DADOS
dados.iesp.uerj.br

[Editorial]

Replicabilidade em DADOS: conquistas, limites e desafios*

Bruno Marques Schaefer[1]

[1]Editor de Replicabilidade de DADOS e Professor de Ciência Política do Instituto de Estudos Sociais e Políticos da Universidade do Estado do Rio de Janeiro (IESP-UERJ), Rio de Janeiro, RJ, Brasil
E-mail: brunoschaefer@iesp.uerj.br ORCID: https://orcid.org/0000-0002-5127-7240

Carlos Freitas[2]

[2]Assistente de Bases de Dados de DADOS e doutorando em Ciência Política pelo Instituto de Estudos Sociais e Políticos da Universidade do Estado do Rio de Janeiro (IESP-UERJ), Rio de Janeiro, RJ, Brasil
E-mail: carlosfreitas@iesp.uerj.br ORCID: https://orcid.org/0000-0002-2493-8154

Luiz Augusto Campos[3]

[3]Editor-Chefe de DADOS e Professor de Sociologia e Ciência Política do Instituto de Estudos Sociais e Políticos da Universidade do Estado do Rio de Janeiro (IESP-UERJ), Rio de Janeiro, RJ, Brasil
E-mail: lascampos@iesp.uerj.br ORCID: https://orcid.org/0000-0003-2153-547X

Marcia Rangel Candido[4]

[4]Editora Adjunta de DADOS e Investigadora Integrada do Centro de Investigação e Estudos de Sociologia do Instituto Universitário de Lisboa (CIES-Iscte), Lisboa, Portugal
E-mail: marcia.candido@iscte-iul.pt ORCID: https://orcid.org/0000-0003-3466-000X

DOI: https://doi.org/10.1590/dados.2026.69.1.000

Dataverse: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/JRKZR6

CC BY

*Gostaríamos de agradecer a Elizabeth Lissovsky pela revisão do manuscrito e a Luana Guedes pela revisão e tradução.

Resumo

Replicabilidade em DADOS: conquistas, limites e desafios

O editorial avalia a implementação de políticas de transparência, replicabilidade e reprodutibilidade na revista *DADOS*. O objetivo é analisar os resultados e desafios da criação, em 2023, de uma Editoria de Replicabilidade dedicada a verificar bases de dados, códigos e materiais empíricos dos artigos aprovados. A análise combina discussão conceitual sobre replicabilidade e reprodutibilidade nas ciências sociais com dados editoriais da revista e das bases disponibilizadas no seu repositório. Os resultados indicam avanços, mas também revelam desafios. Diagnósticos anteriores mostravam que cerca de 5% dos artigos quantitativos da ciência política brasileira eram plenamente replicáveis. Na experiência recente da revista, todos os artigos quantitativos (dos últimos anos) puderam ser reproduzidos, mas apenas dois estudos qualitativos disponibilizaram materiais de pesquisa. Em termos de gênero, homens predominam entre os primeiros autores, embora as taxas de compartilhamento sejam semelhantes para ambos os sexos. Ao termo, apontamos alguns desafios e potenciais soluções para o debate em torno desses temas nas ciências sociais.

Palavras-chave: dados; transparência; replicabilidade; reprodutibilidade.

Abstract

Replicability at DADOS: Achievements, Limitations, and Challenges

This editorial assesses the implementation of transparency, replicability, and reproducibility policies at the journal *DADOS*. The aim is to analyze the outcomes and challenges of establishing, in 2023, a dedicated Replicability Editorial Board tasked with verifying the datasets, code, and empirical materials of accepted articles. The analysis combines a conceptual discussion of replicability and reproducibility in the social sciences with editorial data from the journal and the materials deposited in its repository. The findings indicate progress, but also reveal persistent challenges. Prior assessments had shown that roughly 5% of quantitative articles in Brazilian political science were fully replicable. In the journal's recent experience, all quantitative articles published in the last few years could be reproduced; however, only two qualitative studies made their research materials available. In terms of gender, men predominate among first authors, although sharing rates are similar across sexes. In closing, we outline several challenges and potential solutions for the broader social science debate on these issues.

Keywords: data; transparency; replicability; reproducibility.

Résumé

Réplicabilité chez DADOS : acquis, limites et défis

Cet éditorial évalue la mise en œuvre de politiques de transparence, de reproductibilité et de réplicabilité au sein de la revue DADOS. L'objectif est d'analyser les résultats et les défis liés à la création, en 2023, d'un Comité éditorial de réplicabilité chargé de vérifier les bases de données, les codes et les matériaux empiriques des articles acceptés. L'analyse combine une discussion conceptuelle sur la réplicabilité et la reproductibilité dans les sciences sociales avec des données éditoriales de la revue et des bases mises à disposition dans son dépôt. Les résultats indiquent des progrès, mais révèlent également des défis persistants. Des diagnostics antérieurs montraient qu'environ 5 % des articles quantitatifs de la science politique brésilienne étaient pleinement réplicables. Dans l'expérience récente de la revue, tous les articles quantitatifs (des dernières années) ont pu être reproduits, mais seulement deux études qualitatives ont mis à disposition leurs matériaux de recherche. En termes de genre, les hommes prédominent parmi les premiers auteurs, bien que les taux de partage soient similaires pour les deux sexes. En conclusion, nous soulignons quelques défis et solutions potentielles pour le débat autour de ces thématiques dans les sciences sociales.

Mots-clés : données ; transparence ; réplicabilité ; reproductibilité.

Resumen

Replicabilidad en DADOS: logros, límites y desafíos

Este editorial evalúa la implementación de políticas de transparencia, replicabilidad y reproducibilidad en la revista DADOS. El objetivo es analizar los resultados y desafíos de la creación, en 2023, de un Comité Editorial de Replicabilidad dedicado a verificar las bases de datos, códigos y materiales empíricos de los artículos aprobados. El análisis combina una discusión conceptual sobre replicabilidad y reproducibilidad en las ciencias sociales con datos editoriales de la revista y de las bases disponibles en su repositorio. Los resultados indican avances, pero también revelan desafíos persistentes. Diagnósticos anteriores mostraban que alrededor del 5% de los artículos cuantitativos de la ciencia política brasileña eran plenamente replicables. En la experiencia reciente de la revista, todos los artículos cuantitativos (de los últimos años) pudieron ser reproducidos, pero solo dos estudios cualitativos pusieron a disposición sus materiales de investigación. En términos de género, los hombres predominan entre los primeros autores, aunque las tasas de compartición son similares para ambos sexos. Al final, señalamos algunos desafíos y posibles soluciones para el debate en torno a estos temas en las ciencias sociales.

Palabras clave: datos; transparencia; replicabilidad; reproducibilidad.

Introdução

Um dos projetos mais ousados do chamado Movimento da Ciência Aberta (MCA) é a proposta de tornar transparentes os processos que constituem uma pesquisa. Contra a concepção dominante de ciência, baseada em laboratórios fechados e metodologias herméticas, o MCA propõe que toda a comunidade científica e a sociedade, de modo mais amplo, possam acessar e se apropriar dos processos e resultados do fazer científico. Dessa forma, a transparência dos dados e dos procedimentos de uma pesquisa não só daria mais validade às conclusões delineadas, como também estimularia a participação mais democrática nos debates científicos.

Contudo, há diversos desafios em jogo para que o princípio da transparência avance no mundo da ciência. O acesso a recursos, como programas de computador e infraestruturas gerais de pesquisa, apresenta diferenças substantivas em meio às inúmeras desigualdades que atravessam o mundo da ciência. As legislações nacionais sobre acesso e proteção de dados variam substancialmente. Ademais, os efeitos no debate público de uma maior transparência do conhecimento científico ainda são nebulosos e nem sempre positivos. Como algumas pesquisas vêm mostrando, os chamados movimentos negacionistas da ciência costumam se valer do maior acesso às práticas científicas para colocá-las em xeque (Szwako 2025).

Para além dessas questões, tornou-se particularmente difícil sustentar legitimamente uma concepção fechada de ciência, o que tem colocado os princípios da transparência no centro do mundo editorial. Como parte de sua adesão ao MCA, *DADOS* vem incorporando desde 2019 diferentes práticas editoriais, como a requisição das bases de dados, evidências empíricas e rotinas de análise dos artigos publicados, e a criação, em 2023, de uma Editoria de Replicabilidade especializada em avaliar e promover a transparência, replicabilidade e reprodutibilidade das pesquisas publicadas.

O objetivo deste texto é apresentar uma avaliação geral da experiência da revista e de sua Editoria de Replicabilidade nos últimos anos. Para tal, fazemos uma discussão sobre os conceitos relacionados a esse debate, particularmente sobre a tríade transparência, replicabilidade e reprodutibilidade. Essa discussão emerge a partir do diagnóstico de que haveria no mundo da ciência uma crise de transparência, com efeitos nefastos para a qualidade, a validade e o impacto do que ela produz. Em seguida, apresentamos as diretrizes gerais do trabalho executado em *DADOS*. Depois disso, discutimos alguns números sobre o grau de transparên-

cia dos artigos publicados na revista, desde o crescimento das bases de dados indexadas até as ferramentas computacionais mais utilizadas para construí-las. Na quarta seção, discutimos os desafios para a promoção da transparência das pesquisas em ciências sociais publicadas pela revista para, em seguida, delinear algumas conclusões gerais.

