Poliana Fernanda Giachetto
Felipe Rodrigues da Silva
Leandro Carrijo Cintra
Carla Cristiane Osawa
Alessandra Rodrigues da Silva
Luiz Antonio Falaguasta Barbosa*

Embrapa Agricultura Digital
Campinas, SP
2022

Embrapa Agricultura Digital
Av. Dr. André Tosello, nº 209 - Campus da Unicamp,
Barão Geraldo - Campinas, SP
CEP. 13083-886 - Fone: +55 (19) 3211-5700

www.embrapa.br/agricultura-digital
www.embrapa.br/fale-conosco/sac

Comitê Local de Publicações
da Unidade Responsável

Presidente
Carla Geovana do Nascimento Macário

Secretário-Executivo
Maria Fernanda Moura

Membros
*Adriana Farah Gonzalez, Alexandre de Castro,
Carla Cristiane Osawa, Debora Pignatari
Drucker, Ivan Mazoni, João Camargo Neto,
João Francisco Gonçalves Antunes, Magda
Cruciol*

Revisão de texto
Adriana Farah Gonzalez

Normalização bibliográfica
Carla Cristiane Osawa

Projeto gráfico da coleção
Carlos Eduardo Felice Barbeiro

Editoração eletrônica
Letícia Mathias do Amaral Campos

Foto da capa
Letícia Mathias do Amaral Campos

1ª edição
Publicação digital (2022): PDF

Todos os direitos reservados.

A reprodução não autorizada desta publicação, no todo ou em parte,
constitui violação dos direitos autorais (Lei nº 9.610).

Dados Internacionais de Catalogação na Publicação (CIP)

Embrapa Agricultura Digital

Planos de gestão de dados acionáveis por máquina alinhados aos princípios FAIR
para o Laboratório Multiusuário de Bioinformática da Embrapa / Márcia Izabel
Fugisawa Souza ... [et al.]. – Campinas : Embrapa Agricultura Digital, 2022.
PDF (68 p.) : il. color. - (Documentos / Embrapa Agricultura Digital, ISSN 2764-
2488 ; 183).

1. Gestão de dados de pesquisa. 2. Dados FAIR. 3. Bioinformática. 4. Labora-
tório Multiusuário de Bioinformática da Embrapa. I. Souza, Márcia Izabel Fugisawa.
II. Embrapa Agricultura Digital. III. Série.

CDD (21. ed.) 005.7406

Autores

Márcia Izabel Fugisawa Souza

Bacharel em Biblioteconomia, doutora em Educação, analista da Embrapa Agricultura Digital, Campinas, SP

Marcos Cezar Visoli

Cientista da Computação, mestre em Informatique e Systèmes Spécialité Recherche Modèles, Systèmes, Imagerie, Robotique, pesquisador da Embrapa Agricultura Digital, Campinas, SP

Paula Regina Kuser Falcão

Física, doutora em Cristalografia de Proteínas, pesquisadora da Embrapa Agricultura Digital, Campinas, SP

Antonio Nhani Junior

Biólogo, doutor em Bioquímica, pesquisador da Embrapa Agricultura Digital, Campinas, SP

Poliana Fernanda Giachetto

Zootecnista, doutora em Produção Animal, pesquisadora da Embrapa Agricultura Digital, Campinas, SP

Felipe Rodrigues da Silva

Biólogo, doutor em Genética e Biologia Molecular, pesquisador da Embrapa Agricultura Digital, Campinas, SP

Leandro Carrijo Cintra

Cientista da Computação, doutor em Bioinformática, analista da Embrapa Agricultura Digital, Campinas, SP

Carla Cristiane Osawa

Bacharel em Química, Química Tecnológica e em Biblioteconomia, mestre em Química, analista da Embrapa Agricultura Digital, Campinas, SP

Alessandra Rodrigues da Silva

Bacharel em Biblioteconomia, doutora em Ciência da Informação, analista da Embrapa na Secretaria-Geral, Brasília, DF

Luiz Antonio Falaguasta Barbosa

Cientista da Computação, mestre em Ciência da Computação, analista da Embrapa Agricultura Digital, Campinas, SP

Apresentação

As atividades científicas têm passado por profundas mudanças na sua forma de produzir conhecimento. O modo de fazer ciência, ao longo dos tempos, vem se modificando. Os avanços constantes na produção de conhecimento científico têm propulsionado o desenvolvimento de tecnologias, metodologias, produtos e serviços, voltados ao atendimento de amplos interesses da sociedade.

A passagem de um paradigma da ciência baseada em hipótese para um paradigma de ciência orientada a dados traz implicações substanciais no fazer científico em instituições de Pesquisa, Desenvolvimento e Inovação (PD&I), como a Empresa Brasileira de Pesquisa Agropecuária (Embrapa). Atenta a essas transformações, a Embrapa tem priorizado ações voltadas à gestão de dados, informação e conhecimento, a partir do estabelecimento de mecanismos para a geração, organização, tratamento, acesso, preservação, recuperação, divulgação, compartilhamento e reúso dos seus ativos de informação.

A adoção de planos de gestão de dados é medida essencial à implementação de melhorias aos processos de gestão de dados no Laboratório Multiusuário de Bioinformática. Tais planos são considerados instrumentos de boas práticas de pesquisa, isso porque reúnem as informações sobre procedência, documentação e qualidade, armazenamento, aspectos legais e éticos e códigos de conduta, compartilhamento, preservação, responsabilidades e recursos financeiros para a gestão de dados.

A reunião dessas informações na forma de um plano de gestão de dados cria as condições para que os dados de pesquisa possam ser acessados, preservados e reusados em novas atividades científicas. Para que essas condições

ocorram, os dados assim como os planos de gestão de dados precisam atender aos princípios FAIR, os quais orientam para a descoberta, a acessibilidade, a interoperabilidade e a reusabilidade de dados de pesquisa.

Esta publicação traz uma contribuição efetiva para comunidade de dados ômicos da Embrapa ao apresentar modelos de planos de gestão de dados acionáveis por máquina, alinhados aos princípios FAIR, desenvolvidos pelo Laboratório Multiusuário de Bioinformática da Embrapa.

Stanley Robson Medeiros de Oliveira
Chefe-geral da Embrapa Agricultura Digital

Sumário

Introdução.....	8
Objetivos e Procedimentos Metodológicos	11
Criação de Planos de Gestão Dados Acionáveis por Máquina para o Laboratório Multiusuário de Bioinformática da Embrapa.....	13
Resultados.....	22
Conclusões.....	25
Referências	28
Apêndices.....	31
Lista de siglas.....	67

Introdução

O Laboratório Multiusuário de Bioinformática da Embrapa (LMB) é uma estrutura computacional de vanguarda, criada em 2010 nas dependências da Embrapa Agricultura Digital, localizada na cidade de Campinas, SP, para promover os avanços do conhecimento nas áreas de recursos genéticos, biotecnologia e melhoramento genético.

Com a missão de viabilizar soluções de Bioinformática para projetos de pesquisa, desenvolvimento e inovação na Empresa Brasileira de Pesquisa Agropecuária (Embrapa) em um ambiente colaborativo, o LMB compartilha equipamentos de alto custo, otimiza recursos econômicos e humanos, incorpora e disponibiliza competências na área de computação de alto desempenho, para a Embrapa e para outras instituições de ensino, pesquisa, desenvolvimento e inovação do Brasil e do mundo, relacionadas à agropecuária. (Embrapa Agricultura Digital, 2022).

No LMB a implementação de processos de gestão de dados é essencial à otimização do processamento de alto desempenho, pois além de facilitar a execução de análises que visam a descoberta, também favorece o compartilhamento e o reúso dos dados. Entretanto, para que a gestão de dados ônicos seja eficiente, faz-se necessária a adoção de planos de gestão de dados, que são documentos formais capazes de fornecer informações sobre a origem, a descrição e coleção de dados ou de reúso dos dados existentes, a documentação e a qualidade dos dados, o armazenamento e backup durante o processo de pesquisa, os requisitos legais e éticos e códigos de conduta, o compartilhamento de dados e preservação e as responsabilidades e recursos financeiros para a gestão de dados de pesquisa.

Os planos de gestão de dados são considerados instrumentos de boas práticas, estas entendidas “[...] como condutas adotadas para uma maior divulgação, disseminação e desenvolvimento da ciência, visando a sua abertura à sociedade” (Guandalini et al., 2019, p. 3). Ainda discorrendo sobre o assunto, as autoras afirmam que:

“[...] a elaboração de um plano de gestão de dados pode ser caracterizada como um procedimento inicial para o desenvolvimento da pesquisa científica, documento que propõe ao pes-

quisador gerenciar os dados brutos de sua pesquisa, valorizando o compartilhamento em conjunto com a abertura dos dados e seu reuso em benefício aos pesquisadores e à sociedade.” (Guandalini et al., 2019, p. 3).

Por sua vez, Veiga et al. (2019, p. 277) consideram que um plano de gestão de dados possibilita “[...] que os dados sejam acessados, preservados e reusados em novas pesquisas”. Para que isso ocorra, tanto os dados como os planos de gestão de dados precisam atender aos princípios FAIR (acrônimo de *Findable, Accessible, Interoperable, Reusable*), um conjunto de elementos norteadores para a descoberta, o acesso, a interoperabilidade e o reuso de dados de pesquisa.

Paralelamente, apregoa-se que planos de gestão de dados devem ser acionáveis por máquina para que possibilitem a integração automática entre sistemas, de modo que dados e informações possam ser compartilhados e reutilizados em outros projetos de pesquisa (Miksa et al., 2019). Ademais, um plano de gestão de dados acionável por máquina traz facilidade no preenchimento do formulário, além de otimizar o tempo do pesquisador e racionalizar o trabalho.

O LMB é uma unidade de pesquisa cuja especificidade dos dados científicos requer um modelo de plano de gestão de dados que reflita as interações e fluxos informacionais com potencial de integração com outros sistemas e serviços, e passíveis de automatização. A nítida compreensão desta particularidade por parte dos integrantes do LMB impulsionou a busca por modelos de gestão de dados que fossem capazes de atender aos requisitos técnicos de forma dinâmica, e não meramente burocrática e sem integração, a exemplo de ferramentas tradicionais como DMPTool¹ e DMPonline², e outras iniciativas simples baseadas em formulários de processadores de texto, como o Word e LibreOffice.

Neste documento são relatadas as atividades de construção de modelos de conhecimento e de criação de planos de gestão de dados acionáveis por máquina, realizadas no âmbito de uma ação gerencial local conduzida em

¹ Disponível em: <<https://dmptool.org/>>.

² Disponível em: <<https://dmponline.dcc.ac.uk/>>.

2021, liderada pelo Laboratório Multiusuário de Bioinformática da Embrapa (LMB), em parceria com o Grupo de Pesquisa em Computação Científica, Engenharia da Informação e Automação (GCIA), da Embrapa Agricultura Digital.

A referida ação gerencial local, que se denominou “Plano de Gestão de Dados Acionável por Máquina (PGDam) Alinhado aos Princípios FAIR”, teve origem em outra ação gerencial, intitulada “Implantar Processos de Gestão de Dados no Laboratório Multiusuário de Bioinformática da Embrapa”, conduzida em 2020, quando se evidenciou a necessidade de criação de instrumentos para auxiliar no processo de gestão de dados ômicos neste Laboratório.

O objetivo central desta ação gerencial consistiu na construção de modelos de conhecimento e na criação de planos de gestão de dados acionáveis por máquina, dinâmicos, interpretáveis, recuperáveis e reusáveis, e capazes de ajudar no gerenciamento de dados dos projetos e atividades de pesquisa no LMB.

Os trabalhos aqui relatados referentes à ação gerencial PGDam foram realizados em 2021, conduzidos por uma equipe multidisciplinar formada por membros do LMB e do GCIA da Embrapa Agricultura Digital, da Biblioteca da Embrapa Agricultura Digital e da Secretaria-Geral da Embrapa.

Esta publicação está organizada em quatro seções, além desta introdução. Na primeira seção são expostos os objetivos e os procedimentos metodológicos adotados na condução da ação gerencial. A segunda seção compreende as quatro etapas estabelecidas para executar as atividades desenhadas, bem como as metodologias utilizadas e os resultados alcançados em cada atividade. Na terceira seção são apresentados e discutidos os resultados obtidos pela ação gerencial. As conclusões estão registradas na seção quatro.

O documento também apresenta três apêndices, sendo dois deles relacionados aos modelos de conhecimento construídos pelos autores na ferramenta DSW (Data Stewardship Wizard, 2021, 2022). O terceiro apêndice, por sua vez, traz o Plano de Gestão de Dados Acionável por Máquina construído para o LMB, baseado na ferramenta de gestão de dados *DMP Canvas Generator*³,

³ Disponível em: <<https://www.vital-it.ch/research/software/DMPCanvasGenerator>>.

desenvolvida pelo Vital-IT Group (Vital-IT, 2022), um centro de computação de alto desempenho, vinculado ao Swiss Institute of Bioinformatics (2022).

