

Planos de Gestão de Dados como instrumento para a reprodutibilidade e integridade científica: um estudo a partir da literatura internacional

Lucia Helena Cunha Vidal¹; Rosa Helena Cunha Vidal²; Rene Faustino Gabriel Junior³; Caterina Groposo Pavão⁴

RESUMO

O artigo aborda as ferramentas de elaboração de planos de gestão de dados. A partir do procedimento de pesquisa bibliográfico, é conceituado dados de pesquisa, planos de gestão de dados e ferramentas de elaboração de planos de gestão de dados. Busca aprofundar a discussão do uso das ferramentas de elaboração de planos de gestão de dados como garantidor da qualidade dos dados. Apresenta exemplos de ferramentas e conclui que as ferramentas podem garantir boas práticas para a gestão correta e segura dos dados de pesquisa durante todo o seu ciclo de vida.

Palavras-chave: Ferramentas de elaboração de planos de gestão de dados; Plano de gestão de dados; Dados de pesquisa.

1 INTRODUÇÃO

O contexto internacional da Ciência Aberta demonstra a relevância dos dados de pesquisa para a ciência. Com a intensificação do uso, compartilhamento e reuso dos dados, os planos de gestão de dados (PGDs) ganharam fundamental importância para a correta e segura gestão. Um plano de gestão deve descrever como os dados serão gerenciados ao longo do seu ciclo de vida.

Ferramentas baseadas na *web* e desenvolvidas por instituições que se dedicam a armazenar e compartilhar dados propõem-se a facilitar a tarefa de

¹Mestranda em Ciência da Informação; UFRGS; luciahcvidal@hotmail.com

²Mestranda em Ciência da Informação; UFRGS; rosadeflor@hotmail.com

³Doutor em Ciência da Informação; UFRGS; rene.gabriel@ufrgs.br

⁴Doutora em Comunicação e Informação; UFRGS; caterina@cpd.ufrgs.br

elaborar um plano de gestão de dados. Essas ferramentas podem contribuir para que os planos de gestão de dados atinjam todo o seu potencial, reunindo todas as informações a respeito do ciclo de vida dos dados e qualificando a sua gestão.

Com o auxílio do procedimento de pesquisa bibliográfico, porque pretende buscar maior conhecimento e aprofundamento na discussão do uso das ferramentas de elaboração de planos de gestão de dados como garantidor da qualidade dos dados, esta pesquisa explora as ferramentas existentes para a cobertura de todas as etapas do ciclo de vida dos dados de pesquisa.

1.1 O DESAFIO DE DEFINIR DADOS DE PESQUISA

A complexidade em definir dados de pesquisa está nos variados contextos em que eles estão inseridos, implicando assim nos diversos conceitos que podem ser encontrados. Tanto assim, que Pavão (2022) 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.”.

Um conceito amplamente utilizado quando se trata de dados de pesquisa no ecossistema da Ciência Aberta, tanto no Brasil quanto internacionalmente, é o da *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.” (ORGANIZAÇÃO..., 2007, p. 13).

1.2 O QUE SÃO E PARA QUE SERVEM OS PLANOS DE GESTÃO DE DADOS

Planos de gestão de dados são documentos que descrevem os aspectos de todo o ciclo de vida dos dados de pesquisa (Figura 1), desde os metadados até a preservação dos dados (MIKSA, 2021). O plano assegura que os dados sejam bem gerenciados e preservados para disponibilização e reuso, entretanto, não existe um modelo único.

Os planos de gestão de dados, normalmente demandados por instituições de financiamento, servem não apenas como planejamento, mas também como um instrumento de orientação e avaliação. Usualmente elaborados a partir de listas de quesitos e questionários *online*, apresentam quais dados coletar, como descrever, preservar e assegurar a qualidade, além de definir onde e por quanto tempo armazenar os dados de pesquisa (MIKSA; WALK; NEIH, 2020).

Figura 1 – Plano de gestão de dados



Fonte: Oliveira e Lóscio (2019).

Dentre as inúmeras razões para elaborar um plano de gestão, a Agência de Bibliotecas e Coleções Digitais (DUDZIAK, 2018) da USP apresentou dez motivos, dentre eles: assegurar a integridade da pesquisa e o seu potencial de replicação; assegurar que os dados e demais registros de pesquisa sejam acurados, completos, autênticos e confiáveis; permitir que os seus dados sejam compreensíveis agora e no futuro; economizar tempo e recursos a longo prazo; aumentar a segurança dos dados e minimizar os riscos de perda; evitar a duplicação de esforços na coleta ou regeneração dos dados; aumentar a visibilidade da pesquisa; tornar mais fácil a preservação e o arquivamento.

