

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO**

Lucia Helena Cunha Vidal

**Planos de gestão de dados acionáveis por máquina no ecossistema
da pesquisa brasileira**

Porto Alegre
2024

Lucia Helena Cunha Vidal

Planos de gestão de dados acionáveis por máquina no ecossistema da pesquisa brasileira

Dissertação apresentada como requisito a obtenção do título de mestre em Ciência da Informação pelo Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Rio Grande do Sul.

Orientador: Prof. Dr. Rene Faustino Gabriel Junior.

Coorientadora: Prof^a Dr^a Caterina Marta Groposo Pavão.

Porto Alegre
2024

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Dr. Carlos André Bulhões Mendes

Vice-Reitora: Prof^a Dr^a Patrícia Helena Lucas Pranke

FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO

Diretora: Prof^a Dr^a Ana Maria Mielniczuk de Moura

Vice Diretora: Prof^a Dr^a Vera Regina Schmitt

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

Coordenador: Prof. Dr. Thiago Henrique Bragato Barros

Coordenadora Substituto: Prof^a Dr^a Jussara Borges de Lima

Dados internacionais de catalogação na publicação
Lucia Helena Cunha Vidal CRB 10/2035

A648v Vidal, Lucia Helena Cunha
Planos de gestão de dados acionáveis por máquina no ecossistema da
pesquisa brasileira / Lucia Helena Cunha Vidal. – 2024.
167 f.

Orientador : Rene Faustino Gabriel Júnior.

Coorientadora : Caterina Marta Groposo Pavão.

Dissertação (Mestrado) – Universidade Federal do Rio Grande do Sul,
Faculdade de Biblioteconomia e Comunicação, Programa de Pós-
Graduação em Ciência da Informação, Porto Alegre, RS, 2024.

1. Gestão de dados de pesquisa. 2. Planos de gestão de dados
acionáveis por máquina. 3. Ecossistemas de pesquisa. I. Título. II. Gabriel
Junior, Rene Faustino (orientador) III. Pavão, Caterina Marta Groposo
(coorientadora).

CDU 001.103-047.37

Programa de Pós-Graduação em Ciência da Informação

Rua Ramiro Barcelos, 2705, Prédio 22201

Porto Alegre – RS - CEP: 90035-007

Telefone: (51) 3308-5067

E-mail: ppgcin@ufrgs.br

Lucia Helena Cunha Vidal

**Planos de gestão de dados acionáveis por máquina no ecossistema da
pesquisa brasileira**

Dissertação desenvolvida como requisito para a obtenção do título de mestre em
Ciência da Informação pelo Programa de Pós-Graduação em Ciência da Informação
da Universidade Federal do Rio Grande do Sul.

Orientador: Prof. Dr. Rene Faustino Gabriel Júnior.
Coorientadora: Prof^a Dr^a Caterina Marta Groposo Pavão.

Aprovada em ___/___/___

COMISSÃO EXAMINADORA:

Prof. Dr. Rene Faustino Gabriel Junior (Orientador)
Membro Interno (PPGCIN/UFRGS)

Prof^a Dr^a Caterina Marta Groposo Pavão (Coorientadora)
Membro Interno (PPGCIN/UFRGS)

Prof. Dr. Jesús Pascual Mena-Chalco
Membro Externo (CMCC/UFABC)

Prof^a Dr^a Ana Carolina Simionato Arakaki
Membro Externo (PPGCI/UFSCar)

Prof. Dr. Rafael Port da Rocha
Membro Interno (PPGCIN/UFRGS)

Prof^a Dr^a Ana Maria Mielniczuk de Moura
Suplente (PPGCIN/UFRGS)

Porto Alegre
2024

AGRADECIMENTOS

O mestrado é uma experiência incrível! Prof. Rene e Prof^a Caterina, muito obrigada por todas as oportunidades que vocês me deram! Que sorte a minha encontrar vocês!

Meu mais profundo reconhecimento à Universidade Federal do Rio Grande do Sul por proporcionar a qualidade e os instrumentos para o meu percurso acadêmico e pessoal. É uma honra e uma alegria pertencer a essa comunidade!

Aos colegas que compartilharam comigo a felicidade e as dúvidas do mestrado, obrigada!

A todos com quem dividi essa experiência, que cruzaram meu caminho, que compartilhei os cafés e que agora carrego comigo, obrigada! Cabem muitos aqui, dos professores ao pessoal do bar, obrigada!

Pedala Rosinha, muito obrigada, maninha!

À minha mãe Helena, *in memoriam*, um infinito obrigada!

E mais uma vez, muito, muito, muito obrigada a dupla fantástica que se completaram na minha pesquisa: Prof. Rene e Prof^a Caterina!

RESUMO

Os planos de gestão de dados são instrumentos de planejamento de uso de dados para projetos de pesquisa. Discute a importância dos dados de pesquisa e destaca a necessidade de um plano de gestão adequado para garantir disponibilidade, acesso, processamento e reuso dos dados. Expõe a falta de integração dos planos de gestão de dados com outros atores no ecossistema de pesquisa, o que dificulta o acompanhamento do ciclo de vida dos dados. Propõe analisar a viabilidade da utilização de metodologias e ferramentas interoperáveis com as diversas bases de dados para tornar os planos de gestão de dados acionáveis por máquina e integrados com o ciclo de vida dos dados no ecossistema da pesquisa brasileira. É esperado que essa pesquisa possa contribuir para a otimização dos recursos financeiros, humanos e físicos, tornando os dados mais acessíveis e úteis para outros pesquisadores, instituições, agências de fomento e sociedade em geral. De natureza básica, abordagem qualitativa e classificada quanto aos objetivos em descritiva e explicativa, tem procedimentos de pesquisa bibliográfica e documental. As 14 ferramentas para de elaboração de planos de gestão de dados identificadas foram analisadas de acordo com os dez princípios para acionamento por máquina, destaca as características do DMPonline e DMPTool. Identifica os principais atores do ecossistema da pesquisa brasileira envolvidos em revistas científicas e agregadores, programas de pós-graduação, repositórios digitais, instituições de pesquisa e pesquisadores, bibliotecas digitais, bases de patentes, conjuntos de dados de pesquisa, produção acadêmica e científica, agências de fomento, projetos de pesquisa e identificadores persistentes. Verifica que as ferramentas DMPonline e DMPTool atendem o maior número de princípios da RDA e têm condições de interoperabilidade entre os PGDs e os sistemas de pesquisa. Conclui que o uso desses modelos pode contribuir para a automatização da gestão de dados e para o fluxo de informações em projetos de pesquisa, além de identificar ferramentas que servem de base para a elaboração de outras ferramentas de elaboração de planos de gestão de dados e a importância dos identificadores persistentes para os planos de gestão de dados.

Palavras-chave: Planos de gestão de dados acionáveis por máquina; Ecossistema de pesquisa; Ferramentas de elaboração de planos de gestão de dados; Atores do ecossistema de pesquisa brasileira.

ABSTRACT

Data management plans are tools for planning the use of data for research projects. It discusses the importance of research data and highlights the need for an appropriate management plan to ensure data availability, access, processing, and reuse. It exposes the lack of integration of data management plans with other actors in the research ecosystem, which makes it difficult to monitor the data life cycle. It proposes to analyze the feasibility of using interoperable methodologies and tools with the various databases to make data management plans machine-actionable and integrated with the data life cycle in the Brazilian research ecosystem. It is expected that this research can contribute to the optimization of financial, human, and physical resources, making data more accessible and useful to other researchers, institutions, funding agencies, and society in general. It is basic in nature, with a qualitative approach and classified according to its objectives as descriptive and explanatory, and has bibliographic and documentary research procedures. It analyzes tools for developing DMPs that meet the ten principles for machine-action. It identifies 14 tools for developing machine-actionable data management plans and analyzes their characteristics, highlighting DMPonline and DMPTool. It identifies the main actors in the Brazilian research ecosystem involved in scientific journals and aggregators, postgraduate programs, digital repositories, research institutions and researchers, digital libraries, patent databases, research datasets, academic and scientific production, funding agencies, research projects, and persistent identifiers. It verifies that the DMPonline and DMPTool tools meet the largest number of RDA principles and have interoperability conditions between DMPs and research systems. It concludes that the use of these models can contribute to the automation of data management and the flow of information in research projects, in addition to identifying tools that serve as a basis for the development of other tools for developing data management plans and the importance of persistent identifiers for data management plans.

Keywords: Machine-actionable data management plans; Research ecosystems; Tools for preparing data management plans; Actors in the Brazilian research ecosystem.

LISTA DE FIGURAS

Figura 1 – Diagrama do plano de gestão de dados.....	15
Figura 2 – Ecossistema do universo de desenvolvimento BrCris.....	26
Figura 3 – Modelo do Ciclo de Vida dos Dados de Pesquisa.....	31
Figura 4 – Gestão do ciclo de vida dos dados.....	32
Figura 5 – Interações das partes interessadas.....	42
Figura 6 – Exemplo de PGD acionáveis por máquina em um micro-ecossistema ...	45
Figura 7 – Exemplo de metadados para a propriedade de “contato” em um PGD ...	46
Figura 8 – Dez princípios para maDMPs.....	50
Figura 9 – Ferramenta DMPonline	72
Figura 10 – Ferramenta DMPTool.....	74
Figura 11 – Ferramenta DMPRoadmap	75
Figura 12 – Ferramenta EasyDMP	77
Figura 13 – Ferramenta Data Stewardship Wizard.....	79
Figura 14 – Ferramenta Research Data Management Organiser.....	81
Figura 15 – Ferramenta UQ Research Data Manager.....	83
Figura 16 – Ferramenta DataWiz	84
Figura 17 – Ferramenta ezDMP	86
Figura 18 – Ferramenta OpenDMP	88
Figura 19 – Ferramenta Argos.....	89
Figura 20 – Ferramenta FioDMP	90
Figura 21 – Ferramenta DMP OPIDoR.....	92
Figura 23 – Ferramenta PGD-BR	93

LISTA DE QUADROS

Quadro 1 – Tipos de dados	29
Quadro 2 – Dez princípios para planos de gestão de dados acionáveis por máquina	54
Quadro 3 – Etapas do processo de análise documental	69
Quadro 4 – Síntese das ferramentas de elaboração de planos de gestão de dados	95
Quadro 5 – Atendimento dos princípios RDA de acordo com as ferramentas de PGDs	106
Quadro 6 – Categorias e atores do ecossistema da pesquisa brasileira	108
Quadro 7 – Relação de cada princípio RDA com os atores do ecossistema da pesquisa brasileira	123

LISTA DE SIGLAS

- BDTD - Biblioteca Digital de Teses e Dissertações
- BRAPCI - Base de Dados em Ciência da Informação
- BrCris - *Brazilian Current Research Information System*
- Capex - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
- CBL - Câmara Brasileira do Livro
- CDL - *California Digital Library*
- CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico
- CNRS - *Centre National de la Recherche Scientifique*
- CRIS - *Current Research Information System*
- DATASUS - Departamento de Informática do Sistema Único de Saúde
- DCAT - *Data Catalog Vocabulary*
- DCC - *Digital Curation Centre*
- DBTD - Biblioteca Digital de Teses e Dissertações
- DDI - *Data Documentation Initiative*
- DMP - *Data Management Plan*
- DOAJ - *Directory of Open Access Journals*
- DOI - *Digital Object Identifier*
- DSW - *Data Stewardship Wizard*
- ENEM - Exame Nacional do Ensino Médio
- FAIR - *Findability, Accessibility, Interoperability, Reusability*
- FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo
- FCT - Fundação para Ciência e Tecnologia
- Fiocruz - Fundação Oswaldo Cruz
- GDP - Gestão de Dados de Pesquisa
- IANA - *Internet Assigned Numbers Authority*
- Ibict - Instituto Brasileiro de Informação em Ciência e Tecnologia
- IEDA - *Interdisciplinary Earth Data Alliance*
- INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
- INPI - Instituto Nacional da Propriedade Industrial
- ISBN - *International Standard Book Number*

ISNI - *International Standard Name Identifier*
ISSN - *International Standard Serial Number*
LISTA - *Information Science, Technology Abstracts*
maDMP - *Machine-actionable Data Management Plan*
MEC - *Ministério da Educação*
NDLTD - *Networked Digital Library of Theses and Dissertations*
NSF - *National Science Foundation*
NWO - *Netherlands Organisation for Scientific Research*
OECD - *Organisation for Economic Co-operation and Development*
OMPI - *Organização Mundial da Propriedade Intelectual*
ORCID - *Open Researcher and Contributor ID*
PGD - *Plano de gestão de dados*
PID - *Persistent identifier*
RCAAP - *Repositórios Científicos de Acesso Aberto de Portugal*
RDA - *Research Data Alliance*
RDMO - *Research Data Management Organiser*
RNP - *Rede Nacional de Ensino e Pesquisa*
SciELO - *Scientific Electronic Library*
SNSF - *Swiss National Science Foundation*
TDK - *Template Development Kit*
TIC - *Tecnologia da Informação e Comunicação*
TV - *Televisão*
UC3 - *University of California Curation Center*
UFRGS - *Universidade Federal do Rio Grande do Sul*
UNAM - *Universidad Nacional Autónoma de México*
Unesp - *Universidade Estadual Paulista*
Unicamp - *Universidade Estadual de Campinas*
URL - *Uniform Resource Locator*
UQRDM - *UQ Research Data Management*
W3C - *World Wide Web Consortium*

SUMÁRIO

1	INTRODUÇÃO	13
1.1	TEMA.....	17
1.2	DELIMITAÇÃO DO TEMA	17
1.3	PROBLEMA.....	19
1.4	OBJETIVO GERAL.....	19
1.5	OBJETIVOS ESPECÍFICOS.....	19
1.6	JUSTIFICATIVA.....	19
2	FUNDAMENTAÇÃO TEÓRICA	21
2.1	ECOSSISTEMA DA PESQUISA BRASILEIRA.....	21
2.2	DADOS DE PESQUISA.....	27
2.3	CICLO DE VIDA DOS DADOS DE PESQUISA.....	30
2.4	GESTÃO DE DADOS	33
2.5	PRINCÍPIOS FAIR	35
2.6	PLANO DE GESTÃO DE DADOS	37
2.6.1	Planos de Gestão de Dados Estáticos.....	40
2.6.2	Planos de Gestão de Dados Acionáveis por Máquina.....	41
2.6.3	Dez princípios da <i>Research Data Alliance</i> para planos de gestão de dados acionáveis por máquina.....	44
2.6.4	Componentes e propriedades do Padrão Comum <i>Research Data Alliance</i> para planos de gestão de dados acionáveis máquina.....	55
3	METODOLOGIA E PROCEDIMENTOS.....	65
4	APRESENTAÇÃO E ANÁLISE DOS RESULTADOS	71
4.1	FERRAMENTAS DE ELABORAÇÃO DE PLANOS DE GESTÃO DE DADOS ACIONÁVEIS POR MÁQUINA.....	71
4.2	PLANOS DE GESTÃO DE DADOS ACIONÁVEIS POR MÁQUINA E OS PRINCÍPIOS DA <i>RESEACH DATA ALLIANCE</i>	99
4.3	ATORES DO ECOSSISTEMA DA PESQUISA BRASILEIRA.....	107
4.4	APLICABILIDADE DOS PLANOS DE GESTÃO DE DADOS ACIONÁVEIS POR MÁQUINA NO ECOSSISTEMA DA PESQUISA BRASILEIRA	121
4.4.1	Primeiro princípio para planos de gestão de dados acionáveis por máquina da <i>Research Data Alliance</i>.....	128

4.4.2	Segundo princípio para planos de gestão de dados acionáveis por máquina da <i>Research Data Alliance</i>	130
4.4.3	Terceiro princípio para planos de gestão de dados acionáveis por máquina da <i>Research Data Alliance</i>	131
4.4.4	Quarto princípio para planos de gestão de dados acionáveis por máquina da <i>Research Data Alliance</i>	132
4.4.5	Quinto princípio para planos de gestão de dados acionáveis por máquina da <i>Research Data Alliance</i>	134
4.4.6	Sexto princípio para planos de gestão de dados acionáveis por máquina da <i>Research Data Alliance</i>	136
4.4.7	Sétimo princípio para planos de gestão de dados acionáveis por máquina da <i>Research Data Alliance</i>	138
4.4.8	Oitavo princípio para planos de gestão de dados acionáveis por máquina da <i>Research Data Alliance</i>	140
4.4.9	Nono princípio para planos de gestão de dados acionáveis por máquina da <i>Research Data Alliance</i>	142
4.4.10	Décimo princípio para planos de gestão de dados acionáveis por máquina da <i>Research Data Alliance</i>	143
5	CONSIDERAÇÕES	147
	REFERÊNCIAS	152
	ANEXO – Plano de gestão de dados: Planos de gestão de dados acionáveis por máquina no ecossistema da pesquisa brasileira	162

1 INTRODUÇÃO

A Ciência Aberta fundamentada na colaboração e no compartilhamento de informações torna o conhecimento acessível e reutilizável para todos (United National Educational, Scientific and Cultural, 2022). Enquanto potencial agregador traz mudanças significativas para o modo de produzir e acessar a ciência. Nesse contexto, o uso, o compartilhamento e o reuso dos dados de pesquisa ganham relevância na ciência.

A importância dos dados de pesquisa tem sido reconhecida não apenas para validar os resultados publicados, mas também para permitir o seu reuso, estimulando novas pesquisas a partir dos dados compartilhados, otimizando e potencializando os recursos financeiros, físicos e humanos (Pavão; Rocha; Gabriel Júnior, 2018). Com a intensificação das abordagens para reuso dos dados, a gestão ganha relevância, pois promove o conhecimento, a descoberta, a inovação e o reuso dos dados. A gestão correta confere ainda mais valor aos dados de pesquisa, transformando-os em um ativo importante para a comunidade científica.

No sentido de valorizar os dados e bem geri-los, há um significativo movimento mundial em torno dos princípios FAIR (*Findability, Accessibility, Interoperability, Reusability*) aplicados à gestão dos dados de pesquisa. As boas práticas de gestão compreendem uma série de etapas que asseguram estratégias adequadas para que os dados possam ser corretamente disponibilizados, facilmente encontrados, acessados, processados e, finalmente, reutilizados.

A gestão é um processo que busca garantir a qualidade e a confiabilidade dos dados depositados e disponibilizados em um repositório. Via de regra, a gestão abrange as diversas etapas do ciclo de vida dos dados, incluindo a coleta, o armazenamento, a organização, a preservação e o compartilhamento dos dados. Como forma de garantir a integridade e a confiabilidade dos dados, além de promover a transparência e a reprodutibilidade das pesquisas, os planos de gestão de dados (PGDs) são instrumentos essenciais para a gestão.

Planos de gestão de dados são documentos que descrevem todo o ciclo de vida dos dados de maneira formal e estática. Os planos de gestão deveriam ser elaborados mesmo antes do início de uma pesquisa e atualizados durante e após a realização da pesquisa. Idealmente, os planos deveriam ser orientados para todo o ciclo de vida dos dados e a cobertura de todas as etapas, desde o planejamento até

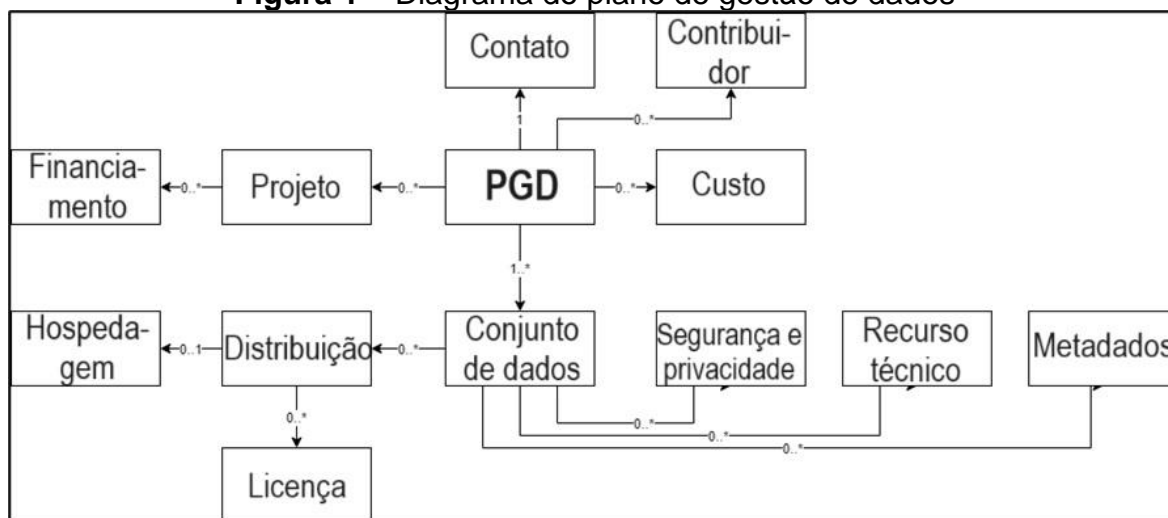
a disponibilização para o reuso dos dados (Almeida, 2019). A documentação dos dados garante detalhes sobre a coleta, o armazenamento e o compartilhamento. Assim, um plano de gestão deve descrever como os dados serão gerenciados ao longo do ciclo de vida dos dados de pesquisa.

Os planos de gestão de dados, em geral, apresentam uma limitação significativa em termos de acessibilidade e tendem a ter uma vida útil mais curta, uma vez que não são facilmente atualizáveis ou adaptáveis às necessidades dos sistemas de informação. No entanto, até o momento, geralmente os PGDs não resultam em processos de planejamento sólidos para a gestão de dados, não integrando informações de várias partes interessadas para reduzir a carga significativa de tempo, esforço e conhecimento técnico de pesquisadores e resultar em planos de gestão de dados eficazes e de alta qualidade.

Dessa forma, o problema que se apresenta é a falta de integração dos planos de gestão de dados com outros dispositivos e serviços, o que faz com que não acompanhem os processos que envolvem o ciclo de vida dos dados dentro do ecossistema da pesquisa. Assim, torna-se necessário buscar a viabilidade de utilizar metodologias e ferramentas interoperáveis com os PGDs acionáveis por máquina para a gestão da pesquisa e que atendam as diversas fases do ciclo de vida dos dados. O acionamento por máquina dá-se pela interoperabilidade, com a estruturação consistente de informações para que máquinas ou computadores previamente programados possam acessar e ler informações (Data, Documentation and Initiative, 2023).

Em um contexto integrado, como nos ecossistemas de pesquisa, os planos de gestão de dados podem alcançar todo o seu potencial ao possibilitar o fluxo de informação e automatizar a gestão de dados. Os ecossistemas de pesquisa envolvem a interoperabilidade entre serviços, indicadores e pesquisadores, reunindo e organizando as informações da pesquisa científica, disponibilizando-os de forma segura e transparente. Então, um ecossistema abrange a interação, colaboração e integração entre diversos atores que assumem funções específicas ao longo da pesquisa (Figura 1).

Figura 1 – Diagrama do plano de gestão de dados



Fonte: traduzido de Miksa, Walk, Neish (2020).

No Brasil, um modelo de ecossistema de pesquisa é o BrCris. Desenvolvido pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), o BrCris busca possibilitar a troca de informações entre os diversos atores do ecossistema da pesquisa científica brasileira. O BrCris visa promover um modelo único de organização que facilite a recuperação e o tratamento dessas informações (Carvalho Segundo *et al.*, 2021). Portanto, o ecossistema BrCris objetiva conectar e integrar os agentes da pesquisa brasileira, relacionando todas as etapas do ciclo de vida dos dados de pesquisa, reunindo pesquisadores, agências financiadoras, instituições de pesquisa, projetos, entre outros.

Dessa forma, buscar soluções para o uso e a integração dos planos de gestão de dados com os atores do ecossistema da pesquisa no Brasil faz-se necessário. Assim, o objetivo de analisar a viabilidade de associação dos PGDs acionáveis por máquina com o ciclo de vida dos dados no ecossistema da pesquisa brasileira foi dividido em identificar na literatura os modelos *machine-actionable Data Management Plans* (maDMP); analisar as características e integrações dos maDMP com os princípios para acionamento por máquina dos PGDs; identificar os atores do ecossistema da pesquisa brasileira; e verificar a aplicabilidade do maDMP no ecossistema da pesquisa brasileira.

Portanto, foram identificadas na literatura 14 ferramentas de elaboração de planos de gestão de dados acionáveis por máquina: DMPonline, DMPTool, DMPRoadmap, EasyDMP, Data Stewardship Wizard (DSW), Research Data Management Organiser (RDMO), UQ Research Data Manager (UQRDM), DataWiz,

ezDMP, OpenDMP, ARGOS, FioDMP, DMP OPIDoR, PGD-BR. As características e particularidades de cada ferramenta foram exploradas para encontrar consonância com os dez princípios da *Research data Alliance*. Igualmente, a partir da literatura, foram identificados 27 atores do ecossistema da pesquisa brasileira que desempenham funções nos fluxos de automatização dos planos de gestão de dados. Os principais atores são revistas científicas e agregadores, programas de pós-graduação, repositórios digitais, instituições de pesquisa e pesquisadores, bibliotecas digitais, bases de patentes, conjuntos de dados de pesquisa, produção acadêmica e científica, agências de fomento, projetos de pesquisa e identificadores persistentes. Dessa forma, com a análise das ferramentas de elaboração de planos de gestão de dados com os dez princípios para PGD acionáveis por máquina da RDA foi possível verificar a aplicabilidade dos planos de gestão de dados acionáveis por máquina no ecossistema da pesquisa brasileira. Duas ferramentas se destacaram com o maior número de princípios observados, a DMPonline e a DMPTool, observando-se que ambas têm condições de trocar informações entre os PGDs e os sistemas de pesquisa, incorporando os planos de gestão de dados de pesquisa aos fluxos de trabalho da gestão dos dados de pesquisa.

Diante do contexto apresentado, esta pesquisa de natureza básica, classificada quanto aos objetivos em descritiva e explicativa, com o auxílio da análise de conteúdo e procedimentos de pesquisa bibliográfica e documental, espera contribuir para discussões e políticas acerca da integração entre os planos de gestão de dados e os ecossistemas de pesquisa de forma acionável por máquina, potencializando recursos aplicados. Justifica-se por colaborar com a gestão de dados, tornando-os mais acessíveis e úteis para outros pesquisadores, instituições, agências de fomento e sociedade em geral. O plano de gestão de dados desta pesquisa está disponibilizado publicamente na plataforma da DMPTool sob o DOI 10.48321/D1F49A1123¹ (Anexo) e os dados coletados foram depositados em formato aberto no repositório de dados de pesquisa CedapDados².

¹<https://doi.org/10.48321/D1F49A1123>.

²<https://cedapdados.ufrgs.br/dataset.xhtml?persistentId=hdl:20.500.11959/dt/25>.

1.1 TEMA

Os dados de pesquisa referem-se às informações produzidas ou adquiridas durante a condução de um estudo. A qualidade dos dados de pesquisa influencia diretamente sua utilização potencial em futuras pesquisas. A gestão adequada dos dados não apenas promove a transparência no processo científico, mas também facilita a validação dos resultados, fomentando o compartilhamento e reutilização dos dados.

No ciclo de vida dos dados de pesquisa, uma das primeiras etapas é o seu planejamento, a escolha das fontes de informações e as ferramentas que serão utilizadas na pesquisa, considerando-se também os aspectos éticos dos dados coletados (Sant'Ana, 2016; Anjos; Dias, 2019).

Os planos de gestão de dados representam um instrumento que auxiliam o pesquisador, mesmo antes do início de sua pesquisa, a planejar e tratar os dados de forma segura e confiável. Todos os aspectos que envolvem a coleta, o tratamento, a análise, a preservação e o acesso aos dados devem ser claramente descritos nos planos de gestão de dados para que, assim, possam ser disponibilizados a todas as partes interessadas, sejam agências de financiamento, bibliotecários, administradores de repositórios de dados ou pesquisadores (Salazar *et al.*, 2022).

Percebe-se então, que a crescente demanda pelo compartilhamento dos dados promove a discussão sobre a gestão correta dos dados de pesquisa por meio de boas práticas como forma de garantir a integridade da pesquisa, sua replicação e reuso dos dados. Assim, esta pesquisa tem como tema a integração dos planos de gestão de dados acionáveis por máquinas com outros sistemas de informação em todo o ciclo de vida dos dados de pesquisa.

1.2 DELIMITAÇÃO DO TEMA

Planos de gestão de dados proporcionam ao pesquisador refletir sobre a pesquisa desde a fase de elaboração do projeto de pesquisa. Documentam o ciclo de vida dos dados de pesquisa e fornecem detalhes sobre a coleta, o armazenamento, o acesso, o compartilhamento e a reprodutibilidade de resultados (Veiga, 2022). Os planos de gestão possibilitam a abertura dos dados de pesquisa, minimizam o risco

de perda dos dados e são parte fundamental da boa prática científica. Assim, ao elaborar um PGD, o pesquisador zela pela qualidade de seus dados em todas as fases da pesquisa (Sant'Ana, 2019).

Os planos de gestão de dados são documentos textuais em formato livre criados com auxílio de processadores de texto ou ferramentas *online* e devem responder perguntas que orientarão todo o percurso do ciclo de vida dos dados (Miksa; Oblasser; Rauber, 2021). Geralmente percebidos como um processo administrativo e com uso limitado, pode ser visto como uma tarefa exaustiva, pois não é um processo automatizado (Miksa *et al.*, 2019).

Apesar de algumas áreas de pesquisa já terem incorporado os planos de gestão de dados à prática de compartilhamento dos dados, “[...] outras comunidades científicas ainda estão hesitantes e a principal força motriz para a adoção de PGDs nessas comunidades geralmente vem de agências de financiamento [...]” (Salazar *et al.*, 2022, não paginado).

Salazar *et al.* (2022) expõem ainda que a falta de treinamento na identificação e uso das ferramentas de elaboração de planos de gestão de dados e a falta de compreensão dos benefícios do gerenciamento dos dados para o pesquisador individual são os principais motivos para a falta de interesse nos PGDs. Esses problemas tornam o plano de gestão de dados um documento estático, difícil de ser atualizado e por isso podendo rapidamente tornar-se obsoleto e sua utilização impraticável durante e ao final da pesquisa.

Um plano de gestão de dados bem elaborado deve conter informações a respeito dos tipos de dados coletados, análise automatizada de arquivos para descrição de dados, suporte automatizado para seleção de licenças, armazenamento, depósito e publicação dos dados ou estimativa de custos para recursos de gestão de dados (Miksa *et al.*, 2019). Dessa forma, a delimitação do tema será:

A problemática da elaboração dos planos de gestão de dados, uma vez que esses planos não têm integração com outros dispositivos e por isso não acompanham os processos que envolvem o ciclo de vida dos dados dentro do ecossistema da pesquisa brasileira.

1.3 PROBLEMA

No contexto apresentado, questiona-se **qual a de associação dos planos de gestão de dados acionáveis por máquina com o ciclo de vida dos dados no ecossistema da pesquisa brasileira?**

1.4 OBJETIVO GERAL

Com a finalidade de responder o problema desta pesquisa, o objetivo geral consiste em **analisar a associação dos planos de gestão de dados acionáveis por máquina com o ciclo de vida dos dados no ecossistema da pesquisa brasileira.**

1.5 OBJETIVOS ESPECÍFICOS

Para aprofundar a pesquisa e atingir o objetivo geral, esse será esmiuçado desdobrando-se em objetivos específicos conforme segue:

- a) identificar na literatura os modelos *machine-actionable Data Management Plans* (maDMP);
- b) analisar as características e integrações dos maDMP com os princípios da *Research Data Alliance* para acionamento por máquina dos PGDs;
- c) identificar os atores do ecossistema da pesquisa brasileira;
- d) verificar a aplicabilidade do maDMP no ecossistema da pesquisa brasileira.

1.6 JUSTIFICATIVA

A produção de dados de pesquisa é cada vez mais profusa em diversas áreas do conhecimento. No entanto, muitas vezes os dados não são corretamente gerenciados, resultando em perda de informação valiosa, replicação inadequada dos resultados de pesquisa ou desperdício financeiro.

Esta pesquisa é motivada pelo interesse em contribuir para o entendimento de como os planos de gestão de dados acionáveis por máquina podem melhorar a gestão

de dados e torná-los mais acessíveis e úteis para outros pesquisadores, instituições, agências de fomento e a sociedade em geral. Assim, justifica-se por colaborar para que os planos de gestão de dados deixem de ser um documento estático, elaborado apenas como requisito para aprovação de uma pesquisa, e passe a ser um documento vivo, ativo e realmente útil para a gestão das pesquisas no Brasil.

Espera-se que esta pesquisa possa contribuir para discussões acerca da integração entre os planos de gestão de dados e os ecossistemas de pesquisa de forma acionável por máquina. Assim, a disposição em realizar esta investigação parte do anseio em estimular a reflexão e o debate sobre políticas e ações que norteiem a questão dos planos de gestão de dados e o entendimento do uso do modelo conceitual de maDMP com sistemas de informação durante o ciclo de vida dos dados de pesquisa.

Então, destaca-se a importância deste estudo na promoção da qualificação das pesquisas científicas, melhorando o compartilhamento, a colaboração entre os pesquisadores e otimizando os recursos financeiros, humanos e físicos. Assim, favorecendo uma gestão de dados eficiente e eficaz.

De forma particular, a autora tem interesse no tema deste estudo, uma vez que participa do grupo de trabalho do projeto de pesquisa Rede de Repositórios de Dados de Pesquisa da Rede Nacional de Ensino e Pesquisa (RNP). A temática trabalhada no grupo de pesquisa suscita questões pertinentes à problemática desta pesquisa, motivando a sua realização.

2 FUNDAMENTAÇÃO TEÓRICA

A fundamentação teórica, baseada na literatura científica, tem foco na gestão de dados de pesquisa e divide-se em capítulos que abordam aspectos do tema na intenção de responder o problema e atingir os objetivos propostos. Os capítulos Ecossistema da Pesquisa Brasileira, Dados de Pesquisa, Ciclo de Vida dos Dados de Pesquisa, Gestão de Dados, Princípios FAIR e Plano de Gestão de Dados, assim como os subcapítulos Planos de Gestão de Dados Estáticos, Planos de Gestão de Dados Acionáveis por Máquina, Dez Princípios da *Research Data Alliance* para Planos de Gestão de Dados Acionáveis por Máquina e Componentes e Propriedades do Padrão Comum *Research Data Alliance* para Planos de Gestão de Dados Acionáveis por Máquina do referencial teórico buscam esclarecer e dar sustentação à escolha do tema da investigação.

2.1 ECOSSISTEMA DA PESQUISA BRASILEIRA

Ecossistema é um conceito oriundo da Ecologia que abrange e agrega comunidades em um determinado local e que estão em interação entre si e entre os demais elementos de seu ambiente (O Eco, 2014). Na literatura científica, o termo "ecossistema" é geralmente definido como um sistema natural ou artificial composto por uma comunidade de organismos interagindo entre si e com seu ambiente físico. Um ecossistema pode também se referir a um conjunto interconectado de sistemas e componentes técnicos que trabalham juntos para alcançar um objetivo comum. No contexto de tecnologia e negócios, isso pode incluir tecnologias, empresas, instituições e seus usuários, todos interagindo dentro de um ambiente compartilhado para promover inovação e desenvolvimento (Moore, 1993).

No contexto social e científico, um ecossistema é um sistema complexo composto por vários atores, como instituições, pesquisadores, políticas públicas, e práticas que interagem e afetam a produção e disseminação de conhecimento. Essas interações podem influenciar e moldar o ambiente científico e acadêmico, refletindo um sistema interdependente e dinâmico (Kizilhan; Bal, 2016). Em ambas definições, o conceito central de "ecossistema" é a interdependência e a interação entre os componentes do sistema, seja no contexto natural, tecnológico ou social.

Desse modo, Oliveira (2019, p. 192-193) afirma que: “É neste sentido que se torna crucial entender estes atores a partir de uma rede de interdependência, observando as transformações que ocorrem a partir dos fatores externos ao ambiente científico”. Para Oliveira (2019), no ecossistema científico, um ator é qualquer indivíduo ou entidade que participa e influencia o processo de produção, disseminação e validação do conhecimento científico. Dessa forma, os sistemas de informação, que armazenam, gerenciam e trocam metadados contextuais com outros sistemas, conectando seus atores, são atores desse mesmo ecossistema (Vidal, 2023).

Um ecossistema de pesquisa envolve integração, colaboração e interação entre diferentes atores que desempenham papéis específicos na pesquisa científica. Para Ponte, Mierzejewska e Klein (2017), alguns dos principais atores de um ecossistema de pesquisa incluem:

- a) pesquisadores: são os principais produtores de conhecimento em um ecossistema de pesquisa. Eles conduzem experimentos, coletam e analisam dados e publicam resultados em revistas científicas;
- b) instituições de pesquisa e universidades: essas instituições fornecem suporte e recursos para os pesquisadores, incluindo financiamento, instalações de laboratório e acesso a equipamentos e tecnologia;
- c) mercados editoriais: são responsáveis pela revisão, edição e publicação de artigos científicos em revistas especializadas;
- d) políticas públicas: governos e outras organizações podem usar a pesquisa científica para informar políticas públicas, programas e projetos;
- e) agências de fomento: organizações que fornecem financiamento para pesquisas. Isso pode incluir agências governamentais, fundações privadas, empresas e indivíduos;
- f) redes sociais acadêmicas e não acadêmicas: são responsáveis pela disseminação de conhecimento científico para o público em geral. Isso pode incluir museus, centros de ciência, canais de TV, revistas e jornais;
- g) além de leitores, cidadãos comuns, políticas científicas, esferas institucionais de decisão política na Ciência, Tecnologia e Inovação etc.

Assim, os atores de um ecossistema trabalham em conjunto para promover a pesquisa científica e contribuir para um contexto científico dinâmico e interconectado.

Esse conjunto de atores diversos favorece o todo, garantindo a troca e a confiabilidade das informações.

No contexto dos dados de pesquisa, os ecossistemas contribuem de forma mais ampla, uma vez que envolvem um conjunto de diferentes atores e extensão (Oliveira, 2019). Dessa forma, fazem convergir os esforços de pesquisadores, bibliotecários, universidades, instituições de pesquisa e financiamento.

Em uma perspectiva de grandes volumes de dados, Demchenko, Laat e Membrey (2014) apontam o ecossistema como um *hub* de técnicas e componentes direcionados aos dados e sua aplicação – armazenamento, processamento, visualização e disponibilização.

Iniciativas estão sendo desenvolvidas para a criação de sistemas de informação de pesquisa, os chamados ecossistemas CRIS. Esses ecossistemas suportam o armazenamento e a gestão dos dados no contexto da pesquisa científica e integram os atores envolvidos no uso e compartilhamento de dados.

Neste tipo de ecossistema, os atores estão conectados e fazem sua contribuição em aspectos pontuais (Oliveira; Lóscio, 2018). Os autores Oliveira e Lóscio (2018, p. 70) afirmam ainda que “[...] todos podem e devem colaborar, quer sejam usuários ou produtores de dados”.

Ubaldi (2013) corrobora ao argumentar que os benefícios aguardados do compartilhamento de dados apenas serão atingidos por meio da criação de um ecossistema de dados no qual produtores e consumidores de dados dialoguem entre si. Desta forma, o principal fator de sucesso dos ecossistemas são a comunicação e a cooperação entre os implicados.

Os ecossistemas propostos para os dados de pesquisa são baseados em *Current Research Information System* e são definidos como:

[...] um ecossistema informacional que agrega informações sobre pesquisas correntes e passadas. Com esse procedimento, possibilita fazer cruzamentos entre as informações agregadas, de maneira a obter novas, de forma consolidada e com maior significância. O CRIS possibilita a obtenção de cenários sobre as pesquisas correntes e passadas, configurando-se como importante ferramenta de gestão. (Cooperação Internacional, 2017, não paginado).

Para alcançar todas as vantagens de uso desse sistema para a gestão das atividades de pesquisa, Alvares *et al.* (2023) elencam uma série de recursos desejáveis em um CRIS: identificadores globais, código aberto, alimentação

automática de dados, conectividade, vários tipos de exportação de dados, interoperabilidade com sistemas internos e externos, taxonomias, modelo de dados comum, *networking* entre usuários, possibilidade de atualização dos próprios perfis de usuários, uso de vocabulários controlados e desambiguação automática de autor.

Um exemplo reconhecido internacionalmente é o sistema europeu euroCRIS, criado em 2002, conta com mais de 200 membros, principalmente de países pertencentes à União Europeia, e reúne parceiros estratégicos, especialistas em informação e sistemas de informação de pesquisa (Cooperação Internacional, 2023; Simons, 2019). O euroCRIS busca oportunizar a cooperação e o compartilhamento de conhecimento na comunidade de pesquisa, promovendo a interoperabilidade das informações de pesquisa com base em um padrão comum (The International Organisation for Research Information, 2020a):

Um dos objetivos do euroCRIS é criar e oferecer serviços para a comunidade internacional de informações de pesquisa. Esses serviços destinam-se mais especificamente a compartilhar experiência e conhecimento sobre sistemas de informação de pesquisa e apoiar e inspirar a comunidade RIS na criação e/ou uso de CRISs. (The International Organisation for Research Information, 2020a, não paginado, tradução nossa).

A estrutura fornecida pelo euroCRIS permite a comutação de informações entre “[...] pesquisadores, gerentes e administradores de pesquisa, conselhos de pesquisa, financiadores de pesquisa, empreendedores e organizações de transferência de tecnologia. O euroCRIS também se dirige aos meios de comunicação e ao público” (The International Organisation for Research Information, 2020b, não paginado, tradução nossa). Com o propósito de consolidar a infraestrutura euroCRIS, foram propostas Melhores Práticas para abordar aspectos específicos do CRIS em diferentes áreas (Castro; Mennielli, 2014).

No Brasil, um exemplo de ecossistema de pesquisa foi iniciado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia, em 2014. O modelo BrCris foi inspirado no modelo proposto por Portugal (Dias *et al.*, 2021), o PTCRIS, financiado pela Fundação para Ciência e Tecnologia (FCT) e membro do euroCRIS (Pinto *et al.*, 2021). Em 2016, o euroCRIS e o Ibict assinaram um Memorando de Entendimento projetando o estabelecimento e a promoção de colaboração para implementar uma infraestrutura de informação científica no Brasil (PTCRIS, 2016). E em 2020, o Ibict

implementou o projeto BrCris para o desenvolvimento de um ecossistema integrado e interoperável, buscando:

[...] viabilizar o intercâmbio de dados entre os diferentes atores do ecossistema da pesquisa científica brasileira, envolvendo o sistema de financiamento, os conteúdos científicos gerados, os dados tecnológicos, os repositórios de teses e dissertações, bem como os demais conteúdos científicos abertos pelas instituições brasileiras de ensino e pesquisa, mas também gerando interfaces de recuperação e tratamento dos dados. (Carvalho Segundo *et al.*, 2021, p. 1).

O modelo brasileiro busca a interoperabilidade entre sistemas com base em quatro dimensões, de acordo com Carvalho Segundo *et al.* (2021, p. 1):

A primeira é voltada à indexação de pesquisadores brasileiros nos sistemas de informação e suas especialidades, contando com um conjunto de dados da Plataforma Lattes e Portal de Dados Abertos da CAPES. A segunda está destinada à organização de dados sobre as revistas científicas que compõem o universo brasileiro de publicação e disseminação de pesquisas científicas [...]. A terceira via está destinada às patentes e seus dados relacionados [...]. A quarta e última está voltada para a consolidação de um sistema de recomendação de especialistas [...] (Carvalho Segundo *et al.*, 2021, p. 1-2).