Objetivos e Procedimentos Metodológicos

A ação gerencial “Plano de Gestão de Dados Acionável por Máquina (PGDam) Alinhado aos Princípios FAIR” teve como objetivo principal: elaborar modelos de plano de gestão de dados de pesquisa acionável por máquina para o LMB. Para o alcance deste objetivo, foram estabelecidos três objetivos específicos: 1) identificar os referenciais teóricos, as experiências e ferramentas de planos de gestão de dados acionáveis por máquina, alinhados aos princípios FAIR; 2) construir modelos de conhecimento de planos de gestão de dados acionáveis por máquina, consoantes à recomendação *RDA DMP Common Standard for Machine-Actionable Data Management Plans*; e 3) criar modelos de planos de gestão de dados acionáveis por máquina para o LMB e para projetos de pesquisa do LMB.

Para a condução dos trabalhos, foram estabelecidos os seguintes procedimentos metodológicos:

- 1) Revisão bibliográfica em bases de dados de periódicos científicos nacionais e internacionais para mapear e contextualizar a produção técnico-científica sobre plano de gestão de dados (PGD) e plano de gestão de dados acionáveis por máquina (PGDam), visando à concepção do referencial teórico sobre o assunto.
- 2) Análise do estado atual do conhecimento existente sobre PGD, com ênfase em PGDam, visando identificar semelhanças, compatibilidades e lacunas entre PGD e PGDam.
- 3) Identificação de necessidades de integração entre sistemas, a partir da análise dos ecossistemas de dados e de pesquisa do LMB.
- 4) Identificação e análise das principais ferramentas de criação de PGD e de PGDam, adotadas por instituições de pesquisa nacionais e internacionais.
- 5) Exame e experimentação com a ferramenta *Data Stewardship Wizard* (DSW) para desenhar um modelo de conhecimento capaz de expressar os

interesses e especificidades de projetos de pesquisa do LMB, com vistas à construção de PGDam.

6) Construção de modelos de conhecimento de PGDam para projetos de pesquisa do LMB, baseados na ferramenta DSW.

7) Criação de PGDam para projetos de pesquisa do LMB, a partir dos modelos de conhecimento construídos.

8) Criação de PGDam para o LMB, baseado na ferramenta de gestão de dados *DMP Canvas Generator*, desenvolvida pelo Vital-IT.

A Tabela 1 apresenta o modelo metodológico adotado para a condução da ação gerencial.

Tabela 1. Modelo metodológico da ação gerencial.

Objetivos	Procedimentos metodológicos
Identificar os referenciais teóricos, as experiências e ferramentas de planos de gestão de dados acionáveis por máquina, alinhados aos princípios FAIR.	<ul style="list-style-type: none"> • Revisão bibliográfica em bases de dados de periódicos científicos nacionais e internacionais para mapear e contextualizar a produção técnico-científica sobre plano de gestão de dados (PGD) e plano de gestão de dados acionáveis por máquina (PGDam), visando à concepção do referencial teórico sobre o tema. • Análise do estado atual do conhecimento existente sobre PGD, com ênfase em PGDam, visando identificar semelhanças, compatibilidades e lacunas entre PGD e PGDam.
Construir modelos de conhecimento de planos de gestão de dados acionáveis por máquina, consoantes à recomendação <i>RDA DMP Common Standard for Machine-Actionable Data Management Plans</i> .	<ul style="list-style-type: none"> • Identificação de necessidades de integração entre sistemas, a partir da análise dos ecossistemas de dados e de pesquisa do LMB. • Identificação e análise das principais ferramentas de criação de PGD e de PGDam, adotadas por instituições de pesquisa nacionais e internacionais. • Exame e experimentação com a ferramenta <i>Data Stewardship Wizard</i> (DSW) para desenhar um modelo de conhecimento capaz de expressar os interesses e especificidades de projetos de pesquisa do LMB, com vistas à construção de PGDam.
Criar modelos de planos de gestão de dados acionáveis por máquina para o LMB e para projetos de pesquisa do LMB.	<ul style="list-style-type: none"> • Construção de modelos de conhecimento de PGDam para projetos de pesquisa do LMB, baseados na ferramenta DSW. • Criação de PGDam para projetos de pesquisa do LMB, a partir dos modelos de conhecimento construídos. • Criação de PGDam para o LMB, baseado na ferramenta de gestão de dados <i>DMP Canvas Generator</i>, desenvolvida pela Vital-IT.

Criação de Planos de Gestão Dados Acionáveis por Máquina para o Laboratório Multiusuário de Bioinformática da Embrapa

Para a criação de planos gestão de dados acionáveis por máquina, foi considerada prioridade a construção de modelos de conhecimento alinhados aos princípios FAIR, considerados capazes de impulsionar e favorecer a preservação, o compartilhamento, o reúso de dados e a reprodutibilidade da ciência. Tal prioridade foi identificada no decorrer de projeto-piloto de implantação de ações de gerenciamento de dados ômicos no LMB, no qual o plano de gestão de dados acionável por máquina foi escolhido como o principal processo a ser aprimorado.

Operacionalmente, as atividades da ação gerencial foram realizadas em quatro etapas:

- 1) Mapeamento da produção técnico-científica sobre PGD e PGDam.
- 2) Levantamento e qualificação das experiências nacionais e internacionais de PGD e de modelos de PGDam.
- 3) Construção de modelos de conhecimento no software DSW e construção de planos de gestão de dados para projetos de pesquisa do LMB.
- 4) Redação do relatório final da ação gerencial.

A seguir, para cada etapa, são descritas as atividades desenvolvidas e as metodologias utilizadas, bem como os resultados parciais obtidos, conforme mostra a Tabela 2.

Tabela 2. Etapas, atividades, metodologias e resultados parciais da ação gerencial

Etapa 1: Mapeamento da produção técnico-científica sobre PGD e PGDam	
Atividade 1: Identificar referenciais teóricos, iniciativas institucionais, ferramentas e sistemas de construção de PGD e PGDam, experiências de PGD FAIR e PGDam FAIR integrados a fluxos de pesquisa.	
Metodologia	Foram realizadas buscas em bases de dados bibliográficas, periódicos, livros, publicações diversas e websites de instituições de pesquisa, nacionais e internacionais, com ênfase na gestão de dados. Foram utilizadas as seguintes expressões de busca, no singular e plural: plano de gestão de dados; <i>data management plan</i> ; plano de gestão de dados de pesquisa; <i>data research management plan</i> ; plano de gestão de dados FAIR; <i>FAIR data management plan</i> ; plano de gestão de dados acionável por máquina; <i>machine-actionable data management plan</i> . Foram recuperados aproximadamente 90 documentos (artigos/textos), dos quais 74 foram selecionados e distribuídos em duas categorias: fundamentos e princípios e ferramentas e modelos. Em ambas as categorias, os textos foram separados em leitura recomendada e leitura complementar. Todos os textos foram distribuídos livremente entre os integrantes da equipe para leitura, fichamento, apresentação e discussão.
Resultado	Conjunto de artigos categorizados por subtemas concernentes a planos de gestão de dados, disponibilizados para a equipe da ação gerencial.
Atividade 2: Realizar leitura dos textos selecionados sobre PGD e PGDam.	
Metodologia	Foram realizadas as leituras dos documentos (artigos/textos) pelos membros da equipe da ação gerencial, seguidas da elaboração dos respectivos fichamentos, utilizando-se um template criado para as anotações-síntese. Posteriormente, os fichamentos foram apresentados e discutidos em reunião da equipe.
Resultado	Fichamentos dos documentos (artigos/textos), na forma de anotação-síntese, apresentados e discutidos em reuniões com a equipe da ação gerencial.

Continua

Atividade 3: Compilar e elaborar as anotações-síntese dos fichamentos realizados pelos integrantes da equipe.

Metodologia	A partir dos fichamentos dos documentos (artigos/textos) lidos pelos integrantes da equipe, foi realizada uma compilação das anotações-síntese desses fichamentos com a finalidade de captar o processo de avanço do conhecimento sobre o tema plano de gestão de dados acionável por máquina e identificar as lacunas de conhecimento a serem preenchidas. Esta prática teve caráter pedagógico e serviu para que todos os integrantes da equipe tivessem uma noção geral acerca do estado do conhecimento do tema PGDam.
Resultado	Documento-síntese dos fichamentos dos artigos e textos lidos, considerados essenciais ao mapeamento do estado do conhecimento sobre o tema PGDam, e norteadores da construção do modelo de conhecimento e da criação de planos de gestão de dados acionáveis por máquina para o LMB.

Etapa 2: Levantamento e qualificação das experiências nacionais e internacionais de PGD e de modelos de PGDam

Atividade 1: Identificar as experiências de PGD e modelos de PGDam para efetuar estudo quanto ao formato, requisitos, princípios, padrões comuns e recomendações para a elaboração do PGDam do LMB.

Metodologia	Foram realizadas buscas de experiências conduzidas em instituições de PD&I nacionais e internacionais sobre a construção de PGDam. A partir dessas experiências, foi efetuado um estudo das plataformas que aportam o PGDam e/ou realizadas conversas específicas com técnicos/pesquisadores que atuam na construção desses modelos. Ao final da execução desta atividade foram apresentadas as experiências consideradas relevantes para os propósitos da ação gerencial, o que permitiu a construção coletiva da base conceitual e prática norteadora da concepção de um modelo conceitual do PGDam do LMB, inspirado em modelos de conhecimento.
Resultado	Reunião virtual realizada com Viviane Veiga e membros da equipe do ICICT/Fiocruz, com a participação dos integrantes da ação gerencial, com o objetivo de conhecer as ações de PGDam da Fiocruz.

Continua

Atividade 2: Identificar requisitos de PGDam para o LMB, tomando como referência a ferramenta <i>Data Stewardship Wizard</i> (DSW).	
Metodologia	Foram identificados os requisitos de PGDam e analisadas as especificidades do LMB e de seus projetos, no que se refere tanto à construção de modelos de conhecimento quanto à criação de planos de gestão de dados acionáveis por máquina, a partir da ferramenta DSW. Ao final da execução da atividade, foram identificados os fluxos e as interações existentes, tanto no que se refere à gestão dos dados quanto aos sistemas de gestão dos projetos de pesquisa.
Resultado	Documento intitulado “Relação de campos <i>DS-Wizard</i> identificados como necessários para serem usados em um PGD pelo LMB”, que contém os requisitos de PGDam identificados para o LMB, a partir da análise dos campos da ferramenta DSW.
Atividade 3: Identificar as integrações desejáveis entre os sistemas da Embrapa e do LMB, de interesse para PGDam, tomando por base a ferramenta DSW.	
Metodologia	A partir da identificação de requisitos de PGDam para o LMB, realizada com base na ferramenta DSW, foi elaborado um levantamento das possibilidades de integração com os sistemas da Embrapa e do LMB, e passíveis de automatização. Neste levantamento foram identificados: a) campos comuns utilizados em ambos os sistemas; b) sistemas da Embrapa e do LMB passíveis de integração; c) pontos de integração.
Resultado	Documento intitulado “Relação de campos <i>DS-Wizard</i> classificados como PGDam e considerações sobre possível conexão com sistemas e ferramentas computacionais – versão 0.1”, que contém a identificação de possíveis integrações entre sistemas, de interesse para o PGDam, considerando-se as análises dos sistemas da Embrapa e do LMB, tendo como referência a ferramenta DSW.
Etapa 3: Construção de modelos de conhecimento no software DSW e criação de planos de gestão de dados para projetos de pesquisa do LMB	
Atividade 1: Construir dois modelos de conhecimento na ferramenta DSW para planos de gestão de dados de pesquisa do LMB.	
Metodologia	A atividade foi desenvolvida a partir da construção dos modelos de conhecimento para PGDam. Os modelos construídos tomaram por base os requisitos e as necessidades de integração entre sistemas da Embrapa e do LMB, utilizando-se a ferramenta DSW. Para construir os modelos foram realizadas quatro reuniões, tendo sido usado como metodologia a técnica de <i>brainstorming</i> .
Resultado	Dois modelos de conhecimento construídos, sendo um para o LMB e outro para o Projeto RmVAC, baseados na ferramenta DSW.

Continua

Atividade 2: Criar dois planos de gestão de dados acionáveis por máquina na ferramenta DSW para projetos de pesquisa do LMB.

