

Dados abertos: o que são e o que eu tenho a ver com isso?

Melina de Souza Leite

Doutoranda e Cientista de Dados
Departamento de Ecologia
IB-USP

DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)



DADOS ABERTOS

o que são  onde vivem
o que comem  como se
reproduzem 

O que eu tenho a ver com isso?

DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)

O que são?



Dados na ciência

- ⊙ Dados são a **base** de quase toda pesquisa ecológica hoje em dia
- ⊙ Diferentes natureza, formatos e tamanhos
- ⊙ São também **produto da sua pesquisa** – Orgulhe-se deles!

Essay

Forgotten treasures: the fate of data in animal behaviour studies

Daniel S. Caetano ^{a,*}, Anita Aisenberg ^{b,1} *Animal Behaviour* 98 (2014) 1–5

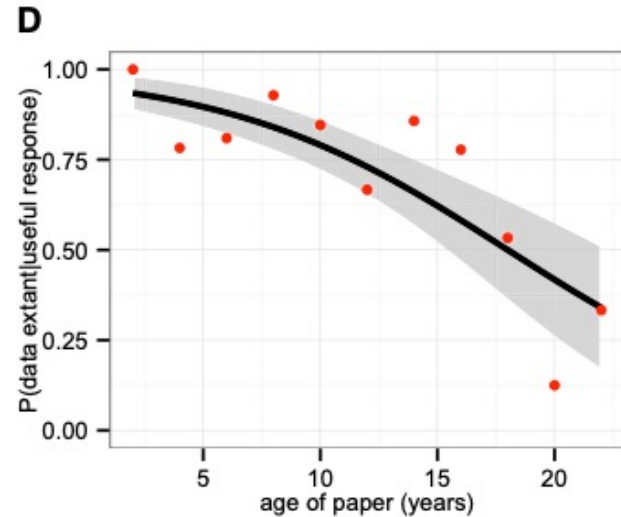
DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)

Vida útil ou longevidade dos dados

The Availability of Research Data Declines Rapidly with Article Age

[Vines et al. 2014](#). Curr. Biol.

- ⊙ Autores são maus gestores dos seus próprios dados
- ⊙ Depois de anos, são incapazes de disponibilizar os dados, mesmo que queiram!
 - por perda de dados
 - arquivos em medias/formatos inacessíveis



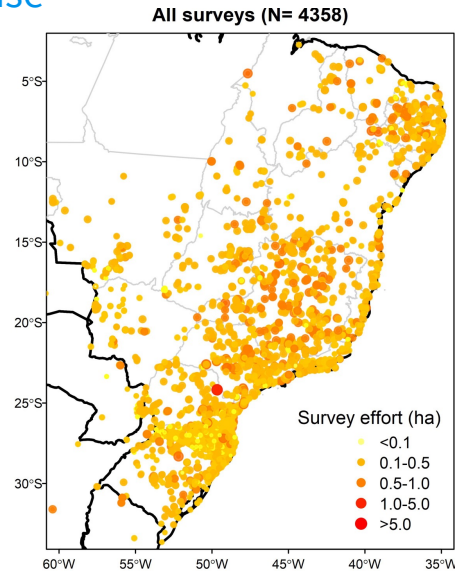
DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)

Resgate de dados antigos

- ◎ Data rescue – diferentes iniciativas em várias áreas (ex. [Bledsoe et al. 2022](#))
- ◎ [Neotropical Tree Community Database](#)



Renato Lima
Alexandre Adalardo
Paulo Inácio Prado



DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)

Por que precisamos de dados abertos?

- ⦿ Fraudes em criação de dados. ex: [#Pruittdata](#)
- ⦿ Retratação de muitos artigos -> **colaboradores** também envolvidos

SCIENTIFIC INTEGRITY

Prominent spider biologist spun a web of questionable data

Two retractions spark scrutiny of dozens of papers

Embattled spider biologist resigns university post

Ecologist Jonathan Pruitt agrees to leave McMaster University, but reasons are still undefined

12 JUL 2022 · 5:30 PM · BY [ELIZABETH PENNISI](#)



Nick DiRienzo
@Niku_DiRienzo · Follow

I may have missed this happening, but it appears Jonathan Pruitt's dissertation has been withdrawn. Not sure when, but Wikipedia was edited early Oct. So is [@McMasterU](#)'s Canada 150 chair not actually a doctor anymore? [#pruittdata](#) [#pruittgate](#).

