

UNIVERSIDADE FEDERAL DE ALAGOAS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE
BACHARELADO EM CIÊNCIAS BIOLÓGICAS

LEONARDO SILVA

QUAL O IMPACTO DE DADOS ABERTOS PARA A PRODUÇÃO CIENTÍFICA
SOBRE ODONATA?

MACEIÓ-AL
2023

LEONARDO SILVA

**QUAL O IMPACTO DE DADOS ABERTOS PARA A PRODUÇÃO CIENTÍFICA
SOBRE ODONATA?**

Trabalho de Conclusão de Curso apresentado à Coordenação do Curso de Ciências Biológicas da Universidade Federal de Alagoas, como requisito para obtenção do título de Bacharel em Ciências Biológicas.

Orientador: Prof. Dr. Marcos Vinícius Carneiro Vital

Maceió-AL

2023

Catálogo na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico
Bibliotecária: Taciana Sousa dos Santos – CRB-4 – 2062

S586q Silva, Leonardo.
Qual o impacto de dados abertos para a produção científica sobre
Odonata? / Leonardo Silva. – 2023.
26 f. : il.

Orientador: Marcos Vinícius Carneiro Vital.
Monografia (Trabalho de Conclusão de Curso em Ciências Biológicas:
Bacharelado) – Universidade Federal de Alagoas. Instituto de Ciências
Biológicas e da Saúde. Maceió, 2023.

Bibliografia: f. 24-26.

1. Replicação – Crise. 2. Ciência aberta. 3. Odonata. 4. Dados abertos. I.
Título.

CDU: 595.733 : 004.65

Dedico

A minha mãe, pelo apoio em todos os momentos difíceis da minha trajetória acadêmica.

AGRADECIMENTOS

Gostaria de expressar minha gratidão a todos os funcionários da Instituição que, de alguma forma, contribuíram para o meu crescimento e desenvolvimento ao longo dos anos, mesmo enfrentando cortes e adversidades durante esse período.

Ao meu orientador, Marcos Vital, por ter me acolhido com carinho no Laboratório de Ecologia Quantitativa - LEQ e por tudo o que fez por mim durante minha graduação. Foi incrível e um privilégio poder trabalhar sob sua orientação, aprendi muito com você e sei que posso contar com o seu apoio a qualquer momento.

Aos meus companheiros de laboratório, Alana, Arthur, Carla, Hellen, e demais colegas de laboratório. Obrigado pelo apoio, conversas e compartilhar seus conhecimentos durante meus anos de graduação. Sinto muito carinho por cada um de vocês e espero que possamos continuar trabalhando juntos.

Por fim, quero agradecer especialmente aos meus amigos e amigas que me acolheram quando voltei ao curso e passaram por todas as situações e momentos difíceis comigo. Mércia, Ellen, Alcía, Thuelly e Fernanda, vocês foram fundamentais para que eu pudesse chegar até aqui e agradeço pelos bons momentos que passamos juntos.

Não tenhamos pressa, mas não percam tempo.
(José Saramago).

RESUMO

Atualmente, a ciência enfrenta uma crise de replicabilidade devido a falhas de transparência e rigor nos métodos de pesquisa. Uma das possíveis soluções para isso é a adoção dos princípios de ciência aberta, que inclui o compartilhamento de dados, códigos e scripts de análise. Este estudo investigou se a utilização de práticas de ciência aberta influencia as publicações científicas na área de Zoologia, especificamente em termos de número de citações. Para isso, foram coletados artigos do repositório digital *Dryad* que foram encontrados através da busca pela palavra-chave “Odonata”. O estudo comparou o efeito de dados e acesso abertos nas publicações científicas. Os dados foram processados no software *R* e o número de citações foi usado como medida para avaliar o impacto da disponibilização de dados em formato aberto. O teste T pareado foi realizado para comparar o número de citações de artigos com dados publicados na *Dryad* com a mediana das citações dos demais artigos publicados no mesmo periódico, volume e edição. Foram obtidos 44 artigos na *Dryad*, mas 19 deles foram descartados por inconsistências. Os demais artigos catalogados, mas que não possuem dados na *Dryad*, somam 335 artigos. O teste T pareado não apresentou evidências estatisticamente significativas de que a publicação de dados abertos na plataforma *Dryad* tenha tido algum impacto no número de citações dos artigos. Este resultado foi diferente de outros estudos anteriores, mas é importante considerar que o número de variáveis analisadas neste estudo difere dos outros. Estas variáveis podem ter influenciado o número de citações. É possível que o impacto dos dados abertos na produção científica seja menor em áreas de estudo específicas, como ocorreu neste estudo sobre Odonata. Isso pode ser devido ao escopo restrito da pesquisa e ao ecossistema de pesquisadores relativamente pequeno. Além disso, o uso do número de citações como métrica para avaliar o impacto de um trabalho científico pode ter influenciado os resultados, pois não considera outras formas de utilização da pesquisa científica. Utilizar métricas alternativas, como altmetrics, pode fornecer uma visão diferente destes dados.

Palavras-chave: Crise de Replicabilidade; Ciência aberta; Dados abertos; Odonata.

