

Pecha kucha

Adoção dos princípios FAIR na gestão de dados científicos: desafios e prática na UFMG

*Adoption of FAIR principles in scientific data management:
challenges and practice at UFMG*

*Adopción de los principios FAIR en la gestión de datos científicos:
retos y práctica en la UFMG*

Viviane Lílian dos Santos Barrozo*

Mestrado em Gestão e Organização do Conhecimento

Universidade Federal de Minas Gerais

ORCID: <https://orcid.org/0000-0002-5893-4354>

Lattes: <http://lattes.cnpq.br/4339720406172473>

E-mail: vivibarrozoead@gmail.com

Adriléia de Moura Lima*

Mestrado em Gestão e Organização do Conhecimento

Universidade Federal de Minas Gerais

ORCID: <https://orcid.org/0000-0002-3086-0015>

Lattes: <http://lattes.cnpq.br/4615913744009144>

E-mail: adrileia.lima@gmail.com

Elisângela Cristina Aganette*

Doutora em Ciência da Informação

Universidade Federal de Minas Gerais

ORCID: <https://orcid.org/0000-0003-4357-8016>

Lattes: <http://lattes.cnpq.br/6224307970906388>

E-mail: elisangelaaganette@gmail.com

Resumo

Este estudo investigou as práticas dos pesquisadores referentes à gestão de dados científicos da Universidade Federal de Minas Gerais. A pesquisa apoia-se nos resultados de uma dissertação de mestrado de natureza aplicada e abordagem qualitativa, cujos objetivos foram exploratórios e explicativos. No presente estudo, adota-se uma metodologia quali-quantitativa. Os procedimentos técnicos incluíram pesquisa bibliográfica e de campo, com coleta de dados por questionários semiestruturados, organizados e analisados por categorização semântica. Os resultados indicaram a necessidade de aprimorar as práticas dos pesquisadores quanto à gestão dos dados científicos, fornecer orientações por meio de instrumentos de apoio aos serviços de dados e incentivar o uso de recursos tecnológicos disponíveis na universidade. As necessidades identificadas decorrem da ausência de normas institucionais para a padronização de metadados e o controle dos dados científicos com o propósito de salvaguardar e disponibilizar os dados para reuso, contribuindo para o avanço da ciência brasileira.

Palavras-chave: Ciência Aberta; Dados científicos; Gestão de dados; Princípios FAIR.

Abstract

This study investigated researchers' practices regarding scientific data management at the Federal University of Minas Gerais. The research was applied using a qualitative-quantitative approach, and had exploratory and explanatory objectives. The technical procedures included bibliographic and field research, with data collection through semi-structured questionnaires, organized and analyzed using semantic categorization. The results indicated the need to improve researchers' practices regarding scientific data management, provide guidance through tools to support data services, and promote the use of technological resources available at the university. The identified needs stem from the lack of institutional standards for standardizing metadata and controlling scientific data with the purpose of safeguarding and making data available for reuse, contributing to the advancement of Brazilian science.

Keywords: Open science; Scientific data; Data management; FAIR principles.

Resumen

Este estudio investigó las prácticas de los investigadores en la gestión de datos científicos de la Universidad Federal de Minas Gerais. La investigación fue aplicada con un enfoque cualitativo-quantitativo y objetivos exploratorios y explicativos. Los procedimientos técnicos incluyeron investigación bibliográfica y de campo, con recopilación de datos mediante cuestionarios semiestruturados, organizados y analizados mediante categorización semántica. Los resultados indicaron la necesidad de mejorar las prácticas de los investigadores en la gestión de datos científicos, brindar orientación mediante herramientas de apoyo a los servicios de datos y fomentar el uso de los recursos tecnológicos disponibles en la universidad. Las necesidades identificadas se derivan de la falta de estándares institucionales para la estandarización de metadatos y el control de datos científicos, con el objetivo de salvaguardar y poner los datos a disposición para su reutilización, contribuyendo así al avance de la ciencia brasileña.

Palabras clave: Ciencia abierta; Datos científicos; Gestión de datos; principios FAIR.

