



# QUALIDADE DOS DADOS DE PESQUISA: aplicação dos princípios FAIR a partir do Altmetric TOP 100

## Francielle Franco dos Santos

 <https://orcid.org/0000-0002-3686-595X>.

✉ [franfranco.santos@gmail.com](mailto:franfranco.santos@gmail.com).

🏢 Universidade Estadual Paulista (UNESP) |

ROR: <https://ror.org/00987cb86> | Marília, Brasil.

## Caterina Marta Groposo Pavão

 <https://orcid.org/0000-0003-3712-7200>.

✉ [caterina@cpd.ufrgs.br](mailto:caterina@cpd.ufrgs.br).

🏢 Universidade Federal do Rio Grande do Sul (UFRGS) |

ROR: <https://ror.org/041yk2d64> | Porto Alegre, Brasil.

## Maurício Coelho da Silva

 <https://orcid.org/0000-0002-7923-9457>.

✉ [mauriciocoelho.hlp@gmail.com](mailto:mauriciocoelho.hlp@gmail.com).

🏢 Universidade Federal de Santa Catarina (UFSC) |

ROR: <https://ror.org/041akq887> | Florianópolis, Brasil.

---

**Eixo temático:** Acesso Aberto, Ciência Aberta e Dados Abertos

**Modalidade:** Resumo expandido

**DOI:** 10.22477/ix.ebbc.243

**Resumo:** Estudo de caráter exploratório que investigou a aderência aos Princípios FAIR dos dados de pesquisa relacionados aos artigos com maior alcance segundo o Altmetric TOP 100 de 2020. O *corpus* de análise é composto de 15 artigos, dos quais oito apresentaram *datasets* em conformidade com os princípios FAIR. Repositórios de dados como *GitHub* e o *FigShare* demonstram significativa aderência aos princípios FAIR, enquanto que *datasets* que são anexados como arquivos suplementares a alguns artigos prejudica a adesão aos princípios. A convergência entre Altmétrica e Ciência Aberta demonstrou potencial para estudos da padronização de diferentes produtos científicos na internet.

**Palavras-Chave:** Ciência Aberta. Dados de Pesquisa. Princípios FAIR. Altmétrica.

## 1 INTRODUÇÃO

Ciência Aberta (CA) é um conjunto de práticas científicas com o objetivo de tornar o processo mais transparente e democratizar o acesso ao conhecimento produzido nas pesquisas. Dentre as recomendações, destaca-se o compartilhamento e reuso dos dados de pesquisa como forças desse movimento, permitindo que diferentes comunidades possam utilizar os dados disponibilizados, fomentando o avanço científico. Considerando que existem diferentes práticas científicas entre as áreas do conhecimento e que isso pode dificultar o compartilhamento dos conjuntos de dados, foram desenvolvidos os princípios FAIR, um conjunto boas práticas desenvolvidas para assegurar a encontrabilidade (*Findable*), acessibilidade (*Accesible*), interoperabilidade (*Interoperable*) e reuso (*Reusable*) do dados.

A internet desempenha papel crucial na CA, permitindo o armazenamento, publicação e distribuição de produtos científicos para diversos públicos. A Altméria, subcampo dos Estudos Métricos da Informação, exerce uma função importante na mensuração do impacto científico, principalmente em plataformas sociais. Assim, este estudo, de caráter exploratório, busca investigar a aderência aos Princípios FAIR dos dados de pesquisa relacionados aos artigos com maior alcance segundo o *Altmetric TOP 100* de 2020. Essa aderência poderá indicar que esses artigos que alcançam um grande público estão alinhados às práticas da CA.

## 2 REFERENCIAL TEÓRICO

A CA representa uma abordagem que busca transformar a prática científica tradicional, promovendo transparência, acessibilidade e colaboração por meio da disponibilização aberta (pública e gratuita) dos dados, métodos e resultados de pesquisa (Albagli, 2015). A CA ainda contempla um conjunto de ações que contribuem para melhorar a replicabilidade, robustez e acessibilidade da pesquisa, aumentando a diversidade na investigação científica (Spellman; Gilbert; Corker, 2018).

O reúso de dados é uma prática significativa de CA para o progresso científico, pois o compartilhamento de conjuntos de dados pode acelerar o desenvolvimento científico, incentivar a transparência, a partilha de recursos, a reprodutibilidade, entre outros. Entretanto, esse esforço requer tempo, organização e padronização dos dados a serem disponibilizados (Bornmann *et al.*, 2021). Este esforço pode ser especialmente oneroso para pesquisadores que possuem tempo ou recursos escassos, fato recorrente entre os pesquisadores brasileiros.