ABSTRACT

Currently, science is facing a replicability crisis due to a lack of transparency and rigour in research methods. One possible solution to this is the adoption of the principles of open science, which includes the sharing of data, code, and analysis scripts. This study investigated whether the use of open science practices influences scientific publications in the field of Zoology, specifically in terms of the number of citations. To do this, articles from the digital repository Dryad were collected and found through a search using the keyword "Odonata." The study compared the effect of open data and access on scientific publications. The data was processed in R and the number of citations was used as a measure to evaluate the impact of data availability in open format. The paired T-test was conducted to compare the number of citations of articles with data published in Dryad to the median of the citations of other articles published in the same journal, volume, and edition. 44 articles were obtained in Dryad, but 19 of them were discarded due to inconsistencies. The remaining catalogued articles, but which do not have data in Dryad, amount to 335 articles. The paired T-test did not show statistically significant evidence that the publication of open data on the Dryad platform had any impact on the number of citations of the articles. This result was different from previous studies, but it is important to consider that the number of variables analysed in this study differs from others. These variables may have influenced the number of citations. It is possible that the impact of open data on scientific production may be smaller in specific areas of study, as occurred in this study on Odonata. This may be due to the restricted scope of the research and the relatively small ecosystem of researchers. In addition, using the number of citations as a metric to evaluate the impact of scientific work may have influenced the results, as it does not consider other forms of use of scientific research. Using alternative metrics, such as altmetrics, may provide a different view of this data.

Keywords: Replicability Crisis; Open Science; Open Data; Odonata.

SUMÁRIO

1 INTRODUÇÃO	9
2 REFERENCIAL TEÓRICO	11
2.1 Ordem odonata	11
2.2 Ciência aberta e acesso aberto	12
2.3 Bibliometria	13
3 METODOLOGIA	15
3.1 Banco de artigos	15
3.2 Análise bibliométrica	15
3.3 Teste de hipótese	16
4 RESULTADOS E DISCUSSÃO	17
5 CONCLUSÃO	23
REFERÊNCIAS	24

1 INTRODUÇÃO

A replicabilidade é considerada um princípio fundamental da ciência e se refere à obtenção dos mesmos resultados quando um novo estudo é realizado utilizando a mesma questão de pesquisa de outro estudo, mas com dados diferentes (HELD; SCHWAB, 2020). Isso permite uma verificação, confirmação ou revisão e correção dos resultados por parte de outros pesquisadores. No entanto, nos últimos anos, tem sido observada uma crescente “crise de replicabilidade”, ou seja, uma dificuldade em se replicar os resultados de muitos estudos científicos publicados (BAKER, 2016; Begley & Ellis, 2012; Ioannidis, 2005). Esta crise pode ser atribuída a vários fatores, incluindo a falta de transparência e rigor nos métodos de pesquisa (NATIONAL ACADEMIES OF SCIENCES, ENGINEERING AND MEDICINE, 2019; Open Science Collaboration, 2015), o uso de técnicas estatísticas inadequadas (Nosek et al., 2015) e a pressão por resultados positivos e inovadores (Fanelli, 2012), limitando a publicação de estudos com resultados considerados negativos ou inconclusivos, por exemplo.

De acordo com HAMPTON et al. (2015), a não disponibilização dos dados brutos e dos detalhes das análises da pesquisa pode ser prejudicial tanto para os pesquisadores quanto para os periódicos. Isso ocorre porque, caso haja algum erro na etapa experimental da pesquisa, será difícil detectá-lo se os dados não forem disponibilizados. Consequentemente, esses erros podem ser replicados em futuras pesquisas ou até mesmo utilizados para chegar a conclusões incorretas. O que pode levar a retratações públicas, aumento da desconfiança da sociedade em relação à ciência (Nosek & Bar-Anan, 2012) e diminuição da confiança nos resultados entre os pesquisadores. Além disso, isto gera um esforço desnecessário por parte de outros pesquisadores ao tentar reproduzir os experimentos ou replicar seus resultados (Fanelli, 2009), fazendo com que o avanço das pesquisas sobre os temas envolvidos leve mais tempo do que seria necessário (Baker, 2016).

Além da limitação de acesso a dados e análises, os textos dos trabalhos científicos em si nem sempre estão imediatamente disponíveis e acessíveis para outros pesquisadores ou para o público geral. Dentre as possibilidades disponíveis aos autores ao concluir uma pesquisa científica, a mais comum é a submissão a um periódico científico. Neste caso, o acesso ao conteúdo da pesquisa, pode ser condicionado a realização de uma assinatura, limitando o alcance do trabalho em

consequência dos custos associados. Também existe a possibilidade de arquivamento em repositórios públicos digitais como repositórios de *preprints*, permitindo o livre acesso ao trabalho (NORRIS; OPPENHEIM; ROWLAND, 2008).

Nesse contexto, a adoção de práticas de ciência aberta (Open Science) surge como uma alternativa à estrutura tradicional de pesquisa e publicação ao promover o compartilhamento dos trabalhos científicos (Open Access), dos dados (Open Data), códigos, scripts de análise e software (Open Source) (McKiernan et al., 2016). Hampton (2015) ainda descreve que a área da ecologia está no limiar para uma crescente adoção dessas práticas, e para isto somam fatores como incremento do poder computacional, surgimento de novas tecnologias e ferramentas da Internet e o interesse dos pesquisadores desta área em tornar suas pesquisas mais abertas e acessíveis.

Diversas áreas são afetadas pela falta de transparência da pesquisa e dentre elas estão a ecologia e a biologia evolutiva; por isso, faz-se necessário o compartilhamento de dados, uma vez que isto aumenta a qualidade dos estudos sobre os temas desses campos. Apesar dos possíveis benefícios decorrentes da aplicação dessa prática, ainda há várias barreiras a serem enfrentadas, visto que no geral a sociedade se beneficia com abertura dos dados de ecologia, concomitantemente tem de lidar com os desafios em consequência da natureza heterogênea de dados (FERNÁNDEZ-JURICIC, 2021). A padronização e desenvolvimento de métodos pode ser a solução para esse desafio e pode aumentar o acesso a esses dados.

