



CODA·BR



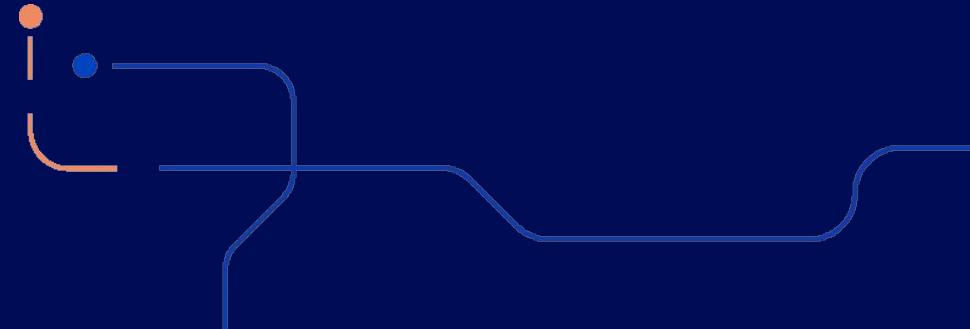
ESCOLA DE **DADOS**

Boas práticas em ciência aberta e reprodutível

BEATRIZ MILZ • milz.bea@gmail.com

OLAVO AMARAL • olavo@bioqmed.ufrj.br

Material suplementar: [Clique aqui](#)



BEATRIZ MILZ

- Doutora em Ciência Ambiental (**USP**)
- Atualmente fazendo o estágio pós-doutoral na **UFABC**
- Associada da **Open Knowledge Brasil (OKBR)**
- *Peer Review Software Editor* na **rOpenSci**
- Equipe da Secretaria Executiva da **Revista Ambiente & Sociedade**
(Coleção **SciELO**)

OLAVO AMARAL

- Médico, doutor em bioquímica (UFRGS).
- Professor associado - Instituto de Bioquímica Médica UFRJ
- Ex-radiologista, ex-neurocientista, escritor
- Metacientista dedicado a melhorar o panorama de confiabilidade e abertura da ciência.
- Fundador da Iniciativa Brasileira de Reprodutibilidade, Rede Brasileira de Reprodutibilidade, BRISA e No-Budget Science.

NESTE WORKSHOP, VAMOS VER...

- Por que ciência aberta?
- Boas práticas
- Exemplos de boas práticas que podem ser úteis no jornalismo de dados
- Discussão

PAINEL

Conhecimento livre e ciência aberta: caminhos para uma sociedade resiliente e desenvolvida

Débora Peres Menezes, Olavo Amaral, Sarita Albagli. Mediação: Alexander Hilsenbeck

Local: Auditório

Por que ciência aberta?

Conhecimento Livre e Ciência Aberta

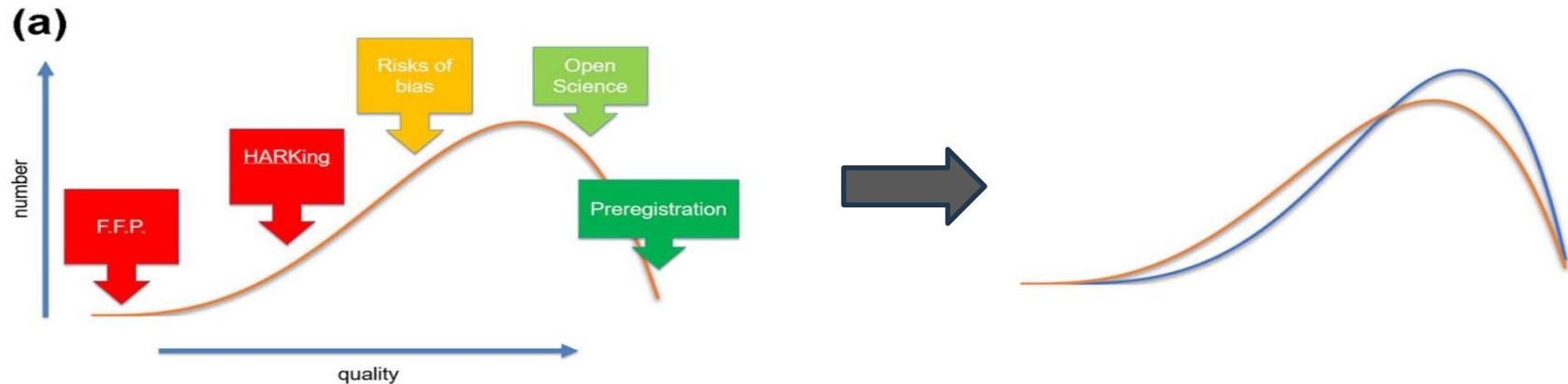
Por que ciência aberta?

- Já falei meus 15 minutos ontem e não vou me repetir, mas em resumo...
 - Mais acessível.
 - Mais inclusiva.
 - Mais eficiente.
 - Mais reutilizável.
 - Mais confiável.
 - Maior impacto.



Uma coisa a ser repetida

- “A Ciência Aberta” não é algo que existe no horizonte.
- Fazer uma ciência aberta é algo que se constrói todos os dias, pouco a pouco.
- Se você está envolvido em pesquisa em qualquer nível, é provável que tenha algo a melhorar nesse sentido.



Transferindo conhecimento

- Boa parte do meu trabalho hoje é ajudar pessoas a dar esses passos dentro da pesquisa acadêmica.
- Eu *não* estou acostumado a fazer isso fora dela, mas imagino que algumas coisas sejam bastante transferíveis.

The screenshot shows the homepage of the piauí magazine website. At the top, there's a navigation bar with links like 'A Revista', 'piauí jogos', 'Áudio', 'piauí recomenda', 'Dossiê piauí', 'Especiais', 'Eventos', 'Herald', 'Vídeos', 'Lupa', 'Faça seu login', and social media icons. Below the navigation is a large red header with the word 'piauí' in white. Underneath the header, there's a search bar and a menu icon. A teal banner labeled 'COLUNISTAS' contains a minus sign. To the left, there's a circular profile picture of a man with curly hair and a beard. In the center, there's a portrait of a man with grey hair and a beard, identified as 'OLAVO AMARAL'. Below his portrait, there's a bio: 'É médico, professor do Instituto de Bioquímica Médica da UFRJ e autor de *Dicionário de Línguas Imaginárias* (Alfaguara)'. To the right, there's a block of text about him: 'OLAVO AMARAL é médico, escritor e pr... e hoje se dedica à promoção de uma ci... Iniciativa Brasileira de Reprodutibili... da ciência biomédica brasileira, e o N... dedicados a construir uma ciência me... Imaginárias e Correnteza e Escombro'. At the bottom left, there's another circular profile picture of a man with curly hair and a beard.

OLAVO AMARAL

na saúde e
na saúde e
na saúde e
na doença
na doença
na doença

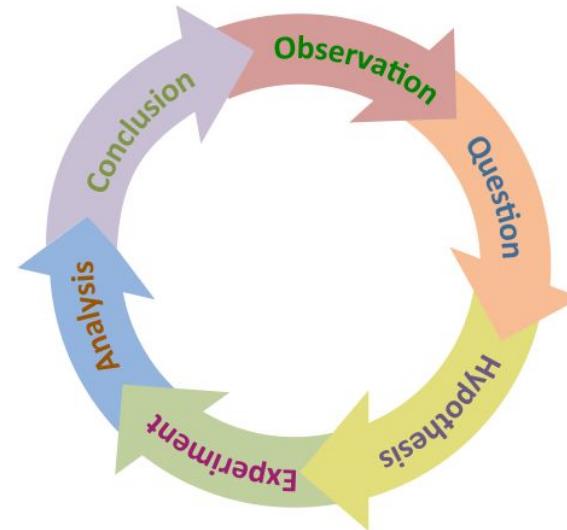
A MEDICINA ENTRE
A CIÊNCIA E A VIDA REAL