Metodologia	A criação dos planos de gestão de dados acionáveis por máquinas foi baseada nos modelos de conhecimento construídos na ferramenta DSW e no plano de gestão de dados desenvolvido pelo Vital-IT, os quais atendem aos requisitos e às necessidades de integração entre sistemas. A criação dos planos foi realizada em quatro reuniões com os integrantes do LMB. As reuniões além de servirem para criação dos planos foram também um espaço dialógico de troca de informações sobre PGDam e de aprendizado.
Resultado	Dois planos de gestão de dados acionáveis por máquina criados, sendo um para o LMB e outro para o Projeto RmVAC, na ferramenta DSW.

Continua

Atividade 3: Detalhar as etapas de construção do modelo de conhecimento para a criação de PGDam para o LMB.

Metodologia	<p>Para construção do PGD do LMB levou-se em conta as informações necessárias para a submissão de um projeto de análises de bioinformática para o LMB, importantes por outras instituições reconhecidas de bioinformática, assim como sobre o caráter FAIR dos dados. Adotou-se por base o modelo de conhecimento <i>Life Sciences Knowledge Model version 2.2.0</i> do <i>DS-Wizard</i>. O primeiro modelo gerado (DMP_LMB_1.0.0) apresenta a maioria das opções presentes no modelo-base. No segundo modelo (DMP_LMB_RmVAC_1.0.1) foram eliminadas questões consideradas desnecessárias ou de menor importância para o PGD do LMB. Embora menos abrangente, manteve a característica de coletar informações essenciais e de atender aos princípios FAIR. O segundo modelo foi exportado usando o <i>Life Sciences Knowledge Model version 2.3.0</i>. Ao longo da elaboração do segundo modelo tomou-se conhecimento do PGD desenvolvido pelo Vital-IT, um modelo ainda mais conciso que os anteriores, porém com o cerne necessário ao LMB. Com base neste modelo, foi criado um plano de gestão de dados para o LMB, que contempla informações de ordem administrativa, como o nome e o contato dos membros da equipe do projeto, informações sobre o formato dos dados, armazenamento dos dados, tipos de dados, reuso dos dados, metadados e licenças usadas. As informações são coletadas em perguntas, algumas com opções limitando o escopo às análises realizadas no LMB e outras de texto livre, visando obter um plano que possa também ser usado em outras instituições, como por exemplo, instituição de financiamento de pesquisa. Foram realizadas quatro reuniões com os integrantes do LMB e quatro rodadas de análise do documento, sendo duas rodadas com os integrantes do LMB e duas rodadas com todos os membros da equipe da ação gerencial.</p>
Resultado	Documento com detalhamento das etapas de construção do modelo de conhecimento para criação de PGDam.
Etapa 4: Relatório final da ação gerencial	
Atividade 1: Elaboração do relatório das atividades realizadas e resultados obtidos.	
Metodologia	Não se aplica
Resultado	Relatório entregue para comprovação de meta institucional.

A seguir, são apresentados e discutidos sucintamente quatro documentos que nortearam a elaboração de modelos de PGDam para o LMB.

Documentos de referência para a construção de modelos de conhecimento e a criação de planos de gestão de dados acionáveis por máquina para o Laboratório Multiusuário de Bioinformática da Embrapa

A partir da análise da literatura, das discussões e da compreensão do estado do conhecimento sobre planos de gestão de dados, a equipe da ação gerencial apontou quatro documentos considerados de referência e norteadores na construção de modelos de conhecimento e de criação de PGDam para o LMB. São eles:

- 1) *Turning FAIR into reality* (European Commission Expert Group on FAIR Data, 2018).
- 2) *Ten principles for machine-actionable data management plans* (Miksa et al., 2019).
- 3) *Practical guide to the International alignment of research data management* (Science Europe Working Group on Research Data, 2018).
- 4) Plano de gestão de dados FAIR: uma proposta a Fiocruz (Veiga et al., 2019).

Na sequência são apresentadas breves sínteses desses documentos:

Turning FAIR into reality (European Commission Expert Group on FAIR Data, 2018) apresenta uma análise das necessidades para implementar os princípios FAIR, bem como traz 27 recomendações, agrupadas em prioritárias e de apoio. Todavia, as recomendações de apoio devem ser consideradas a partir das recomendações prioritárias, com adição de especificações, detalhes ou etapas mais avançadas para a implementação FAIR.

Notadamente, a Recomendação 22 “Use as informações contidas nos planos de gestão de dados” é de especial interesse para o propósito do presente estudo, pois reforça a importância e a necessidade do uso de planos de gestão de dados de pesquisa acionáveis por máquina. Sabe-se que os planos de gestão de dados (PGD) contêm informações valiosas sobre os dados e outros insumos deles decorrentes; tais informações devem ser estruturadas automaticamente para que possam ser compartilhadas e reutilizadas em ou-

tros projetos de pesquisa. Também é destacada a necessidade de adoção de padrões comuns para apoiar PGD ativos, entendidos como PGD capazes de evoluírem e de serem monitorados, para que permitam a troca de informações em todo o ecossistema de dados FAIR. (European Commission Expert Group on FAIR Data, 2018, p. 72-73). Para tornar efetiva a implementação dos princípios FAIR, as ações enunciadas na Recomendação 22 orientam que:

- 1) Os PGD sejam explicitamente referenciados em sistemas que contenham informações sobre projetos de pesquisa e seus resultados.
- 2) Haja um padrão PGD extensível (por exemplo, o Dublin Core) por disciplina (por exemplo, o Darwin Core); ou pelas características dos dados (por exemplo, escala, sensibilidade); ou por tipo de dados (características específicas e requisitos de codificação).
- 3) O trabalho necessário para tornar os PGD legíveis e acionáveis por máquina inclui o desenvolvimento de conceitos e ferramentas para apoiar a criação de PGDam úteis, utilizáveis e integrados aos fluxos de trabalho de pesquisa.
- 4) Os PGD estejam em conformidade com os princípios FAIR e abertos sempre que possível.
- 5) As informações coletadas no processo de implementação e avaliação de PGD relacionadas à conformidade, desafios e boas práticas sejam usadas para a melhoria da prática.

Ten principles for machine-actionable data management plans (Miksa et al., 2019) apresenta os princípios norteadores para o desenvolvimento de planos de gestão de dados acionáveis por máquina, mostrados na Tabela 3.

Tabela 3. Dez princípios para planos de gestão de dados acionáveis por máquina.

Descrição dos Princípios	
Princípio 1	Integrar PGD com os fluxos de trabalho de todos os interessados no ecossistema de dados da pesquisa
Princípio 2	Permitir que sistemas automatizados funcionem em benefício dos interessados
Princípio 3	Criar regras para as máquinas também, não apenas para as pessoas
Princípio 4	Descrever - para máquinas e para humanos - os componentes do ecossistema de gerenciamento de dados
Princípio 5	Usar IDP (identificadores digitais persistentes) e vocabulários controlados
Princípio 6	Seguir um modelo comum para PGDam
Princípio 7	Tornar os PGD disponíveis para humanos e máquinas
Princípio 8	Apoiar a avaliação do gerenciamento e monitoramento dos dados
Princípio 9	Fazer dos PGD documentos atualizáveis, vivos e versionados
Princípio 10	Tornar os PGD disponíveis ao público

Fonte: Miksa et al. (2019, p. 5, tradução nossa).

Os autores afirmam que os PGD, em vez de serem instrumentos estáticos, precisam ser atualizados nos estágios dos projetos. Enfatizam ainda a necessidade de promover mudanças na cultura do uso dos PGD, para que eles sejam aceitos como recurso de apoio à pesquisa e aos órgãos de fomento.

Practical guide to the International alignment of research data management (Science Europe Working Group on Research Data, 2018) contém orientações para os pesquisadores desenvolverem seus planos de gestão de dados, guiando-os a responderem perguntas em formato de tópicos que os levem ao atendimento de seis requisitos básicos. Esses tópicos que se desdobram em várias perguntas precisam ser abordados e respondidos em um plano de gestão de dados.

A seguir, estão descritos os requisitos básicos estabelecidos no documento citado (Science Europe Working Group on Research Data (2018, p. 9-10)):

- 1) Descrição de dados e coleta ou reutilização de dados existentes.
- 2) Documentação e qualidade de dados.
- 3) Armazenamento e backup durante o processo de pesquisa.
- 4) Requisitos legais e éticos, códigos de conduta.
- 5) Compartilhamento de dados e preservação de longo prazo.
- 6) Responsabilidade e recursos de gerenciamento.

Plano de gestão de dados FAIR: uma proposta para a Fiocruz (Veiga et al., 2019) trata-se de documento de notória importância, por ser uma iniciativa pioneira no Brasil, e também porque sinaliza para o desenvolvimento de um PGD acionável por máquina. Nesse sentido, o estudo realizado no LMB em busca da criação de um PGDam encontra ressonância e aderência, sobretudo quando Veiga et al. (2019, p. 283) pontuam que:

As ferramentas tradicionais disponíveis para automatizar os planos de gestão de dados utilizadas pela comunidade científica internacional, como DMPonline⁷ ou DMPTool⁸, por exemplo, não atendem satisfatoriamente às demandas das pesquisas orientadas em dados complexos dos dias de hoje. Essas [ferramen-

tas] oferecem um formulário online com questões referentes aos PGD, com possíveis explicações, com campos para respostas gerando, ao final, um relatório em .doc ou .pdf. A comunidade científica deseja muito mais do que um documento estático com finalidade administrativa. [...] o desafio do momento está em desenvolver um modelo de PGD acionável por máquina [...]"

O Plano de Gestão de Dados FAIR para a Fiocruz, apresentado por Veiga et al. (2019), consiste de 35 questões, sendo 14 delas de cunho administrativo e as demais inerentes à gestão de dados de pesquisa.

O exame dos referenciais teóricos e conceituais supramencionados, associado à análise de outros textos, que teve o objetivo de ampliar os níveis de informação e de conhecimento da equipe da ação gerencial, confirmou a premissa de que os modelos de PGDam para o LMB devem ser FAIR.

Resultados

As atividades conduzidas para a realização deste estudo produziram seis resultados, sendo três considerados intermediários, por terem contribuído fortemente para a formulação dos outros três resultados, de caráter finalístico.

Os resultados intermediários obtidos consistiram em:

- 1) Documento-síntese dos textos lidos e discutidos (artigos, monografias, teses, apresentações, recomendações técnicas etc.).
- 2) Documento descritivo sobre integrações desejáveis entre sistemas de gestão de dados do LMB e da Embrapa, de interesse para PGDam.
- 3) Documento descritivo de requisitos de PGDam para o LMB, a partir da análise da ferramenta DSW.

Já os três resultados finais devem ser entendidos como aqueles que responderam ao objetivo principal traçado para o estudo, qual seja, desenvolver modelos de conhecimento de plano de gestão de dados acionáveis por máquina, alinhados aos princípios FAIR para o LMB:

1) Construção de dois modelos de conhecimento alinhados aos princípios FAIR, acionáveis por máquina, no software DSW, tomando por base os ecossistemas de dados e de pesquisa, os requisitos e as necessidades do Projeto RmVAC em relação a PGDam, conforme mostram os Apêndices A e B

2) Criação de planos de gestão de dados acionáveis por máquina para projetos de pesquisa do LMB, baseados nos modelos de conhecimento construídos no software DSW, o qual atende aos interesses e necessidades do LMB e do Projeto RmVAC (Apêndice A e Apêndice B).

3) Criação de plano de gestão de dados do Laboratório Multiusuário de Bioinformática da Embrapa, a partir da ferramenta de gestão de dados *DMP Canvas Generator*, desenvolvida pelo Vital-IT, considerada capaz de atender aos requisitos, interesses e necessidades do LMB, como mostra o Apêndice C.

Os dois modelos de conhecimento construídos na ferramenta DSW, passíveis de configuração para integração com o software Dataverse (Dataverse, 2022), são: 1) modelo de conhecimento para o Laboratório Multiusuário de Bioinformática da Embrapa (Apêndice A); e 2) modelo de conhecimento do projeto de pesquisa “Obtenção de formulações vacinais contra o carrapato bovino *Rhipicephalus microplus* e espécies relacionadas”, cuja sigla é “Projeto RmVAC” (Apêndice B).

Ambos os modelos de conhecimento foram construídos por pesquisadores e analistas participantes desta ação gerencial, utilizando-se a funcionalidade de gerar novos modelos de conhecimento, oferecida pela ferramenta DSW. O primeiro modelo de conhecimento tem um caráter genérico e abrangente, e, por isso, é mais extenso e mais completo, já que incorpora um conjunto de perguntas e respostas que procuram cobrir os interesses, as necessidades e as especificidades do LMB.