1. Introdução

Os dados científicos são fundamentais para validar pesquisas, mas sua gestão enfrenta desafios devido à diversidade de formatos e à heterogeneidade. Isso exige uma gestão ativa que assegure qualidade, acesso, compartilhamento, reuso e abertura dos dados (Silva, 2019). Este estudo investiga a adoção dos princípios FAIR (Findable, Accessible, Interoperable e Reusable) na Universidade Federal de Minas Gerais - UFMG, com foco em identificar lacunas e desafios a partir da análise de questionários aplicados a pesquisadores.

Os objetivos específicos incluem: (i) analisar práticas de gestão de dados; (ii) comparar abordagens entre áreas do conhecimento; (iii) identificar barreiras à implementação dos princípios FAIR; (iv) avaliar a relação entre práticas institucionais e diretrizes de governança digital; e (v) propor recomendações para aprimorar a gestão e a preservação dos dados.

A ausência de políticas institucionais de ciência aberta, constatada por Dias (2024), compromete práticas como o uso de metadados e a interoperabilidade (Borgman, 2015; Treloar et al., 2013). A falta de normas também dificulta o alinhamento com diretrizes internacionais, como o modelo OAIS e os princípios FAIR (Wilkinson et al., 2016).

A pesquisa responde à pergunta: quais são os desafios e as lacunas na adoção dos princípios FAIR pelos pesquisadores da UFMG?

Esta pesquisa justifica-se pela urgência em alinhar a universidade às melhores práticas de ciência aberta, promovendo transparência, reuso e visibilidade aos dados científicos. Ao propor diretrizes institucionais, o estudo contribui para a criação de políticas que fortaleçam a preservação digital e a inovação acadêmica no Brasil.

2. Dados Científicos na Perspectiva da Ciência Aberta

A gestão de dados científicos é fundamental para o avanço da ciência, sobretudo diante das exigências da ciência aberta. Sales e Sayão (2018) apontam que a produção de dados ocorre majoritariamente em pequenas equipes e laboratórios de pesquisa, que compõem a chamada “cauda longa da ciência”. Essa produção descentralizada gera conjuntos de dados diversos, que precisam ser organizados, documentados e integrados aos grandes sistemas científicos para garantir visibilidade, reutilização e continuidade da pesquisa.

A heterogeneidade dos dados exige estratégias eficazes de gestão, especialmente quando envolvem financiamento público. Nesse sentido, a FAPESP (2017) reforça a importância de que os dados sejam gerenciados e compartilhados de modo a ampliar seus benefícios para a ciência e a sociedade. Isso inclui o uso de padrões e práticas que assegurem sua preservação, acessibilidade e interoperabilidade, conforme preveem os princípios FAIR.

Tais fundamentos evidenciam a necessidade de políticas institucionais que orientem os pesquisadores quanto à gestão responsável dos dados, alinhando-se às diretrizes internacionais e fortalecendo a transparência e a inovação científica, conforme constatado no estudo de Barrozo (2022).

3. Metodologia

A pesquisa adotou uma abordagem quali-quantitativa, fundamentada na análise de questionários aplicados a pesquisadores da Universidade Federal de Minas Gerais (UFMG). O instrumento de coleta foi um questionário on-line, elaborado e disponibilizado por meio da plataforma Microsoft Forms (Office 365), com acesso restrito aos membros da comunidade acadêmica por meio de e-mails institucionais vinculados ao domínio @ufmg.br e com período delimitado para participação.

A coleta de dados envolveu a sistematização das respostas, permitindo a categorização das práticas de gestão de dados científicos e a identificação dos principais desafios enfrentados. A análise comparativa entre diferentes áreas do conhecimento visa oferecer uma compreensão mais ampla sobre a adoção dos princípios FAIR na universidade, bem como sua relação com as diretrizes institucionais de governança digital.