O Debate sobre Transparência, Reprodutibilidade e Replicabilidade

O conceito de replicação, ou replicabilidade, acaba por ser utilizado como sinônimo de reprodução ou transparência. Ambos são igualmente importantes, mas distintos. Para complexificar ainda mais o debate, há que se considerar diferenças na definição desses conceitos em distintas áreas das ciências (Barba 2018; Brodeur et al. 2024; Figueiredo Filho et al. 2019; Ioannidis 2005; Janz 2016; King 1995; Miske et al. 2026; National Academies of Sciences et al. 2019; Tyner et al. 2026). Aqui focamos o debate das ciências sociais, campo em que tais conceitos ainda estão longe de ser estabilizados. Para nossos fins, e em diálogo com a literatura especializada nesses temas, trabalharemos com as seguintes definições:

Transparência é um conceito mais amplo, que envolve comunicar de maneira clara e aberta como os procedimentos da pesquisa foram realizados, como se dá o caminho entre a pergunta de pesquisa e os resultados encontrados. Uma pesquisa é transparente quando procede à: "(...) divulgação completa do desenho da pesquisa, que inclui os métodos usados para coletar e analisar dados, disponibilização de informações brutas e limpas [no caso de banco de dados], além de códigos empregados ao longo do caminho" (Figueiredo Filho et al. 2019)[1]. A definição de Figueiredo et al (2019), apesar de bastante ampla, não esgota o conceito, dado que é focada na pesquisa quantitativa. Pesquisas qualitativas, que não utilizam bancos de dados, e/ou possuem outras concepções epistemológicas sobre como conhecer o mundo social e político, também podem ser transparentes (Domingos, Rocha, e Marciano 2024). A questão-chave aqui é a descrição minuciosa dos procedimentos de coleta e análise dos dados, preferencialmente com a disponibilização de anexos metodológicos, documentos, notas de campo e entrevistas (em áudio, vídeo e/ou transcritas). Voltaremos a esse ponto.

Reprodutibilidade, por seu turno, tem a ver com a possibilidade de se chegar aos mesmos resultados e conclusões a partir do mesmo conjunto de dados (Brodeur et al. 2024; Miske et al. 2026). Ou seja, a reanálise dos

mesmos dados, a partir do mesmo caminho metodológico, possibilita a reafirmação dos resultados. Esse processo usualmente se dá via a disponibilização da base de dados e do código computacional (*script*). Logo, uma pesquisa reprodutível é transparente, mas o contrário não necessariamente é verdadeiro. A dizer, pesquisas transparentes podem não ser reprodutíveis *stricto sensu*, na medida em que não é possível reproduzir a realização de observação participante, entrevistas ou outras estratégias metodológicas qualitativas[2].

Já **Replicação** significa o: "(...) processo pelo qual os achados de um artigo publicado são reanalisados para confirmar, avançar ou desafiar os resultados originais" (Janz 2016)[3]. É o exercício de testar se os resultados de uma pesquisa original serão encontrados novamente utilizando diferentes dados, métodos ou conceitos (Brodeur et al. 2024; Figueiredo Filho et al. 2019; Paranhos et al. 2013). Novamente, a transparência aqui é condição necessária para que tenhamos a possibilidade de replicação, mas o mesmo não se pode dizer da reprodutibilidade.

Em síntese, uma pesquisa reprodutível é replicável; ambas necessariamente são transparentes, mas nem tudo o que é replicável é necessariamente reprodutível. Neste último caso, é possível considerar como replicáveis pesquisas que não disponibilizam forçosamente códigos computacionais, mas são claras e transparentes sobre processos de coleta de dados, codificação do material empírico e estratégias de análise. Este é o caso, por exemplo, de clássicos como a pesquisa sobre os Azande de E. E. Evans-Pritchard (1976), que, mesmo antes da difusão do debate sobre ciência aberta, registrou os procedimentos de sua etnografia de forma suficientemente sistemática para permitir a reconstituição da construção de seu argumento e servir de base para análises posteriores. Ainda que Evans-Pritchard não tenha disponibilizado seu caderno de campo à época da pesquisa, ele possibilitou que outras pesquisas como a de Sharon Hutchinson retomassem esse legado em *Nuer Dilemmas: Coping with Money, War, and the State* (Hutchinson 1996), mostrando como transformações históricas e econômicas alteraram práticas e significados sociais descritos anteriormente.

A preservação de arquivos etnográficos, com cadernos de campo e outros materiais que permitam reconstituir a construção dos argumentos, é, portanto, factível, ainda que mais desafiadora por razões práticas e éticas. Repositórios de dados voltados especificamente para pesquisas qualitativas, como o Qualitative Data Repository, da Syracuse University[4], têm

sido criados para enfrentar esses desafios, que envolvem, entre outros aspectos, a proteção da identidade de participantes, o manejo de informações sensíveis e a definição de diferentes níveis de acesso aos dados.

Tomando de empréstimo os termos da discussão sobre inferência causal (Keele 2015), pode-se dizer que a reprodutibilidade garante a validade interna de uma pesquisa (consigo chegar aos mesmos resultados, com o mesmo conjunto de dados – os achados são consistentes) e a replicabilidade garante a validade externa (consigo chegar aos mesmos resultados com outros conjuntos de dados, métodos ou conceitos – os achados são generalizáveis).

Vale notar, entretanto, que esses conceitos emergiram de um intenso debate sobre a "crise de replicabilidade" nas ciências. Há quase duas décadas esse debate é um espectro que ronda a prática científica, seja nas ciências mais "duras" (química, física, biologia), seja nas ciências sociais (psicologia, economia, ciência política, sociologia, entre outras) (Ioannidis 2005; Piper 2020). O artigo de Ioannidis, *Why Most Published Research Findings Are False*, publicado em 2005, postula que grande parte dos estudos randomizados no campo da epidemiologia tem resultados falsos, dado que não são replicáveis. Esse achado envolveu principalmente a constatação de dois problemas: pesquisas com baixo número de casos analisados e viés de significância estatística. Tratou-se, portanto, da recorrência de levantamentos com poucas evidências empíricas e que "forçaram a barra" para chegar a um p-valor de menos de 5%, critério de confiabilidade. No campo da psicologia, um esforço de replicação de 100 experimentos que pressupunham inferências causais obteve os mesmos resultados em menos da metade deles (Open Science Collaboration 2015).

A crise de replicabilidade pode ser entendida, portanto, como a dificuldade sistemática de reproduzir resultados científicos sob diferentes critérios de validação. Essa dificuldade pode se manifestar de distintas formas: (i) incapacidade de obter novamente significância estatística; (ii) redução substantiva dos tamanhos de efeito; (iii) inconsistência nos sinais ou direções dos efeitos; e (iv) baixa potência estatística para detectar efeitos originalmente reportados. Evidências recentes mostram que mesmo quando replicações são conduzidas com alto poder estatístico e seguindo protocolos rigorosos, os efeitos encontrados tendem a ser significativamente menores do que os originais, sugerindo um padrão de superestimação inicial e regressão à média (Tyner et al. 2026).

Nessa agenda de pesquisa, diferentes trabalhos têm se concentrado em dimensões distintas do problema. O estudo de Miske et al. (2026), por exemplo, volta-se para a consistência interna dos resultados publicados, examinando em que medida achados das ciências sociais (negócios, economia, educação, psicologia, sociologia e ciência política) podem ser reproduzidos a partir dos dados e procedimentos originais. Seus resultados indicam que a disponibilidade de dados ainda é limitada (apenas 24% dos artigos selecionados aleatoriamente em *top journals* das seis áreas) e que, mesmo quando acessíveis, apenas parte dos achados pôde ser reproduzida com exatidão (em torno de 53,6%). Isso sugere que falhas operacionais, como erros de codificação, lacunas na documentação ou ausência de materiais, continuam sendo um obstáculo relevante para a confiabilidade da evidência empírica. A ausência de dados e códigos também é um problema considerável, apesar do cenário estar se modificando na medida em que mais revistas adotam políticas de compartilhamento de dados (Rainey et al. 2025).

Em contraste, o trabalho de Tyner et al. (2026) foca na robustez das afirmações empíricas, avaliando se os mesmos padrões se mantêm quando testados com novos dados. Os resultados apontam que cerca de metade das hipóteses encontra suporte consistente, ao mesmo tempo em que evidenciam uma redução expressiva nos tamanhos de efeito e forte sensibilidade aos critérios utilizados para definir sucesso de replicação. Em conjunto, esses achados indicam que evidências iniciais tendem a ser mais fortes do que aquelas acumuladas posteriormente, reforçando a necessidade de cautela na interpretação de resultados isolados.

Casos empíricos ajudam a ilustrar esse ponto. No estudo sobre a relação entre a intensidade do sorriso e a longevidade, análises subsequentes mostram que o efeito desaparece quando controles adicionais são incorporados (Dufner et al. 2018). De forma semelhante, a hipótese de que ataques de tubarão influenciam resultados eleitorais perde sustentação quando reavaliada com novos dados e especificações alternativas (Fowler e Hall 2018). Esses exemplos sugerem que muitos resultados dependem de escolhas teórico-metodológicas ou contextos específicos, e não necessariamente capturam regularidades mais gerais.