Invariavelmente, os planos de gestão de dados procuram documentar e proteger os dados de pesquisa, bem como a preservação e o compartilhamento dos dados.

1.3 FERRAMENTAS DE ELABORAÇÃO DE PLANOS DE GESTÃO DE DADOS

As ferramentas de elaboração de planos de gestão de dados são recursos que auxiliam pesquisadores no planejamento, coleta, documentação, preservação e disponibilização eficiente dos dados de pesquisa. Essas ferramentas fornecem orientações, modelos e suporte para garantir que os planos de gestão de dados sejam abrangentes, claros e alinhados com as melhores práticas. As ferramentas facilitam o processo de criação de um PGD bem estruturado e em conformidade com as políticas e regulamentos relevantes (PRÍNCIPE *et al.*, 2020). Ao utilizar essas ferramentas, os pesquisadores podem garantir a transparência, a reprodutibilidade e a qualidade de seus dados de pesquisa, contribuindo para a integridade e a eficiência do trabalho científico.

Assim, as ferramentas de elaboração de planos de gestão cobrem todas as etapas do ciclo de vida dos dados de pesquisa. Toma-se como exemplo o modelo de ciclo de vida dos dados proposto pelo Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT (CICLO..., 2022, vídeo), que sugere quatro quadrantes (Figura 2), cada um deles referentes à fase do ciclo de vida e assim distribuídos:

- a) planejamento da pesquisa, que inclui a perspectiva de reuso e planejamento da gestão dos dados;
- b) desenvolvimento da pesquisa, que compreende a coleta, a análise e o processamento dos dados coletados;
- c) publicação dos resultados;
- d) acesso e preservação, que consiste na conservação e salvaguarda dos dados por intermédio dos repositórios de dados.

Ainda de acordo com o IBICT (CICLO..., 2022, vídeo), o modelo do ciclo de vida dos dados de pesquisa pode ser representado como um movimento contínuo, sendo cada etapa imprescindível para a pesquisa científica e consiste no planejamento, desenvolvimento da pesquisa, publicação dos resultados e salvaguarda para o futuro acesso.

Figura 2 – Modelo do Ciclo de Vida dos Dados de Pesquisa

Fonte: Ciclo de vida dos dados de pesquisa (2022).

Na medida em que as ferramentas documentam o planejamento, a coleta e a disponibilização dos dados de uma pesquisa, as ferramentas, se bem focadas no ciclo de vida dos dados de pesquisa, têm potencial para oferecer a documentação correta, ou seja, o registro sólido e completo de cada ator participante do ciclo.

2 METODOLOGIA

Este artigo, de natureza básica porque: “Objetiva gerar conhecimentos novos, úteis para o avanço da Ciência [...]” (GERHARDT; SILVEIRA, 2009, p. 34), tem como objetivo explorar as ferramentas de elaboração de planos de gestão de dados. Assim, propõe-se o uso do procedimento de pesquisa bibliográfica, “[...] 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).

A coleta deu-se a partir das ferramentas de elaboração de planos de gestão de dados mais citadas na literatura e, conseqüentemente, de maior conhecimento por parte dos pesquisadores são as internacionais, assim partiu-se da busca no *Google Scholar*, Base de Dados em Ciência da Informação (BRAPCI), *Information Science*, *Technology Abstracts* (LISTA), *Scientific Electronic Library* (SciELO), Portal

de Periódicos da Capes, *Scopus* e Biblioteca Digital Brasileira de Teses e Dissertações (BDTD).

A pesquisa foi realizada pela busca dos termos em português “dados de pesquisa”, “plano de gestão de dados” e “ferramenta de elaboração de planos de gestão de dados”, bem como suas traduções em inglês: “*research data*”, “*data management plan*” e “*data management plan tool*”, com recorte temporal de 2019 a 2023.

A construção do referencial teórico deu-se com a recuperação e seleção pertinente da literatura científica a respeito do tema. Portanto, no Quadro 1 estão expostas as bases de dados bibliográficas pesquisadas e a quantidade de documentos recuperados na busca a partir dos termos utilizados.

