

Panorama dos planos de gestão de dados na pesquisa científica: uma análise cientométrica na BRAPCI

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

RESUMO

Os planos de gestão de dados (PGDs) são documentos que descrevem como os dados de pesquisa serão gerenciados ao longo do ciclo de vida dos dados. Esta pesquisa de natureza básica, abordagem qualitativa e procedimento bibliográfico, versa sobre os PGDs. Busca responder como são abordados os PGDs na pesquisa científica dentro da literatura brasileira em Ciência da Informação. A coleta se deu na Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação – BRAPCI, por meio da busca do termo em português “plano de gestão de dados”. A pesquisa é justificada pela importância dos PGDs como instrumento de confiabilidade da qualidade dos dados de pesquisa disponibilizados. Conclui que a abordagem do tema se dá pela ótica da gestão de dados e ainda é pouco explorada mesmo com os esforços dos pesquisadores em aprimorar os planos.

Palavras-chave: Plano de gestão de dados; Ciência aberta; BRAPCI; Cientometria.

1 INTRODUÇÃO

O avanço das tecnologias da informação e da Ciência Aberta provocou a intensificação do uso de dados levando às mudanças significativas no modo de criar, tratar, analisar, preservar, acessar e reutilizar os dados disponíveis. O tratamento correto dos dados, por meio de boas práticas, como forma de garantir a integridade dos dados da pesquisa, sua replicação e reuso dos dados impeliram os planos 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

gestão de dados (PGDs) – que são documentos que descrevem todo o ciclo de vida dos dados de maneira formal (UNIVERSIDADE..., 2020) – como nova prática à comunidade científica.

A importância dos dados de pesquisa tem sido reconhecida não apenas para validar os resultados publicados, mas também para permitir o reuso dos dados, 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.

Via de regra, a gestão abrange as diversas etapas do ciclo de vida dos dados, que corresponde o planejamento da pesquisa, coleta, processamento e análise dos dados, publicação e compartilhamento de dados, preservação e reuso dos dados (UK DATA..., 2019), assim futuras pesquisas poderiam aproveitar os dados previamente coletados e devidamente preservados.

A documentação dos dados garante detalhes sobre a coleta, o armazenamento, publicação e o compartilhamento. Assim, os PGDs otimizam recursos, garantem a reprodutibilidade e viabilizam novas pesquisas com segurança e confiabilidade (TORINO *et al.*, 2022).

Assim, esta pesquisa justifica-se pela importância dos PGDs enquanto instrumento de confiabilidade da qualidade dos dados de pesquisa disponibilizados. Ademais, a demanda por PGDs incentivada pela abertura dos dados de pesquisa no contexto da Ciência Aberta tem impulsionado a adoção de práticas de governança e infraestrutura de dados mais robustas, promovendo a integridade, segurança e acessibilidade dos dados científicos.

Diante desse contexto, esta pesquisa busca responder como estão sendo abordados os PGDs na pesquisa científica (GERHARDT; SILVEIRA, 2009). Foi utilizada a Base de Dados em Ciência da Informação (BRAPCI), por ser da Ciência da Informação, para recuperação de artigos e anais de eventos a partir da expressão de busca “plano de gestão de dados”. Essa pesquisa descritiva baseia-se nos procedimentos de pesquisa bibliográfica por ser realizada a partir de meios escritos (FONSECA, 2002). Por tratar-se de uma pesquisa cientométrica, de

natureza básica, tem abordagem quantitativa, os dados foram analisados globalmente para atingir o objetivo de identificar o panorama geral dos PGDs na pesquisa científica dentro da base de dados BRAPCI.

2 PLANO DE GESTÃO DE DADOS

A dificuldade em entender os dados de outros pesquisadores faz com que eles não sejam reutilizados (CHOI; LEE, 2020). Nos últimos anos há um investimento significativo no desenvolvimento e uso de PGDs para pesquisa, impulsionado pelo incentivo por parte das agências de financiamento à abertura dos dados antes da concessão de recursos (MIKSA; WALK; NEIH, 2019).