Por fim, é importante ressaltar que divergências entre estudos não devem ser interpretadas automaticamente como evidência de erro ou invalidação de resultados anteriores. Como argumentam Janz e Freese (2019), resultados distintos em estudos de replicação podem decorrer de múltiplas

fontes (diferenças de dados, escolhas metodológicas, contextos empíricos ou mesmo variação aleatória) e nem sempre permitem uma explicação conclusiva. Nesse sentido, classificações binárias como "replicou" ou "não replicou" tendem a simplificar excessivamente um processo que é, por natureza, cumulativo e interpretativamente aberto. Mais do que arbitrar "quem está certo", o valor das replicações está em qualificar o grau de robustez dos achados, explicitar suas condições de validade e orientar agendas futuras de pesquisa. Assim, o debate sobre replicabilidade deve ser entendido menos como um veredicto definitivo sobre estudos individuais e mais como parte de um processo coletivo de refinamento do conhecimento científico, no qual diferentes evidências são continuamente confrontadas, reinterpretadas e integradas.

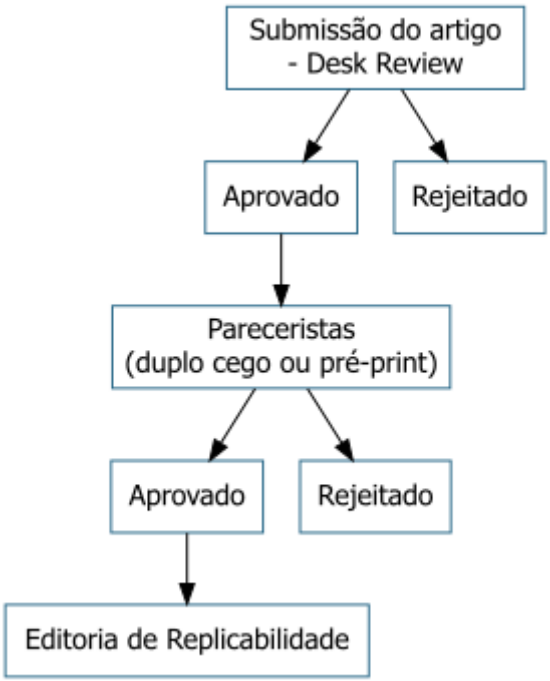
A seguir, especificamos como tratamos desta "crise" em *DADOS*, que teve no artigo de Avelino, Desposato e Mardegan (2021), publicado na própria revista, um primeiro diagnóstico público. Em *Transparency and Replication in Brazilian Political Science: A First Look*, os autores chamavam a atenção da comunidade de cientistas políticos no Brasil para os baixos níveis de transparência e replicabilidade na produção da área. Com base na análise de artigos quantitativos publicados em periódicos da disciplina, os autores mostraram que apenas uma pequena parcela dos estudos disponibilizava dados e códigos suficientes para permitir a reprodução integral das análises. Na prática, foi possível replicar completamente cerca de 5% dos trabalhos examinados[5]. O diagnóstico evidenciava, assim, limitações importantes nas práticas de documentação e compartilhamento de dados na área, reforçando a necessidade de políticas editoriais mais claras voltadas para a transparência e a replicabilidade da pesquisa.

A Editoria de Replicabilidade em *DADOS*

Nesta seção, descrevemos os procedimentos adotados em *DADOS* para a promoção da replicabilidade de seus artigos. Como mostra a Figura 1, os manuscritos submetidos à revista passam por diversos filtros. Em uma primeira etapa, chamada avaliação de *desk review*, eles são avaliados de acordo com sua forma, estilo e adequação às normas e ao escopo da revista. Caso vençam tal etapa, são encaminhados para a análise do conselho de redação, composto por editoras associadas, que fazem uma avaliação mais profunda de seu mérito e sua pertinência ao escopo editorial da revista e indicam pareceristas anônimos para os casos aprovados (Campos e Candido 2021). O processo de avaliação nessa etapa pode ser tanto duplo cego (pareceristas não conhecem autores/as e autores/as não

conhecem pareceristas) quanto parcialmente aberto, quando os manuscritos são previamente submetidos a repositórios de preprints, o que torna a identidade dos/das autores/as disponível aos pareceristas, mas não a dos pareceristas aos/às autores/as (Campos 2021). Este último caso, por sua vez, pode ser divulgado mediante autorização dos pareceristas. Para isso, adotamos o procedimento de perguntar a eles se optam por divulgar suas identidades ou não. Em caso de aprovação após o processo de revisão por pares, o artigo passa pela Editoria de Replicabilidade, incumbida de avaliar se o manuscrito utiliza dados passíveis de algum tipo de replicação.

Figura 1
Processo de avaliação da Editoria de Replicabilidade



Fonte: Os autores.

A Editoria de Replicabilidade é composta por um editor de replicação, atualmente o professor Bruno Schaefer, e um assistente de replicação, posto atualmente ocupado por Carlos Freitas. Uma vez com a lista de artigos aprovados para um número específico da revista, estes passam para a Editoria de Replicabilidade, que inicia o processo de verificação composto pelas etapas que descrevemos a seguir.

A revista, em suas normas, incentiva a submissão de todas as informações pertinentes para garantir o máximo de transparência. No caso de artigos quantitativos, isso envolve todos os bancos de dados utilizados, códigos de análise e/ou descrições que possibilitem a reprodução dos achados. No caso de artigos qualitativos, as entrevistas (em forma de áudio, vídeo ou transcrição), documentos, notas de campo, entre outras informações. Embora incentive o envio dessas informações no momento da submissão do manuscrito, *DADOS* só submete à análise da Editoria de Replicabilidade as bases dos manuscritos aprovados para publicação.

A primeira etapa do trabalho da Editoria de Replicabilidade é a análise prévia do material submetido em relação aos achados reportados no artigo. Para artigos quantitativos, isso significa verificar se os dados e rotinas disponibilizados são suficientes para se chegar aos mesmos valores, gráficos e tabelas que constam do artigo. No caso de artigos qualitativos, verifica-se a existência do material que serve de base para análise (aqui a variedade é ampla, mas alguns exemplos são corpus de textos analisados, documentos históricos, fotografias e transcrições de entrevistas). Caso seja constatado que o material é suficiente, passa-se à próxima etapa do processo. Caso contrário, a Editoria entra em contato com os/as autores/as para solicitação de eventuais complementos.

Após a coleta do material junto aos/às autores/as, o próximo passo é o teste de reprodutibilidade propriamente dito. Esse processo foca artigos e/ou achados de natureza quantitativa e é adaptado ao tipo de informação disponível. Autores/as que documentam o processo de coleta, tratamento e análise dos dados em softwares livres (R, Python) tornam o processo mais fácil, em comparação com o uso de softwares pagos (SPSS, Stata), ou a ausência de qualquer documentação.

A reprodução é feita a partir das instruções, comparando os achados disponíveis no artigo (tabelas, gráficos, figuras) com as informações enviadas pelos/as autores/as. É importante mencionar que se trata aqui, usualmente, de buscar a reprodução dos resultados e não de etapas anteriores da pesquisa, como a própria criação do banco de dados utilizado. Em alguns casos, a documentação enviada não é adequada, por exemplo: ausência de sintaxe no SPSS, de passo a passo no Excel ou TerraView, uso de software point-and-click (Jamovi), o que demanda a reprodução das informações via o uso de outro software (preferencialmente o R), e/ou pedidos aos/às autores/as para envio de informações adicionais.

Esse tipo de pedido é comum durante essa etapa do processo, dado que somente durante os testes é possível ter a noção completa dos materiais e informações necessários. Há dois possíveis resultados dessa etapa: a verificação integral dos achados, ou a constatação de divergências e outros tipos de erros.

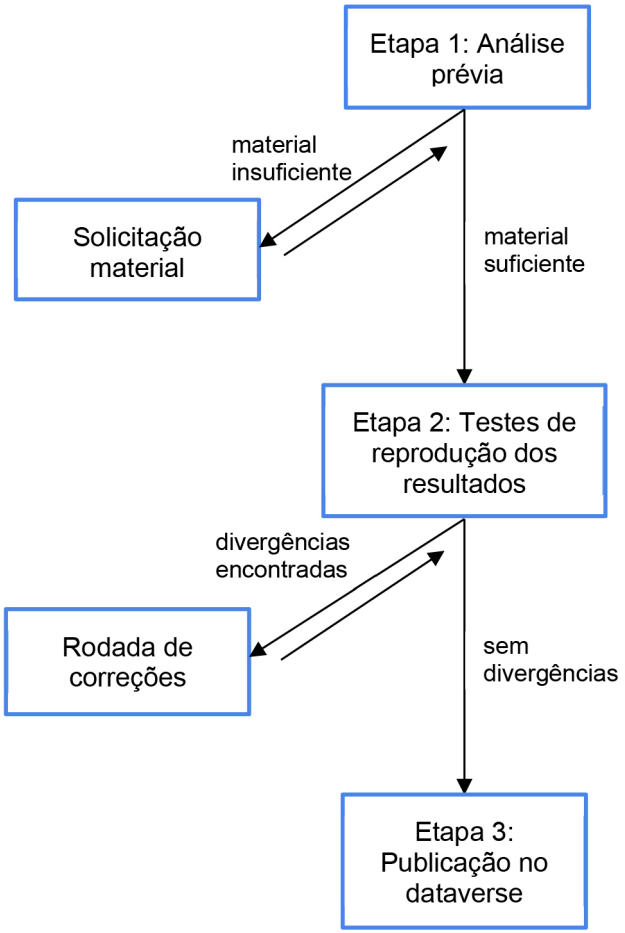
Divergências ocorrem quando os resultados obtidos pela Editoria durante os testes divergem de alguma forma dos resultados que constam do artigo, seja por alguma incongruência nos dados utilizados, ou erro de scripts ou rotinas. Nesses casos, entra-se em contato com os/as autores/as para realização de eventuais correções nos materiais ou no texto do artigo, e uma nova rodada de testes é realizada, até que todas as divergências sejam sanadas. Incluindo todas as etapas do processo, em média, são trocados três e-mails entre a Editoria de Replicabilidade e autores/as (variando de zero, quando é possível reproduzir os achados sem informações adicionais dos/das autores/as, a 24, o máximo até hoje).