Quadro 1 – Documentos recuperados na base de dados bibliográficos

Termo de busca Base de dados	2019 a 2023					
	Dados de pesquisa	<i>Research data</i>	Plano de gestão de dados	<i>Data management plan</i>	Ferramenta de elaboração de planos de gestão	<i>Data management plan tools</i>
Google Scholar	12.300	336.00	321	7.660	0	30
BRAPCI	44	84	12	09	0	0
LISTA	12	742	02	774	13	01
SciELO	19	03	0	0	0	0
Capes	1.038	13.900.745	998	17	0	0
Scopus	13	9.903	0	153	0	01
BDTD	881	481	05	04	0	0

A busca nas bases citadas foi realizada de forma exata aos termos descritos e com o filtro por ano. Os aspectos não contemplados pela literatura recuperada foram pesquisados em sites e plataformas de bibliotecas e instituições dedicadas ao tema, como no caso das ferramentas de elaboração de planos de gestão de dados.

3 RESULTADOS DA PESQUISA: CONSIDERAÇÕES SOBRE AS FERRAMENTAS DE ELABORAÇÃO DE PLANOS DE GESTÃO DE DADOS

Se a formulação de um plano de gestão de dados é um exercício que muitos pesquisadores ainda não dominam e o consideram desnecessário e enfadonho, hoje, o pesquisador dispõe de diversas ferramentas que podem facilitar a atividade de elaboração de planos de gestão de dados de pesquisa. Iniciativas internacionais de instituições de ensino e de produção científica criaram e disponibilizam ferramentas na *web* para ajudar pesquisadores a desenvolver seus planos de gestão de forma que garanta a administração e a reutilização dos dados descritos.

As ferramentas propõem-se a auxiliar pesquisadores na atividade de elaborar seus planos de gestão de dados por meio de modelos personalizáveis. Uma das ferramentas mais conhecidas e utilizadas é o DMPTool⁵, ferramenta editável da Universidade da Califórnia, fornece um guia para a elaboração do plano de gestão e o pesquisador pode optar em compartilhar seu plano de gestão de dados publicamente ou não. Com essa ferramenta o pesquisador pode acessar exemplos de planos de gestão, os requisitos das agências de financiamento e salvar versões à medida que os requisitos mudam (DMPTOOL, 2023).

O *Digital Curation Center* do Reino Unido desenvolveu outra ferramenta, amplamente utilizada, o *DMPonline*⁶ para ajudar pesquisadores a criar e compartilhar PGDs. O *DMPonline* oferece orientações personalizadas, exemplos e modelos que visam atender aos requisitos das agências de financiamento não só do Reino Unido como também de outros países (DMPonline, 2023).

Existe ainda, o *Data Stewardship Wizard*⁷, uma ferramenta de código aberto publicamente disponível para criar planos de gestão de dados de forma colaborativa fazendo uso de questionários inteligentes e customizados (DATA..., 2023) e o *DMP Canvas Generator* (DMP..., 2023)⁸, utilizado principalmente por pesquisadores na Suíça, também baseado em um questionário estruturado de acordo com as instruções para PGDs da Fundação Nacional de Ciência da Suíça.

⁵<https://dmptool.org/>.

⁶<https://dmponline.dcc.ac.uk/>.

⁷<https://ds-wizard.org/>.

⁸<https://dmp.vital-it.ch/>.

Outra ferramenta disponível é o *EasyDMP* (EASY..., 2023)⁹, fornecida pela rede pan-europeia colaborativa de infraestrutura de dados EUDAT. O *EasyDMP* é gratuito e acessível para pesquisadores da Noruega e Europa, tem como objetivo fornecer um instrumento que permita até mesmo para pesquisadores com pouco conhecimento em gestão de dados criar um PGD de forma simples. O *EasyDMP* (EASY..., 2023) converte as diretrizes de gestão de dados de agências financiadoras em perguntas fáceis de responder e contendo respostas prontas. Depois de pronto, o plano pode ser usado como modelo para outros pesquisadores.

Já a ferramenta *Argos*¹⁰, de código aberto e desenvolvido pelo projeto *OpenAIRE*, promove apoio nos processos de criação, gestão, de partilha e ligação dos planos aos resultados da pesquisa.

RDMO (RESEARCH..., 2023)¹¹, uma iniciativa alemã que organiza a gestão de dados de pesquisa e permite que instituições e pesquisadores planejem e realizem sua gestão de dados. A ferramenta conta com funções e interface para instituições e repositórios. O RDMO reúne todas as informações de planejamento relevantes de todo o ciclo de vida dos dados de pesquisa.