Segundo Roche (2015), mesmo periódicos que apresentem uma forte política de acesso aberto apresentam inconsistências. Ao pesquisar em um conjunto de 100 trabalhos associados à ecologia evolutiva, 56% deles possuíam dados incompletos e 64% foram arquivados de forma a impossibilitar a reutilização parcial ou completa dos dados. Para lidar com estas necessidades e barreiras seria interessante, então, explorar as potenciais vantagens diretas obtidas por pesquisadores que adotam boas práticas de compartilhamento de dados.

O objetivo deste trabalho é compreender como as iniciativas de ciência aberta trazem impactos para as publicações científicas sobre a ordem Odonata. Além de testar a hipótese de que o número de citações de artigos que possuem dados disponibilizados abertamente na plataforma Dryad seja maior que artigos que não disponibilizam dados abertos.

2 REFERENCIAL TEÓRICO

2.1 Ordem odonata

A ordem dos insetos conhecida como Odonata incluem as libélulas, zigue-zagues ou donzelinhas (como são conhecidas popularmente no Brasil), e são caracterizados por insetos hemimetábolos com adultos terrestre-aéreos e larvas aquáticas, além de seus olhos grandes e multifacetados, dois pares de asas transparentes, corpos alongados e cores usualmente vibrantes.

Esta ordem é dividida em três subordens: Anisoptera, Zygoptera e Anisozygoptera, esta última sendo restrita à região asiática (DIJKSTRA et al., 2013). Estão presentes nos mais variados habitats como corpos d'água, desde rios, riachos e lagos, até poças temporárias ou brejos (CORBET, 1999). Os estudos sobre a Odonata têm implicações importantes para uma variedade de áreas, incluindo ecologia, evolução, comportamento animal e conservação. Por exemplo, os odonatos são frequentemente usados como indicadores de saúde ambiental, visto que são muito sensíveis às mudanças ambientais e alterações em suas populações podem sinalizar mudanças na qualidade dos habitats aquáticos (CARVALHO & NESSIMIAN, 1998; ROSEMBERG & RESH, 1993). Além disso, a pesquisa sobre Odonata ajudou a melhorar a compreensão dos mecanismos de voo de insetos e contribuiu para o desenvolvimento de novas tecnologias para estudar e controlar populações de insetos. Os odonatos são exclusivamente predadores (MUGNAI et al., 2010), os adultos podem se alimentar de insetos considerados prejudiciais ao ser humano, como, por exemplo, *Aedes aegypti* e mosquitos do gênero *Culex*, vetores de agentes patogênicos para dengue e filariose (MANDAL et al., 2008; AKRAM & ALI-KHAN, 2016). Já as ninfas podem se alimentar das larvas dos peixes, gerando prejuízo econômico em tanques de piscicultura (PINTO, 2016).

Apesar de ser um grupo com grande potencial para estudos ecológicos, evolutivos, comportamentais e conservacionistas e ter mais de 6300 espécies já documentadas (BYBEE et al. 2021; CÓRDOBA-AGUILAR, 2008) ainda existem muitas questões sem resposta sobre esta ordem e carência de informação sobre muitas espécies, especialmente em regiões tropicais (DE MARCO & VIANNA, 2005).

2.2 Ciência aberta e acesso aberto

A ciência não é estática, ela está em constante transformação, assim como os processos de apropriação e construção do conhecimento, a iniciativa de abertura da ciência é uma das alternativas utilizadas para contornar o paradigma atual de publicação científica. Conceitualmente, ciência aberta (*Open Science*) refere-se às abordagens de trabalho adotadas ao longo do processo de elaboração, redação, publicação e pós-publicação de artigos científicos que possuem como objetivo geral a maior abertura e transparência de uma ou mais destas etapas (Watson, 2015).

Embora o conceito de ciência aberta exista há décadas, ele ganhou força crescente nos últimos anos com o surgimento da internet e a disponibilidade de ferramentas digitais para compartilhamento e colaboração. De acordo com Stiglitz (1999), o conhecimento científico deve ser considerado um bem público global, uma vez que sua disponibilidade para outros membros da sociedade não é afetada em função de seu uso e não é possível impedir uma pessoa de usufruir daquele bem.

Kuramoto (2006) e Rodrigues (2004), destacam que os pesquisadores muitas vezes realizam seus estudos utilizando recursos públicos fornecidos por organizações de fomento, através de bolsas de pesquisa ou mesmo projetos de financiamento da organização onde trabalham. A questão é que os periódicos científicos recebem gratuitamente os resultados destes estudos (além do trabalho de editoração e revisão, quase sempre feitos de forma voluntária por cientistas) — que acabam sendo produzidos na forma de artigos e revendidos para as instituições de pesquisa.

Um dos principais pontos do movimento da ciência aberta é a ideia de que a pesquisa com financiamento público deve ser disponibilizada ao público (Piwowar et al., 2013). Isso significa que artigos de pesquisa, dados e outros materiais devem ser disponibilizados gratuitamente on-line, permitindo que qualquer pessoa possa acessá-los e usá-los. Isto inclui a disponibilização de artigos e dados de pesquisa gratuitamente online, permitindo práticas de pesquisa transparentes e reproduzíveis (BOECKHOUT; ZIELHUIS; BREDENOORD, 2018) e incentivando a colaboração entre disciplinas científicas, ajudando a reduzir o problema de fraude e viés científico (Fanelli, 2009) e melhorar a confiabilidade e a credibilidade das descobertas científicas.