Uma vez garantido que os resultados são reprodutíveis, passamos à etapa final, com a disponibilização das informações no Dataverse da Revista. O Dataverse é um repositório online em que os dados e materiais referentes a cada artigo ficam disponíveis gratuitamente para serem acessados, baixados e utilizados por qualquer pesquisador/a. É esse acesso livre e facilitado que efetivamente concretiza os princípios de Ciência Aberta que motivaram a criação da Editoria.

Além dos materiais, o Dataverse ainda inclui as informações básicas dos artigos (também chamadas de metadados), como nome dos/das autores/as, resumo e o número e volume da revista em que foi publicado. Atualmente, os artigos publicados por *DADOS* e seus materiais correspondentes estão hospedados no Dataverse da Universidade de Harvard, o mais rico repositório desse tipo, com cerca de 300 mil conjuntos de dados de publicações do mundo todo, em diversas áreas do conhecimento. Há também o projeto em curso de expandir a disponibilização para o SciELO Data, o repositório para artigos publicados em periódicos da Rede SciELO. O objetivo é que os artigos e os materiais que utilizam tenham o maior alcance possível, e possam servir de inspiração e insumo técnico para o desenvolvimento de novas pesquisas.

O diagrama na figura 2 resume as etapas do processo:

Figura 2:
Etapas do processo de reprodução dos resultados



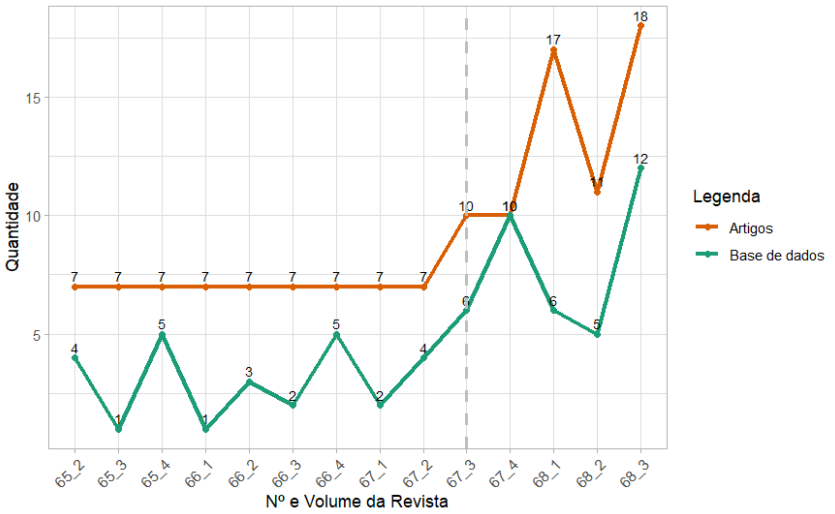
Fonte: Os autores.

Replicabilidade em Números

Desde o estabelecimento do Dataverse da revista em 2021 (antes mesmo da criação da Editoria de Replicabilidade), 129 artigos foram publicados em *DADOS*. Entre eles, 66 possuem algum tipo de informação disponível no repositório. Considerando que a primeira base de dados incluída no Dataverse data de janeiro de 2021 (volume 65, número 2, de 2022 - 65_2) (Reis e Lopes, 2021), este valor indica 51% do total (Gráfico 1).

A Editoria de Replicabilidade em *DADOS* foi instituída em 2023 (volume 67, número 3 - 67_3). De lá para cá, 59 artigos foram analisados. Desses, 34 (57,62%) tinham alguma base de dados ou outro tipo de informação que foi disponibilizada no Dataverse (transcrições de entrevistas, documentos históricos e notícias). Além do crescimento total de artigos publicados, há um crescimento importante do total de artigos com dados disponíveis, conforme vemos no gráfico 1.

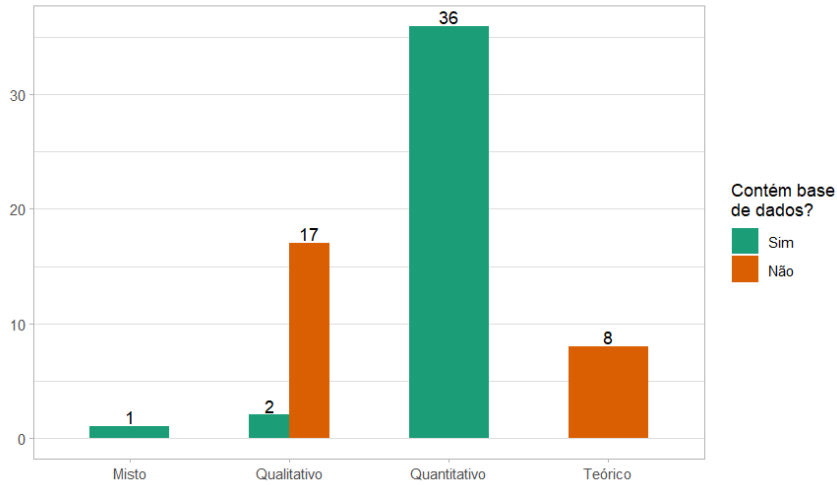
Gráfico 1
Quantidade de artigos publicados e bases de dados indexadas em DADOS



Fonte: Os autores.

No gráfico 2, observamos a relação entre a disponibilização de bases de dados e a metodologia empregada no trabalho. A classificação metodológica foi realizada a partir da leitura do artigo e se divide em quatro tipos: artigos teóricos, quantitativos, qualitativos e mistos. Como apontado por outros trabalhos, há predominância da replicação (como consequência da disponibilidade de dados e rotinas computacionais) entre os artigos quantitativos (Avelino, Desposato, Mardegan 2021; Domingos, Rocha, Marciano 2024; Figueiredo Filho et al. 2019; King 1995; Tuval-Mashiach 2021). Do total de 19 artigos qualitativos aprovados no período, apenas dois disponibilizaram algum tipo de informação sobre a produção dos dados. Por sua vez, todos os artigos quantitativos foram reproduzidos.

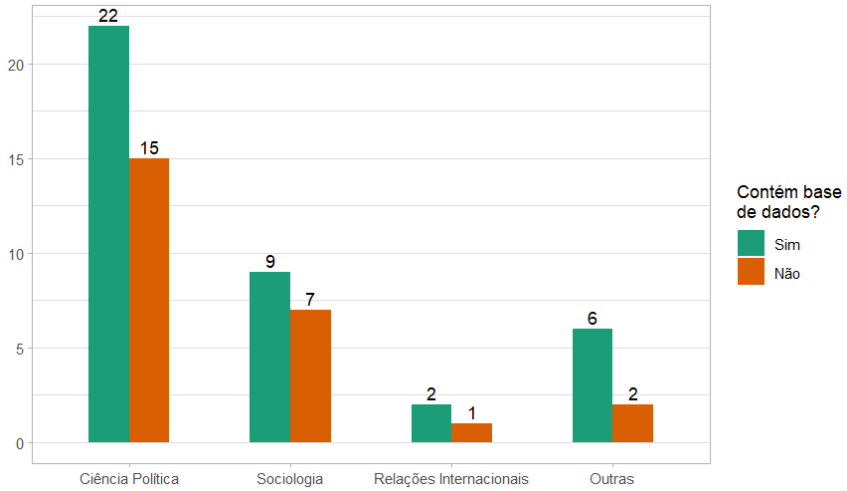
Gráfico 2
Relação entre disponibilização de bases de dados e metodologia



Fonte: Os autores.

Quando comparada a relação entre a disponibilização da base de dados e a área de formação de quem assina a primeira autoria, é interessante notar a predominância da ciência política, seja no total de artigos aprovados (36 de 59), seja na disponibilização das bases (21 de 34). A sociologia está dividida entre disponibilização e não disponibilização, como explícito no gráfico 3. A explicação reside, muito provavelmente, no uso mais intensivo de métodos quantitativos entre cientistas políticos/as, do que entre outros/as cientistas sociais (Figueiredo et al. 2021).