Stanford University (2023), University College London (2023) e European Commission (2023) definem PGD como um documento que descreve os dados que se espera adquirir ou gerar durante o curso de um projeto de pesquisa, como serão gerenciados, descritos, analisados e armazenados e quais mecanismos serão usados no final do projeto para compartilhar e preservar os dados. Aventurier e Alencar (2016) indicam que um PGD 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.

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.*, 2023) 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.

Nos anos 2000, a internet provocou mudanças significativas na pesquisa científica gerando uma grande produção de dados, levando à ampla discussão as mudanças na natureza da pesquisa e as respostas necessárias das instituições e do governo diante dessas mudanças (SMALE *et al.*, 2018).

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

PGD 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..., 2017a). Desde então, 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..., 2017b), sendo o Estado de São Paulo pioneiro a implantar o PGD de pesquisa nacionalmente.

Como forma de garantir as boas práticas de pesquisa e gestão dos dados o Digital Curation Centre (2013) elaborou perguntas que norteiam a construção de um PGD. Essas orientações consideram quais 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.

3 PROCEDIMENTOS METODOLÓGICOS

Para responder o problema desta pesquisa e atingir o objetivo de identificar o panorama dos PGDs na pesquisa científica, a base de dados escolhida foi a Base de Dados em Ciência da Informação (BRAPCI), considerando sua relevância na área da Ciência da Informação. Como estratégia de busca foi utilizado o termo em português “plano* de gestão de dados”, com o operador booleano “*” a fim de garantir que todos os registros fossem recuperados, tanto no singular quanto no plural. Os resultados obtidos após a busca contemplavam a expressão e foram considerados para a pesquisa, sem a exclusão de datas.

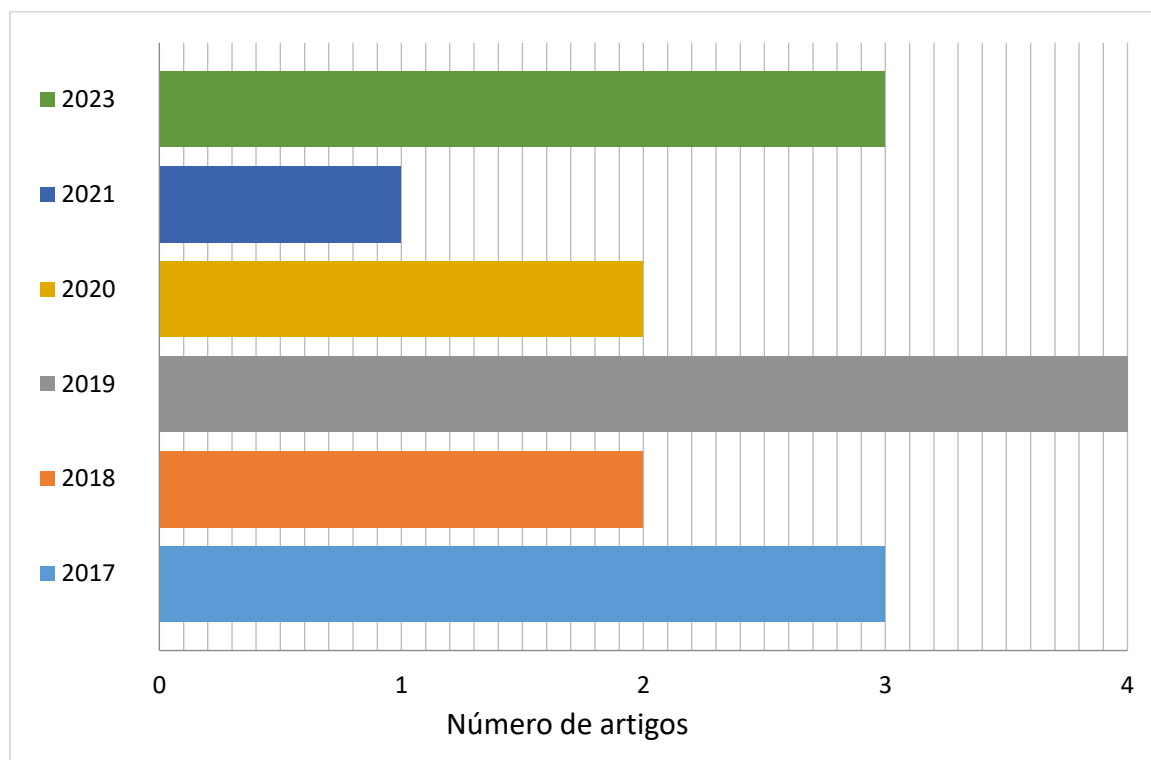
Para cada documento recuperado foi feita análise detalhada das revistas, títulos, resumos, palavras-chave, ano, autor(es) e instituição do(s) autor(es). Os dados relevantes de cada artigo obtidos com a análise foram registrados em planilha eletrônica Excel para facilitar o processo de análise, buscando identificar padrões, tendências e características expressivas relacionadas ao contexto dos PGDs.