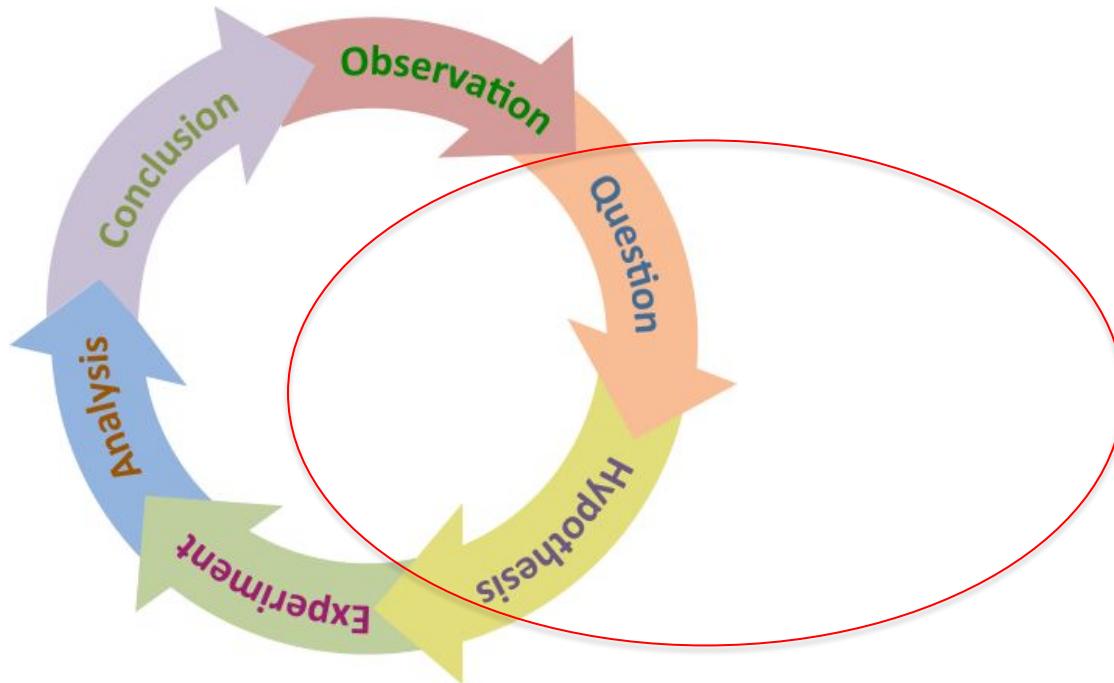
BOAS PRÁTICAS

Ao longo do processo

- Ciência aberta *não* é uma mera questão de compartilhamento de informações ao final e um processo.
- A transparência e reproduzibilidade podem ser aumentadas por pesquisadores individuais em inúmeros passos do ciclo de pesquisa:
 - Ao elaborar um projeto
 - Ao coletar dados
 - Ao analisar dados
 - Ao publicar resultados
 - Na vida acadêmica em geral

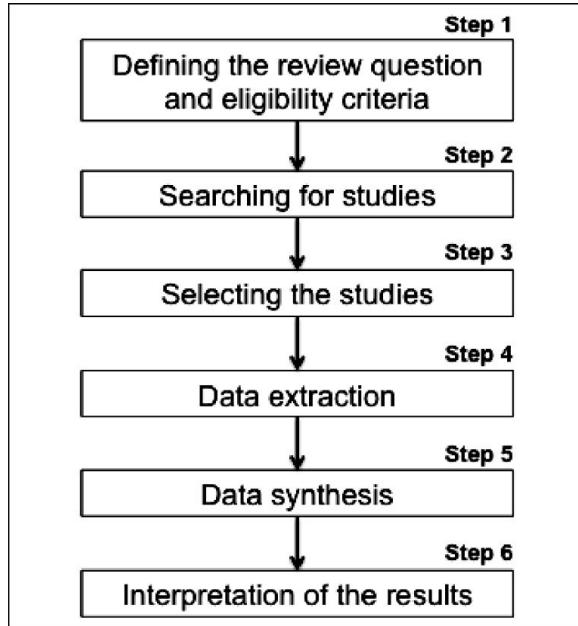


Ao elaborar um projeto



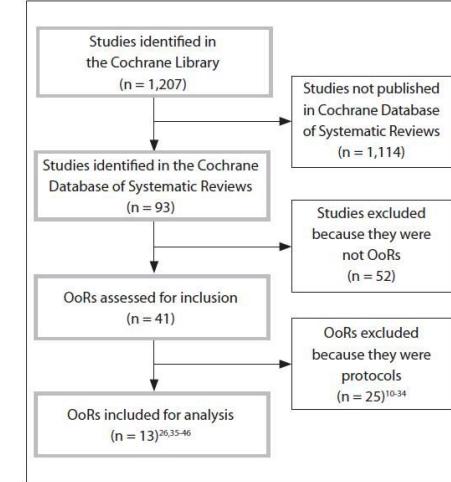
Revisões sistemáticas

- Quase todo projeto começa por uma revisão da literatura, e sistematizá-la a torna mais completa, menos enviesada e reproduzível.



Search terms (including truncations)

'Sport* inj*' OR 'athlet* inj*' (ab)
AND
Psychosocial OR psycholog* OR emotion* (ab)
AND
Rehabilitat* OR recover* OR outcome* OR return (ab)
AND
athlet* OR player* OR individual*OR patient* (ab)



Pré-registro de protocolos

- Observação exploratória e testagem de hipóteses são passos diferentes do processo científico.
- Um conjunto de dados não pode gerar e testar uma hipótese ao mesmo tempo.



A falácia do atirador do Texas

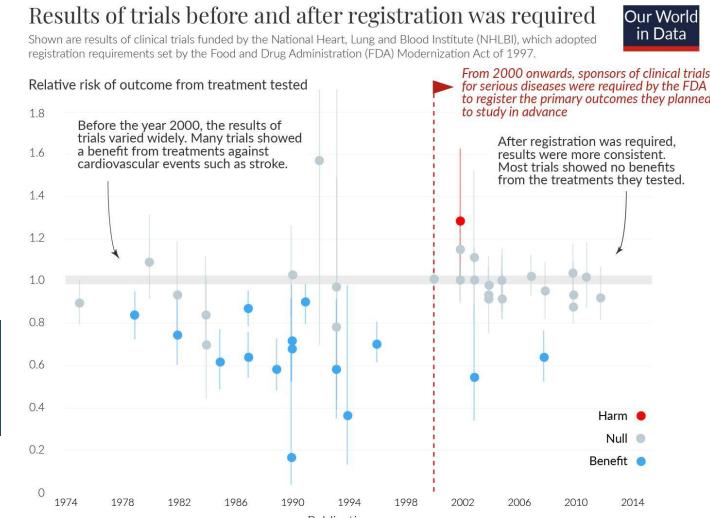
Pré-registro de protocolos

- Pré-Registrar um protocolo permite ao leitor diferenciar análises exploratórias de confirmatórias.
- Além disso, permite a identificação de estudos realizados e não publicados.



The preregistration revolution

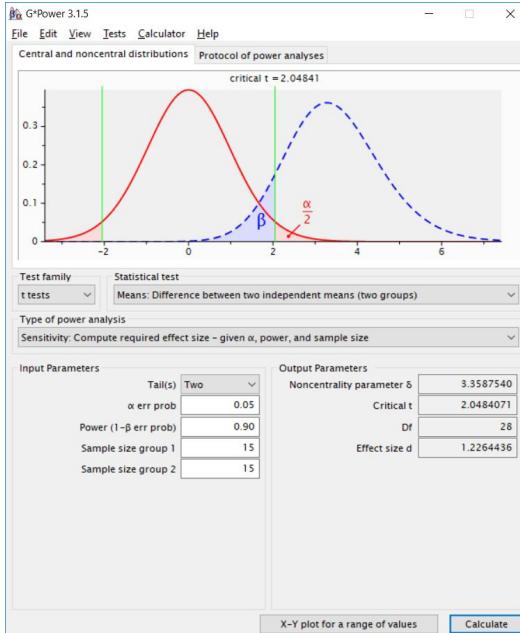
Brian A. Nosek^{a,b,1}, Charles R. Ebersole^b, Alexander C. DeHaven^a, and David T. Mellor^a



Note: Relative risks and 95% confidence intervals are shown for treatments against cardiovascular outcomes.
Data source: Kaplan and Irvin (2015). Likelihood of null effects of large NHLBI clinical trials has increased over time. PLOS ONE.
OurWorldInData.org - Research and data to make progress against the world's largest problems.
Licensed under CC-BY by the author Saloni Dattani.

Cálculo de tamanho amostral

- Se você vai trabalhar com dados quantitativos, conhecer o tamanho de amostra necessário para fazer inferências além dela é um passo óbvio.



Example: Sample Size Justification

- In the article the authors state:

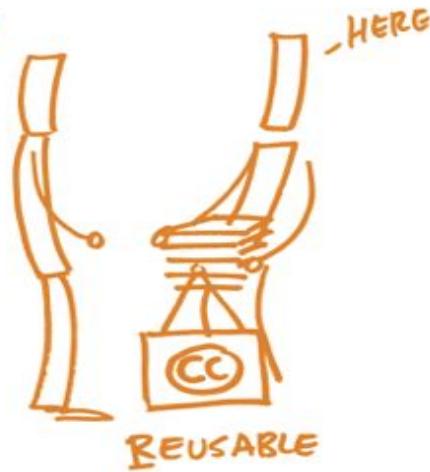
"The sample size was based on an estimated standard deviation of 15 for the change in the percentage of the predicted FEV₁ in the Prednisone group. In order to allow detection of a 10 percentage point difference between the groups in the degree of improvement in FEV₁ (as a percentage of the predicted value) from base line to 240 minutes and to maintain an alpha (α) error of 0.05 and a beta (β) error of 0.10, the required size of the sample was 94 children."