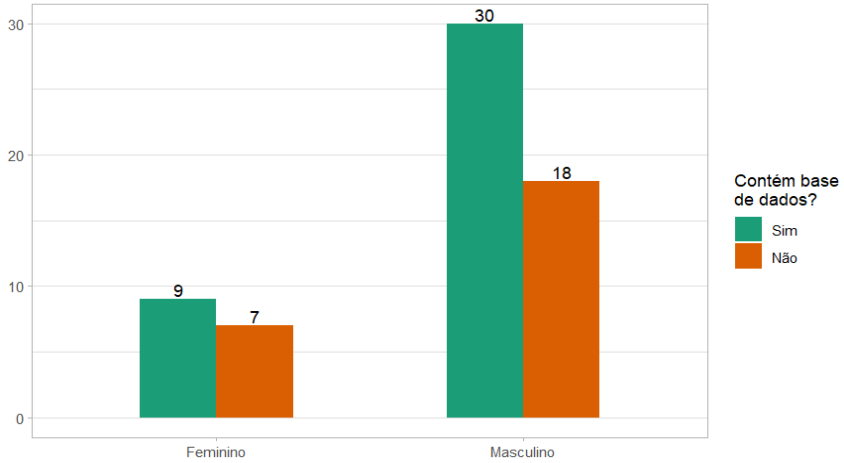
Gráfico 3
Relação entre disponibilização da base de dados e área de formação do/a primeiro/a autor/a



Fonte: Os autores.

Outra dimensão analisada é o gênero do primeiro autor. Os resultados indicam predominância masculina no total de artigos aprovados no período e na disponibilização das bases. O que, novamente, parece ser produto de homens utilizarem mais métodos quantitativos e serem maioria entre cientistas políticos. Ou, invertendo a relação: a ciência política é mais masculina que outras áreas das ciências sociais (Candido, Feres e Campos 2020). A ciência política utiliza mais métodos quantitativos e, com a aplicação de métodos quantitativos, há maior probabilidade de compartilhamento de dados (Pownall 2022). Em termos relativos, porém, não há diferença de gênero. Do total de artigos aprovados com liderança de mulheres, 60% possuem bases disponíveis, enquanto no caso dos homens esse número é de 58%. O problema, portanto, é anterior: na desigualdade de gênero na submissão de artigos (Campos e Candido 2021).

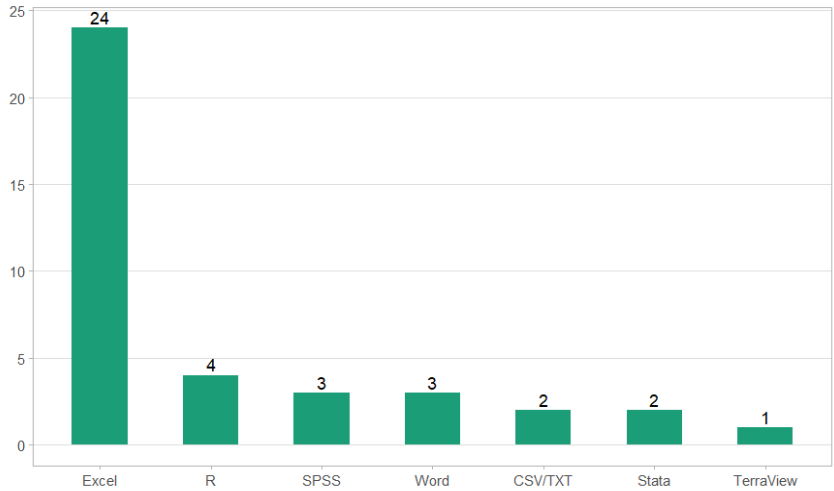
Gráfico 4
Relação entre disponibilização da base de dados e gênero do/a primeiro/a autor/a



Fonte: Os autores.

No gráfico 5 apresentamos os tipos de formatos de bases apresentadas à Revista. O formato mais utilizado é o Microsoft Excel (.xlsx), seguido do R (.RDa ou .RData), Word (.doc) e, então, Stata (.dta), Terra View e TXT (.txt). Excel, .RDa/.RData e .dta são os formatos de base mais compatíveis com os softwares de análise de dados mais utilizados (R e Stata). Outros formatos muitas vezes precisam ser adaptados para passar pelos testes de reprodução, o que pode gerar erros e dificultar o processo.

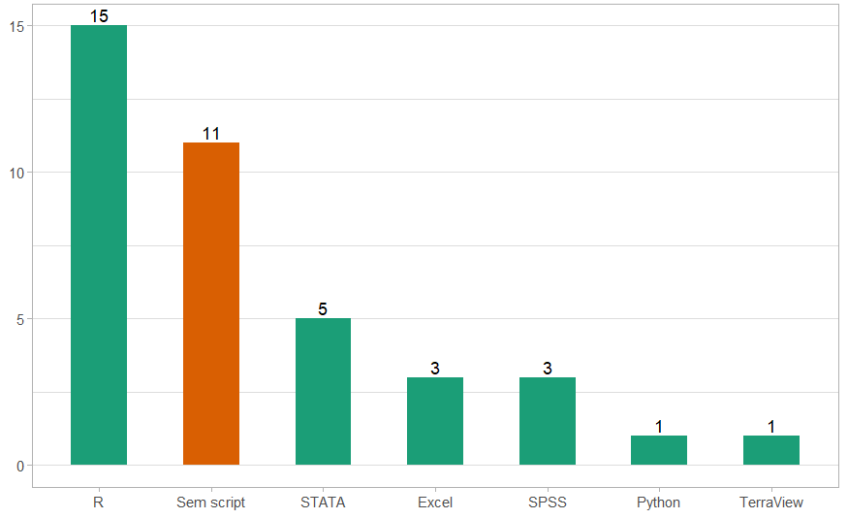
Gráfico 5
Tipos de bases disponibilizadas de acordo com o software utilizado



Fonte: Os autores.

No gráfico 6, apresentamos o tipo de código computacional empregado na análise de dados. Dos 39 artigos que apresentaram bases de dados, 11 (28%) não enviaram nenhum tipo de código que permitisse seguir o passo a passo da análise. Do restante, o código mais comum é feito em R. Softwares mais populares em revistas internacionais, R e Stata, são disponibilizados em conjunto com algum tipo de script. Além de facilitar o processo de replicação, esse tipo de disponibilização é mais útil para fins pedagógicos (Figueiredo Filho et al. 2019; Janz 2016). Bases apresentadas no Word não tinham nenhuma rotina computacional. Isso é produto, por um lado, de que essas bases eram arquivos não diretamente replicáveis (entrevistas e discursos); e, por outro, da simples ausência de qualquer instrução sobre a realização da análise.

Gráfico 6
Tipos de códigos computacionais das bases disponibilizadas



Fonte: Os autores.

Os resultados que apresentamos até aqui ajudam a elucidar como a etapa de profissionalização da editoria com foco em replicabilidade demanda trabalho e tempo. Dentro de processos de avaliação que cada vez mais demandam celeridade dos editores, tendo em vista a frequente demora no retorno de pareceristas, isso impõe um desafio particular na adoção dos melhores procedimentos de ciência aberta. No entanto, esses não são os únicos desafios em jogo. Antes de concluirmos este editorial, abordamos abaixo outras dificuldades para promover transparência.

Desafios

Apesar de todos os números apresentados e da rápida melhora nos parâmetros da Replicabilidade em *DADOS*, os desafios ainda são substanciais. A seguir, discutimos dez deles: 1) o tempo consumido na replicação dos manuscritos aceitos; 2) a aguda assimetria entre as pesquisas quantitativas e qualitativas; 3) o alto número de pesquisas quantitativas que não registram a contento suas rotinas computacionais; 4) os avanços ainda incipientes de reprodutibilidade de dados; 5) o uso ainda baixo das plataformas disponibilizadas pelas ciências sociais; 6) o engajamento dos/as pareceristas no teste das bases de dados; 7) a concentração das práticas de replicabilidade nas pesquisas lideradas por homens; 8) os riscos ligados

à segurança e soberania dos dados; 9) incorporação de ferramentas de inteligência artificial (IA) ao processo; e a 10) reestruturação das práticas de ciência aberta de modo a criar incentivos à disseminação de uma cultura de replicabilidade.

Diferença de tempo entre a aceitação do artigo e a replicação

O processo editorial é demorado por padrão, seja no Brasil ou internacionalmente (Huisman e Smits 2017). O tempo médio entre a submissão e alguma decisão pode variar de seis meses a vários anos. Todo esse processo é frustrante para autores/as e emperra o avanço de debates científicos, além de sobrecarregar o corpo editorial, pareceristas, entre outros envolvidos. No caso da Editoria de Replicabilidade, o trabalho só começa com o artigo aprovado, o que diminui a carga de trabalho, mas pode gerar problemas: o que fazer com um artigo que não é replicável? Como encontrar os dados após o arquivamento dos itens submetidos no sistema da Scielo?

A resposta à primeira pergunta envolve considerações substantivas; a segunda, pontos técnicos. Até agora todos os artigos com bases de dados foram reproduzidos, e, quando os dados foram arquivados, entramos em contato com os/as autores/as para requisitá-los novamente. Ainda assim, esses procedimentos não eliminam completamente o problema. A realização da verificação apenas após a aprovação implica que eventuais inconsistências, lacunas na documentação ou ausência de materiais só sejam identificadas em um estágio avançado do processo editorial, podendo gerar retrabalho, atrasos na publicação e custos adicionais de coordenação entre editoria e autores/as. Em casos mais críticos, coloca-se o dilema entre a publicação de um artigo não plenamente reprodutível e a necessidade de revisão substantiva dos materiais, o que pode tensionar prazos e expectativas.