De outra maneira, o DMP CAT (LIBER..., 2023)¹², ou Catálogo de Plano de Gestão de Dados, desenvolvido pelo Grupo de Trabalho de Gerenciamento de Dados de Pesquisa LIBER, fornece um *hub* para planos de diferentes disciplinas com revisões e avaliações da qualidade das diferentes partes do plano. Este catálogo serve como modelo para pesquisadores e outros envolvidos no processo de escrever um PGD. O DMP CAT (LIBER..., 2023) oferece metadados e *links* para a versão PDF citável com DOI¹³ dos planos no Zenodo¹⁴. Isso porque os planos são publicados via Zenodo e, assim, os autores e colaboradores são devidamente vinculados ao seu ORCID¹⁵.

⁹<https://easydmp.eudat.eu/>.

¹⁰<https://argos.openaire.eu/>.

¹¹<https://rdmorganiser.github.io/en/>.

¹²<https://libereurope.eu/working-group/research-data-management/plans/>.

¹³DOI, *Digital Object Identifier* (Identificador de Objeto Digital), é um padrão de números e letras que identificam publicações. Ele faz essa identificação exclusivamente em ambiente virtual, dando ao objeto singularidade e permanência reconhecida na web.

¹⁴Zenodo é um repositório digital de dados multidisciplinar de Acesso Aberto desenvolvido sob o programa europeu OpenAIRE e operado pelo CERN.

¹⁵ORCID é um identificador digital para acadêmicos/pesquisadores, que permite a distinção de autoria, bem como conecta de forma fácil e correta os autores a sua produção científica.

Em âmbito nacional, a Fundação Oswaldo Cruz concebeu o FioDM¹⁶, que consiste em um *template* que contém perguntas que ajudam o pesquisador na gestão dos seus dados durante todo o ciclo de vida da pesquisa. Disponibilizado na página da Fiocruz, o *template* está dividido em seis seções, são elas:

- a) seção A: informação administrativa – referente ao pesquisador e sua pesquisa;
- b) seção B: descrição dos dados coletados ou reuso de dados existentes;
- c) seção C: documentação e qualidade dos dados;
- d) seção D: armazenamento e *backup* durante o processo de pesquisa;
- e) seção E: requisitos legais, éticos e de códigos de conduta;
- f) seção F: compartilhamento de dados.

Esses são alguns dos exemplos de ferramentas disponíveis e desenvolvidas por instituições de pesquisa e agências financiadoras a partir de seus requisitos e experiências que visam auxiliar na criação de um PGD.

4 CONSIDERAÇÕES FINAIS

Percebe-se que o avanço da Ciência Aberta provocou a intensificação do uso dos dados levando às mudanças significativas no modo de criar, tratar, analisar, preservar, acessar e reutilizar os dados disponíveis. Muito utilizados em diversas áreas do conhecimento para validar os resultados das pesquisas científicas, os dados de pesquisa ganham destaque com o avanço das novas formas de colaboração e compartilhamento.

Com esse avanço, diversas iniciativas corroboram com a promoção do compartilhamento no sentido de garantir boas práticas para a gestão correta e segura dos dados de pesquisa durante todo o seu ciclo de vida. Os planos de gestão de dados ocupam-se justamente em descrever como serão coletados e tratados os dados durante toda a pesquisa e, inclusive, após a sua conclusão.

No propósito de facilitar e apoiar a criação de um plano de gestão de dados, diversas instituições desenvolveram ferramentas que auxiliam o pesquisador a criar

¹⁶<https://fiodmp.fiocruz.br/>

um documento que detalhe o ciclo de vida dos dados. Algumas ferramentas e modelos citadas ao longo desta pesquisa buscam exatamente contemplar todas as características dos dados para bem registrá-los e preservá-los.

Hoje, a publicação de dados de pesquisa pode ocorrer em três fases da pesquisa: a) quando o pesquisador já finalizou a pesquisa e quer somente depositar seus dados; b) o pesquisador já está em fase de coleta, e precisa modificar a modelagem dos dados para publicar; c) o projeto está em fase de elaboração (planejamento) – criar a cultura do plano de gestão de dados.

Assim, percebe-se que as ferramentas de elaboração de planos de gestão de dados podem servir de facilitador, tanto para pesquisadores, bibliotecários, arquivistas, curadores, quanto para serviços especializados, nos diferentes estágios do gerenciamento dos dados de pesquisa.