Este modelo está estruturado em cinco seções compostas por questões relevantes para a elaboração de planos de gestão de dados de pesquisa. Para cada questão, o modelo de conhecimento fornece diversas opções de respostas, o que possibilita o fornecimento de informações detalhadas para o plano. O Apêndice A apresenta em detalhes o modelo de conhecimento genérico, que se compõe dos seguintes tópicos:

I – Administrative information

II – Re-using data

III – Creating and collecting data

IV – Preserving data

V – Giving access to data

O segundo modelo de conhecimento, de abrangência mais restrita, inclui perguntas e respostas que atendem às necessidades e especificidades do Projeto RmVAC. O modelo está estruturado em sete tópicos formados por questões que, ao serem respondidas, ensejam novas perguntas e novas respostas que vão se articulando entre si, de modo a torná-las (perguntas e respostas) cada vez mais claras, detalhadas e específicas. O Apêndice B trata do modelo de conhecimento do Projeto RmVAC, que é formado pelas seguintes seções:

I – Administrative information

II – Data that will be collect/generated

III – Analysis that will be generated

IV – Questions on data collection

V – Data documentation and metadata

VI – Ethics and legal issues

VII – Data preservation, sharing and reusability

Ambos os modelos foram exercitados na criação de dois planos de gestão de dados acionáveis por máquina baseados na ferramenta DSW, objetos igualmente dos Apêndices A e B.

Por sua vez, o Apêndice C refere-se exclusivamente a um plano de gestão de dados criado para o LMB, que foi construído com base na ferramenta de plano de gestão de dados *DMP Canvas Generator*, desenvolvida pelo Vital-IT. Esta ferramenta é composta de seis sessões:

I - Data that will be collected / generated

II - Analysis that will be generated

III - Questions on data collection

IV - Data documentation and metadata

V - Ethics and legal issues

VI - Data preservation, sharing and reusability

O PGD criado para o LMB está estruturado em seis blocos:

Bloco 1 – Informações administrativas

Bloco 2 – Reúso dos dados

Bloco 3 – Geração e coleta de dados

Bloco 4 – Processamento dos dados

Bloco 5 – Preservação de dados

Bloco 6 – Acesso aos dados gerados a partir das análises

Diferentemente dos dois planos mencionados anteriormente, criados a partir da ferramenta DSW, o PGD do LMB não se baseia em modelo de conhecimento.

Conclusões

A adoção de planos de gestão de dados é medida essencial à implementação de melhorias aos processos de gestão de dados no LMB. Tais planos são considerados instrumentos de boas práticas de pesquisa, isso porque reúnem as informações sobre procedência, documentação e qualidade, armazenamento, aspectos legais e éticos e códigos de conduta, compartilhamento, preservação, responsabilidades e recursos financeiros para a gestão de dados. A reunião dessas informações na forma de um plano de gestão de dados cria as condições para que os dados de pesquisa possam ser acessados, preservados e reusados em novas atividades científicas. Para que essas

condições ocorram, os dados assim como os planos de gestão de dados precisam atender aos princípios FAIR, os quais orientam para a descoberta, a acessibilidade, a interoperabilidade e o reusabilidade de dados de pesquisa.

No entanto, os planos de gestão de dados precisam ser necessariamente acionáveis por máquina para que sejam FAIR, de modo que seja possível a integração automática entre sistemas e as informações e os dados possam ser compartilhados e reutilizados em outros projetos de pesquisa, inclusive em outros planos de gestão de dados.

O LMB é uma unidade de pesquisa cuja especificidade dos dados científicos requer um modelo de plano de gestão de dados que reflita as interações e fluxos informacionais com potencial de integração com outros sistemas e serviços, e passíveis de automatização. A nítida compreensão desta particularidade por parte dos integrantes do LMB impulsionou a busca por modelos de planos gestão de dados que fossem capazes de atender aos requisitos técnicos, porém de forma dinâmica, e não meramente burocrática e sem integração, como fazem os modelos baseados nas ferramentas DMPTool e DMPonline.

À luz da análise da literatura e das experiências nacionais e internacionais, e principalmente, da assimilação do conhecimento construído coletivamente pela equipe da ação gerencial, propôs-se a construção de modelos de planos de gestão de dados acionáveis por máquina, alinhados aos princípios FAIR.

Os objetivos estabelecidos para a condução da ação gerencial foram plenamente atendidos. Os dois modelos de conhecimento e os planos de gestão de dados construídos na ferramenta DSW, bem como o plano desenvolvido a partir da ferramenta *DMP Canvas Generator*, do Vital-IT, traduzem as especificidades e necessidades dos pesquisadores do LMB, inerentes ao ecossistema de dados e de pesquisa da Embrapa.

Os dois modelos de conhecimento, assim como os respectivos planos construídos na ferramenta DSW, foram produzidos por meio da utilização da funcionalidade de gerar novos modelos de conhecimento. O primeiro modelo tem caráter genérico e abrangente, com questões e respostas voltadas aos interesses, necessidades e especificidades do LMB. O segundo modelo construído teve abrangência mais restrita, tendo sido projetado para abar-

car perguntas e respostas inerentes às necessidades e especificidades do Projeto RmVAC.

Diferentemente, o plano de gestão de dados baseado na ferramenta *DMP Canvas Generator*, desenvolvida pelo Vital-IT, não é baseado em modelo de conhecimento. A ferramenta foi desenvolvida para ajudar cientistas a gerar planos de gestão de dados para projetos do Swiss National Science Foundation (SNSF). O centro Vital-IT coordena o uso de recursos de tecnologia de informação e bioinformática e, em particular, dá suporte às pesquisas das áreas de genômica e proteômica, que são também as áreas dos projetos de pesquisa do LMB.

Tanto os modelos de conhecimento construídos quanto os planos de gestão de dados criados no escopo desta ação gerencial estão embasados nos referenciais teóricos que refletem a análise do estado do conhecimento sobre o tema, que orientam para a adoção de planos de gestão de dados alinhados aos princípios FAIR e acionáveis por máquina.

Cabe destacar, contudo, que o escopo do estudo empreendido é temporal e gerencialmente limitado, razão pela qual novas frentes de pesquisa serão necessárias para ampliar as possibilidades de melhoria, integração e interoperabilidade entres sistemas.

Ao término das atividades da ação gerencial vislumbraram-se as seguintes possibilidades de continuação dos trabalhos, com vistas a realizar:

- 1) Customizações da ferramenta DSW para incorporar os modelos de conhecimento construídos e os planos de gestão de dados acionáveis por máquina criados para o LMB e para o Projeto RmVAC.
- 2) Integrações automatizadas entre o Repositório de Dados de Pesquisa da Embrapa (Redape), uma instalação do software Dataverse, e os modelos de conhecimento construídos e os planos de gestão criados na ferramenta DSW.
- 3) Avaliação dos planos de gestão de dados construídos para o LMB e para o Projeto RmVAC, no tocante ao nível *FAIRness*.

Referências

- DATA STEWARDSHIP WIZARD. **Data management plans**. Disponível em: <https://ds-wizard.org/data-management-plans>. Acesso em: 14 nov. 2021.
- DATA STEWARDSHIP WIZARD. **Knowledge models**. Disponível em: <https://ds-wizard.org/knowledge-models>. Acesso em: 20 mar. 2022.
- DATAVERSE. **Harvard Dataverse**. [Cambridge, MA: Harvard College, 2022]. Disponível em: <https://dataverse.harvard.edu>. Acesso em: 11 maio 2022.
- EMBRAPA AGRICULTURA DIGITAL. **Laboratório Multiusuário de Bioinformática**. Disponível em: <https://www.embrapa.br/agricultura-digital/lmb>. Acesso em: 14 abr. 2022.
- EUROPEAN COMMISSION EXPERT GROUP ON FAIR DATA. **Turning FAIR into reality: final report and action plan from the European Commission Expert Group on FAIR Data**. Brussels, 2018. 76 p. Disponível em: https://ec.europa.eu/info/sites/default/files/turning_fair_into_reality_1.pdf. Acesso em: 12 abr. 2022.
- GUANDALINI, C. A.; FURNIVAL, A. C. M.; ARAKAKI, A. C. S. Boas práticas científicas na elaboração de planos de gestão de dados. **RDBCI: Revista Digital Biblioteconomia e Ciência da Informação**, v. 17, p. 1-20, 2019. DOI: [10.20396/rdbci.v17i0.8655895](https://doi.org/10.20396/rdbci.v17i0.8655895).
- MIKSA, T.; SIMMS, S.; MIETCHEN, D.; JONES, S. Ten principles for machine-actionable data management plans. **PLoS Computational Biology**, v. 15, n. 3, e1006750, 2019. DOI: [10.1371/journal.pcbi.1006750](https://doi.org/10.1371/journal.pcbi.1006750).
- SCIENCE EUROPE WORKING GROUP ON RESEARCH DATA. **Practical guide to the international alignment of research data management**. Brussels, 2018. 36 p. Disponível em: https://www.scienceeurope.org/media/jezkhnoo/se_rdm_practical_guide_final.pdf. Acesso em: 12 abr. 2022.
- SWISS INSTITUTE OF BIOINFORMATICS. SIB: Swiss Institute of Bioinformatics. **SIB: Swiss Institute of Bioinformatics**. Disponível em: <https://www.sib.swiss/>. Acesso em: 23 maio 2022.
- VEIGA, V.; HENNING, P.; DIB, S.; PENEDO, E.; LIMA, J. da C.; SILVA, L. O. B. da; PIRES, L. F. Plano de gestão de dados FAIR: uma proposta para a Fiocruz. **Liinc em Revista**, v. 15, n. 2, p. 275-286, nov. 2019. Disponível em: <http://revista.ibict.br/liinc/article/view/5030/4349>. Acesso em: 12 abr. 2022.

APÊNDICE A – Modelo de Conhecimento e Plano de Gestão de Dados - DMP_LMB_1.0.0

Organization DSW (demo)
Created by Antonio Nhani, Jr.
(antonio.nhani@embrapa.br)
Based on LMB, 1.0.0 (*dsw.demo:LMBteste:1.0.0*)
Project phase Before Submitting the Proposal
Created at 25 Nov. 2021

Summary

Answered indication

Answered (current phase)	39 / 40	97.50 %
Answered	127 / 136	93.38 %

Metrics

Metric	Score
Findability	0.64
Accessibility	0.81
Interoperability	0.67
Reusability	0.53
Good DMP Practice	0.58
Openness	0.27

I. Administrative information

Summary

Answered indication

Answered (current phase)	4 / 4	100.00 %
Answered	20 / 22	90.91 %

1. Contributors

Each person contributing to creating or executing the data management plan should be added as a contributor. A project probably should have a Contact Person, and a Data Curator.

1.a.1. Name

- ✓ Poliana Fernanda Giachetto

1.a.2. E-mail address

- ✓ poliana.giachetto@embrapa.br

1.a.3. ORCID Identifier

- ✓ <https://orcid.org/0000-0002-9991-6566>

1.a.4. Affiliation

- ✓ Brazilian Agricultural Research Corporation (ROR Organizations):
(<https://ror.org/0482b5b22>)

1.a.5. Role

Roles in a project should be given as they are defined by [datacite](#).

You should specify at least one "Contact Person". If your project has a work package for data management, identify the leader of that work package as "Data Curator".

- ✓ h. Project Leader

1.b.1. Name

- ✓ Antonio Nhani Jr

1.b.2. E-mail address

- ✓ antonio.nhani@embrapa.br

1.b.3. ORCID Identifier

1.b.4. Affiliation

- ✓ Brazilian Agricultural Research Corporation (ROR Organizations):
(<https://ror.org/0482b5b22>)

1.b.5. Role

Roles in a project should be given as they are defined by [datacite](#).

You should specify at least one "Contact Person". If your project has a work package for data management, identify the leader of that work package as "Data Curator".

- ✓ k. Researcher

2. Research Project(s)

Add each of the research project(s) that are you will be working on and for which the data and work are described in this DMP. Give each project a small identifying name for yourself.