Reference: Schuh et al., (2000) NEJM. 343(10):689-694

Plano de gestão de dados

- Lembre que para serem úteis, seus dados devem ser encontráveis, acessíveis, interoperáveis e reutilizáveis.
- O momento para assegurar que isso aconteça não é depois de coletá-los, e sim antes.

FAIR DATA PRINCIPLES



Preservação de dados

- Isso deve valer não só para você mas para outros, e não só para hoje como para daqui a 10 ou 20 anos.
- Há inúmeros repositórios (gerais ou específicos) para esse fim.



Known and Predicted Protein-Protein Interactions

NCBI Gene
Entrez Gene

Compartilhamento ético

- Se você vai trabalhar com seres humanos, lembre que a autorização para compartilhar dados têm de ser pedida de antemão e aprovada por um comitê de ética.

Editable Template

INFORMED CONSENT

Informed Consent

Title of Study/Procedure

Introduction

This section briefly explains the purpose of the informed consent form, introducing the study or procedure, and the organization responsible for it.

Welcome to [Name of Organization/Study Team]. This document serves to inform you about [Name/Type of Study or Procedure] you are considering participating in. Our goal is to ensure that you understand the nature of the procedure, its benefits, risks, and your rights as a participant.

Investigator/Provider Contact Information

- Name: _____
- Position: _____

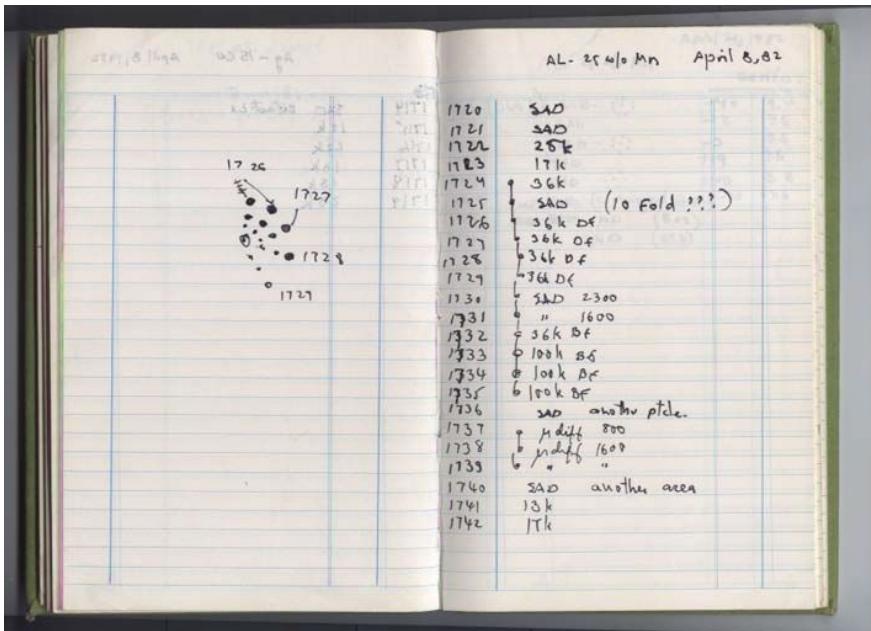


Ao coletar dados



Documentação do processo

- O registro da coleta dos dados é uma parte do próprio dado e deve ser preservado.



Documentação do processo

- Mais do que ferramentas específicas, isso envolve organização e boas práticas de gerenciamento, versionamento e nomenclatura de arquivos.

Guia de boas práticas em uso de dados na ciência

Um manual prático para você se atualizar

<https://osf.io/7su8g/>



Controle de viés

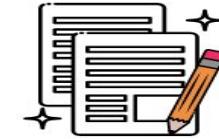
- Todos temos um conflito de interesses com os resultados de um estudo, e somos as piores pessoas para coletar dados de forma isenta.

Don't touch THIS
should be beyond your control as a scientist?

The results

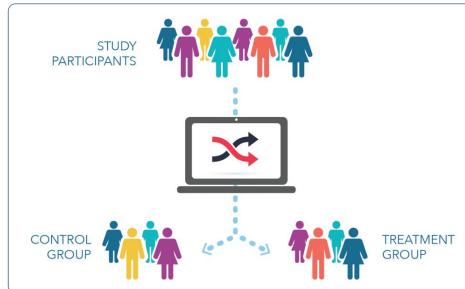
But make sure THIS is amazing

The results

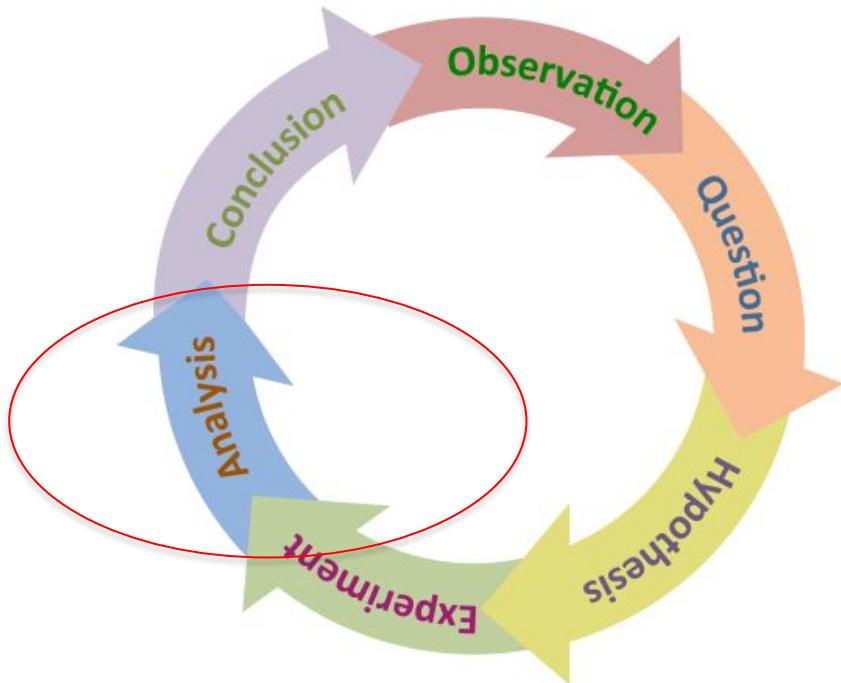


Controle de viés

- Aceitar que vieses sempre existem não significa que não devamos tentar controlá-los.
- Formas de controle de viés vão variar conforme o tipo de experimento:
 - Randomização de sujeitos
 - Cegamento do experimentador
 - Monitoramento e análises independentes
 - (...)



Ao analisar dados



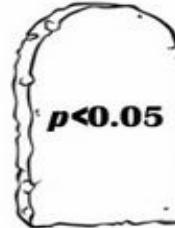
Entendendo estatística

- Se você for usar análise estatística, entenda o que você está fazendo (ao contrário da maioria dos cientistas).

Statistical Rituals: The Replication Delusion and How We Got There

Gerd Gigerenzer

Harding Center for Risk Literacy, Max-Planck Institute for Human Development, Berlin, Germany



“Lavagem de incerteza”

Análise reproduzível

- Trocar botões por código da análise e compartilhá-lo é a forma mais eficiente de garantir reproduzibilidade de análise.
- Num mundo de LLMs, isso se tornou acessível a um universo mais amplo (mas não exime você de uma alfabetização mínima).



Sem soluções mágicas

- Usar código não exime você de entender o que está acontecendo com seus dados.
- Documente a análise passo a passo, e se certifique que o que está acontecendo é o que você queria.
- Ferramentas de documentação como Jupyter ou Markdown podem facilitar esse processo.