Neste panorama os princípios FAIR tencionam aumentar a capacidade das máquinas de encontrar e sugerir a utilização de dados de pesquisa de forma automatizada (Wilkinson *et al.*, 2016), mas também para uso por pessoas, e se forem devidamente adotados viabilizam a interoperabilidade entre diferentes ambientes de dados, auxiliando o pesquisador na reutilização mesmos. O Quadro 1 apresenta as quatro diretrizes propostas e os princípios que derivam de cada uma delas.

Quadro 1 - Princípios FAIR.

Diretriz	Princípio
<b>Findable</b> (Encontrável)	<b>F1.</b> Os (meta) dados recebem um identificador globalmente exclusivo e persistente.
	<b>F2.</b> Os dados são descritos com metadados ricos (definidos por R1 abaixo).
	<b>F3.</b> Os metadados incluem clara e explicitamente o identificador dos dados que descrevem.
	<b>F4.</b> Os (meta) dados são registrados ou indexados em um recurso pesquisável.
<b>Accessible</b> (Acessível)	<b>A1.</b> Os (meta) dados são recuperáveis pelo seu identificador usando um protocolo de comunicação padronizado.
	<b>A1.1.</b> O protocolo é aberto, gratuito e universalmente implementável.
	<b>A1.2.</b> O protocolo permite um procedimento de autenticação e autorização.
	<b>A2.</b> Os metadados são acessíveis, mesmo quando os dados não estão mais disponíveis.
<b>Inter-operable</b> (Interoperável)	<b>I1.</b> Os (meta) dados utilizam uma linguagem formal, acessível, partilhada e amplamente aplicável para representação do conhecimento.
	<b>I2.</b> Os (meta) dados usam vocabulários que seguem os princípios FAIR.
	<b>I3.</b> Os (meta) dados incluem referências qualificadas a outros (meta) dados.
<b>Reusable</b> (Reuso)	<b>R1.</b> Os (meta) dados são ricamente descritos com uma pluralidade de atributos precisos.
	<b>R1.1.</b> Os (meta) dados apresentam uma licença de uso de dados clara e acessível.
	<b>R1.2.</b> Os (meta) dados fornecem proveniência detalhada
	<b>R1.3.</b> Os (meta) dados atendem aos padrões da comunidade relevantes para o domínio.

Fonte: Traduzido e adaptado de Wilkinson *et al.*, 2016.

Bornmann *et al.* (2021) analisou os princípios FAIR do ponto de vista da cientometria e conclui que o requisito *Findable* pode ser considerado satisfatório em conjuntos de dados disponibilizados pelas plataformas mais populares como *Clarivate Analytics*, *Elsevier*, *Dimensions* e *Microsoft Academic*; por outro lado a interoperabilidade e o reuso do dados não são problemas relevantes porque geralmente cada estudo cientométrico faz a sua própria coleta. Bornmann *et al.* (2021) descrevem resistência no compartilhamento de conjuntos de dados cientométricos, o que pode ocorrer em função da sua má documentação e sua variedade de fontes. Nesse sentido, os dados altmétricos enfrentam igual desafio em relação à adesão aos princípios FAIR, uma vez que a padronização desse tipo de dado está submetida a volatilidade e diversidade de fontes que a *web social* oferece (Bornmann *et al.*, 2021).

A internet é uma das principais propulsoras da CA (Bornmann *et al.*, 2021), além disso “[...] em muitos aspectos, a acelerada evolução da internet e da *web social* ocorre em paralelo ao movimento de concepção e expansão da CA [...]” (Araujo, 2014, p. 32), portanto disseminar um artigo em acesso aberto aumenta as chances de citação e de menção na *web* (Brainard, 2024). Nesse sentido, apresentam-se as possíveis relações entre Altmétrie e CA na medida em que esse subcampo dos estudos métricos está baseado em indicadores de plataformas sociais e sua coleta e posterior análise está fortemente relacionada ao acesso aberto.

### 3 MÉTODO E RESULTADOS

A partir da lista dos 100 artigos com maior *Score* Altmétrico<sup>1</sup>, verificou-se quais disponibilizam os conjuntos de dados e os *datasets* identificados foram analisados para constatar sua aderência e alinhamento aos princípios FAIR (conforme o Quadro 1). O *Altmetric Top 100* é uma lista organizada pela plataforma *Altmetric* e divulga anualmente os 100 artigos com maior atenção *online* segundo os critérios da própria plataforma, não definindo qualidade ou impacto das pesquisas, apenas o alcance na *web social*. A classificação apresenta os cinco artigos mais mencionados distribuídos em 20 áreas temáticas. Para fins de análise, optou-se por selecionar pelo menos um artigo por área, desse modo, foram selecionados os artigos com maior *score* alométrico de cada área e que disponibilizassem o conjunto de dados utilizado no estudo. O Quadro 2 apresenta os artigos selecionados.