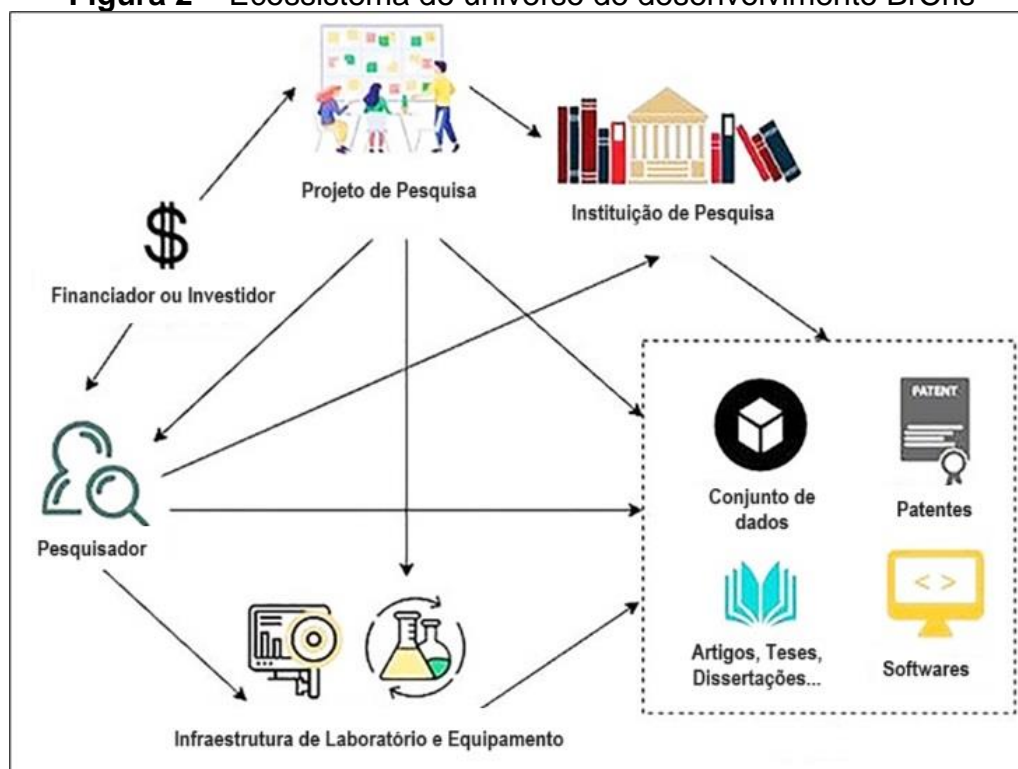
O BrCris é um modelo de infraestrutura tecnológica que procura agregar os sistemas de informação existentes e propor outros que atendam “[...] todas as necessidades de informação do ecossistema, e descrever tanto o modelo para integração dos sistemas de informação que compõem o ecossistema (CRIS locais), quanto o CRIS Nacional, que se torna o nó integrador do ecossistema” (Cooperação Internacional, 2017, não paginado). A iniciativa BrCris propõe-se a armazenar, gerenciar e trocar metadados contextuais “[...] para atividades de pesquisa financiadas por um órgão governamental ou realizadas em uma organização executora de pesquisa” (Cooperação Internacional, 2017, não paginado), assim:

O projeto BrCRIS é um ecossistema de sistemas, com foco em executar a interoperabilidade de dados científicos e tecnológicos. Este sistema integra a troca de dados de vários atores, como agências de financiamento, bases de dados científicas e tecnológicas, repositórios institucionais, plataformas governamentais, infraestrutura de pesquisa, bases de teses e dissertações entre outros atores. (Cooperação Internacional, 2017, não paginado).

O ecossistema proposto pelo Ibict objetiva executar a interoperabilidade de dados científicos e tecnológicos existentes. Por meio da integração e intercâmbio de

dados de diversos atores (Figura 2), como pesquisadores, projetos, instituições e agências financiadoras, repositórios e bancos de dados, ferramentas governamentais, suportes de pesquisa etc. (Dias *et al.*, 2021).

Figura 2 – Ecosistema do universo de desenvolvimento BrCris



Fonte: traduzido de Pinto *et al.* (2022).

Na medida em que o ecossistema BrCris abarca todos os atores implicados e interessados na gestão dos dados, a infraestrutura tem potencial para oferecer a documentação correta, ou seja, o registro sólido e completo de cada parte implicada no ciclo.

Conforme bem expõem Pinto *et al.* (2022, p. 3) a respeito do BrCris:

Esse sistema integra a troca de dados de diversos atores, como agências financiadoras, bancos de dados científicos e técnicos, repositórios institucionais, plataformas governamentais, infraestrutura de pesquisa, bancos de dados de teses e dissertações.

De acordo com Dias *et al.* (2021, p. 276), o modelo proposto pelo Ibict seria alimentado “[...] principalmente pelos repositórios institucionais, bibliotecas digitais de teses e dissertações, revistas eletrônicas de acesso aberto e repositórios de dados de pesquisa brasileiros, reunidos nos portais da BDTD e do OasisBr”. Portanto, o BrCris

estabeleceria um modelo único de organização da informação científica de todo o ecossistema da pesquisa brasileira (Carvalho Segundo *et al.*, 2022).

Diante do exposto, o modelo BrCris serve de representação do ecossistema de pesquisa brasileira, exemplificando de forma clara e objetiva os diversos atores que se conectam no ciclo de vida dos dados de pesquisa.

2.2 DADOS DE PESQUISA

A Ciência Aberta chega à sociedade com práticas e ferramentas que não somente divulgam e estimulam a participação social, como também propõem uma forma mais colaborativa de fazer ciência: “Trata-se de refletir sobre os desafios que essas mudanças trazem às dinâmicas [...], seus valores e práticas, e sobre os novos olhares que se impõem para melhor compreender e lidar com tais desafios” (Albagli; Maciel; Abdo, 2015, p. 9). A Ciência Aberta engloba uma série de iniciativas, dentre elas, os dados de pesquisa.

Os dados de pesquisa provocam uma mudança significativa na práxis científica mundial. As tecnologias e as ferramentas digitais levam os dados de pesquisa às novas formas de aplicação e vêm sendo usados como recurso de validação das pesquisas científicas (Borgman, 2015), embora ainda não seja uma prática muito comum (Salazar *et al.*, 2022).

Definir dados de pesquisa implica diretamente no contexto ao qual estão inseridos, por isso sua definição pode sofrer variações. Pavão (2022, não paginado) salienta que: “Os dados da pesquisa abrangem todas as disciplinas e, portanto, a definição pode variar em relação às abordagens dos diferentes atores (pesquisadores, instituições, órgão de fomento, etc.) e aos diferentes contextos nacionais”. Em uma tentativa mais geral, Silva (2019, p. 21) estabelece que dados “[...] são as informações registradas ou produzidas através de qualquer forma ou meio durante o decurso de uma pesquisa”.

Um conceito amplamente utilizado quando se trata de dados de pesquisa, tanto no Brasil quanto internacionalmente, é o da *Organisation for Economic Co-operation and Development* – OECD (Organisation for Economic Co-Operation and Development, 2007), que define dados de pesquisa como:

[...] registros factuais (pontuações numéricas, registros textuais, imagens e sons) usados como fontes primárias para pesquisa científica e que são comumente aceitos na comunidade científica como necessários para validar os resultados da pesquisa. Um conjunto de dados de pesquisa constitui uma representação sistemática e parcial do assunto que está sendo investigado. (Organisation for Economic Co-Operation and Development, 2007, p. 13).

Para a Fiocruz (Fundação Oswaldo Cruz, 2022, não paginado), dados de pesquisa: “São os objetos informacionais digitais comumente registrados e aceitos pela comunidade científica como necessários para validar os resultados de pesquisa, sendo dados produzidos para diversos fins e utilizados em pesquisas”. Observa-se, entretanto, que a amplitude do entendimento de dados de pesquisas sugere um conceito complexo que se manifesta em múltiplas formas.

Conforme Dudziak (2016), dados de pesquisa:

[...] incluem: fatos e estatísticas recolhidas para posterior referência ou análise, documentos (texto, Word), planilhas (Excel, etc), cadernos de laboratório, cadernos de campo, diários, questionários, transcrições, fitas de áudio, fitas de vídeo, fotografias, filmes, sequências de proteínas ou genéticos, respostas de teste, slides, artefatos, amostras, coleção de objetos digitais adquiridos e gerados durante o processo de pesquisa, conteúdos de banco de dados (vídeo, áudio, texto, imagens), modelos, algoritmos, scripts, arquivos de log, software de simulação, metodologias e fluxos de trabalho, procedimentos operacionais, padrões e protocolos. (Dudziak, 2016, tradução nossa, não paginado).

Já Semeler e Pinto (2019, p. 115) afirmam que “[...] dados de pesquisa científica são o resultado de qualquer investigação sistemática que envolva processos de observação, experimentação ou simulação de procedimentos de pesquisa científica”.

Outra definição usualmente encontrada na literatura é a de Sayão e Sales (2020) que esclarecem que:

[...] dado de pesquisa é todo e qualquer tipo de registro coletado, observado, gerado ou utilizado no âmbito da pesquisa científica, que pode ser interpretado, tratado e aceito como evidência pela comunidade científica e necessário para analisar, validar e produzir resultados de pesquisa. (Sayão; Sales, 2020, p. 32).

Os autores prosseguem:

Essa definição é suficientemente ampla para abarcar todas as possibilidades de dados de pesquisa. No entanto, é importante destacar que para que o registro se configure como dado de pesquisa, ele precisa ser tratado e aceito pela comunidade. Isso coloca em pauta a obrigação de uma gestão mínima, com atribuição de metadados que

tornem o registro compreensível para a comunidade científica. (Sayão; Sales, 2020, p. 32).

Uma característica importante é que os dados de pesquisa se distinguem de acordo com diferentes critérios e são classificados conforme o procedimento de coleta, o caráter primário ou secundário, o grau de estruturação, o nível de cobertura e o formato. Conforme a *National Science Foundation* (2007), os dados podem ser tipificados de acordo com o Quadro 1.

Quadro 1 – Tipos de dados

CLASSIFICADOS	TIPOS	EXEMPLOS
De acordo com a sua origem	Dados observacionais	São obtidos por meio de observações diretas, que podem ser associadas a lugares e tempo específicos e não podem ser coletados uma segunda vez
	Dados computacionais	São resultados da execução de modelos computacionais ou de simulações
	Dados experimentais	São provenientes de situações controladas em bancadas de laboratórios
De acordo com a sua natureza	Números, imagens, vídeos ou áudios, <i>software</i> , algoritmos, equações, animações ou modelos e simulações	
De acordo com a fase da pesquisa	Dados brutos, crus ou preliminares	São dados que vêm diretamente dos instrumentos científicos
	Dados derivados	São resultados do processamento ou combinação de dados brutos ou de outros dados
	Dados canônicos ou dados referenciais	São coleções de dados consolidados e arquivados geralmente em grandes centros de dados, por exemplo, sequência genética, estrutura química etc.

Fonte: adaptado de *National Science Foundation* (2007).

Sayão e Sales (2015, p. 7) afirmam “[...] que os dados são gerados para diferentes propósitos, por diferentes comunidades acadêmicas e científicas e por meio de diferentes processos intensifica ainda mais essa percepção de diversidade”.

Nota-se, portanto, que os dados de pesquisa têm grande valor para a comunidade científica e devido à crescente demanda, Pavão (2022) expõe que pesquisadores e instituições aspiram ter acesso aos dados de pesquisa em si. Nesse sentido, o movimento dos dados de pesquisa encaminha-se para a sua abertura de acesso:

A divulgação dos dados de pesquisa vem sendo reconhecida como de grande importância, não só para validar os resultados obtidos e publicados, mas propiciar seu reuso, como também para impulsionar novas pesquisas e socializar o conhecimento. (Pavão; Rocha; Gabriel Junior, 2018, p. 3).

Percebe-se então, que os dados de pesquisa se mostram centrais no desenvolvimento de pesquisas porque são a matéria-prima obtida por meio dos processos de coleta de dados para a representação de uma realidade (Silva, 2019). Assim, é necessário observar todas as fases do ciclo de vida dos dados de pesquisa para garantir a qualidade, a transparência, a reprodutibilidade e a preservação dos dados.

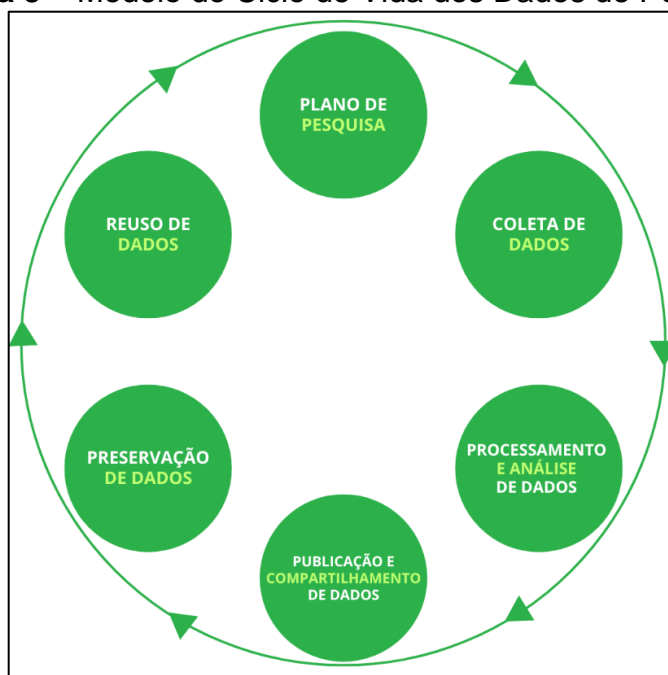
2.3 CICLO DE VIDA DOS DADOS DE PESQUISA

A vida dos dados se estende por todo o projeto da pesquisa ampliando-se para além desses projetos. Todas essas etapas fazem parte do ciclo de vida dos dados de pesquisa e contam com a colaboração de diversos personagens. Esse ciclo é fundamental para que os conjuntos de dados possam ser perfeitamente analisados, atualizados, citados e reutilizados por outros pesquisadores.

Cada uma das etapas do ciclo de vida dos dados de pesquisa é essencial para a gestão responsável dos dados. Planejar como serão coletados, arquivados, preservados e, posteriormente, compartilhados responde questões fundamentais para a gestão de dados. Esse planejamento confere qualidade e sustentabilidade ao longo do tempo (Anna; Dias; Maculan, 2019).

De acordo com o *UK Data Service* (2019), o modelo do ciclo de vida dos dados de pesquisa pode ser representado como um movimento cíclico (Figura 3), sendo cada fase importante para o desenvolvimento da pesquisa científica.

Figura 3 – Modelo do Ciclo de Vida dos Dados de Pesquisa



Fonte: traduzido de *UK Data Service* (2019).

O modelo proposto pelo *UK Data Service* (2019) sugere seis fases, cada uma delas aborda as etapas do ciclo de vida e são elas:

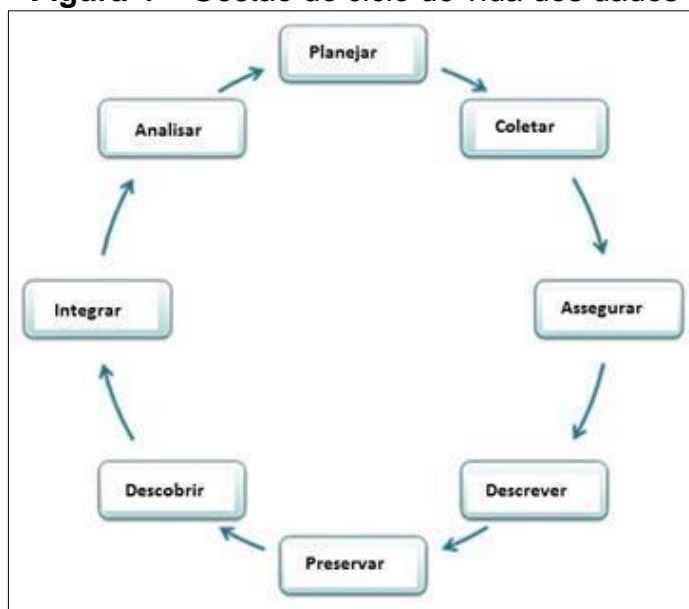
- planejamento da pesquisa: inclui o projeto de pesquisa, o plano de gestão de dados, o consentimento para compartilhamento, o plano para a coleta de dados;
- coleta de dados: coleta dos dados com metadados e ou aquisição de dados de terceiros;
- processamento e análise dos dados: registrar, digitalizar, transcrever e traduzir os dados, verificar, validar, limpar, anonimizar, descrever e documentar os dados, gerenciar e armazenar os dados, produzir resultados de pesquisa e citar fontes;
- publicação e compartilhamento de dados: estabelecer os direitos autorais, criar documentação do usuário, criar metadados para descoberta, selecionar acesso apropriado aos dados, publicar e compartilhar dados, promover os dados;
- preservação dos dados: migração dos dados para o melhor formato e/ou mídia, armazenamento e *backup* dos dados, criação da documentação de preservação, preservação e curadoria dos dados;

- f) reuso dos dados: realizar análises secundárias, pesquisas de acompanhamento, realizar revisões de pesquisa, examinar as descobertas e usar os dados para ensinar e aprender.

Todas as fases do ciclo de vida dos dados acima descritas pelo modelo *UK Data Service* (2019) observam aspectos importantes para assegurar a integridade, a qualidade, a transparência, a reprodutibilidade, o compartilhamento e a preservação dos dados. Dessa forma, a gestão do ciclo de vida dos dados ao longo de todo o seu processo garante o valor dos dados de pesquisa.

Sob a perspectiva dos pesquisadores, Dudziak (2016) elenca oito passos para a gestão do ciclo de vida dos dados de pesquisa (Figura 4): 1) planejar o dado; 2) gerar ou coletar o dado de pesquisa; 3) assegurar e aprimorar a qualidade do dado; 4) descrever o dado a partir de metadados apropriados; 5) preservar em repositórios adequados; 6) possibilitar a descoberta; 7) integrar com outros dados; 8) analisar para reiniciar o ciclo.

Figura 4 – Gestão do ciclo de vida dos dados



Fonte: traduzido de Dudziak (2016).

Portanto, observa-se que o ciclo de vida dos dados de pesquisa pode ser representado sob a perspectiva da pesquisa e dos próprios dados, as suas características e fases de desdobramento ao longo da pesquisa, bem como sob ótica

da sua gestão, valendo-se de instrumentos e metodologias que garantam a preservação, disponibilização e reuso dos dados.

2.4 GESTÃO DE DADOS

A intensificação do uso de dados leva às mudanças significativas no modo de criar, tratar, analisar, preservar, acessar e reutilizar os dados disponíveis. Os dados podem ser gerados a qualquer momento da pesquisa, por esse motivo, documentá-los adequadamente é essencial para a sua preservação, acesso, uso e reutilização. Assim, a gestão otimiza recursos, garante a reprodutibilidade e viabiliza novas pesquisas com segurança e confiabilidade (Torino *et al.*, 2022).

A gestão de dados de pesquisa começa ainda na fase inicial do projeto de pesquisa, seguindo na fase de coleta de dados, passando pelo plano de gestão de dados. Com o propósito de disponibilizar da melhor forma possível os conjuntos de dados produzidos pelos pesquisadores e atendendo aos princípios de excelência, além de responder as demandas das agências de financiamento e apoio, a gestão de dados de pesquisa se ocupa da organização dos dados, compreendendo todas as fases de seu ciclo de vida, desde o projeto, a coleta, o tratamento, a preservação e a disseminação para o acesso e reuso (Silva, 2019).

Evidencia-se, portanto, a necessidade de acompanhamento dos processos durante todo o ciclo de vida dos dados, gerando um documento contendo a descrição dos dados e metadados definidos que serão produzidos durante a coleta dos dados na pesquisa.

Segundo Silva (2019), os dados precisam ser organizados e entendidos, pois:

Requerem um processo contínuo que identifique as dimensões, as categorias, as tendências, os padrões e as relações, revelando seu significado. Este processo é complexo e implica um trabalho de redução, organização e interpretação de dados, que se inicia previamente na fase exploratória e que continua durante todo o ciclo de pesquisa. (Silva, 2019, p. 3).

O adequado planejamento, organização e documentação do processo de coleta, preservação e compartilhamento dos dados possibilita, para além da pesquisa em que foram produzidos, o prolongamento do ciclo de vida dos conjuntos de dados.

Wilkinson *et al.* (2016) salientam que a gestão dos dados é uma prática estratégica para o conhecimento, descoberta, inovação, posterior integração e reuso dos dados.

Gestão de dados de pesquisa é “[...] um conjunto de práticas para lidar com a informação coletada e criada durante a pesquisa [...]” (Universidade Estadual Paulista, 2022, não paginado) que garantem a integridade e a reprodutibilidade das pesquisas. Para Almeida (2019, p. 17), gestão de dados de pesquisa “[...] inclui todo o ciclo de vida dos dados desde o planejamento, coleta, análise, preservação e compartilhamento visando o reuso [...]”.

A Fiocruz (Fundação Oswaldo Cruz, 2022, não paginado) sintetiza bem as duas definições acima ao estabelecer no Glossário da Ciência Aberta que gestão de dados de pesquisa é o: “Conjunto de práticas de gestão voltadas para o tratamento de dados de pesquisa durante o seu ciclo de vida; inclui todos os aspectos de manutenção, compartilhamento, segurança e preservação”.

As etapas do processo de gestão dos dados envolvem diferentes atores, perpassando pesquisadores, bibliotecários, arquivistas, curadores e serviços especializados nos diferentes estágios do gerenciamento. Este envolvimento interdisciplinar na gestão dos dados permite que todo o ciclo seja adequadamente concebido e executado.

A gestão de dados competente propicia a operacionalização de recursos, dá suporte ao pesquisador e visibilidade aos dados. Ademais, garante o armazenamento, a preservação, o compartilhamento e a confiabilidade e, ainda, a disponibilização à sociedade, promovendo o conhecimento científico. A gestão começa ainda na fase inicial do projeto de pesquisa, seguindo na fase de coleta de dados, chegando ao plano de gestão de dados. Cada uma dessas fases de produção, coleta e guarda de dados, exige requisitos específicos que devem ser observados de acordo com o processo implicado. Assim, conforme afirma Ribeiro, Oliveira e Araújo (2019, p. 163): “A gestão adequada de dados [...] pode apresentar-se como potencializadores para gerar o conhecimento científico, e como consequência favorecer avanços para a pesquisa científica e para a sociedade [...]”.

Para Rosa, Arakaki e Furnival (2022, p. 9-10), “[...] tanto o depósito de dados num repositório correto quanto à elaboração de um PGD como peças centrais à Gestão de Dados de Pesquisa (GDP) no geral, que se preocupa com os dados de pesquisa desde a sua entrada até a sua saída”, são partes importantes do processo de gerenciamento dos dados.

Com as boas práticas, o cumprimento da função da gestão se estabelece ao preservar e tornar os dados localizáveis e disponíveis para reuso. A função da gestão se dá justamente na importância de identificação, organização, armazenamento, acesso e disponibilização dos conjuntos de dados de pesquisa. A gestão implica pensar no modo como serão armazenados e descritos os dados para que possa, de fato, haver o compartilhamento e reuso, resultando mais à frente, na colaboração com outra importante métrica dos dados de pesquisa, a de citação.

2.5 PRINCÍPIOS FAIR

O acrônimo FAIR significa *findable, accessible, interoperable* e *reusable*, e foram traduzidos para o português como “encontrável”, “acessível”, “interoperável” e “reutilizável”. O termo FAIR foi cunhado em 2014 em um *workshop* com a participação do grupo Elixir³ em cooperação com as instituições *Netherlands eScience Center, Dutch Techcentre for Life Sciences* e Centro Lorentz, na Holanda (Santos, 2019). Os Princípios FAIR foram publicados em 2016, resultante dos esforços para criar um selo de aprovação que garantisse que os dados arquivados pudessem ser encontrados, compreendidos e usados no futuro, tanto por pessoas quanto por máquinas, pela comunidade de estudiosos, bibliotecários, arquivistas, editores e financiadores de pesquisas (FORCE11, 2013).

Os Princípios FAIR (Wilkinson *et al.*, 2016) foram propostos para gestão de dados de pesquisa com o propósito de garantir a utilidade dos conjuntos de dados, bem como dos dados em si, tanto para humanos quanto para máquinas e envolvem quatro facetas principais (Salazar *et al.*, 2022): Localizável, com uso de identificadores únicos e persistentes, além de metadados ricos; Acessível, ou seja, recuperável por seu identificador persistente; Interoperável, através de uma linguagem amplamente aplicável para representação de dados; e Reutilizável, sob uma licença de uso clara e acessível.

Além da coleta, registro e arquivamento adequados, o gerenciamento de dados inclui a noção de “cuidado em longo prazo” de ativos digitais valiosos, com o objetivo de que eles sejam descobertos e reutilizados em pesquisas posteriores, sozinhos ou

³Elixir é uma infraestrutura europeia de ciências da vida que reúne cientistas e institutos de pesquisa: <https://elixir-europe.org/>.

em combinação com novos dados gerados (Wilkinson *et al.*, 2016). Os resultados de uma boa gestão e administração de dados, portanto, são publicações digitais de alta qualidade que facilitam e simplificam esse processo contínuo de descoberta, avaliação e reutilização em estudos posteriores. O que constitui o bom gerenciamento de dados é, ainda, em grande parte indefinido e geralmente fica a critério do proprietário dos dados ou do repositório.

Para que os dados de pesquisa possam ser FAIR, os metadados desempenham um papel importante e devem estar bem descritos e detalhados. Os metadados, informações que contêm os detalhes da organização, da coleta e do pré-processamento dos dados, são fundamentais para os processos adequados de encontrar, usar e citar arquivos em repositórios. Portanto, para os dados serem acessíveis, conforme *Columbia University Irving Medical Center (2022)*, os metadados e os dados devem ser compreensíveis por humanos e máquinas, e os dados devem ser armazenados em um repositório confiável. Para serem interoperáveis, os metadados devem usar uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento, como vocabulários controlados convencionados. Para serem reutilizáveis, os dados e as coleções devem ter uma licença de uso clara e fornecer informações precisas sobre a proveniência.

A *Columbia University Irving Medical Center (2022)* e a *Utrecht University (2022)* convergem quanto aos inúmeros benefícios de tornar os dados FAIR para pesquisadores, comunidades de pesquisa e organizações de pesquisa. Dados FAIR pretendem maximizar o impacto da pesquisa, aumentar a visibilidade e as citações de pesquisas, melhorar a reprodutibilidade e confiabilidade da pesquisa, além de atrair novas parcerias com pesquisadores, empresas, políticas e comunidades, permitindo que novas questões de pesquisa sejam respondidas.

Dados FAIR beneficiam criadores e usuários de dados:

À medida que os humanos dependem cada vez mais do suporte computacional das máquinas, os dados FAIR podem permitir que os sistemas computacionais encontrem, acessem, interoperem e reutilizem dados sem ou com mínima intervenção humana. Em um nível pessoal, os Princípios de Dados FAIR fornecem uma estrutura de gerenciamento de dados para ajudar os pesquisadores a gerenciar seus dados. Além disso, ao compartilhar dados FAIR, os pesquisadores facilitam a descoberta de conhecimento e aumentam a chance de uma possível colaboração, o que é benéfico especialmente para pesquisadores em início de carreira. (Columbia University Irving Medical Center, 2022, não paginado).

É importante ressaltar que existem diferentes níveis para tornar os dados FAIR e nem sempre será possível alcançar todos. Aplicar os princípios FAIR aos dados, mesmo que não na sua totalidade, aumentará a capacidade de localização, acessibilidade, interoperabilidade e reutilização dos dados de pesquisa. Portanto, apesar de os princípios serem relacionados, eles são independentes e podem ser aplicados de forma separada (Wilkinson *et al.*, 2016).

Dados FAIR que atendem aos princípios tornam-se explícitos, aumentando a sua capacidade de descobertas e geração de novos conhecimentos. Observar esses princípios facilita a elaboração do plano de gestão de dados, pois ajudam a descrever a forma de criar, armazenar, compartilhar, gerenciar e preservar dados de pesquisa.

Os princípios FAIR são essenciais para os planos de gestão de dados, pois melhoram a qualidade dos dados, possibilitando a interoperabilidade e a sua reutilização, além de fortalecer a transparência e a confiabilidade da pesquisa.

2.6 PLANO DE GESTÃO DE DADOS

A dificuldade em entender os dados de outros pesquisadores faz com que eles não sejam reutilizados em outras pesquisas ou mesmo permitam a validação das pesquisas apresentadas (Choi; Lee, 2020). Nos últimos anos tem havido um investimento significativo no desenvolvimento e uso de planos de gestão de dados para pesquisa, impulsionado pelo incentivo por parte das agências de financiamento à abertura dos dados antes da concessão de recursos (Kim; Stanton, 2016).

Planos de gestão de dados descrevem os aspectos de todo o ciclo de vida dos dados de pesquisa, desde a descrição dos metadados até a preservação dos dados. O plano assegura que os dados sejam bem gerenciados e preservados para disponibilização e reuso, entretanto, não existe um modelo único para elaboração de um plano de gestão de dados.

Originalmente criados, em 1966, para gerenciar a coleta e análise de dados de projetos aeronáuticos e de engenharia da NASA, os planos documentavam as atividades de pesquisa e os fatores importantes de projetos complexos (Smale *et al.*, 2018). Nas décadas de 1970 e 1980, os planos de gestão foram incorporados nas disciplinas científicas (Smale *et al.*, 2018) e serviam de instrumentos de gestão para

a coleta e análise dos dados em grandes pesquisas, com o objetivo de auxiliar outros pesquisadores da área em pesquisas semelhantes.

Nessas décadas iniciais, os planos de gestão foram utilizados para atender uma necessidade de planejamento para atingir objetivos:

[...] geralmente podem ser vistos como abordando o 'como' da implementação técnica dos processos de gerenciamento de dados de um estudo específico, geralmente nos estágios de coleta ou análise de dados. Tais publicações raramente consideram a questão do 'por que', pois há pouca consideração explícita sobre os motivos que levam os pesquisadores a usar DMPs [...] (Smale *et al.*, 2018, p. 4).

Smale *et al.* (2018) afirmam que até o início dos anos 2000, os planos de gestão de dados eram limitados. Utilizados para projetos de pesquisa de grande complexidade técnica e com a finalidade de registrar a coleta e o processamento dos dados. A partir dos anos 2000, os PGDs ganharam visibilidade devido às pesquisas em meio eletrônico e as políticas econômicas (Smale *et al.*, 2018). Entre as décadas de 2000 e 2010, o uso de planos de gestão de dados foi impulsionado pela adoção de políticas de órgãos e instituições de financiamento à pesquisa.

Nos anos 2000, a internet provocou mudanças significativas na pesquisa científica gerando uma grande produção de dados, o que levou a novos estudos que discutiram e especularam de forma mais ampla as mudanças na natureza da pesquisa e a maneira pela qual as instituições e o governo deveriam responder a essas mudanças (Smale *et al.*, 2018).

A frágil preservação dos dados à época e a falta de planos para armazenamento em longo prazo alertou a comunidade científica. Na tentativa de sanar o problema, estratégias voltadas a governos e instituições de gestão, compartilhamento e curadoria de dados foram propostas por Lord e MacDonald (2003).

Desde outubro de 2003, o Instituto Nacional de Saúde dos Estados Unidos requisita de pesquisadores que buscam subsídios um plano de gestão de dados que explicita a possibilidade de compartilhamento de dados (National Institutes of Health, 2023).

A “Declaração sobre acesso aberto a dados de pesquisa com financiamento público” (Organisation for Economic Co-Operation and Development, 2022), da Organização para Cooperação e Desenvolvimento Econômico, trouxe visibilidade à questão dos dados. O relatório abordou os benefícios econômicos de compartilhar os

dados, imputando a responsabilidade aos pesquisadores em fazer a gestão e, por conseguinte a disponibilização de seus dados.

A *National Science Foundation*, corroborando a prática de demandar planos de gestão de dados para a solicitação de financiamentos, começou em janeiro de 2011 a requerer juntamente as propostas submetidas a ela a inclusão de um plano de gestão de dados (National Science Foundation, 2007).

Percebe-se, principalmente no Reino Unido e nos Estados Unidos, a partir dos anos 2010, movido por órgãos e escritórios de ciência e tecnologia, um incentivo crescente por parte das agências de financiamento pelo compartilhamento dos dados. Passou-se também, a solicitar a elaboração de um plano de gestão de dados baseado no ciclo de vida dos dados aos pesquisadores beneficiados com financiamento.

Assim, evidencia-se que o movimento no sentido de requisitar um plano de gestão de dados aos pesquisadores que demandam financiamento para projetos de pesquisa, conferiu uma nova atribuição, principalmente, às bibliotecas universitárias e instituições dedicadas à pesquisa, demandando serviços especializados para a organização, gestão e acesso à informação (Burnette; Williams; Imker, 2016). Essa demanda impulsionou a criação dos planos de gestão de dados, bem como o desenvolvimento de ferramentas para a elaboração desses planos.

No Brasil, em 2017, a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) passou a demandar um documento complementar: um plano de gestão de dados para pesquisadores interessados em solicitar financiamento por meio da Chamada de Proposta de Pesquisa para o Programa FAPESP de Pesquisa sobre Mudanças Climáticas Globais (Fundação de Amparo à Pesquisa do Estado de São Paulo, 2017a). A partir de então, em 2018, progressivamente, a demanda foi incorporada a outros tipos de apoio, entre elas as de Auxílio à Pesquisa – Regular, Jovem Pesquisador, bolsas de Doutorado e de Pós-doutorado, exigência já praticada por agências públicas e privadas de apoio à pesquisa dos Estados Unidos, Europa e Austrália (Fundação de Amparo à Pesquisa do Estado de São Paulo, 2017b). Diante desse panorama, o Estado de São Paulo foi o pioneiro a implantar o plano de gestão de dados de pesquisa nacionalmente.

Os planos de gestão de dados, normalmente exigidos por instituições de financiamento como condição para a concessão de fundos (Gajbe *et al.*, 2021), são usualmente elaborados a partir de listas de quesitos e questionários *online* e

apresentam quais dados coletar, como descrever, preservar e assegurar a qualidade e definem onde e por quanto tempo armazenar os dados de pesquisa:

Os Planos de Gerenciamento de Dados são documentos de texto em formato livre que descrevem os dados que são usados e produzidos durante as atividades de pesquisa. Eles especificam onde os dados serão arquivados, quais licenças e restrições se aplicam e a quem o crédito deve ser dado, etc. (Miksa; Walk; Neih, 2020, p. 1).

É importante ressaltar que planos de gestão de dados podem variar no conteúdo e orientação, direcionando o foco de acordo com as exigências do solicitante, seja agência de financiamento ou gestor de repositório de dados de pesquisa.

Hoje, discute-se o desenvolvimento de um modelo que permita expressar as informações dos planos de gestão de dados tradicionais de maneira acionável por máquina. Assim, facilita-se a troca de informações entre sistemas que atuam em nome das partes interessadas envolvidas no ciclo de vida da pesquisa, como pesquisadores, financiadores, gestores de repositório, bibliotecários etc.

2.6.1 Planos de Gestão de Dados Estáticos

O relatório de Simms *et al.* (2017) ressalta que os planos de gerenciamento de dados estão se tornando comuns em todo o mundo como resultado das políticas de agências de financiamento que os solicitam com propostas de subsídios, mas não estão sendo empregados de maneira que realmente apoiem o empreendimento de pesquisa. O modelo atual de um PGD, que se caracteriza como um documento estático geralmente criado antes do início de um projeto, apenas contribui para a percepção de que os planos são uma tarefa administrativa desagradável (Simms *et al.*, 2017; Miksa *et al.*, 2019). O que os PGDs realmente são, ou, pelo menos deveriam ser, é parte integrante da prática de pesquisa, já que hoje a maioria das pesquisas em todas as disciplinas envolve dados, códigos e outros componentes digitais.

A Universidade Federal de São Paulo (UNIFESP) (Universidade Federal de São Paulo, 2020, não paginado) caracteriza o *Data Management Plan* (DMP) ou Plano de Gestão de Dados, como “[...] um documento formal que descreve os dados produzidos durante um projeto de pesquisa e descrevem estratégias de gerenciamento de dados que serão implementadas durante e após a fase ativa do

projeto de pesquisa”. Mais especificamente, *Stanford University* (2022), *University College London* (2022) e *European Commission* (2022) definem plano de gestão de dados como um documento escrito que detalha os dados que se espera adquirir ou gerar durante o curso de um projeto de pesquisa, como serão gerenciados, descritos, analisados e armazenados esses dados e quais mecanismos serão usados no final do projeto para compartilhar e preservar os dados.

Como forma de garantir as boas práticas de pesquisa e gestão dos dados o *Digital Curation Centre* (2013) elaborou algumas perguntas que norteiam a construção de um PGD. Essas orientações consideram que e como serão coletados os dados, a documentação, metadados e questões éticas e de direito legal que acompanharão os dados, gerenciamento e preservação em longo prazo, além das questões de compartilhamento.

Os planos de gestão de dados estabelecem um roteiro para os dados, desde o planejamento até a preservação, pois descrevem os dados e as ferramentas empregadas nas investigações científicas. De modo geral, *Aventurier e Alencar* (2016) indicam que um plano de gestão de dados inclui a descrição do projeto de pesquisa, autores e afiliações, exigências das agências de fomento, tipo de dados produzidos, regras para distribuição e reutilização e planos para conservação e arquivamento.

Documentar o ciclo de vida dos dados em um PGD é uma prática que resulta em uma gestão mais eficiente, colaborativa e confiável dos dados de pesquisa. Na medida em que o plano é um documento vivo que deve acompanhar e incorporar as alterações que o curso da pesquisa pode sofrer, torná-los legíveis, tanto para humanos quanto para máquinas, permite que dados e informações sobre pesquisas sejam comunicados e compartilhados entre as partes interessadas (*Datacite*, 2023).

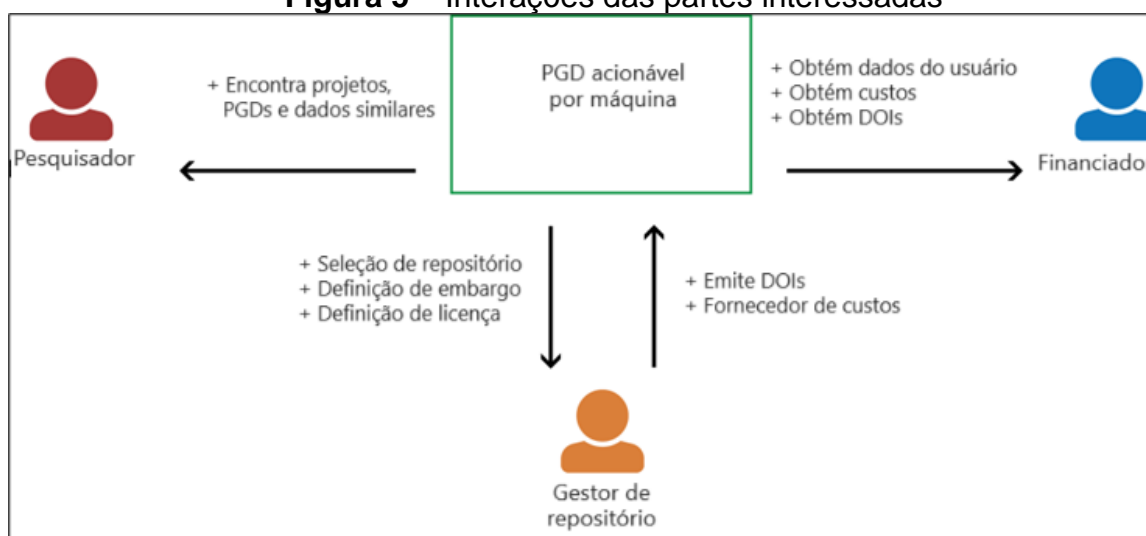
2.6.2 Planos de Gestão de Dados Acionáveis por Máquina

Muitos esforços são direcionados no sentido de tornar os planos de gestão de dados vivos. À medida que o curso da pesquisa avança, o plano de gestão de dados precisa ser alterado para contemplar todas as etapas do ciclo de vida dos dados. Assim, estudos buscam discutir as limitações dos PGDs e fornecer uma forma ativa de representar as informações dos planos de gestão de modo acionável por máquina. Esboços no estudo de *Miksa, Walk e Neish* (2020) apresentam um padrão de estrutura

que contempla a cobertura de diversos casos de uso de PGDs. Esses planos de gestão incorporam informações de todo o ciclo de vida dos dados e da sua gestão, apresentando as ações já realizadas e planejadas.

A pesquisa de Simms *et al.* (2017) revela a necessidade de uma estrutura de planos de gestão com informações consistentes pré-definidas que possam ser entendidas por sistemas computadorizados, ou seja, acionáveis por máquina (Datacite, 2023). Em um esforço colaborativo para melhorar e aperfeiçoar o trabalho de todos os envolvidos (Figura 5) discute-se o aprimoramento dos planos para maior riqueza temática e valor agregado para todas as partes implicadas. Nota-se o PGD integrando os vários atores do ecossistema científico. Conforme o *Data Documentation Initiative* (Datacite, 2023), planos de gestão de dados acionáveis por máquina trocam informações entre ferramentas e sistemas de pesquisa e incorporam PGDs em fluxos de trabalho existentes.

Figura 5 – Interações das partes interessadas



Fonte: traduzido de Miksa *et al.* (2019).

Na intenção de tornar todas as etapas que compõem os planos de gestão de dados automatizadas, Miksa *et al.* (2019) listaram em suas pesquisas dez princípios para que os planos possam ser acessados e legíveis por humanos e máquina. Os princípios prezam pela integração dos PGDs nos fluxos de trabalho dos participantes do ecossistema de dados; permitem a ação de sistemas automatizados em nome das “partes interessadas”; elaboram políticas para as máquinas e não apenas para as pessoas; descrevem para máquinas e humanos os componentes do sistema de gestão de dados; usam identificadores persistentes e vocabulários controlados;

adotam um modelo comum de dados para PGDs acionáveis por máquina; disponibilizam o plano para uso humano e por máquina; apoiam a avaliação e monitoramento da gestão de dados; tornam os planos atualizáveis, vivos e versionáveis; e os tornam acessíveis publicamente.

É consenso que os planos de gestão de dados reúnem informações de interesse para as partes envolvidas na pesquisa (que vão de financiadores a administradores institucionais). Uma futura geração de planos de gestão acionáveis por máquina favorecerá a interoperabilidade entre ferramentas e sistemas de informação.

Neste sentido, ferramentas da *web*, como DMPTool⁴, DMPonline⁵ e EasyDMP⁶, propõem-se a facilitar a tarefa de elaborar um plano de gestão de dados, orientando o pesquisador por meio de uma lista de perguntas pré-elaboradas ou de texto livre, normalmente criado no início da pesquisa e dificilmente retomado para atualização.

Diversas ferramentas de elaboração de planos de gestão de dados estão disponíveis em todo o mundo. As primeiras ferramentas surgiram por volta do ano de 2010, como o *DMPonline* no Reino Unido e o *DMPTool* nos Estados Unidos (Donnelly; Jones; Pattenden-Fail, 2010). Nos últimos anos surgiram outras ferramentas de elaboração de PGDs. Muitas delas oferecem funcionalidades para criar, compartilhar, exportar e revisar PGDs, embora cada uma delas enfatize diferentes aspectos (Jones *et al.*, 2020).

Jones e coautores (2020) salientam que algumas ferramentas se concentram em fornecer perguntas fechadas em vez de respostas em formato livre, e também disponibilizam bases de conhecimento para orientar os usuários. Outras ferramentas têm foco mais específico em projetos ou conjuntos de dados, e muitas delas estão integradas a outras ferramentas de documentação de dados e alocamento de armazenamento, o que apoia a efetiva implementação dos planos de gerenciamento de dados.

Príncipe e outros autores (2020) listam em suas pesquisas as principais ferramentas de elaboração de planos de gestão de dados. O relatório (Príncipe *et al.*,

⁴DMPTool é um aplicativo *online* gratuito oferecido pela Biblioteca Digital da Califórnia que ajuda os pesquisadores a criar planos de gestão de dados e pode ser acessado em <https://dmptool.org/>.

⁵DMPonline é uma ferramenta fornecida pelo Centro de Curadoria Digital que ajuda a criar, revisar e compartilhar planos de gestão de dados e pode ser acessado em <https://dmponline.dcc.ac.uk/>.

⁶EasyDMP é gratuito e foi desenvolvido pela Sigma2 em colaboração com o EUDAT2020 e pode ser acessado em <https://easydmp.no/>.

2020) avalia aspectos relevantes das ferramentas sobre a abertura ou não do *software*; o número de utilizadores; modelos disponibilizados que contemplem aspectos dos financiadores; casos de uso; formação, suporte e apoio técnico; e planos de custos. Assim como Jones e outros autores (2020) que também elencam ferramentas de elaboração de planos de gestão de dados em seu artigo.

Os planos de gestão de dados descrevem métodos de pesquisa que necessariamente evoluirão ao longo de um projeto, portanto, para ser uma ferramenta útil para pesquisadores, instituições financiadoras, gestores e outros, o conteúdo deve ser atualizado para registrar os métodos empregados e os dados produzidos. Nesse sentido, os planos de gestão de dados acionáveis por máquina podem integrar e aperfeiçoar a experiência de todos os envolvidos, trocando informações entre ferramentas e sistemas de informação e incorporando PGDs em fluxos de trabalho existentes, permitindo que partes do plano sejam geradas e compartilhadas automaticamente, por exemplo, com colaboradores e financiadores.

Os planos de gestão de dados acionáveis por máquina (maDMPs) levam o conceito dos planos de gestão adiante usando serviços de identificadores persistentes para conectar todos os recursos associados a um PGD. O objetivo de um PGD acionável por máquina é facilitar o sistema de gerenciamento de dados de pesquisa para permitir que dados e informações sobre pesquisas sejam comunicados e compartilhados entre as partes interessadas, vinculando metadados, repositórios e instituições (Datacite, 2023).