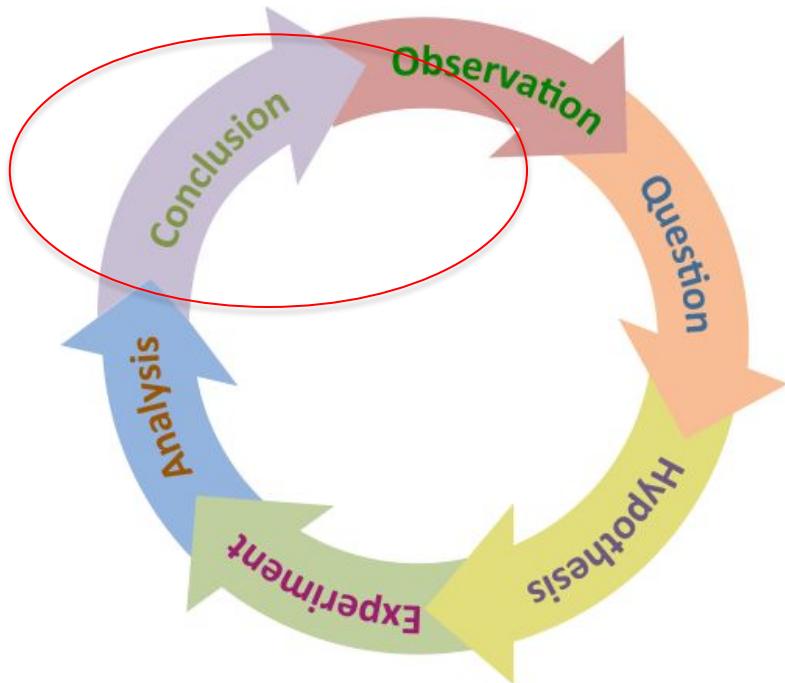
PLOS COMPUTATIONAL BIOLOGY

Good enough practices in scientific computing

Greg Wilson  , Jennifer Bryan , Karen Cranston , Justin Kitzes , Lex Nederbragt , Tracy K. Teal 



Ao publicar resultados



Acesso aberto

- Acesso aberto: no mínimo, seu artigo tem que poder ser lido agora que o Sci-Hub parou de ser atualizado em tempo real.



nature > commentary > article

MENU ▾ nature

Commentary | Published: 30 April 1992

The growing inaccessibility of science

Donald P. Hayes

Nature 356, 739–740(1992) | Cite this article

1015 Accesses | 44 Citations | 27 Altmetric | Metrics

Access options

Rent or Buy article
Get time limited or full article access on ReadCube.
from \$8.99
[Rent or Buy](#)

Subscribe to Journal
Get full journal access for 1 year
\$199.00
only \$3.83 per issue
[Subscribe](#)

Acesso aberto

- Isso *não* significa ter que pagar taxas de publicação: existem inúmeras formas de tornar um artigo disponível que não dependem disso.



Gold open access

Sherpa Romeo



Green open access

arXiv.org



DOAJ DIRECTORY OF
OPEN ACCESS
JOURNALS

SciELO

 OSF **PREPRINTS**

bioRxiv
beta
THE PREPRINT SERVER FOR BIOLOGY

medRxiv
TM
THE PREPRINT SERVER FOR HEALTH SCIENCES

 **Research Square**

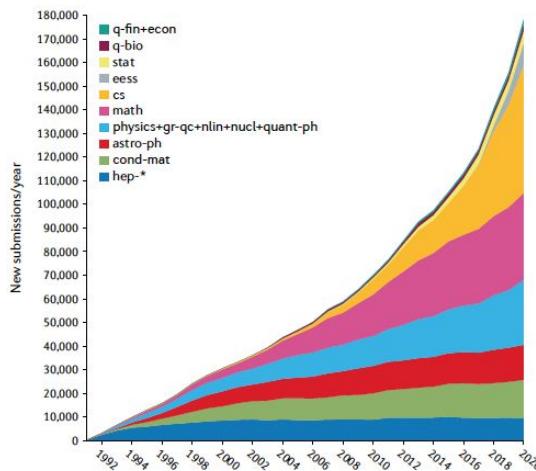
ChemRxiv
TM

Preprints

- O uso de preprints como forma de disponibilização é cada vez mais aceito em diferentes áreas, acelera o processo científico e liberta você de editores e revisores.

Lessons from arXiv's 30 years
of information sharing

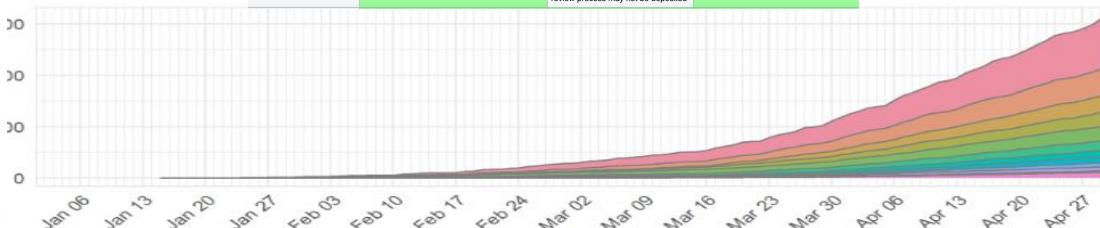
Paul Ginsparg



List of academic publishers by preprint policy

From Wikipedia, the free encyclopedia

Publisher	Location	Restrictions	License
American Association for the Advancement of Science	Not-for-profit servers (e.g. arXiv, bioRxiv, chemRxiv, medRxiv)	Unrestricted	Unrestricted
American Association for Cancer Research	Unrestricted	Must not post revised manuscript (after peer review or editorial comment)	Unrestricted
American Association for Physics in Medicine	Non-commercial servers (e.g. arXiv, Open Science Framework, Zenodo)	Unrestricted	Unrestricted
American Chemical Society	Unrestricted	Unrestricted	Unrestricted
Association for Computing Machinery	Non-commercial servers (e.g. arXiv, Open Science Framework, Zenodo)	Unrestricted	Unrestricted
American Geophysical Union	Unrestricted	Unrestricted	Unrestricted
American Heart Association	Unrestricted	Versions of a manuscript that have been altered as a result of the peer review process may not be deposited	Unrestricted



medRxiv
SSRN
ResearchGate
Research Square
bioRxiv
Preprints.org
OSF Preprints
JMIR
PsyArXiv (OSF)
ChemRxiv
SocArXiv (OSF)
Other

Diretrizes de relato

- Comunidades específicas podem (e devem) entrar em acordo sobre o que precisa ser relatado para tipos específicos de artigos.



PRISMA 2020 Statement. MetaarXiv preprint (September 2020)

Section and Topic	Item #	Checklist Item
TITLE		
Title	1	Identify the report as a systematic review.
ABSTRACT		
Abstract	2	See the PRISMA 2020 for Abstracts checklist (Table 2).
INTRODUCTION		
Rationale	3	Describe the rationale for the review in the context of existing knowledge.
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.
METHODS		
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.
Search strategy	7	Present the full search strategies for all databases, registers and websites, along with filters and limits used.
Selection process	8	Specify the methods used to decide whether the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis.
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.
	13e	Describe any methods used to explore possible causes of heterogeneity among study results.
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.



Reporting guidelines for main study types

Randomised trials	CONSORT	Extensions
Observational studies	STROBE	Extensions
Systematic reviews	PRISMA	Extensions
Study protocols	SPIRIT	PRISMA-P
Diagnostic/prognostic studies	STARD	TRIPOD
Case reports	CARE	Extensions
Clinical practice guidelines	AGREE	RIGHT
Qualitative research	SRQR	COREQ
Animal pre-clinical studies	ARRIVE	
Quality improvement studies	SQUIRE	
Economic evaluations	CHEERS	

Disponibilização de dados

- Seus dados são tão ou mais importantes do que seu artigo, e disponibilizá-los da forma mais estruturada possível faz uma diferença enorme para a ciência.
- Repositórios apropriados à área tendem a ser a forma mais útil de compartilhamento, mas o mais importante é que eles estejam algum lugar e acessá-los não dependa de você.

NEWS | 21 June 2022

Many researchers say they'll share data – but don't

Reasons included a lack of informed consent or ethics approval to share; misplaced data; and that others had moved on from the project.

COMMENT | VOLUME 35, 100795, JULY 2024

 Download Full Issue

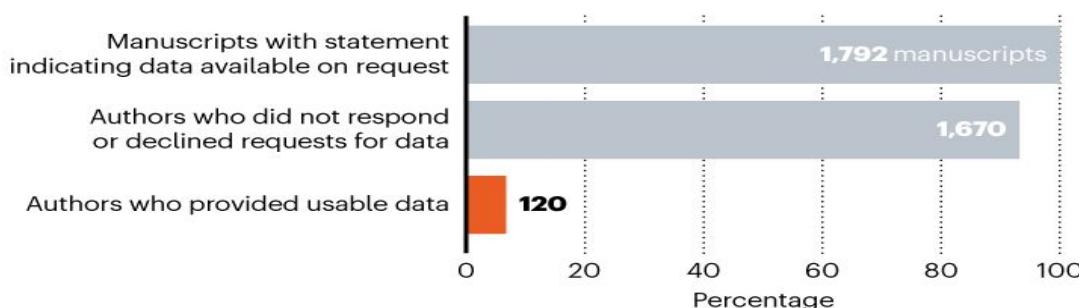
A reasonable request for true data sharing

Tiago Lubiana • Helder I. Nakaya  

Open Access • Published: May 28, 2024 • DOI: <https://doi.org/10.1016/j.lana.2024.100795>

DATA-SHARING BEHAVIOUR

Of almost 1,800 manuscripts for which the authors stated they were willing to share their data, more than 90% of corresponding authors either declined or did not respond to requests for data. Only about 7% of authors actually handed over data.