**Quadro 2** - Artigos por área temática e que disponibilizavam os conjuntos de dados.

Nº	Score	Title	Periódico	Área	Autores
1	26745	Global shifts in mammalian population trends reveal key predictors of virus spillover risk	Proceedings of the Royal Society B: Biological Sciences	Ciências Agrárias e Veterinárias	JOHNSON, C. K. <i>et al.</i>
2	6516	Phosphine gas in the cloud decks of Venus	Nature Astronomy	Química	GREAVES, J. S. <i>et al.</i>
3	4488	Comparative Greenhouse Gas Footprinting of Online versus Traditional Shopping for Fast-Moving Consumer Goods: A Stochastic Approach	American Chemical Society (ACS)	Comércio, Gestão, Turismo e Serviços	SHAHMOHAMMADI, S. <i>et al.</i>
4	4427	Global human-made mass exceeds all living biomass	Nature	Ciências da Terra	ELHACHAM, E. <i>et al.</i>
5	3669	The effect of large-scale anti-contagion policies on the COVID-19 pandemic	Nature	Ciências Econômicas	HSIANG, S. <i>et al.</i>
6	2924	Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math	Proceedings of the National Academy of Sciences of the United States of America	Educação	THEOBALD, E. J. <i>et al.</i>
7	2101	Cross-species transmission of the newly identified coronavirus 2019-nCoV	Journal of Medical Virology	Ciências Ambientais e Biológicas	Jl, W. <i>et al.</i>

1 Disponível em: <https://www.altmetric.com/blog/top-100-2021-feat-subjects/>.



Nº	Score	Title	Periódico	Área	Autores
8	2060	Origins and genetic legacy of prehistoric dogs	Science	História e Arqueologia	BERGSTRÖM, A. <i>et al.</i>
9	1968	A scalable pipeline for designing reconfigurable organisms	Proceedings of the National Academy of Sciences of the United States of America	Ciências da Informação e da Computação	KRIEGMAN, S.; <i>et al.</i>
10	1720	Effect of voicing and articulation manner on aerosol particle emission during human speech	PLoS ONE	Linguagem, Comunicação e Cultura	ASADI, S. <i>et al.</i>
11	1239	Variability in the analysis of a single neuroimaging dataset by many teams	Nature	Ciências Matemáticas	BOTVINIK-NEZER. R.. <i>et al.</i>
12	607	Periodic activity from a fast radio burst source	Nature	Física	AMIRI, M. <i>et al.</i>
13	564	Tracking historical changes in trustworthiness using machine learning analyses of facial cues in paintings	Nature Communications	Psicologia e Ciências Cognitivas	SAFRA, L. <i>et al.</i>
14	414	Conductors' tempo choices shed light over Beethoven's metronome	PLoS ONE	Artes e escrita	MARTIN-CASTRO, A.; UCAR, I.
15	376	The association between early career informal mentorship in academic collaborations and junior author performance	Nature Communications	Estudos em Ciências Sociais	ALSHEBLI, B.; MAKOVI, K.; RAHWAN, T.

Fonte: *Altmetric Engineering* (2021).

Dentre as 20 áreas relacionadas na *Altmetric Top 100*, cinco não disponibilizam os artigos com seus respectivos conjuntos de dados de forma abertas e acessível, sendo elas: Arquitetura e *Design*; Engenharia e Tecnologia; Medicina e Ciências Médicas; Filosofia e Estudos Religiosos. Cabe ressaltar que alguns artigos disponibilizavam os *datasets* mediante solicitação, o critério adotado nesta investigação é de que o acesso aos dados deveria ser aberto, portanto descartaram-se aqueles onde era necessário solicitar acesso. As coletas e análises aconteceram em dezembro de 2023. Conforme os dados no Quadro 2, observa-se a presença relevante da *Nature* (sete artigos), nota-se que o alcance desse periódico independe da área. Destaca-se aqui a importância do papel do periódico no compartilhamento dos dados de pesquisa em conformidade com os princípios FAIR, conforme mencionado por Bornmann *et al.* (2021). No Quadro 3, a seguir, apresenta-se a análise dos dados de pesquisa que puderam ser acessados.