4. Resultados e discussões

Esta seção apresenta os procedimentos de análise dos dados coletados entre outubro de 2021 e maio de 2022, a partir de questionário respondido por 42 pesquisadores dos 92 programas de pós-graduação da UFMG, identificados como Produtores/Consumidores de Dados Científicos (PCDC). A análise buscou compreender as práticas de gestão de dados à luz dos princípios FAIR, com base no método de Bardin (2016), organizando as questões em quatro categorias correspondentes aos princípios.

O questionário foi composto por 22 questões fechadas. Para a análise, foram consideradas 16 questões que tinham aderência aos princípios FAIR, sendo 9 binárias (sim/não), 7 de múltipla escolha (*) e um campo aberto para complementações. A análise foi dividida entre as respostas binárias e de múltipla escolha.

Na dimensão “Encontrável”, conforme a Tabela 1, a Q12 revelou divisão equilibrada: 50% dos participantes afirmam manter descrição dos campos nos arquivos de dados. Já na Q15, 76,19% dizem não criar metadados. Na Q14, 78,57% relatam não manter descrições em mais de um idioma. Esses dados indicam fragilidades no atendimento a aspectos básicos de encontrabilidade, conforme os princípios FAIR.

Tabela 1. Categorização das práticas docentes na gestão de dados científicos

CATEGORIAS	SUBCATEGORIA	QUESTÃO	SIM		NÃO		TOTAL
			Freq	%	Freq	%	
Encontrável	Uso de metadados persistentes	12	21	50,00	21	50,00	42
		14	9	21,43	33	78,57	42
		15	10	23,81	32	76,19	42
Acessível	Protocolos padronizados, abertos e gratuitos	7	36	85,71	6	14,29	42
		8	33	78,57	9	21,43	42
		9*					
		17	10	23,81	32	76,19	42
		22*					
Interoperável	Formato de arquivo padronizado	10*, 13*, 20*					
Reutilizável		11	28	66,67	14	33,33	42
	Compartilhamento	16	19	45,24	23	54,76	42
		19*					
	Licenciamento	18*					
	Segurança	21	31	281,82	11	26,19	42

Fonte: dados da pesquisa (2025).

A análise da categoria “Acessível” mostrou que a maioria dos respondentes (85,71%) mantém mais de uma cópia dos dados, sendo que 78,57% armazenam essas cópias em locais distintos. No entanto, 76,19% não mantêm cópias on-line disponíveis para terceiros, evidenciando uma limitação no acesso público aos dados.

Na categoria “Reutilizável”, 66,67% verificam os arquivos contra vírus, mas apenas 45,24% disponibilizam os dados para reuso por terceiros. A ausência de controle de versionamento também é significativa, com 73,81% dos pesquisadores afirmando não manter versões organizadas dos arquivos.

A segunda parte da análise considerou respostas de múltipla escolha, conforme a Tabela 2. A maioria armazena cópias em dispositivos pessoais (84,62%), utiliza majoritariamente os formatos XLS (62,5%) e PDF (37,5%) e descreve os campos dos dados em português (84,85%). Quanto ao compartilhamento, 50% não disponibilizam os arquivos e 97,79% não os publicam em repositórios. Além disso, 81,25% não utilizam licenças de uso, o que compromete a reutilização dos dados. Por outro lado, 92,31% concordam que um repositório institucional para dados com suporte técnico facilitaria o compartilhamento e o reuso dos dados.

Tabela 2. Categorização das práticas docentes na gestão de dados científicos - múltipla escolha

CATEGORIAS	SUBCATEGORIA	QUESTÃO	Maior Freq	%	2ª Freq	%	TOTAL
Acessível	Protocolos padronizados, abertos e gratuitos	9	88	84,62	16	15,38	104
		22	24	92,31	2	7,69	26
Interoperável	Formato de arquivo padronizado	10	15	62,50	9	37,50	24
		13	28	84,85	5	15,15	33
		20	11	50,00	11	50,00	22
Reutilizável		19	30	96,77	1	3,23	31
	Licenciamento	18	13	81,25	3	18,75	16
TOTAL			209	81,64	47	18,36	256

Fonte: dados da pesquisa (2025).