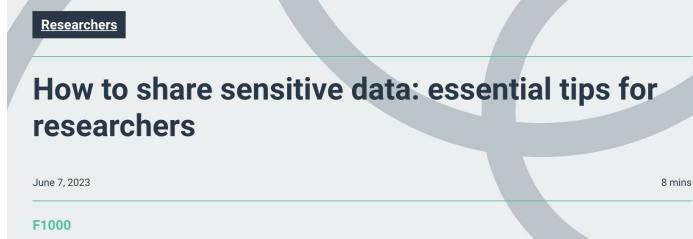


Disponibilização de dados

- É óbvio que existem inúmeros casos de dados sensíveis que não devem ser totalmente compartilhados, mas esses são a exceção e não a regra.
- Mesmo para esses, no entanto, existem alternativas para anonimização e/ou disponibilização parcial de dados e metadados.

Anonymising and sharing individual patient data

Khaled El Emam,^{1,2} Sam Rodgers,³ Bradley Malin⁴



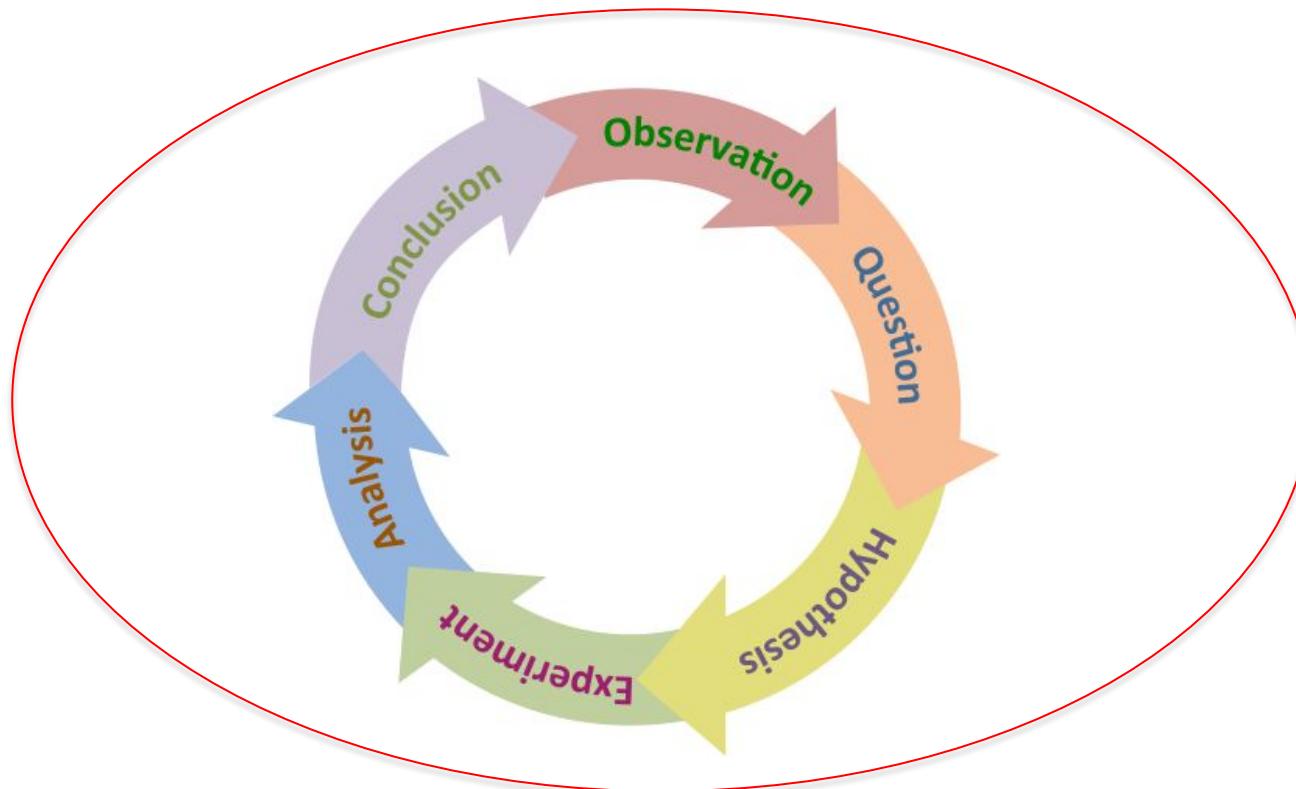
Gates Open Research

[Gates Open Research](#) » [Blog](#) » How can researchers share sensitive data openly?

Blog

[How can researchers share sensitive data openly?](#)

Na vida acadêmica em geral



Autoria e contribuições

- Discuta critérios de autoria ao iniciar um projeto, e explice as contribuições de cada participante.



**Authorship and
contributorship**

**CRediT – Contributor Roles
Taxonomy**



14 Contributor Roles

Conceptualization	Resources
Data curation	Software
Formal Analysis	Supervision
Funding acquisition	Validation
Investigation	Visualization
Methodology	Writing – original draft
Project administration	Writing – review & editing

Revisões abertas

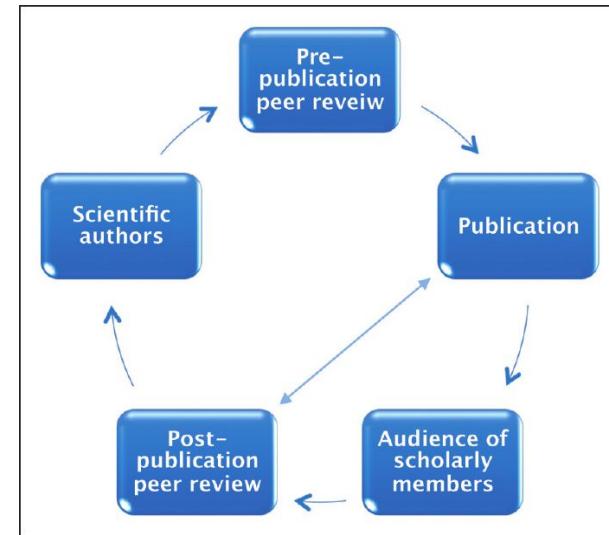
- Revisões de artigos são contribuições ao processo científico e não existe razão para que permaneçam escondidas.
- Engaje-se na revisão e discussão pós-publicação de forma a fazer uma contribuição pública e visível.

the **bmj**

eLife

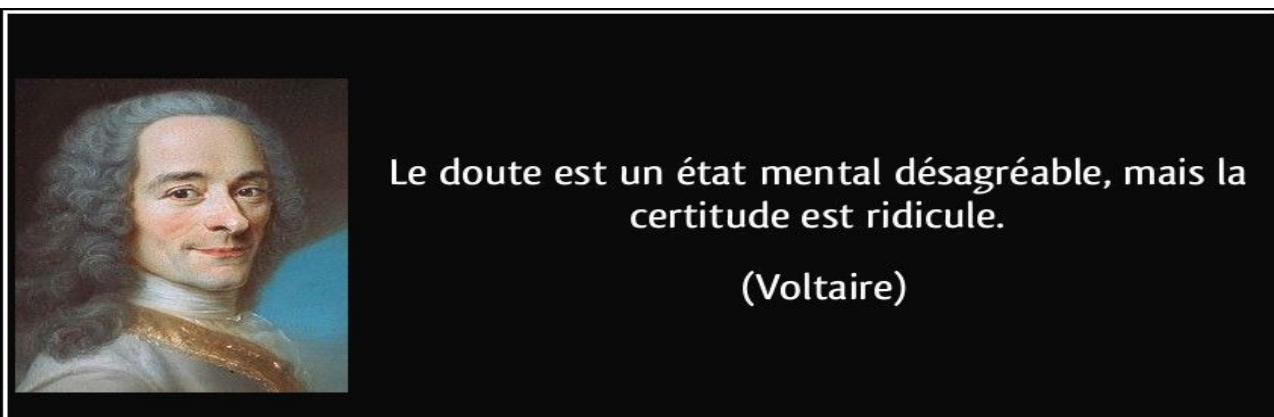


bioRxiv
THE PREPRINT SERVER FOR BIOLOGY



Na comunicação com público e pares

- Assuma a ciência como um processo de refinamento de incertezas em revisão permanente que não termina com a publicação.