2.a.1. Project name

- ✓ Obtenção de formulações vacinais contra o carrapato bovino *Rhipicephalus (B.) microplus*

2.a.2. Project abstract

- ✓ A pecuária brasileira apresenta hoje perdas anuais estimadas em mais de 3 bilhões

de dólares, decorrentes da infestação de bovinos pelo carrapato *Rhipicephalus microplus*. Os prejuízos causados pelo parasita incluem a sua ação direta e indireta, com a transmissão de patógenos que provocam graves enfermidades nos rebanhos e resultam em altos índices de morbidade e mortalidade. A aplicação de acaricidas é hoje a principal estratégia adotada pelos criadores no controle das infestações. No entanto, a seleção de cepas de carrapatos resistentes à maioria dos princípios ativos utilizados têm ocorrido em uma velocidade maior do que o desenvolvimento de novos químicos para o controle do parasita. Ainda, além do alto custo associado e problemas de resistência há ainda a presença de resíduos de acaricidas nos alimentos e no meio ambiente, o que traz a necessidade de desenvolvimento de métodos alternativos de controle. Diante deste cenário, a descoberta de vacinas tem sido o objeto principal das pesquisas na área. Atualmente há duas vacinas disponíveis comercialmente, TickGARD e GavacTM, ambas baseadas na proteína Bm86 recombinante isolada a partir do intestino do carrapato. A eficácia dessas vacinas, no entanto, tem se mostrado bastante variável contra diferentes cepas de *R. microplus*: resultados obtidos em teste conduzido pela Embrapa Gado de Corte com um isolado regional do carrapato evidenciou uma baixa eficácia, 46,4% e 49,2%, respectivamente, para TickGARD e GavacTM. Sabe-se que o desenvolvimento de uma vacina é um processo caro e demorado, mas diante das abordagens existentes hoje para a identificação de alvos, a vacinologia reversa tem se mostrado bastante promissora na busca por antígenos, em função, principalmente, do tempo e custo reduzidos, quando comparada às metodologias tradicionais. A vacinologia reversa é uma estratégia que tem início com a análise in silico de genomas, transcriptomas ou outras sequências biológicas do organismo alvo. Com o auxílio de ferramentas de bioinformática, genomas de patógenos são interrogados com o objetivo de se identificar potenciais alvos para o desenvolvimento de vacinas. As proteínas preditas são selecionadas, baseadas em atributos desejáveis, como a presença de domínios na interface molecular parasita-hospedeiro (porção extracelular) e capacidade de reconhecimento pelo sistema imune do hospedeiro (presença de epítomos de células B e T). Uma vez identificados e priorizados, os alvos são testados em laboratório e depois a campo. A partir de um estudo liderado pela Embrapa Gado de Corte, dentro da Rede Genômica Animal II, e finalizado em 2017, nós geramos, por meio da plataforma Illumina/HiSeq2500, mais de 600 milhões de sequências a partir de diferentes tecidos/estágios de desenvolvimento do *R. microplus*. Além da obtenção do perfil transcricional dessas amostras, identificamos também os genes diferencialmente expressos entre amostras do carrapato crescido em hospedeiros resistentes (Nelore), susceptíveis (Holandês) e cruzados (Nelore x Holandês). Somado a esse extenso repertório de dados, resultado do projeto acima citado, contamos também um banco de dados contendo 8.487 ESTs sequenciadas a partir de amostras de larvas infestantes e ninfas de *R. microplus* e 11.965 ESTs obtidas a partir do sequenciamento de cDNAs do intestino, ovário e glândula salivar do parasita. Por fim, com a recente publicação do draft do genoma do carrapato bovino por Barrero et al. (2017), temos em mãos uma rica e vasta base de dados, para, com a abordagem de vacinologia reversa e juntamente com informações da função biológica

das proteínas preditas a partir dos bancos de dados, identificar potenciais antígenos para a elaboração de vacina(s) eficaz(es) que serão utilizadas em estratégia de controle do carrapato bovino.

2.a.3. Date the project will start

✓ 2019/07/01

2.a.4. Date the project will end

✓ 2023/06/01

2.a.5. Funding

Add all the funding that is part of this project.

2.a.5.a.1. Funder

Specify the name of the funder that you ask for funding for your project. If the funder is not present in the suggested list, please specify a complete URL to the funder web site.

✓ Empresa Brasileira de Pesquisa Agropecuária (Crossref Funders):
(<http://dx.doi.org/10.13039/501100003046>)

2.a.5.a.2. Funding status

✓ c. Granted

2.a.5.a.3. Grant number

3. To execute the DMP, is additional specialist expertise required?

✓ a. No

4. Do you require hardware or software in addition to what is usually available in the institute?

✓ a. No

II. Re-using data

Before you decide to embark on any new study, it is good practice to check all options to re-use existing available data, either collected or generated by yourself in an earlier project, or data from others (Barend Mons calls this "Other People's Data And Services" or OPEDAS). This can include reusable data that have been created for an earlier study, and also so-called "reference data" which is used by many projects. It is not because we can generate massive amounts of data that we always need to do so. Creating data with public money is bringing with it the responsibility to treat those data well and (if potentially useful) make them available for re-use by others. And the circle is only complete if such data is actually re-used.

Summary

Answered indication

Answered (current phase)

4 / 4

100.00 %

1. Is there any pre-existing data?

 **External Links:** [Google dataset search](#), [Datacite Search](#)

Are there any data sets available in the world that are relevant to your planned research?

✓ b. Yes

1.b.1. Will you be using any pre-existing data (including other people's data)?

Will you be referring to any earlier measured data, reference data, or data that should be mined from existing literature? Your own data as well as data from others?

✓ a. No

1.b.2. Do you need to harmonize different sources of existing data?

If you are combining data from different sources, harmonization may be required. You may need to re-analyse some original data.

✓ a. No

1.b.3. Will you be using data that needs to be (re-)made computer readable first?

Some old data may need to be recovered, e.g. from tables in scientific papers or may be punch cards.

✓ a. No

III. Creating and collecting data

We will make sure that we know what data will be coming together in the project, when it will be coming. We also need to make sure that we have adequate storage space to deal with it, and that all the responsibilities have been taken care of.

Summary

Answered indication

Answered (current phase)	15 / 15	100.00 %
Answered	41 / 43	95.35 %

Metrics

Metric	Score
Findability	0.50
Accessibility	0.50
Interoperability	0.67
Reusability	0.49

1. What existing data formats/types will you be using?

Have you identified types of data that you will use that are used by others too? Some types of data (for example "images" or "tables") are used by many different projects. For such data, often common standards exist (in our example "JPG" and "CSV" [comma separated values]) that help to make these data reusable. Are you using such common data formats?

1.a.1. Data format/type

✓ fastq fasta

1.a.2. Is this a standard data format widely used by researchers in this field?

1.a.3. Does this data format enable sharing and long term archiving?

Complicated (binary) file formats tend to change over time, and software may not stay compatible with older versions. Also, some formats (e.g. DOC, XLS) hamper long term usability by making use of patents or being hampered by restrictive licensing. Ideally, a format should be simple, text only, completely described, not restricted by copyrights, and implemented in different software packages.

✓ b. Yes

1.a.4. What volume of data of this type will you be working with?

✓ c. I can specify the number of files/subjects and the size of each

1.a.4.c.1. Number of files/subjects

✓ 9

1.a.4.c.2. Rough average size of each file/subject in gigabytes

✓ 0.05

1.a.5. Is this data format completely described?

Formats like XLS or SQL are very flexible; they can be adapted to many different uses, and this makes them good for interoperability. However, their flexibility also makes that it is not immediately obvious from the file structure how it can be used. The data needs a proper *description* in order for others (or yourself at a later time) to be able to unambiguously understand what it contains.

✓ b. Yes

2. What existing encodings/terminologies/vocabularies/ontologies will you be using?

2.a.1. Name

✓ Agro

2.a.2. If you use a standard that is not in FAIRsharing, please specify its PID or URL

3. Will you be using new types of data?

Sometimes the type of data you collect can not be stored in a commonly used data format. In such cases you may need to make your own, keeping interoperability as

high as possible.

✓ a. No, all of my data will fit in common formats

4. How will you be collecting and keeping your metadata?

For the re-usability of your data by yourself or others at a later stage, a lot of information about the data, how it was collected and how it can be used should be stored with the data. Such data about the data is called metadata, and this set of questions are about this metadata

✓ a. Explore

4.a.1. What standard(s) will you use to specify author/title/keyword information?

There are a few different standards that are often used to give basic information about your dataset. Which ones of these will you be using?

✓ a. Explore

4.a.1.a.1. Will you document the data with Dublin Core metadata?

🔗 **External Links:** [Dublin Core Metadata Terms](#), [Dublin Core Initiative](#)

Dublin Core is a standard documenting domain independent aspects of a resource; including who has created it, audience, function, formatting and licensing. Does your documentation follow the Dublin Core standard?

✓ a. No

4.a.1.a.2. Will you document the data with DataCite metadata

🔗 **External Links:** [DataCite metadata schema](#)

✓ a. No

4.a.1.a.3. Will you document the data with DDI metadata

🔗 **External Links:** [DDI metadata documentation](#)

DDI metadata is more extensive than Dublin Core and DataCite, it details more of what is in the data and really can help other researchers locate your data set as an interesting source.

✓ a. No

4.a.2. Do suitable 'Minimal Metadata About ...' (MIA...) standards exist for your experiments?

🔗 **External Links:** [FAIRsharing repository of standards](#)

Many research fields have worked together to define what kind of metadata should really be collected when an experiment of a certain kind is performed and described. That information is described in a Minimal Metadata Standard. Often, these standards describe both what kind of information needs to be collected as well as the format in which it is expected.

✓ a. No

4.a.2.a.1. Do you have a good idea of what metadata is needed to make it possible for others to read and interpret your data in the future?

🔗 **External Links:** [FAIRsharing repository of standards](#)

Imagine yourself trying to interpret a similar data set from a colleague in a few years. Is there any way in which the data could be misinterpreted? Does everyone in the field do things in exactly the same way? And will that still be known in 10 years? How can such misinterpretation be avoided?

✓ a. No

4.a.3. Do you know how and when you will be collecting the necessary metadata?

🔗 **External Links:** [FAIRsharing repository of standards](#)

Often it is easiest to make sure you collect the metadata as early as possible.

✓ b. Yes

4.a.4. Will you consider re-usability of your data beyond your original purpose?

Adding more than the strict minimum metadata about your experiment will possibly allow more wide re-use of your data, with associated higher data citation rates. Please note that it is not easy for yourself to see all other ways in which others could be reusing your data.

✓ b. Yes, I will add "optional" metadata where I can

4.a.4.b.1. How will you balance the extra efforts with the potential for added reusability?

✓ c. I will collect all metadata I can gather and document the data set beyond minimal standards

4.a.4.b.2. Do you need to exchange your data with others?

✓ b. Yes

4.a.5. Did you consider how to monitor data integrity?

Working with large amounts of heterogeneous data in a larger research group has implications for the data integrity. How do you make sure every step of the workflow is done with the right version of the data? How do you handle the situation when a mistake is uncovered? Will you be able to redo the strict minimum data handling?

✓ a. Explore

4.a.5.a.1. Will you be keeping a master list with checksums of certified/correct/canonical/verified data?

Data corruption or mistakes can happen with large amounts of files or large files.

Keeping a *master list* with data checksums can be helpful to prevent expensive mistakes, because it will help detect early when data files are damaged or mixed up. It can also be helpful to keep the list under version control so that all changes are well documented.

✓ b. Yes

4.a.5.a.2. Will you define a way to detect file or sample swaps, e.g. by measuring something independently?

✓ a. No

4.a.6. Do all datasets you work on have a license?

It is not always clear to everyone in the project (and outside) what can and can not be done with a data set. It is helpful to associate each data set with a license as early as possible in the project. A data license should ideally be as free as possible: any restriction like 'only for non-commercial use' or 'attribution required' may reduce the reusability and thereby the number of citations. If possible, use a computer-readable and computer actionable license.

- ✓ a. No

4.a.7. How will you do file naming and file organization?

Putting some thoughts into file naming can save a lot of trouble later.

- ✓ a. Explore

4.a.7.a.1. Did you make a SOP (Standard Operating Procedure) for file naming?

It can help if everyone in the project uses the same naming scheme.

- ✓ a. No

4.a.7.a.2. Will you be keeping the relationships between data clear in the file names?

- ✓ b. Yes

4.a.7.a.3. Will all the metadata that is embedded in the file names also be available in the proper metadata?

The file names are very useful as metadata for people involved in the project, but to computers they are just identifiers. To prevent accidents with e.g. renamed files metadata information should always also be available elsewhere and not only through the file name.

Also note that if metadata could need to change, embedding it in the file names may require renaming files during the project; and this may have implications for references to those files.

- ✓ b. Yes, all metadata is also explicitly available elsewhere

4.a.7.a.4. Will you be using persistent identifiers to refer to data within the project?

🔗 **External Links:** [The Handle System](#), [Handbook on Persistent Identifiers](#)

Especially for large projects, referring to data internally via a persistent identifier system can be helpful as such a system can help to keep track of data that moves to a new location.

- ✓ b. Yes

4.a.8. How will you be keeping track of the "provenance" of the data?

Data analysis is normally done step-by-step. It is essential to make sure all steps are properly documented, otherwise results will not be reproducible. Re-users of the data also need this information to decide whether the data can be used for their purpose.