Uma forma importante de mitigar esse tipo de problema é incentivar a disponibilização prévia dos dados, códigos e materiais de pesquisa em repositórios públicos, ainda nas etapas iniciais do processo editorial. O depósito antecipado reduz a dependência de arquivos submetidos em sistemas fechados e facilita tanto a recuperação quanto a verificação dos materiais ao longo do tempo. Além disso, repositórios amplamente utilizados (como Dataverse, Open Science Framework [OSF], Zenodo,

entre outros) oferecem infraestrutura estável para armazenamento, versionamento e compartilhamento de dados, o que contribui para maior padronização e rastreabilidade das pesquisas.

As recomendações recentes para o depósito de dados indicam, ainda, que diferentes áreas do conhecimento podem recorrer tanto a repositórios especializados quanto a repositórios multidisciplinares, que permitem distintos níveis de acesso (aberto, restrito ou sob embargo), a depender das exigências éticas e legais envolvidas. No caso das ciências humanas e sociais, essa flexibilidade é particularmente relevante, dado que os materiais de pesquisa frequentemente incluem documentos, entrevistas, notas de campo e outros insumos que exigem cuidados adicionais quanto à privacidade e ao consentimento. Ainda assim, mesmo quando o compartilhamento integral não é possível, a disponibilização de descrições detalhadas, protocolos e materiais complementares já representa um avanço significativo em termos de transparência e possibilidade de verificação independente (Scielo 2025).

Assimetrias entre bases quali e quanti

Há grande discrepância na disponibilização de bases de dados (ou outros materiais) entre os artigos quantitativos e qualitativos. Isso se deve, em parte, às próprias diferenças epistemológicas entre metodologias (Domingos et al. 2024; Haverland e Yanow 2012; Silva 2015), como já mencionamos, mas também a considerações de privacidade e transparência.

Nesse período, apenas dois artigos qualitativos aprovados na revista disponibilizaram dados. O primeiro utilizou como método principal o *process tracing* e forneceu informações de notícias de jornais, discursos e outras notas que embasaram as conclusões. O segundo utilizou entrevistas em profundidade e forneceu as transcrições (anônimas) (Do Vale 2023; Fernandes e Silva 2024).

Apesar de concordarmos com as diferenças entre quali e quanti no quesito replicabilidade e reprodutibilidade, bem como com a necessidade de preservação da privacidade de pesquisados/as e pesquisadores/as (norma exigida por Comissões de Ética em pesquisa e legislações sobre proteção de dados), as pesquisas qualitativas não estão imunes aos princípios da transparência. Por conta disso, temos demandado, já no momento de submissão, que os autores expliquem as razões para não disponibilizarem as bases de dados. Além disso, temos buscado gerar mais incentivos

para que pesquisadoras/es qualitativos/as compartilhem documentos complementares e/ou demais informações que auxiliem na verificação de seus resultados. A disponibilização do material pode auxiliar pesquisas futuras, bem como incentivar o reaproveitamento de dados a partir de outras perspectivas teóricas e metodológicas (Small 2011).

Ausência de rotinas computacionais (scripts)

É possível realizar pesquisas quantitativas com bases de dados sem necessariamente registrar formalmente as rotinas de análise. Porém, a ausência de rotinas computacionais torna o processo de replicação mais custoso. A ideia por trás das rotinas é que qualquer pesquisador/a que leia o artigo e acesse os dados consiga reproduzir todo o processo e obter os mesmos resultados. Softwares como R e Stata funcionam a partir de scripts computacionais que, por definição, servem como um passo a passo executável. No entanto, outros softwares não funcionam a partir de *scripts*, como gráficos e tabelas gerados no Excel, por exemplo. Nesses casos, é preciso que os/as autores/as elaborem um documento do tipo "read-me" ou "leia-me", descrevendo textualmente o passo a passo para chegar aos resultados (por exemplo: "primeiro tire a média das colunas A e B, em seguida selecione as colunas C e D e crie um gráfico de barras"). Esse procedimento auxilia o trabalho da Editoria, mas também é um recurso importante para os/as próprios/as pesquisadores/as. Criar um passo a passo metodológico facilita a escrita do trabalho posteriormente (Figueiredo Filho et al. 2019).

Limites da reprodutibilidade e documentação das bases de dados

Um desafio adicional diz respeito aos limites da própria reprodutibilidade. No âmbito da Editoria de Replicabilidade, o foco recai sobre a reprodução dos achados reportados nos artigos, e não sobre a reconstrução integral do processo de geração das bases de dados utilizadas. Esse processo envolve uma série de decisões por parte dos/as autores/as — como critérios de coleta, seleção, limpeza, agregação e codificação — que nem sempre são plenamente documentadas ou passíveis de reconstrução.

Diante disso, recomenda-se que os/as autores/as disponibilizem, sempre que possível, documentação detalhada sobre o processo de construção das bases de dados, incluindo descrição das fontes, critérios de inclusão

e exclusão e etapas de tratamento dos dados. Essa prática contribui para a própria organização da pesquisa, amplia sua transparência, facilita a avaliação independente e potencializa o uso futuro das bases em exercícios de replicação e reanálise (Becker 2020).

O uso ainda limitado das bases disponibilizadas

O Dataverse da revista é uma fonte rica de dados sobre variados temas nas ciências sociais. No entanto, o número de downloads ainda é baixo, indicando que esses materiais permanecem subutilizados. Esse cenário reflete um problema mais amplo da pesquisa (não só no Brasil): a relativa ausência de uma cultura consolidada de replicação e reanálise, em que pesquisadoras e pesquisadores se engajem de forma mais ativa com evidências já produzidas. Em grande medida, isso decorre do fato de que trabalhos de replicação ainda tendem a ser subvalorizados na produção científica, sobretudo quando comparados a estudos que apresentam achados novos ou considerados inovadores (Brodeur et al. 2024; Miske et al. 2026). Contudo, o avanço cumulativo do conhecimento depende precisamente da capacidade de verificar, qualificar e, quando necessário, revisar resultados existentes. Em outras palavras, só é possível avançar de forma consistente se houver clareza sobre aquilo que, de fato, sabemos.

Nesse sentido, a revista *DADOS* busca contribuir para o fortalecimento dessa agenda ao estimular explicitamente a submissão de comentários críticos[6] sobre artigos publicados nos últimos cinco anos. Esses textos podem incluir exercícios de replicação que utilizem novos dados, metodologias alternativas ou diferentes operacionalizações de conceitos para testar afirmações empíricas. Trata-se de um espaço para o debate acadêmico qualificado, orientado não à desqualificação de trabalhos anteriores, mas ao refinamento das evidências e ao avanço do conhecimento. Os comentários devem manter caráter estritamente acadêmico e respeitoso, sendo avaliados pelo mesmo processo de revisão por pares aplicado aos demais artigos, e garantindo aos autores originais o direito de resposta.

A revista e outros periódicos também podem avançar ainda mais nesse esforço, por exemplo, por meio da realização de eventos voltados para a exploração do Dataverse (e de outras bases de dados) e para a disseminação de boas práticas de replicação. Do mesmo modo, é possível incentivar o uso desses dados em disciplinas de graduação e pós-graduação,

bem como em orientações de dissertações e teses, demonstrando que trabalhos originais podem ser produzidos a partir de dados secundários — seja por meio de reanálises, inclusão de novas variáveis ou adoção de especificações alternativas (Korell, Reinecke e Lott 2023). Além de fomentar um ambiente de pesquisa mais ativo e reflexivo, essas iniciativas contribuem para ampliar o alcance, o impacto e a credibilidade das publicações da revista.

Engajamento dos/as pareceristas na replicabilidade

Um dos desafios centrais para o avanço das práticas de replicabilidade diz respeito às diretrizes para pareceristas. Em princípio, seria desejável que a revisão por pares incorporasse também a verificação dos dados e dos códigos, permitindo identificar inconsistências já na etapa de avaliação. No entanto, essa expectativa esbarra em um problema estrutural mais amplo: a crescente sobrecarga do sistema de revisão por pares. Com o aumento contínuo no número de submissões e a ausência de incentivos proporcionais para a atividade de parecerista, observa-se uma dificuldade crescente em recrutar revisores e obter avaliações em tempo hábil, fenômeno frequentemente descrito como "fadiga de pareceristas" (Adam 2025).

Uma alternativa possível seria antecipar a atuação da Editoria de Replicabilidade, deslocando essa verificação para um momento anterior à revisão por pares. Em tese, isso permitiria identificar problemas técnicos antes mesmo da avaliação substantiva do manuscrito. Contudo, na prática, essa solução tende a ser inviável, especialmente no contexto brasileiro, em que grande parte das atividades editoriais é realizada de forma voluntária. A incorporação de uma etapa adicional de checagem técnica antes da revisão ampliaria significativamente a carga de trabalho editorial, tornando difícil sua implementação sistemática.

Nesse contexto, parece pouco realista exigir que pareceristas assumam, de forma sistemática, a revisão detalhada de bases de dados e códigos, sobretudo em um cenário em que já há dificuldades significativas de retorno. Uma alternativa mais factível é a especialização dessa função, com a designação de pareceristas específicos para a verificação técnica, quando possível[7].