Cite this as: BMJ 2020;371:m3979

<http://dx.doi.org/10.1136/bmj.m3979>

Published: 19 October 2020

Covid-19's known unknowns

The more certain someone is about covid-19, the less you should trust them

George Davey Smith,¹ Michael Blastland,² Marcus Munafò¹

É menos fácil do que parece

- Opor “ciência” e “anticiência” ou “informação” e “desinformação” virou um hábito persistente no debate público, e raramente ajuda.

CPI: 'Ciência não tem dois lados', diz especialista a senador defensor do tratamento precoce

Microbiologista Natalia Pasternak deu declaração durante audiência pública na CPI da Covid. Eduardo Girão (Pode-CE) sugeriu debate com especialistas pró e contra tratamento precoce.

Por Gustavo Garcia, G1 — Brasília
11/06/2021 17h17 · Atualizado há 4 anos



Natalia Pasternak: 'Ciência não é uma questão de opinião'



Você não é a ciência

- Boa parte do meu trabalho como jornalista nos últimos tempos tem sido apontar para isso.

OLAVO AMARAL

na saúde e
na saúde e
na saúde e
na doença
na doença
na doença

A MEDICINA ENTRE
A CIÊNCIA E A VIDA REAL



Checagem de fatos científicos: crônica de um fracasso anunciado

27 de setembro de 2022

Estabelecer o que é um fato científico não é simples, e é estranho pensar que delegar a tarefa aos jornalistas possa dar certo

A ivermectina é comprovadamente inefica... ops

16 de abril de 2024

O maior ensaio clínico testando a droga contra a covid-19 passou completamente em branco, mas seus resultados deveriam nos fazer refletir sobre nossas certezas

Estamos desinformando sobre a desinformação?

10 de julho de 2024

A desinformação científica é um espantalho conveniente para a pesquisa acadêmica não ter que abordar seus próprios problemas

A revanche silenciosa da cloroquina

29 de outubro de 2024

Revisitar a improvável saga do medicamento mais odiado da última década é um exercício de humildade para o establishment científico

EXEMPLOS DE BOAS PRÁTICAS

Metodologia, dados e código

MATERIAL DE REFERÊNCIA



Guia de boas práticas em ciência aberta e reproduzível, publicado por [Instituto Serrapilheira](#) e escrito por [Olavo Amaral](#).

"QUAIS FATORES CONTRIBUEM PARA A PESQUISA IRREPRODUTÍVEL?"

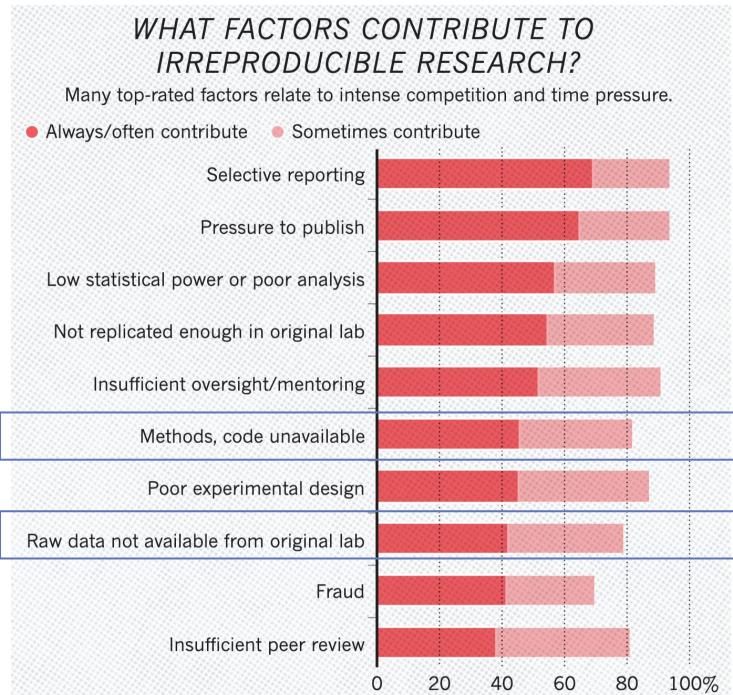
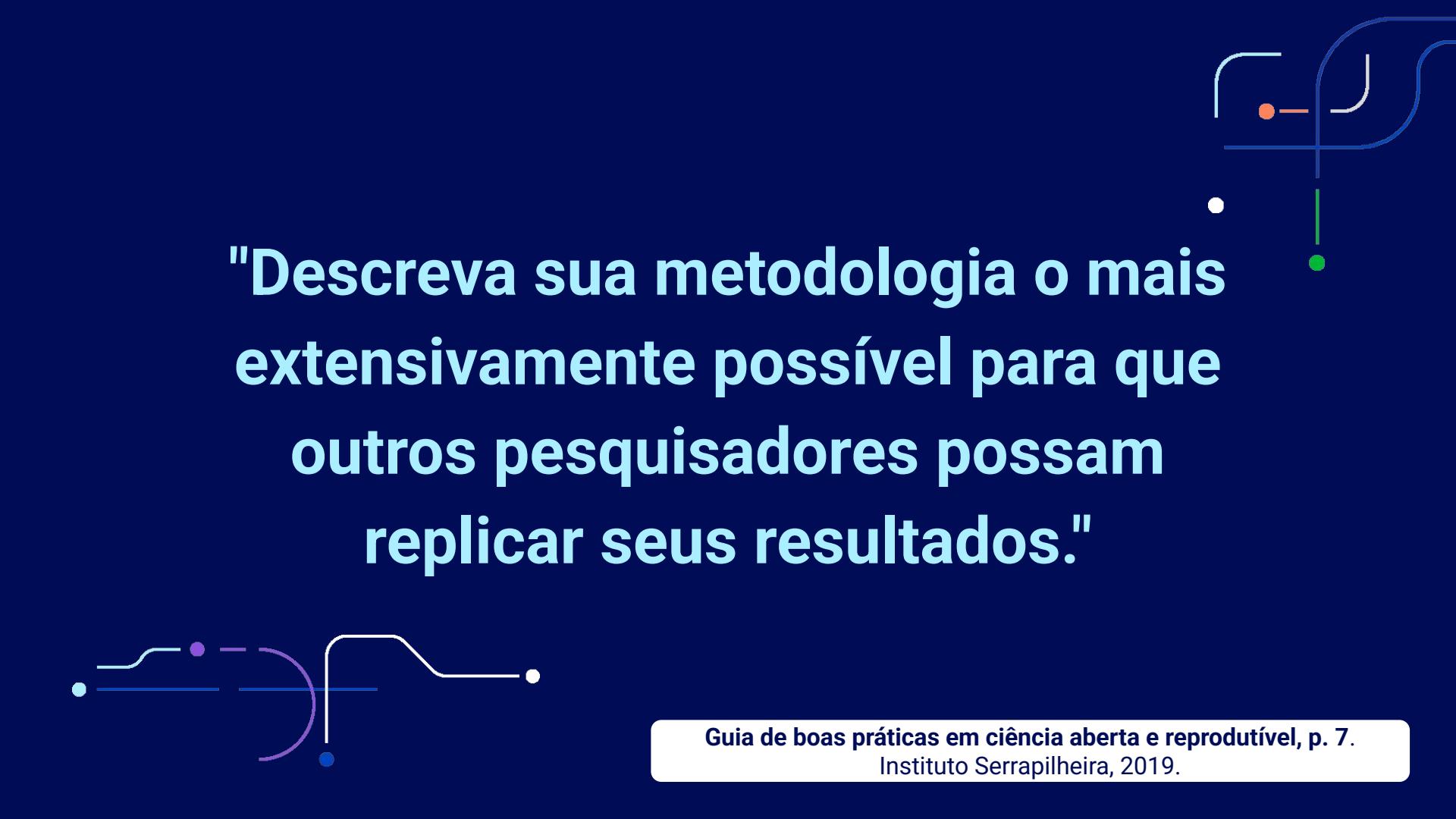


Figura apresentada no artigo "[1,500 scientists lift the lid on reproducibility](#)", publicado na revista Nature em 2016, de autoria de Monya Baker.

METODOLOGIA

Documentar o que foi realizado!



"Descreva sua metodologia o mais extensivamente possível para que outros pesquisadores possam replicar seus resultados."

EXEMPLO DE REPORTAGEM

The screenshot shows a news article from the website apublica.com.br. The top navigation bar includes the logo 'apublica' and links for 'APOIE' and a search function. On the left side, there is a vertical column of social media sharing icons for Facebook, Twitter, LinkedIn, and others. The main headline is 'Em São Paulo, corte de luz na periferia é até 8 vezes mais frequente que no centro'. Below the headline, a subtext reads 'Levantamento com dados da Enel mostra que, mesmo antes do apagão, falta de luz é comum nos bairros periféricos'. The publication details at the bottom indicate it was published on November 11, 2023, at 08:00, by Bianca Muniz, Bruno Fonseca, Danilo Queiroz, Gabriel Gama, and José Cícero. A small note at the bottom left states 'Este artigo tem mais de 1 ano'.