O acesso aberto (*Open Access*) surgiu como uma resposta às políticas de acesso restrito e de pagamento de taxas pelo acesso a conteúdo científico, que

restringiam o acesso e o uso de informação científica pelos pesquisadores e pela sociedade em geral (Suber, 2012). O movimento do acesso aberto promove a disponibilização de informação científica de forma acessível e gratuita, visando promover o avanço do conhecimento e o desenvolvimento da pesquisa (Harnad, 2001).

Os dados abertos (*Open Data*) são dados científicos disponibilizados de forma acessível e gratuita para cientistas e público em geral, sem restrições de acesso ou de uso. A prática de disponibilizar dados abertos visa promover a transparência e a reprodutibilidade da pesquisa científica, além de também incentivar a colaboração entre pesquisadores e disciplinas científicas.

2.3 Bibliometria

A bibliometria é uma técnica que se utiliza de métodos quantitativos e estatísticos para avaliar a produção e disseminação do conhecimento científico. De acordo com Araújo (2006), a bibliometria pode ser definida como uma técnica que permite medir índices de produção e disseminação de conhecimento científico. Inicialmente, o objetivo da bibliometria era quantificar as características de livros, como o número de edições ou o espaço ocupado em uma biblioteca. No entanto, com o passar do tempo, a bibliometria expandiu seu alcance para incluir outros formatos de informação, como artigos acadêmicos, teses e dissertações, e até mesmo produção científica encontrada em redes sociais e bases de dados online. Além disso, as técnicas bibliométricas começaram a ser utilizadas para analisar citações e avaliar a produção dos pesquisadores. O objetivo dessas medições é contribuir com novas informações para o corpo de conhecimento existente e, assim, promover o avanço da pesquisa em várias áreas do conhecimento científico.

A análise bibliométrica tem uma série de implicações para a comunidade científica. Além de ser útil para avaliar o impacto e a importância relativa de diferentes artigos de pesquisa e periódicos, ajudando os pesquisadores a identificar os trabalhos mais importantes e influentes em seu campo de estudo. Essa técnica também pode ser utilizada por agências e instituições de financiamento para avaliar a produtividade e o impacto dos pesquisadores.

No entanto, é importante considerar que a análise bibliométrica também tem sido alvo de críticas por possuir limitações e vieses. Por exemplo, o uso exclusivo de citações como métrica de impacto pode levar à subestimação de trabalhos de

pesquisa que não são amplamente citados, mas importantes para o avanço do conhecimento em uma determinada área (Piwowar et al., 2018). Além disso, a análise bibliométrica também pode ser influenciada por fatores como o idioma e o país de publicação, o que pode resultar em vieses em relação à visibilidade e impacto de trabalhos de pesquisa.

A utilização de índices bibliométricos para avaliação da produção científica pode ter diversos efeitos na comunidade acadêmica (Bordons & Zulueta, 1999). Um deles é o fenômeno do Publish or Perish, que se refere à pressão que existe para publicar artigos científicos para prosperar na carreira acadêmica (Bornmann & Daniel, 2008). Isso pode levar ao desenvolvimento de um grande número de artigos, embora nem sempre de alta qualidade, visando o reconhecimento de pares e avaliações positivas em relação aos índices de produtividade. Além disso, a análise bibliométrica pode ser afetada por diversos fatores, como o tamanho e a produtividade de um grupo de pesquisa e a disponibilidade de financiamento para pesquisa.

Apesar de suas limitações, a análise bibliométrica continua sendo uma ferramenta importante para avaliar o impacto e a importância da pesquisa científica. Gläser e Laudel (2007) mencionam que os tomadores de decisão têm usado avaliações bibliométricas com mais frequência porque consomem menos tempo e dinheiro quando comparada com revisão por pares. Assim, reduzindo juízos de valor e ajudando os tomadores de decisão a visualizar informações apresentando resultados em gráficos e tabelas.

3 METODOLOGIA

3.1 Banco de artigos

Compilamos um banco de dados com artigos provenientes do repositório digital *Dryad* (<https://datadryad.org/>), obtidos através da busca pela palavra-chave "Odonata". *Dryad* funciona como um repositório online de acesso aberto para dados de pesquisas científicas publicadas, de forma que cada base de dados disponível está associada a um artigo. Desta forma, nosso banco de dados reuniu artigos sobre Odonata com dados abertos disponíveis nesta plataforma. Posteriormente, triamos os artigos manualmente visando eliminar inconsistências.

Para cada artigo encontrado, buscamos identificar em qual periódico este foi publicado, bem como volume e edição. Em posse destas informações, também catalogamos todos os artigos publicados conjuntamente àqueles encontrados na *Dryad*, para que pudéssemos comparar o efeito dos dados e acesso aberto dentro de um mesmo periódico, volume e edição, uma vez que esses artigos que não possuem dados na *Dryad* podem ter ou não acesso aberto. Com isso, os possíveis efeitos da revista ou ano de publicação sobre o número de citações pode ser ignorado, já que as comparações foram realizadas sempre entre artigos publicados em um mesmo número.