- ✓ c. We use other arrangements

4.a.8.c.1. What other arrangements?

- ✓ A dataverse project

4.a.9. Will you be documenting the data with W3C PROV provenance?

🔗 **External Links:** [W3C Prov primer](#)

The W3C Prov standard documents processes (workflow) that were used to produce a resource. This can be used to document e.g. the software (including version) and parameters you use to analyze the data. Will your documentation follow the W3C Prov standard?

- ✓ a. No

4.a.10. Will you use a workflow system that automatically keeps track of the steps in the analysis?

Some workflow systems automatically keep track of which steps were done in data analysis and what options were selected. This can help document the data for reproducibility.

- ✓ b. Yes

5. Will you be acquiring data using measurement equipment?

- ✓ a. No

6. Do you have any non-equipment data capture?

Does the data you collect contain non-equipment captured data such as questionnaires, case report forms, electronic patient records?

- ✓ a. No

7. Is there a data integration tool that can handle and combine all the data types you are dealing with in your project?

- ✓ a. No

7.a.1. Can all data be brought into the same format, e.g. RDF?

- ✓ a. No

8. Will you be storing physical samples?

Will you be collecting artefacts like minerals, biological samples, specimens?

- ✓ a. No

9. Will you need consent for any newly collected personal data?

- ✓ a. No, we do not collect any new personal data

10. How is the ownership of the collected data arranged?

- ✓ d. We have a consortium agreement that arranges Intellectual Property

IV. Preserving data

In this chapter, issues regarding data publication and long term archiving are addressed.

Summary

Answered indication

Answered (current phase)	9 / 9	100.00 %
Answered	43 / 47	91.49 %

Metrics

Metric	Score
Findability	0.88
Accessibility	0.80
Reusability	0.62
Good DMP Practice	0.50

1. Specify a list of data sets you will be producing
Add all the data sets you will be producing. Give each a short name, sufficient for yourself to know what data it is about. It is useful to think about a data set as some collection of data that will be ending up in the same place.

1.a.1. Data set:

✓ fastq files

1.a.2. Description of the data set

What type of data is in this data set? Examples could be "Field observations", "raw instrument data", "genomic variants".

1.a.3. Identifier of the data set

Please add all "formal" identifiers you have for this data set: these can be handles or DOIs or any other type. One important purpose of these identifiers is to be able to find the dataset back.

1.a.4. Will this data set be published?

Will you publish the data set somewhere? Note that this does not necessarily mean that the data set becomes openly available, conditions for access and use may apply.

✓ b. Yes

1.a.4.b.1. Specify where you will distribute this data from

Add each of the locations where this data set will be made available. Give each of these a short name that identifies it to yourself.

1.a.4.b.1.a.1. What repository will this data be stored in?

🔗 **External Links:** [FAIRSharing](#), [Registry of Research data Repositories](#)

Domain repositories often have the best functionality to make the data findable and reusable: even though it may look like a database that could be reused in a completely different field would be better findable in a generic repository, the limited availability of domain-specific metadata make that less valuable.

Many repositories are listed in <https://fairsharing.org/>

If a repository offers to give your data set a DOI or alternative persistent identifier it is a good idea to use that option.

✓ c. Our institutional repository

1.a.4.b.1.a.2. Who will the data in this place be shared with?

✓ b. Shared: The data will be shared with a predefined list of people, e.g. coworkers

1.a.4.b.1.a.3. Licenses under which this distribution of the data set will be available

Please add each license that this data set will be available as. For each license you will be able to specify when it starts being applicable, so that you can e.g. specify that the data is restricted for a few months and open afterwards.

1.a.4.b.1.a.3.a.1. Under what license will the data set be made available?

✓ c. They will be available under some restrictions

1.a.4.b.1.a.3.a.1.c.1. What conditions?

Please summarise the restrictions. Note that it is a good idea to publish the final conditions for use of the data openly, so that interested researchers can know beforehand whether they will be able to get access to the data before actually placing a request.

✓ CC-BY-NC-SA

1.a.4.b.1.a.3.a.1.c.2. Link to the license text

If the license is available as a text on the internet, please specify its URL here.

✓ <https://creativecommons.org/licenses/by-nc-sa/3.0>

1.a.4.b.1.a.3.a.1.c.3. How will re-users be able to get access?

Consider that users of a data set will need to be able to get access several years after the project is finished. The best way to arrange this is to put the responsibility for checking the rights to access the data with a fixed body in an organisation, a data access committee, rather than one of the investigators of the project.

✓ a. By request to the contact person

1.a.4.b.1.a.3.a.1.c.4. Will the conditions be published as part of open metadata?

✓ b. Yes

1.a.4.b.1.a.3.a.2. Starting date

From which date will data be available under this license?

✓ 2024/01/01

1.a.4.b.2. Will you be adding a reference to the published data to at least one data catalogue?

🔗 **External Links:** [What is a data catalog \(Dataversity\)](#)

Data is sometimes difficult to locate, especially if it is not in a domain-specific repository. Data catalogues may increase findability: they put the metadata for data sets that are located in different places together in one location that is well known to

researchers in a particular field.

Data sets can be mentioned in multiple data catalogues without the need for duplicating the data: the data catalogues point to where the actual data can be found.

✓ c. Yes

1.a.4.b.3. Will there be different versions of this data over time?

✓ a. No

1.a.4.b.4. Will a user of this data need specific software to be able to use it?

Some data types may be closely coupled to specific software packages. If that is the case for this data set, please name and point out the required software.

✓ a. No

1.a.5. Does this dataset contain sensitive information?

Personal information can be sensitive if it is for instance about the health, sexual orientation, religion of a person. But there are also other classes of sensitive information: e.g. locations of rare species in biodiversity could be sensitive and should not leak to poachers.

✓ b. Yes

2. Will you be archiving data (using so-called 'cold storage') for long term preservation already during your project?

Much of the raw data you have will need to be archived for your own later use somewhere. This is often done off-line on tape, not on the disks of the compute facility. Please note that this does not refer to the data publication.

✓ b. Yes

2.b.1. Is the archived data changing over time, needing re-archival?

✓ a. No

2.b.2. Will the archive be stored on disk or on tape?

✓ b. Tape

2.b.3. Will the archive be stored in a remote location, protecting the data against disasters?

✓ b. Yes

2.b.4. Will the archive need to be protected against loss or theft?

✓ b. Yes

2.b.4.b.1. Will the archive be encrypted?

✓ a. No

2.b.4.b.2. Is it clear who has physical access to the archives?

✓ b. Yes

2.b.5. Will your project require the archives to be available on-line?

✓ b. Yes

2.b.5.b.1. Will data integrity be guaranteed?

If the 'master copy' of the data is available on line, it should probably be protected against being tampered with.

✓ b. Yes

2.b.5.b.2. Is there an interface and a defined process for people to request access to the data?

✓ b. Yes

2.b.6. Has it been established who has access to the archive, and how fast?

✓ b. Yes

2.b.6.b.1. Has it been established who can ask for a restore during the project?

✓ b. Yes

2.b.6.b.2. If the data is voluminous, will the project be able to cope with the time needed for a restore?

✓ b. Yes

2.b.6.b.3. Has authority over the data been arranged for when the project is finished (potentially long after)?

✓ b. Yes

2.b.7. Has it been established how long the archived data need to be kept? For each of the different parts of the archive (raw data / results)?

✓ a. No

2.b.8. Will the data still be understandable and reusable after a long time?

See also all questions about keeping metadata and data formats. Make sure the metadata is kept close to the data in the archive, and that community supported data formats are used for all long term archiving.

✓ b. Yes

3. Will you be archiving your data in 'cold storage' after the project finishes? Will you be storing (in cold storage) copies of your own data for a longer period after the project has ended? Possibly as a continuation of archival as part of data storage strategy during the project? Data archival is distinct from data publishing, an archive is usually strictly limited in who can access the data.

✓ a. No

4. Will any of the repositories you use charge you for their services?

✓ a. No

5. Are there any other recurring fees to keep data or documents available? Are you using any commercially licensed products to keep data, software or documents available, for which a regular fee must be paid?

✓ a. No

6. Did you budget for the time and effort it will take to prepare the data for publication?

✓ a. No

7. Will you also publish data if the results of your study are negative/inconclusive or unpublishable?

Even if you do not obtain the results you had foreseen from your own study, the data can still be valuable for reuse in another context. Also, publishing the data can avoid that someone else collects a similar data set with a similar negative result.

✓ b. Yes

8. Will you be making sure that blocks of data deposited in different repositories can be recognized as belonging to the same study?

✓ b. Yes, all data sets will have links to the related data

9. Specify a list of software packages you will be publishing
Specify a short name for each software package.

9.a.1. Software package:

✓ In house scripts

9.a.2. Will you be adding a proper open-source license?

✓ c. Yes, we will decide on an open source license

9.a.3. Where will the software package be available?

9.a.4. Will this software be listed in a catalogue?

10. Will reference data be created?

Will any of the data that you will be creating form a reference data set for future research (by others)?

Much of today's data is used in comparison with reference data. You may be comparing your own data with a "standard set" which is maintained as a collection by someone else. Or you could be determining differences to a standard (for example in bioinformatics, a genome is often compared with a reference genome to identify genomic variants). Will you be creating any data that will be reference data for other researchers?

✓ a. No

V. Giving access to data

This chapter deals with the information needed by people who will re-use your data, and with the access conditions they will need to follow.

Summary

Answered indication

Answered (current phase)

7 / 8

87.50 %

Answered 19 / 20 95.00 %

Metrics

Metric	Score
Findability	0.30
Accessibility	1.00
Good DMP Practice	0.75
Openness	0.27

1. Can all of your data become completely open immediately?
Some data may be subject to a temporary embargo, or need to stay closed for specific reasons.

✓ a. No

1.a.1. Are there legal reasons why (some of your) data can not be completely open?

✓ b. Yes

1.a.1.b.1. Are there privacy reasons why your data can not be open?

✓ b. Yes

1.a.1.b.1.b.1. Are there restrictions on where the data need to be stored?

✓ d. Yes, they must stay in the same institute

1.a.1.b.1.b.2. Could pseudonymization be used to make the data more openly available?

Legally, pseudonymous data (which means that someone has the key to reverse the process) is still considered privacy sensitive information. However, the EU is working on special cases where the data can still be opened as long as the key availability is sufficiently limited.

✓ a. No

1.a.1.b.1.b.3. Could anonymization be used to make the data more openly available?

Different anonymization techniques exist. Disadvantage of anonymization is that data integration becomes virtually impossible, but it may be the only way to open up your data for other research

✓ a. No

1.a.1.b.1.b.4. Could you use data aggregation to make the data openly available?

Aggregated data, where typically at least 15 individuals are in any data point, are considered sufficiently anonymous. This is an alternative way of making data openly available for future research

✓ a. No

1.a.1.b.2. Are there IP reasons why your data can not be open?

- ✓ b. Yes

1.a.1.b.2.b.1. Is it clear who owns data and documents?

- ✓ b. Yes

1.a.1.b.2.b.2. Will someone be given decision power to move documents or data to a new place after the project has finished?

In one case in the past, all documents that had been assembled by a project in a documentation system had to be deleted because not a single person could decide to move them to a new platform when the documentation system was going off-line.

- ✓ b. Yes

1.a.1.b.3. Will you be allowing authenticated access to the data?

- ✓ b. Yes

1.a.1.b.3.b.1. Where will the data be stored?

- ✓ b. In a national or institutional repository that arranges restricted access

1.a.1.b.3.b.2. Who will take care of authentication of potential users?

- ✓ a. We will use username/password authentication, possibly augmented with two-factor security

1.a.1.b.3.b.3. Who will take care of authorization of potential users?

- ✓ c. We will make use of an existing data access committee

1.a.1.b.3.b.3.c.1. Which existing Data Access Committee?

- ✓ Project leader, Data manager

1.a.1.b.3.b.4. Are the criteria for application to access the data openly available?

- ✓ b. Yes

1.a.1.b.3.b.5. Has auditing for the re-use been arranged?

- ✓ b. Yes

1.a.2. Are there business reasons why (some of your) data can not be completely open?

- ✓ b. Yes, patents still need to be applied for

1.a.3. Are there other reasons why (some of your) data can not be completely open?

- ✓ c. Yes, other reasons

1.a.3.c.1. What other reasons are there not to open all data immediately?