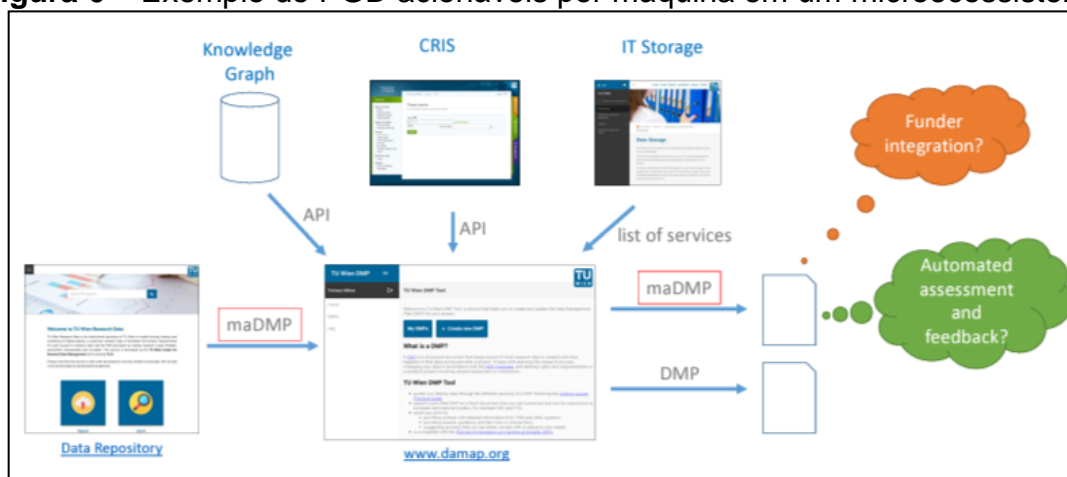
2.6.3 Dez princípios da *Research Data Alliance* para planos de gestão de dados acionáveis por máquina

Os planos de gestão de dados são instrumentos que ajudam os pesquisados a gerenciar seus dados. Por sua vez, os planos de gestão de dados acionáveis por máquina, conforme a Figura 6, são a ligação entre diferentes serviços de informação (Research Data Alliance, 2023).

Jones *et al.* (2020) narra que durante a 9ª reunião plenária da *Research Data Alliance*, em Barcelona, houve uma proposta para formar o grupo de trabalho de Padrões Comuns do PGD. Esse movimento gerou o *white paper* de Simms *et al.* (2017) sobre planos de gestão de dados acionáveis por máquina, construído a partir

de um *workshop* internacional realizado em Edimburgo no ano de 2017. O documento destaca a necessidade de um padrão comum para PGDs acionáveis por máquina e definindo “acionável por máquina” como informação estruturada para programação por computadores.

Figura 6 – Exemplo de PGD acionáveis por máquina em um microecossistema



Fonte: Research Data Alliance (2022).

O padrão comum da RDA para planos de gestão de dados acionáveis por máquina modela propriedades em campos principais da estrutura geral. Essas estruturas não se baseiam apenas em questionários ou *templates*, mas permitem que mais campos sejam adicionados à estrutura do PGD (Research Data Alliance, 2022).

Os planos de gestão de dados acionáveis por máquina facilitam a troca de informações entre sistemas de informações. Com o uso do padrão comum desenvolvido pela RDA, é possível a troca, integração e validação automática das informações fornecidas pelos PGDs (Miksa; Walk; Neish, 2020).

O padrão comum RDA para planos de gestão de dados acionáveis por máquina é um conjunto de metadados destinado a facilitar a interoperabilidade entre sistemas envolvidos na criação ou utilização de planos de gestão de dados que podem ser processados automaticamente (Research Data Alliance, 2023).

O acionamento por máquina, conforme a *Data Documentation Initiative* (2023, não paginado, tradução nossa), “[...] é definido como ‘informação que é estruturada de maneira consistente para que máquinas, ou computadores, possam ser programados de acordo com a estrutura’ [...]”. Esta é a principal característica dos planos de gestão de dados acionáveis por máquina e que os tornam documentos vivos. Wilkinson *et al.* (2016), afirmam que esse termo está associado aos princípios

Findable, Accessible, Interoperable, Reusable (FAIR) para expressar que máquinas devem ser capazes de agir de forma autônoma em objetos digitais.

De forma geral, o padrão comum da RDA aborda os aspectos sobre a autoria e a pesquisa à qual o PGD pertence, os dados em si, o armazenamento e a disponibilização dos dados. O padrão comum fornece um perfil de aplicação de metadados que contemplam a identificação de propriedade dos PGDs; a descrição da pesquisa associada ao PGD; o detalhamento sobre o financiamento da pesquisa; o contato do responsável pelos dados coletados; a relação de todas as partes envolvidas no processo de gestão de dados e criação do PGD; a informação sobre os custos relacionados à gestão dos dados; as informações de alto nível sobre o conjunto de dados; a forma como os dados serão disponibilizados; a indicação de licença sob a qual os dados serão disponibilizados (ou se ficarão embargos); as informações sobre o sistema onde os dados são armazenados; os requisitos específicos relacionados à segurança e privacidade de um conjunto de dados específico; a especificação do equipamento necessário/utilizado para criar ou processar os dados; e fornece um indicador padrão de metadados para descrever os dados.

Cada um desses aspectos descreve as propriedades a serem incorporadas na estrutura geral do PGD (Figura 7).

Figura 7 – Exemplo de metadados para a propriedade de “contato” em um PGD

Nome	Descrição	Tipo de dados	Cardinalidade	Valor de exemplo
ID_de_contato	Identificador de uma pessoa de contato	Estrutura de dados aninhada	1	
mbox	Endereço de email	Corda	1	cc@exemplo.com
nome	Nome da pessoa de contato	Corda	1	Charles Chaplin

Fonte: *Research Data Alliance* (2022, não paginado).

Na página *web* da *Research Data Alliance* existem recomendações de algumas ferramentas de elaboração de PGD, todas baseadas nas recomendações acionáveis por máquina (Miksa; Walk; Neish, 2020), a saber:

- a) DMPOnline, do Digital Curation Centre no Reino Unido;
- b) DMPTool, da California Digital Library nos Estados Unidos;
- c) DMP OPIDoR, do *Centre National de la Recherche Scientifique* na França;

- d) RDMO, do Leibniz-Institut für Astrophysik da Alemanha.
- e) Data Stewardship Wizzard, da *Elixir Research Infrastructure* na União Europeia;
- f) Argos-OpenDMP, da OpenAIRE e da EUDAT da União Europeia;
- g) F1000 Research, do Reino Unido;
- h) Norwegian Open Research Data Infrastructure, da Noruega;
- i) Haplo Repository, no Reino Unido;
- j) TUWien, TUGraz, UniWien via projeto FAIR Data Austria.

A RDA apresenta uma recomendação oficial de padrão comum para o acionamento por máquina de planos de gestão de dados. Este perfil de aplicação destina-se a cobrir uma ampla gama de uso e representa informações sobre todo o ciclo de vida do plano de gestão de dados. Com o mapeamento das limitações dos PGDs convencionais, o modelo padronizado de metadados acionáveis por máquina torna ativos os planos e possibilitam a troca de informações entre sistemas de pesquisa durante a gestão dos dados (Miksa; Oblasser; Rauber, 2021).

Para que os planos de gestão de dados acionáveis por máquina desempenhem todo o seu potencial, os diferentes atores envolvidos na gestão dos dados devem trocar informações de forma automatizada. Dessa forma, esses atores devem disponibilizar informações estruturadas que sejam passíveis de serem processadas por máquinas, permitindo que essas informações se integrem aos planos de gestão de dados.

O acionamento por máquina se dá na estruturação de informações modeladas semanticamente com o uso de vocabulários e padrões controlados, além do uso de identificadores persistentes para referenciar pessoas, instituições, financiadores, conjunto de dados, repositórios etc. (Miksa; Oblasser; Rauber, 2021).

Os planos de gestão de dados têm potencial para tornarem-se registros dinâmicos, altamente interativos com sistemas computacionais e mais flexíveis em termos de programação. Em estudos da *Research Data Alliance* foram definidos dez princípios para o acionamento por máquina dos planos de gestão de dados. Conforme Príncipe *et al.* (2020), essas regras buscam:

- 1) integrar os PGDs com os fluxos de trabalho de todos os intervenientes no ecossistema de dados de investigação;

- 2) permitir que os sistemas automatizados atuem em representação dos intervenientes no ecossistema de dados de investigação;
- 3) elaborar políticas que integrem os sistemas e não apenas as pessoas;
- 4) descrever, tanto para pessoas como para sistemas, os componentes do ecossistema de gestão de dados;
- 5) utilizar identificadores persistentes e vocabulários controlados;
- 6) seguir um modelo de dados comum para planos de gestão de dados acionáveis por máquina;
- 7) disponibilizar PGDs para utilização por parte de pessoas e sistemas;
- 8) apoiar a avaliação e o monitoramento da gestão de dados;
- 9) tornar os PGDs documentos atualizáveis, vivos e com possibilidade de versionamento;
- 10) tornar os PGDs publicamente disponíveis.

Os dez princípios da RDA para o acionamento por máquina dos planos de gestão de dados integram os PGDs aos fluxos de trabalho, facilitando a troca de informações entre diversas ferramentas e sistemas de pesquisa. Essa integração automatiza a geração e o compartilhamento de partes dos PGDs aos segmentos interessados, facilitando, por exemplo, o envio automático para colaboradores e financiadores. Assim, os planos demandam a implementação de modelos padronizados para facilitar a troca de informações.

Em um contexto de ecossistema de pesquisa compartilhado de serviços e atores que possam enviar e receber informações e atuar em questões que os pesquisadores não tenham condições de responder, os PGDs acionáveis por máquina têm a capacidade de automatizar a sugestão de licenças adequadas, estimar os custos de armazenamento e notificar o gestor de repositórios para o armazenamento dos dados, por exemplo, atuando em nome dos pesquisadores.

Os autores Miksa e colaboradores (2019) elencam alguns benefícios para o acionamento por máquina dos planos. A disponibilidade de informações estruturadas, como identificadores de autores, que reconhecem os produtores dos dados, e identificadores de repositórios que oferecem a possibilidade aos financiadores de monitorar a conformidade de maneira automatizada. Elementos pertinentes do PGD podem oferecer informações importantes sobre o consentimento e outros aspectos relevantes no início de um projeto, criando um registro rastreável da aprovação pelo

devido conselho de ética e assegurando a integridade da pesquisa. Os benefícios estendem-se aos aspectos jurídicos, pois elementos pertinentes do plano podem ser aplicados em pedidos de patente.

Ainda conforme Mika e outros autores (2019), os planos de gestão de dados acionáveis por máquina facilitam a interação de pesquisadores com especialistas ao longo do projeto de pesquisa, disponibilizando orientação e suporte no gerenciamento de dados. Assim, os procedimentos automatizados simplificam a elaboração e atualização do PGD. Na perspectiva do editor, viabiliza a criação automática de uma declaração referente à disponibilidade de dados, facilitando a associação e a citação apropriada de artigos, conjuntos de dados e outros resultados.

Em âmbito gerencial do operador de repositório, os benefícios do acionamento por máquina são o fornecimento de informações antecipadas sobre licenças, requisitos de metadados etc. Para o provedor de infraestrutura, a transferência de informações entre sistemas é otimizada, evitando inserções repetidas. As partes interessadas adequadas podem atualizá-lo em nome dos pesquisadores, aprimorando assim a qualidade das informações (Miksa *et al.*, 2019).

Outros benefícios apontados por Miksa *et al.* (2019) são para a equipe de apoio à pesquisa, que está habilitada a avaliar a qualidade das informações presentes no plano e fornecer um *feedback* relevante ao pesquisador; e para o administrador institucional, que podem obter uma visão abrangente dos dados utilizados, processados e criados na instituição.

Os dez princípios (Figura 8) relacionam-se a etapas distintas para o acionamento por máquina dos planos de gestão de dados e podem ser executadas de forma independentes, sem a necessidade de implantação de todos os princípios; entretanto, alguns princípios dependem da implementação de outros (Miksa *et al.*, 2019). Portanto, é importante observar as dependências e relações entre os princípios.

Figura 8 – Dez princípios para maDMPs

	1 Integrar os PGDs aos fluxos de trabalho no ecossistema dos dados de pesquisa
	2 Permitir que sistemas automatizados atuem em nome das partes interessadas
	3 Fazer políticas (também) para máquinas, não apenas para pessoas
	4 Descrever – tanto para máquinas quanto para humanos – os componentes do ecossistema de gerenciamento de dados
	5 Usar PIDs e vocabulários controlados
	6 Seguir um modelo de dados comum para PGD acionáveis por máquina
	7 Disponibilizar DMPs para consumo humano e de máquinas
	8 Apoiar a avaliação e monitoramento da gestão de dados
	9 Tornar os PGDs documentos atualizáveis, vivos e versionáveis
	10 Disponibilizar publicamente PGDs

Fonte: traduzido de Miksa *et al.* (2019).

Os princípios combinam questões técnicas, organizacionais e sociais, assim, abrangem metadados básicos que devem ser incorporados. É importante ressaltar que os princípios levam em conta três funções que englobam as partes interessadas, sendo elas, a formulação de políticas e a infraestrutura; a criação e a atualização de PGDs; e o uso e reuso (Miksa *et al.*, 2019). Assim, os princípios listados por Miksa e outros autores (2019) para tornar os planos de gestão de dados acionáveis por

máquina permitem que os planos sejam incorporados ao fluxo de trabalho dos atores envolvidos no ciclo de vida dos dados de pesquisa.

O **primeiro princípio** se aplica a todas as partes interessadas e objetiva integrar os planos de gestão de dados aos fluxos de trabalho de todas as partes interessadas no ecossistema de dados de pesquisa. Uma gestão eficiente de dados exige informações precisas abrangendo aspectos metodológicos, técnicos, legais e éticos da coleta e reutilização de dados. A criação de planos de gestão de dados deve ser colaborativa, envolvendo várias partes interessadas com conhecimento em áreas relevantes, garantindo a qualidade das informações, por exemplo, PGDs são usados por operadores de repositórios para estabelecer embargos e licenças, e pelos financiadores para verificar a conformidade com princípios FAIR.

O **segundo princípio** permite que sistemas automatizados atuem em nome das partes interessadas. Qualquer parte que gere informação em sistemas relacionados com planos de gestão de dados, como operador de repositório, fornecedor de infraestrutura, administrador institucional, revisão ética, perito jurídico e editor beneficiam-se desse princípio. Algumas das informações capturadas em um plano de gestão de dados já estão disponíveis eletronicamente, portanto, em vez de inseri-las novamente, seria útil se os bits relevantes pudessem ser obtidos de fontes apropriadas, talvez após verificações de consistência com outras fontes para garantia de qualidade: coleta de dados administrativos; estimativa de custos; seleção de licença; reserva de armazenamento; depósito de dados; validação e conformidade.

O **terceiro princípio** demanda políticas para máquinas e não apenas para pessoas. É aplicável a todas as partes interessadas que fornecem políticas relacionadas a dados (financiador, operador de repositório, fornecedor de infraestrutura, administrador institucional, revisão ética, especialista jurídico e editor). A característica comum desse princípio é que os principais requisitos da política devem ser expressos num formato em que as máquinas possam agir, ou seja, utilizando uma estrutura predefinida consistente e um vocabulário controlado.

O **quatro princípio** tem por objetivo descrever, tanto para máquinas quanto para humanos, os componentes do ecossistema de gerenciamento de dados e se aplica a todas as partes interessadas que fornecem sistemas relacionados ao PGD. Ao fornecer uma forma comum de descrever componentes específicos de um ecossistema de gestão de dados, como repositórios, esses componentes poderão ser facilmente descobertos por humanos e máquinas.

O **quinto princípio**, refere-se ao uso de identificadores persistentes e vocabulários controlados que podem ser empregados por todas as partes interessadas. Para tornar os planos de gestão de dados explícitos e compreensíveis para todas as partes interessadas, é preciso termos bem definidos e uma identificação precisa dos recursos: vocabulários controlados e identificadores persistentes sempre que possível.

O **sexto princípio** demanda um modelo de dados comum para planos de gestão de dados acionáveis por máquina. Financiador, operador de repositório, fornecedor de infraestrutura, administrador institucional, especialista jurídico e editor, todos que forneçam sistemas relacionados ao PGD. Um modelo de dados comum é um meio para troca de informações entre as partes interessadas. O modelo de dados comum não é um modelo prescritivo ou um questionário, mas fornece uma forma reutilizável de representar informações acionáveis por máquina de forma estruturada sobre temas cobertos pelos planos de gestão de dados. Ele modela informações, o que contrasta com as informações de texto livre coletadas pelos questionários conhecidos nas ferramentas tradicionais de PGD.

O **sétimo princípio** tem o propósito de disponibilizar planos de gestão de dados para consumo humano e de máquinas, sendo aplicável a todas as partes interessadas. Esse princípio enfatiza a importância de fornecer informações claras sobre a estrutura e a tecnologia subjacente que suporta os PGDs, facilitando assim a interpretação e a ação automatizada desses planos por sistemas e máquinas. Ele contribui para a interoperabilidade e a reutilização eficiente dos planos de gestão de dados em vários contextos e sistemas.

O **oitavo princípio** busca apoiar a avaliação e o monitoramento da gestão de dados e se destina a todas as partes interessadas do processo. Esse princípio enfatiza a importância de tornar os planos de gestão de dados acionáveis por máquina capazes de fornecer informações que permitam a avaliação e o acompanhamento das políticas de dados e dos requisitos específicos de uma instituição ou organização. Isso contribui para garantir a conformidade com as políticas e regulamentações relacionadas aos dados.

O **nono princípio** aplica-se a todas as partes interessadas e objetiva tornar os PGDs documentos atualizáveis, vivos e com controle de versão. Este princípio destaca a importância de tornar os planos de gestão de dados acionáveis por máquina capazes de fornecer informações que permitam a avaliação de custos e a estimativa

dos recursos necessários para a implementação das práticas de gestão de dados delineadas no plano. Isso ajuda as organizações e pesquisadores a planejar e alocar recursos de forma eficaz para gerenciar e preservar os dados de pesquisa adequadamente.

O **décimo princípio** tem o objetivo de disponibilizar publicamente os PGDs e é aplicável a todas as partes interessadas. Compartilhar e co-criar o plano com a equipe do projeto durante as fases de concepção e planejamento é fundamental, pois auxilia na definição da metodologia de pesquisa, na estimativa dos recursos necessários e na elaboração de um cronograma viável para a divulgação dos dados. Além disso, compartilhar o PGD com partes externas à equipe do projeto, como instituições, repositórios, financiadores e comitês de revisão ética, desde o início, simplifica as interações relacionadas aos dados entre as várias partes interessadas ao longo do desenvolvimento do projeto.

O Quadro 2 apresenta de forma resumida os princípios, seus objetivos e a quem se aplica. O padrão comum da *Research Data Alliance* para planos de gestão de dados acionáveis por máquina apresenta-se como um conjunto de elementos que define o escopo de informações cobertas pelos planos.

Quadro 2 – Dez princípios para planos de gestão de dados acionáveis por máquina

PRINCÍPIO	OBJETIVO	APLICAÇÃO	OBSERVAÇÃO
1	Integrar os PGDs aos fluxos de trabalho de todas as partes interessadas no ecossistema de dados de pesquisa	Todas as partes interessadas	A criação de planos de gestão de dados deve ser colaborativa, envolvendo várias partes interessadas com conhecimento em áreas relevantes, garantindo a qualidade das informações.
2	Permitir que sistemas automatizados atuem em nome das partes interessadas	Qualquer parte interessada que gere informação em sistemas relacionados com PGD	Algumas das informações capturadas num PGD já estão disponíveis eletronicamente.
3	Fazer políticas (também) para máquinas, não apenas para pessoas	Todas as partes interessadas que fornecem políticas relacionadas a dados	Os principais requisitos da política devem ser expressos num formato em que as máquinas possam agir.
4	Descrever – tanto para máquinas quanto para humanos – os componentes do ecossistema de gestão de dados	Todas as partes interessadas que fornecem sistemas relacionados ao PGD	Ao fornecer uma maneira comum de descrever componentes específicos de um ecossistema de gestão de dados, esses componentes poderão ser facilmente descobertos por humanos e máquinas.
5	Usar identificadores persistentes e vocabulários controlados	Todas as partes interessadas	Para tornar os PGDs explícitos e compreensíveis para todas as partes interessadas, precisa-se de termos bem definidos e de identificação precisa dos recursos.
6	Siga um modelo de dados comum para PGDs acionáveis por máquina	Todas as partes interessadas que fornecem sistemas relacionados ao PGD	Um modelo de dados comum é um meio para troca de informações entre as partes interessadas.
7	Disponibilizar PGDs para consumo humano e de máquinas	Todas as partes interessadas	Esse princípio enfatiza a importância de fornecer informações claras sobre a estrutura e a tecnologia subjacente que suporta os PGDs.
8	Apoiar a avaliação e monitoramento da gestão de dados	Todas as partes interessadas	Esse princípio enfatiza a importância de tornar os planos de gestão de dados acionáveis por máquina capazes de fornecer informações que permitam a avaliação e o acompanhamento das políticas de dados e dos requisitos específicos de uma instituição ou organização.
9	Tornar os PGDs documentos atualizáveis, vivos e com controle de versão	Todas as partes interessadas	Este princípio destaca a importância de tornar os planos de gestão de dados acionáveis por máquina capazes de fornecer informações que permitam a avaliação de custos e a estimativa dos recursos necessários para a implementação das práticas de gestão de dados delineadas no plano.
10	Disponibilizar publicamente os DMPs	Todas as partes interessadas	Compartilhar e co-criar o plano de gestão de dados com a equipe do projeto durante as fases de concepção e planejamento é fundamental, pois auxilia na definição da metodologia de pesquisa, na estimativa dos recursos necessários e na elaboração de um cronograma viável para a divulgação dos dados.

Fonte: adaptado de Miksa *et al.* (2019).

O acionamento por máquina pode transformar os planos de gestão de dados em um documento dinâmico, pois informações podem ser incorporadas de diferentes fontes e atualizadas de forma automática. O acionamento por máquina facilita a interoperabilidade de informações entre sistemas de pesquisa, porém, é necessário aplicar modelos e parâmetros internacionais para descrever como os dados são

gerenciados e é nesse sentido que o Padrão Comum da *Research Data Alliance* pode ser aplicado.

2.6.4 Componentes e propriedades do Padrão Comum *Research Data Alliance* para planos de gestão de dados acionáveis máquina

O perfil de aplicação de metadados do Padrão Comum RDA foi desenvolvido com a finalidade de tornar possível a expressão de informações contidas em planos de gestão de dados tradicionais de forma acionável por máquina. O propósito é estabelecer a interoperabilidade entre sistemas que produzem ou consomem planos de gestão de dados acionáveis por máquina. É importante ressaltar que este padrão comum foi concebido para abranger uma ampla variedade de casos de uso e não estabelece requisitos comerciais específicos, tais como aqueles impostos por financiadores ou outras partes interessadas. Seu escopo abrange informações relacionadas a todo o ciclo de vida dos dados de pesquisa, visando à flexibilidade e a adaptabilidade em diversos contextos de pesquisa e gestão de dados, possibilitando a troca, integração e validação automática das informações contidas nos planos e promovendo uma maior eficiência na gestão dos dados de pesquisa.

A implementação desse padrão comum facilita a troca de informações entre os diversos sistemas de informação que atuam em nome das partes envolvidas no ciclo de vida da pesquisa. Dentre essas partes envolvidas, destacam-se os pesquisadores, financiadores, gestores de repositórios, fornecedores de tecnologia da informação e comunicação (TIC) e bibliotecários.

O Padrão Comum RDA para Planos de Gestão de Dados Acionáveis por Máquina apresenta uma estrutura de metadados organizados para a expressão de informações em um PGD, abrangendo todas as fases da pesquisa. Cada metadado dessa estrutura subdivide-se em propriedades. O metadado **DMP** (PGD), tem como objetivo fornecer informações de alto nível sobre o plano de gestão de dados, incluindo elementos essenciais como título, data de modificação e outros detalhes relevantes. Essas informações garantem uma base sólida para a construção do perfil de aplicação.

No metadado DMP, há a descrição das propriedades relacionadas à pessoa de contato para um PGD, com a identificação dos indivíduos que desempenham funções

na gestão de dados associados ao plano. Isso inclui a lista de pessoas responsáveis pela execução das ações delineadas no plano, bem como a especificação dos custos envolvidos na gestão de dados. A possibilidade de fornecer múltiplas instâncias de custos permite uma divisão detalhada dos custos ou uma agregação para apresentar uma visão geral dos gastos.

Outro aspecto relevante é a inclusão da data e hora da primeira versão do PGD, um elemento que não deve ser modificado em PGDs subsequentes, garantindo uma linha temporal precisa. Além disso, o padrão define a descrição dos dados em um nível não técnico, possibilitando a inserção de informações em formato livre no PGD. Também são abordadas questões éticas, permitindo a descrição direta desses aspectos e indicando a presença de questões éticas relacionadas aos dados descritos.

A padronização é estendida para incluir a identificação de um protocolo de comitê de ética, bem como a especificação do idioma do plano usando a norma ISO 639-3, sendo necessário sempre que o PGD for modificado. Além disso, o metadado aborda elementos como a indicação da versão do plano, o projeto associado e identificadores fundamentais para um plano, completando uma estrutura abrangente e detalhada para expressar as informações essenciais em um plano de gestão de dados.

O metadado **Project** (Projeto) do Padrão Comum RDA representa uma faceta fundamental para contextualizar o plano de gestão de dados. Esse metadado oferece a possibilidade de descrever detalhes sobre o projeto associado ao PGD, se aplicável. O termo “project” é abordado em sua amplitude, abrangendo não apenas projetos financiados, mas também outros tipos de iniciativas, como projetos internos e teses de doutorado. Isso reflete a flexibilidade do padrão, permitindo a adaptação a uma variedade de contextos de pesquisa e desenvolvimento.

Uma das propriedades importantes dentro deste componente é a descrição do projeto, que possibilita uma narrativa detalhada sobre a natureza, escopo e objetivos do projeto relacionado ao plano. Essa descrição enriquece o entendimento do contexto do PGD, fornecendo informações para a sua interpretação e implementação.

Outro aspecto essencial é a data de início do projeto, que registra o momento em que o projeto teve início. Essa informação permite estabelecer uma linha temporal e compreender a evolução do projeto desde sua fase inicial. Além disso, a data de término do projeto é uma propriedade que oferece informações sobre quando o projeto

foi concluído, contribuindo para a compreensão da vigência e duração do projeto em questão.

A propriedade financiamento relacionada com o projeto permite a indicação de recursos financeiros vinculados ao projeto. Essa informação é vital para entender a sustentabilidade e o suporte financeiro associado à realização do projeto, o que pode influenciar a disponibilidade de recursos para a gestão de dados e a execução do plano delineado no PGD. O componente *Project* enriquece o Padrão Comum RDA, tornando-o mais abrangente e aplicável a uma variedade de contextos de pesquisa.

O metadado **Funding** (Financiamento) do Padrão Comum RDA descreve aspectos relacionados à fonte de financiamento associada ao projeto. Esse metadado visa fornecer detalhes específicos sobre os projetos que foram financiados.

A propriedade ID do financiador é um dado chave dentro deste contexto, recomendando o uso do *CrossRef Funder Registry* para especificar o financiador do projeto. Essa especificação estabelece uma referência clara e única ao financiador, permitindo a identificação precisa e padronizada que facilita a integração e interoperabilidade entre diferentes sistemas e plataformas.

Além disso, o padrão prevê a indicação do tipo de identificador associado ao financiamento, oferecendo três valores permitidos: referência de fundos, *URL*⁷ ou outro. Essa flexibilidade permite adaptar-se a diferentes práticas e requisitos de identificação de financiamento, garantindo uma abordagem mais abrangente e inclusiva.

Outra propriedade essencial é a ID da concessão, que possibilita a especificação do identificador associado à concessão financeira. Essa propriedade permite rastrear e documentar os recursos financeiros atribuídos ao projeto, oferecendo transparência e uma trilha de auditoria clara.

O metadado **Contact** (Contato) do Padrão Comum RDA define a parte responsável por fornecer informações sobre o PGD. A função principal é estabelecer a identificação da pessoa ou organização que pode oferecer detalhes relevantes relacionados ao PGD. É importante ressaltar que essa parte identificada como o contato não necessariamente corresponde ao criador original do PGD, ampliando a flexibilidade e abrangência do padrão.

⁷*Uniform Resource Locator* é um termo técnico que se refere ao endereço de rede no qual se encontra algum recurso informático.

A propriedade do identificador da pessoa de contato visa fornecer um identificador único associado à pessoa de contato. Pode ser um ORCID⁸, ISNI⁹ ou um OpenID, garantindo a identificação precisa e padronizada da parte responsável pelo plano. Além disso, o padrão prevê a inclusão do endereço de *e-mail* e o nome da pessoa de contato, permitindo uma forma adicional de contato e identificação dessa parte.

As propriedades do componente Contato possibilitam a comunicação eficaz e direta com a pessoa ou organização responsável por fornecer esclarecimentos ou informações adicionais sobre o PGD, facilitando a colaboração entre as partes interessadas envolvidas na gestão dos dados.

O metadado **Contributor** (Contribuinte) visa mapear e registrar todas as partes envolvidas tanto no processo de gestão de dados descrito no PGD quanto na criação e administração do próprio plano.

A propriedade endereço de e-mail é essencial, pois permite a identificação única e a forma de contato com cada contribuinte envolvido. Isso facilita a comunicação eficaz e direta, possibilitando esclarecimentos adicionais sobre a gestão de dados. Além disso, a propriedade nome apresenta o nome do contribuinte, proporcionando uma referência clara e identificação inequívoca.

Outra propriedade é o tipo de contribuição, que especifica a natureza da contribuição de cada indivíduo ou organização no contexto do PGD. Essa propriedade pode incluir informações sobre seu papel no processo de gestão de dados, como criador de dados, responsável pela preservação ou outro papel relevante. Incorporar identificadores como ORCID, ISNI ou códigos abertos para o nome reforça a padronização e a unicidade na identificação dos contribuintes.

O metadado **Cost** (Custo) é importante para a gestão financeira e a transparência associada à gestão de dados. Ao fornecer uma lista de custos relacionados a essa gestão, o padrão permite uma análise detalhada dos recursos financeiros necessários para a implementação efetiva do plano de gestão de dados.

A propriedade dos valores permitidos, seguindo a norma ISO 4217 com códigos de moeda, garante a padronização e uniformidade na representação dos custos. Essa

⁸Open Researcher and Contributor ID é um código alfanumérico não proprietário para identificar exclusivamente cientistas e outros autores acadêmicos e contribuidores.

⁹ISNI é um identificador que identifica de forma exclusiva as identidades públicas dos contribuidores de conteúdo de mídia.

propriedade possibilita a comunicação eficaz e clara dos recursos financeiros envolvidos, evitando ambiguidades e garantindo uma interpretação precisa dos custos em diferentes contextos e entre diferentes partes interessadas.

Essa padronização na representação dos custos em termos de moeda também facilita a comparação e o gerenciamento financeiro, permitindo uma avaliação eficaz dos recursos necessários para a implementação do PGD. Além disso, a disponibilidade de uma lista detalhada de custos oferece uma visão abrangente dos aspectos financeiros relacionados ao gerenciamento de dados, promovendo a transparência e facilitando a alocação adequada de recursos para garantir a eficácia do plano de gestão de dados proposto.

O metadado **Dataset** (Conjunto de Dados) do Padrão Comum RDA adota a definição de *dataset* conforme especificado na W3C¹⁰ DCAT¹¹, fornecendo uma representação lógica de dados. Essa entidade apresenta informações de alto nível sobre os dados, destacando a sua importância para o contexto do plano de gestão de dados. A granularidade do conjunto de dados pode variar de acordo com a configuração específica, podendo ser um arquivo isolado ou uma coleção de arquivos em diferentes formatos, demonstrando flexibilidade e adaptabilidade.

A propriedade garantia de qualidade de dados é fundamental para assegurar a integridade e confiabilidade dos dados apresentados. O ID do conjunto de dados proporciona uma identificação única para o conjunto de dados, facilitando sua referência e rastreabilidade. A descrição, também presente tanto no conjunto de dados quanto na distribuição, é alinhada com as diretrizes da W3C DCAT, enfatizando uma abordagem abstrata para o conjunto de dados e uma referência específica para a distribuição.

A inclusão de informações sobre a publicação, codificada conforme a ISO 8601, oferece uma marcação temporal relevante para o conjunto de dados, contribuindo para a compreensão do seu ciclo de vida e relevância. As palavras-chave e o idioma

¹⁰O *World Wide Web Consortium* é uma comunidade de especialistas que desenvolvem padrões para a *web*.

¹¹O DCAT é o vocabulário do catálogo de dados recomendado pela W3C e fornece às pessoas e às máquinas uma abordagem específica e independente de domínio para criar catálogos que expressam os principais elementos de uma descrição de conjunto de dados de maneira padronizada, adequada para publicação na Web, e permite a interoperabilidade entre domínios. Veja mais em https://www.w3c.br/Noticias/traducao_DCAT.html#:~:text=O%20DCAT%20fornece%20%C3%A0s%20pessoas,dom%C3%ADnios%2C%20sendo%20usada%20por%20conta.

do conjunto de dados expressos usando ISO 639-3 são elementos importantes para a categorização e entendimento dos dados apresentados.

Para abordar aspectos como a privacidade e segurança dos dados, o Padrão considera a declaração de preservação e a lista de problemas e requisitos relacionados à segurança e privacidade, destacando a preocupação com a proteção de dados sensíveis. Além disso, a lista de recursos técnicos necessários para implementar um PGD é uma propriedade essencial que contribui para uma implementação bem-sucedida do plano de gerenciamento de dados, garantindo a integridade e acessibilidade dos dados.

O título e o tipo de conjunto de dados são propriedades que reforçam a abstração do conceito de conjunto de dados e a especificidade da distribuição. E, o identificador para um conjunto de dados e o tipo de identificador para o conjunto de dados possibilitam uma identificação única e diversificada do conjunto de dados, adequando-se a diferentes sistemas de referência e identificação.

O metadado ***Distribution*** (Distribuição) baseia-se na definição proveniente do perfil amplamente utilizado de metadados W3C DCAT. Distribuição aqui se refere a uma instância específica de um conjunto de dados que foi disponibilizada ou pretende ser disponibilizada de alguma forma. É fundamental distinguir a noção lógica de conjunto de dados de suas distribuições, que podem variar, especialmente para anexar propriedades de metadados mais específicas, como tamanho e licença. A relação entre o ciclo de vida dos dados de pesquisa e a distribuição é destacada, indicando que uma distribuição pode ser definida independentemente da realização efetiva do PGD.

Uma das propriedades essenciais é a *URL* do recurso, que fornece o acesso à distribuição do conjunto de dados, geralmente representando uma página de destino. É importante indicar a disponibilidade da distribuição, especificando por quanto tempo essa distribuição estará ou deverá estar disponível. O tamanho em *bytes* é uma informação significativa, pois disponibiliza informações sobre o volume de dados da distribuição.

Além disso, o modo de acesso aos dados é uma propriedade que indica se os dados estão em acesso aberto, compartilhado ou fechado. A *URL* do arquivo para *download* em um formato facilita o acesso e a compreensão dos dados. A

padronização do formato, conforme os padrões da IANA¹², contribui para uma representação precisa.

A propriedade qualidade do serviço fornece informações sobre a qualidade do serviço prestado pela infraestrutura onde os dados estão armazenados. Isso é relevante para a avaliação e a garantia da confiabilidade dos dados. A lista de licenças aplicadas à distribuição é fundamental para a transparência sobre as permissões e restrições associadas à utilização dos dados.

A propriedade título está presente tanto no conjunto de dados quanto na distribuição. A diferenciação entre esses títulos ressalta a abstração do conceito de conjunto de dados e a especificidade da distribuição, proporcionando uma visão clara e organizada.

O metadado **License** (Licença) abordado no Padrão Comum RDA para PGDs acionáveis por máquina indica a licença sob a qual os dados de cada distribuição específica serão disponibilizados. Esse metadado esclarece os termos e condições que governam o uso e a distribuição dos dados associados. Além disso, a inclusão da possibilidade de modelar embargos fornece uma abordagem flexível, considerando situações em que a disponibilização dos dados pode estar sujeita a restrições temporais. A referência a um *link* para documento de licença é uma prática importante, visto que permite que os usuários acessem informações detalhadas sobre a licença e seus termos. E a consideração de uma data definida no futuro como indicativo de um período de embargo destaca a adaptabilidade do padrão para contemplar situações em que a disponibilização dos dados está programada para um momento posterior.

O metadado **Host** (Hospedar) apresenta propriedades que fornecem informações detalhadas sobre o sistema onde os dados estão armazenados. O metadado permite uma visão mais clara sobre o ambiente em que os dados estão inseridos, abrangendo desde repositórios específicos onde os dados são depositados até os sistemas em que são armazenados e processados durante a pesquisa. Isso é de grande relevância, pois oferece transparência sobre a infraestrutura envolvida na gestão e manipulação dos dados.

A propriedade disponibilidade indica o tempo em que o sistema de hospedagem está acessível para os usuários. Já a frequência de *backup* e o tipo de *backup* são

¹²Internet Assigned Numbers Authority, organização mundial, supervisiona a atribuição dos números das portas, os endereços IP, sistemas autônomos, servidores-raiz e outros recursos relativos aos protocolos de Internet.

propriedades que revelam práticas importantes de gestão de dados e proporciona informações sobre a segurança e a preservação dos dados.

A especificação do repositório certificado de acordo com um padrão reconhecido ressalta a conformidade do sistema de hospedagem com padrões reconhecidos, reforçando a credibilidade e a confiabilidade do armazenamento dos dados. A descrição e a localização física dos dados são propriedades que enriquecem a compreensão da infraestrutura, permitindo uma representação precisa do ambiente de armazenamento dos dados.

A propriedade tipo de armazenamento necessário é de grande importância para especificar os requisitos específicos de armazenamento. O suporte de versionamento é outra propriedade importante, visto que indica se o sistema de hospedagem oferece suporte a versões dos dados, permitindo o controle e a rastreabilidade das alterações nos dados.

A *URL* do sistema que hospeda uma distribuição de um conjunto de dados é uma propriedade que fornece o acesso direto ao sistema de hospedagem de uma distribuição de dados, facilitando a localização e a obtenção dos dados.

O metadado ***Security and Privacy*** (Segurança e Privacidade) é um elemento que garante a integridade e a proteção dos dados. Essa propriedade é utilizada para indicar requisitos específicos relacionados à segurança e privacidade de um conjunto de dados, permitindo uma abordagem mais granular sobre o tratamento dessas questões sensíveis. A inclusão da descrição e do título como propriedades são elementos que fornecem informações detalhadas e identificação das condições de segurança e privacidade associadas a um conjunto de dados específico.

A descrição é a propriedade que fornece elementos sobre os requisitos de segurança e privacidade relacionados ao conjunto de dados. Essa propriedade pode incluir aspectos como o nível de anonimização dos dados, a aplicação de técnicas de criptografia ou qualquer outra medida de segurança específica. A transparência nessa descrição garante que os envolvidos tenham uma compreensão completa das práticas adotadas para proteger os dados.

O título é uma propriedade que oferece a identificação clara e concisa dos requisitos de segurança e privacidade relacionados ao conjunto de dados. Esse título pode ser útil para categorizar ou rotular os diferentes conjuntos de dados de acordo com suas exigências de segurança e privacidade, facilitando a organização e a gestão eficaz desses dados sensíveis.

O metadado ***Technical Resource (Recurso Técnico)*** trabalhado no Padrão Comum RDA para planos de gestão de dados acionáveis por máquina tem por objetivo fornecer informações sobre o equipamento necessário ou utilizado durante a criação ou processamento dos dados. O metadado descreve os recursos físicos, como equipamentos e instrumentos, que desempenham um papel fundamental na geração e manipulação dos dados. A descrição do recurso técnico fornece informações detalhadas sobre a função e as características específicas desse recurso, permitindo uma compreensão completa de como ele é utilizado no contexto dos dados.

O nome do recurso técnico é outra propriedade relevante, que oferece uma identificação clara e concisa do recurso técnico em questão. O nome do recurso técnico proporciona uma forma de referência e categorização eficaz dos recursos técnicos, o que facilita a organização e a gestão do conjunto de equipamentos necessários para o tratamento e a produção dos dados.

Outro metadado explorado no Padrão Comum RDA são os ***Metadata*** (Metadados). Os metadados oferecem uma referência para o padrão de metadados utilizado para descrever os dados associados. É relevante destacar que este componente não contém metadados reais relacionados ao conjunto de dados, mas serve como um ponto de acesso ou indicação do padrão de metadados adotado. A propriedade descrição é uma parte integral desse componente, fornecendo informações descritivas sobre a função e a finalidade do padrão de metadados referenciado.

O idioma dos metadados expressos especifica o idioma utilizado para a expressão dos metadados no padrão referenciado, o que garante que os metadados estejam compreensíveis e acessíveis para os usuários que podem necessitar de diferentes idiomas para interpretação. O ID padrão de metadados e o identificador do padrão de metadados utilizado são propriedades essenciais para identificar o padrão de metadados referenciado e proporcionar uma forma de referência única e inequívoca.

O tipo de identificador é outra propriedade fundamental para o componente Metadados, pois especifica o tipo de identificador utilizado para o padrão de metadados referenciado. Os valores permitidos incluem *URL* e outros tipos de identificadores. Essa propriedade contribui para a precisão e a consistência na referência ao padrão de metadados.

O Padrão Comum RDA para planos de gestão de dados acionáveis por máquina apresenta um conjunto abrangente e estruturado de propriedades essenciais para descrever e gerenciar dados de forma acionável e interoperável. Cada metadado, desde informações sobre o conjunto de dados e sua distribuição até detalhes sobre metadados, segurança, privacidade e recursos técnicos, desempenha um papel fundamental na articulação de práticas eficazes de gestão de dados. A inclusão de propriedades como “Hospedar”, “Conjunto de Dados”, “Licença” e “Metadados” oferece uma visão abrangente do ecossistema de dados, promovendo transparência, padronização e eficiência na gestão e compartilhamento de informações. Ao considerar os requisitos específicos e nuances do ciclo de vida dos dados, o Padrão Comum RDA se mostra uma ferramenta valiosa para a comunidade científica e profissionais de dados na busca pela otimização da gestão e da utilidade dos dados ao longo do tempo.

3 METODOLOGIA E PROCEDIMENTOS

Esta pesquisa tem natureza básica, também conhecida como pura ou fundamental por alguns autores (Cervo; Bervian, 1983), pois gera conhecimento com o objetivo de contribuir com novos entendimentos que sejam relevantes para a comunidade científica. A pesquisa de natureza básica: “Objetiva gerar conhecimentos novos, úteis para o avanço da Ciência, sem aplicação prática prevista. Envolve verdades e interesses universais” (Gerhardt; Silveira, 2009, p. 34). Ou seja, pretende introduzir o tema de planos de gestão de dados acionáveis por máquina, no contexto do ecossistema da pesquisa brasileira.

A pesquisa busca se aprofundar na problemática da elaboração dos planos de gestão de dados estáticos e a necessidade de integração com outros dispositivos, mostrando como a falta de automatização afeta o acompanhamento dos processos que envolvem o ciclo de vida dos dados. Para Flick (2004, p. 28): “A pesquisa qualitativa é orientada para a análise de casos concretos em sua particularidade temporal e local [...]”. Desta forma, a abordagem qualitativa tem como princípio analisar e interpretar características mais detalhadamente, para assim, descrever e entender complexidades de uma investigação (Marconi; Lakatos, 2010).

Assim, para Leite (2008, p. 100): “Uma das características da aplicabilidade dos métodos qualitativos são as situações em que se necessita realizar classificações comparativas e que se pretende identificar proporção, grau, ou intensidade de um determinado fenômeno”. Neste tipo de abordagem, Flick (2009) salienta que a qualidade deve predominar acima de qualquer viés. Portanto, a qualidade deste tipo de abordagem precisa ser objeto constante de atenção dos pesquisadores.

Em razão de descrever e apresentar características do universo estudado e suas relações: “A metodologia qualitativa preocupa-se em analisar e interpretar aspectos mais profundos [...]. Fornece análise mais detalhada sobre as investigações [...]” (Marconi; Lakatos, 2008, p. 269).

Quanto aos seus objetivos, esta pesquisa é classificada como descritiva, uma vez que, conforme Triviños (1987, p. 110), exige “[...] do pesquisador uma série de informações sobre o que se deseja pesquisar. [...] [e] [...] pretende descrever ‘com exatidão’ os fatos e fenômenos de determinada realidade”. Esta pesquisa também é classificada em explicativa em razão de relacionar teoria e prática, pois de acordo com Vergara (2000, p. 47) este tipo de pesquisa “[...] tem como principal objetivo tornar

algo inteligível, justificar-lhe os motivos. Visa, portanto, esclarecer quais fatores contribuem, de alguma forma, para a ocorrência de determinado fenômeno”.