4 RESULTADOS DA PESQUISA

Os dados foram coletados a partir da busca pelo termos em português “plano* de gestão de dados” na BRAPCI. Não foi aplicado nenhum filtro para o refinamento da pesquisa. Assim, foram recuperados 16 documentos, sendo que dois (2) deles são iguais em publicações diferentes. Dessa forma, o recorte da pesquisa foi de 15 publicações.

Dos 15 artigos publicados entre os anos de 2017 e 2023 (Figura 1), verifica-se que em 2019 foram quatro (4) artigos publicados sobre a temática. 2017 e 2023 também foram anos significativos, uma vez que cada um possui três (3) artigos publicados, em contrapartida, nenhum artigo na BRAPCI do ano de 2022 foi encontrado.

Figura 1 – Anos das publicações

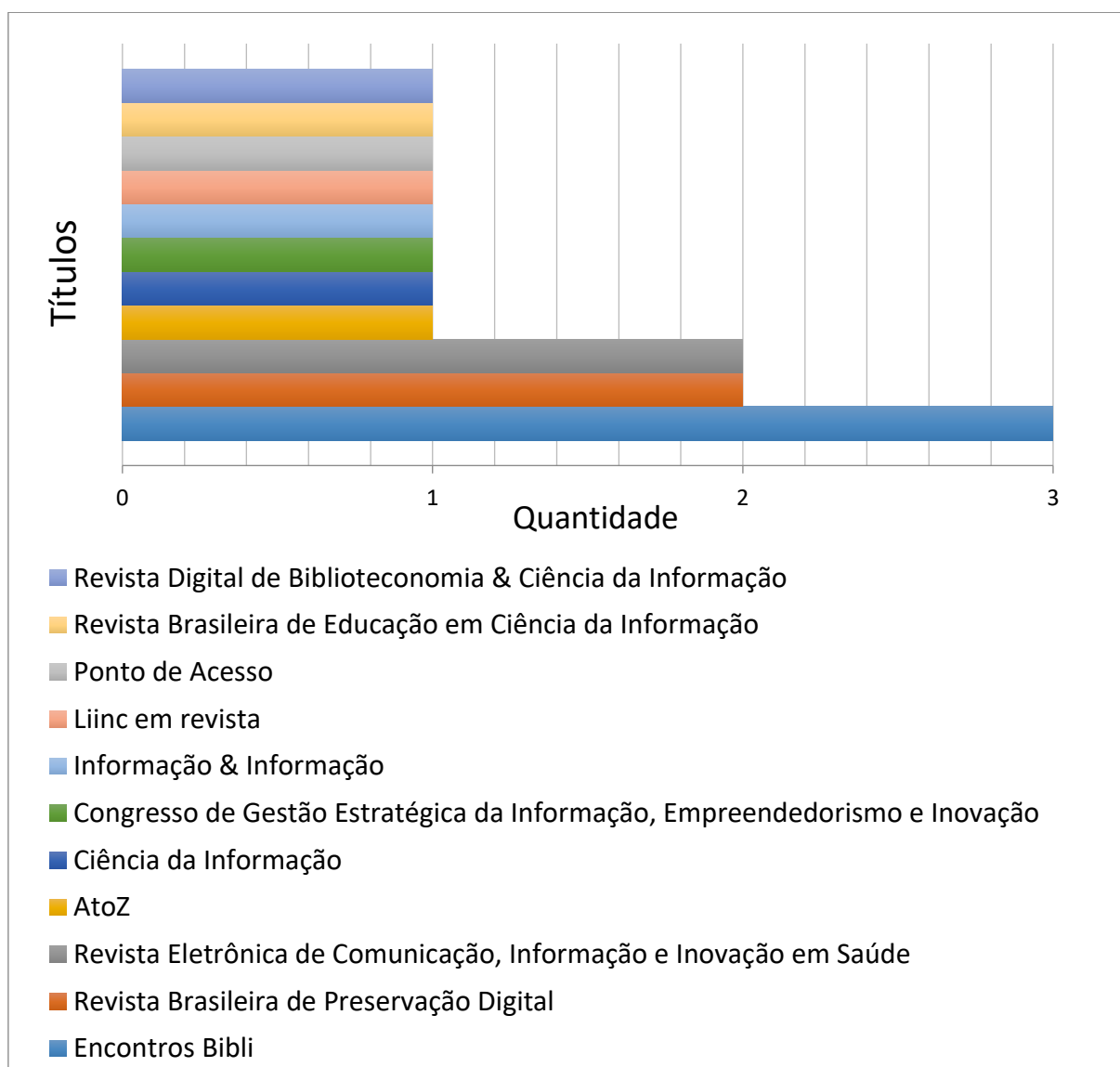


Fonte: autores.

Em relação aos títulos de onde os trabalhos foram publicados (Figura 2), encontrou-se 11 diferentes, evidenciando-se o interesse pelo tema. As revistas científicas Encontros Bibli e a Ciência da Informação têm três (3) dos artigos acerca

da temática. Revista Brasileira de Preservação Digital e Revista Eletrônica de Comunicação, Informação e Inovação em Saúde somam quatro (4) publicações. Outra análise pertinente sobre os títulos de onde os trabalhos foram publicados, é que no ano em que houve maior número de publicações de artigos, 2019, nenhum periódico publicou mais do que um artigo. Contudo, em 2017, onde três (3) artigos foram publicados, dois (2) são na Revista Eletrônica de Comunicação, Informação e Inovação em Saúde.