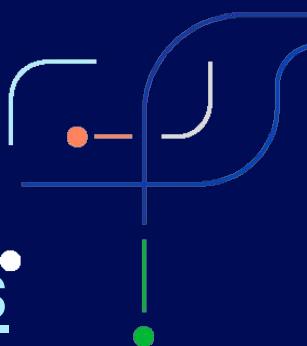
Reportagem publicada por [apublica](#), em 2023, de autoria de Bianca Muniz, Bruno Fonseca, Danilo Queiroz, Gabriel Gama, José Cícero

METODOLOGIA DA REPORTAGEM

METODOLOGIA	-
<ul style="list-style-type: none">▪ Os dados de “<u>Interrupções</u> de Energia Elétrica nas Redes de Distribuição” foram baixados do <u>portal</u> de dados abertos da Aneel no dia 8 de novembro de 2023. Os dados podem ser consultados aqui;▪ A localização dos conjuntos de unidades consumidoras foram exportados no portal de <u>dados georreferenciados</u> da Aneel. Selecionamos apenas os dados do ano de 2022;▪ Os dados georreferenciados dos <u>distritos</u> de São Paulo foram obtidos no portal de dados abertos da Prefeitura do município;▪ Os arquivos georreferenciados de distritos de São Paulo e de conjuntos de unidades consumidoras foram sobrepostos utilizando o software QGIS;	<ul style="list-style-type: none">▪ Em seguida, inserimos os indicadores de continuidade (DEC e FEC) no mapa, nas camadas sobrepostas;▪ Um conjunto pode abastecer diferentes distritos do município e cidades vizinhas, bem como um distrito pode ser atendido por mais de um conjunto de unidades consumidoras. Levamos em consideração essa definição durante a análise: incluímos os dados de indicadores de DEC e FEC por conjunto de unidades consumidoras que atendem distritos de São Paulo. Se um conjunto da Enel não atende o município, ele foi desconsiderado; se atende um distrito, mas atende outras cidades e distritos vizinhos, ele foi considerado;▪ Os códigos para o tratamento dos dados e análise estão hospedados no <u>Github</u> da Pública.

DADOS

Organização e disponibilização



**"Mantenha seus bancos de dados
organizados e documente as
análises realizadas da mesma forma
que você faria com resultados
experimentais."**

BOAS PRÁTICAS

- Manter os dados organizados
- Salvar uma cópia dos dados originais (sem modificações)
- Documentar as formas de obtenção dos dados
- Compartilhar os dados em um repositório (ex: Zenodo)

Em caso de uso de dados secundários (ex. dados abertos)

- Além de salvar uma cópia dos dados originais (sem modificações), registre também informações como:
 - Data de acesso
 - Data de referência dos dados (caso isso esteja informado)
 - Website/plataforma de acesso (se possível com a URL)
 - Caso o website/plataforma tenha filtros de pesquisa para acessar os dados, registrar quais foram utilizados
 - [...] outras informações que sejam relevantes para que outras pessoas possam obter os mesmos dados (ou uma versão mais atualizada, caso o site tenha sido atualizado)

EXEMPLO DE REPORTAGEM

The screenshot shows a news article from the website PÚBLICA. At the top, there is a yellow header bar with the PÚBLICA logo, a menu icon, a purple 'APOIE' button, and a search icon. Below the header is a large photograph of several men in military-style uniforms standing next to large logs. One log has the number '615' written on it. A yellow 'REPORTAGEM' box is overlaid on the photo. The main title of the article is 'Prefeito infrator: políticos multados por danos ambientais disputam votos em 2024'. To the left of the title, there are social media sharing icons for Facebook, Twitter, LinkedIn, WhatsApp, and Telegram. To the right, there is a purple donation box with the text 'Leia de graça, retribua com uma doação.' and a 'Doe aqui!' button. At the bottom of the article, there is a small note in Portuguese: 'Levantamento da PÚBLICA mostra que 69 prefeitos eleitos em 2020 têm quase R\$ 35 milhões em multas ambientais'.

PUBLICA

APOIE

REPORTAGEM

Prefeito infrator: políticos multados por danos ambientais disputam votos em 2024

Leia de graça, retribua com uma doação.
Doe aqui!

Levantamento da PÚBLICA mostra que 69 prefeitos eleitos em 2020 têm quase R\$ 35 milhões em multas ambientais

14 de agosto de 2024

12:00

Este artigo tem mais
de 1 ano

Por Bianca Muniz, Bruno Fonseca, Rafael Oliveira

Reportagem publicada por [apublica](#), em 2024, de autoria de Bianca Muniz, Bruno Fonseca, Rafael Oliveira

EXEMPLO DE REPORTAGEM

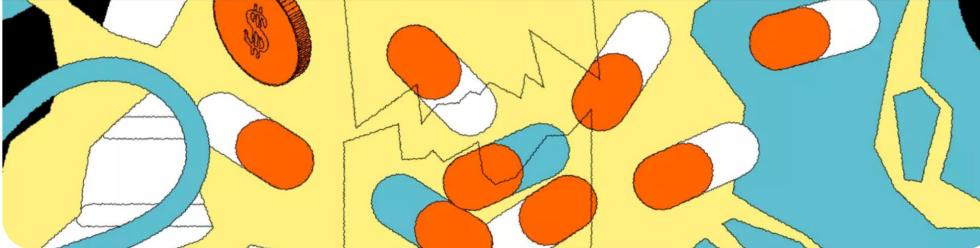
- As informações de candidatos de 2024 foram coletadas diretamente do repositório de dados do TSE, com última atualização às 19h30 do dia 12 de agosto. O código da coleta pode ser visto no [link](#);
- As multas ambientais aplicadas pelo Ibama foram extraídas do [Portal de Dados Abertos](#) do órgão, com última atualização no dia 12 de agosto. Selecionamos multas aplicadas entre janeiro de 2015 e agosto de 2024, excluindo os seguintes status: “Baixado – Defesa deferida”, “Baixado – Recurso de ofício indeferido”, “Baixado – Recurso deferido”, “Baixado c/base na Lei 9.873/99 (Prescrito)”, “Baixado por determinação judicial”, “Baixado por prescrição intercorrente durante anál. Jurídica”, “Cancelado”, “Cancelado na homologação (AI sem defesa)”, “Substituição de multa por advertência”, “Substituído na homologação por outro AI”. O código e o arquivo resultante podem ser vistos no [link](#);

EXEMPLO DE REPORTAGEM

The screenshot shows a news article from the website 'aos fatos'. At the top, there's a navigation bar with icons for menu, search, and support, along with a '10+' rating. Below the header is a colorful illustration of various medical pills and capsules on a yellow background. The author's name, 'Méuri Elle', is visible at the bottom right of the illustration. The publication date is listed as '26 de novembro de 2024, 17:33'. The main title of the article is 'Como anúncios usam sites jornalísticos para espalhar desinformação em saúde e curas falsas'. Below the title, the names of the authors are mentioned: 'Por Bianca Bortolon, Bruno Fávero, Milena Mangabeira, Rhenan Bartels e Yvna Sousa'. In the bottom right corner, there's a logo for 'fátima' with a small circular icon.

aos fatos: 10+

apoie +



Méuri Elle

26 de novembro de 2024, 17:33

Como anúncios usam sites jornalísticos para espalhar desinformação em saúde e curas falsas

Por Bianca Bortolon, Bruno Fávero, Milena Mangabeira, Rhenan Bartels e Yvna Sousa

fátima

Reportagem publicada por , em 2024, de autoria de Bianca Bortolon, Bruno Fávero, Milena Mangabeira, Rhenan Bartels e Yvna Sousa

EXEMPLO DE REPORTAGEM

The screenshot shows a news article from the 'aos fatos' website. The header includes the site's logo, a '10+' rating, a search icon, and a 'apoie' button. The main content starts with a section titled 'Coleta dos dados' (Data Collection) which describes the automated process of collecting and analyzing data from multiple news websites. It mentions using 'Playwright' for automation and 'Playwright' for geolocation. A bar chart at the bottom shows the average number of ads per URL, with 'Globo' having the highest average of 48.32.

Coleta dos dados

O processo de coleta e análise automatizada dos dados ocorreu em três etapas.

De 17 de junho a 30 de julho de 2024, um servidor acessou cada domínio múltiplas vezes por dia e extraiu as URLs de links presentes em sua página inicial. Em seguida, o mesmo servidor acessou cada URL coletada e, usando a biblioteca de automação [Playwright](#), fez download dos textos e imagens dos anúncios nativos encontrados. No total, foram extraídas 15.164 URLs únicas, das quais foram coletados 242.812 anúncios, como mostra o gráfico abaixo.