Ainda, de acordo com Vergara (2000), a pesquisa explicativa corrobora com a pesquisa descritiva, uma vez que esta serve de base para suas explicações. Assim, explora o tema da pesquisa na tentativa de relacionar conhecimentos para compreender determinado fenômeno, ajudando a aumentar a compreensão de um determinado tópico.

No que tange a coleta de dados, as técnicas utilizadas foram o levantamento bibliográfico e a análise de documentos. Os dados gerados, tanto do levantamento bibliográfico como da análise documental, foram registrados, analisados, classificados e interpretados. Lakatos e Marconi (2001) salientam que a coleta de dados é um conjunto de regras ou processos que a ciência se vale para coletar informações para a pesquisa.

Utilizou-se dos procedimentos de pesquisa bibliográfica por ser “[...] feita a partir do levantamento de referências teóricas já analisadas, e publicadas por meios escritos e eletrônicos, como livros, artigos científicos, páginas de web sites” (Fonseca, 2002, p. 32), e documental, uma vez que as informações foram coletadas a partir de diversos tipos de documentos existentes, com o propósito de selecionar, tratar e interpretar as informações, a fim de extrair valores para as mesmas. Para Severino (2007), a pesquisa documental é:

[...] fonte [de] documentos no sentido amplo, ou seja, não só de documentos impressos, mas, sobretudo de outros tipos de documentos, tais como jornais, fotos, filmes, gravações, documentos legais. Nestes casos, os conteúdos dos textos ainda não tiveram nenhum tratamento analítico, são ainda matéria-prima, a partir da qual o pesquisador vai desenvolver sua investigação e análise. (Severino, 2007, p. 122).

Dessa forma, descreve-se o método de caráter científico que busca dar estrutura consistente à pesquisa para obtenção do resultado.

O procedimento de coleta de dados deu-se com a pesquisa nas bases de dados bibliográficas nacionais e internacionais Base de Dados em Ciência da Informação (BRAPCI), *Scientific Electronic Library* (SciELO), Portal de Periódicos da Capes, Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), *Information Science, Technology Abstracts* (LISTA) e *Google Scholar*, das áreas do conhecimento que contribuíram para a pesquisa, tais como a Ciência da Informação, a Ciência da

Computação e a Administração, para recuperar publicações sobre o tema. Os passos para coleta incluíram o estabelecimento dos limites para o estudo (Creswell, 2010), ou seja, a recuperação de artigos, relatórios, anais de eventos e documentos foram contemplados pelas expressões de busca em português e inglês para “planos de gestão de dados” e “planos de gestão de dados acionáveis por máquina”, e foram admitidos ao longo da pesquisa, a inclusão das expressões “ecossistema de pesquisa” e “ferramentas de elaboração de planos de gestão de dados acionáveis por máquina”, que colaboraram para a elucidação do problema desta pesquisa. Por tratar-se de um tema ainda em discussão no Brasil e no exterior, a pesquisa não delimitou as bases de dados para consulta.

A principal vantagem da coleta de dados a partir dos documentos reunidos é que, conforme Creswell (2010, p. 212): “Podem ser acessados em um momento conveniente para o pesquisador – uma fonte de informações pertinente”. Os dados coletados contemplaram as ferramentas de elaboração de planos de gestão de dados disponíveis *online* e suas características e os atores do ecossistema da pesquisa brasileira.

A literatura científica recuperada apoiou a exploração do tema, buscando abordar a aplicação dos planos de gestão de dados acionáveis por máquina no ecossistema da pesquisa. Segundo os autores Quivy e Campenhoudt (1995), a literatura científica pertinente serve:

[...] primeiramente para nos informarmos das pesquisas já realizadas sobre o tema e obtermos contribuições para o projeto de pesquisa. Graças a essas leituras, o pesquisador poderá evidenciar a perspectiva que lhe parece mais pertinente para abordar seu objeto de estudo. A escolha das leituras requer ser feita em função de critérios precisos: ligações com a questão inicial, dimensão razoável de leituras [...] (Quivy; Campenhoudt, 1995, p. 44).

Os dados coletados a partir da literatura científica recuperada, ou seja, as características, possibilidades de integração e aplicações das ferramentas de elaboração de maDMPs foram registrados em planilhas no formato CSV e, posteriormente, comparadas, analisadas e interpretadas.

A fim de melhor extrair significado dos dados obtidos, eles foram manipulados de forma descritiva, pois de acordo com Dencker e Viá (2001, p. 135), esse tipo de análise identifica todas as características do que se quer estudar: “As diferentes características consideradas são transformadas em variáveis operacionalizadas e

medidas por meio de procedimentos de pesquisa para depois serem submetidas à análise [...]”.

O processo de análise dos dados compreendeu o que Creswell (2010, p. 217) considera a extração de sentido dos dados: “Envolve preparar os dados para a análise, conduzir diferentes análises, ir cada vez mais fundo no processo de compreensão dos dados [...], representar os dados e realizar uma interpretação do significado mais amplo dos dados”.

Portanto, de posse dos dados analisados e interpretados, foi possível correlacionar a cobertura das ferramentas com o fluxo do ciclo de vida dos dados e verificar a aplicabilidade de processos automáticos nos planos de gestão de dados acionáveis por máquina no ecossistema da pesquisa brasileira. Esta é a fase que Bardin (1997) chama de fase da codificação, ou seja, a transformação dos dados em grupos para o tratamento e posterior esclarecimento da informação.

A partir da coleta e tratamento dos dados, o estudo foi interpretado com o auxílio da análise documental, que revelou aspectos novos do tema (Ludke; André, 1986). Conforme Matos (2023), a análise documental é realizada a partir de documentos, sejam contemporâneos ou retrospectivos, que são considerados cientificamente autênticos. Os documentos podem ser de fontes primárias e secundárias; fontes escritas ou não”. Iglesias e Gómez (2004) vão além ao conceituar a análise documental como um conjunto de operações intelectuais, propondo-se a descrição e representação dos documentos de forma unificada e sistemática.

De acordo com alguns teóricos, a análise documental está centrada em duas concepções: em método (Pimentel, 2001) e em técnica (Bardin, 1997). Para Matos (2023) a prova concreta de dados, a independência do pesquisador e a maior separação entre o sujeito e o objeto da pesquisa são vantagens para esse tipo de análise.

Portanto, esta pesquisa utilizou a perspectiva da análise documental enquanto técnica para o tratamento dos dados a fim de torná-los precisos, pois, este método de análise fornece uma representação objetiva das informações contidas no documento, descrevendo e representando o conteúdo dos documentos de maneira distinta (Iglesias; Gómez, 2004).

O processo de análise documental seguiu as etapas do Quadro 3.

Quadro 3 – Etapas do processo de análise documental

Etapa analítica	Identificação da Manifestação e seu Plano de Conteúdo: abstração das diferentes manifestações (visuais, gestuais, verbais ou sincréticas) e identificação do seu plano de conteúdo.
	Leitura Técnica do documento: em que o documentalista adentra na estrutura do documento, buscando tomar contato com as partes que revelem maior conteúdo temático valendo-se, para tanto, de um conjunto de estratégias metacognitivas.
	Identificação de conceitos: uma vez identificadas as partes mais significativas tematicamente, aplica-se ao documento um conjunto de categorias conceituais, visando à construção de enunciados de assunto.
Etapa sintética	Seleção de conceitos: os enunciados de assunto são, então, categorizados em si, secundários e periféricos, e ordenados logicamente, tendo como parâmetros a estrutura, a função e os usos (tipo de busca informacional a que se presta o documento).
	Condensação documentária: redução do documento original a um micro documento (resumos).
	Representação documentária: tradução do conteúdo temático do documento em linguagem de indexação, representando-o por meio de índices.

Fonte: elaborado a partir de Sabbag (2013).

Na pesquisa qualitativa, os resultados são validados com a verificação da precisão dos resultados empregando a documentação dos procedimentos da investigação e assim como os passos dos procedimentos (Yin, 2003). Como estratégia de validade foi utilizada a descrição rica e densa para comunicar os resultados (Creswell, 2010).

Para atingir o objetivo de identificar na literatura os modelos *machine-actionable Data Management Plans* (objetivo “a”), foram feitas buscas nas bases de dados bibliográficas nacionais e internacionais com a aplicação de filtros por áreas da Ciência da Informação, Ciência da Computação e a Administração para recuperar artigos e relatórios que explorem os modelos de ferramentas de elaboração de planos de gestão acionáveis por máquina, assim como buscas em *sites* de instituições de pesquisa e fomento.

De posse da identificação dos maDMPs foram elencadas as possibilidades, benefícios e aplicações desses modelos com o ciclo de vida dos dados. A partir das fontes recuperadas nas buscas, foi possível qualificar os maDMPs e o ciclo de vida dos dados de pesquisa e relacioná-los com ecossistemas de pesquisa, respondendo, dessa forma, o objetivo “b”.

Em resposta ao objetivo “c”, que é o de identificar os atores do ecossistema da pesquisa brasileira, fez-se buscas em relatórios, artigos, diagramas e anais de eventos recuperados nas pesquisas às bases de dados bibliográficas, os atores do ecossistema de pesquisa, de forma a identificá-los e listá-los. A identificação partiu, mais especificamente, do mapeamento dos atores descritos nos trabalhos de Pinto *et al.* (2021) e Carvalho Segundo *et al.* (2022). Em conseqüente, foram apresentados os papéis que cada ator desempenha no ecossistema.

Após identificar os modelos *machine-actionable Data Management Plans* e exploradas as suas características, assim como verificada a cobertura dos *maDMPs* quanto ao ciclo de vida dos dados de pesquisa para a integração com ecossistema de pesquisa e, por fim, identificados os atores e os seus papéis no ecossistema, foi analisada teoricamente a aplicabilidade do *maDMP* no ecossistema da pesquisa brasileira, atendendo o objetivo “d”.

Dessa forma, objetivou-se analisar a associação do plano de gestão de dados acionáveis por máquina com o ciclo de vida dos dados no ecossistema da pesquisa brasileira, respondendo o problema desta pesquisa.

4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Esta seção apresenta os resultados dos dados coletados sobre as ferramentas de elaboração de planos de gestão de dados, correspondente ao objetivo específico “a”; a análise das características e integrações dos maDMP com os princípios para acionamento por máquina da *Research Data Alliance*, respondendo o objetivo específico “b”; a identificação dos atores, contemplando o objetivo específico “c” e a verificação da aplicabilidade dos planos de gestão de dados acionáveis por máquina no ecossistema da pesquisa brasileira, relativo ao objetivo “d”.

4.1 FERRAMENTAS DE ELABORAÇÃO DE PLANOS DE GESTÃO DE DADOS ACIONÁVEIS POR MÁQUINA

As ferramentas arroladas nesta pesquisa correspondem ao objetivo específico “a”, que busca identificar na literatura os modelos *machine-actionable Data Management Plans*. Abaixo estão relacionadas e detalhadas as 14 ferramentas identificadas na revisão de literatura.

A ferramenta **DMPonline**¹³ (Figura 9) é uma iniciativa desenvolvida seguindo os princípios para acionamento por máquina pela *Digital Curation Centre* (DCC), pela *University of California Curation Center* (UC3) e financiada pelo *Netherlands Organisation for Scientific Research* (NWO), com a colaboração da *California Digital Library* (CDL). Lançada em 2010, é baseada no código aberto da DMPRoadmap. A DMPonline oferece funcionalidades para a criação e administração de planos de gestão de dados. Dentre essas funcionalidades, destacam-se a capacidade de criar, editar, compartilhar, revisar e publicar PGDs, fornecendo uma abordagem personalizada para atender às necessidades específicas de pesquisas, financiadores e instituições. Permite ainda, a escolha de um padrão de metadados e disponibiliza glossários durante a criação do PGD. A ferramenta é multilíngue, garantindo maior acessibilidade e utilidade global.

¹³<https://dmponline.dcc.ac.uk/>.

Figura 9 – Ferramenta DMPonline

Info:
As part of our routine maintenance, we have upgraded our SSO login to enhance security.

- If your account was not linked to your institutional credentials, please log in as normal.
- If your account was linked to your institutional credentials, you will now need to re-link your account as part of this upgrade. To do this, please log in using your **DMPonline** email and password.
- Next, go to **Edit profile** > scroll down to the point **Institutional credentials**, and select the **Link your institutional credentials** option.
- After re-linking your account, you need to refresh your browser to complete the process. Remember to save your updated settings.

Plan to make data work for you
Data Management Plans that meet institutional funder requirements.

DMPonline helps you to create, review, and share data management plans that meet institutional and funder requirements. It is provided by the Digital Curation Centre (DCC).

126 096 Users 329 Organisations 141 727 Plans 89 Countries

Nota: ferramenta disponível em <https://dmponline.dcc.ac.uk/>.

Com o intuito de proporcionar um suporte eficaz aos usuários, a DMPonline disponibiliza uma variedade de recursos, incluindo uma central de atendimento, manual do usuário, treinamento e uma comunidade ativa de desenvolvedores. Sua ampla adoção é evidenciada por uma base significativa de usuários, totalizando 89 países, entre eles Reino Unido, África do Sul, Austrália e países da Europa, 338 organizações, 133.054 usuários e, aproximadamente, 149.604 PGDs.

A infraestrutura de entrega da ferramenta é robusta, proporcionando flexibilidade aos usuários. Oferece serviço principal multilocatário, instâncias privadas hospedadas no DCC e a possibilidade de autoimplementação. A viabilidade financeira é sustentada por diferentes fontes de financiamento, incluindo subsídios, taxas para instâncias hospedadas e assinaturas de financiadores e instituições. Esse modelo garante a continuidade e a evolução da ferramenta ao longo do tempo.

No que diz respeito à integração com outras plataformas e à exportação de dados, a ferramenta apresenta a capacidade de atribuir um Identificador de Objeto Digital ao plano de gestão de dados, promovendo a unicidade e a permanência da identificação dos dados. Ademais, a exportação em formatos populares como PDF,

DOC e XML, além da possibilidade de compartilhamento com colaboradores da pesquisa via e-mail, amplia a interoperabilidade e portabilidade dos dados gerenciados.

Cabe ressaltar que o DMPonline adota uma abordagem de código aberto, disponível no *GitHub*¹⁴, utiliza *Ruby on Rails*¹⁵ com uma licença MIT¹⁶. Essa escolha fomenta a colaboração, o aprimoramento constante e a personalização da ferramenta para atender às necessidades específicas de diferentes contextos de pesquisa.

A **DMPTool**¹⁷ (Figura 10) é um projeto iniciado pela *California Digital Library* em colaboração com as instituições *University of California Curation Center*, *DataONE*, *Digital Curation Centre*, *Smithsonian Institution* e diversas bibliotecas universitárias de acordo com os requisitos para acionamento por máquina. Personalizada para necessidades de pesquisa, financiadores e instituições, disponibiliza orientações de organizações. Baseada no código aberto da DMPRoadmap e lançada em 2011, oferece funcionalidades para a criação, edição, compartilhamento, revisão e publicação de planos de gestão de dados. A DMPTool possui assistente *click-through*¹⁸ para criar PGD que atenda aos requisitos do financiador e *links* para *sites* de financiadores, texto de ajuda para responder perguntas e recursos de práticas recomendadas de gestão de dados de pesquisa. Há ainda, a possibilidade de atribuição de DOI ao PGD.

A DMPTool é uma ferramenta multilíngue e oferece suporte que alavanca sua eficiência. Com amplos recursos, como central de atendimento, guia de início rápido, treinamento e uma comunidade ativa de desenvolvedores, a ferramenta garante um apoio abrangente e eficaz aos usuários. Além disso, a presença de uma base significativa de usuários nos Estados Unidos e na América do Sul, somando 396 organizações, aproximadamente 116.024 usuários e mais de 111.539 planos de gestão de dados, ilustra sua aceitação e relevância em comunidades acadêmicas e de pesquisa.

¹⁴*Github* é uma plataforma de desenvolvedor com tecnologia de inteligência artificial para criar, dimensionar e fornecer software seguro. Disponível em: <https://github.com/>.

¹⁵*Ruby on Rails* é um *framework* livre que visa facilitar o desenvolvimento de *sites* orientados a banco de dados. Disponível em: <https://rubyonrails.org/>.

¹⁶A licença MIT é uma licença para softwares, criada pelo Instituto de Tecnologia de Massachusetts (MIT).

¹⁷<https://dmptool.org/>.

¹⁸É a proporção de cliques em um *link* específico para o número de vezes que uma página, e-mail ou anúncio é exibido.

Figura 10 – Ferramenta DMPTool

Best practice made easier

With a free, community-supported service that makes it easier to create machine-actionable data management and sharing plans (DMSPs) that meet funder requirements and follow open science best practice.

[Get Started >>](#)

How does the DMP Tool work?

Simplifying data management and sharing plans for researchers and institutions with the DMP Tool's efficient, best practice-oriented, and compliance-designed approach

- Customization tools for researchers and institutions
- A mechanism for registering a DMP ID
- Best practice guidance to ensure plans are structured and optimized
- Participating organizations can provide feedback on plans
- A click-through wizard for creating a DMP that complies with funder requirements
- Direct links to funder websites, help text for answering questions, and Research Data Management (RDM) best practice resources

Getting started with the DMP Tool

Create data management and sharing plans that meet requirements and promote research.

[Get Started >>](#)

Nota: ferramenta disponível em <https://dmptool.org/>.

O mecanismo de entrega da ferramenta é composto por um serviço hospedado pela *California Digital Library* e a possibilidade de autoimplementação. Essa flexibilidade na entrega permite que os usuários escolham a melhor abordagem de acordo com suas necessidades e infraestrutura, garantindo uma maior adoção e adaptabilidade. A DMPTool disponibiliza também sua API no *GitHub*¹⁹.

No âmbito do financiamento, a DMPTool é respaldada por subsídios e pelo suporte da *California Digital Library*, assegurando sua continuidade e evolução. Essa diversificação nas fontes de financiamento fortalece a estabilidade e a sustentabilidade da ferramenta no longo prazo. A escolha de utilizar *Ruby on Rails* e

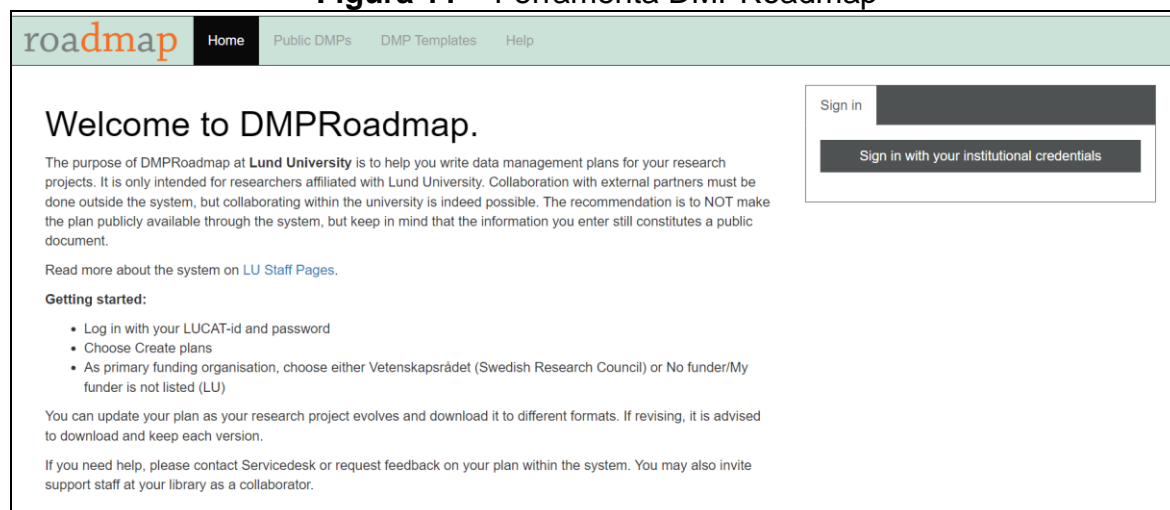
¹⁹Disponibilizado em: <https://github.com/CDLUC3/dmptool?tab=security-ov-file#readme>.

uma licença MIT para o desenvolvimento da DMPTool contribui para a melhoria contínua da ferramenta, atendendo a diferentes necessidades e cenários de pesquisa.

A DMPTool é uma ferramenta baseada no código aberto do DMPRoadmap e se destaca pela parceria entre instituições, suporte eficaz e flexibilidade de entrega, consolidando-se como uma escolha sólida para a gestão eficaz de dados de pesquisa e a criação de PGDs. A ferramenta permite a exportação nos formatos CSV, HTML, PDF, TXT, DOCX, JSON e compartilhamento via *e-mail* com colaboradores da pesquisa.

A ferramenta de elaboração de planos de gestão de dados **DMPRoadmap**²⁰ (Figura 11) é uma iniciativa desenvolvida pelo *Digital Curation Centre* e *University of California Curation Center*, além de provedores dos países de Edimburgo, Escócia e Reino Unido, lançada em 2015 e em constante evolução desde então. Atualmente, a DMPRoadmap, da Universidade de Lund, está sendo encerrada e migrada para a DMPonline. Essa colaboração entre diferentes provedores indica um esforço conjunto para criar uma solução de gestão de dados abrangente e multifacetada, capaz de atender a uma variedade de necessidades e contextos globais. A ferramenta acionável por máquina está alinhada com as demandas atuais de gestão de dados e reflete a contínua evolução e adaptação às exigências mutáveis da comunidade científica.

Figura 11 – Ferramenta DMPRoadmap



Nota: ferramenta disponível em <https://dmp.research.lu.se/>.

²⁰<https://dmp.research.lu.se/>.

Quanto às suas funcionalidades, Jones *et al.* (2020) indicam que suas funções, às quais incluem recursos essenciais para a criação, edição, compartilhamento, revisão e publicação de planos de gestão de dados, estão alinhadas com as necessidades de pesquisa, financiadores e instituições, disponibilizando orientações dessas instituições. A DMPRoadmap oferece também, a possibilidade de atribuição de DOI e ORCID e permite a exportação nos formatos CSV, HTML, PDF, TXT, DOCX, JSON e o compartilhamento com colaboradores por *e-mail*.

O suporte oferecido para essa ferramenta é abrangente, com *helpdesks*, treinamento, guia do usuário e multilíngue, demonstrando compromisso com a assistência e a orientação aos usuários, facilitando a adoção e a utilização efetiva da ferramenta. A presença de usuários em 89 países, incluindo Canadá, Finlândia, França, Bélgica, Espanha, Dinamarca e Japão, entre outros, indica uma ampla aceitação e adoção internacional da ferramenta. Com cerca de 17.600 usuários e 203 organizações, juntamente com o total de 23.000 PGDs, essa ferramenta possui uma base de usuários significativa.

No aspecto do financiamento, a DMPRoadmap é sustentada pelo apoio de ministérios dos países de desenvolvimento, projetos e financiamento de consórcios. Essa diversificação nas fontes de financiamento confere estabilidade e assegura a continuidade do desenvolvimento e aprimoramento da ferramenta.

A DMPRoadmap é disponibilizada por meio de instâncias hospedadas, possibilitando que organizações levem em consideração suas infraestruturas e requisitos específicos. A abordagem de código aberto e a ênfase na colaboração e transparência refletem na escolha das tecnologias de desenvolvimento (*GitHub*²¹), *Ruby on Rails* e licença MIT. Essas escolhas possibilitam a participação da comunidade na evolução e customização da ferramenta para atender às necessidades da gestão de dados em diferentes contextos e regiões.

A **EasyDMP**²² (Figura 12) apresenta-se como uma opção de plano de gestão de dados gratuita com enfoque na simplicidade e personalização lançada no ano de 2015. Desenvolvida pela Sigma2 em colaboração com EUDAT2020, permite o acionamento por máquina. A proposta da ferramenta é permitir a criação de planos de gestão de dados com base em modelos personalizáveis, incluindo modelos específicos de financiadores como H2020 e *Science Europe*. As diretrizes de gestão

²¹<https://github.com/DMPRoadmap/roadmap>.

²²<https://easydmp.eudat.eu>.

de dados da agência financiadora são transformadas em uma série de perguntas fáceis de responder, muitas delas contendo respostas prontas. Esse aspecto ajuda a adequar os PGDs às exigências e padrões de diferentes financiadores, facilitando o processo de conformidade.

Figura 12 – Ferramenta easyDMP



Nota: ferramenta disponível em <https://easydmp.eudat.eu/>.

O suporte oferecido pela ferramenta tem foco especial na facilidade de uso para atender tanto a usuários iniciantes quanto experientes. A presença de uma central de atendimento Sigma2 e documentação do usuário revela o esforço em proporcionar auxílio e orientação aos usuários, promovendo a eficácia na utilização da ferramenta.

A EasyDMP, embora mais concentrada na Noruega e Europa, conta com uma base de 1.558 usuários, 1.637 planos de gestão de dados e 258 organizações. Esses números demonstram uma adoção considerável da ferramenta nesses locais. Os planos de gestão de dados gerados podem ser exportados nos formatos HTML, PDF e JSON, além de permitir enviar convites por *e-mail* para colaboradores para a visualização e edição dos PGDs. Por outro lado, a easyDMP não permite a atribuição de DOI.

O mecanismo de entrega da EasyDMP, por meio de um serviço hospedado pela UNINETT Sigma2 AS & EUDAT, contribui para a facilidade de acesso e utilização. A opção de hospedagem do serviço permite aos usuários focarem no uso

da ferramenta sem se preocuparem com questões de infraestrutura, o que é um ponto positivo.

A EasyDMP disponibiliza modelos de PGDs de acordo com as diretrizes do *Horizon 2020 Expert*, *Science Europe*, *Sigma2 Data Management Plan*, *Horizon 2020 e NINA data management plan*. Em relação ao financiamento, a informação não especifica fontes de financiamento, o que pode ser um aspecto a ser melhor detalhado. A clareza sobre as fontes de financiamento é importante para compreender a sustentabilidade e continuidade da ferramenta.

A escolha de tecnologias de desenvolvimento de código aberto, como *Python*²³, e a intenção de adotar a licença do MIT reforçam o compromisso com a colaboração na evolução da ferramenta. A presença de uma API²⁴ é uma vantagem adicional, pois permite a integração com outros sistemas e personalização de funcionalidades.

Outra ferramenta, a **Data Stewardship Wizard**²⁵ (DSW) (Figura 13), representa uma abordagem inovadora para a gestão de dados, sendo desenvolvida pelo *Dutch Techcentre for Life Sciences* e pela *Czech Technical University*, em Praga, com lançamento no ano de 2015. Criada com recursos para o acionamento por máquina²⁶, seu principal diferencial é a utilização de um questionário dinâmico hierárquico, possibilitando a elaboração colaborativa de planos de gestão de dados por meio de questionários inteligentes e personalizáveis. Esses questionários permitem a avaliação da *FAIRness*²⁷ dos dados, contribuindo para o cumprimento de padrões de qualidade e acessibilidade dos dados. A ferramenta permite a elaboração de PGDs de acordo com modelos específicos de financiadores, como o *Horizon Europe*, *Horizon 2020* e *Science Europe*.

O suporte oferecido pela DSW é abrangente, inclui central de atendimento, documentação detalhada, canal flexível (comentar perguntas) e tutoriais sobre oficinas. A variedade de recursos de suporte demonstra o compromisso em auxiliar os usuários em diferentes aspectos da utilização da ferramenta, desde a compreensão até a aplicação prática.

²³Python é uma linguagem de programação de alto nível. Disponível em: <https://www.python.org/>.

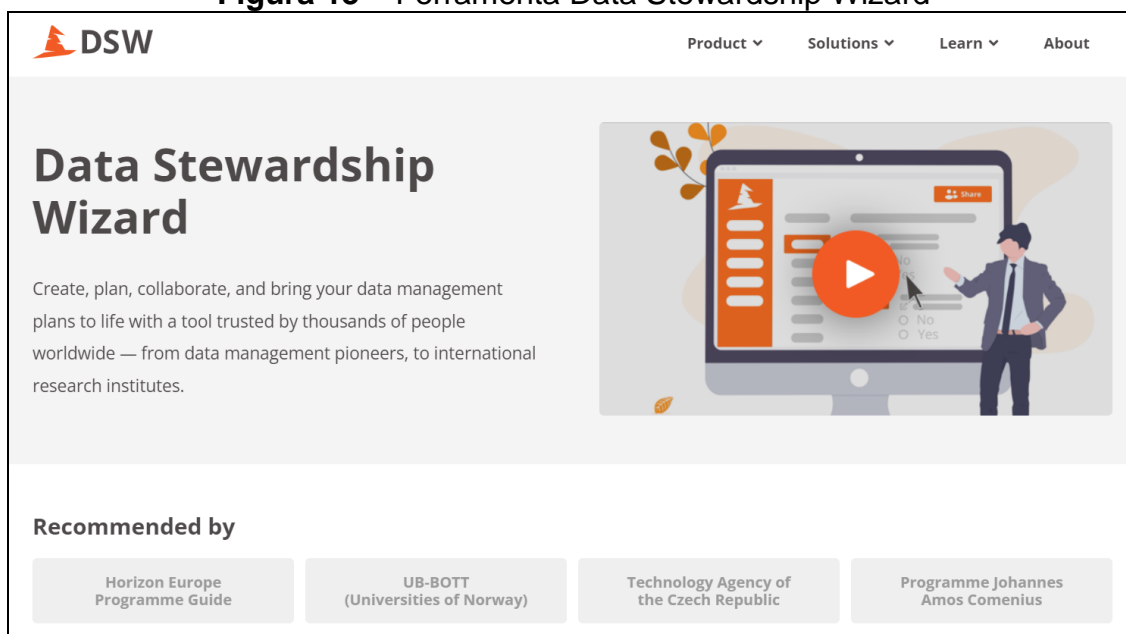
²⁴<https://easydmp.readthedocs.io/en/latest/api.html>.

²⁵<https://ds-wizard.org/>.

²⁶Existe o modelo *Machine-Actionable DMP* (maDMP) de acordo com o *RDA DMP Common Standard*.

²⁷Tratamento dos dados de acordo com os princípios FAIR.

Figura 13 – Ferramenta Data Stewardship Wizard



Nota: ferramenta disponível em <https://ds-wizard.org/>.

A presença de usuários na Europa, Brasil e parcerias com a indústria e instituições relacionadas ao domínio das Ciências da Vida destaca a relevância da ferramenta em um amplo contexto. A Data Stewardship Wizard oferece uma versão de demonstração, instâncias privadas e autoimplementação. A ferramenta é mantida por subsídios, financiamento da ELIXIR e taxas de personalização e hospedagem.

A Data Stewardship Wizard permite a atribuição de DOI ao PGD, o que é uma característica importante, pois fornece uma identificação única e permanente para os planos, contribuindo para a rastreabilidade e referenciamento adequado dos dados. A ferramenta permite ainda, o controle de versões. Além disso, a exportação em formatos como PDF, DOC, LaTeX, JSON, HTML, Markdown, OpenDocument, RDF/XML, Turtle, TRiG ou JSON-LD amplia a portabilidade dos dados e a integração com outros sistemas e ferramentas.

Com código aberto e licença Apache 2²⁸, reflete a adesão a princípios de colaboração, transparência e liberdade de customização. A aplicação integrada num servidor da CESNET²⁹ ou via Docker³⁰ demonstra versatilidade e adaptabilidade tecnológica para diferentes ambientes de implantação. A DSW permite também, que

²⁸A Licença Apache é uma licença de *software* livre permissiva de autoria da *Apache Software Foundation*.

²⁹CESNET é desenvolvedora e operadora de infraestrutura eletrônica nacional para ciência, pesquisa, desenvolvimento e educação na República Tcheca.

³⁰Docker é uma plataforma projetada para ajudar os desenvolvedores a construir, compartilhar e executar aplicativos contêineres.

qualquer pessoa possa desenvolver um modelo customizado que transforme as respostas em um questionário usando a linguagem de marcação de modelos Jinja2. Outro recurso é o editor de modelo de documento que consiste em uma ferramenta para criar novos modelos de documentos ou ajustar os existentes às necessidades do usuário. Já o *Template Development Kit* (TDK) é uma ferramenta de linha de comando conveniente para gerenciar projetos-modelo.

Assim, a Data Stewardship Wizard apresenta-se como uma ferramenta flexível para a elaboração colaborativa de PGDs, com enfoque especial na avaliação da *FAIRness* dos dados. A ferramenta propõe-se a orientar a elaboração de um PGD e, principalmente, instruir pesquisadores como lidar corretamente com os dados, torná-los FAIR, mantê-los confiáveis durante o projeto e curá-los a longo prazo. Seu suporte abrangente, presença internacional e mecanismos de financiamento diversificados reforçam sua relevância e potencial para contribuir para a eficácia na gestão de dados em diversas áreas científicas.

A ferramenta **Research Data Management Organiser**³¹ (RDMO) (Figura 14), lançada em 2017, se destaca por oferecer uma abordagem flexível, permitindo a autoimplementação por várias instituições e o acionamento por máquina. Financiada pela *German Research Foundation* em duas etapas do projeto, antes de ser repassada para um consórcio comunitário. Durante a etapa inicial, o *Leibniz-Institut für Astrophysik Potsdam*, *Potsdam University of Applied Sciences* e *Karlsruhe Institute of Technology* estiveram envolvidos; na segunda fase, a biblioteca do Instituto de Tecnologia de Karlsruhe (KIT) também participou.

³¹<https://rdmorganiser.github.io/en/>.

Figura 14 – Ferramenta Research Data Management Organiser



Nota: ferramenta disponível em <https://rdmorganiser.github.io/en/>.

Existem três funções diferentes na RDMO: usuário, gestores e administradores. As funcionalidades são voltadas para o planejamento, implementação e organização de dados de pesquisa, proporcionando uma abordagem estruturada por meio de questionários multilíngues e diversos formatos de exportação.

O suporte oferecido para a RDMO inclui tutoriais, canais de troca para desenvolvedores e uma lista de discussão. Esses recursos demonstram um esforço em fornecer auxílio e orientação aos usuários, tanto na implementação quanto na utilização efetiva da ferramenta. A ferramenta disponibiliza visualizações das especificações de planos de algumas agências de financiamento, por exemplo, *Federal Ministry of Education and Reserach, German Research Association, Swiss*

National Fonds, FWF Science Fund, Horizon Europe, European Research Council e Horizon 2020.

A presença da DSW predominante na Alemanha indica uma forte base de usuários no país. O mecanismo de entrega da ferramenta oferece opção de versão de demonstração e autoimplementação em instituições.

A atribuição de DOI aos dados de pesquisa é uma funcionalidade recomendada, pois confere identificação única e permanente aos planos, o que é essencial para a rastreabilidade e citabilidade adequadas dos dados de pesquisa. A RDMO permite a exportação em formatos como XML, CSV e JSON. Essa variedade de formatos amplia a utilidade da ferramenta, atendendo a diferentes requisitos e preferências dos usuários.

Em relação aos aspectos técnicos, a escolha de utilizar tecnologias de código aberto, disponibilizado no *GitHub*³², com documentação disponível em inglês, indica uma abordagem transparente e colaborativa. Além disso, a disponibilidade de uma API³³ abre possibilidades para integrações personalizadas e aprimoramento da interoperabilidade. A ferramenta disponibiliza ainda uma instância de demonstração³⁴. Destaca-se o aspecto de custo associado à ferramenta, a RDMO estima 5.000 euros para instalação inicial e 3.000 euros por ano.

A **UQ Research Data Manager**³⁵ (UQRDM) (Figura 15) é uma solução de gestão de dados desenvolvida e mantida pela *University of Queensland*, tendo sido lançada no ano de 2017, suporta o acionamento por máquina. Seu foco está na produção de registros de gestão de dados, com o objetivo de capturar metadados e alocar armazenamento durante a execução de projetos de pesquisa. Essa abordagem destaca a importância de documentar e gerenciar informações cruciais sobre os dados, permitindo uma melhor organização e utilização eficiente dos recursos de armazenamento.

³²<https://github.com/rdmorganiser>.

³³<https://rdmo.readthedocs.io/en/latest/administration/api.html>.

³⁴<https://rdmo.aip.de/>.

³⁵https://cloud.rdm.uq.edu.au/simplesaml/module.php/uqmultiauth/selectsource.php?AuthState=_c85652e7fb80ffd088fec2e54766cd95a2c7aef2bb%3Ahttps%3A%2F%2Fcloud.rdm.uq.edu.au%2Fsimplesaml%2Fsaml2%2Fidp%2FSSOService.php%3Fspentityid%3Dhttps%253A%252F%252Ftoken.rdm.uq.edu.au%252Fonelogin%252Fendpoints%252Fmetadata.php%26RelayState%3Dhttps%253A%252F%252Frdm.uq.edu.au%252Flogin-success%253FredirectTo%253D%25252F%26cookieTime%3D1715100515.

Figura 15 – Ferramenta UQ Research Data Manager

The screenshot shows the UQ Research Data Manager website. The main content area is titled "UQ Research Data Manager" and includes a search bar. A sidebar on the left contains a "UQRDM" dropdown menu with options like "UQRDM 2.0 - What's New", "Access UQRDM", "Related guides", and "Why use UQ Research Data Manager?". The main content area features a section titled "UQRDM 2.0 - What's New" with a "Did you know?" callout box. The callout box contains a table of services and permissions:

Services	Permissions
Storage 1	Read/Write, Read/Write, Read only
Storage 2	Read/Write, No access, No access
Notebook	Admin, User, No access
Impact Tracker	Owner, No access, Co-investigator

Nota: ferramenta disponível em <https://guides.library.uq.edu.au/for-researchers/uq-research-data-manager/uqrdm-access>.

A ferramenta auxilia no cumprimento dos requisitos do órgão financiador *Australian Research Council* e atende aos requisitos de financiadores, periódicos e órgãos reguladores da Austrália. Há também a viabilidade de atribuir um DOI ao plano de gestão de dados. Por outro lado, a literatura não informa os formatos de exportação possíveis³⁶, bem como a existência de custos de utilização da ferramenta.

Em termos de suporte, a UQRDM oferece uma central de ajuda, guia do usuário e treinamento para auxiliar os pesquisadores na sua utilização. Esse suporte permite aos usuários tirar o máximo proveito da ferramenta, especialmente considerando a complexidade e a importância da gestão de dados de pesquisa.

³⁶A ferramenta não permite acesso externo à comunidade da *University of Queensland*, por isso não foi possível verificar em testes na ferramenta os formatos de exportação

A base de usuários da UQRDM inclui a comunidade da *University of Queensland*, na Austrália, bem como colaboradores de pesquisa associados à instituição. O mecanismo de entrega é um serviço hospedado, o que indica que a ferramenta é disponibilizada remotamente e gerenciada pela equipe da *University of Queensland*.

A UQRDM é financiada institucionalmente pela *University of Queensland*. É relevante notar que o código atual é fechado, embora esteja sendo considerada a possibilidade de torná-lo aberto. Nesse sentido, não há uma API disponível. Essa mudança para um modelo de código aberto pode expandir a colaboração e a adesão à ferramenta, permitindo contribuições da comunidade e promovendo maior transparência.

A **DataWiz**³⁷ (Figura 16) apresenta-se como uma ferramenta gratuita de elaboração de planos de gestão de dados lançada em 2017 e é desenvolvida pelo *Leibniz Institute for Psychology Information (ZPID)*. Seu principal objetivo é auxiliar na elaboração de PGDs que estejam em conformidade com os requisitos dos financiadores, oferecendo suporte para o processamento e a documentação de dados psicológicos conforme os padrões estabelecidos, como o *Data Documentation Initiative (DDI)*. Essa orientação específica para dados psicológicos destaca uma abordagem especializada, focada em atender às necessidades desse domínio científico.

A ferramenta oferece funções para preparação, documentação e arquivamento de dados, bem como um ambiente de trabalho digital colaborativo para pesquisadores. Propõe-se ainda a reduzir o tempo gasto na gestão de dados de pesquisa pelos pesquisadores, aliviar a carga nos repositórios de dados de pesquisa, aumentando a qualidade dos dados depositados e estabelecer gestão de dados de pesquisa adequado ao domínio e práticas de Ciência Aberta em Psicologia.

A DataWiz oferece suporte multilíngue, controle de versão e uma comunidade diversificada de usuários. O guia do usuário é um recurso que auxilia na compreensão e na utilização efetiva da ferramenta, proporcionando orientações detalhadas sobre suas funcionalidades e processos. A DataWiz preocupa-se com formatos e interfaces padronizados para permitir e incentivar a disponibilização de dados de pesquisa.

³⁷<https://datawizkb.leibniz-psychology.org/index.php/project-start/dmp/>.

Nesse sentido, vale-se de metadados e vocabulários controlados DDI ou *Dublin Core* e permite a exportação nos formatos XML, PDF e CSV.

Figura 16 – Ferramenta DataWiz



Nota: ferramenta disponível em <https://datawizkb.leibniz-psychology.org/index.php/project-start/dmp/>.

O DataWiz ajuda a gerar planos de gestão de dados alinhados às informações exigidas pelas diretrizes das agências de financiamento *Horizon 2020* da União Europeia, *Bundesministerium für Bildung und Forschung* e *German Research Foundation*. Outra possibilidade da ferramenta é a realização do *upload* do arquivo dos dados de pesquisa para o PGD. Embora permita registrar um número interno de identificação do projeto, bem como o código de concessão, não há a possibilidade de atribuir um DOI.

Quanto ao mecanismo de entrega, a DataWiz é hospedada pelo ZPID, mas também oferece a opção de autoimplementação. Essa flexibilidade no modo de entrega permite que os usuários escolham a melhor forma de acesso e utilização da ferramenta, considerando suas necessidades e infraestrutura.

A menção de uma API planejada para a versão 2, disponibilizada no *GitHub*³⁸, sugere uma visão de evolução e expansão da ferramenta de código aberto, considerando a integração com outros sistemas e aprimoramento de funcionalidades por meio de interfaces programáticas.

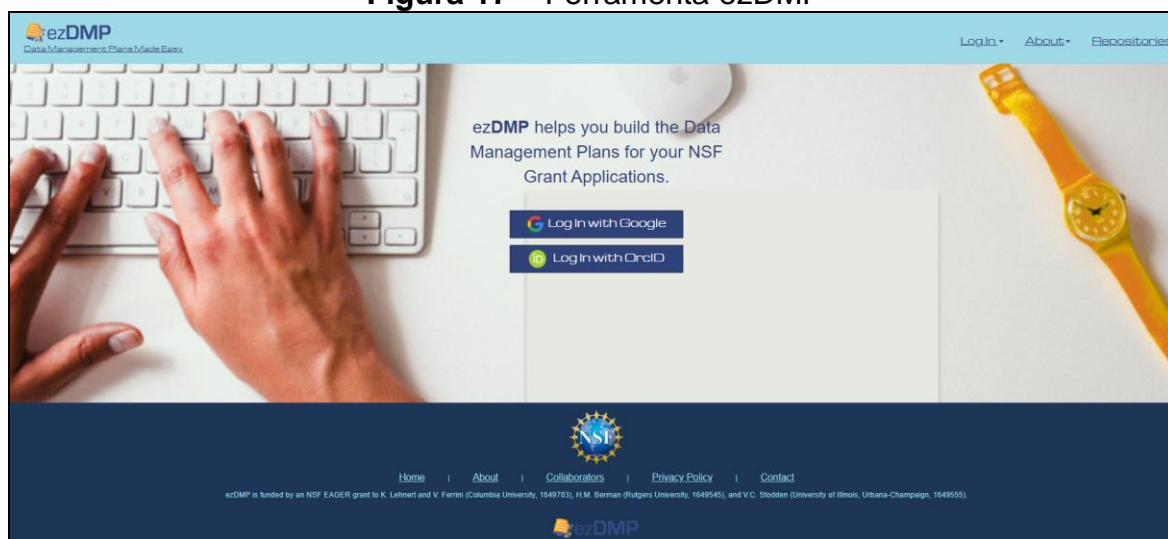
Percebe-se que a DataWiz se apresenta como uma ferramenta especializada e voltada para a gestão de dados no contexto da psicologia, proporcionando

³⁸<https://github.com/ZPID/DataWiz/>.

orientação específica para processamento e documentação de dados nesse domínio. Sua compatibilidade com importantes financiadores e organizações, bem como a flexibilidade no mecanismo de entrega e a visão de desenvolvimento futuro com a API planejada, sugerem um potencial significativo para aprimorar a gestão de dados na área da psicologia e em domínios relacionados, por meio do acionamento por máquina.

A ferramenta **ezDMP**³⁹ (Figura 17) é uma ferramenta de elaboração de planos de gestão de dados gratuita, lançada no ano de 2019. Baseada na ferramenta de gestão de dados da *Interdisciplinary Earth Data Alliance* (IEDA) e financiada pela *National Science Foundation* (NSF).

Figura 17 – Ferramenta ezDMP



Nota: ferramenta disponível em <https://ezdmp.org/index>.