Figura 2 – Título dos periódicos ou eventos recuperados na pesquisa

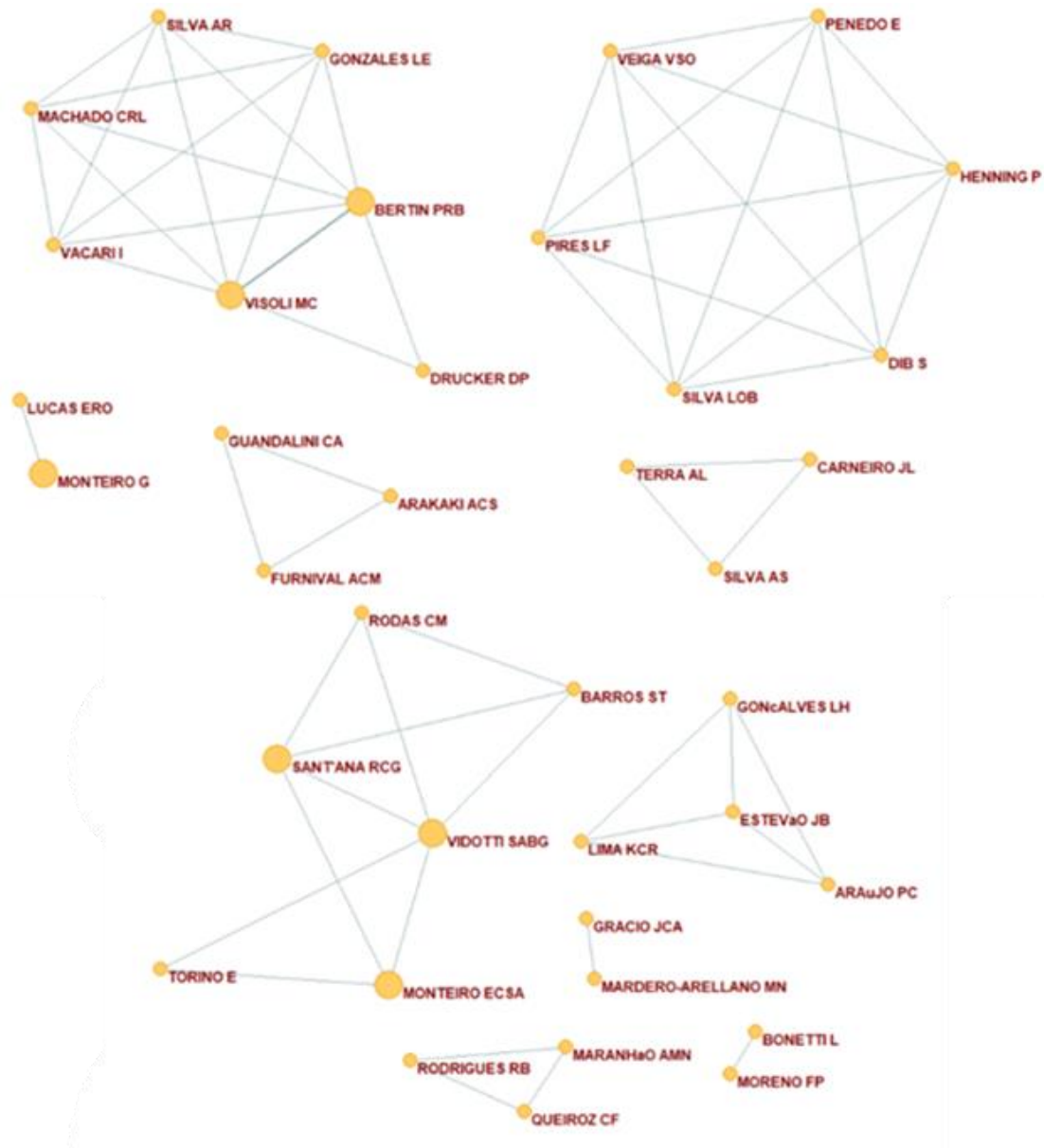


Fonte: autores.

Analisando-se a autoria, 46 autores foi o total distribuído entre os 15 artigos, sendo 40 nomes distintos. Desses autores únicos, seis (6) estão em dois (2) artigos. Sobre as instituições, 15 diferentes foram encontradas de acordo com a autoria. Nove (9) autores são da Empresa Brasileira de Pesquisa Agropecuária (Embrapa), oito (8) da Universidade Estadual Paulista “Júlio de Mesquita Filho” (Unesp) e sete (7) da Fundação Oswaldo Cruz (Fiocruz). Importante ressaltar que das 15 instituições únicas, quatro (4) são de fora do país: Holanda citadas duas vezes e Portugal citada três vezes. Por meio das instituições foi possível verificar que, além dos dois países citados (Holanda e Portugal), as regiões Sul, Sudeste, Centro-oeste e Nordeste estão representadas, faltando apenas a região Norte.