Diversidade e replicabilidade

A diversidade é uma preocupação central na editoria de *DADOS*. Temos produzido relatórios frequentes sobre a participação de gênero na revista. O presente editorial deu exemplo disso, pois divulgamos aqui indicadores sociais inéditos sobre a adesão de homens e mulheres à política de dados abertos que adotamos. No entanto, gênero não é a única variável que importa para refletir sobre desigualdades na ciência. Se este aspecto ainda é pouco mensurado nas orientações e pesquisas sobre ciência aberta, elementos como a desigualdade racial nas publicações ou questões de acessibilidade seguem praticamente sem mensuração. Pretendemos continuar a acompanhar o debate sobre inclusão na ciência e tentar incorporar – ou inovar – na discussão dessa pauta quando possível.

Soberania de dados

Outro desafio emergente diz respeito à soberania de dados, especialmente no contexto da crescente dependência de infraestruturas internacionais para o armazenamento e compartilhamento de dados de pesquisa. Embora repositórios amplamente utilizados — como Dataverse, OSF e Zenodo — desempenhem papel fundamental na promoção da ciência aberta, a centralização desses serviços fora do país levanta questões sobre autonomia, segurança e governança das informações científicas. Nesse sentido, iniciativas nacionais ganham relevância, como a transição da *DADOS* para o SciELO Data, que se insere em um esforço mais amplo de fortalecimento de repositórios locais e alinhamento com diretrizes da ciência aberta no Brasil. Documentos recentes indicam, inclusive, a importância de orientar autores na escolha de repositórios adequados, considerando não apenas aspectos técnicos, mas também políticas de acesso, tipos de dados e áreas do conhecimento (Scielo 2025).

Incorporação de ferramentas de IA ao processo

Desde a criação da Editoria, ferramentas de IA foram usadas apenas de forma secundária no processo de replicação — quase exclusivamente para adaptar scripts de código para a linguagem R. No entanto, atualmente, os avanços dos modelos generativos de linguagem (LLMs) e dos agentes de IA se dão num ritmo de meses, não de anos. Se desde a publicação do último número da revista essas ferramentas já se desenvolveram signifi-

cativamente, o que dirá desde a criação da Editoria, em 2023. Isso impõe aos editores a atualização constante para entender como as ferramentas podem ajudar (ou criar riscos) no processo de replicação.

Um possível teste no curto prazo é incorporar o uso dessas ferramentas na etapa da análise prévia dos artigos (Figura 2, Etapa 1), a fim de fazer a primeira verificação dos dados disponibilizados pelos autores, e se são suficientes para reproduzir as análises do artigo. A rápida identificação de arquivos faltantes nessa etapa pode ajudar a agilizar o processo, uma vez que os autores podem ser imediatamente notificados para suprir as lacunas. Outro uso potencial é para uma segunda rodada de testes (Figura 2, Etapa 2), quando a verificação humana não revelar divergências. Isso pode ajudar a detectar inconsistências em dados extensos (por exemplo, tabelas de regressão com muitos resultados) que escaparam à primeira análise.

Sejam quais forem os novos usos, entendemos que a mão humana ainda será necessária para o crivo final em todas as etapas do processo. A IA é rápida, mas comete erros com a mesma confiança que acertos. Assim, usos para a criação de redundâncias e duplicação de checagens, e não para a substituição da avaliação humana, tendem a aumentar a segurança e a confiabilidade do processo.

Criação de uma cultura de replicabilidade

No futuro, será necessário reavaliar a política de replicabilidade da revista. Um aspecto relevante a investigar é se os artigos que disponibilizam dados abertos tendem a receber mais citações (Christensen et al. 2019). Outro ponto igualmente importante diz respeito à institucionalização das práticas de replicação — tanto para sua disseminação em outras revistas quanto para seu aprofundamento na *DADOS*, com o envolvimento de novos colaboradores.

Nesse sentido, um desafio central é alterar os incentivos acadêmicos que atualmente não privilegiam a verificação sistemática de achados existentes como prática científica legítima e importante. A consolidação de uma cultura de replicação passa, portanto, não apenas por diretrizes editoriais, mas também pelo reconhecimento dessas práticas como contribuições científicas. Isso envolve, por exemplo, valorizar publicações de replicação, incorporar essas atividades à formação de estudantes de graduação e pós-graduação e estimular seu uso em projetos de pesquisa e orientações acadêmicas.

Adicionalmente, a expansão dessas práticas depende da construção de rotinas mais padronizadas e da difusão de infraestruturas adequadas para armazenamento e compartilhamento de dados, como os repositórios públicos já mencionados. A experiência da *DADOS* (e de outras revistas brasileiras, como a *Brazilian Political Science Review*) sugere que políticas editoriais podem desempenhar um papel importante nesse processo, mas sua consolidação exige um esforço mais amplo da comunidade científica. Em última instância, trata-se de avançar na direção de um sistema em que a produção de novos conhecimentos esteja necessariamente articulada a verificação, reanálise e acúmulo de evidências previamente existentes.

Considerações finais

O objetivo deste editorial foi avaliar a implementação de políticas de transparência, replicabilidade e reprodutibilidade na revista *DADOS* no contexto do Movimento da Ciência Aberta. Mais do que meramente construir um balanço sobre o nosso trabalho, o intuito aqui foi publicizar os parâmetros e desafios dessa experiência com a comunidade acadêmica, estimulando o debate e a própria visibilidade do tema.

Os resultados apresentados indicam avanços importantes, especialmente no que diz respeito à reprodução de artigos quantitativos e à ampliação da disponibilização de bases de dados. Ao mesmo tempo, eles evidenciam desafios persistentes, como a baixa adesão entre estudos qualitativos às práticas de transparência, a ausência de rotinas computacionais em parte das pesquisas quantitativas, a sobrecarga do sistema de revisão por pares e a ainda limitada utilização dos repositórios disponíveis. Esses elementos mostram que a institucionalização de práticas de ciência aberta é um processo gradual, que depende tanto de mudanças normativas quanto de transformações mais amplas nos incentivos e na cultura acadêmica (Avelino et al. 2021; Figueiredo Filho et al. 2019).

Nesse contexto, a experiência da *DADOS* sugere que periódicos científicos podem desempenhar um papel central na promoção da transparência, não apenas por meio de exigências formais, mas também pela criação de espaços de debate, como os comentários críticos, e pelo estímulo à circulação e reutilização de dados. Ao mesmo tempo, o fortalecimento de infraestruturas nacionais, a consolidação de práticas de replicação e a incorporação dessas atividades na formação de novas gerações de pesquisadoras e pesquisadores são elementos fundamentais para o avanço dessa agenda.

Por fim, cabe reiterar que a transparência, a replicabilidade e a reprodutibilidade não devem ser compreendidas como fins em si mesmas, mas como meios para o aprimoramento contínuo da produção científica (Brodeur et al. 2024; Miske et al. 2026; Open Science Collaboration 2015; Tyner et al. 2026), bem como princípios inseridos na proposta mais ampla de ciência aberta. Em um contexto marcado por crescentes demandas por rigor, credibilidade e impacto social da pesquisa, a construção de um ambiente mais aberto, cumulativo e reflexivo é não apenas desejável, mas necessária para o desenvolvimento das ciências sociais.

Editor responsável: Aldaberto Cardoso.
(Recebido para publicação em 2 de maio de 2026)
(Aprovado para publicação em 1 de julho de 2026)

Notas

1. Todos os trechos em inglês foram traduzidos.

2. Por vezes, o termo reprodutibilidade é substituído por replicação direta (Open Science Collaboration 2015): "Direct replication is the attempt to recreate the conditions believed sufficient for obtaining a previously observed finding and is the means of establishing reproducibility of a finding with new data".

3. Tradução dos autores. No original: "Process by which a published article's findings are reanalyzed to confirm, advance or challenge the original results" (Janz, 2016, p. 3).

4. Disponível em: https://qdr.syracuse.edu/. Acesso em: 6 mar. 2026.

5. Do total de 650 artigos publicados entre 2012 e 2016 nos cinco principais periódicos da área (Brazilian Political Science Review, Revista de Ciência Política, Dados – Revista de Ciências Sociais, Opinião Pública e Revista Brasileira de Ciências Sociais), 197 reportaram achados quantitativos. Desses, os autores só conseguiram recuperar 28% dos dados e códigos (n = 56). A partir desse conjunto, foi possível tentar replicar apenas 14% dos estudos (n = 28), sendo que a reprodução bem-sucedida dos resultados ocorreu em cerca de 5% do total de artigos quantitativos analisados (n = 10).

6. Desde 2022, *DADOS* incentiva a submissão de comentários críticos (ou réplicas) aos textos que publica. Conforme nossas normas, comentários críticos são "textos que formulam avaliações críticas de manuscritos publicados pela própria revista nos últimos cinco anos, nos quais resultados de pesquisa são debatidas de modo qualificado e cortês com o objetivo único de fazer avançar um determinado debate acadêmico. Mais informações em nossas normas técnicas: https://www.scielo.br/journal/dados/about/#instructions

7. Outro ponto interessante é a incorporação de ferramentas de IA e Modelos de Linguagem (Large Language Models - LLM) no processo de revisão. Uma iniciativa interessante é o Proof Patrol for Empirical Econ (GPT), criado por Danilo Santa Cruz Coelho (2025).