APÊNDICE B – Modelo de Conhecimento e Plano de Gestão de Dados - DMP_LMB_RmVAC_1.0.1

Organization DSW (demo)
Created by Antonio Nhani, Jr. (antonio.nhani@embrapa.br)
Based on LMB, 1.0.1 (*dsw.demo:LMBteste:1.0.1*)
Project phase Before Submitting the Proposal
Created at 25 Nov. 2021

Summary

Answered indication

Answered (current phase)	20 / 25	80.00 %
Answered	65 / 74	87.84 %

Metrics

Metric	Score
Findability	0.75
Accessibility	1.00
Reusability	0.78
Openness	0.50

I. Administrative information

Summary

Answered indication

Answered (current phase)	4 / 4	100.00 %
Answered	26 / 27	96.30 %

1. Contributors

Each person contributing to creating or executing the data management plan should be added as a contributor. A project probably should have a Contact Person, and a Data Curator.

1.a.1. Name

✓ Poliana Fernanda Giachetto

1.a.2. E-mail address

✓ poliana.giachetto@embrapa.br

1.a.3. ORCID Identifier

- ✓ <https://orcid.org/0000-0002-9991-6566>

1.a.4. Affiliation

- ✓ Brazilian Agricultural Research Corporation (ROR Organizations):
(<https://ror.org/0482b5b22>)

1.a.5. Role

Roles in a project should be given as they are defined by [datacite](#).

You should specify at least one "Contact Person". If your project has a work package for data management, identify the leader of that work package as "Data Curator".

- ✓ h. Project Leader

1.b.1. Name

- ✓ Antonio Nhani Jr

1.b.2. E-mail address

- ✓ antonio.nhani@embrapa.br

1.b.3. ORCID Identifier

1.b.4. Affiliation

- ✓ Brazilian Agricultural Research Corporation (ROR Organizations):
(<https://ror.org/0482b5b22>)

1.b.5. Role

Roles in a project should be given as they are defined by [datacite](#).

You should specify at least one "Contact Person". If your project has a work package for data management, identify the leader of that work package as "Data Curator".

- ✓ k. Researcher

1.c.1. Name

- ✓ Poliana Fernanda Giachetto

1.c.2. E-mail address

- ✓ poliana.giachetto@embrapa.br

1.c.3. ORCID Identifier

- ✓ Brazilian Agricultural Research Corporation

1.c.4. Affiliation

- ✓ Brazilian Agricultural Research Corporation (ROR Organizations):
(<https://ror.org/0482b5b22>)

1.c.5. Role

Roles in a project should be given as they are defined by [datacite](#).

You should specify at least one "Contact Person". If your project has a work package for data management, identify the leader of that work package as "Data Curator".

✓ a. Contact Person

2. Research Project(s)

Add each of the research project(s) that are you will be working on and for which the data and work are described in this DMP. Give each project a small identifying name for yourself.

2.a.1. Project name

✓ Obtenção de formulações vacinais contra o carrapato bovino *Rhipicephalus (B.) microplus*

2.a.2. Project abstract

✓ A pecuária brasileira apresenta hoje perdas anuais estimadas em mais de 3 bilhões de dólares, decorrentes da infestação de bovinos pelo carrapato *Rhipicephalus microplus*. Os prejuízos causados pelo parasita incluem a sua ação direta e indireta, com a transmissão de patógenos que provocam graves enfermidades nos rebanhos e resultam em altos índices de morbidade e mortalidade. A aplicação de acaricidas é hoje a principal estratégia adotada pelos criadores no controle das infestações. No entanto, a seleção de cepas de carrapatos resistentes à maioria dos princípios ativos utilizados tem ocorrido em uma velocidade maior do que o desenvolvimento de novos químicos para o controle do parasita. Ainda, além do alto custo associado e problemas de resistência há ainda a presença de resíduos de acaricidas nos alimentos e no meio ambiente, o que traz a necessidade de desenvolvimento de métodos alternativos de controle. Diante deste cenário, a descoberta de vacinas tem sido o objeto principal das pesquisas na área. Atualmente há duas vacinas disponíveis comercialmente, TickGARD e GavacTM, ambas baseadas na proteína Bm86 recombinante isolada a partir do intestino do carrapato. A eficácia dessas vacinas, no entanto, tem se mostrado bastante variável contra diferentes cepas de *R. microplus*: resultados obtidos em teste conduzido pela Embrapa Gado de Corte com um isolado regional do carrapato evidenciou uma baixa eficácia, 46,4% e 49,2%, respectivamente, para TickGARD e GavacTM. Sabe-se que o desenvolvimento de uma vacina é um processo caro e demorado, mas diante das abordagens existentes hoje para a identificação de alvos, a vacinologia reversa tem se mostrado bastante promissora na busca por antígenos, em função, principalmente, do tempo e custo reduzidos, quando comparada às metodologias tradicionais. A vacinologia reversa é uma estratégia que tem início com a análise in silico de genomas, transcriptomas ou outras sequências biológicas do organismo alvo. Com o auxílio de ferramentas de bioinformática, genomas de patógenos são interrogados com o objetivo de se identificar potenciais alvos para o desenvolvimento de vacinas. As proteínas preditas são selecionadas, baseadas em atributos desejáveis, como a presença de domínios na interface molecular parasita-hospedeiro (porção extracelular) e capacidade de reconhecimento pelo sistema imune do hospedeiro (presença de epítomos de células B e T). Uma vez identificados e priorizados, os alvos são testados em laboratório e depois a campo. A partir de um estudo liderado pela Embrapa Gado de Corte, dentro da Rede Genômica Animal II, e finalizado em 2017, nós geramos, por meio da

plataforma Illumina/HiSeq2500, mais de 600 milhões de sequências a partir de diferentes tecidos/estágios de desenvolvimento do *R. microplus*. Além da obtenção do perfil transcricional dessas amostras, identificamos também os genes diferencialmente expressos entre amostras do carrapato crescido em hospedeiros resistentes (Nelore), susceptíveis (Holandês) e cruzados (Nelore x Holandês). Somado a esse extenso repertório de dados, resultado do projeto acima citado, contamos também um banco de dados contendo 8.487 ESTs sequenciadas a partir de amostras de larvas infestantes e ninfas de *R. microplus* e 11.965 ESTs obtidas a partir do sequenciamento de cDNAs do intestino, ovário e glândula salivar do parasita. Por fim, com a recente publicação do draft do genoma do carrapato bovino por Barrero et al. (2017), temos em mãos uma rica e vasta base de dados, para, com a abordagem de vacinologia reversa e juntamente com informações da função biológica das proteínas preditas a partir dos bancos de dados, identificar potenciais antígenos para a elaboração de vacina(s) eficaz(es) que serão utilizadas em estratégia de controle do carrapato bovino.

2.a.3. Date the project will start

✓ 2019/10/01

2.a.4. Date the project will end

✓ 2023/09/01

2.a.5. Funding

Add all the funding that is part of this project.

2.a.5.a.1. Funder

Specify the name of the funder that you ask for funding for your project. If the funder is not present in the suggested list, please specify a complete URL to the funder web site.

✓ Empresa Brasileira de Pesquisa Agropecuária (Crossref Funders):
(<http://dx.doi.org/10.13039/501100003046>)

2.a.5.a.2. Funding status

✓ c. Granted

2.a.5.a.3. Grant number

✓ 30.19.90.007.00.00

3. To execute the DMP, is additional specialist expertise required?

✓ a. No

4. Do you require hardware or software in addition to what is usually available in the Embrapa Multiuser Laboratory (LMB)?

✓ a. No

II. Data that will be collected / generated

Information on the data type categories that best fit the data that you plan to collect/generate during the course of your project.

Summary

Answered indication

Answered (current phase)	1 / 1	100.00 %
Answered	1 / 2	50.00 %

1. Please choose the general data type categories that best fit the data that you plan to collect/generate

- Genomic data
- Transcriptomic data
- Metabolomic data
- Proteomic data
- Lipidomic data
- Molecular Biology data
- Microscopy images
- Flow cytometry data
- Other

2. In case of Other type of data (if option on the previous question), please describe it

Describe your data

III. Analysis that will be generated

Please choose the approach that will be used to analyze your data

Summary

Answered indication

Answered (current phase)	1 / 2	50.00 %
Answered	1 / 2	50.00 %

1. Type of analysis

- Comparative Genomics
- Metagenomics
- Quantitative Trait Loci
- Microarray
- Differential Expression
- Other

2. Please, describe your analysis if other type was selected

Describe your analysis here

IV. Questions on data collection

We will make sure that we know what data will be coming together in the project, when it will be coming. We also need to make sure that we have adequate storage space to deal with it, and that all the responsibilities have been taken care of.

Summary

Answered indication

Answered (current phase)	7 / 8	87.50 %
Answered	29 / 32	90.62 %

Metrics

Metric	Score
Findability	0.75
Accessibility	1.00
Reusability	0.78

1. Will there be any use of human subjects or samples?

The answer should be YES or NO

✓ b. No

2. Will there be any use of vertebrate subjects or samples?

The answer should be YES or NO

✓ b. No

3. What is your estimation of total produced data volume?

✓ About 1,5TB

4. How will you be collecting and keeping your metadata?

For the re-usability of your data by yourself or others at a later stage, a lot of information about the data, how it was collected and how it can be used should be stored with the data. Such data about the data is called metadata, and this set of questions are about this metadata

✓ a. Explore

4.a.1. What standard(s) will you use to specify author/title/keyword information?

There are a few different standards that are often used to give basic information about your dataset. Which ones of these will you be using?

✓ a. Explore

4.a.1.a.1. Will you document the data with Dublin Core metadata?

🔗 **External Links:** [Dublin Core Metadata Terms](#), [Dublin Core Initiative](#)

Dublin Core is a standard documenting domain independent aspects of a resource; including who has created it, audience, function, formatting and licensing. Does your

documentation follow the Dublin Core standard?

✓ a. No

4.a.1.a.2. Will you document the data with DataCite metadata

🔗 **External Links:** [DataCite metadata schema](#)

✓ a. No

4.a.1.a.3. Will you document the data with DDI metadata

🔗 **External Links:** [DDI metadata documentation](#)

DDI metadata is more extensive than Dublin Core and DataCite, it details more of what is in the data and really can help other researchers locate your data set as an interesting source.

✓ b. Yes

4.a.2. Do suitable 'Minimal Metadata About ...' (MIA...) standards exist for your experiments?

🔗 **External Links:** [FAIRsharing repository of standards](#)

Many research fields have worked together to define what kind of metadata should really be collected when an experiment of a certain kind is performed and described. That information is described in a Minimal Metadata Standard. Often, these standards describe both what kind of information needs to be collected as well as the format in which it is expected.

✓ b. Yes

4.a.2.b.1. Which "Minimal Information" standards will you use?

4.a.3. Do you know how and when you will be collecting the necessary metadata?

🔗 **External Links:** [FAIRsharing repository of standards](#)

Often it is easiest to make sure you collect the metadata as early as possible.

✓ b. Yes

4.a.4. Will you consider re-usability of your data beyond your original purpose?

Adding more than the strict minimum metadata about your experiment will possibly allow more wide re-use of your data, with associated higher data citation rates. Please note that it is not easy for yourself to see all other ways in which others could be reusing your data.

✓ b. Yes, I will add "optional" metadata where I can

4.a.4.b.1. How will you balance the extra efforts with the potential for added reusability?

✓ a. I will see what I can do

4.a.4.b.2. Do you need to exchange your data with others?

✓ b. Yes

4.a.5. Did you consider how to monitor data integrity?

Working with large amounts of heterogeneous data in a larger research group has implications for the data integrity. How do you make sure every step of the workflow is done with the right version of the data? How do you handle the situation when a mistake is uncovered? Will you be able to redo the strict minimum data handling?

✓ a. Explore

4.a.5.a.1. Will you be keeping a master list with checksums of certified/correct/canonical/verified data?

Data corruption or mistakes can happen with large amounts of files or large files. Keeping a *master list* with data checksums can be helpful to prevent expensive mistakes, because it will help detect early when data files are damaged or mixed up. It can also be helpful to keep the list under version control so that all changes are well documented.

✓ b. Yes

4.a.5.a.2. Will you define a way to detect file or sample swaps, e.g. by measuring something independently?

✓ b. Yes

4.a.6. Do all datasets you work on have a license?

It is not always clear to everyone in the project (and outside) what can and can not be done with a data set. It is helpful to associate each data set with a license as early as possible in the project. A data license should ideally be as free as possible: any restriction like 'only for non-commercial use' or 'attribution required' may reduce the reusability and thereby the number of citations. If possible, use a computer-readable and computer actionable license.

✓ b. Yes

4.a.6.b.1. Will you store the licenses with the data at all time?

It is very likely that data will be moved and copied. At some point, people may lose track of the origins. It can be helpful to have the licenses (of course as open as possible) stored in close association with the data.

✓ b. Yes

4.a.7. How will you do file naming and file organization?

Putting some thoughts into file naming can save a lot of trouble later.

✓ a. Explore

4.a.7.a.1. Did you make a SOP (Standard Operating Procedure) for file naming?