A ezDMP concentra-se em perguntas estruturadas em vez de texto livre e permite a exportação em PDF. A ferramenta não apresenta suporte ao usuário durante a elaboração do PGD e tampouco a atribuição de um DOI. Não foram localizadas informações sobre a concordância com requisitos de agências de financiamento.

O mecanismo de entrega da ferramenta envolve hospedagem pela *Interdisciplinary Earth Data Alliance* (IEDA), indicando que a ferramenta é disponibilizada remotamente para os usuários. Essa abordagem simplifica a utilização da ferramenta, pois elimina a necessidade de configuração local ou manutenção

³⁹<https://ezdmp.org>.

técnica. A ezDMP disponibiliza documentação e API⁴⁰ em sua página para customizações.

A ferramenta **OpenDMP**⁴¹ (Argos) (Figura 18) representa uma abordagem inovadora para a gestão de dados, sendo desenvolvida em colaboração entre OpenAIRE e EUDAT. A ferramenta é uma plataforma de código aberto altamente configurável, projetada para consumir APIs públicas/abertas. O conceito fundamental é o perfil do conjunto de dados, o que significa que um PGD pode ter uma ou várias descrições de conjunto de dados, permitindo flexibilidade na gestão dos dados associados a diferentes projetos ou pesquisas. Em fase beta em 2020, a OpenDMP serviu de base para a ferramenta Argos.

A OpenDMP oferece exportação nos formatos PDF, DOC, XML, JSON e a atribuição de DOI. Disponibiliza também modelos de PGDs para financiadores e permite a criação de novos à medida dos requisitos da organização ou financiador. A oferta de assistência técnica por meio da OpenAIRE revela o compromisso de proporcionar suporte e orientação aos usuários. A disponibilidade para hospedagem do serviço, autoimplementação e instalação de uma instância local oferece aos usuários opções flexíveis de implementação, permitindo que escolham a melhor adequação ao seu contexto e requisitos. A colaboração entre OpenAIRE e EUDAT no financiamento da ferramenta também destaca uma abordagem colaborativa, reunindo recursos e esforços para impulsionar a eficácia da gestão de dados.

Do ponto de vista técnico, a OpenDMP é baseada em tecnologias *Java* e *Python* e é uma aplicação que promove a interoperabilidade aplicando o padrão comum RDA sob a licença Apache 2. A disponibilidade de APIs permite referenciar entidades (pesquisadores, organizações, conjuntos de dados, licenças, políticas, padrões etc.) e indica uma ênfase na interoperabilidade e integração com outros sistemas e ferramentas, promovendo uma maior flexibilidade no uso e na personalização da ferramenta, devido à possibilidade de acionamento por máquina.

⁴⁰<https://ezdmp.org/api>.

⁴¹<https://www.openaire.eu/opendmp-software-expands>.

Figura 18 – Ferramenta OpenDMP

The screenshot shows the OpenAIRE website interface. At the top, there is a navigation bar with 'Services', 'Support', 'Open Science In Europe', 'About Us', and a search icon. The main header area features a large 'News' graphic on the left and a central banner for 'ma-DMP' (Data Management Plan) with a 'Publish and retrieve' section listing integrations with Zenodo, Dataverse, ckan, and SPACE. Below the banner, a news article titled 'OpenDMP software expands' is displayed. The article includes a 'Published' date of Apr 5, 2023, an 'Author' section for Elli Unknown, and a '# Topics' section for 'dmp, data management plan, argos'. The main text of the article discusses the release of a new version (v1.6.1) of the OpenDMP software, highlighting new connectivity features and the ability to publish FAIR DMPs. A 'Follow Us' button is visible at the bottom left of the article content.

Nota: ferramenta disponível em <https://www.openaire.eu/opendmp-software-expands>.

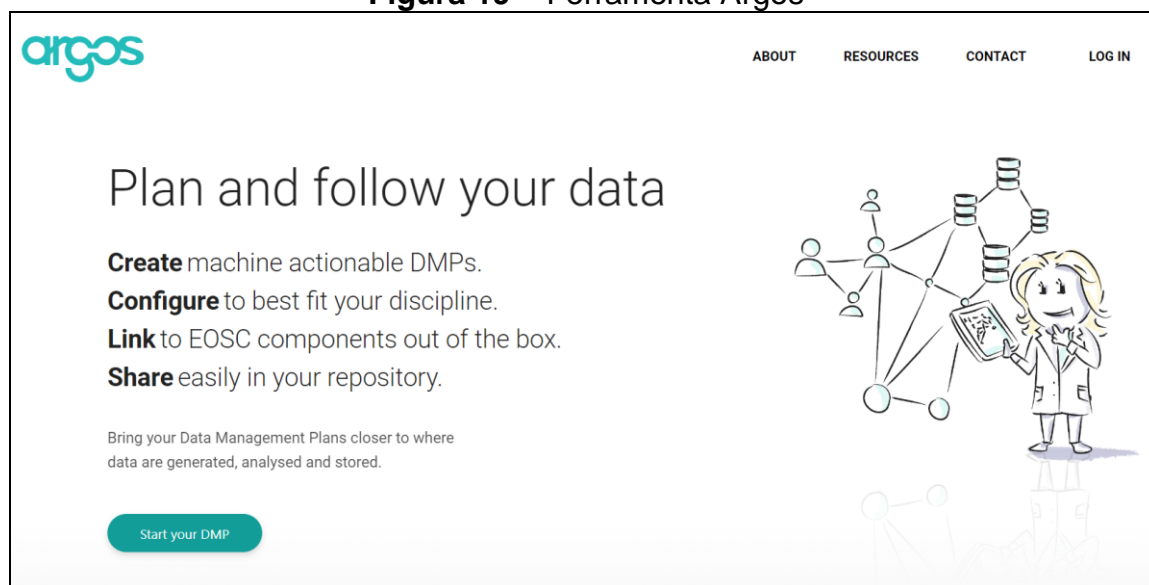
Em 2020, indicava-se que os custos ainda não estavam definidos e que seriam estabelecidos de acordo com o nível de suporte requerido pela organização.

A ferramenta **Argos**⁴² (Figura 19), lançada em 2017, foi desenvolvida pela OpenAIRE e EUDAT. Oferece funcionalidades abrangentes para a criação, gestão, validação, monitoramento, manutenção, compartilhamento e conexão dos planos de gestão de dados com os resultados de pesquisa correspondentes. Isso sugere que a ferramenta não apenas facilita a elaboração dos PGDs, mas também os integra de maneira eficaz com os dados resultantes da pesquisa, promovendo uma gestão coesa e integrada dos dados ao longo do ciclo de vida da pesquisa, por meio do acionamento por máquina.

O suporte fornecido pela OpenAIRE garante que os usuários tenham acesso à assistência técnica e orientação durante a utilização da ferramenta. Essa assistência assegura que os usuários possam aproveitar ao máximo os recursos oferecidos pela ferramenta e cumprir as boas práticas de gestão de dados.

⁴²<https://argos.openaire.eu/>.

Figura 19 – Ferramenta Argos



Nota: ferramenta disponível em <https://argos.openaire.eu/splash/>,

Em 2020, a Argos possuía 132 usuários, 85 PGDs e 91 descrições de *datasets*. A possibilidade de atribuir PIDs, como o DOI, ao PGD é uma característica relevante, pois facilita a citação e a referência adequada no contexto da pesquisa e publicação. Isso contribui para a melhor visibilidade e acessibilidade dos planos de gestão de dados. A Argos disponibiliza *templates* de financiadores e a possibilidade de financiadores adicionar, atualizar e modificar modelos de PGDs.

O fato de a ferramenta ter código aberto e estar disponível no *GitLab* do OpenDMP demonstra o compromisso com a transparência, colaboração e contribuição da comunidade. O mecanismo de entrega permite a instalação de uma instância local e serviço hospedado. A Argos possui licença MIT, que permite a customização do *software*, desde que mantenham a atribuição dos direitos autorais e a licença original.

Os PGDs podem ser exportados nos formatos PDF, DOC, XML e JSON. A presença de uma API⁴³ indica a preocupação com a interoperabilidade e integração com outros sistemas e ferramentas. Essa flexibilidade na integração é crucial para a utilização eficaz da ferramenta em um ecossistema diversificado de pesquisa.

A ferramenta **FioDMP**⁴⁴ (Figura 20) representa uma iniciativa concebida pela equipe do Instituto de Comunicação e Informação Científica e Tecnológica em Saúde com apoio do Grupo de Trabalho Ciência Aberta Fiocruz da Fundação Oswaldo Cruz.

⁴³<http://develop.openaire.eu/>.

⁴⁴<https://fiodmp.fiocruz.br/>.

As funcionalidades oferecidas pela FioDMP incluem uma estrutura de perguntas, facilitando a elaboração de planos de gestão de dados, com possibilidade de acionamento por máquina. Além disso, a possibilidade de compartilhamento com outras pessoas da equipe destaca uma abordagem colaborativa na criação dos PGDs, o que pode ser crucial em ambientes de pesquisa onde o trabalho em equipe é comum.

Figura 20 – Ferramenta FioDMP

Nota: ferramenta disponível em <https://fiodmp.fiocruz.br/>.

A integração com o repositório Arca Dados (2022) e a integração com vocabulários controlados como DeCS/Mesh⁴⁵ demonstram a preocupação com a padronização e a qualidade dos metadados associados aos dados de pesquisa. Essa padronização pode facilitar a busca e o entendimento dos dados, promovendo a reutilização e a interoperabilidade.

A integração com o *Crossref*⁴⁶ para afiliação dos autores e financiadores mostra um esforço em garantir a precisão e a conformidade dos metadados relacionados às fontes de financiamento e afiliações, aspectos fundamentais na gestão de dados. A capacidade de versionamento e a estrutura para múltiplos idiomas são elementos positivos, fornecendo flexibilidade e adaptabilidade à ferramenta para atender a

⁴⁵<https://decs.bvsalud.org/>.

⁴⁶*Crossref* é uma organização de infraestrutura digital aberta e lucrativa para a comunidade global de pesquisa acadêmica.

diversas necessidades de pesquisa e contextos multiculturais. A exportação dá-se em formatos comuns como PDF, JSON e DOCX. E há ainda, a possibilidade de atribuição de DOI.

Quanto aos usuários, a FioDMP é gratuita e direcionada principalmente para pesquisadores da Fiocruz e da área da saúde. Essa orientação específica reflete a afinidade da ferramenta com o domínio da saúde e da pesquisa realizada pela Fiocruz. O suporte à ferramenta é oferecido pelo Núcleo de Ciência Aberta, da Fiocruz. Não foram encontradas informações sobre o mecanismo de entrega, a fonte financiadora e se há conformidade com as diretrizes de agências de financiamento, bem como a disponibilidade de APIs. Da mesma forma, não foram encontradas informações sobre o *software*.

Recentemente lançada, em agosto de 2023, a FioDMP atinge o número de 36 usuários e sete instituições que utilizam a ferramenta, totalizando 110 planos. É fundamental observar a expansão desses números ao longo do tempo, pois isso pode indicar a crescente aceitação e relevância da ferramenta no cenário da gestão de dados de pesquisa no Brasil.

A **DMP OPIDoR**⁴⁷ (Figura 21) é uma ferramenta gratuita de elaboração de planos de gestão de dados baseada no código aberto da DMPRoadmap, que foi desenvolvida pela DCC e pela UC3, com o *software Ruby* e licença MIT. O código-fonte da ferramenta está disponível no GitHub⁴⁸ e sua API⁴⁹ em seu site. A DMP OPIDoR, lançada em 2019, foi financiada pela instituição de pesquisa francesa *Centre National de la Recherche Scientifique* (CNRS) e personalizada para pesquisadores franceses.

Hospedada pelo *Institut de l'information scientifique et technique* do *Centre National de la Recherche Scientifique* (Inist-CNRS), a DMP OPIDoR possui modelos de PGDs de acordo com financiadores e oferece suporte técnico do *Inist-CNRS Institut de l'Information Scientifique et Technique*. É possível informar um DOI, salvar diferentes versões para o controle do plano e exportar nos formatos HTML, PDF, DOCX, JSON, além da possibilidade de convidar colaboradores para a edição do PGD.

⁴⁷<https://dmp.opidor.fr/>.

⁴⁸<https://github.com/OPIDoR/DMPOPIDoR>.

⁴⁹<https://dmp.opidor.fr/api-docs/index.html>.

Figura 21 – Ferramenta DMP OPIDoR

The screenshot shows the DMP OPIDoR website interface. At the top, there is a navigation menu with 'DMP OPIDoR', 'DMP publics', 'Modèles de DMP', 'Aide', and 'Plus'. A language selector on the right indicates 'Français'. The main content area is titled 'Bienvenue !' and includes a welcome message: 'DMP OPIDoR vous accompagne à travers l'élaboration et la mise en pratique de plans de gestion de données et de logiciels.' Below this, there are five key features listed with icons: 'Accessible à la communauté scientifique de l'ESR et à ses partenaires français ou étrangers', 'Personnalisable par tout organisme de recherche pour la mise en place de sa politique de données', 'Enrichi par des exemples et des recommandations adaptés à l'environnement de recherche', 'Collaboratif pour permettre les échanges entre les partenaires d'un même projet et les services d'accompagnement', and 'Machine actionnable pour faciliter la saisie et les interactions avec les services impliqués dans la gestion des données'. At the bottom, there is a footer with the text: 'DMP OPIDoR évolue grâce à vos retours. Les développements s'inscrivent dans le cadre d'une collaboration internationale autour du logiciel open source DMPRoadmap'. On the right side, there is a sidebar for user authentication. It has two tabs: 'Se connecter' and 'Créer un compte'. Under 'Accès institutionnel', there is a 'Se connecter' button. Under 'Accès individuel', there are input fields for '* Courriel' and '* Mot de passe', a 'Mot de passe oublié ?' link, a checkbox for 'Se souvenir de l'adresse', and a 'Se connecter' button.

Nota: ferramenta disponível em <https://dmp.opidor.fr/>.

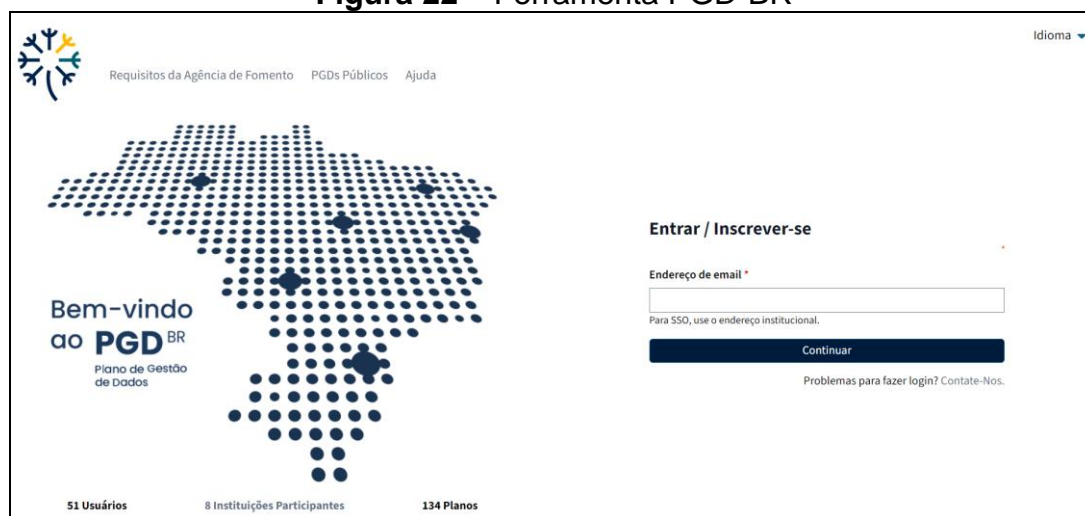
O fato de ser acionável por máquina indica que é uma ferramenta em conformidade com o cenário atual da pesquisa, onde a gestão eficaz de dados é essencial para garantir a qualidade e a reutilização dos dados de pesquisa.

A **PGD-BR**⁵⁰ (Figura 22) é uma ferramenta gratuita orientada para atender os pesquisadores brasileiros quanto à gestão dos dados de pesquisa. Com código aberto, modulada pelo Ibict a partir da adaptação da ferramenta DMPTool, da *California Digital Library*, é uma ferramenta que permite o acionamento por máquina. A disponibilidade de uma API⁵¹ reforça essa característica.

⁵⁰<https://pgd.ibict.br/>.

⁵¹<https://github.com/CDLUC3/dmptool/wiki/API-Overview>.

Figura 22 – Ferramenta PGD-BR



Nota: ferramenta disponível em <https://pgd.ibict.br/>.

O desenvolvimento da PGD-BR teve início em 2011 com o estabelecimento de uma parceria entre oito instituições, incluindo a *University of California Curation Center*, *California Digital Library*, *DataONE*, *Digital Curation Centre*, *Instituto Smithsonian* e várias bibliotecas universitárias dos Estados Unidos. Essas instituições uniram esforços para consolidar experiências e reduzir custos, visando atender às crescentes necessidades de gestão de dados. A partir da primeira versão bem-sucedida o projeto recebeu financiamento da *Alfred P. Sloan Foundation*⁵² para desenvolver uma segunda versão da ferramenta, lançada em 2014. A última versão da ferramenta foi anunciada em 2023, durante o VI WIDaT⁵³, e foi desenvolvido em colaboração com a Universidade Federal do Rio Grande do Sul (UFRGS), Universidade Estadual de Campinas (Unicamp) e Universidade Estadual Paulista (Unesp).

A PGD-BR possui uma rede de 77 Usuários, 8 instituições e 153 planos⁵⁴ e permite a exportação nos formatos CSV, HTML, PDF, TXT, DOCX e JSON, assim como o compartilhamento com colaboradores do projeto. O suporte fornecido pela ferramenta é um assistente de preenchimento para criar o PGD. A PGD-BR apresenta, ainda, a opção de idioma inglês e espanhol, além do português. Os modelos de planos de gestão de dados disponibilizados pela PGD-BR incorporam requisitos específicos

⁵²<https://sloan.org/>.

⁵³O Workshop de Informação, Dados e Tecnologia (WIDaT), concebido para proporcionar ambiente propício para o intercâmbio de ideias e experiências entre profissionais da indústria e pesquisadores das áreas de Ciência da Informação, Ciência da Computação, Engenharias e campos correlatos.

⁵⁴Dados coletados em: 08 de maio de 2024.

de PGDs de várias agências de fomento. Outra característica importante é a possibilidade de atribuição de DOI.

As 14 ferramentas de elaboração de planos de gestão de dados desta seção apresentam um conjunto diversificado de recursos, conforme síntese do Quadro 4:

Quadro 4 – Síntese das ferramentas de elaboração de planos de gestão de dados (Continua...)

FERRAMENTA	PROPRIETÁRIO/DESENVOLVIDOR	ANO	FUNÇÕES	SUORTE	USUÁRIOS	FINANCIAMENTO	SOFTWARE/LICENÇA DE DISTRIBUIÇÃO	INSTANCIAMENTO DA APLICAÇÃO	EXPORTAÇÃO
DMPonline	Digital Curation Centre (DCC) e pela University of California Curation Center (UC3).	2010	Criar, editar, compartilhar, revisar e publicar PGDs. Personalizado para necessidades de pesquisa, financiadores e instituições. Multilíngue. Escolha de padrão de metadados e disponibilização de glossários.	Central de Atendimento; Manual do usuário; Treinamento; Comunidade de desenvolvedores.	89 países: Reino Unido, Europa, África do Sul, Austrália; 133.054 usuários, 338 organizações, 149.604 PGDs.	Netherlands Organisation for Scientific Research, com a colaboração da California Digital Library (CDL) e DMPRoadmap. Subsídios; Taxas para instâncias hospedadas; Assinaturas de financiadores/instituições.	GitHub; Ruby on Rails; licença MIT; Baseado DMPRoadmap, de código aberto.	Nuvem, instâncias privadas.	PDF, DOC, XML e compartilhamento com colaboradores (via e-mail).
DMPTool	Projeto iniciado pela California Digital Library (CDL) em parceria com University of California Curation Center (UC3), DataONE, Digital Curation Centre (DCC), Smithsonian Institution, University of California, Los Angeles Library, University of California, San Diego Libraries, University of Illinois, Urbana-Champaign Library e University of Virginia Library.	2011	Criar, editar, compartilhar, revisar e publicar PGDs. Personalizado para necessidades de pesquisa, financiadores e instituições. Multilíngue.	Central de Atendimento; Guia de início rápido; Treinamento; Comunidade de desenvolvedores. Assistente <i>click-through</i> para criar um PGD que atenda aos requisitos do financiador e <i>links</i> para sites de financiadores, texto de ajuda para responder perguntas e recursos de práticas recomendadas de gestão de dados de pesquisa.	EUA, América do Sul; 36.800 usuários; 396 organizações; 111.539 PGDs.	Subsídios; suporte CDL.	GitHub; Ruby on Rails; licença MIT. Versão baseada no código aberto do DMPRoadmap.	Local, customização.	CSV, HTML, PDF, TXT, DOCX, JSON. e compartilhamento com colaboradores (via e-mail).
DMPRoadmap	Digital Curation Centre e University of California Curation Center, além de provedores dos países de Edimburgo, Escócia e Reino Unido.	2015	Criar, editar, compartilhar, revisar e publicar PGDs. Personalizado para necessidades de pesquisa, financiadores e instituições. Multilíngue.	Helpdesks de nível nacional; Treinamento; Guia do usuário.	89 países: Canadá, Finlândia, França, Bélgica, Espanha, Dinamarca, Japão, África do Sul; Cerca de 17.600 utilizadores; 203 organizações; 23.000 PGDs.	Projetos ou financiamento de consórcios.	GitHub; Ruby on Rails; licença MIT.	Não localizado.	CSV, HTML, PDF, TXT, DOCX, JSON. e compartilhamento com colaboradores (via e-mail).

Fonte: elaborado pela autora.

Quadro 5 – Síntese das ferramentas de elaboração de planos de gestão de dados (continuação)

FERRAMENTA	PROPRIETÁRIO/ DESENVOLVIDOR	ANO	FUNÇÕES	SUPORTE	USUÁRIOS	FINANCIAMENTO	SOFTWARE/ LICENÇA DE DISTRIBUIÇÃO	INSTANCIAMENTO DA APLICAÇÃO	EXPORTAÇÃO
EasyDMP	Sigma2 em colaboração com EUDAT2020.	2015	Criar PGDs de acordo com modelos personalizáveis.	Central de atendimento Sigma2; Documentação do usuário.	Noruega, Europa. 1.558 usuários, 1637 planos, 258 organizações.	UNINETT Sigma2 AS & EUDAT.	GitHub; Python; Licença do MIT planejada (atualmente não licenciada). <i>Open Source</i> .	Local, customizado.	HTML, PDF, JSON e convidar para editar e visualizar.
Data Stewardship Wizard (DSW)	<i>Dutch Techcentre for Life Sciences (DTL, ELIXIR NL) e Czech Technical University (CTU, ELIXIR CZ).</i>	2015	Questionário dinâmico hierárquico. Customização do modelo de conhecimento. Avaliação FAIR. Criação do PGD de forma colaborativa. Controle de versões. Permite desenvolver um modelo customizado que transforma as respostas em um questionário usando a linguagem de marcação de modelos Jinja2.	Central de atendimento; Documentação; canal frouxo; tutoriais sobre oficinas. GitHub.	Europa e Brasil; Parceiros da indústria. Instituições majoritariamente relacionadas com o domínio das Ciências da Vida.	Subsídios, financiamento ELIXIR, taxas de personalização e hospedagem.	GitHub; Elm, Haskell, Python; licença Apache 2. <i>Open Source</i> . Aplicação integrada num servidor da CESNET ou Via Docker.	Nuvem, instâncias privadas.	PDF, DOC, LaTeX, JSON, HTML, Markdown, OpenDocument, PDF, JSON, RDF/XML, Turtle, TRIG ou JSON-LD.
Research Data Management Organiser (RDMO)	<i>Leibniz-Institut für Astrophysik Potsdam (AIP), Potsdam University of Applied Sciences (FHP) e Karlsruhe Institute of Technology (KIT).</i>	2017	Existem três funções diferentes no RDMO: usuário, gerente e administradores. Planejamento, implementação e organização de dados de pesquisa. Multilíngue. Questionário estruturado, muitos formatos de saída.	Tutoriais; Canais de folga do desenvolvedor; lista de discussão.	Alemanha.	<i>German Research Foundation</i> em duas etapas, antes de ser repassado para um consórcio comunitário. Durante a etapa inicial, o <i>Leibniz-Institut für Astrophysik Potsdam, Potsdam University of Applied Sciences e Karlsruhe Institute of Technology</i> estiveram envolvidos; na segunda fase, a biblioteca do <i>Instituto de Tecnologia de Karlsruhe (KIT)</i> também participou.	GitHub. Código aberto.	Local, customização.	XML, CSV, JSON.

Fonte: elaborado pela autora.

Quadro 6 – Síntese das ferramentas de elaboração de planos de gestão de dados (continuação)

FERRAMENTA	PROPRIETÁRIO/ DESENVOLVIDOR	ANO	FUNÇÕES	SUPORTE	USUÁRIOS	FINANCIAMENTO	SOFTWARE/ LICENÇA DE DISTRIBUIÇÃO	INSTANCIAMENTO DA APLICAÇÃO	EXPORTAÇÃO
UQ Research Data Manager (UQRDM)	<i>University of Queensland.</i>	2017	Concentre-se na produção de registros de gestão de dados para capturar metadados e alocar armazenamento durante a entrega do projeto.	Central de Ajuda; Guia do usuário; Treinamento.	<i>University of Queensland</i> e colaboradores de pesquisa.	Financiado institucionalmente.	Código fechado, mas considerando tornar o código aberto.	Nuvem, instâncias privadas.	Não foi possível verificar.
DataWiz	<i>Leibniz Institute for Psychology Information (ZPID).</i>	2017	PGDs em conformidade com o financiador, suporte para processar e documentar dados psicológicos de acordo com os padrões, por exemplo, DDI. Multilíngue.	Guia do usuário. Metadados, vocabulários controlados DDI ou <i>Dublin Core</i> .	Alemanha.	<i>Leibniz Institute for Psychology Information (ZPID).</i>	GitHub; Java Empresarial; Licença GNU GPL. Código aberto.	Local.	XML, PDF e CSV.
ezDMP	<i>Interdisciplinary Earth Data Alliance (IEDA) e National Science Foundation.</i>	2019	Perguntas estruturadas em vez de texto livre para permitir análises sobre conformidade de dados.	Multilíngue.	EUA.	Financiado pela NSF.	Baseada na ferramenta de gestão de dados da <i>Interdisciplinary Earth Data Alliance (IEDA)</i> .	Local, customizado.	PDF.
OpenDMP	OpenAIRE & EUDAT.	Em beta em 2020.	O PGD pode ter uma ou várias descrições de conjunto de dados, podem ser co-autoria e publicada.	Assistência técnica OpenAIRE.	A ser lançado.	OpenAIRE & EUDAT.	GitHub; Java/Spring, Angular, Python/Django; sem licença. <i>Open Source</i> . Apache 2.	Nuvem e local.	PDF, DOC, XML, JSON.
ARGOS	OpenAIRE e EUDAT.	2017	Gestão, validação, monitorização e manutenção. Criação, gestão, partilha e ligação do PGD com os resultados de investigação correspondentes. Vocabulários.	OpenAIRE.	132 Utilizadores, 85 PGDs, 91 Descrições de Datasets.	OpenAIRE & EUDAT.	<i>Open Source</i> . Licença MIT.	Não localizado.	PDF, DOC, XML, JSON.

Fonte: elaborado pela autora.

Quadro 7 – Síntese das ferramentas de elaboração de planos de gestão de dados (continuação)

FERRAMENTA	PROPRIETÁRIO/ DESENVOLVIDOR	ANO	FUNÇÕES	SUPORTE	USUÁRIOS	FINANCIAMENTO	SOFTWARE/ LICENÇA DE DISTRIBUIÇÃO	INSTANCIAMENTO DA APLICAÇÃO	EXPORTAÇÃO
FioDMP	Fundação Oswaldo Cruz - Fiocruz. Concebido pela equipe do Instituto de Comunicação e Informação Científica e Tecnológica em Saúde com apoio do Grupo de Trabalho Ciência Aberta Fiocruz.	2020	Estrutura de perguntas; compartilha com outros da equipe; Integração com o repositório Arca Dados (em desenvolvimento); Integrado com vocabulário controlado DeCS/Mesh; Integração com <i>Crossref</i> ; Versionamento e estrutura para multi-idiomas; Integração com Crossref para afiliação e financiadores.	Núcleo de Ciência Aberta, da Fiocruz.	Pesquisadores da Fiocruz e do campo da saúde.	Não localizado.	Desenvolvido um próprio.	Não localizado.	PDF, JSON e DOCX.
DMP OPIDoR	<i>Centre National de la Recherche Scientifique</i> (CNRS).	2019	Modelos que atendem aos requisitos de órgãos e instituições financiadoras. Permite salvar versões.	Suporte técnico <i>Inist-CNRS Institut de l'Information Scientifique et Technique</i> .	Para pesquisadores franceses.	<i>Centre National de la Recherche Scientifique</i> .	Código aberto. Ruby e Licença MIT. Baseado no DMPRoadmap. https://github.com/OPIDoR/DMPOPIDoR	Não localizado.	Permite convidar colaboradores. HTML, PDF, DOCX, JSON.
PGD-BR	<i>California Digital Library da University of California Office of the President</i> e modulado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict).	2023	Português, inglês e espanhol.	A ferramenta fornece um assistente de preenchimento para criar um PGD.	Pesquisadores brasileiros. 77 Usuários, 8 instituições e 153 planos.	1ª versão: As instituições colaboradoras originais foram: <i>University of California Curation Center (UC3) em California Digital Library, DataONE, Digital Curation Centre (DCC-UK), Smithsonian Institution, University of California, Los Angeles Library, University of California, San Diego Libraries, University of Illinois, Urbana-Champaign Library</i> , e o <i>University of Virginia Library</i> . 2ª versão: financiamento da <i>Alfred P. Sloan Foundation</i> .	Adaptada a partir do software de código aberto DMPTool	Não localizado.	CSV, HTML, PDF, TXT, DOCX, JSON, compartilhamento com colaboradores.

Fonte: elaborado pela autora.

Essas ferramentas desempenham papel essencial no processo de planejamento e gestão dos dados de pesquisa. Cada ferramenta possui características distintas, abordagens e funcionalidades, mas todas convergem para o objetivo de facilitar a criação de planos de gestão de dados eficazes e de promover a transparência e a integridade na pesquisa. Assim, revela-se a variedade de ferramentas de elaboração de planos de gestão de dados com diversas características de funcionalidade e flexibilidade.

4.2 PLANOS DE GESTÃO DE DADOS ACIONÁVEIS POR MÁQUINA E OS PRINCÍPIOS DA *RESEARCH DATA ALLIANCE*

Para responder o objetivo “b”, que determina analisar as características e integrações dos maDMPs com os princípios para acionamento por máquina da *Research Data Alliance*, buscou-se explorar as particularidades das ferramentas identificadas a partir do objetivo “a” e encontrar consonância com cada um dos princípios para acionamento por máquina da RDA.

As ferramentas DMPonline, DMPTool, DMPRoadmap, EasyDMP, Data Stewardship Wizard (DSW) e Research Data Management Organiser (RDMO) são soluções que se alinham com o **primeiro princípio** de integrar os PGDs aos fluxos de trabalho das partes interessadas no ciclo de vida dos dados de pesquisa (**primeiro princípio**). Essas ferramentas facilitam a criação colaborativa dos planos de gestão, envolvendo atores de domínios específicos e adjacentes do ecossistema de gestão de dados. Além disso, fornecem suporte abrangente para o consumo de informações por diferentes atores, desde operadores de repositórios até financiadores de pesquisa.

Da mesma forma, as ferramentas OpenDMP, Argos, FioDMP, DMP OPIDoR e PGD-BR também compartilham o objetivo de integrar os PGDs nos fluxos de trabalho da comunidade de pesquisa. A OpenDMP, por exemplo, destaca-se por sua flexibilidade e facilidade de customização, permitindo a adaptação dos PGDs às necessidades específicas de projetos e pesquisas. A OpenDMP permite a interoperabilidade e integração com outros sistemas por meio de sua API, o que pode facilitar a integração com os demais atores do ecossistema de dados de pesquisa.

A Argos vai além da elaboração de planos de gestão de dados ao integrá-los eficazmente com os dados resultantes da pesquisa, promovendo uma gestão coesa

e integrada dos dados ao longo do ciclo de vida da pesquisa. Seu suporte técnico garante que os usuários possam aproveitar todas as funcionalidades da ferramenta, facilitando sua adoção e utilização.

Já a FioDMP destaca-se pela padronização e qualidade dos metadados associados aos dados de pesquisa, facilitando a integração dos PGDs a outros sistemas de informação, como é o caso da integração com o repositório Arca e os recursos para colaboração entre as partes interessadas. A DMP OPIDoR oferece flexibilidade e controle do plano de gestão. Por fim, a PGD-BR, por ser modelada a partir do código da DMPTool herda a possibilidade de integração de seus planos pelos diferentes atores da pesquisa.

As ferramentas acima citadas promovem mudanças organizacionais para tornar a gestão de dados mais dinâmica, pois contam com guias, tutoriais, central de atendimento e integração com outros sistemas, instituições e centros de informação. Com funcionalidades que incluem atribuição de identificadores persistentes aos planos, exportação de dados em diferentes formatos e até mesmo automação de partes do processo de gestão de dados, essas ferramentas ajudam a estruturar a informação e a tornar a gestão de dados mais eficiente em todo o ciclo de vida da pesquisa.

As ferramentas UQRDM, DataWiz e ezDMP não oferecem evidências claras de integração dos planos de gestão de dados aos fluxos de trabalho das partes interessadas. Por outro lado, a DataWiz destaca-se por sua flexibilidade de entrega e uma API planejada, sugerindo uma abordagem orientada para integração com outros sistemas de informação e fluxos de trabalho. Já a ezDMP, embora focada em perguntas estruturadas, carece de informações específicas sobre integração, mas pode herdar funcionalidades relacionadas à integração de sua base na DMPTool.

Contemplando o **segundo princípio** para atender os atores do ecossistema da pesquisa, as ferramentas DMPonline, DMPTool, DMPRoadmap, EasyDMP, DSW e RDMO apresentam características que permitem que sistemas automatizados atuem diretamente na gestão de dados. Essas ferramentas oferecem funcionalidades que possibilitam a automatização de diversas tarefas ao longo do ciclo de vida dos dados de pesquisa. Por exemplo, a DMPonline e a DMPTool permitem a coleta automática de dados administrativos, como afiliação e números de concessão. Ferramentas como a EasyDMP, DSW e RDMO apresentam funcionalidades automatizadas que ajudam a otimizar tempo, reduzir custos e fornecer informações mais precisas.

A ferramenta UQRDM satisfaz o **segundo princípio** da RDA, pois demonstra facilidade para a captura automatizada de metadados e a alocação de armazenamento durante os projetos de pesquisa, reduzindo assim a necessidade de intervenção humana adicional. Da mesma forma, a DataWiz apresenta características que se alinham ao **segundo princípio** da RDA. Sua capacidade de oferecer suporte multilíngue e integração com outros sistemas sugere que a ferramenta está preparada para permitir a automação na gestão de dados de pesquisa.

Outra ferramenta que atende o **segundo princípio** é a ezDMP, uma vez que possui a capacidade de atuar de forma automatizada na coleta e verificação de informações. A Argos e a OpenDMP também estão alinhadas com o **segundo princípio**, pois oferecem API para integração com outros sistemas e tecnologias modernas, permitindo que sistemas automatizados realizem tarefas em nome dos atores implicados. Além disso, sua capacidade de integrar planos de gestão de dados com os resultados de pesquisa contribui para a automação de processos relacionados à gestão de dados.

A ferramenta FioDMP, desenvolvida pela Fiocruz, demonstra conformidade com o princípio, dado que a integração com repositórios e vocabulários controlados indicam a capacidade da ferramenta de atuar de forma automatizada na gestão dos dados de pesquisa. Da mesma forma, a DMP OPIDoR permite a automação de tarefas relacionadas à gestão.

Das ferramentas analisadas, a DMPonline, a DMPTool, a DSW, a RDMO e a UQRDM explicitam o atendimento ao **terceiro princípio** ao permitirem o acionamento por máquina de políticas relacionadas a dados. Essas ferramentas foram projetadas não apenas para atender às necessidades dos pesquisadores e instituições na gestão de dados, mas também para garantir que essas políticas sejam aplicadas de forma consistente e eficaz para as máquinas que lidam com os planos de gestão de dados. Elas promovem a interoperabilidade e integração de dados ao permitir a atribuição de identificadores persistentes, pois facilitam, dessa forma, a rastreabilidade e a citabilidade adequadas. Essas ferramentas baseiam-se em código aberto, o que promove a incorporação de políticas voltadas para máquinas.

As ferramentas UQRDM, DataWiz, ezDMP, OpenDMP e FioDMP descrevem características que possibilitam o acionamento por máquina, entretanto não demonstram o atendimento claro à possibilidade de interpretação e implementação de

políticas voltadas aos dados, garantindo que a gestão dos dados de pesquisa seja mais eficiente e transparente.

Para o **quarto princípio** da RDA, que envolve descrever tanto para máquinas quanto para humanos os componentes do ecossistema de gestão de dados, a DMPonline, a DMPTool e a DMPRoadmap fornecem funcionalidades para a criação e administração de planos de gestão de dados, permitindo o compartilhamento das informações do PGD. Tanto a DMPonline, a DMPTool quanto a DMPRoadmap oferecem infraestrutura robusta, com acionamento por máquina.

A DMPonline, por exemplo, oferece funcionalidades para a criação e administração de planos de gestão de dados de forma personalizada para atender às necessidades específicas de pesquisas, financiadores e instituições. Essas funcionalidades garantem que os PGDs sejam compreendidos e executados corretamente tanto por humanos quanto por máquinas. O código aberto da DSW possibilita a personalização das ferramentas para atender às necessidades específicas de diferentes atores e pesquisa, garantindo que os componentes do ecossistema de gestão de dados sejam adaptáveis e eficazes tanto para humanos quanto para máquinas.

Outras ferramentas que atendem o **quarto princípio** são a RDMO e a ezDMP, pois oferecem a disponibilização de API, indicando a ênfase na interoperabilidade e integração com outros sistemas. A UQRDM se destaca na simplificação dos fluxos de trabalho do projeto, na facilidade de colaboração com pesquisadores externos de qualquer instituição ou organização. A FioDMP incorpora na descrição de seus planos alguns diferenciais. Entre eles estão a integração com um repositório de dados, vocabulário controlado e a integração com identificadores persistentes de afiliações e financiadores.

De acordo com as descrições das ferramentas, pôde-se identificar as que facilitam o uso de Identificadores de Objetos Persistentes e vocabulários controlados para tornar os PGDs explícitos e compreensíveis para todas as partes interessadas, satisfazendo o **quinto princípio** da RDA.

As ferramentas DMPonline, DMPTool, OpenDMP, Argos, FioDMP, DMP OPIDoR e PGD-BR permitem, declaradamente, a atribuição do identificador persistente (PID) DOI ao plano de gestão de dados, evocando a unicidade e permanência da identificação dos dados. Da mesma forma, a DMPRoadmap também permite a atribuição de DOI, além de ORCID aos planos, garantindo a identificação

única e permanente dos recursos. Assim como a DMPonline, a DMPTool adota uma abordagem de código aberto, promovendo a sua customização. Outras ferramentas que atendem ao **quinto princípio** são a DSW e a RDMO. As ferramentas facilitam a atribuição de DOIs aos planos de gestão de dados, contribuindo para a identificação única e permanente dos recursos gerenciados.

A OpenDMP destaca o uso de tecnologias modernas e é uma aplicação de código aberto, o que revela a possibilidade de integrar PIDs e vocabulários controlados. Ademais, a ênfase na interoperabilidade e integração com outros sistemas promove o uso de padrões bem definidos. A UQRDM destaca a importância de capturar metadados durante os projetos de pesquisa. Já a ferramenta FiodMP menciona a integração com vocabulários controlados, como DeCS/Mesh e com o Crossref, para afiliação dos autores e financiadores, manifestando a padronização e precisão dos metadados e a contemplação do uso de PIDs.

No que se refere ao **sexto princípio** da RDA, que envolve seguir um modelo de dados comum para maDMPs, permitindo a interoperabilidade entre diferentes partes interessadas, as ferramentas DMPonline, DMPTool, DAMPRoadmap, EasyDAMP, RDMO, UQRDM, DataWiz, OpenDMP, Argos, DAMP OPIDoR e PGD-BR destacam-se pela capacidade de fornecer um modelo de dados comum para maDMPs.

A RDMO, mesmo que não manifeste seguir o **sexto princípio**, tem a capacidade de atribuir DOI aos dados gerenciados e dispõe de uma API para integrações personalizadas. Essas capacidades permitem seguir um modelo de dados comum e promover a interoperabilidade entre diferentes sistemas e serviços relacionados ao PGD.

As ferramentas UQRDM, DataWiz, DPMTTool e OpenDMP, por declarar-se acionáveis por máquina, também satisfazem o **sexto princípio**, uma vez que fazem uso do código aberto, o que sugere seguir um modelo de dados comum para promover a interoperabilidade. A UQRDM, por exemplo, expõe que é projetada para interoperabilidade com outros sistemas e serviços, alinhando-se ao princípio de seguir um modelo de dados comum. E a DataWiz, com base na sua orientação para seguir padrões estabelecidos, pode indicar conformidade implícita com um modelo de dados comum, especialmente considerando a sua ênfase na documentação estruturada de dados.

A FioDMP disponibiliza estrutura de perguntas para a elaboração de planos de gestão de dados e integração com o repositório Arca Dados e vocabulários controlados, essa integração e padronização pressupõe que a ferramenta segue um modelo de dados comum. Assim, essas ferramentas parecem estar alinhadas com o princípio de seguir um modelo de dados comum para maDMPs, seja implicitamente por seguir padrões reconhecidos na área de gestão de dados ou explicitamente através de interoperabilidade e integração com outros sistemas.

O **sétimo princípio** preconiza disponibilizar os planos de gestão de dados para consumo humano e de máquinas devido à sua capacidade de acionamento por máquina. A DMPonline e a DMPTool satisfazem o **sétimo princípio** em razão do fornecimento de funcionalidades que permitem a revisão, compartilhamento e publicação de planos de gestão de dados.

As ferramentas DSW, EasyDMP, DMPRoadmap, Argos e RDMO também oferecem acionamento por máquina e o uso de tecnologias de código aberto e a disponibilidade de API para integração com outros sistemas são aspectos positivos que contribuem para a capacidade de atender às demandas de consumo humano e de máquinas. A UQRDM demonstra forte compromisso com a disponibilização dos planos de gestão de dados para consumo humano e de máquinas, pois suporta o acionamento por máquina e a ênfase na documentação e gestão de metadados revelam a interoperabilidade e reutilização dos dados.

Apoiar a avaliação e monitoramento da gestão dos dados de pesquisa é a premissa do **oitavo princípio**. As ferramentas DMPonline, DMPTool, DataWiz e FioDMP são exemplos de ferramentas que se destacam no cumprimento desse princípio. A capacidade de acionamento por máquina e a possibilidade de compartilhamento colaborativo do plano de gestão de dados promovem o monitoramento automatizado da gestão dos dados de pesquisa. Igualmente, a ferramenta RDMO, com a disponibilização de três funções para a ferramenta, a saber usuário, gestor e administrador, cumpre o **oitavo princípio**, pois cobre as informações pertinentes à avaliação e monitoramento.

Da mesma forma, a FioDMP suporta o acionamento por máquina, integração com repositórios e vocabulários controlados, demonstrando o compromisso com a padronização e qualidade dos metadados associados aos dados de pesquisa e permitindo a execução do **oitavo princípio**. Também a Argos e a DMP OPIDoR

permitem a integração com os resultados da pesquisa, colaborando com o pressuposto do princípio.

Analisando as descrições das ferramentas apresentadas, identificou-se aquelas que satisfazem o **nono princípio**: tornar os planos de gestão de dados documentos atualizáveis, vivos e com controle de versão. As ferramentas DMPonline e a DMPTool demonstram a capacidade de atender esse princípio. Ambas as ferramentas permitem a criação, edição, revisão e publicação de planos de gestão de dados, mantendo um controle de versão que registra as alterações ao longo do tempo. Além disso, oferecem suporte para notificações configuráveis que informam as partes interessadas sobre mudanças nos planos, como condições de disponibilização de dados ou atualizações nos conjuntos de dados depositados.

A DMPRoadmap, EasyDMP, DSW e FioDMP também oferecem funcionalidades que permitem a atualização e o controle de versão dos planos de gestão de dados. Essas ferramentas permitem a criação e edição de PGDs, acompanhando as mudanças ao longo do tempo e registrando os principais eventos do projeto.