Ainda sobre autoria, a partir da Figura 3 observa-se três grandes *clusters* de autoria, onde se destacam cinco (5) dos seis (6) autores com mais publicações. A interação entre esses autores se dá também no âmbito de diferentes instituições, como por exemplo entre pesquisadores da Universidade Estadual Paulista “Júlio de Mesquita Filho” e Embrapa.

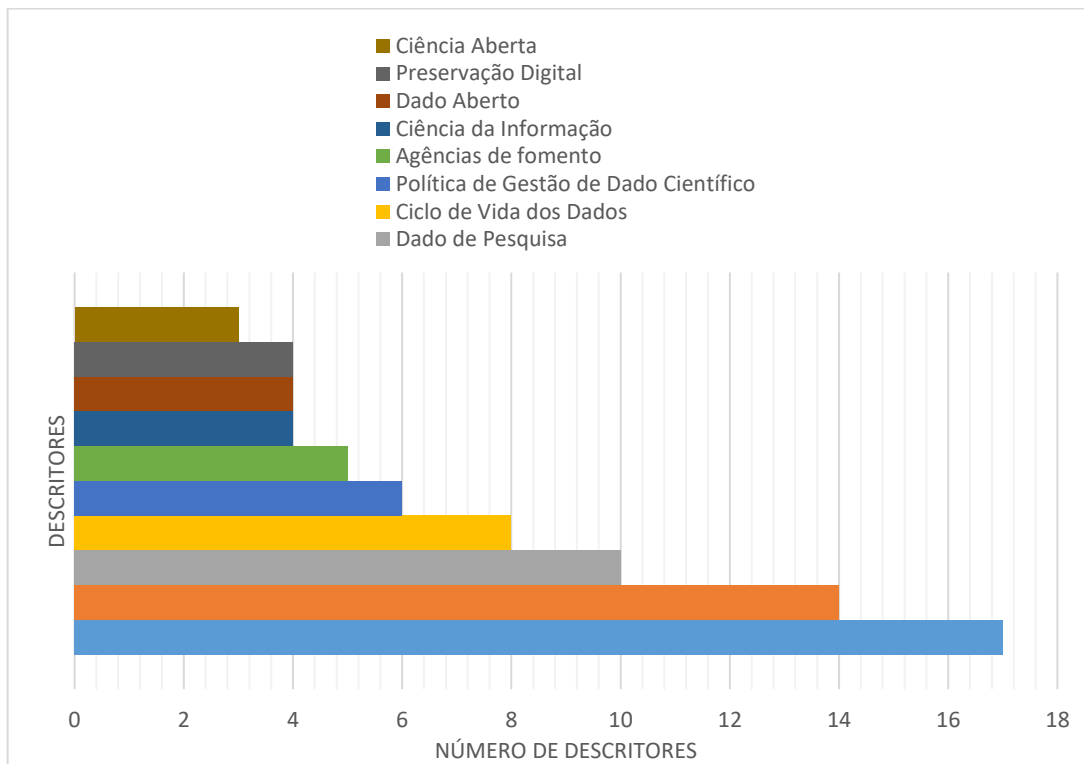
Figura 3 – *Cluster de autores*



Fonte: autores.

Em relação às palavras-chave (Figura 4), 46 descritores foram coletados, onde “Gestão de Dados” é a expressão que mais aparece, 17 vezes do total de descritores. “Plano de Gestão de Dados” aparece 14 vezes e “Dado de Pesquisa” dez (10) vezes, esses três termos juntos contabilizam 41 manifestações do total recuperado. Termos que aparecem uma única vez somam 25 palavras e variam bastante, desde nomes de instituições até termos ligados a áreas específicas.

Figura 4 – Termos listados dos artigos recuperados



Fonte: autores.

Verifica-se que os trabalhos recuperados acerca das palavras-chave são bastante diversos em relação a autoria, instituição, descritores e título das publicações.