**Quadro 3** - Artigos analisados de acordo com os princípios FAIR.

Artigo	Diretriz/Princípio														
	Findable				Accessible				Interoperable			Reusable			
	F1	F2	F3	F4	A1.	A1.1	A1.2	A2	I.1	I.2	I.3	R.1	R.1.1	R.1.2	R.1.3
1	•	•	•	•	•	•	•	-	-	-	-	•	•	•	•
2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	-	-	-	-	-	-	-	-	-	-	-	•	•	•	•
5	•	•	•	•	•	•	•	-	-	-	-	•	•	•	•
6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	•	•	•	•	•	•	•	-	-	-	-	•	•	•	•
11	•	•	•	•	•	•	•	-	-	-	-	•	•	•	•
12	•	•	•	•	•	•	•	-	-	-	-	•	•	•	•
13	•	•	•	•	•	•	•	-	-	-	-	•	•	•	•
14	•	•	•	•	•	•	•	-	-	-	-	•	•	•	•
15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

**Fonte:** Dados da pesquisa, 2023.

**Legenda:** • = em conformidade; - = não conformidade.

A respeito da diretriz de encontrabilidade (*Findable*) destaca-se que pouco menos da metade, sete artigos, atendem este item, contudo aqueles que não atendem esses princípios de maneira satisfatória, deve-se principalmente pelo fato dos dados estarem anexados como arquivos suplementares aos artigos, desse modo não possuíam um identificador único que os representasse. Portanto a encontrabilidade destes *datasets* depende da manutenção da indexação dos artigos nas bases de dados, e assim não estão em conformidade com o item F1.

Sobre o requisito de acessibilidade (*Accessible*) não foi possível avaliar o item A2 porque este ponto refere-se a disponibilidade dos metadados mesmo quando os dados não estiverem mais disponíveis, contudo, os dados de todos os artigos selecionados estavam disponíveis, bem como seus metadados. Por outro lado, sete artigos alcançaram os princípios de acessibilidade conforme definido, denota-se que todos os artigos com encontrabilidade também puderam ser considerados acessíveis.

A interoperabilidade, conforme descrita pelos princípios FAIR, não pode ser alcançada por nenhum dos artigos analisados, entretanto observou-se a extensão dos arquivos nos quais os dados eram disponibilizados. Destacou-se o uso dos formatos *xls* e *csv* (*Microsoft Excel*) e *PDF*. Além de não estarem em conformidade com os princípios que definem a diretriz de interoperabilidade, os dados ainda foram apresentados em formatos proprietários de arquivo, o que pode ainda dificultar o reuso. Quanto aos princípios de reuso, observa-se que, conforme as demais diretrizes, estiveram presentes em oito dos 15 artigos analisados.



As plataformas de compartilhamento de dados mais mencionadas nos artigos foram *GitHub* e *FigShare*, diferentemente do *Figshare* e do *Zenodo*, o *GitHub* não fornece um identificador persistente, o que tem impacto direto nas diretrizes de encontrabilidade e acesso, por outro lado o *GitHub* é uma plataforma reconhecida entre a comunidade de desenvolvedores por oferecer diversos recursos que facilitam o compartilhamento de códigos desenvolvidos nos estudos ou que tenha sido utilizados na coleta dos dados.

Destaca-se que a relação entre Altmtria e CA não se dá apenas pelas duas se favoreceram das ferramentas e dinâmicas da *web social*. A Altmtria possui um papel significativo na transição para um ecossistema científico nos moldes propostos pela CA, pois permite recompensar diferentes produtos científicos e práticas propostas por esse movimento, resultando em um sistema de avaliação e recompensa científica mais inclusivo e multidimensional, sendo que a métrica ainda está alinhada com a Escola das Métricas e a demanda acadêmica por métricas responsáveis para uma ciência inclusiva (Araujo, 2020). Esse estudo possui resultados que podem indicar uma convergência entre a métrica e o movimento da CA nesse sentido, pois a mensuração e avaliação de datasets como um produto científico por meio de abordagens alométricas permite reconhecer e incentivar o compartilhamento de dados de pesquisa para reuso.

## 4 CONSIDERAÇÕES FINAIS

O estudo buscou compreender como são caracterizados, de acordo com os princípios FAIR, os dados compartilhados junto aos artigos relacionados no *Altmtric TOP 100* do ano de 2020. O *corpus* de análise foi composto de 15 artigos de diferentes áreas temáticas com conjuntos de dados passíveis de análise. Identificou-se que pelo menos metade dos artigos apresentavam *datasets* em conformidade com os princípios FAIR.