Os anúncios coletados, por site

Foram analisados **242.812 anúncios**, em **15.164 URLs de notícias**, média de **16 anúncios** por URL

Média de anúncios por URL: 1 48,32

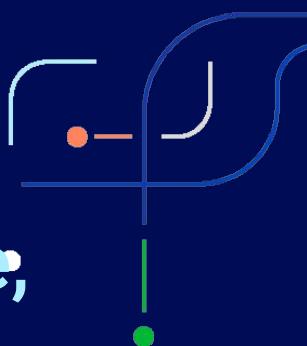
Sítio	Média de anúncios por URL
Globo	48,32
Metrópoles	1

O código fonte usado para a coleta e análise automatizada dos dados pode ser encontrado no [Github do Aos Fatos](#). Para a coleta usada neste relatório, o servidor foi georreferenciado em São Paulo. Como anúncios programáticos levam em consideração a localização e o histórico dos usuários para decidir qual publicidade exibir, outras configurações levariam a resultados diferentes.

CÓDIGO

BOAS PRÁTICAS

- Utilizar linguagens de código aberto (ex. Python, R, etc)
- Disponibilizar os códigos em um repositório (ex. GitHub, GitLab...)
- Documentar o código:
 - quais são as ferramentas utilizadas
 - o que o código realiza
 - decisões tomadas que sejam relevantes para entender o contexto
 - o que outras pessoas precisam saber para reproduzir a análise
 - [...] entre outros



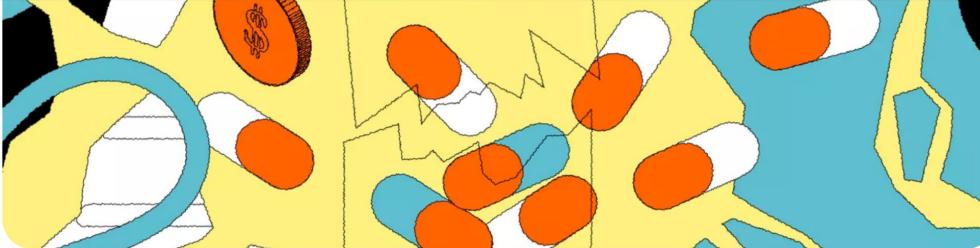
**"Caso use código próprio para análise,
mantenha-o documentado e
compreensível para terceiros. Considere
integrar o código a seus resultados com
ferramentas [...] que permitem repetir
análises e simulações facilmente."**

EXEMPLO DE REPORTAGEM

The screenshot shows a news article from the website 'aos fatos'. At the top, there's a navigation bar with icons for menu, search, and support, along with a '10+' rating. Below the header is a colorful illustration of various medical pills and capsules on a yellow background. The author's name, 'Méuri Elle', is visible at the bottom right of the illustration. The publication date is listed as '26 de novembro de 2024, 17:33'. The main title of the article is 'Como anúncios usam sites jornalísticos para espalhar desinformação em saúde e curas falsas'. Below the title, the names of the authors are mentioned: 'Por Bianca Bortolon, Bruno Fávero, Milena Mangabeira, Rhenan Bartels e Yvna Sousa'. At the bottom right, there's a logo for 'fátima' with a small circular icon.

aos fatos: 10+

apoie +



Méuri Elle

26 de novembro de 2024, 17:33

Como anúncios usam sites jornalísticos para espalhar desinformação em saúde e curas falsas

Por Bianca Bortolon, Bruno Fávero, Milena Mangabeira, Rhenan Bartels e Yvna Sousa

fátima

Reportagem publicada por , em 2024, de autoria de Bianca Bortolon, Bruno Fávero, Milena Mangabeira, Rhenan Bartels e Yvna Sousa

EXEMPLO DE REPORTAGEM

The screenshot shows a GitHub repository page for 'aosfatos/check-up'. The repository is public and has 1 branch and 0 tags. The most recent commit was made by 'rhenanbartels' 4 months ago, with 221 commits. The commits are listed in a table:

Commit	Message	Date
docker	Revert "Update infra names"	last year
ilm	Set new ILM model	last year
plays	Close RBS banner	4 months ago
plog	Fix logging	last year
spiders	Add ILM analysis	
tests	Move crawlers and scraper to the same folder le...	
utils	Move crawlers and scraper to the same folder le...	
.dockerignore	First commit	

The 'About' section indicates 'No description, website, or topics provided.' It also lists repository details: Readme, View license, Activity, Custom properties, 12 stars, and 1 watching.

A callout box highlights the following text:

O código fonte usado para a coleta e análise automatizada dos dados pode ser encontrado no [Github do Aos Fatos](#). Para a coleta usada neste relatório, o servidor foi georreferenciado em São Paulo. Como anúncios programáticos levam em consideração a localização e o histórico dos usuários para decidir qual publicidade exibir, outras configurações levariam a resultados diferentes.

Reportagem publicada por [aos fatos](#), em 2024, de autoria de Bianca Bortolon, Bruno Fávero, Milena Mangabeira, Rhenan Bartels e Yvna Sousa

EXEMPLO DE REPORTAGEM

The screenshot shows a news article from the website PÚBLICA. The header features a yellow bar with the PÚBLICA logo and navigation links. Below the header is a large photograph of several men in military-style uniforms standing near large logs, some of which have markings like '615'. A 'REPORTAGEM' box is overlaid on the photo. The main title of the article is 'Prefeito infrator: políticos multados por danos ambientais disputam votos em 2024'. To the right of the title is a purple donation box with the text 'Leia de graça, retribua com uma doação.' and a 'Doe aqui!' button. On the left side of the photo, there are social media sharing icons for Facebook, Twitter, LinkedIn, WhatsApp, and Telegram. At the bottom of the page, there is a footer with publication details and a note about the article's age.

PÚBLICA

APOIE

REPORTAGEM

Prefeito infrator: políticos multados por danos ambientais disputam votos em 2024

Leia de graça, retribua com uma doação.
Doe aqui!

f
b
in
w
t
i

14 de agosto de 2024
12:00

Este artigo tem mais
de 1 ano

Reportagem publicada por [apublica](#), em 2024, de autoria de Bianca Muniz, Bruno Fonseca, Rafael Oliveira

EXEMPLO DE REPORTAGEM

The screenshot shows a GitHub repository page for 'prefeitos_multados'. The repository is public and has one branch ('main') and no tags. It contains several files: 'code', '.gitignore', 'README.md', 'multas_prefeitos (1).xlsx', and 'prefeitos_multados.Rproj'. The 'code' file has an 'analise.R' script. The repository was last updated 'last year' by user 'biamuniz'. The commit history shows 7 commits. The README file is present but empty.

About

Metodologia da reportagem: Prefeito infrator: políticos multados por danos ambientais disputam votos em 2024

apublica.org/2024/08/prefeito-infrator...

Readme
Activity
Custom properties
1 star
0 watching
0 forks
Report repository

Releases

Metodologia da reportagem: Prefeito infrator:

Reportagem publicada por [apublica](#), em 2024, de autoria de Bianca Muniz, Bruno Fonseca, Rafael Oliveira

DISCUSSÃO

REFLEXÕES E DESAFIOS

Alguns desafios

- Cultura...
- Tempo...
- Dedicação para aprender novas ferramentas...

OBRIGADO !

Obrigado!



REDE BRASILEIRA DE REPRODUTIBILIDADE

[Sobre nós](#)[Membros](#)[Notícias](#)[Atividades](#)[Recursos](#)[Participe!](#)

PT

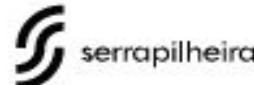
Reunimos aqui uma série de recursos para quem está interessado em reproduzibilidade. Os materiais incluídos estão classificados por idioma, conteúdo e tipo - use os filtros abaixo para visualizar apenas conteúdos relevantes para você. Buscamos manter essa lista sempre atualizada, mas não conseguimos encontrar todos os recursos interessantes que existem. Se quiser recomendar um recurso, por favor escreva para redereprodutibilidade@gmail.com.

Filtrar por Conteúdo

[acesso aberto](#) [ciência aberta](#) [dados abertos](#) [desenho experimental](#) [estatística](#) [integridade científica](#) [metaciência](#)
[pesquisa colaborativa](#) [reproduzibilidade](#) [reproduzibilidade computacional](#) [reproduzibilidade de métodos](#)

Filtrar por Idioma

**guia de boas práticas
em ciência aberta e
reprodutível**



<http://reprodutibilidade.org>

Slides: <https://osf.io/tk5dp>

<http://bit.ly/guiacienciaaberta>