3.2 Análise bibliométrica

O número de citações foi escolhido como medida comparativa para avaliar os efeitos da publicação em formato de acesso aberto. Para isso, realizamos pesquisas através do título dos artigos de nosso banco na plataforma indexadora Clarivate Analytics *Web of Science* (<https://www.webofknowledge.com/>), com o intuito de coletar os números de citações de cada artigo. Assumimos zero como número de citação para os artigos indexados na plataforma, mas que não dispunham de número de citação registrado. Artigos que não puderam ser encontrados durante nossas buscas na *Web of Science* foram excluídos do banco de artigos e análises subsequentes. Após a triagem, os dados do banco de artigos foram baixados em formato CSV para serem processados e analisados no software R 4.2.1 (R Core Team, 2021).

Para cada publicação (considerando aqui como publicação um número de uma revista com todos os artigos publicados naquela edição), calculamos o primeiro

quartil, a mediana e o terceiro quartil da quantidade de citações de cada artigo sendo comparado com o “artigo alvo” (isto é, o artigo daquele número com dados publicados na plataforma *Dryad*).

3.3 Teste de hipótese

Para avaliar se o número de citações de artigos com dados publicados na plataforma *Dryad* é maior do que o esperado para os demais artigos publicados no mesmo número da mesma revista, realizamos um teste T pareado comparando o número de citações do artigo alvo com a mediana do número de citações dos artigos para comparação. Antes da realização da análise, verificamos o pressuposto de normalidade dos dados, e os transformamos em escala logarítmica quando necessário.

4 RESULTADOS E DISCUSSÃO

A pesquisa inicial na *Dryad* teve como resultado 44 artigos, contudo, 19 artigos foram descartados por apresentarem inconsistências como: artigos publicados fora de um volume em um periódico (apenas como um artigo único, sem atribuição de um número da revista com outros artigos publicados), e aqueles não indexados na *Web of Science*. Já para os demais artigos catalogados, mas que não possuem dados na *Dryad* somam-se 335 artigos, para os quais foram calculadas as medianas, 1º e 3º quartis do número de citações para cada periódico em que foram publicados (usando apenas artigos do mesmo número e volume que o artigo escolhido) (Tabela 1).

Tabela 1 - Comparativo entre a quantidade de citações de artigos com e sem dados publicados na plataforma Dryad. A coluna "Dryad" indica o número de citações dos artigos encontrados na Dryad, enquanto as colunas "1º Quartil", "Mediana" e "3º Quartil" se referem aos artigos que não têm dados disponíveis na Dryad.

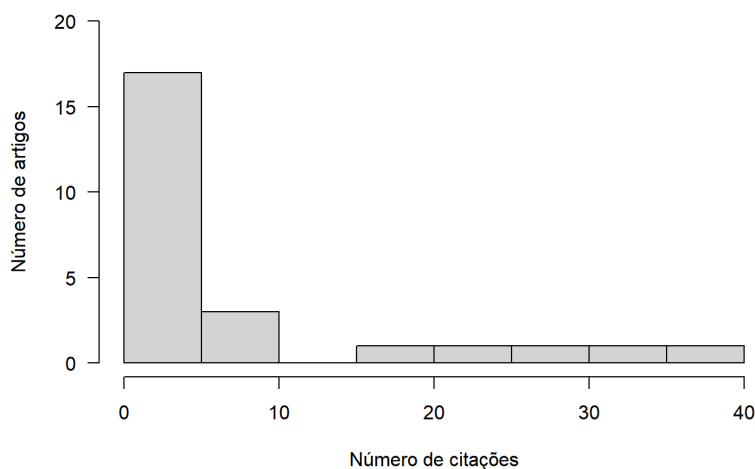
Periódicos	1º Quartil	Mediana	3º Quartil	Dryad
Alcheringa v.44 i.2	1.00	2.0	3.00	1
Arthropod Structure & Development v.47 i.4	4.00	6.0	9.00	25
Biological Journal of Linnean Society v.130 i.3	1.25	3.5	4.75	1
Biological Journal of the Linnean Society v.134 i3	0.00	0.0	1.00	3
Biology Letters v.15 i.5	3.50	5.0	10.50	6
Biology Letters v.17 i.6	1.00	2.0	3.00	5
Biotropica v.52 i.2	2.75	5.5	8.25	3
Biotropica v.53 i.2	0.50	2.0	2.50	0
BMC Ecology and Evolution v.22	0.00	0.0	0.00	0
Diversity and Distributions v.27 i.12	0.00	1.0	1.00	0

Ecology and Evolution v.10 i.14	2.00	3.0	5.50	3
Ecology and Evolution v.10 i.21	1.00	2.0	3.00	3
Ecology and Evolution v.8 i.23	6.00	7.0	15.00	2
Fossil Record v.24 i.2	0.00	0.0	0.50	0
Freshwater Biology v.65 i.4	3.00	5.0	6.00	6
Global Ecology and Biogeography v.30 i.7	0.25	1.0	3.75	8
Historical Biology v.33 i.3	1.50	3.0	4.00	3
Journal of Evolutionary Biology v.31 i.8	4.00	5.0	7.50	2
Molecular Ecology v.25 i.6	20.50	33.0	43.50	36
Molecular Ecology v.26 i.5	9.00	22.0	29.00	20
PNAS v.118 i.44	2.25	3.0	3.75	5
Systematic Biology v.62 i.2	17.50	56.0	83.00	32
Systematic Biology v.71 i.3	0.00	1.0	3.00	3
The Royal Society Open Science v.5 i.6	6.00	7.0	10.00	30
The Royal Society Open Science v.7 i.12	1.00	1.5	2.75	0

Fonte: Elaborado pelo autor (2023)

O número de citações dos artigos com dados publicados na plataforma *Dryad* foi menor ou igual a 5 para a 17 dos trabalhos, com números mais elevados de citações distribuídos nos demais (Figura 1 e Tabela 1).