It can help if everyone in the project uses the same naming scheme.

✓ a. No

4.a.7.a.2. Will you be keeping the relationships between data clear in the file names?

Advice: Use the same identifiers for sample IDs etc. throughout the entire project.

- ✓ b. Yes

4.a.7.a.3. Will all the metadata that is embedded in the file names also be available in the proper metadata?

The file names are very useful as metadata for people involved in the project, but to computers they are just identifiers. To prevent accidents with e.g. renamed files metadata information should always also be available elsewhere and not only through the file name.

Also note that if metadata could need to change, embedding it in the file names may require renaming files during the project; and this may have implications for references to those files.

- ✓ b. Yes, all metadata is also explicitly available elsewhere

4.a.7.a.4. Will you be using persistent identifiers to refer to data within the project?

🔗 **External Links:** [The Handle System](#), [Handbook on Persistent Identifiers](#)

Especially for large projects, referring to data internally via a persistent identifier system can be helpful as such, a system can help to keep track of data that moves to a new location.

- ✓ b. Yes

4.a.8. How will you be keeping track of the "provenance" of the data?

Data analysis is normally done step-by-step. It is essential to make sure all steps are properly documented, otherwise results will not be reproducible. Re-users of the data also need this information to decide whether the data can be used for their purpose.

- ✓ c. We use other arrangements

4.a.8.c.1. What other arrangements?

- ✓ Redape DataVerse

4.a.9. Will you be documenting the data with W3C PROV provenance?

🔗 **External Links:** [W3C Prov primer](#)

The W3C Prov standard documents processes (workflow) that were used to produce a resource. This can be used to document e.g. the software (including version) and parameters you use to analyze the data. Will your documentation follow the W3C Prov standard?

- ✓ a. No

4.a.10. Will you use a workflow system that automatically keeps track of the steps in the analysis?

Some workflow systems automatically keep track of which steps were done in data analysis and what options were selected. This can help document the data for reproducibility.

- ✓ b. Yes

5. How is the ownership of the collected data arranged?

- ✓ c. All data will be owned by the institute

6. Will you reuse already generated datasets, owned by someone other than your group?

 **External Links:** [See also](#)

Please choose the general data type categories that best fit the data that you plan to collect/generate during the course of the project

i. Other

6.i.1. Please specify

7. Will you reuse already generated datasets, owned by your group?

V. Data documentation and metadata

This chapter deals with the information needed by people who will re-use your data

Summary

Answered indication

Answered (current phase) 2 / 4 50.00 %

Answered 2 / 4 50.00 %

Metrics

Metric Score

Openness 0.00

1. Can all of your data become completely open immediately?

Some data may be subject to a temporary embargo, or need to stay closed for specific reasons.

a. No

2. How will the consistency and quality of the collected data be controlled?

Instrument calibration

Data recording standards

Data peer review

File naming conventions

Electronic Laboratory Notebook

Training

Repeated measurements

Data entry validation

Database

Folder structure

Laboratory Information Management System

Other

3. What metadata standards will the data comply with?
4. Please, describe your data documentation system if other

VI. Ethics and legal issues

Chapter text

Summary

Answered indication

Answered (current phase)	2 / 2	100.00 %
Answered	3 / 3	100.00 %

Metrics

Metric	Score
Accessibility	1.00
Openness	1.00

1. How will ethical issues be addressed and handled?
 - ✓ We have the necessary permission to obtain, process, preserve and share the data
2. Will you work with sensitive research data?
 - ✓ a. Yes
3. Which license will apply to the data?
 - 🔗 **External Links:** [Creative Commons Licenses](#)
 - ✓ c. CC-BY-NC-SA

VII. Data preservation, sharing and reusability

In this chapter, questions on data reusability and preservation are addressed.

Summary

Answered indication

Answered (current phase)	3 / 4	75.00 %
Answered	3 / 4	75.00 %

1. Where will your data be stored?
 - ✓ LMB
 - ✓ Other
2. Data sharing limitations
3. Repositories where the generated data will be shared
 - ✓ Redape, NCBI, LMB

4. Please, describe your data preservation method if none of the above

- ✓ Tape backup, Redape backup

APÊNDICE C – Plano de Gestão de Dados do Laboratório Multiusuário de Bioinformática da Embrapa Agricultura Digital - a partir do *DMP Canvas Generator*

Bloco 1. Informação administrativa

1.1 Equipe do projeto (criar um bloco para cada membro)

1.1.1 Nome

1.1.2 E-mail

1.1.3 ORCID

1.1.4 Filiação

1.1.5 Função (pode ter mais do que uma função)

Opções: pessoa de contato; coletor de dados; curador de dados; gerente de dados; *data steward*; distribuidor; editor; produtor; líder de projeto; gerente de projeto; membro de projeto; pesquisador; detentor de direitos autorais; detentor de direitos patrimoniais; patrocinador; supervisor; outro.

1.2 Projeto(s) de Pesquisa

1.2.1 Nome do Projeto

1.2.2 Sigla

1.2.3 Abstract

1.2.4 Data de início do projeto

1.2.5 Data de término do projeto

1.2.6 Possui financiamento

Opções:

() Não

() Sim

Se SIM:

1.2.6.1 Nome da financiadora

1.2.6.2 Status do financiamento

Opções:

() em planejamento;

() submetido;

() aprovado;

() rejeitado.

1.2.6.3 Número do processo

1.3 Você precisa de um especialista para executar o Plano de Gerenciamento de Dados?

Opções:

Não

Sim

Se SIM:

Opções:

Sim, temos pessoal treinado.

Sim, vamos treinar membros da equipe.

Sim, vamos contratar pessoal com essa capacidade.

1.4 Você necessita de infraestrutura de hardware ou software além da que já é oferecida pelo LMB?

Opções:

Não

Sim

Se SIM:

1.4.1 Qual(is)?

Bloco 2. Reúso de dados

2.1 Existem dados previamente coletados que serão usados nas análises?

Opções:

Não

Sim

Se SIM:

2.1.1. Os dados existentes pertencem ao seu grupo ou a você?

Opções:

Não

Sim

Se NÃO:

2.1.1.1 Qual é a fonte dos dados?

2.1.1.2 Existe licença de uso? Qual?

2.1.2 Os dados estão depositados? Possuem licença?

Opções:

Não

Sim

Se SIM:

2.1.2.1. Qual a licença de uso?

2.1.2.2 Onde os dados estão depositados?

2.2. Você vai combinar dados de diferentes fontes?

Opções:

() Não

() Sim

Se SIM:

2.2.1 Cite as fontes: _____.

Bloco 3. Geração e coleta de dados

3.1 Haverá coleta de dados de amostras humanas?

Opções:

() Não

() Sim

Se SIM:

3.1.1 Existe aprovação do Comitê de Ética em Pesquisa?

3.2 Haverá coleta de dados de amostras de vertebrados?

Opções:

() Não

() Sim

Se SIM:

3.2.1 Existe aprovação do Comitê de Ética no uso de animais?

Opções:

() Não

() Sim

3.3 Haverá dados de amostras envolvendo Patrimônio Genético e Conhecimento Tradicional Associado?

Opções:

() Não

() Sim

Se SIM:

3.3.1 Existe cadastro de acesso no Sistema Nacional de Gestão do Patrimônio Genético e do Conhecimento Tradicional Associado (SisGen)?

Opções:

Não

Sim

3.4 Dos tipos de dados descritos abaixo assinale os que serão coletados/gerados

Opções:

Dados genômicos;

Dados transcriptômicos;

Dados metabolômicos;

Dados de biologia molecular;

Dados proteômicos;

Outros (descrever): _____.

3.5 Qual será o formato dos dados?

3.5.1 Esse é um formato padrão de dados?

Opções:

Não

Sim

Se NÃO:

3.5.1.1 Qual é o formato dos dados?

3.5.1.2 Por que você está usando um formato que não é padrão?

3.5.2 Este formato de dados permite compartilhamento e armazenamento de dados de longo prazo?

Opções:

Não

Sim

Se NÃO:

3.5.2.1 Você vai converter os dados para um formato mais adequado ao armazenamento?

Opções:

Não

Sim

3.6 Qual é o volume dos dados que você usará? (pode especificar em quantidade - Gigabytes ou número de arquivos com um tamanho aproximado de Gigabytes).

3.7 Quais vocabulários/ontologias/terminologias existentes você vai usar?

3.8 Qual padrão/esquema de metadados será utilizado para descrever os dados?

3.9 Como está definida a propriedade dos dados coletados?

Opções:

- () Nosso dado não tem proprietário legal;
- () Todos os dados serão de propriedade do pesquisador líder do projeto;
- () Todos os dados serão de propriedade da empresa/instituto/fundação/universidade;
- () Nós temos um acordo que define a propriedade intelectual.
- () Outro (definir): _____.

3.10 Como a consistência e a qualidade dos dados gerados serão controladas? Qual é o desenho experimental da coleta de dados?

Bloco 4. Processamento dos Dados

4.1 Quais análises serão realizadas? (O que você está querendo responder com as análises?)

Opções:

- () Genômica comparativa;
- () Metagenômica;
- () Microarranjo;
- () Expressão Diferencial;
- () QTL;
- () Outros (descrever): _____.

4.2 Quais ferramentas serão usadas para o processamento dos dados brutos? (nome, link para a página do software e versão dos softwares usados).

não obrigatório

4.3 Qual estrutura de arquivos será utilizada no processamento de dados? (você pode definir um identificador único para o projeto que contenha a sigla, a data e a versão da documentação (Exemplo: YYYYMMDD_ProjetoIDX_V01).

Na árvore de arquivos gere um arquivo README descrevendo a hierarquia do diretório.

Bloco 5. Preservação de Dados

5.1 Nome do *dataset* (criar um bloco para cada *dataset*).

5.2 Descrição do *dataset*.

5.3 Identificador do *dataset* (Ex.: DOI, URL).

5.4 Esse *dataset* será publicado?

Opções:

Não

Sim

Se SIM:

5.4.1 Onde você vai publicar esses dados?

5.4.2 Por quanto tempo esses dados serão mantidos?

5.5 Você vai querer armazenar seus dados (*cold storage*) para preservação de longo prazo durante o projeto?

Opções:

Não

Sim

Se SIM:

5.5.1 O seu projeto dispõe de recursos para o custeio de armazenamento?

5.5.2 Você vai armazenar seus dados em local diferente do LMB ou Redape?

Opções:

Não

Sim

Se SIM:

5.5.2.1 Onde? Explique: _____.

5.6 Você vai armazenar seus dados (*cold storage*) após o término do projeto?

Opções:

Não

Sim

Bloco 6. Acesso aos dados gerados a partir das análises

6.1 Você vai trabalhar com a filosofia de publicar seus dados e deixá-los tão abertos quanto possível?

Opções:

Não

Sim

Se SIM:

6.1.1 Você já escolheu o repositório para publicar seus dados?

6.2. Os seus dados podem se tornar públicos imediatamente?

Opções:

Não

Sim

Se NÃO:

6.2.1 Há razões legais para não publicar seus dados?

6.2.2 Há razões comerciais para não publicar seus dados?

Opções:

Não.

Há expectativa de registro de patentes.

Outras razões comerciais.

6.2.3 Há outras razões para seus dados não serem publicados?

6.3 Que licenças serão aplicadas aos dados?

Opções:

CC-BY-NC-ND;

CC BY-ND Attribution-Noderivs;

CC-BY-NC-SA;

CC BY-NC Attribution-Noncommercial;

CC-BY-SA;

CC-BY;

CC0 (Public Domain).

Observação:

Quando o PGD for gerado a partir do sistema DSW que está sendo implementado localmente no LMB, as versões produzidas terão a data da criação e as datas registradas de todas as novas versões criadas.

Lista de siglas

DOI – Digital Object Identifier

DSW – Data Stewardship Wizard

Embrapa – Empresa Brasileira de Pesquisa Agropecuária

FAIR – Findable, Accessible, Interoperable, Reusable

GCIA – Grupo de Pesquisa em Computação Científica, Engenharia da Informação e Automação

LMB – Laboratório Multiusuário de Bioinformática da Embrapa

ORCID – Open Researcher and Contributor ID

PD&I – Pesquisa, Desenvolvimento e Inovação

PGD – Plano(s) de Gestão de Dados

PGDam – Plano de Gestão de Dados Acionável por Máquina

PGDIC – Política de Governança de Dados, Informação e Conhecimento da Embrapa RmVAC – Projeto

SEG (Sistema Embrapa de Gestão) que tem por objetivo identificar antígenos para o desenvolvimento de vacina contra o carrapato bovino *Rhipicephalus microplus*

URL – Uniform Resource Locator