Dentre as diretrizes pode-se destacar a encontrabilidade, a acessibilidade e o reuso como pontos fortes dos *datasets* analisados, bem como a importância de repositórios de dados como *GitHub* e o *FigShare* que dispõem de funcionalidades que propiciam a correta aplicação dos princípios FAIR. Um fator preponderante para o não alcance dos princípios foi o fato dos dados estarem anexados junto aos artigos como arquivos suplementares, destacando a importância dos periódicos definirem políticas que incentivem o depósito correto dos dados de pesquisa, em conformidade com os princípios FAIR. A aproximação entre Altmtria e CA permitiu analisar aspectos da comunicação de um produto científico específico, mesmo se tratando da etapa inicial do estudo. Pode-se destacar como limitação a quantidade de artigos analisadas, uma amostra maior permitiria análises aprofundadas. Como estudos futuros pretende-se ampliar a amostra e aprofundar o debate por meio da comparação entre o TOP 100 de diferentes anos.



## REFERÊNCIAS

- ALBAGLI, Sarita. Ciência aberta em questão. *In*: ALBAGLI, Sarita; MACIEL, Maria Lucia; ABDO, Alexandre Hannud (org.). **Ciência aberta, questões abertas**. Rio de Janeiro: UNIRIO, 2015. p. 9-25. Disponível em: <https://livroaberto.ibict.br/handle/1/1060>. Acesso em: 8 mar. 2024.
- ALTMETRIC ENGINEERING. **Data for the 2020 Altmetric Top 100**, dataset posted on 19 Jan. 2021. [Site]. DOI: <https://doi.org/10.6084/m9.figshare.13607312.v2>. Disponível em: [https://altmetric.figshare.com/articles/dataset/Data\\_for\\_the\\_2020\\_Altmetric\\_Top\\_100/13607312/2](https://altmetric.figshare.com/articles/dataset/Data_for_the_2020_Altmetric_Top_100/13607312/2). Acesso em: 3 dez. 2023.
- ARAUJO, Ronaldo Ferreira. Ciência 2.0 e a presença online de pesquisadores: visibilidade e impacto. **Ciência da Informação em Revista**, Maceió, v. 1, n. 3, p. 32-40, set./dez. 2014. Disponível em: <https://www.seer.ufal.br/index.php/cir/article/view/1608>. Acesso em: 3 dez. 2023.
- ARAUJO, Ronaldo Ferreira. Ciência aberta e altmetria: aproximações e desafios. *In*: MOREIRA, Luciana Albuquerque; SOUZA, Jacqueline Aparecida de; TANUS, Gabrielle Francinne de Souza Carvalho (org.). **Informação na sociedade contemporânea**. Florianópolis: Rocha Gráfica e Editora, 2020. p. 39-50. Disponível em: <https://repositorio.ufrn.br/handle/123456789/31021>. Acesso em: 12 mar. 2024.
- BORNMANN, Lutz; GUNS, Raf; THELWALL, Michael, WOLFRAM, Dietmar. Which aspects of the Open Science agenda are most relevant to scientometric research and publishing? **An opinion paper**, Cambridge, v. 2, n. 2, 2021, 438-453. DOI: [https://doi.org/10.1162/qss\\_e\\_00121](https://doi.org/10.1162/qss_e_00121). Disponível em: <https://direct.mit.edu/qss/article/2/2/438/96160/Which-aspects-of-the-Open-Science-agenda-are-most>. Acesso em: 15 dez. 2023.
- BRAINARD, Jeffrey. Open-access papers draw more citations from a broader readership. **Science**, 24 Jan. 2024. [Site]. DOI: <https://doi.org/10.1126/science.zb4sw6i>. Disponível em: <https://www.science.org/content/article/open-access-papers-draw-more-citations-broader-readership>. Acesso em: 1 fev. 2024.
- SPELLMAN, Barbara A.; GILBERT, Elizabeth; CORKER, Katherine S. **Open Science**: Stevens' handbook of experimental psychology and cognitive neuroscience. New Jersey: Wiley, 2018. v. 5. DOI: <https://doi.org/10.1002/9781119170174.epcn519>. Disponível em: <https://onlinelibrary.wiley.com/doi/epdf/10.1002/9781119170174.epcn519>. Acesso em: 13 dez. 2023.
- WILKINSON, Mark D. et al. The FAIR Guiding principles for scientific data management and stewardship. **Scientific Data**, London, v. 3, 160018, 2016. DOI: <https://doi.org/10.1038/sdata.2016.18>. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 13 dez. 2023.