Figura 1 - Histograma da distribuição do número de citações pelo número de artigos.



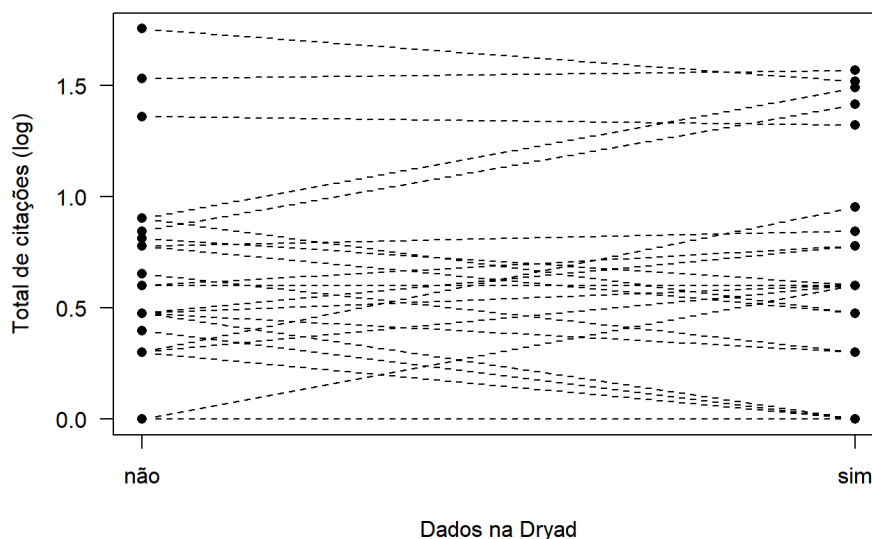
Fonte: Elaborado pelo autor (2023)

Ao examinar o escopo de cada uma das 17 revistas listadas nas quais os artigos foram publicados, fica claro que nenhuma delas é totalmente especializada na ordem Odonata, mas publicam trabalhos de áreas correlatas a estes organismos que envolvam ecologia, genética, taxonomia e afins. Vale ressaltar que existem dois periódicos de grande importância para a pesquisa sobre odonatos: *Odonatologica* e *International Journal of Odonatology*. No entanto, nenhum deles está representado na amostra em questão.

Ao comparar o número de citações dos artigos que possuem dados na *Dryad* com os valores de primeiro e terceiro quartis e mediana do número de citações dos artigos que não possuem dados abertos na *Dryad* foi possível observar que número de citações em 16 dos 25 artigos com dados abertos na *Dryad* foi superior aos valores de 1º quartil de artigos sem dados abertos. Em relação à mediana, esse número cai para 15, e para o 3º quartil apenas 5 possuem número de citação superior àqueles sem dados abertos na *Dryad* publicados no mesmo periódico e volume (Tabela 1).

O teste T pareado realizado para testar a hipótese principal não indicou evidências estatisticamente significativas ($t = -0,503$; $g.l = 24$; $p = 0,62$) de que a publicação de dados de forma aberta na plataforma *Dryad* tenha tido algum tipo de impacto no número de citações nos artigos (Figura 2).

Figura 2 - Resultados do teste t-pareado ($t = -0,503$; $g.l = 24$; $p = 0,62$) comparando o total de citações entre artigos com dados abertos na Dryad e a mediana do total de citações dos demais artigos de um mesmo número da mesma revista.



Fonte: Elaborado pelo autor (2023)

É importante lembrar que os cientistas que trabalham em áreas menos estudadas podem enfrentar mais dificuldades para ter acesso à literatura disponível, fazendo com que esses pesquisadores precisem realizar um esforço extra para ter acesso ao conhecimento existente, seja através do contato com outros autores ou mesmo através de plataformas digitais como o Sci-Hub para fazer o download de artigos e outros materiais. Com isso, os pesquisadores podem terminar por buscar, ler e utilizar boa parte da literatura relacionada aos organismos e abordagens de seu interesse, sem necessariamente privilegiar trabalhos com determinadas características (como dados disponíveis). Quando se busca por trabalhos sobre a ordem Odonata, especialmente relacionados a um grupo, família ou espécies específicas, é provável que o pesquisador acabe lendo tudo o que está disponível sobre o assunto, pois a literatura sobre o tema não é extensa, assim possibilitando sua leitura por completo. Neste sentido, não consideramos que o resultado encontrado em nosso estudo seja uma falha, mas sim uma consequência da escolha de trabalhar com um grupo tão específico como a ordem Odonata para realizar a análise.

O presente estudo apresentou resultados opostos aos de outros estudos anteriores, como Gleditsch, Metelits & Strand (2003), Ioannidis et al. (2009),

Henneken & Accomazzi (2011) e Dorch (2012). Um estudo realizado por Piwowar (2013) observou um aumento de 9% nas citações de artigos que tornaram seus dados disponíveis em repositórios públicos, em comparação a estudos similares que não tornaram esses dados disponíveis. É importante mencionar que o número de variáveis analisadas neste estudo difere de outros estudos, como o de Piwowar (2013), que examinou a data de publicação, o fator de impacto da revista, o status de acesso aberto, o número de autores, o histórico de publicação do primeiro e último autor e o país do autor correspondente. Essas variáveis podem exercer algum impacto no número de citações recebidas por um artigo, e é importante considerá-las quando se comparam os resultados de diferentes estudos. No presente estudo, mesmo sem medir estas variáveis, parte do seu impacto (como fator de impacto da revista e ano de publicação) foi controlado pelo desenho pareado utilizado neste estudo. É possível, entretanto, que outros efeitos não controlados possam ter afetado nossos resultados.