REFERÊNCIAS

ARGOS. **Plan and follow your data**. Disponível em: <https://argos.openaire.eu/>. Acesso em: 06 jul. 2023.

CICLO de vida dos dados de pesquisa: #2. [S. l.: s. n.], 2022. 1 vídeo (6 min). Publicado pelo canal Instituto Brasileiro de Informação em Ciência e Tecnologia – IBICT. Disponível em: <https://www.youtube.com/watch?v=WtHc0HxmmPg>. Acesso em: 07 jul. 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: 07 jul. 2023.

DATA STEWARDSHIP WIZARD. [Home]. Disponível em: <https://ds-wizard.org/>. Acesso em: 09 jul. 2023.

DATA STEWARDSHIP WIZARD. **About**. Disponível em: <https://ds-wizard.org/about>. Acesso em: 08 jul. 2023.

DMP CANVAS GENERATOR. [Home]. Disponível em: <https://dmp.vital-it.ch/>. Acesso em: 09 jul. 2023.

DMPONLINE. **Planto make data work for you**. Disponível em: <https://dmponline.dcc.ac.uk/>. Acesso em: 07 jul. 2023.

DMPTOOL. [Home]. Disponível em: <https://dmptool.org/>. Acesso em: 07 jul. 2023.

DUDZIAK, E. **Gestão de dados de pesquisa: o que precisamos saber hoje!** 2018. Disponível em: <http://www.abcd.usp.br/noticias/gestao-de-dados-de-pesquisa-o-que-precisamos-saber-hoje/>. Acesso em: 07 jul. 2023.

EASY. DMP. **Data management plan generator**. Disponível em: <https://easydmp.eudat.eu/>. Acesso em: 08 jul. 2023.

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

FUNDAÇÃO OSWALDO CRUZ. **Ciência aberta**. Disponível em: <https://portal.fiocruz.br/glossario-da-ciencia-aberta>. Acesso em: 08 jul. 2023.

FUNDAÇÃO OSWALDO CRUZ. **Plano de gestão de dados**. Disponível em: <https://fiopgd.icict.fiocruz.br/area-do-pesquisador>. Acesso em: 06 jul. 2023.

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

LIBER. **LIBER research data management**. Disponível em: <https://libereurope.eu/working-group/research-data-management/plans/>. Acesso em: 09 jul. 2023.

MIKSA, T. *et al.* **Application Profile for Machine-Actionable Data Management Plans**. Data Science Journal, n. 20, v 1, p.32, 2021. DOI: <http://doi.org/10.5334/dsj-2021-032>.

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.

MONTEIRO, E. C. S. A. **Por que usar a DMPTool para a gestão de dados de pesquisa?** Disponível em: <https://www.sibi.ufscar.br/arquivos/apresentacao-por-que-usar-dmptool-para-gestao-de-dados.pdf>. Acesso em: 07 jul. 2023.

OLIVEIRA, M. I. S.; LÓSCIO, B. F. Ecosistemas de dados na web: da teoria aos desafios. In: BRAZILIAN SYMPOSIUM ON DATABASES, 34., Ceará, 2019. **Proceedings [...]** Ceará, 2019. Disponível em: <https://sol.sbc.org.br/livros/index.php/sbc/catalog/download/62/273/515-1?inline=1>. Acesso em: 06 jul. 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., NewYork, 2018. **Proceedings [...]** NewYork, 2018.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. OECD Principles and Guidelines for Access to Research Data from Public Funding. Paris:

OECD, 2007. 24 p. Disponível em: <http://www.oecd.org/sti/inno/38500813.pdf>. Acesso em: 09 jul. 2023.

PAVÃO, C. G. **Dados abertos**. 19 out. 2022. Material de aula disponibilizado pela professora, 37 slides.

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.

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: 08 jul. 2023.

PRÍNCIPE, P. *et al.* **Relatório técnico sobre ferramentas para a elaboração de Planos de Gestão de Dados**. 2020. Disponível em: <https://hdl.handle.net/1822/67530>. Acesso em: 21 jul. 2023.

RESEARCH DATA MANAGEMENT ORGANISER. **RDMO**. Disponível em: <https://rdmorganiser.github.io/en/>. Acesso em: 09 jul. 2023.

SILVA, F. C. C. da. **Gestão de dados científicos**. Rio de Janeiro: Interciência, 2019.