Verificou-se também, que a UQRDM oferece atualizações automáticas e notificações configuráveis para partes interessadas. A DataWiz, da mesma forma, satisfaz esse princípio, possibilitando atualizações automáticas e integração com outros sistemas, o que contribui para a vivacidade e controle de versão dos PGDs. A ezDMP apresenta atualizações automáticas com base em eventos do ecossistema de pesquisa. Sua base na ferramenta DMPTool e a integração com sistemas de acionamento por máquina garantem que os PGDs possam evoluir conforme necessário ao longo do ciclo de vida da pesquisa. A OpenDMP e a FioDMP, embora permita a criação de perfis de conjunto de dados e múltiplas descrições, não deixa claro o controle de versão.

O **décimo princípio** da *Research Data Alliance* que recomenda disponibilizar publicamente os planos de gestão de dados é contemplado pelas ferramentas DMPonline, DMPTool, DMPRoadmap, DSW, OpenDMP, FioDMP, DMP OPIDoR e PGD-BR. Essas ferramentas permitem compartilhar, editar e publicar PGDs.

A FioDMP é destacada por sua capacidade de compartilhar PGDs com outras pessoas da equipe, revelando a abordagem colaborativa na criação dos planos. Além disso, integração com o repositório Arca Dados e com o Crossref demonstra um

esforço em disponibilizar os planos de gestão de dados publicamente e garantir sua visibilidade.

Diante da análise das características e integrações dos maDMP com os princípios para acionamento por máquina da *Research Data Alliance*, o Quadro 5 apresenta uma síntese geral do atendimento dos dez princípios da RDA de acordo com as ferramentas de elaboração de planos de gestão de dados elencadas nesta pesquisa.

Quadro 8 – Observância dos princípios RDA de acordo com as ferramentas

Ferramenta	Princípios <i>Research Data Alliance</i>										Número de princípios observados
	1 ⁵⁵	2 ⁵⁶	3 ⁵⁷	4 ⁵⁸	5 ⁵⁹	6 ⁶⁰	7 ⁶¹	8 ⁶²	9 ⁶³	10 ⁶⁴	
DMPonline	✓	✓	✓	✓	✓	✓	✓	✓		✓	9
DMPTool	✓	✓	✓	✓	✓	✓	✓	✓		✓	9
DMPRoadmap	✓	✓		✓	✓	✓	✓		✓	✓	8
Data Stewardship Wizard (DSW)	✓	✓	✓	✓	✓		✓		✓	✓	8
Research Data Management Organiser (RDMO)	✓	✓	✓	✓	✓	✓	✓	✓			8
FioDMP	✓	✓		✓	✓	✓		✓	✓	✓	8
UQ Research Data Manager (UQRDM)		✓	✓	✓	✓	✓	✓		✓		7
ARGOS	✓	✓			✓	✓	✓	✓			6
DMP OPIDoR	✓	✓			✓	✓		✓		✓	6
EasyDMP	✓	✓				✓	✓		✓		5
OpenDMP	✓	✓			✓	✓				✓	5
DataWiz		✓				✓		✓	✓		4
PGD-BR	✓				✓	✓				✓	4
ezDMP		✓		✓					✓		3

Fonte: elaborado pela autora.

Percebe-se que as ferramentas listadas, mesmo produzindo planos de gestão de dados acionáveis por máquina, não observam todos os princípios da *Research*

⁵⁵1: integrar os PGDs aos fluxos de trabalho de todas as partes interessadas no ecossistema de dados de pesquisa.

⁵⁶2: permitir que sistemas automatizados atuem em nome das partes interessadas.

⁵⁷3: faça políticas (também) para máquinas, não apenas para pessoas.

⁵⁸4: descrever – tanto para máquinas quanto para humanos – os componentes do ecossistema de gerenciamento de dados.

⁵⁹5: use PIDs e vocabulários controlados.

⁶⁰6: siga um modelo de dados comum para maDMP.

⁶¹7: disponibilizar PGDs para consumo humano e de máquinas.

⁶²8: apoiar a avaliação e monitorização da gestão de dados.

⁶³9: tornar os PGDs documentos atualizáveis, vivos e com controle de versão.

⁶⁴10: disponibilizar publicamente os PGDs.

Data Alliance. Embora os dez princípios propostos beneficiem cada parte interessada do ecossistema da pesquisa científica, os princípios são independentes e podem ser implementados de forma progressiva.

4.3 ATORES DO ECOSSISTEMA DA PESQUISA BRASILEIRA

Os atores identificados neste capítulo ilustram os principais indivíduos, entidades e sistemas de informação que participam e influenciam (Oliveira, 2019; Vidal, 2023) o ecossistema da pesquisa brasileira encontrados na literatura, satisfazendo o objetivo “c”. O cenário do ecossistema da pesquisa brasileira é composto por uma gama de atores que podem vir a desempenhar diversos papéis ao longo dos fluxos de automatização dos planos de gestão de dados. Esses atores englobam projetos de pesquisa, revistas científicas e agregadores, repositórios digitais, bibliotecas digitais, programas de pós-graduação, instituições de pesquisa, agências de fomento, identificadores persistentes, pesquisadores, bases de patentes, conjuntos de dados de pesquisa e produção acadêmica e científica.

Carvalho Segundo *et al.* (2022) elencam uma série de atores que compõem o ecossistema da pesquisa brasileira no modelo BrCris, interagindo de forma complexa e interdependente nesse ecossistema. Assim, esta pesquisa vale-se do estudo de Carvalho Segundo *et al.* (2022) para a apresentação dos atores do ecossistema da pesquisa, além da apresentação de outros atores encontrados dispersos na literatura. Os atores identificados acompanham o fluxo da pesquisa científica e foram categorizados de acordo com o papel que desempenham no ecossistema da pesquisa brasileira.

As categorias listadas foram baseadas no trabalho de Carvalho Segundo *et al.* (2022) e nas características dos atores encontrados, conforme o Quadro 6. É importante ressaltar que devido aos atributos e particularidades de alguns atores, e as funções que exercem, eles podem aparecer em mais de uma categoria. Assim, posteriormente, os atores foram apresentados individualmente.

Quadro 9 – Categorias e atores do ecossistema da pesquisa brasileira

CATEGORIA	ATORES
Revistas científicas e agregadores	Oasisbr
	Diadorim
	Latindex
	DOAJ
	UlrichsWeb
	Revistas científicas
Programas de pós-graduação	Plataforma Sucupira
	Capes
	Plataforma Lattes
Repositórios digitais	Plataforma Sucupira
	BDTD
	Capes
Instituições de pesquisa e pesquisadores	Plataforma Sucupira
	Instituições do INEP
	Capes
	Plataforma Lattes
	OpenAIRE Research Graph
Bibliotecas digitais	Oasisbr
	Capes
	BDTD
Bases de patentes	Instituto Nacional da Propriedade Industrial (INPI)
	Latipat
	Espacenet
	Patentscope
Conjuntos de dados de pesquisa	OpenAIRE Research Graph
	Portal Wikidata
Produção acadêmica e científica	Oasisbr
	BDTD
	Capes
	OpenAIRE Research Graph
	La Referencia
	NDLTD
	RCAAP
Agências de fomento	CNPq
Projetos de pesquisa	OpenAIRE Research Graph
Identificadores persistentes	DOI
	ISSN
	ISBN
	ORCID
	ISNI
	ROR

Fonte: elaborado pela autora.

A partir das características dos atores, foram criadas 11 categorias, são elas: **Revistas científicas e agregadores**, que abrange atores que desempenham papéis importantes na disseminação e indexação da produção científica periódica. Entre os atores, encontram-se o Oasisbr, o Diadorim, o Latindex, o DOAJ, o UlrichsWeb e as revistas científicas. Esses atores promovem o conhecimento científico, a produção e a acessibilidade científica por meio de suas plataformas. A categoria de **Programas**

de pós-graduação apresenta atores que desempenham papel no contexto acadêmico e científico. Os atores são a Plataforma Sucupira, a Capes e a Plataforma Lattes.

Outra categoria, a de **Repositórios digitais**, inclui plataformas e iniciativas voltadas a preservação e a disseminação da produção científica. São elas, a Plataforma Sucupira, utilizada para acompanhar e avaliar programas de pós-graduação no Brasil, a BDTD e a CAPES, que mantém repositórios digitais como parte de suas iniciativas de apoio à pesquisa e à pós-graduação, oferecendo um espaço para o armazenamento e acesso a documentos científicos.

Há ainda, a categoria **Instituições de pesquisa e pesquisadores**, a qual pertence a Plataforma Sucupira, as instituições vinculadas ao Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), a CAPES, a Plataforma Lattes e o *OpenAIRE Research Graph*. Esses atores promovem a pesquisa e o avanço do conhecimento científico. Já a categoria **Bibliotecas Digitais** compreende atores que disponibilizam e oferecem acesso à informação científica e acadêmica. Esta categoria é composta pelo Oasisbr, a Capes e a Biblioteca Digital de Teses e Dissertações (BDTD).

A categoria de **Bases de patentes** apresenta atores relacionados a busca, consulta e análise de informações relacionadas a patentes e propriedade intelectual. Os atores são o Instituto Nacional da Propriedade Industrial (INPI), que fornece acesso a bancos de dados de patentes nacionais e internacionais, a Latipat, que reúne informações sobre patentes de países da América Latina, o Espacenet, que oferece acesso gratuito a documentos de patentes de todo o mundo e o Patentscope, mantida pela Organização Mundial da Propriedade Intelectual (OMPI).

Encontra-se também, a categoria **Conjuntos de dados de pesquisa**, com atores que disponibilizam e promovem o acesso a conjuntos de dados de pesquisa, a exemplo do *OpenAIRE Research Graph* e do Portal Wikidata. Tanto o *OpenAIRE Research Graph* quanto o Portal Wikidata promovem a transparência, a reprodutibilidade e o compartilhamento de dados de pesquisa. A categoria **Produção acadêmica e científica** envolve uma variedade de atores que dão acesso a artigos científicos, teses, dissertações, livros, capítulos de livro e relatórios científicos. A categoria compreende o Oasisbr, a Biblioteca Digital de Teses e Dissertações (BDTD), a Capes, o *OpenAIRE Research Graph*, a Rede Latino-Americana de Ciência Aberta

(La Referencia), o *Networked Digital Library of Theses and Dissertations* (NDLTD) e o Repositórios Científicos de Acesso Aberto de Portugal (RCAAP).

As categorias **Agências de fomento e Projetos de pesquisa** contam apenas com um ator cada uma. A categoria de “Agências de Fomento” destaca o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) enquanto o ator de apoio financeiro e o estímulo à pesquisa científica e tecnológica. E a categoria “Projetos de Pesquisa” apresenta o *OpenAIRE Research Graph*. Por fim, a categoria **Identificadores persistentes** compreende o DOI (*Digital Object Identifier*), o ISSN (*International Standard Serial Number*), o ISBN (*International Standard Book Number*), o ORCID (*Open Researcher and Contributor ID*), o ISNI (*International Standard Name Identifier*) e o ROR (*Research Organization Registry*).

Explorando as categorias, tem-se entre os atores do ecossistema da pesquisa brasileira a **Plataforma Lattes**⁶⁵, do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), reúne cerca de sete milhões de currículos de pesquisadores, além de grupos de pesquisa e diretórios de instituições (Carvalho Segundo *et al.*, 2022). A Plataforma Lattes é um sistema de currículos acadêmicos desenvolvido e mantido pelo CNPq no Brasil. É amplamente utilizada por pesquisadores, professores e profissionais acadêmicos brasileiros para registrar e divulgar informações sobre suas carreiras acadêmicas e atividades de pesquisa.

A Plataforma Lattes (Conselho Nacional de Desenvolvimento Científico e Tecnológico, 2023) é uma ferramenta que possibilita o acompanhamento e a avaliação da produção científica e acadêmica no Brasil. Os currículos Lattes contêm uma variedade de informações, incluindo: dados pessoais e informações de contato; formação acadêmica, incluindo graduações, pós-graduações e outras qualificações; produção científica, como artigos em periódicos, apresentações em conferências, livros e capítulos de livros; atividades de pesquisa, projetos de pesquisa e participação em grupos de pesquisa; atividades de ensino, incluindo disciplinas ministradas e orientações de estudantes de graduação e pós-graduação; participação em eventos científicos e comissões avaliadoras; além da experiência profissional e atuação em instituições acadêmicas e de pesquisa.

A Plataforma Lattes é usada para avaliação acadêmica, avaliação da pesquisa, alocação de bolsas de estudo e financiamento de pesquisa, entre outros fins

⁶⁵<http://lattes.cnpq.br/>.

(Conselho Nacional de Desenvolvimento Científico e Tecnológico, 2023). Permite que as instituições de ensino, agências de fomento à pesquisa e outros órgãos tenham acesso a informações abrangentes sobre o perfil acadêmico e científico dos indivíduos, facilitando a tomada de decisões relacionadas à pesquisa e ao ensino no Brasil. É uma ferramenta valiosa para a gestão da pesquisa e da ciência no país.

O **Oasisbr**⁶⁶ é o Portal Brasileiro de Publicações e Dados Científicos em Acesso Aberto do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) (Carvalho Segundo *et al.*, 2022), que apoia os princípios de acesso aberto à informação científica e contribui para o desenvolvimento da pesquisa e do conhecimento no Brasil.

O Oasisbr desempenha papel importante na disseminação e acessibilidade de informações científicas e acadêmicas produzidas no país, pois oferece acesso a uma ampla variedade de recursos, incluindo artigos científicos, teses, dissertações, relatórios técnicos, livros e outros tipos de publicações acadêmicas e científicas. O portal também disponibiliza dados de pesquisa em acesso aberto, o que é fundamental para promover a transparência e a reprodutibilidade na pesquisa científica (Instituto Brasileiro de Informação em Ciência e Tecnologia, 2023).

O **Diadorim**⁶⁷, serviço que oferece informações sobre as autorizações concedidas para o armazenamento e acesso dos artigos das revistas brasileiras em repositórios digitais de acesso aberto, integra o conjunto de serviços de acesso aberto do Ibict. As informações disponibilizadas são obtidas diretamente dos editores das revistas científicas brasileiras, cuja colaboração é valorizada para a atualização e precisão dos dados. O objetivo principal do Diadorim é identificar, sistematizar e disponibilizar as políticas estabelecidas pelas editoras em relação ao armazenamento dos artigos nos repositórios institucionais.

O **Latindex**⁶⁸ é um sistema resultante da colaboração de uma rede de instituições que trabalham em conjunto para coletar e disseminar informações sobre publicações científicas seriadas produzidas na América Latina. Criada em 1995, na Universidade Nacional Autônoma do México (UNAM), evoluiu em 1997 para uma rede de cooperação regional. O Latindex oferece dois produtos de informação distintos: o diretório, que contém dados bibliográficos e de contato de todas as revistas impressas

⁶⁶<https://oasisbr.ibict.br/vufind/>.

⁶⁷<https://diadorim.ibict.br/>.

⁶⁸<https://latindex.org/latindex/>.

e *online* cadastradas, e o catálogo 2.0, que engloba apenas revistas *online* que atendem aos padrões de qualidade, conforme a metodologia do Latindex.

O sistema abrange revistas de pesquisa científica, técnico-profissional e de divulgação científica e cultural publicadas na América Latina, Caribe, Espanha e Portugal, além de fornecer informações sobre revistas com conteúdo latino-americano publicadas em qualquer lugar do mundo e abrangem todas as disciplinas científicas.

A **Ulrichsweb**⁶⁹ é uma ferramenta que auxilia a busca por informações sobre publicações seriadas em um único local. A Ulrichsweb fornece detalhes bibliográficos e de fornecedores atualizados. Compreende mais de 300.000 periódicos de todos os tipos, abrangendo vasta gama de assuntos e áreas do conhecimento, seja em publicações periódicas ou não. A Ulrichs inclui tanto publicações regulares quanto irregulares, e cobre periódicos de acesso gratuito e por assinatura. Fornece ainda, dados de edição das revistas, indica em quais bases de dados elas estão indexadas e oferece visão abrangente e detalhada do cenário editorial científico.

As **Revistas Científicas** são publicações periódicas que servem como meio para a disseminação e validação do conhecimento científico, permitindo que pesquisadores possam publicar seus achados científicos e teóricos (Garfield, 1979). As revistas científicas são essenciais para o avanço do conhecimento em diversas áreas do conhecimento. Os artigos submetidos passam por um rigoroso processo de revisão por pares, onde especialistas na área avaliam a qualidade, a relevância e a originalidade dos trabalhos antes de sua publicação.

Ademais, as revistas científicas facilitam a comunicação e a colaboração entre pesquisadores de diferentes partes do mundo. Ao disponibilizar resultados de pesquisas para a comunidade científica global, elas promovem o compartilhamento de conhecimentos, incentivando colaborações que podem levar a novos avanços.

A **Biblioteca Digital Brasileira de Teses e Dissertações**⁷⁰ (BDTD) de acesso aberto, desenvolvida e coordenada pelo Ibict (Carvalho Segundo *et al.*, 2022), é uma plataforma *online* que reúne e disponibiliza teses e dissertações de instituições brasileiras de ensino superior produzidas no país e no exterior. Foi criada com o objetivo de democratizar o acesso ao conhecimento acadêmico. A BDTD oferece um repositório centralizado e de fácil consulta para estudantes, pesquisadores e

⁶⁹<https://ulrichsweb.serialssolutions.com/login>.

⁷⁰<https://bdttd.ibict.br/vufind/>.

interessados em diversas áreas do saber (Biblioteca Digital Brasileira de Teses e Dissertação, 2023).

A BDTD é constituída por uma estrutura descentralizada de sistemas de gestão de informações que gerenciam trabalhos acadêmicos completos, abrangendo tanto teses quanto dissertações. Essa rede é composta por dois principais participantes: os provedores de dados, responsáveis pela administração dos depósitos e pela publicação dos documentos, disponibilizando os metadados para serem automaticamente coletados (*harvesting*); e os provedores de serviços, que oferecem serviços informativos com base nos metadados coletados dos provedores de dados (Biblioteca Digital Brasileira de Teses e Dissertação, 2023).

A **Plataforma Sucupira**⁷¹ é uma ferramenta digital utilizada no contexto acadêmico brasileiro para a gestão e avaliação de programas de pós-graduação. Desenvolvida pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). O nome “Sucupira” é uma sigla que corresponde a “Sistema Único de Comunicação Pública da Produção Acadêmica”, refletindo sua finalidade de centralizar informações sobre a produção científica dos programas de pós-graduação (Brasil, 2023c).

A plataforma Sucupira é essencial para o monitoramento e aprimoramento da qualidade da pós-graduação no Brasil, permitindo a coleta, armazenamento e análise de dados relacionados à produção acadêmica, corpo docente, infraestrutura, entre outros aspectos relevantes para a avaliação dos programas de pós-graduação (Brasil, 2023c). Os programas de pós-graduação utilizam a plataforma Sucupira para enviar informações sobre seus projetos, atividades, publicações, teses, dissertações, além de dados sobre seus docentes e discentes. Esses dados são fundamentais para a avaliação dos programas pela CAPES, contribuindo para a definição de critérios de qualidade e para a distribuição de recursos de financiamento e bolsas de estudo.

As **instituições do INEP**⁷² estão reunidas em um portal de instituições de ensino do país credenciadas no Ministério da Educação (Carvalho Segundo *et al.*, 2022). O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira é uma autarquia federal brasileira vinculada ao Ministério da Educação (MEC) que desempenha um papel fundamental na coleta, análise e disseminação de informações sobre a educação no Brasil. O INEP realiza pesquisas e estudos educacionais,

⁷¹<https://sucupira.capes.gov.br/sucupira/>.

⁷²<http://portal.mec.gov.br/instituicoes-credenciadas>.

desenvolve avaliações e exames nacionais, e promove a produção de dados relevantes para o sistema educacional do país (Brasil, 2022). O INEP é formado por várias instituições e órgãos que desempenham funções específicas relacionadas à educação e à pesquisa educacional. O INEP trabalha para fornecer dados confiáveis e informações que auxiliam na tomada de decisões e no aprimoramento do sistema educacional do país.

Alguns dos principais órgãos e instituições ligados ao INEP incluem o ENEM (Exame Nacional do Ensino Médio), que é o exame nacional aplicado anualmente para avaliar o desempenho dos estudantes no ensino médio e é utilizado como critério de seleção para o ingresso em instituições de ensino superior no Brasil; o SAEB (Sistema de Avaliação da Educação Básica), que é um sistema de avaliação que avalia a qualidade da educação básica no país, aplicando provas e coletando dados sobre o desempenho dos alunos em diversas disciplinas; o censo da educação, que coleta informações demográficas, educacionais e financeiras sobre as instituições de ensino, alunos e profissionais da educação em todo o Brasil; o INEP-DATASUS, que consiste em uma parceria entre o INEP e o Departamento de Informática do Sistema Único de Saúde (DATASUS) e busca integrar informações sobre saúde e educação (Brasil, 2022).

O **OpenAIRE Research Graph**⁷³ é uma iniciativa que visa criar uma infraestrutura aberta e interconectada para a pesquisa acadêmica. É uma grande base de dados aberta que integra e conecta várias fontes de informação relacionadas à pesquisa, incluindo publicações científicas, dados de pesquisa, projetos de pesquisa, organizações de pesquisa e autores (Openaire Graph, 2023). Ao conectar e agregar informações de diversas fontes, a *OpenAIRE Research Graph* oferece uma visão mais completa e integrada da pesquisa, contribuindo para a promoção da ciência aberta e facilitando a interoperabilidade entre sistemas e serviços relacionados à pesquisa (Openaire Graph, 2023).

A iniciativa *OpenAIRE Research Graph* faz parte do movimento de acesso aberto à pesquisa, simplificando a descoberta de recursos de pesquisa e metadados relevantes. Assim, garante que os resultados da pesquisa estejam disponíveis ao público gratuitamente sempre que possível. Além disso, o *OpenAIRE Research Graph* enfatiza a disponibilização de dados de pesquisa, apoia a conformidade com políticas

⁷³<https://www.openaire.eu/>.

de pesquisa e integra-se a outras infraestruturas de pesquisa, possibilitando a interoperabilidade e a colaboração entre sistemas. O uso de vocabulários controlados e padrões de metadados melhoram a qualidade e a precisão da descrição de recursos de pesquisa, tornando-o uma peça essencial no ecossistema de pesquisa aberta e no compartilhamento de informações científicas (Openaire Graph, 2023).

A **Rede Latino-Americana de Ciência Aberta**⁷⁴ (La Referencia), apoia as estratégias de Acesso Aberto na América Latina e na Espanha por meio de uma plataforma que adota padrões de interoperabilidade, compartilhamento e visibilidade da produção científica gerada por instituições de ensino superior e pesquisa científica. Integra artigos científicos e teses e dissertações de mais de uma centena de universidades e instituições de pesquisa dos dez países membros – Argentina, Brasil, Chile, Colômbia, Costa Rica, Equador, El Salvador, México, Panamá, Peru e Uruguai. Sua criação, em 2012, por meio de um Acordo de Cooperação assinado em Buenos Aires, reflete o compromisso político de oferecer acesso aberto à produção científica como um bem público regional, destacando os resultados financiados com recursos públicos. Essa iniciativa é fruto de acordos técnicos e organizacionais entre agências de ciência e tecnologia dos países membros, em colaboração com a RedCLARA⁷⁵, demonstrando um esforço conjunto para fortalecer a pesquisa e a inovação na região.

O **Digital Object Identifier**⁷⁶ (DOI) é um identificador único e permanente atribuído a um objeto digital (DOI Foundation, 2023), como um artigo de pesquisa, um conjunto de dados, uma imagem ou qualquer outro tipo de recurso digital. O DOI foi criado para proporcionar uma forma estável e confiável de referenciar e acessar conteúdos digitalmente. Composto por um prefixo que identifica a organização ou editora responsável por atribuir o identificador, seguido de um sufixo único que identifica especificamente o objeto digital. Esse identificador é exclusivo e não se repete, o que facilita a localização e o acesso ao objeto em questão, mesmo que sua localização física ou endereço *online* seja alterada (DOI Foundation, 2023).

O DOI é amplamente utilizado na comunidade acadêmica e científica para citar trabalhos de pesquisa de forma precisa e confiável. Ao utilizar o DOI em citações, pesquisadores garantem que outras pessoas possam facilmente encontrar e acessar

⁷⁴<https://www.lareferencia.info/pt/>.

⁷⁵Cooperación Latino Americana de Redes Avanzadas é uma organização internacional cujo objetivo é conectar as redes de pesquisa e educação da América Latina.

⁷⁶<https://www.doi.org/>.

o mesmo objeto digital, promovendo a transparência e a rastreabilidade na divulgação científica (Ferreira *et al.*, 2015).

O **Portal Wikidata**⁷⁷ é um projeto colaborativo mantido pela *Wikimedia Foundation* que visa criar uma base de dados aberta e gratuita, acessível a todos, para armazenar e compartilhar informações estruturadas e interconectadas sobre uma ampla variedade de tópicos. Ele foi lançado em 2012 como uma extensão da Wikipédia, embora seja um projeto independente (Wikimedia Foundation, 2023).

A proposta central do Wikidata é servir como uma fonte centralizada de dados que pode ser utilizada por diferentes projetos da Wikimedia e por qualquer pessoa ou aplicação interessada em informações estruturadas. Os dados no Wikidata são organizados em entidades, que podem ser pessoas, locais, conceitos, obras de arte, eventos e muito mais. Cada entidade tem uma página individual que contém informações e referências sobre o tópico em questão (Wikimedia Foundation, 2023).

O **DOAJ**⁷⁸, *Directory of Open Access Journals*, é um diretório *online* que indexa e fornece acesso a revistas acadêmicas de acesso aberto. Busca dar visibilidade, acessibilidade e uso às revistas científicas de acesso aberto, promovendo assim a disseminação global e gratuita do conhecimento científico (Carvalho Segundo *et al.*, 2022).

O DOAJ avalia as revistas que deseja incluir em seu diretório, garantindo que atendam a padrões de qualidade e transparência editorial. As revistas listadas no DOAJ oferecem acesso aberto imediato ao seu conteúdo, permitindo que pesquisadores e público em geral leiam, baixem, copiem, distribuam e imprimam os artigos sem restrições financeiras ou técnicas (Open Global Trusted, 2023).

Além disso, o DOAJ contribui para a promoção de práticas éticas na publicação científica, exigindo que as revistas cumpram critérios específicos, como revisão por pares, transparência de políticas editoriais e preservação digital. Isso ajuda a fortalecer a confiança nos periódicos acadêmicos e a melhorar a integridade da pesquisa (Open Global Trusted, 2023).

Ademais dos atores citados, elencados por Carvalho Segundo *et al.* (2022), as agências de fomento, como o Conselho Nacional de Desenvolvimento Científico e Tecnológico e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, são

⁷⁷https://www.wikidata.org/wiki/Wikidata:Community_portal.

⁷⁸<https://doaj.org/>.

responsáveis por financiar e incentivar projetos de pesquisa e formação acadêmica, viabilizando a realização de pesquisas.

O **CNPq**⁷⁹ é uma agência governamental fundamental para a promoção da pesquisa científica e tecnológica no Brasil. Desempenha papel central no fomento da ciência, tecnologia e inovação no país. Suas principais atribuições incluem a concessão de bolsas de pesquisa, financiamento de projetos de pesquisa, avaliação da qualidade da produção científica, e o desenvolvimento de políticas públicas relacionadas à pesquisa. Assim, o CNPq auxilia na formação de recursos humanos qualificados, ao conceder bolsas de pesquisa para estudantes de graduação e pós-graduação, bem como para pesquisadores (Conselho Nacional de Desenvolvimento Científico e Tecnológico, 2023).

A **CAPES**⁸⁰ é uma agência vinculada ao Ministério da Educação brasileiro. Sua principal missão é promover a excelência da pós-graduação no país, supervisionando e financiando programas de mestrado e doutorado em instituições de ensino superior (Brasil, 2023a). A CAPES desempenha um papel fundamental na avaliação e regulação da qualidade da educação de pós-graduação, contribuindo para a formação de recursos humanos altamente qualificados, a produção de pesquisa científica de alto nível e o avanço do conhecimento em diversas áreas acadêmicas. A agência também concede bolsas de estudo e apoio financeiro a estudantes e pesquisadores, promovendo a internacionalização da pesquisa brasileira e o intercâmbio acadêmico (Brasil, 2023a).

Incluem-se também entre os atores do ecossistema da pesquisa brasileira, as bases de **patentes**. As bases de patentes são repositórios de informações que contêm um conjunto de documentos relacionados a patentes de invenção e inovações técnicas. Elas servem como recursos valiosos para pesquisadores, inventores, empresas e instituições acadêmicas que desejam acessar informações sobre invenções e avanços tecnológicos. Essas bases fornecem dados detalhados sobre patentes concedidas em diversos campos, incluindo descrições técnicas, desenhos, datas de concessão, nomes dos inventores e titulares de patentes, entre outras informações relevantes.

As bases de patentes disponibilizam a pesquisa, oferecendo informações de patentes necessárias para pesquisar o estado da arte em determinadas áreas,

⁷⁹<https://www.gov.br/cnpq/pt-br>.

⁸⁰<https://www.gov.br/capes/pt-br>.

identificar tecnologias existentes e descobrir lacunas em pesquisa e no desenvolvimento tecnológico. Oferecem ainda a análise de patentes e de suas citações cruzadas para ajudar a avaliar o impacto e a inovação de uma determinada tecnologia ou campo de pesquisa. Permitem a proteção e o registro de patentes, além do licenciamento e a transferência de tecnologia, facilitando a identificação de oportunidades de transferência de tecnologia.

O **Instituto Nacional da Propriedade Industrial**⁸¹ do Ministério do Desenvolvimento, Indústria, Comércio e Serviços disponibiliza tutoriais para acesso em algumas bases de patentes, sendo elas a base do INPI, que fornece acesso a pedidos de patente depositados no Brasil, tanto por residentes no país, quanto por não-residentes; a base Latipat, que inclui pedidos de patentes depositados na Espanha e em países da América Latina, possibilitando as pesquisas em português e espanhol; a base Espacenet com mais de 140 milhões de documentos de patentes de mais de 100 países, tais como EUA, China, Japão, Coreia do Sul, Alemanha, Brasil; a base Patentscope com milhões de documentos de patente de dezenas de países. Permite buscas em texto completo, no idioma original, em documentação patentária e não patentária de vários países (inclusive Brasil), apresentando resultados em gráficos e tabelas.

A rede **ISSN**⁸² é uma organização intergovernamental que gerencia globalmente a identificação e descrição de publicações periódicas e recursos contínuos, sejam impressos ou *online*, abrangendo diversos temas. Fundado oficialmente em 1976, conforme acordo entre a UNESCO e a França. Atualmente, conta com 93 países membros e tem como função coordenar as ações desses países, além de manter e publicar o Portal Internacional ISSN. Também é responsável pela atribuição de Identificadores Internacionais de Publicações Seriadas (ISSNs) a publicações de alcance internacional ou emitidas em nações sem um Centro Nacional ISSN.

O **ISBN**⁸³ (*International Standard Book Number*) é um sistema padronizado de registro numérico, criado para atribuir uma identidade única a publicações monográficas como livros, artigos e apostilas. Composto por uma sequência de 13 números, o ISBN representa o título, autor, país, editora e edição de uma obra. Essa

⁸¹<https://www.gov.br/inpi/pt-br/assuntos/informacao/guia-pratico-para-buscas-de-patentes>.

⁸²<https://www.issn.org/the-centre-and-the-network/our-organization/le-reseau-issn-en/>.

⁸³<https://www.cbiservicos.org.br/isbn/>.

numeração facilita sua identificação e catalogação em escala global. Com o ISBN é possível individualizar e catalogar informações específicas de cada obra, melhorando os processos de produção, distribuição, análise de vendas e armazenamento de dados bibliográficos. Além disso, o ISBN é convertido em código de barras, eliminando barreiras linguísticas e facilitando sua circulação e comercialização. A consulta ao cadastro ISBN para obras cadastradas pode ser realizada no *site* da Câmara Brasileira do Livro, agência brasileira do ISBN, fornecendo informações como o ISBN, título da obra, nome da editora e participação.

O ***Open Research and Contributor ID***⁸⁴ (ORCID) é um identificador digital único para pesquisadores, utilizado principalmente na área acadêmica e científica. Esse identificador é composto por uma sequência de 16 dígitos, que permite distinguir de maneira precisa um pesquisador de outros com nomes similares ou iguais. O ORCID não apenas identifica unicamente um autor, mas também está vinculado a um perfil *online* onde podem ser armazenadas informações profissionais, como publicações, afiliações, prêmios e outros dados relevantes sobre a carreira acadêmica do pesquisador.

A principal finalidade do ORCID em artigos científicos é garantir a correta atribuição de autoria e facilitar a rastreabilidade das contribuições de um pesquisador. Ao incluir o ORCID em um artigo, os autores asseguram que suas obras sejam facilmente associadas a eles, independentemente de mudanças de nome, variações na grafia ou inconsistências nas afiliações ao longo do tempo. O uso do ORCID beneficia a interoperabilidade entre diferentes sistemas de gestão de informações acadêmicas.

O ***International Standard Name Identifier***⁸⁵ (ISNI) é um identificador global único certificado pela ISO⁸⁶ destinado a identificar entidades públicas, como pesquisadores, autores, artistas e outras pessoas ou organizações envolvidas na produção de conteúdo intelectual ou criativo. O ISNI consiste em um código numérico de 16 dígitos, que garante a distinção precisa entre indivíduos ou entidades com nomes semelhantes ou iguais, evitando confusões e garantindo a correta atribuição de obras e contribuições.

⁸⁴<https://orcid.org/>.

⁸⁵<https://isni.org/>.

⁸⁶A ISO (*International Organization for Standardization*) é uma entidade que desenvolve e publica normas técnicas internacionais para garantir a qualidade, segurança e eficiência de produtos, serviços e sistemas.

A principal utilidade do ISNI em artigos científicos é assegurar a identificação precisa e inequívoca dos autores e colaboradores envolvidos. O ISNI melhora a interoperabilidade entre vários sistemas de informação, como bibliotecas, editoras, instituições acadêmicas e bases de dados de pesquisa. Utilizando um identificador padrão como o ISNI, essas entidades conseguem sincronizar e atualizar dados de forma mais eficiente, minimizando o risco de erros e duplicações nos registros.

A **Research Organization Registry**⁸⁷ (ROR) é um sistema que fornece identificadores únicos e persistentes para organizações envolvidas em atividades de pesquisa. Esses identificadores ajudam a distinguir claramente as instituições de pesquisa, como universidades, laboratórios, centros de estudo e outras entidades similares, permitindo uma identificação precisa e consistente em um contexto mundial. O ROR foi criado para resolver o problema de ambiguidade e variação na nomenclatura das instituições, garantindo que cada organização seja reconhecida de forma única, independentemente de alterações no nome ou fusões.

A principal função do ROR em artigos científicos é facilitar a correta atribuição de afiliações institucionais dos autores. O ROR promove a integração eficiente entre diversos sistemas de gerenciamento de informações de pesquisa. Editores de revistas científicas, financiadores de projetos de pesquisa e plataformas de gerenciamento de dados podem usar o ROR para harmonizar informações sobre afiliações institucionais, facilitando a análise e o acompanhamento das colaborações e do impacto das pesquisas. A adoção do ROR contribui para uma melhor organização e acesso aos dados, melhorando a transparência e a qualidade da comunicação científica.

A **Networked Digital Library of Theses and Dissertations**⁸⁸ (NDLTD) é uma organização internacional, sem fins lucrativos, dedicada à promoção e suporte das teses e dissertações eletrônicas, visando aprimorar o compartilhamento global de conhecimento. A NDLTD oferece recursos e tecnologias inovadoras relacionados a teses e dissertações e incentiva a participação das instituições de ensino superior em suas atividades, além de promover o movimento do Acesso Aberto, incluindo o compartilhamento rápido de resultados de pesquisa.

O **Repositórios Científicos de Acesso Aberto de Portugal**⁸⁹ (RCAAP) tem como missão promover, apoiar e facilitar o Acesso Aberto ao conhecimento científico

⁸⁷<https://ror.org/>.

⁸⁸<https://ndltd.org/>.

⁸⁹<https://www.rcaap.pt/>.

em Portugal, ao mesmo tempo em que armazena, disponibiliza e preserva a produção científica do país. A RCAAP promove a Ciência Aberta em Portugal, incentivando a adoção do Acesso Aberto ao conhecimento científico e aumentando a visibilidade dos resultados da atividade acadêmica e científica. Por meio de serviços como o Portal RCAAP e o Serviço de Alojamento de Repositórios Institucionais (SARI), o RCAAP oferece suporte para o depósito e acesso a teses, dissertações e outros trabalhos científicos. O Portal RCAAP, como componente central do serviço, reúne e indexa conteúdos científicos de repositórios e revistas de entidades nacionais e, desde 2010, também inclui conteúdo do Brasil. Além disso, o RCAAP integra-se com outras plataformas, como a OpenAIRE, proporcionando um acesso amplo e facilitado à produção científica e acadêmica para a comunidade acadêmica e científica, bem como para a sociedade em geral.

Nota-se que os atores apresentados desempenham papéis essenciais na promoção do avanço do conhecimento e da inovação no país, no contexto do ecossistema da pesquisa brasileira. Cada um deles contribui de forma polivalente, e a colaboração eficaz entre pesquisadores, instituições de pesquisa, agências de fomento, indústria e governo fortalece o ecossistema.

4.4 APLICABILIDADE DOS PLANOS DE GESTÃO DE DADOS ACIONÁVEIS POR MÁQUINA NO ECOSSISTEMA DA PESQUISA BRASILEIRA

A partir da análise das ferramentas de elaboração de planos de gestão de dados com os dez princípios para PGD acionáveis por máquina da *Research Data Alliance* foram observadas as ferramentas que mais atendem esses princípios, a fim de verificar a aplicabilidade dos planos de gestão de dados acionáveis por máquina no ecossistema da pesquisa brasileira, respondendo, dessa forma, o objetivo “d”.

Verificou-se, portanto, que duas ferramentas se destacam com o maior número de princípios observados. A DMPonline e a DMPTool somam nove princípios observados cada uma. Dessa forma, entende-se que, tanto a DMPonline como a DMPTool teriam condições de trocar informações entre os PGDs e os sistemas de pesquisa, incorporando-os aos fluxos de trabalho da gestão dos dados de pesquisa.

Os dez princípios do Padrão Comum RDA cobrem o ciclo de vida da pesquisa científica e orientam os planos de gestão de dados para o acionamento por máquina.

Esses princípios definem diretrizes técnicas e tratam aspectos organizacionais, refletindo uma abordagem abrangente para a gestão dos dados de pesquisa. Diante do exposto, cada princípio da RDA relaciona-se com os atores do ecossistema da pesquisa brasileira, conforme o Quadro 7.

Quadro 10 – Relação de cada princípio RDA com os atores do ecossistema da pesquisa brasileira (Continua...)

PRINCÍPIO	OBJETIVO	PARA QUEM SE APLICA	ATORES DO ECOSISTEMA DA PESQUISA BRASILEIRA
1	Integrar os PGDs aos fluxos de trabalho de todas as partes interessadas no ecossistema de dados de pesquisa	Todas as partes interessadas	Revistas científicas e agregadores: Oasisbr, Diadorim, Latindex, DOAJ, UlrichsWeb, revistas científicas
			Programas de pós-graduação: Plataforma Sucupira, Capes, Plataforma Lattes
			Repositórios digitais: Plataforma Sucupira, BDTD, Capes
			Instituições de pesquisa e pesquisadores: Plataforma Sucupira, Instituições do INEP, Capes, Plataforma Lattes, OpenAIRE Research Graph
			Bibliotecas digitais: Oasisbr, Capes, BDTD
			Bases de patentes: Instituto Nacional da Propriedade Industrial (INPI), Latipat, Espacenet
			Conjuntos de dados de pesquisa: OpenAIRE Research Graph, Portal Wikidata
			Produção acadêmica e científica: Oasisbr, BDTD, Capes, OpenAIRE Research Graph, La Referencia, NDLTD, RCAAP
			Agências de fomento: CNPq
			Projetos de pesquisa: OpenAIRE Research Graph
2	Permitir que sistemas automatizados atuem em nome das partes interessadas	Qualquer parte interessada que gere informação em sistemas relacionados com PGD	Identificadores persistentes: DOI, ISSN, ISBN, ORCID, ISNI, ROR
			Programas de pós-graduação: Plataforma Sucupira, Capes, Plataforma Lattes
			Repositórios digitais: Plataforma Sucupira, BDTD, Capes
			Instituições de pesquisa e pesquisadores: Plataforma Sucupira, Instituições do INEP, Capes, Plataforma Lattes, OpenAIRE Research Graph
			Bases de patentes: Instituto Nacional da Propriedade Industrial (INPI), Latipat, Espacenet
			Agências de fomento: CNPq
3	Faça políticas (também) para máquinas, não apenas para pessoas	Todas as partes interessadas que fornecem políticas relacionadas a dados	Identificadores persistentes: DOI, ISSN, ISBN, ORCID, ISNI, ROR
			Repositórios digitais: Plataforma Sucupira, BDTD, Capes
			Agências de fomento: CNPq

Fonte: elaborado pela autora.

Quadro 7 – Relação de cada princípio RDA com os atores do ecossistema da pesquisa brasileira (continuação)

PRINCÍPIO	OBJETIVO	PARA QUEM SE APLICA	ATORES DO ECOSISTEMA DA PESQUISA BRASILEIRA
4	Descrever – tanto para máquinas quanto para humanos – os componentes do ecossistema de gestão de dados	Todas as partes interessadas que fornecem sistemas relacionados ao PGD	Programas de pós-graduação: Plataforma Sucupira, Capes, Plataforma Lattes
			Repositórios digitais: Plataforma Sucupira, BDTD, Capes
			Instituições de pesquisa e pesquisadores: Plataforma Sucupira, Instituições do INEP, Capes, Plataforma Lattes, OpenAIRE Research Graph
			Bibliotecas digitais: Oasisbr, Capes, BDTD
			Bases de patentes: Instituto Nacional da Propriedade Industrial (INPI), Latipat, Espacenet
			Conjuntos de dados de pesquisa: OpenAIRE Research Graph, Portal Wikidata
			Agências de fomento: CNPq
			Projetos de pesquisa: OpenAIRE Research Graph
5	Use identificadores persistentes e vocabulários controlados	Todas as partes interessadas	Identificadores persistentes: DOI, ISSN, ISBN, ORCID, ISNI, ROR
			Revistas científicas e agregadores: Oasisbr, Diadorim, Latindex, DOAJ, UlrichsWeb, revistas científicas
			Programas de pós-graduação: Plataforma Sucupira, Capes, Plataforma Lattes
			Repositórios digitais: Plataforma Sucupira, BDTD, Capes
			Instituições de pesquisa e pesquisadores: Plataforma Sucupira, Instituições do INEP, Capes, Plataforma Lattes, OpenAIRE Research Graph
			Bibliotecas digitais: Oasisbr, Capes, BDTD
			Bases de patentes: Instituto Nacional da Propriedade Industrial (INPI), Latipat, Espacenet
			Conjuntos de dados de pesquisa: OpenAIRE Research Graph, Portal Wikidata
			Produção acadêmica e científica: Oasisbr, BDTD, Capes, OpenAIRE Research Graph, La Referencia, NDLTD, RCAAP
			Agências de fomento: CNPq
Projetos de pesquisa: OpenAIRE Research Graph			
Identificadores persistentes: DOI, ISSN, ISBN, ORCID, ISNI, ROR			

Fonte: elaborado pela autora.

Quadro 7 – Relação de cada princípio RDA com os atores do ecossistema da pesquisa brasileira (continuação)

PRINCÍPIO	OBJETIVO	PARA QUEM SE APLICA	ATORES DO ECOSISTEMA DA PESQUISA BRASILEIRA
6	Siga um modelo de dados comum para PGDs acionáveis por máquina	Todas as partes interessadas que fornecem sistemas relacionados ao PGD	Revistas científicas e agregadores: Oasisbr, Diadorim, Latindex, DOAJ, UlrichsWeb, revistas científicas
			Programas de pós-graduação: Plataforma Sucupira, Capes, Plataforma Lattes
			Repositórios digitais: Plataforma Sucupira, BDTD, Capes
			Instituições de pesquisa e pesquisadores: Plataforma Sucupira, Instituições do INEP, Capes, Plataforma Lattes, OpenAIRE Research Graph
			Conjuntos de dados de pesquisa: OpenAIRE Research Graph, Portal Wikidata
			Agências de fomento: CNPq
7	Disponibilizar PGDs para consumo humano e de máquinas	Todas as partes interessadas	Identificadores persistentes: DOI, ISSN, ISBN, ORCID, ISNI, ROR
			Revistas científicas e agregadores: Oasisbr, Diadorim, Latindex, DOAJ, UlrichsWeb, revistas científicas
			Programas de pós-graduação: Plataforma Sucupira, Capes, Plataforma Lattes
			Repositórios digitais: Plataforma Sucupira, BDTD, Capes
			Instituições de pesquisa e pesquisadores: Plataforma Sucupira, Instituições do INEP, Capes, Plataforma Lattes, OpenAIRE Research Graph
			Bibliotecas digitais: Oasisbr, Capes, BDTD
			Bases de patentes: Instituto Nacional da Propriedade Industrial (INPI), Latipat, Espacenet
			Conjuntos de dados de pesquisa: OpenAIRE Research Graph, Portal Wikidata
			Produção acadêmica e científica: Oasisbr, BDTD, Capes, OpenAIRE Research Graph, La Referencia, NDLTD, RCAAP
			Agências de fomento: CNPq
Projetos de pesquisa: OpenAIRE Research Graph			
Identificadores persistentes: DOI, ISSN, ISBN, ORCID, ISNI, ROR			

Fonte: elaborado pela autora.