Por outro lado, o impacto dos dados abertos na produção científica pode não ser tão evidente em áreas de estudo bastante específicas, como os estudos sobre Odonata. Desta forma, uma das possíveis razões para o resultado obtido pode ter sido a escolha de restringir o escopo da pesquisa apenas para trabalhos da ordem Odonata. Isso fez com que estivéssemos estudando um ecossistema de pesquisadores bastante pequeno quando comparado com grandes áreas de pesquisa ou grupos de organismos mais estudados, o que pode ter contribuído para que o impacto dos dados abertos não fosse tão evidente. Áreas de pesquisa mais específicas, ou seja, que são menos estudadas em comparação com outras áreas amplamente pesquisadas, como Ecologia ou Biologia da Conservação, geralmente possuem um total de artigos publicados menor. Por outro lado, nas áreas amplamente estudadas, é comum que haja uma grande quantidade de artigos publicados. Isso pode dificultar para os pesquisadores ficarem a par de tudo o que foi publicado, o que os leva a filtrar os artigos que vão ler com base em áreas de seu interesse pessoal, autores, revistas ou particularidades dos artigos. Seguindo este raciocínio, pensamos que o impacto causado por aspectos potencialmente “atraentes” como a disponibilização de dados possa se tornar mais evidente em áreas de pesquisas com maior quantidade de trabalhos publicados, gerando então um possível impacto no número de citações destes trabalhos ou outras métricas potenciais se “sucesso”.

O uso do número de citações como uma métrica para avaliar o impacto de um trabalho científico pode ter influenciado os resultados deste estudo. Embora o número de citações seja um indicador amplamente utilizado para calcular o fator de impacto de uma revista científica, ele desconsidera outras formas e padrões de utilização da pesquisa científica, limitando-se a contar apenas as publicações em veículos tradicionais. Partindo deste ponto, a utilização de métricas alternativas ou "altmetrics" poderia oferecer uma visão mais completa dos dados, avaliando o número de menções individuais para os artigos e onde essas menções ocorreram, cruzando dados de redes sociais como Facebook, Twitter e outras plataformas, bancos de dados online ou documentos de domínio público. Dessa forma, seria possível ampliar a compreensão da influência científica ao avaliar parâmetros específicos do ambiente digital. Além disso, é fundamental considerar que artigos científicos podem ter impactos não mensuráveis, mas que sejam importantes. Por exemplo, um artigo específico pode ser uma peça fundamental para o início de outra pesquisa ou projeto, sem que isso transpareça em seu número de citações. E de maneira inversa ao que se espera que o número de citações costume representar, um artigo considerado "problemático" poderia receber um número alto de citações que sejam mais críticas, por exemplo (TODD; LADLE, 2008). Desta forma, consideramos que investigar o potencial impacto dos dados abertos e de outras práticas abertas também deva ser feito por outras abordagens.

5 CONCLUSÃO

Em geral, acreditamos que o impacto dos dados abertos na produção científica pode ser menos evidente em áreas de estudo específicas, como os estudos sobre Odonata. Embora, esta visão possa estar relacionada ao escopo da pesquisa foi restrito apenas a trabalhos da ordem Odonata, que possivelmente resultou em um ecossistema de pesquisadores relativamente pequeno. Além disso, destacamos que o uso do número de citações como uma métrica para avaliar o impacto de um trabalho científico pode ter influenciado os resultados do estudo, pois desconsidera outras formas de utilização da pesquisa científica. A utilização de métricas alternativas, como altmetrics, pode oferecer uma perspectiva diferente destes dados. Uma possível solução para os problemas levantados seria ampliar o escopo da pesquisa para incluir outras áreas de estudo e grupos de pesquisadores, para que se tenha uma amostra maior e mais diversa. Além disso, é importante considerar o uso de métricas alternativas, para avaliar o impacto dos dados abertos.

REFERÊNCIAS

- AKRAM, W.; ALI-KHAN, H. A. Odonate Nymphs: Generalist Predators and Their Potential in the Management of Dengue Mosquito, *Aedes aegypti* (Diptera: Culicidae). **Journal of Arthropod-Borne Diseases**, v. 10, n. 2, p. 252–257, 1 jun. 2016.
- ARAÚJO, C. A. A. Bibliometria: evolução histórica e questões atuais. **Em Questão**, v. 12, n. 1, p. 11-32, 2006.
- BAKER, M. 1,500 scientists lift the lid on reproducibility. **Nature**, v. 533, n. 7604, p. 452–454, maio 2016.
- BOECKHOUT, M.; ZIELHUIS, G. A.; BREDENOORD, A. L. The FAIR guiding principles for data stewardship: fair enough? **European Journal of Human Genetics**, v. 26, n. 7, p. 931–936, 17 maio 2018.
- BORDONS, M.; ZULUETA, M. A. Evaluación de la actividad científica a través de indicadores bibliométricos. **Revista Española de Cardiología**, v.52, n.10, p.790-800, 1999.
- BORNMANN, L.; DANIEL, H. What do citation counts measure? A review of studies on citing behavior. **Journal of Documentation**, v. 64, n. 1, p. 45–80, 18 jan. 2008.
- BYBEE, S. M. et al. Phylogeny and classification of Odonata using targeted genomics. **Molecular Phylogenetics and Evolution**, v. 160, p. 107115, 1 jul. 2021.
- CARVALHO, A. L.; NESSIMIAN, J. L. Odonata do Estado do Rio de Janeiro, Brasil: Hábitats e hábitos das larvas, p. 3-28. In: NESSIMIAN, J. L.; CARVALHO, A. L. Ecologia de Insetos Aquáticos. **Series Oecologia Brasiliensis**, Rio de Janeiro, v. 5, 309 p. 1998.
- CORBET, P. S. 1999. **Dragonflies - Behaviour and Ecology of Odonata**. Harley Books, Colchester, 829 p.
- CÓRDOBA-AGUILAR, A. **Dragonflies and Damselflies**. [s.l.] OUP Oxford, 2008.
- DE MARCO, P., VIANNA, D.M. 2005. Distribuição do esforço de coleta de Odonata no Brasil - Subsídios para escolha de áreas prioritárias para levantamentos faunísticos. **Lundiana**. p.13–26.
- DIJKSTRA, K.-D. B. et al. The classification and diversity of dragonflies and damselflies (Odonata). **Zootaxa**, v. 3703, n. 1, p. 36, 30 ago. 2013.
- DORCH, B. On the Citation Advantage of linking to data. Disponível em: <<https://hal-hprints.archives-ouvertes.fr/hprints-00714715>>. Acesso em: 20 dez. 2022.
- FANELLI, D. How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data. **PLoS ONE**, v. 4, n. 5, p. e5738, 29 maio 2009.