Esses resultados evidenciam práticas limitadas quanto à acessibilidade e à reutilização, reforçando a necessidade de políticas institucionais que promovam infraestrutura e orientação adequadas para a gestão de dados científicos.

5. Considerações

A análise das respostas dos pesquisadores da UFMG evidencia que a ausência de diretrizes institucionais consolidadas impacta diretamente a gestão de dados científicos. Práticas como a descrição dos dados, a criação de metadados, o uso de repositórios e a aplicação de licenças ainda são adotadas de forma limitada e despadronizada, dificultando o alinhamento aos princípios FAIR. Apesar disso, os resultados revelam uma predisposição positiva por parte dos pesquisadores, especialmente quanto ao uso de repositórios institucionais, caso fossem disponibilizados. Isso indica que, com políticas claras, capacitação e infraestrutura adequada, é possível avançar significativamente na gestão e na abertura dos dados científicos, promovendo maior visibilidade e impacto para a produção científica da universidade, que direciona para pesquisas futuras.

Agradecimentos

O presente trabalho foi realizado com o apoio da Fundação de Amparo à Pesquisa de Minas Gerais – FAPEMIG (Processo 23072.248435/2024-71, Autor 1). Agradecimento também à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, Brasil (Código de Financiamento 001, Autor 2) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (Processo 306386/2025-9, Autor 3).

Conflito de Interesses

Os autores declaram não haver conflitos de interesses.

Disponibilização dos Dados de Investigação

Barrozo, V. L. S., & Marcondes, C. H. (2025). Coleta_Dados_Preservacao_Dados_Científicos_UFMG. Zenodo. <https://doi.org/10.5281/zenodo.15185940>

CRedit – Contribuições dos Autores

Viviane Lílian dos Santos Barrozo | Concetualização, Escrita – redação original, Recolha de dados, Curadoria de dados

Adriléia de Moura Lima | Curadoria de dados, Escrita – revisão e edição

Elisângela Cristina Aganette | Escrita – revisão e edição, Supervisão

Referências

Bardin, L. (2016). Análise de conteúdo (1ª edição). Edições 70.

Barrozo, V. L. S. (2022). Curadoria e preservação digital: diretrizes para a gestão de dados científicos. [Dissertação de mestrado]. Universidade Federal de Minas Gerais. <http://hdl.handle.net/1843/48991>

Barrozo, V. L. S., & Marcondes, C. H. (2025). Coleta_Dados_Preservacao_Dados_Científicos_UFMG. Zenodo. <https://doi.org/10.5281/zenodo.15185940>

Borgman, C. L. (2015). Big data, little data, no data: Scholarship in the networked world. MIT Press. <https://mitpress.mit.edu/9780262028561/big-data-little-data-no-data/>

Dias, C. G. S. (2024). Políticas Públicas e institucionais de ciência aberta no Brasil de 2020-2023. [Comunicação oral]. 9º Encontro Brasileiro de Bibliometria e Cientometria. Brasília-DF, Brasil. <https://cip.brapci.inf.br/download/306094>

Fundação de Amparo à Pesquisa do Estado de São Paulo (b). (2017, novembro). Fapesp começa a exigir plano de gestão de dados. https://www.acessoaberto.usp.br/fapesp-comeca-a-exigir-plano-de-gestao-de-dados/?doing_wp_cron=1656728574.2968099117279052734375

Sales, L. F., & Sayão, L. F. (2018). A ciência invisível: Revelando os dados da cauda longa da pesquisa. 19º Encontro Nacional de Pesquisa em Ciência da Informação, Londrina, Brasil. <https://brapci.inf.br/index.php/res/download/124831>

Silva, F. C. C. (2019). Gestão de dados científicos. Interciência.

Treloar, A., Harbourne, S., Wilkinson, R. (2013). Research Data Management Policies: Bringing Libraries and Academics Together. International Journal of Digital Curation, 8(1), 105–113.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. Scientific Data, 3(1), 1–9. <https://doi.org/10.1038/sdata.2016.18>