5 CONSIDERAÇÕES FINAIS

Percebe-se que a abordagem dos PGDs está atrelada a gestão dos dados de pesquisa. Os PGDs promovem a eficiência na coleta e organização dos dados, pois ajudam os pesquisadores a planejar e organizar a coleta e o armazenamento de dados desde o início de um projeto de pesquisa.

Por tratar-se de uma temática relativamente recente se comparada com outras pesquisas da área da Ciência da Informação, nota-se um número pequeno de especialistas pesquisando o tema PGDs. Estudar os PGDs é importante para compreender e promover a transparência e integridade científica da pesquisa, além de garantir a sua reprodutibilidade e avaliação da qualidade dos dados,

proporcionando a promoção do compartilhamento por meio da correta gestão de dados e conformidade com políticas e regulamentos.

Do universo estudado, constata-se o empenho da comunidade científica em promover a conscientização e o uso dos PGDs, de forma a estimular pesquisadores e instituições sobre a sua importância. Em um contexto de Ciência Aberta, apesar da relevância, nota-se que ainda é pouco explorado o PGD enquanto instrumento vivo, interconectando todos os principais atores de um ecossistema de pesquisa.

Em pesquisas internacionais, verifica-se que a discussão acerca dos PGDs está mais avançada, uma vez que já está consolidada a importância da elaboração de planos para a gestão dos dados de pesquisa, visto que agências de fomento internacionais exigem apresentação de PGD para financiamento (UNIVERSIDADE..., 2023). No Brasil, a partir desta pesquisa, evidenciou-se que o tema ainda é pouco explorado.

É importante ressaltar que universidades, agências de financiamento e instituições de pesquisa têm se engajado cada vez mais na promoção dos PGDs, com a criação de políticas e diretrizes, bem como a oferta de treinamentos e capacitações para os pesquisadores.

Dessa forma, espera-se que esta pesquisa incentive e contribua para o entendimento e aperfeiçoamento dos PGDs, uma vez que eles ajudam a garantir a integridade da pesquisa e capacidade de replicação desses dados. Assim, os PGDs auxiliam nas boas práticas científicas, principalmente em um tempo onde há grande volume de dados.

REFERÊNCIAS

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. p. 1-19, 2016.

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: 07 jul. 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: 06 jul. 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: 07 jul. 2023.

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

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO (FAPESP). **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: 06 jul. 2023.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO (FAPESP). **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: 08 jul. 2023.

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

MIKSA, T.; WALK, P.; NEISH, P. **RDA DMP common standad for machine-actionable data management plans**. 2019. Disponível em:

<https://doi.org/10.15497/rda00039>. Acesso em: 07 jul. 2023.

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.

SMALE, N. *et al.* **The history, advocacy and efficacy of data management plans**.

Disponível em: <https://www.biorxiv.org/content/10.1101/443499v1.full>. Acesso em: 01 jan. 2023.

SMALE, N. *et al.* The History, Advocacy and Efficacy of Data Management Plans. **Scientific Communication and Education**, 2018. DOI:10.1101/443499.

STANFORD UNIVERSITY. Libraries. **Data management plans**. Disponível em:

<https://library.stanford.edu/research/data-management-services/data-management-plans>. Acesso em: 09 jul. 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: 09 jul. 2023.

UK DATA SERVICE. **Research Data Lifecycle**. 2019. Disponível em: <https://ukdataservice.ac.uk/learning-hub/research-data-management/>. Acesso em: 18 abr. 2023.

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: 08 jul. 2023.

UNIVERSIDADE FEDERAL DO ABC. **Plano de gestão de dados empesquisa**. Disponível em: <https://propes.ufabc.edu.br/perfis-de-acesso/pesquisador/ciencia-e-dados-abertos/plano-de-gestao-de-dados-em-pesquisa>. Acesso em: 27 jul. 2023.

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: 09 jul. 2023.