Referências

Adam, Daniel. 2025. “The peer-review crisis: how to fix an overloaded system”. https://www.nature.com/articles/d41586-025-02457-2.

Avelino, George, Scott Desposato, e Ivan Mardegan. 2021. “Transparency and Replication in Brazilian Political Science: A First Look”. *Dados* 64:e20190304. doi:10.48550/arXiv.1802.03311.

Becker, Megan. 2020. “Qualitative Replication as a Pedagogical Approach to Teaching Research Methods”. *PS: Political Science & Politics* 53(4):802–6. doi:10.1017/S1049096520000864.

Brodeur, Abel, Kevin Esterling, Jörg Ankel-Peters, Natália S. Bueno, Scott Desposato, Anna Dreber, Federica Genovese, Donald P. Green, Matthew Hepplewhite, Fernando Hoces de la Guardia, Magnus Johannesson, Andreas Kotsadam, Edward Miguel, Yamil R. Velez, e Lauren Young. 2024. “Promoting Reproducibility and Replicability in Political Science”. *Research & Politics* 11(1):20531680241233439. doi:10.1177/20531680241233439.

Campos, Luiz Augusto. 2021. “O que são pré-prints?”. *Blog DADOS,* 10 de maio, Disponível em: https://dados.iesp.uerj.br/es/o-que-sao-preprints.

Campos, Luiz Augusto, e Marcia Rangel Candido. 2021. “Transparência em DADOS: submissões, pareceristas e diversidade no fluxo editorial dos últimos anos *”. *Dados* 65:e20220000. doi:https://doi.org/10.1590/dados.2022.65.1.000.

Candido, Marcia Rangel, João Feres, e Luiz Augusto Campos. 2020. “Desigualdades na Elite da Ciência Política Brasileira”. *Civitas - Revista de Ciências Sociais* 19:564–82. doi:10.15448/1984-7289.2019.3.33488.

Christensen, Garret, Allan Dafoe, Edward Miguel, Don A. Moore, e Andrew K. Rose. 2019. “A Study of the Impact of Data Sharing on Article Citations Using Journal Policies as a Natural Experiment”. *PLOS ONE* 14(12):e0225883.

Coelho, Danilo Santa Cruz. 2025. “ChatGPT - Proof Patrol for Empirical Econ (GPT)”. https://chatgpt.com/pt-BR/.

Do Vale, Helder Ferreira. 2023. *Replication Data for: “When Time Matters: Federal Changes, Temporality and Recentralization in Brazil”*: Versão 1.1. Harvard Dataverse. doi:10.7910/DVN/HKSTMU.

Domingos, Amanda, Virginia Rocha, e Palloma Marciano. 2024. “A Estrada dos Tijolos Amarelos: desafios e sugestões para produzir pesquisas qualitativas mais transparentes”. *Revista Brasileira de Ciência Política* 43:e275124. doi:10.1590/0103-3352.2024.43.275124.

Dufner, Michael et al. 2018. “Does Smile Intensity in Photographs Really Predict Longevity? A Replication and Extension of Abel and Kruger (2010)”. *Psychological Science* 29(1):147–53. doi:10.1177/0956797617734315.

Evans-Pritchard, E. E. 1976. *Witchcraft, Oracles and Magic Among the Azande*. Oxford: Oxford University Press.

Fernandes, Eduardo Georjão, e Marcelo Kunrath Silva. 2024. *Replication Data For: Controlling Protest: Repertoire, Strategies, and Lessons from Policing the 2013-2014 Protests Cycle*: Versão 1.2. Harvard Dataverse. doi:10.7910/DVN/2J3Q2N.

Figueiredo Filho, Dalson et al. 2019. “Seven Reasons Why: A User’s Guide to Transparency and Reproducibility”. *Brazilian Political Science Review* 13:e0001. doi:10.1590/1981-3821201900020001.

Figueiredo, Dalson, Antônio Fernandes, Lucas Borba, e Thaís Helena Aguiar. 2021. "Metodologias de Pesquisa em Ciência Política: uma Breve Introdução". *BIB - Revista Brasileira de Informação Bibliográfica em Ciências Sociais* (94). https://bibanpocs.emnuvens.com.br/revista/article/view/175.

Fowler, Anthony, e Andrew B. Hall. 2018. "Do Shark Attacks Influence Presidential Elections? Reassessing a Prominent Finding on Voter Competence". *The Journal of Politics* 80(4):1423–37. doi:10.1086/699244.

Haverland, Markus, e Dvora Yanow. 2012. "A Hitchhiker's Guide to the Public Administration Research Universe: Surviving Conversations on Methodologies and Methods". *Public Administration Review* 72(3):401–8. doi:10.1111/j.1540-6210.2011.02524.x.

Huisman, Janine, e Jeroen Smits. 2017. "Duration and Quality of the Peer Review Process: The Author's Perspective". *Scientometrics* 113(1):633–50. doi:10.1007/s11192-017-2310-5.

Hutchinson, Sharon E. 1996. *Nuer Dilemmas: Coping with Money, War, and the State*. Berkeley: University of California Press.

Ioannidis, John P. A. 2005. "Why Most Published Research Findings Are False". *PLOS Medicine* 2(8):e124.

Janz, Nicole. 2016. "Bringing the Gold Standard into the Classroom: Replication in University Teaching". *International Studies Perspectives* 17(4):392–407.

Janz, Nicole, e Jeremy Freese. 2019. "Good and bad replications in political science: how replicators and original authors (should) talk to each other". Apresentação pessoal. Disponível em: https://www.mzes.uni-mannheim.de/openscience/wp-content/uploads/2019/01/Janz-Freese_-Good-and-Bad-Replications-1.pdf.

Keele, Luke. 2015. "The Statistics of Causal Inference: A View from Political Methodology". *Political Analysis* 23(3):313–35.

King, Gary. 1995. "Replication, Replication". *PS: Political Science & Politics* 28(3):444–52.

Korell, Daniel, Niklas Reinecke, e Lars Lott. 2023, "Student-Led Replication Studies in Comparative Politics: New Findings by Fresh Eyes?". *Zeitschrift Für Vergleichende Politikwissenschaft* 17(3):261–73. doi:10.1007/s12286-023-00578-4.

Miske, Olivia, et al. 2026. "Investigating the Reproducibility of the Social and Behavioural Sciences". *Nature* 652(8108):126–34. doi:10.1038/s41586-026-10203-5.

National Academies of Sciences et al. 2019. "Understanding Reproducibility and Replicability". em *Reproducibility and Replicability in Science*. National Academies Press (US).

Open Science Collaboration. 2015. "Estimating the reproducibility of psychological science". *Science* 349(6251):aac4716. doi:10.1126/science.aac4716.

Paranhos, Ranulfo et al. 2013. "A Importância da Replicabilidade na Ciência Política: O Caso do SIGOBR". *Revista Política Hoje* 22(2):213–29.

Piper, Kelsey. 2020. "Science Has Been in a 'Replication Crisis' for a Decade. Have We Learned Anything?" Disponível em: https://www.vox.com/future-perfect/21504366/science-replication-crisis-peer-review-statistics.

Pownall, Madeleine. 2022. "Is replication possible in qualitative research? A response to Makel et al.". *Educational Research and Evaluation* 29(1–2):104–110. doi:10.1080/13803611.2024.2314526.

Rainey, Carlisle et al. 2025. “Data and Code Availability in Political Science Publications from 1995 to 2022”. *PS: Political Science & Politics* 58(2):339–45. doi:10.1017/S1049096524001276.

Reis, Guilherme Simões, e Nathália LOPES. 2021. *Replication Data for: “Uruguai: Novo Realinhamento Partidário ou Toda Democracia de Partidos Sólida se Desmancha no Ar?”*. Versão 1.1. Harvard Dataverse. doi:10.7910/DVN/6384NR.

Scielo. 2025. “Lista de Repositórios para Depósito de Dados de Pesquisa”. Disponível em: https://www.scielo.org/pt-br/sobre-o-scielo/metodologias-e-tecnologias/lista-de-repositorios-para-deposito-de-dados-de-pesquisa/.

Silva, Glauco Peres da. 2015. “Desafios Ontológicos e Epistemológicos para os Métodos Mistos na Ciência Política”. *Revista Brasileira de Ciências Sociais* 30:115–28. doi:10.17666/3088115-128/2015.

Small, Mario Luis. 2011. “How to Conduct a Mixed Methods Study: Recent Trends in a Rapidly Growing Literature”. *Annual Review of Sociology* 37(1):57–86. doi:10.1146/annurev.soc.012809.102657.

Szwako, José. 2025. *Negacionismos & Extrema-direita*. Rio de Janeiro: Telha.

Tuval-Mashiach, Rivka. 2021. “Is Replication Relevant for Qualitative Research?”. *Qualitative Psychology* 8(3):365–77. doi:10.1037/qup0000217.

Tyner, Andrew H. et al. 2026. “Investigating the Replicability of the Social and Behavioural Sciences”. *Nature* 652(8108):143–50. doi:10.1038/s41586-025-10078-y.