Quadro 7 – Relação de cada princípio RDA com os atores do ecossistema da pesquisa brasileira (continuação)

PRINCÍPIO	OBJETIVO	PARA QUEM SE APLICA	ATORES DO ECOSISTEMA DA PESQUISA BRASILEIRA
8	Apoiar a avaliação e monitoramento da gestão de dados	Todas as partes interessadas	Revistas científicas e agregadores: Oasisbr, Diadorim, Latindex, DOAJ, UlrichsWeb, revistas científicas
			Programas de pós-graduação: Plataforma Sucupira, Capes, Plataforma Lattes
			Repositórios digitais: Plataforma Sucupira, BDTD, Capes
			Instituições de pesquisa e pesquisadores: Plataforma Sucupira, Instituições do INEP, Capes, Plataforma Lattes, OpenAIRE Research Graph
			Bibliotecas digitais: Oasisbr, Capes, BDTD
			Bases de patentes: Instituto Nacional da Propriedade Industrial (INPI), Latipat, Espacenet
			Conjuntos de dados de pesquisa: OpenAIRE Research Graph, Portal Wikidata
			Produção acadêmica e científica: Oasisbr, BDTD, Capes, OpenAIRE Research Graph, La Referencia, ND LTD, RCAAP
			Agências de fomento: CNPq
			Projetos de pesquisa: OpenAIRE Research Graph
9	Tornar os PGDs documentos atualizáveis, vivos e com controle de versão	Todas as partes interessadas	Identificadores persistentes: DOI, ISSN, ISBN, ORCID, ISNI, ROR
			Revistas científicas e agregadores: Oasisbr, Diadorim, Latindex, DOAJ, UlrichsWeb, revistas científicas
			Programas de pós-graduação: Plataforma Sucupira, Capes, Plataforma Lattes
			Repositórios digitais: Plataforma Sucupira, BDTD, Capes
			Instituições de pesquisa e pesquisadores: Plataforma Sucupira, Instituições do INEP, Capes, Plataforma Lattes, OpenAIRE Research Graph
			Bibliotecas digitais: Oasisbr, Capes, BDTD
			Bases de patentes: Instituto Nacional da Propriedade Industrial (INPI), Latipat, Espacenet
			Conjuntos de dados de pesquisa: OpenAIRE Research Graph, Portal Wikidata
			Produção acadêmica e científica: Oasisbr, BDTD, Capes, OpenAIRE Research Graph, La Referencia, ND LTD, RCAAP
			Agências de fomento: CNPq
Projetos de pesquisa: OpenAIRE Research Graph			
Identificadores persistentes: DOI, ISSN, ISBN, ORCID, ISNI, ROR			

Fonte: elaborado pela autora.

Quadro 7 – Relação de cada princípio RDA com os atores do ecossistema da pesquisa brasileira (continuação)

PRINCÍPIO	OBJETIVO	PARA QUEM SE APLICA	ATORES DO ECOSISTEMA DA PESQUISA BRASILEIRA
10	Disponibilizar publicamente os DMPs	Todas as partes interessadas	Revistas científicas e agregadores: Oasisbr, Diadorim, Latindex, DOAJ, UlrichsWeb, revistas científicas
			Programas de pós-graduação: Plataforma Sucupira, Capes, Plataforma Lattes
			Repositórios digitais: Plataforma Sucupira, BDTD, Capes
			Instituições de pesquisa e pesquisadores: Plataforma Sucupira, Instituições do INEP, Capes, Plataforma Lattes, OpenAIRE Research Graph
			Bibliotecas digitais: Oasisbr, Capes, BDTD
			Bases de patentes: Instituto Nacional da Propriedade Industrial (INPI), Latipat, Espacenet
			Conjuntos de dados de pesquisa: OpenAIRE Research Graph, Portal Wikidata
			Produção acadêmica e científica: Oasisbr, BDTD, Capes, OpenAIRE Research Graph, La Referencia, NDLTD, RCAAP
			Agências de fomento: CNPq
			Projetos de pesquisa: OpenAIRE Research Graph
Identificadores persistentes: DOI, ISSN, ISBN, ORCID, ISNI, ROR			

Fonte: elaborado pela autora.

4.4.1 Primeiro princípio para planos de gestão de dados acionáveis por máquina da *Research Data Alliance*

Ao atender o primeiro princípio, as ferramentas **DMPonline** e **DMPTool** contemplam os atores do ecossistema da pesquisa brasileira. Projetando que os planos de gestão de dados sejam documentos dinâmicos, atualizáveis e acessíveis a todas as partes envolvidas, tanto humanas quanto máquinas, o **primeiro princípio** preconiza a integração dos planos de gestão de dados nos fluxos de trabalho de todos os atores do ecossistema da pesquisa brasileira. O princípio estabelece uma base para gestão dos dados de pesquisa e reconhece que uma boa gestão de dados requer a colaboração e o envolvimento de múltiplos atores ao longo de todo o ciclo de vida dos dados, desde a concepção do projeto até a sua conclusão e preservação.

Os atores Plataforma Sucupira, BDTD e Capes, da categoria “Repositórios digitais”, beneficiam-se do primeiro princípio ao integrar os PGDs aos seus fluxos de trabalho. Essa integração permite, por exemplo, estabelecer critérios para embargos e licenças do conteúdo do repositório com base nas informações contidas nos PGDs, fornecidos pelos pesquisadores. Isso ajuda a garantir o cumprimento das políticas de acesso e reutilização de dados.

O ator CNPq, da categoria “Agências de fomento”, pode verificar se os resultados da pesquisa publicados ou depositados em repositórios seguem políticas e diretrizes relevantes, como os princípios FAIR. Por outro lado, os atores Oasisbr, Diadorim, Latindex, DOAJ, UlrichsWeb e revistas científicas, da categoria “Revistas científicas e agregadores”, valem-se das informações dos planos de gestão de dados para avaliar a qualidade e a integridade dos dados oriundos dos artigos científicos submetidos. Da mesma forma, os atores Plataforma Sucupira, Capes, Plataforma Lattes, da categoria “Programas de pós-graduação”, garantem que os planos de gestão de dados guiem os pesquisadores nas melhores práticas de gestão de dados em seus projetos de pesquisa.

Juntamente com os atores da categoria “Agências de fomento”, ao integrar os PGDs ao fluxo de trabalho, os atores DOI, ISSN, ISBN, ORCID, ISNI e ROR, da categoria “Identificadores persistentes”, podem avaliar o impacto e a eficácia dos financiamentos concedidos. Portanto, ambas categorias, “Agências de fomento” e “Identificadores persistentes”, podem usar os PGDs para avaliar o andamento, a

qualidade e a relevância dos projetos de pesquisa financiados, promovendo uma pesquisa de maior impacto e benefício público.

A Plataforma Sucupira, Instituições do INEP, Capes, Plataforma Lattes e OpenAIRE Research Graph, da categoria “Instituições de pesquisa e pesquisadores”, também podem garantir uma gestão eficaz dos dados desde o planejamento até a disponibilização. As duas ferramentas que seguem o primeiro princípio proporcionam para a categoria “Bibliotecas digitais” e seus atores, Oasisbr, Capes e BDTD, a melhor organização e disseminação dos dados de pesquisa, facilitando, dessa forma, o acesso aos dados e promovendo a descoberta de recursos relevantes para os usuários. Os atores da categoria “Bases de patentes” (Instituto Nacional da Propriedade Industrial, Latipat, Espacenet) guiam-se pelos PGDs para a gestão e documentação adequada dos dados relacionados a patentes.

Com base no primeiro princípio de integrar os PGDs aos fluxos de trabalho de todas as partes interessadas no ecossistema da pesquisa brasileira, os atores OpenAIRE Research Graph e Portal Wikidata (categoria “Conjuntos de Dados de Pesquisa”) garantem a documentação adequada dos metadados e das políticas de uso dos dados.

A integração dos PGDs também permite aos atores Oasisbr, BDTD, Capes, OpenAIRE Research Graph, NDLTD, RCAAP e La Referencia, da categoria “Produção acadêmica e científica”, a correlação entre os dados coletados e os resultados publicados. Dessa forma, as informações contidas nos planos garantem a replicabilidade e reprodutibilidade dos resultados apresentados. A categoria assegura a transparência e a credibilidade da pesquisa, além de facilitar a replicação e a validação dos resultados. Para a categoria “Projetos de pesquisa”, a integração dos PGDs permite que pesquisadores possam usá-los para planejar e documentar adequadamente o ciclo de vida dos dados ao longo do projeto, o que facilita a organização, a preservação e o compartilhamento futuro desses dados. Essa prática não apenas aumenta a transparência e a reutilização dos dados, mas também contribui para a conformidade com políticas de financiamento e diretrizes éticas.

A aplicação do primeiro princípio do Padrão Comum RDA para a integração dos planos de gestão de dados nos fluxos de trabalho de todos os atores promove a gestão de dados colaborativa, beneficiando toda a comunidade científica e garantindo a qualidade, acessibilidade e reuso dos dados.

4.4.2 Segundo princípio para planos de gestão de dados acionáveis por máquina da *Research Data Alliance*

Observando o **segundo princípio**, as **ferramentas** DMPonline e DMPTool permitem que sistemas automatizados atuem em nome de qualquer parte interessada que gere informação em sistemas relacionados com o PGD, promovendo benefícios significativos para todos os atores envolvidos. Essa capacidade de sistemas automatizados atuarem em nome dos atores é essencial para a eficiência e a eficácia da gestão de dados de pesquisa.

Os atores Plataforma Sucupira, Capes e Plataforma Lattes, da categoria “Programas de pós-graduação”, têm a possibilidade de coletar dados administrativos dos pesquisadores a partir de sistemas institucionais para preencher automaticamente os planos de gestão de dados. Essa automação reduz a carga administrativa e garante a precisão das informações. Para os atores da categoria “Repositórios digitais” (Plataforma Sucupira, BDTD e Capes), a automatização pode pré-reservar espaço de armazenamento com base nas informações dos PGDs, garantindo que os repositórios tenham capacidade suficiente para os dados gerados. Além disso, sistemas automatizados podem depositar dados e metadados associados conforme especificado nos planos de gestão de dados, agilizando o processo.

A automação pode fornecer estimativas de custos de armazenamento e preservação com base nos parâmetros fornecidos nos PGDs para os atores Plataforma Sucupira, Instituições do INEP, Capes, Plataforma Lattes e OpenAIRE Research Graph (categoria “Instituições de pesquisa e pesquisadores”). Esse princípio ajuda os atores a planejarem os custos de gestão dos dados de pesquisa de forma mais precisa e a validarem a conformidade com as políticas institucionais e de financiamento, economizando tempo e recursos humanos.

O segundo princípio também traz uma série de benefícios tangíveis para os atores das categorias “Bases e patentes”, “Agências de fomento” e “Identificadores persistentes”. A automação na coleta de dados administrativos pode facilitar a atualização e integração de informações nos sistemas das bases dos atores INPI, Latipat e Espacenet, da “Bases de patentes”, reduzindo erros e o tempo necessário para atualizações manuais. Projeções precisas de demanda e uso de recursos de

armazenamento estimadas de forma automatizada podem ajudar esses atores a planejar seus orçamentos de maneira mais eficaz.

Para o CNPq, da categoria “Agências de fomento”, os serviços automatizados de validação e conformidade podem garantir que os pesquisadores financiados estejam cumprindo os requisitos de políticas de dados estabelecidos pela agência. Da mesma forma, as notificações automatizadas podem alertar a agência quando ações humanas são necessárias, garantindo uma intervenção oportuna e eficaz para resolver quaisquer problemas ou esclarecer questões.

E, para os atores DOI, ISSN, ISBN, ORCID, ISNI e ROR, da categoria “Identificadores persistentes”, a automação no depósito de dados pode simplificar e agilizar o processo de atribuição de DOI, ISSN, ISBN, ORCID, ISNI e ROR facilitando a identificação e referenciamento de recursos de pesquisa. O mesmo serve para a seleção de licença, garantindo que os dados de pesquisa compartilhados estejam em conformidade com as políticas de acesso aberto e licenciamento específicas dessas organizações, promovendo Ciência Aberta.

4.4.3 Terceiro princípio para planos de gestão de dados acionáveis por máquina da *Research Data Alliance*

As ferramentas DMPonline e DMPTool também incorporam o **terceiro princípio**, preconizando a elaboração de políticas para máquinas, e não apenas para pessoas. No contexto do ecossistema da pesquisa brasileira, aplica-se aos atores que fornecem políticas relacionadas a dados: Plataforma Sucupira, BDTD e Capes, da categoria “Repositórios digitais”, e CNPq, da categoria “Agências de fomento”.

Para garantir que os dados depositados nos repositórios da categoria “Repositórios digitais” (atores Plataforma Sucupira, BDTD e Capes) estejam em conformidade com as políticas de dados, as políticas devem ser traduzidas em formatos acionáveis por máquinas. Assim, a adoção de políticas legíveis por humanos e máquinas pode facilitar a incorporação e a gestão dos conjuntos de dados, pois permitem a verificação automatizada da convergência com as políticas de acesso, licenciamento e preservação.

O ator CNPq, da categoria “Agências de fomento”, pode igualmente estabelecer políticas que sejam diretamente interpretáveis por sistemas automatizados para verificar a conformidade dos projetos de pesquisa com os requisitos da gestão dos dados de pesquisa. Dessa forma, a categoria pode garantir que os fundos sejam alocados de acordo com os requisitos de gestão.

4.4.4 Quarto princípio para planos de gestão de dados acionáveis por máquina da *Research Data Alliance*

O **quarto princípio** incorporado às ferramentas DMPonline e DMPTool destaca a importância de descrever os componentes do ecossistema de gestão dos dados de forma compreensível para máquinas e humanos. Essa abordagem beneficia todos atores envolvidos no cenário de gestão de dados de pesquisa. No contexto do ecossistema da pesquisa brasileira, esse princípio facilita a descoberta e seleção de repositórios adequados para o armazenamento e compartilhamento dos dados de pesquisa produzidos pelos pesquisadores, maximizando a atuação dos atores Plataforma Sucupira, Capes e Plataforma Lattes (categoria “Programas de pós-graduação”). Assim, são considerados a integridade e a preservação a longo prazo dos resultados da pesquisa financiada.

Os atores da categoria “Repositórios digitais” (Plataforma Sucupira, BDTD e Capes) podem beneficiar-se da descrição padronizada de seus serviços, facilitando a descoberta por parte dos pesquisadores e a integração com outros sistemas automatizados. Da mesma forma, a categoria “Instituições de pesquisa e pesquisadores” (atores Plataforma Sucupira, Instituições do INEP, Capes, Plataforma Lattes e OpenAIRE Research Graph) pode ser beneficiada com a possibilidade de colaboração com outras partes interessadas da gestão dos dados de pesquisa.

O quarto princípio facilita a descoberta e o acesso a conteúdos relevantes para pesquisadores e usuários da categoria “Bibliotecas digitais” (atores Oasisbr, Capes e BDTD). Da mesma forma, a descrição padronizada dos dados e serviços oferecidos pelos atores da categoria “Bases de patentes” pode melhorar a interoperabilidade e facilitar a integração com outros sistemas relacionados à gestão dos dados de pesquisa, como sistemas de pesquisa de patentes e análise de tendências

tecnológicas. Também o CNPq (categoria “Agências de fomento”) pode valer-se do quarto princípio para a conformidade dos projetos de pesquisa com os requisitos de gestão de dados e promover a prestação de contas no uso de fundos públicos.

O benefício para os atores INPI, Latipat e Espacenet, da categoria “Bases de patentes”, é a descrição padronizada e compreensível dos repositórios disponíveis para compartilhamento e preservação de dados, permitindo que pesquisadores saibam onde depositar os dados resultantes de suas pesquisas, orientando para a acessibilidade e a preservação desses recursos. Os mesmos atores beneficiam-se ainda, ao utilizar uma linguagem comum para descrever os requisitos e políticas de cada repositório, da compreensão por parte dos pesquisadores das condições associadas ao depósito de dados.

Para os atores OpenAIRE Research Graph e Portal Wikidata, da categoria “Conjuntos de dados de pesquisa”, a descrição detalhada dos repositórios disponíveis pode ajudar esses atores a conectar os pesquisadores aos recursos de compartilhamento de dados relevantes para suas áreas de interesse. Ao integrar informações sobre os repositórios em suas plataformas, esses conjuntos de dados de pesquisa podem oferecer aos usuários uma visão abrangente das opções disponíveis para o compartilhamento e preservação de seus próprios dados, facilitando assim o processo de gestão dos dados de pesquisa.

O uso de PIDs e vocabulários controlados para descrever os recursos de pesquisa de maneira precisa e compreensível tanto para humanos quanto para máquinas, tem implicações significativas para o ator OpenAIRE Research Graph, da categoria “Projetos de pesquisa”. Uma descrição clara e abrangente dos repositórios disponíveis pode auxiliar pesquisadores na seleção dos melhores locais para depositar e preservar os dados gerados em suas pesquisas. Isso significa que os pesquisadores podem encontrar facilmente os repositórios mais adequados para seus dados, promovendo a preservação e acessibilidade a longo prazo. Além disso, ao fornecer informações sobre os requisitos e políticas de cada repositório, os PGDs podem ajudar os pesquisadores a compreender as implicações do compartilhamento de dados e tomar decisões informadas sobre a gestão de seus próprios dados.

Para os atores DOI, ISSN, ISBN, ORCID, ISNI e ROR, da categoria “Identificadores persistentes”, a descrição padronizada dos repositórios disponíveis pode ajudar essa categoria a orientar os pesquisadores sobre onde depositar seus

dados. Ao integrar informações sobre os repositórios em seus próprios sistemas, essa categoria proporciona que os pesquisadores terão recursos mais claros e acessíveis para depositar seus dados, enquanto as organizações poderão garantir uma maior conformidade com as políticas de compartilhamento de dados e promover a reutilização dos dados em diferentes contextos.

4.4.5 Quinto princípio para planos de gestão de dados acionáveis por máquina da *Research Data Alliance*

O **quinto princípio** destaca a importância do uso de identificadores persistentes e vocabulários controlados para tornar os planos de gestão de dados mais explícitos e compreensíveis para humanos e máquinas. Esse princípio beneficia todos os atores envolvidos no ecossistema da pesquisa.

O uso de identificadores persistentes e vocabulários controlados na categoria “Revistas científicas e agregadores” ajuda a garantir que os dados associados aos artigos publicados sejam identificados de forma única e clara, facilitando sua localização e reutilização por outros pesquisadores. Além disso, o uso de vocabulários controlados pode melhorar a indexação e a descoberta dos artigos, garantindo a consistência nos termos utilizados para descrever os dados e os métodos de pesquisa.

Para “Programas de pós-graduação” o uso desse princípio simplifica a tarefa dos atores Plataforma Sucupira e Capes na descrição de teses e dissertações, garantindo que os dados de pesquisa sejam identificados de maneira precisa e que os padrões de metadados adequados sejam aplicados. O uso de PIDs ajuda a identificar de forma única os estudantes, suas teses e dissertações, facilitando a referência e o compartilhamento dos resultados de pesquisa. Além disso, o uso de vocabulários controlados pode padronizar a descrição dos dados e dos métodos utilizados nas pesquisas dos estudantes. Isso facilita a citação e o acesso aos dados, além de melhorar sua visibilidade e rastreabilidade. O uso de vocabulários controlados pode padronizar a descrição dos conjuntos de dados, facilitando sua descoberta e compreensão. Igualmente para a categoria “Repositórios digitais”, o uso de identificadores persistentes facilita a identificação única de conjuntos de dados, o que

possibilita a rastreabilidade e a interoperabilidade com outros sistemas e infraestruturas.

Para “Instituições de pesquisa e pesquisadores” o uso de identificadores persistentes e vocabulários controlados pode ajudar a garantir a integridade e a qualidade dos dados gerenciados, pois torna mais fácil a referência e a citação de conjuntos de dados específicos em publicações acadêmicas. O mesmo benefício recebe o ator CNPq, da categoria “Agências de fomento”, pois viabiliza rastrear e monitorar o impacto dos investimentos em pesquisa, assegurando que os resultados da pesquisa financiada sejam adequadamente atribuídos e reconhecidos. Nota-se que o quarto princípio adere as características de cada ator e, por consequência, resulta em benefícios e facilidades aos usuários e clientes desses atores.

Os atores Oasisbr, Capes e BDTD, da categoria “Bibliotecas digitais”, podem empregar identificadores persistentes para identificar de forma única os recursos digitais que disponibilizam, como coleções de teses, dissertações, artigos e conjuntos de dados. Isso facilita a gestão e o acesso aos recursos, garantindo sua identificação única e permanente. O uso de vocabulários controlados pode padronizar a descrição dos recursos, melhorando sua descoberta e compreensão pelos usuários.

Utilizar PIDs e vocabulários controlados proporciona aos atores INPI, Latipat e Espacenet, da categoria “Bases de patentes”, a localização precisa de recursos de propriedade intelectual, como patentes, tornando-os mais acessíveis para pesquisadores, empresas e outras partes interessadas. Além disso, o uso de vocabulários controlados promove a descrição padronizada e clara dos recursos, reduzindo ambiguidades e facilitando a compreensão das informações.

O uso de PIDs permite aos atores OpenAIRE Research Graph e Portal Wikidata, da categoria “Conjuntos de dados de pesquisa”, identificar os conjuntos de dados de pesquisa, possibilitando uma referência precisa e única a esses recursos, facilitando a sua descoberta, citação e reutilização por outros pesquisadores e sistemas. O uso de vocabulários controlados pode padronizar os termos utilizados para descrever conjuntos de dados, assegurando consistência e interoperabilidade entre diferentes repositórios e plataformas de dados.

Para a categoria “Produção acadêmica e científica” a aplicação do quinto princípio oferece vantagens para os atores Oasisbr, BDTD, Capes, OpenAIRE Research Graph, La Referencia, NDLTD e RCAAP. Empregar PIDs facilita o acesso

a produções acadêmicas e científicas, como teses, dissertações, artigos e outros documentos, ampliando assim sua disponibilidade para pesquisadores, estudantes e demais interessados. A utilização de vocabulários controlados contribui para a padronização na descrição desses recursos, reduzindo ambiguidades e garantindo que as informações sejam compreendidas de forma consistente por diferentes partes interessadas.

A referência inequívoca a projetos de pesquisa em planos de gestão de dados facilita o acompanhamento e a atribuição de créditos aos responsáveis pelos projetos. E o emprego de vocabulários controlados ajuda a padronizar a descrição de projetos de pesquisa, tornando as informações mais claras e compreensíveis para todas as partes interessadas envolvidas.

O ator OpenAIRE Research Graph, da categoria “Projetos de pesquisa” proporciona a referência de recursos relacionados à pesquisa, como conjuntos de dados, publicações, protocolos e outros artefatos, ampliando a acessibilidade e a visibilidade desses recursos para a comunidade científica.

O benefício do uso de PIDs para os atores DOI, ISSN, ISBN, ORCID, ISNI e ROR, da categoria “Identificadores persistentes”, é a identificação de recursos como publicações e conjuntos de dados para promover a citabilidade e a rastreabilidade desses recursos na comunidade científica. Também os vocabulários controlados podem ser usados para padronizar termos relacionados à descrição e categorização de recursos.

4.4.6 Sexto princípio para planos de gestão de dados acionáveis por máquina da *Research Data Alliance*

O **sexto princípio**, que preconiza o uso de um modelo de dados comum para planos de gestão de dados acionáveis por máquina, contribui com todas as partes envolvidas no ciclo de vida da pesquisa. O uso desse princípio simplifica a descrição e o compartilhamento de teses e dissertações da categoria “Programas de pós-graduação”. Os atores Plataforma Sucupira e Capes amparam-se na consistência e na interoperabilidade dos dados gerados pelos estudantes de pós-graduação. A adoção de um modelo de dados comum pode ajudar os programas de pós-graduação

a padronizar a coleta e a gestão de informações sobre as pesquisas conduzidas pelos estudantes. Isso pode facilitar a avaliação da qualidade e do impacto das pesquisas, além de promover a transparência e a integridade acadêmica.

Do mesmo modo, os atores da categoria “Repositórios digitais” aproveitam-se da padronização e da integração e interoperabilidade dos metadados associados aos conjuntos de dados depositados para tornar mais fácil a descoberta e o acesso a esses recursos. Esse princípio garante para a categoria a consistência e a interoperabilidade dos dados depositados, promovendo uma prática de gestão de dados mais eficiente.

Outra categoria auxiliada pelo sexto princípio é a “Agências de fomento”, uma vez que o emprego do modelo de dados comum permite uma avaliação mais eficiente e transparente dos planos de gestão de dados submetidos pelos pesquisadores e facilitando a tomada de decisões quanto ao financiamento de projetos de pesquisa. Esse princípio permite ainda uma avaliação mais eficiente dos resultados da pesquisa e o monitoramento do cumprimento de políticas relacionadas à gestão de dados.

Para a categoria “Instituições de pesquisa e pesquisadores”, o sexto princípio sustenta a consistência e a qualidade das informações contidas nos PGDs asseguram a representação clara e precisa das políticas e práticas de gerenciamento de dados adotadas. A utilização de um modelo de dados comum pode permitir que instituições de pesquisa e pesquisadores comuniquem informações sobre suas atividades de pesquisa de forma padronizada e interoperável.

Ao seguir um modelo de dados comum para planos de gestão de dados acionáveis por máquina, a categoria “Revistas científicas e agregadores” (Oasisbr, Diadorim, Latindex, DOAJ, UlrichsWeb e revistas científicas) pode garantir a consistência e a qualidade das informações relacionadas aos dados de pesquisa submetidos para publicação. Essa garantia melhora a transparência e a reprodutibilidade dos estudos publicados, além de facilitar a identificação e o acesso aos dados provenientes das pesquisas.

Por sua vez, os atores OpenAIRE Research Graph e Portal Wikidata, da categoria “Conjuntos de dados de pesquisa”, favorecem-se com a facilitação da descrição e o compartilhamento de conjuntos de dados de pesquisa em diferentes plataformas e domínios, o que pode melhorar a visibilidade e o acesso dos dados.

O sexto princípio relaciona-se aos atores DOI, ISSN, ISBN, ORCID, ISNI e ROR, da categoria “Identificadores persistentes”, ao estabelecer um modelo de dados comum, pois esse princípio pode facilitar a comunicação entre diferentes partes interessadas, proporcionando a interoperabilidade de ferramentas e serviços relacionados ao PGDs. Isso significa que as informações podem ser trocadas de forma mais suave e integrada, permitindo uma colaboração mais eficaz e uma melhor utilização dos recursos de pesquisa. Além disso, a adoção de um modelo de dados comum garante consistência e padronização na representação das informações relacionadas aos planos de gestão de dados.

A utilização de padrões, vocabulários controlados e modelos existentes para organizar as informações de forma sistemática também é facilitada pelo modelo de dados comum, uma vez que os atores DOI, ISSN, ISBN, ORCID, ISNI E ROR possam garantir a qualidade e a precisão dos dados. A abordagem modular do modelo de dados comum permite a inclusão de extensões para descrever aspectos específicos da gestão de dados ou atender a requisitos de domínio específicos, oferecendo flexibilidade e adaptabilidade às necessidades em constante evolução do ambiente de pesquisa.

4.4.7 Sétimo princípio para planos de gestão de dados acionáveis por máquina da *Research Data Alliance*

A disponibilização de planos de gestão de dados para consumo humano e de máquinas, preconizada pelo **sétimo princípio**, envolve e beneficia todos os atores do ecossistema de pesquisa brasileira. Para a categoria “Revistas científicas e agregadores”, que conta com os atores Oasisbr, Diadorim, Latindex, DOAJ, UlrichsWeb e revistas científicas, a disponibilidade de PGDs permite uma melhor compreensão do contexto dos dados associados aos artigos publicados, facilita a revisão por pares e promove a transparência na pesquisa. Portanto, o sétimo princípio melhora a qualidade dos processos de revisão e aumentam a confiabilidade nos resultados publicados.

O sexto princípio também facilita o acompanhamento e a avaliação dos projetos de pesquisa conduzidos por estudantes de pós-graduação, o que ajuda a certificar a

conformidade com as diretrizes institucionais e os requisitos de financiamento dos atores Plataforma Sucupira e Capes, da categoria “Programas de pós-graduação”, promovendo a cultura de boas práticas em relação aos dados de pesquisa. Da mesma forma, a categoria “Instituições de pesquisa e pesquisadores” estimula a responsabilidade na gestão de dados de pesquisa e alavanca a colaboração entre os membros da comunidade acadêmica.

Esse princípio oportuniza ainda, para os atores Plataforma Sucupira, BDTD e Capes, da categoria “Repositórios digitais”, a melhor integração e interoperabilidade dos metadados associados aos conjuntos de dados depositados. Essa integração promove a descoberta e o acesso a esses recursos por outros pesquisadores, viabilizando a colaboração e a inovação na pesquisa. E a categoria “Agências de fomento”, representada pelo CNPq, recorre ao sexto princípio para oferecer uma avaliação mais abrangente e transparente dos projetos de pesquisa financiados. Desse modo, os recursos públicos podem ser utilizados de forma eficiente e responsável. Conseqüentemente, a categoria também pode avaliar o impacto e a eficácia dos investimentos em pesquisa. Isso pode informar decisões futuras de financiamento e promover uma pesquisa de alta qualidade e transparente.

A categoria “Bibliotecas digitais” (atores Oasisbr, Capes e BDTD) pode se beneficiar da disponibilização dos planos de gestão de dados ao facilitar o acesso e a preservação dos dados de pesquisa. Para a categoria “Bases de patentes” (atores Instituto Nacional da Propriedade Industrial – INPI, Latipat e Espacenet), a disponibilização dos PGDs para consumo humano e de máquinas pode ajudar as bases de patentes a avaliar melhor a originalidade e a relevância das invenções patenteadas, facilitando o processo de concessão de patentes.

Outros atores beneficiados são o OpenAIRE Research Graph e o Portal Wikidata, da categoria “Conjuntos de dados de pesquisa”, que podem aumentar a visibilidade e a acessibilidade dos conjuntos de dados de pesquisa. Assim como os atores DOI, ISSN, ISBN, ORCID, ISNI e ROR, da categoria “Identificadores persistentes”, que podem acompanhar e promover boas práticas de gestão de dados de pesquisa em diferentes setores e regiões, contribuindo para o desenvolvimento de políticas e iniciativas que promovam a inovação na pesquisa.

A disponibilização dos PGDs para consumo tanto humano quanto de máquinas beneficia a categoria “Produção acadêmica e científica”, representada por instituições

como Oasisbr, BDTD, Capes, OpenAIRE Research Graph, La Referencia, NDLTD e RCAAP e a categoria “Projetos de pesquisa”, cujo ator é OpenAIRE Research Graph. Ao converter os planos de gestão de dados em documentos vivos e acessíveis, é provável que sejam consultados e atualizados ao longo do ciclo de vida da pesquisa, o que aumenta sua utilidade e relevância. A atualização é especialmente importante porque PGDs tradicionais não são revisados rotineiramente como parte das propostas de financiamento, resultando em sua subutilização.

4.4.8 Oitavo princípio para planos de gestão de dados acionáveis por máquina da *Research Data Alliance*

O **oitavo princípio** do Padrão Comum da RDA concentra-se em apoiar a avaliação e monitoramento da gestão de dados e aplica-se a todos os atores do ecossistema da pesquisa brasileira. Os atores Oasisbr, Diadorim, Latindex, DOAJ, UlrichsWeb e revistas científicas da categoria “Revistas científicas e agregadores”, aproveitam-se do oitavo princípio para avaliação mais precisa da qualidade e integridade dos dados associados às publicações científicas. Os planos de gestão de dados podem ser usados como parte do processo de revisão para avaliar a conformidade com políticas de dados e práticas de gestão de dados de pesquisa. Essa ação é a base da transparência e da replicabilidade da pesquisa e fortalece a credibilidade das publicações.

Atores, como Plataforma Sucupira e Capes, da categoria “Programas de pós-graduação”, podem utilizar os PGDs para o monitoramento da gestão e realizar uma avaliação mais eficaz dos projetos de pesquisa conduzidos por estudantes de pós-graduação e assegurar que os padrões de gestão de dados sejam atendidos. O mesmo princípio facilita para a categoria “Repositórios digitais”, da qual fazem parte a Plataforma Sucupira, BDTD e Capes, a avaliação da adequação dos conjuntos de dados depositados. A categoria pode valer-se dos PGDs para verificar se os dados depositados estão em conformidade com as políticas e diretrizes estabelecidas. Dessa forma, garante-se que os dados sejam gerenciados e compartilhados de acordo com as políticas e diretrizes estabelecidas.

O oitavo princípio propicia ainda, que a categoria “Instituições de pesquisa e pesquisadores” impulsionem os planos de gestão de dados a instrumentos de acompanhamento da gestão dos dados de pesquisa. Os atores Plataforma Sucupira, Instituições do INEP, Capes, Plataforma Lattes e OpenAIRE Research Graph, dessa categoria, podem utilizar os PGDs monitorar e avaliar a conformidade com políticas de dados e práticas de gestão de dados ao longo do ciclo de vida da pesquisa.

Para o ator identificado na categoria “Agências de fomento”, o CNPq, a aplicação desse princípio permite a avaliação global dos projetos de pesquisa financiados. Os PGDs permitem que a categoria monitore a concordância dos projetos contemplados com a concessão de financiamento com as políticas de dados e diretrizes de gestão de dados, usando informações fornecidas neles. Percebe-se que a aplicação do oitavo princípio eleva os PGDs a promotores de qualidade, integridade e transparência da gestão de dados de pesquisa.

O oitavo princípio proporciona ainda, para os atores Oasisbr, Capes e CDTD, da categoria “Bibliotecas digitais”, e Instituto Nacional da Propriedade Industrial - INPI, Latipat e Espacenet, da categoria “Bases de patentes”, a possibilidade de avaliar a qualidade e a integridade dos conjuntos de dados armazenados e disponibilizados para acesso. Os atores OpenAIRE Research Graph e Portal Wikidata, da categoria “Conjuntos de dados de pesquisa”, beneficiam-se do oitavo princípio ao apoiar a avaliação e monitorização dos conjuntos de dados de pesquisa quanto à conformidade com políticas de dados e boas práticas, garantindo a qualidade dos conjuntos de dados para reutilização por outros pesquisadores.

Para os atores DOI, ISSN, ISBN, ORCID, ISNI e ROR, da categoria “Identificadores persistentes”, o oitavo princípio permite que os planos de gestão de dados sirvam de instrumento para verificar a conformidade com políticas e regulamentos relacionados à gestão de dados de pesquisa.

A presença do oitavo princípio nas ferramentas de elaboração de planos de gestão de dados DMPonline e DMPTool oferece benefícios para os atores Oasisbr, BDTD, Capes, OpenAIRE Research Graph, La Referencia, NDLTD e RCAAP, da categoria “Produção acadêmica e científica” e OpenAIRE Research Graph, da categoria “Projetos de pesquisa”. Sendo os PGDs explicitamente avaliados durante as fases de proposta de subvenção e ao longo da pesquisa ativa, assegura-se que esses documentos sejam considerados de forma séria e efetiva. A utilização de

vocabulários controlados, PIDs em um modelo de dados comum contribui para a qualidade da informação contida nos planos de gestão de dados. Envolvendo-se os atores interessados no processo de criação e utilização dos planos, assegura-se que as informações fornecidas sejam estruturadas e precisas, o que facilita a automação de muitas tarefas associadas, como a verificação da conformidade com as políticas do financiador e a monitorização do cumprimento do PGD ao longo do ciclo de vida da pesquisa.

Ferramentas automatizadas podem verificar se a licença selecionada para compartilhamento de dados está em conformidade com as políticas do financiador no início do processo de criação do PGD. Posteriormente, durante o depósito dos dados em um repositório, essas ferramentas podem verificar automaticamente se os dados foram depositados corretamente e se estão acessíveis e licenciados conforme exigido pela política aplicável.

4.4.9 Nono princípio para planos de gestão de dados acionáveis por máquina da *Research Data Alliance*

Embora as ferramentas DMPonline e DMPTool não contemplem princípio de tornar os planos de gestão de dados documentos atualizáveis, vivos e com controle de versão, o **nono princípio**, que implica todas as categorias e atores do ecossistema da pesquisa, o acesso aos PGDs atualizados e em constante evolução permite que todos os atores do ecossistema da pesquisa brasileira tenham a compreensão do contexto e da qualidade dos dados provenientes das pesquisas científicas. O nono princípio torna os planos de gestão de dados dinâmicos e atualizáveis, oferecendo supervisão e acompanhamento dos projetos de pesquisa conduzidos pelos pesquisadores. Esse princípio permite a intervenção proativa para garantir que as boas práticas de gestão de dados sejam atendidas e, assim, promovendo o desenvolvimento de habilidades de gestão de dados entre os pesquisadores em formação.

4.4.10 Décimo princípio para planos de gestão de dados acionáveis por máquina da *Research Data Alliance*

O **décimo princípio** confere aos planos de gestão de dados a primeira indicação concreta de quais dados serão criados no âmbito de um projeto de pesquisa e como esses dados serão geridos. Assim, esse princípio enfatiza a importância de disponibilizar publicamente os PGDs e aplica-se a todas as partes envolvidas na pesquisa científica.

O décimo princípio também permite aos atores da categoria “Revistas científicas e agregadores” – sendo eles o Oasisbr, Diadorim, Latindex, DOAJ, UlrichsWeb e revistas científicas, o acesso público aos PGDs permite a compreensão ampla das circunstâncias e das abordagens da pesquisa que sustentam a coleta de dados dos artigos publicados. Esse princípio pode aprimorar a clareza e a capacidade de reprodução da pesquisa. Para a categoria “Programas de pós-graduação”, a supervisão e o acompanhamento dos projetos de pesquisa têm potencial de fornecer oportunidades para colaboração e troca de boas práticas entre as instituições.

Para os atores Plataforma Sucupira, BDTD e Capes, da categoria “Repositórios digitais”, o acesso aos PGDs publicados permite que se avalie de forma mais precisa a relevância dos conjuntos de dados depositados e garantir que os dados sejam gerenciados e compartilhados de acordo com as políticas e diretrizes estabelecidas. Também a categoria “Instituições de pesquisa e pesquisadores” conta com esse princípio para a facilitação do reuso de dados, a replicação de estudos e a colaboração interinstitucional.

Ao incluir o décimo princípio, as ferramentas de elaboração de PGDs DMPonline e DMPTool, permitem que os atores das categorias “Instituições de pesquisa e pesquisadores” (Plataforma Sucupira, Instituições do INEP, Capes, Plataforma Lattes e OpenAIRE Research Graph) e “Bibliotecas digitais” (Oasisbr, Capes e BDTD) estimulem seus pesquisadores a compartilharem publicamente os PGDs, promovendo a colaboração, o acesso e a reprodutibilidade da pesquisa.

Os atores Instituto Nacional da Propriedade Industrial - INPI, Latipat e Espacenet, da categoria “Bases de patentes”, podem incluir informações sobre os planos de gestão de dados associados a patentes para aumentar a transparência e a compreensão do contexto de pesquisa por trás das invenções patenteadas. Assim

também, os atores OpenAIRE Research Graph e Portal Wikidata, da categoria “Conjuntos de dados de pesquisa”, podem vincular publicamente os PGDs aos dados depositados para fornecer informações sobre o planejamento e a gestão dos dados associados.

A implementação do décimo princípio oferece uma série de vantagens para os atores da categoria “Produção acadêmica e científica”, Oasisbr, BDTD, Capes, OpenAIRE Research Graph, La Referencia, NDLTD e RCAAP. Ao compartilhar os PGDs publicamente desde o início do projeto, os pesquisadores podem fornecer uma visão objetiva das intenções e metodologias de pesquisa, facilitando a colaboração e a compreensão entre pesquisadores. Essa transparência permite que outros pesquisadores compreendam e usem os dados de forma eficaz, contribuindo assim para o conhecimento científico. Além disso, o compartilhamento público dos PGDs simplifica as interações centradas em dados entre várias partes interessadas ao longo do projeto, como instituições, repositórios e financiadores, promovendo a cooperação e a conformidade com as políticas de gerenciamento de dados. Ademais, ao agregar e analisar os planos de gestão de dados compartilhados publicamente, os usuários podem extrair informações valiosas para prestação de serviços de gestão de dados de pesquisa.

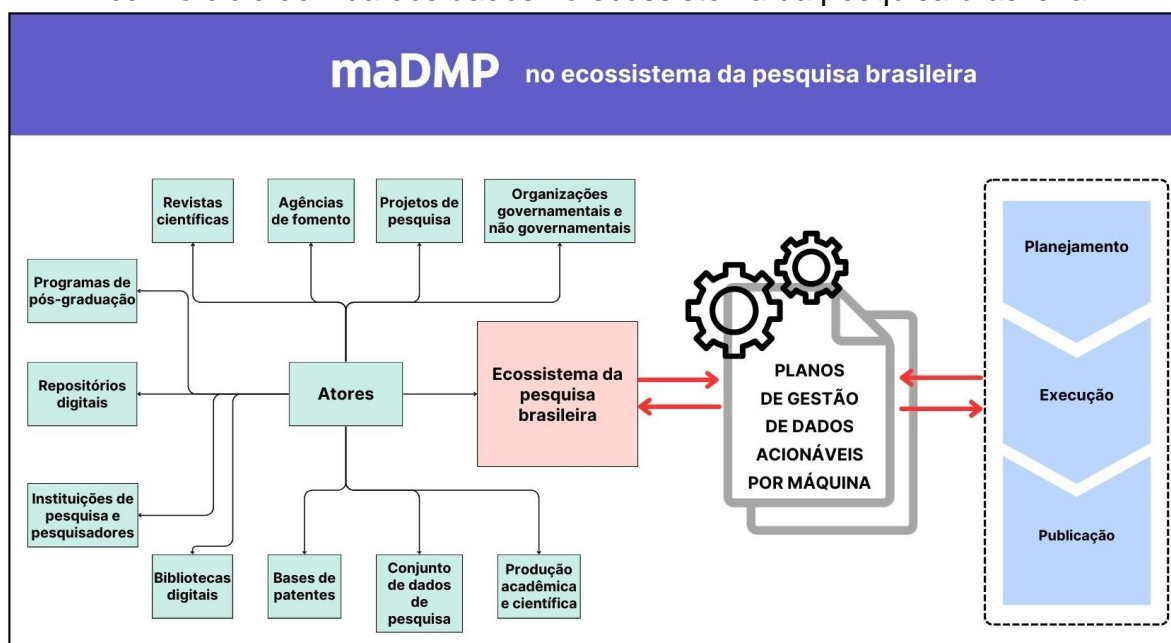
Para o ator OpenAIRE Research Graph, da categoria “Projetos de pesquisa”, compartilhar os planos de gestão de dados publicamente permite o detalhamento de como os dados serão coletados, gerenciados e disponibilizados ao longo do ciclo de pesquisa. As ferramentas DMPonline e DMPTool oferece ao ator CNPq, da categoria “Agências de fomento”, o benefício da transparência e da prestação de contas em relação ao uso de recursos públicos. Semelhantemente, os atores DOI, ISSN, ISBN, ORCID, ISNI e ROR, da categoria “Identificadores persistentes”, valem-se da promoção da divulgação pública dos PGDs para apoiar a transparência e a colaboração na pesquisa.

Disponibilizar publicamente os PGDs permite compartilhá-los dentro e fora da equipe do projeto desde o início, uma vez que os planos ajudam a especificar a metodologia de pesquisa, estimar os recursos necessários e produzir um cronograma viável para a divulgação dos dados. Isso simplifica as interações centradas em dados entre diversas partes interessadas ao longo do projeto, incluindo instituições, repositórios, financiadores e conselhos de revisão ética.

Compartilhar os planos de gestão de dados precoce e frequentemente ao longo do processo de pesquisa, e tão amplamente quanto possível, permite que outros pesquisadores e instituições o utilizem, compartilhem e agreguem as informações resultantes. Assim, compartilhar esses PGDs publicamente e sob uma licença aberta, os tornam acessíveis a qualquer pessoa.