FERNÁNDEZ-JURICIC, E. Why sharing data and code during peer review can enhance behavioral ecology research. **Behavioral Ecology and Sociobiology**, v. 75, n. 7, 25 jun. 2021.

GLÄSER, J.; LAUDEL, G. The social construction of bibliometric evaluation. In: WHITLEY, R.; GLÄSER, J. **The changing governance of the sciences: the advent of research evaluation systems**. Dordrecht, Springer, 2007. p.101-126.

GLEDITSCH, N. P.; METELITS, C.; STRAND, H. 2003. Posting your data: will you be scooped or will you be famous? **International Studies Perspectives** 4(1):89–97.

HAMPTON, S. E. et al. The Tao of open science for ecology. **Ecosphere**, v. 6, n. 7, p. art120, jul. 2015.

HARNAD, S. The self-archiving initiative. **Nature**, v. 410, n. 6832, p. 1024–1025, abr. 2001.

HELD, L.; SCHWAB, S. Improving the reproducibility of science. **Significance**, v. 17, n. 1, p. 10–11, 29 jan. 2020.

HENNEKEN, E. A.; ACCOMAZZI, A. Linking to Data - Effect on Citation Rates in Astronomy. arXiv:1111.3618 [astro-ph], 15 nov. 2011.

IOANNIDIS, J. P. A. Why Most Published Research Findings Are False. **PLoS Medicine**, v. 2, n. 8, p. e124, 30 ago. 2005.

IOANNIDIS, J. P. A. et al. Repeatability of published microarray gene expression analyses. **Nature Genetics**, v. 41, n. 2, p. 149–155, 28 jan. 2009.

MANDAL, S. K. et al. Biocontrol efficiency of odonate nymphs against larvae of the mosquito, *Culex quinquefasciatus* Say, 1823. **Acta Tropica**, v. 106, n. 2, p. 109–114, maio 2008.

MCKIERNAN, E. C. et al. How open science helps researchers succeed. **eLife**, v. 5, 7 jul. 2016.

MUGNAI, R.; NESSIMIAN, J. L. & BAPTISTA, D. F. 2010. **Manual de identificação de macroinvertebrados aquáticos do Estado do Rio de Janeiro**. Rio de Janeiro, Technical Books. 176p.

NATIONAL ACADEMIES OF SCIENCES, ENGINEERING, AND MEDICINE. **Reproducibility and Replicability in Science**. Washington, D.C.: National Academies Press, 2019.

NORRIS, M.; OPPENHEIM, C.; ROWLAND, F. The citation advantage of open-access articles. **Journal of the American Society for Information Science and Technology**, v. 59, n. 12, p. 1963–1972, out. 2008.

NOSEK, B. A.; BAR-ANAN, Y. Scientific Utopia: I. Opening Scientific Communication. **Psychological Inquiry**, v. 23, n. 3, p. 217–243, 2012.

PINTO, A. P. A fauna de libélulas da América do Sul: a última fronteira a ser desvendada. **Informativo Sociedade Brasileira de Zoologia**, p. 07-09, 2016.

PIWOWAR, H. A.; DAY, R. S.; FRIDSMA, D. B. Sharing Detailed Research Data Is Associated with Increased Citation Rate. **PLoS ONE**, v. 2, n. 3, p. e308, 21 mar. 2007.

PIWOWAR, H. A.; VISION, T. J. Data reuse and the open data citation advantage. **PeerJ**, v. 1, p. e175, 1 out. 2013.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

ROCHE, D. G. et al. Public Data Archiving in Ecology and Evolution: How Well Are We Doing? **PLOS Biology**, v. 13, n. 11, p. e1002295, 10 nov. 2015.

ROSENBERG, D. M.; RESH, V. H. Introduction to freshwater biomonitoring and benthic macroinvertebrates. In: **Freshwater Biomonitoring and Benthic Macroinvertebrates**. New York: Chapman Hall, 1993. p. 1-9.

SUBER, P. **Open access**. Cambridge, Mass.: Mit Press, 2012.

WATSON, M. When will “open science” become simply “science”? **Genome Biology**, v. 16, n. 1, 19 maio 2015.