De forma geral, os princípios propõem integrar os planos de gestão de dados em todas as etapas do fluxo de trabalho das partes interessadas e torná-los acessíveis e utilizáveis ao longo do ciclo de vida da pesquisa (Figura 23). Os princípios destacam também a necessidade de apoiar a avaliação e monitoramento da gestão de dados, tornar os PGDs documentos atualizáveis e vivos, com controle de versão, e disponibilizá-los publicamente para consumo humano e de máquinas. Isso implica fornecer uma série de recursos e políticas acionáveis por máquinas.

Figura 23 – Modelo conceitual da associação dos PGDs acionáveis por máquina com o ciclo de vida dos dados no ecossistema da pesquisa brasileira



Fonte: elaborado pela autora.

Observa-se a importância da aplicação dos princípios da *Research Data Alliance* para planos de gestão de dados acionáveis por máquina ao relacionar cada um dos princípios com os atores do ecossistema da pesquisa brasileira. O conjunto de diretrizes explorado promove a gestão de dados consistente, capaz de

acompanhar o desenvolvimento da pesquisa e atender às necessidades dos diversos atores identificados no ecossistema da pesquisa no Brasil.

5 CONSIDERAÇÕES

O percurso desenvolvido nesta pesquisa permitiu analisar a associação do plano de gestão de dados acionáveis por máquina com o ecossistema da pesquisa brasileira. Portanto, os dados coletados a partir da literatura científica possibilitaram identificar e analisar as características das ferramentas de elaboração de planos de gestão de dados e como elas atendem os dez princípios da *Research Data Alliance* para acionamento por máquina dos PGDs. Assim também, o mapeamento dos atores do ecossistema da pesquisa brasileira contribuiu para a sistematização das etapas do fluxo da pesquisa científica, ajudando a encontrar eco em cada um dos princípios da RDA. Dessa forma, foi possível verificar a aplicabilidade do maDMP no ecossistema da pesquisa brasileira.

Os dados coletados sobre as ferramentas de elaboração de PGDs acionáveis por máquina e os atores que compõem o ecossistema da pesquisa brasileira foram depositados no repositório de dados de pesquisa CedapDados⁹⁰ e o plano de gestão de dados disponibilizado publicamente na DMPTool, sob o DOI 10.48321/D1F49A1123⁹¹.

Em resposta ao objetivo “a”, que demandava identificar na literatura os modelos *machine-actionable Data Management Plans*, encontrou-se na pesquisa bibliográfica e documental 14 ferramentas com características variadas e abrangentes que atendem a esse critério. As ferramentas identificadas foram analisadas de acordo com as suas particularidades e possibilidades de integrações conforme os princípios para PGDs acionáveis por máquina da *Research Data Alliance*, satisfazendo o objetivo “b”.

A análise do objetivo “b” revelou que os planos de gestão de dados acionáveis por máquina são eficazes na promoção da interoperabilidade, reutilização e automação dos dados de pesquisa. Ferramentas como DMPonline e DMPTool, por exemplo, permitem a integração automatizada com sistemas institucionais e repositórios digitais, facilitando a gestão eficiente e colaborativa dos dados. Estas ferramentas também suportam a documentação dos metadados e a aplicação de padrões de metadados, essenciais para garantir a qualidade e acessibilidade dos dados.

⁹⁰<https://cedapdados.ufrgs.br/dataset.xhtml?persistentId=hdl:20.500.11959/dt/25>.

⁹¹<https://doi.org/10.48321/D1F49A1123>.

Na análise das 14 ferramentas, duas se sobressaíram às demais, a DMPOnline e a DMPTool. O destaque dessas ferramentas dá-se pelo fato de atenderem a nove dos dez princípios propostos pela RDA, contemplando, assim, um número maior de atores interessados do ecossistema da pesquisa. O estudo da DMPOnline e DMPTool apontou que essas ferramentas são capazes de promover uma gestão de dados eficaz, garantindo que os planos de gestão de dados sejam documentos dinâmicos, atualizáveis e acessíveis tanto para humanos quanto para sistemas automatizados.

No contexto do ecossistema da pesquisa brasileira, planejada no objetivo “c”, foram identificados os principais atores envolvidos e constatou-se uma ampla variedade de beneficiários potenciais da implementação dos planos de gestão de dados acionáveis por máquina. Os atores representativos estão envolvidos em projetos de pesquisa, revistas científicas e agregadores, repositórios digitais, bibliotecas digitais, programas de pós-graduação, instituições de pesquisa, agências de fomento, identificadores persistentes, pesquisadores, bases de patentes, conjuntos de dados de pesquisa e produção acadêmica e científica e foram considerados na análise.

Os atores identificados desempenham diversos papéis nos processos automatizados dos planos de gestão de dados. A integração dos maDMP aos fluxos de trabalho desses atores assegura não apenas a conformidade com políticas de dados e diretrizes éticas, mas também promove uma maior transparência, reprodutibilidade e reuso dos dados de pesquisa. Ao analisar os atores do ecossistema da pesquisa brasileira, foi possível mapear um cenário complexo e interdependente, composto por atores que desempenham papéis importantes na gestão dos dados de pesquisa desde a criação até a preservação e compartilhamento, reforçando a necessidade de uma abordagem integrada e colaborativa para a gestão de dados.

Portanto, ao verificar o objetivo “d” sobre a aplicabilidade do maDMP no ecossistema da pesquisa brasileira, respondeu-se o problema desta pesquisa a propósito da associação dos planos de gestão de dados acionáveis por máquina com o ciclo de vida dos dados no ecossistema da pesquisa brasileira. O objetivo “d” permitiu apurar que os maDMP facilitam o planejamento e a organização dos dados desde o início de um projeto de pesquisa e mesmo após a sua conclusão. Os maDMP colaboram com a padronização dos processos de coleta, armazenamento, análise e compartilhamento de dados.

Além disso, verificou-se que a aplicabilidade dos maDMPs no ecossistema da pesquisa poderá promover a interoperabilidade entre diferentes sistemas de informação e plataformas de pesquisa no Brasil. Essa característica é particularmente importante em um ambiente de pesquisa cada vez mais colaborativo e interdisciplinar, onde a troca eficiente de informações é essencial para o avanço científico.

Outro aspecto relevante da aplicabilidade do maDMP no ecossistema da pesquisa brasileira é a possibilidade de melhorar a transparência e a reprodutibilidade da pesquisa. Ao proporcionar um registro detalhado e acessível das práticas de gestão de dados, o maDMP proporcionará que outros pesquisadores revisem e repliquem estudos com maior precisão, contribuindo para a credibilidade e o impacto das pesquisas realizadas no Brasil.

Assim sendo, verificou-se que os planos de gestão de dados acionáveis por máquina, a exemplo das ferramentas DMPonline e DMPTool, quando implementados, facilitarão a integração dos PGDs aos processos de gestão de dados existentes no ecossistema da pesquisa brasileira, melhorando a coordenação entre as diferentes partes interessadas e promovendo uma cultura de dados mais colaborativa e eficiente. A capacidade de automatizar processos administrativos, como preenchimento de dados, validação de conformidade e estimativas de custos, reduzirá a carga administrativa e aumentará a precisão das informações.

A implementação dos maDMPs pode transformar significativamente a maneira como os dados de pesquisa são geridos, promovendo uma maior eficiência, transparência e colaboração entre os diferentes atores do ecossistema de pesquisa. A adoção desses planos alinhados aos princípios da RDA assegura que os dados de pesquisa sejam bem gerenciados desde a sua concepção até a preservação final, beneficiando toda a comunidade científica e a sociedade em geral.

Assim, a análise da aplicabilidade do maDMP no contexto brasileiro não só destaca vantagens operacionais, mas também o papel estratégico na promoção de uma cultura de dados aberta e eficiente. Portanto, a associação dos planos de gestão de dados acionáveis por máquina com o ciclo de vida dos dados no ecossistema da pesquisa brasileira é altamente benéfica e recomendada.

Cabe ressaltar ainda, a importância dos identificadores persistentes no contexto da gestão dos dados de pesquisa no Brasil. Os identificadores persistentes garantem a unicidade, a estabilidade de identificações e a citabilidade para diversos elementos dentro do ecossistema da pesquisa, como *datasets*, publicações, autores,

instituições e projetos de pesquisa. Essa unicidade é fundamental para assegurar que os dados sejam inequivocamente referenciados e rastreados ao longo do tempo.

No que diz respeito à gestão e ao rastreamento de dados, os PIDs são fundamentais para a eficiência e precisão ao longo do ciclo de vida dos dados. Com o auxílio dos PIDs, os maDMPs permitem monitoramento preciso de onde os dados estão armazenados, quem tem acesso a eles e como estão sendo utilizados. A rastreabilidade não só melhora a organização e a acessibilidade dos dados, mas também facilita a conformidade com as políticas de gestão de dados e as exigências dos financiadores.

Ademais, uma característica importante revelada nesta pesquisa é que existem ferramentas que se prestam a “ferramentas-base” para outras ferramentas de elaboração de planos de gestão de dados. Essa característica é encontrada nas ferramentas DMPRoadmap, que derivou as ferramentas DMPonline, DMPTool e DMP OPIDoR, e OpenDMP, que derivou a ferramenta Argos. Por sua vez, a ferramenta DMPTool derivou a PGD-BR.

Verificou-se que o objetivo dessas ferramentas-base é reduzir os sistemas de gestão de dados personalizados implementados para cada organização, programa, projeto e a concomitante duplicação de esforços. Nesse sentido, essas ferramentas-base buscam criar um sistema que faça o mínimo possível de suposições sobre como os usuários trabalham, armazenam, processam e analisam dados; apoie a integração com ferramentas; o código é intuitivo, pois público-alvo são cientistas de dados e não engenheiros de *software*; o sistema principal deve ser de código aberto; e o serviços de informação principais definem interfaces e servem para gerenciar a execução de algoritmos, processos e ferramentas definidos pelo usuário.

Diante do exposto, entende-se que a associação dos planos de gestão de dados acionáveis por máquina com o ciclo de vida dos dados no ecossistema da pesquisa brasileira apresenta-se como uma necessidade premente e uma solução viável para impulsionar a pesquisa científica no país. Para isso, sugere-se novas pesquisas que aprofundem o entendimento da convergência de tecnologias, políticas e práticas para maDMP no ecossistema da pesquisa brasileira. O avanço de infraestruturas digitais, o estabelecimento de padrões e protocolos, e o apoio de iniciativas governamentais e institucionais expõe um cenário favorável para a integração dos maDMPs com o ciclo de vida dos dados.

Durante esta pesquisa, encontrar e reunir literatura científica que abordasse e aprofundasse a apresentação das características de cada uma das ferramentas de elaboração de planos de gestão de dados, mostrou-se um desafio significativo devido à escassez de estudos específicos sobre o tema. Da mesma forma, a pouca oferta de literatura que abordasse o ecossistema da pesquisa brasileira de forma ampla e, por conseguinte, seus atores, revelou-se um fator provocador na apresentação dos resultados. Diante desse cenário, sugere-se para trabalhos futuros a exploração e o aprofundamento nas características das ferramentas de elaboração de planos de gestão de dados e os atores que compõem o ecossistema da pesquisa brasileira.

REFERÊNCIAS

- ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (org.). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: Unirio, 2015. Disponível em: [https://livroaberto.ibict.br/bitstream/1/1060/1/Ciencia%20aberta_questoes%20abertas_PORTUGUES_DIGITAL%20\(5\).pdf](https://livroaberto.ibict.br/bitstream/1/1060/1/Ciencia%20aberta_questoes%20abertas_PORTUGUES_DIGITAL%20(5).pdf). Acesso em: 21 ago. 2022.
- ALMEIDA, F. G. **Suporte à gestão de dados de pesquisa**: uma ampliação dos serviços oferecidos pelas bibliotecas. Belo Horizonte, 2019. Tese (Doutorado em Gestão e Organização do Conhecimento) – Programa de Pós-Graduação em Gestão e Organização do Conhecimento, Universidade Federal de Minas Gerais, Belo Horizonte, 2019.
- ALVARES, L. *et al.* **Desafios e perspectivas para a gestão de identidade no meio acadêmico nacional**. Instituto Brasileiro de Informação em Ciência e Tecnologia. Disponível em: <http://lillianalvares.fci.unb.br/phocadownload/Apresentacoes/CRIS%2022022018A.pdf>. Acesso em: 10 mar. 2023.
- ANJOS, R. L. D.; DIAS, G. A. Atuação dos profissionais da informação no ciclo de vida dos dados – dataone: um estudo comparado. **Informação & Informação**, v. 24, n. 1, p. 80-101, 2019. DOI: 10.5433/1981-8920.2019v24n1p80. Acesso em: 07 mar. 2023.
- ANNA, J. S.; DIAS, C. C.; MACULAN, B. C. M. D. S. A gestão dos dados de pesquisa nas universidades e o papel dos serviços informacionais oferecidos nas bibliotecas: uma revisão narrativa. **Múltiplos Olhares em Ciência da Informação**, v. 9, n. 2, 2019. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/137095>. Acesso em: 26 set. 2022.
- ARCA dados e FioDMP. [S. l.: s. n.], 2022. 1 vídeo (30 seg). Publicado pelo canal Video Saúde Distribuidora da Fiocruz. Disponível em: <https://www.youtube.com/watch?v=am5jPy21KrA>. Acesso em: 01 jan. 2023.
- AVENTURIER, P.; ALENCAR, M. de C. F. Os desafios dos dados de pesquisa abertos. **Revista Eletrônica de Comunicação & Inovação em Saúde**, v. 10, n. 3, jul./set. 2016.
- BARDIN, L. **Análise de conteúdo**. Lisboa: Edições 70, 1997.
- BIBLIOTECA DIGITAL BRASILEIRA DE TESES E DISSERTAÇÃO. **Sobre a BDTD**. Disponível em: <https://bdtd.ibict.br/vufind/>. Acesso em: 05 nov. 2023.
- BORGMAN, C. **Big data, little data, no data**: scholarship in the networked world. Cambridge: MIT Press, 2015.
- BRASIL. Ministério da Educação. **Capes**. Disponível em: <https://www.gov.br/capes/pt-br/acesso-a-informacao/institucional>. Acesso em: 05 nov. 2023a.

BRASIL. Ministério da Educação. **Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep**: institucional. 2022. Disponível em: <https://www.gov.br/inep/pt-br/acesso-a-informacao/institucional>. Acesso em: 05 nov. 2023.

BRASIL. Ministério da Educação. **Novo desenho garante melhorias à Plataforma Sucupira da Capes**. Disponível em: <http://portal.mec.gov.br/component/tags/tag/35995>. Acesso em: 05 nov. 2023c.

BURNETTE, M. H.; WILLIAMS, S. C.; IMKER, H. J. From Plan to Action: successful data management plan implementation in a multidisciplinary project. **Journal of ESscience Librarianship**, v. 5, n. 1, 2016. Disponível em: <https://experts.illinois.edu/en/publications/from-plan-to-action-successful-data-management-plan-implematio>. Acesso em: 12 mar. 2023.

KIZILHAN, T.; BAL, S. The rise of the network society - the information age: economy, society, and culture. **Contemporary Educational Technology**, 7, 2016. Disponível em: [10.30935/cedtech/6177](https://doi.org/10.30935/cedtech/6177).

CASTRO, P. de; MENNIELLI, M. Recent work by the euroCRIS Best Practice/DRIS Task Group: a way forward for engaging with the CRIS community. **Procedia Computer Science**, v. 33, p. 60-67, 2014.

CERVO, A. L.; BERVIAN, P. A. **Metodologia científica**: para uso dos estudantes universitários. 3. ed. São Paulo: McGraw-Hill do Brasil, 1983.

CHOI, M-S.; LEE, S. Current status and issues of data management plan in Korea. **The Journal of the Korea Contents Association**, v. 20, n. 6, p. 220-229, 2020. Disponível em: <https://koreascience.kr/article/JAKO202018955008023.page>. Acesso em: 19 mar. 2023.

COLUMBIA UNIVERSITY IRVING MEDICAL CENTER. **What are the FAIR Data Principles?** 19 abr. 2022. Disponível em: <https://library.cumc.columbia.edu/insight/what-are-fair-data-principles>. Acesso em: 14 nov. 2022.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO. **Sobre a plataforma Lattes**. 2023. Disponível em: <https://www.lattes.cnpq.br/>. Acesso em: 05 nov. 2023.

COOPERAÇÃO INTENACIONAL. Instituto Brasileiro de Informação em Ciência e Tecnologia. **BR-CRIS é tema de palestra no Ibict**. Jul. 2017. Disponível em: <http://cint.ibict.br/?p=1791>. Acesso em: 23 jan. 2023.

COOPERAÇÃO INTENACIONAL. Instituto Brasileiro de Informação em Ciência e Tecnologia. **Ibict celebra memorando de entendimento com o euroCris**. Disponível em: <http://cint.ibict.br/?p=1487>. Acesso em: 19 mar. 2023.

CRESWELL, J. W. **Projeto de pesquisa**: métodos qualitativo, quantitativo e misto. Porto Alegre: Artmed, 2010.

DATA, DOCUMENTATION AND INITIATIVE. **Machine-actionable**. Disponível em: <https://ddialliance.org/taxonomy/term/198>. Acesso em: 05 nov. 2023.

DATAcite. **Introduction to machine actionable DMPs (maDMPs)**. Disponível em: <https://support.datacite.org/docs/introduction-to-machine-actionable-dmps-madmps>. Acesso em: 15 nov. 2023.

DEMCHENKO, Y.; GRUENGARD, E.; KLOUS, S. Instructional model for building effective Big Data curricula for online and campus education. *In: IEEE STC CC AND RDA WORKSHOP ON CURRICULA AND TEACHING METHODS IN CLOUD COMPUTING, BIG DATA, AND DATA SCIENCE*, 1., 2014, Singapore. **Proceedings [...]**. Singapore, 2014. Disponível em: https://www.researchgate.net/publication/273945502_Instructional_Model_for_Building_Effective_Big_Data_Curricula_for_Online_and_Campus_Education. Acesso em: 25 jan. 2023.

DENCKER, A. de F. M.; VIÁ, S. C. da. **Pesquisa empírica em ciências humanas: com ênfase em comunicação**. 2. ed. São Paulo: Futura, 2001.

DIAS, T. M. R. *et al.* A informação patentária no contexto do BrCRIS. *In: ÁLVARES, L. M. A. de R.; ITABORAHY, A. I. C. (org.). Os múltiplos cenários da informação tecnológica no Brasil do século XXI*. Rio de Janeiro: IBICT, 2021. v. 1, p. 275-290. Disponível em: https://repositorio.unb.br/bitstream/10482/43602/1/LIVRO_MultiplosCenariosInformacao.pdf. Acesso em: 13 mar. 2023.

DIGITAL CURATION CENTRE. **Checklist for a Data Management Plan**. v4.0. Edinburgh: DCC, 2013. Disponível em: https://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf. Acesso em: 15 nov. 2022.

DOI FOUNDATION. [Página inicial]. Disponível em: <https://www.doi.org/>. Acesso em: 05 nov. 2023.

DONNELLY, M.; JONES, S.; PATTENDEN-FAIL, J.W. DMP online: the digital curation centre's web-based tool for creating, maintaining and exporting data management plans. **International Journal of Digital Curation**, v. 5, p. 187-193, 2010. Disponível em: 10.2218/ijdc.v5i1.152. Acesso em: 05 nov. 2023.

DUDZIAK, E. **Dados de pesquisa agora devem ser armazenados e citados**. Jun. 2016. Disponível em: <https://www.abcd.usp.br/noticias/dados-materiais-metodos-revistas-exigem-dados-pesquisa-estejam-disponiveis/>. Acesso em: 01 jan. 2023.

EUROPEAN COMMISSION. **Data management**. Disponível em: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm. Acesso em: 15 nov. 2022.

FERREIRA, E. *et al.* Digital Object Identifier (DOI): o que é, para que serve, como se usa? **AtoZ: novas práticas em informação e conhecimento**, v. 4, n. 1, p. 5-9, 2015. Disponível em: <http://dx.doi.org/10.5380/atoz.v4i1.42369>. Acesso em: 05 nov. 2023.

FLICK, U. **Qualidade na pesquisa qualitativa**. Porto Alegre: Bookman, 2009.

FLICK, U. **Uma introdução à pesquisa qualitativa**. 2. ed. Porto Alegre: Bookman, 2004.

FONSECA, J. J. S. **Metodologia da pesquisa científica**. Fortaleza: UEC, 2002. Apostila.

FORCE11. **About FORCE11**. Disponível em: <https://force11.org/info/about-force11/>. Acesso em: 01 jan. 2013.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO. **Planos de gestão de dados se incorporam a projetos de pesquisa no Brasil**. São Paulo: Pesquisa FAPESP, 2017b. Disponível em: <http://revistapesquisa.fapesp.br/2017/10/25/planos-de-gestao-de-dados-se-incorporam-a-projetos-de-pesquisa-no-brasil/>. Acesso em: 29 dez. 2022.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO. **Programa FAPESP de pesquisa sobre mudanças climáticas globais**: chamada de propostas de pesquisa 2017. São Paulo, 2017a. Disponível em: <http://www.fapesp.br/11068>. Acesso em: 29 dez. 2022.

FUNDAÇÃO OSWALDO CRUZ. **Ciência aberta**. Disponível em: <https://portal.fiocruz.br/glossario-da-ciencia-aberta>. Acesso em: 14 nov. 2022.

GAJBE, S. B. *et al.* Evaluation and analysis of Data Management Plan tools: a parametric approach. **Information Processing & Management**, v. 58, n. 3, 2021. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0306457320309699>. Acesso em: 16 mar. 2023.

GARFIELD, E. **Citation indexing**: its theory and application in science, technology, and humanities. New York: John Wiley & Sons, 1979.

GERHARDT, T. E.; SILVEIRA, D. T. (org.). **Métodos de pesquisa**. Porto Alegre: Ed. UFRGS, 2009.

IGLESIAS, M. E. D.; GÓMEZ, A. M. M. Análisis documental y de información: dos componentes de un mismo proceso. **ACIMED**, Ciudad de La Habana, v. 12, n. 2, p. 1-5, mar./abr. 2004.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Ibict lança Manifesto de Acesso Aberto a Dados da Pesquisa Brasileira para Ciência Cidadã**. 03 out. 2016. Disponível em: <http://sitehistorico.ibict.br/Sala-de-Imprensa/noticias/2016/ibict-lanca-manifesto-de-acesso-aberto-a-dados-da-pesquisa-brasileira-para-ciencia-cidada>. Acesso em: 01 jan. 2023.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Oasisbr: portal brasileiro de publicações e dados científicos em acesso aberto**. Disponível em: <https://oasisbr.ibict.br/vufind/about/home>. Acesso em: 05 nov. 2023.

JONES, S. *et al.* Data management planning: how requirements and solutions are beginning to converge. **Data Intelligence**, v. 2, n. 1/2, p. 208–219, 2020. Disponível em: <https://direct.mit.edu/dint/article/2/1-2/208/10002/Data-Management-Planning-How-Requirements-and>. Acesso em: 18 nov. 2023.

KIM, Y.; STANTON, J. M. Institutional and individual factors affecting scientists' data sharing behaviors: a multilevel analysis. **Journal of the Association for Information Science and Technology**, v. 67, n. 4, p. 776-799, 2016.

KRIPPENDORFF, K. **Metodologia de análisis de contenido**. Barcelona: Paidós, 1990.

LAKATOS, E. M.; MARCONI, M. A. **Fundamentos metodologia científica**. 4. ed. São Paulo: Atlas, 2001.

LEITE, F. T. **Metodologia científica: métodos e técnicas de pesquisa** (monografias, dissertações, teses e livros). Aparecida: Ideias e Letras, 2008.

LORD, P.; MACDONALD, A. **E-Science curation report: data curation for e-Science in the UK: an audit to establish requirements for future curation and provision**. Digital Archiving Consultancy Limited. Twickenham, 2003. Disponível em: <https://docplayer.net/9861532-E-science-curation-report.html>. Acesso em: 12 mar. 2023.

LUDKE, M.; ANDRÉ, M. E. D. A. **Pesquisa em educação: abordagens qualitativas**. São Paulo: E.P.U. 1986.

MARCONI, M. de A.; LAKATOS, E. M. **Fundamentos de metodologia científica**. 7. ed. São Paulo: Atlas, 2010.

MARCONI, M. de A.; LAKATOS, E. M. **Metodologia científica**. 5. ed. São Paulo: Atlas, 2008.

MATOS, J. S. **Análise documental**. Disponível em: http://www.sabercom.furg.br/bitstream/123456789/1739/1/An%C3%A1lise_documental.pdf. Acesso em: 19 mar. 2023.

MIKSA, T. *et al.* Ten principles for machine-actionable data management plans. **PLOS Computational Biology**, v. 15, n. 3, p. e1006750, 2019. DOI: <https://doi.org/10.1371/journal.pcbi.1006750>.

MIKSA, T.; OBLASSER, S.; RAUBER, A. Automating Research Data Management Using Machine-Actionable Data Management Plans. **ACM Transactions on Management Information Systems**, v. 13, n. 2, article 18, dec. 2021. Disponível em: <https://dl.acm.org/doi/10.1145/3490396>. Acesso em: 15 nov. 2022.

MIKSA, T.; WALK, P.; NEISH, P. **RDA DMP common standad for machine-actionable data management plans**. 2020. Disponível em: <https://doi.org/10.15497/rda00039>. Acesso em: 01 jan. 2023.

MOORE, J. F. Predators and prey: a new look at the competitive dynamics of market ecosystems. **Harvard Business Review**, v. 3, n. 71, 1993. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/10126156/>. Acesso em: 04 ago. 2024.

NATIONAL INSTITUTES OF HEALTH. **Data management & sharing policy overview**. Disponível em:

http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm. Acesso em: 01 jan. 2023.

NATIONAL SCIENCE FOUNDATION. **Cyberinfrastructure vision for 21st century discovery**. Mar. 2007. Disponível em:

http://escience.caltech.edu/workshop/CI_Vision_March07.pdf. Acesso em: 23 jan. 2023.

O ECO. **O que é um ecossistema e um bioma**. 25 jul. 2014. Disponível em: <https://oeco.org.br/dicionario-ambiental/28516-o-que-e-um-ecossistema-e-um-bioma/>. Acesso em: 23 jan. 2023.

OLIVEIRA, M. I. S.; LÓSCIO, B. F. What is a data ecosystem? *In*: ANNUAL INTERNATIONAL CONFERENCE ON DIGITAL GOVERNMENT RESEARCH: GOVERNANCE IN THE DATA AGE, 19., 2018, New York. **Proceedings [...]**. New York, 2018.

OLIVEIRA, T. As políticas científicas na era do conhecimento: uma análise de conjuntura sobre o ecossistema científico global. **Perspectivas em Ciência da Informação**, v. 24, n. 1, p. 191-215, 2019. Disponível em:

<http://hdl.handle.net/20.500.11959/brapci/112221>. Acesso em: 14 mar. 2023.

OPEN GLOBAL TRUSTED. **About DOAJ**. Disponível em: <https://doaj.org/about/>. Acesso em: 05 nov. 2023.

OPENAIRE GRAPH. **Open**. Disponível em: <https://graph.openaire.eu/>. Acesso em: 05 nov. 2023.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT.

Declaration on access to research data from public funding. 2022. Disponível em: <https://legalinstruments.oecd.org/public/doc/157/157.en.pdf>. Acesso em: 15 maio 2023.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **OECD principles and guidelines for access to research data from public funding**.

Paris: OECD, 2007. Disponível em: <http://www.oecd.org/sti/inno/38500813.pdf>. Acesso em: 05 dez. 2022.

PAVÃO, C. G. **Dados abertos**. 19 out. 2022. Apresentação de Power Point. 37 slides. Disciplina Ciência Aberta e Reprodutividade, Programa de Pós-Graduação em Ciência da Informação, Universidade Federal do Rio Grande do Sul.

PAVÃO, C. G.; ROCHA, R. P. da; GABRIEL JÚNIOR, R. F. Proposta de criação de uma rede de dados abertos da pesquisa brasileira. **Revista Digital Biblioteconomia e Ciência da Informação**, Campinas, v. 16, n. 2, p. 329-343, maio/ago. 2018.

PIMENTEL, A. O método da análise documental: seu uso numa pesquisa historiográfica. **Cadernos de Pesquisa**, São Paulo, n. 114, p. 179-195, nov. 2001.

PINTO, A. L. *et al.* Brazil developing current research information systems (BrCRIS) as data sources for studies of research. **Iberoamerican Journal of Science Measurement and Communication**, v. 2, n. 1, p. 1-12, 2022. Disponível em: <https://doi.org/10.47909/ijsmc.135>. Acesso em: 23 jan. 2023.

PINTO, A. L. *et al.* Brcris como um sistema de recomendação científico-tecnológica. *In: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO*, 21., 2021, Rio de Janeiro. **Proceedings [...]**. Rio de Janeiro, 2021. GT 8 – Informação e Tecnologia, 2021. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/192997>. Acesso em: 13 mar. 2023.

PONTE, D.; MIERZEJEWSKA, B. I.; KLEIN, S. The transformation of the academic publishing market: multiple perspectives on innovation. **Electronic Markets**, v. 27, n. 2, p. 97-100, 2017.

PRÍNCIPE, P. *et al.* **Relatório técnico sobre ferramentas para a elaboração de Planos de Gestão de Dados**. Repositórios Científicos de Acesso Aberto, 2020. Disponível em: https://repositorium.sdum.uminho.pt/bitstream/1822/67530/1/RelatorioTecnico_Ferramentas_PlanosGestao_DadosV1_final_23092020.pdf. Acesso em: 05 nov. 2023.

PTCRIS. **Nova parceria pelo futuro dos CRIS**: euroCRIS e Brasil. 22 jan. 2016. Disponível em: <https://ptcris.pt/tag/eurocris/>. Acesso em: 15 maio 2023.

QUIVY, R.; CAMPENHOUDT, L. V. **Manuel de recherche en sciences sociales**. Paris: Dunod, 1995.

RESEARCH DATA ALLIANCE. **DMP common standards WG webinar**: integrating research data management services using machine-actionable data management plans. 2022 Webinar Series Highlighting RDA Outputs, 22 nov. 2022. Disponível em: <https://www.rd-alliance.org/sites/default/files/Nov%202022%202022%20-%20DMP%20Common%20Standards%20WG%20Webinar%20Slide%20Deck.pdf>. Acesso em: 05 nov. 2023.

RESEARCH DATA ALLIANCE. **RDA DMP common standard for machine-actionable data management plans**. Disponível em: <https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard>. Acesso em: 05 nov. 2023.

RESOURCE DESCRIPTION AND ACCESS. **maDMP-diagram.png** [diagrama]. Disponível em: <https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/blob/master/docs/diagrams/maDMP-diagram.png>. Acesso em: 01 jan. 2023.

RIBEIRO, N. C.; OLIVEIRA, D. A.; ARAÚJO, R. F. Conjecturas da ciência aberta na contemporaneidade do big data. **BIBLOS**, v. 33, n. 2, p. 163-179, 2019.

ROSA, A. B. A. R.; ARAKAKI, A. C. S.; FURNIVAL, A. C. M. Gestão de dados de pesquisa nas bibliotecas universitárias públicas paulistas. **Revista Brasileira de Biblioteconomia e Documentação**, v. 18, maio 2022.

SABBAG, D. M. A. **Análise documental em textos narrativos de ficção**: subsídios para o processo de análise. Marília, 2013. 193 f. Tese (Doutorado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, 2013.

SALAZAR, A. *et al.* How research data management plans can help in harmonizing open science and approaches in the digital economy. **Chemistry – A European Journal**, v. 29, n. 9, p. e202202720, feb. 2022. Disponível em: <https://chemistry-europe.onlinelibrary.wiley.com/doi/full/10.1002/chem.202202720>. Acesso em: 07 mar. 2023.

SANT'ANA, R. C. G. Campo informacional resultante da interação de ciclos de vida dos dados. *In*: DIAS, G. A.; OLIVEIRA, B. M. J. F. **Dados científicos: perspectivas e desafios**. João Pessoa: Ed. UFPB, 2019. p. 13-31. Disponível em: <https://www.redalyc.org/journal/4656/465669358007/html/>. Acesso em: 07 mar. 2023.

SANT'ANA, R. C. G. Ciclo de vida dos dados: uma perspectiva a partir da Ciência da Informação. **Informação & Informação**, v. 21, n. 2, p. 116–142, 2016. Disponível em: <https://doi.org/10.5433/1981-8920.2016v21n2p116>. Acesso em: 07 mar. 2023.

SANTOS, P. X. dos *et al.* GO FAIR e os princípios FAIR: o que representam para a expansão dos dados de pesquisa no âmbito da Ciência Aberta. **Em Questão**, Porto Alegre, v. 25, n. 2, p. 389-412, maio/ago. 2019.

SAYÃO, L. F.; SALES, L. F. Afinal, o que é dado de pesquisa? **BIBLOS**, v. 34, n. 2, 2020. Disponível em: <https://doi.org/10.14295/biblos.v34i2.11875>. Acesso em: 12 mar. 2023.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa para bibliotecários e pesquisadores**. Rio de Janeiro: CNEN, 2015. Disponível em: <http://www.icb.usp.br/~sbibicb/images/guia%20gestaoPDF/Guia%20de%20gestao%20dados%20de%20pesquisa.pdf>. Acesso em: 09 mar. 2023.

CARVALHO SEGUNDO, W. L. R. C. *et al.* O BrCris como ferramenta de apoio à Ciência Aberta. **Cadernos BAD**, n. 1/2, p. 1-11, 2021. Disponível em: <https://doi.org/10.48798/cadernosbad.2735>. Acesso em: 13 mar. 2023.

CARVALHO SEGUNDO, W. L. R. C. *et al.* Uma estratégia para coleta, integração e tratamento de dados científicos no contexto do BrCris. *In*: DIAS, T. M. R. (org.). **Informação, dados e tecnologia**. Tallinn: ColNes Publishing, 2022. v. 2. DOI: 10.47909/anis.978-9916-9760-3-6.117.

SEMELER, A. R.; PINTO, A. L. Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados. **Ciência da Informação**, v. 48, n. 1, 2019. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/111525>. Acesso em: 09 mar. 2023.

SEVERINO, A. J. **Metodologia do trabalho científico**. São Paulo: Cortez, 2007.

SILVA, F. C. C. da. **Gestão de dados científicos**. Rio de Janeiro: Interciência, 2019.

SIMMS, S. *et al.* Machine-actionable data management plans (maDMPs). **Research Ideas and Outcomes**, v. 3, e13086, abr. 2017. Disponível em: <https://riojournal.com/article/13086/list/1/>. Acesso em: 15 nov. 2022.

SIMONS, E. Introduction to euroCRIS and CRIS Systems. *In: ANNUAL VIVO CONFERENCE*, 10., 2019, Montenegro. **Proceedings [...]**. Montenegro: Radboud University, 2019. Disponível em: https://dspacecris.eurocris.org/bitstream/11366/1167/1/euroCRIS_CRIS_Presentation_VIVO_Montenegro_2019_Ed_S.pdf. Acesso em: 19 mar. 2023.

SMALE, N. *et al.* **The history, advocacy and efficacy of data management plans**. 2018. Disponível em: <https://www.biorxiv.org/content/10.1101/443499v1.full>. Acesso em: 01 jan. 2023.

STANFORD UNIVERSITY. Libraries. **Data management plans**. Disponível em: <https://library.stanford.edu/research/data-management-services/data-management-plans>. Acesso em: 15 nov. 2022.

THE INTERNATIONAL ORGANISATION FOR RESEARCH INFORMATION. **Services**. 2020a. Disponível em: <https://eurocris.org/services>. Acesso em: 19 mar. 2023.

THE INTERNATIONAL ORGANISATION FOR RESEARCH INFORMATION. **What is euroCRIS?** 2020b. Disponível em: <https://eurocris.org/what-eurocris>. Acesso em: 19 mar. 2023.

TORINO, E. *et al.* Infraestrutura e serviços de gestão de dados de pesquisa: uma avaliação no âmbito da Universidade Estadual Paulista (Unesp). **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 27, p. 1-26, 2022. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/85188/51221>. Acesso em: 14 nov. 2022.

TRIVIÑOS, A. N. S. **Introdução à pesquisa em ciências sociais**: a pesquisa qualitativa em educação. São Paulo: Atlas, 1987.

UBALDI, B. **Open government data**: towards empirical analysis of open government data initiatives. Technical report, Organisation for Economic Cooperation and Development, 2013.

UK DATA SERVICE. **Research Data Lifecycle**. 2019. Disponível em: <https://ukdataservice.ac.uk/learning-hub/research-data-management/>. Acesso em: 18 abr. 2023.

UNITED NATIONAL EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION. **Recomendação da UNESCO sobre Ciência Aberta**. 2022. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000379949_por. Acesso em: 23 abr. 2023.

UNIVERSIDADE ESTADUAL PAULISTA. **Plano de gestão de dados**: definições. Disponível em: <https://www2.unesp.br/portal#!/acessoaberto/plano-de-gestao-de-dados/definicao/>. Acesso em: 14 nov. 2022.

UNIVERSIDADE FEDERAL DE SÃO PAULO. Escola de Filosofia, Letras e ciências Humanas. **Gestão de dados de pesquisa**. 11 nov. 2020. Disponível em: [https://www.unifesp.br/campus/gua/dados-de-pesquisa/plano-de-gestao-de-dados#:~:text=O%20Data%20Management%20Plan%20\(DMP,ativa%20do%20projeto%20de%20pesquisa](https://www.unifesp.br/campus/gua/dados-de-pesquisa/plano-de-gestao-de-dados#:~:text=O%20Data%20Management%20Plan%20(DMP,ativa%20do%20projeto%20de%20pesquisa). Acesso em: 15 nov. 2022.

UNIVERSITY COLLEGE LONDON. **Writing a data management plan**. Disponível em: <https://www.ucl.ac.uk/library/open-science-research-support/research-data-management/policies/writing-data-management-plan>. Acesso em: 15 nov. 2022.

UTRECHT UNIVERSITY. **How to make your data FAIR**. Disponível em: <https://www.uu.nl/en/research/research-data-management/guides/how-to-make-your-data-fair>. Acesso em: 14 nov. 2022.

VEIGA, V. **Políticas e planos de gestão de dados**. Rede Go Fair Agro. Ciclo de Oficinas Agosto, Setembro. 2022.

VERGARA, S. C. **Projetos e relatórios de pesquisa em administração**. 3. ed. São Paulo: Atlas, 2000.

VIDAL, L. H. C., CARVALHO SEGUNDO, W. L. R., MENA-CHALCO, J. P., & GABRIEL JUNIOR, R. F. Brcris: visualização de indicadores científicos. Encontro nacional de pesquisa e pós-graduação em ciência da informação, 2023. Disponível em: <https://cip.brapci.inf.br/download/257789>. Acesso em: 06 ago. 2024.

WIKIMEDIA FOUNDATION. **Boas-vindas à Wikidata**. 2023. Acesso em: <https://www.wikidata.org/?uselang=pt>. Acesso em: 05 nov. 2023.

WILKINSON, M. D. *et al.* The FAIR guiding principles for scientific data management and stewardship. **Scientific Data**, v. 3, p. 160018, mar. 2016. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 23 out. 2022.

YIN, R. K. **Case study research: design and methods**. 2nd. ed. Thousand Oaks: Sage, 2003.

ANEXO – Plano de gestão de dados: Planos de gestão de dados acionáveis por máquina no ecossistema da pesquisa brasileira

Plan Overview

A Data Management Plan created using DMPTool

DMP ID: <https://doi.org/10.48321/D1F49A1123>

Title: Planos de gestão de dados acionáveis por máquina no ecossistema da pesquisa brasileira

Creator: Lucia helena Cunha vidal - **ORCID:** [0009-0000-1431-5351](https://orcid.org/0009-0000-1431-5351)

Affiliation: Federal University of Rio Grande do Sul (ufrgs.br)

Principal Investigator: Lucia Helena Cunha Vidal

Data Manager: Lucia Helena Cunha Vidal

Project Administrator: Lucia Helena Cunha Vidal, Rene Faustino Gabriel Junior, Caterina Marta Groposo Pavão

Funder: Federal University of Rio Grande do Sul (ufrgs.br)

Template: Digital Curation Centre

Project abstract:

Os planos de gestão de dados são instrumentos de planejamento de uso de dados de pesquisa para projetos de pesquisa. Discute a importância dos dados de pesquisa e destaca a necessidade de um plano de gestão de dados adequado para garantir disponibilidade, acesso, processamento e reuso dos dados. Expõe a falta de integração dos planos de gestão de dados com outros atores no ecossistema de pesquisa, o que dificulta o acompanhamento do ciclo de vida dos dados. Propõe analisar a viabilidade da utilização de metodologias e ferramentas interoperáveis com as diversas bases de dados para tornar os planos de gestão de dados acionáveis por máquina e integrados com o ciclo de vida dos dados no ecossistema da pesquisa brasileira. É esperado que essa pesquisa possa contribuir para a otimização dos recursos financeiros, humanos e físicos, tornando os dados mais acessíveis e úteis para outros pesquisadores, instituições, agências de fomento e sociedade em geral. De natureza básica, abordagem qualitativa e procedimentos de pesquisa bibliográfica e documental, será descritiva e explicativa.

Analisa ferramentas para elaboração de PGDs que atendem os dez princípios para acionamento por máquina. Identifica os atores do ecossistema da pesquisa brasileira. Verifica que as ferramentas DMPonline e DMPTool atendem o maior número de princípios da RDA e têm condições de trocar informações entre os PGDs e os sistemas de pesquisa. Conclui que o uso desses modelos pode contribuir para a automatização da gestão de dados e para o fluxo de informações em projetos de pesquisa, além de identificar ferramentas que servem de base para a elaboração de outras ferramentas de elaboração de planos de gestão de dados.

Start date: 04-18-2022

End date: 06-30-2024

Last modified: 05-25-2024

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Planos de gestão de dados acionáveis por máquina no ecossistema da pesquisa brasileira

Data Collection

What data will you collect or create?

Os dados coletados são classificados derivados e de origem observacionais, dados textuais.

How will the data be collected or created?

O procedimento de coleta de dados deu-se com a pesquisa em bases de dados bibliográficas nacionais e internacionais das áreas do conhecimento que poderão contribuir para a pesquisa, tais como a Ciência da Informação, a Ciência da Computação e a Administração. A recuperação de artigos, relatórios, anais de eventos e documentos contemplaram as expressões de busca em português e inglês para “planos de gestão de dados” e “planos de gestão de dados acionáveis por máquina”, e foram admitidos ao longo da pesquisa, a inclusão das expressões “ecossistema de pesquisa” e “ferramentas de elaboração de planos de gestão de dados acionáveis por máquina”, que colaboraram para a elucidação do problema desta pesquisa. Por tratar-se de um tema ainda em discussão no Brasil e no exterior, a pesquisa não delimitará as bases de dados para consulta.

Os dados foram coletados a partir da literatura científica recuperada, ou seja, as características, possibilidades de integração e aplicações das ferramentas de elaboração de maDMPs foram registrados em planilhas no formato CSV e, posteriormente, comparadas, analisadas e interpretadas.

Os dados encontrados contemplaram as características das ferramentas de elaboração de planos de gestão de dados disponíveis on-line e os atores implicados no ecossistema da pesquisa brasileira.

Documentation and Metadata

What documentation and metadata will accompany the data?

Nenhuma documentação.

Ethics and Legal Compliance

How will you manage any ethical issues?

Não há dados ou informações particulares. Os dados foram coletados a partir da literatura disponível em acesso aberto e não dados sensíveis.

How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?

Os dados tem licença CC-BY.

Storage and Backup

How will the data be stored and backed up during the research?

Em computador pessoal e nuvem.

How will you manage access and security?

Os dados serão armazenados em computador pessoal e os colaboradores terão acesso via drive.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

Os dados não têm restrições.

What is the long-term preservation plan for the dataset?

Os dados serão depositados em formato aberto em repositório de dados de pesquisa.

Data Sharing

How will you share the data?

Os dados serão depositados em repositório de dados de pesquisa.

Are any restrictions on data sharing required?

Não há dados sensíveis. Não há restrições.

Responsibilities and Resources

Who will be responsible for data management?

O autor principal é o responsável pelo PGD, coleta e gestão.

What resources will you require to deliver your plan?

Os dados são de acesso livre, em formato aberto, sem necessidade de programa excepcional.

Planned Research Outputs

Text - "Análise da viabilidade de associação dos planos de gestão de dados acionáveis por máquina com o ciclo de vida dos dados no ecossistema da pesquisa brasileira"

Espera-se verificar a viabilidade de associação dos planos de gestão de dados acionáveis por máquina com o ciclo de vida dos dados no ecossistema da pesquisa brasileira.

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Análise da viabilidade de associação dos planos de ...	Text	2024-05-17	Open	None specified		None specified	None specified	No	No