DOCTO

Sociality in the Spider *Anelosimus studiosus*: Behavioral Correlates and Adaptive Consequences

[Jonathan N. Pruitt](#), *University of Tennessee - Knoxville*

DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)

FAIR Data Principle

FAIR Data Principle (2016)

- Findable / “encontrável”
 - Accessible / acessível
 - Interoperable / interoperável
 - Reusable / reutilizável
- ◎ Para garantir:
- transparência,
 - reprodutibilidade e
 - reusabilidade
 - confiança
- ◎ Compartilhar, ter crédito e reutilizar

SCIENTIFIC DATA 

Amended: Addendum

OPEN

SUBJECT CATEGORIES

- » Research data
- » Publication characteristics

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*[#]

DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)

Onde vivem?



Onde vivem os SEUS dados?

⦿ Dados em cativeiro

- Computadores pessoais
- HDs externos, pendrive, disquete
- Papel – caderno de campo/lab



⦿ Dados abertos

- Repositórios
- Data papers/packages
- Material suplementar
- Sites pessoais/institucionais



DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)



Riscos de dados em cativeiro

- ⦿ Cadê o backup??
- ⦿ **Perda** por:
 - morte ou roubo
 - Apagar arquivo/pasta não intencionalmente
 - Formato incompatível
- ⦿ Esquecimento: Qual a **versão** mais atual?
 - dados_vesao2.xlsx;
 - dados_versaofinal_final3_aMissao.xlsx
- ⦿ Falta de **metadados**:
 - Achar que vai se lembrar de tudo depois de 6 meses sem ver os dados...

DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)

Repositórios de dados

- ◎ Temáticos:
 - [Genbank](#), [MG rast](#), [Movebank](#), [GBIF](#)
- ◎ Genéricos/gerais:
 - [Figshare](#), [Zenodo](#), [Dryad](#)...
- ◎ Institucionais
 - Da [USP](#)!
- ◎ Redes de pesquisa/pesquisadores
 - [ForestGEO](#)
 - [Amazon Tree Diversity Network](#)
 - [ForestPlots.Net](#)
 - [bioTIME](#)

DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)





for animal tracking data

Search **Advanced Search**

chironectes minimus

All Sensor Types

Only studies where I can see data

Search

Search result

Sort by Animal Identifier

<input checked="" type="checkbox"/>	Spatial ecology of the water opossum (data from Leite et al. 2016)	i Q
<input type="checkbox"/>	F1-ALE, [n=35], Chironectes minimus	i Q
<input type="checkbox"/>	F2-FLOR, [n=66], Chironectes minimus	i Q
<input type="checkbox"/>	F3-MARG, [n=16], Chironectes minimus	i Q
<input type="checkbox"/>	F4-PACO, [n=27], Chironectes minimus	i Q
<input type="checkbox"/>	M1-GODO, [n=60], Chironectes minimus	i Q
<input type="checkbox"/>	M2-MALA, [n=19], Chironectes minimus	i Q
<input type="checkbox"/>	M3-PANQ, [n=54], Chironectes minimus	i Q
<input type="checkbox"/>	M4-RIVA, [n=28], Chironectes minimus	i Q
<input type="checkbox"/>	M5-ROBI, [n=35], Chironectes minimus	i Q
<input type="checkbox"/>	M6-TONI, [n=17], Chironectes minimus	i Q

Select Zoom Options Link Google Maps Open Layers

Brasil
Rio de Janeiro
São Paulo
Paraguay
Asunción
Uruguay
Buenos Aires

[Link aqui](#)



- [Spatial ecology of the water opossum from Leite et al. 2016](#)
[Contact owner](#)
[Show Movement Data on Map](#)
[Open in Studies Page](#)

When using this dataset, please cite the original article. [Link aqui](#)

Leite MS, Galliez M, Queiroz TL, Fernandez FAS (2016) Spatial ecology of the water opossum *Chironectes minimus* in Atlantic Forest streams. *Mammalian Biology*. doi:10.1016/j.mambio.2016.05.003

Additionally, please cite the Movebank data package:

Leite MdS, Galliez M, Queiroz TL, Pinto de Mendonça P, Fernandez FAS (2016) Data from: Spatial ecology of the water opossum *Chironectes minimus* in Atlantic Forest streams. Movebank Data Repository. doi:10.5441/001/1.4rr97k10

Repositórios genéricos e ligação com GitHub

PNAS [2020](#)

Global tree-ring analysis reveals rapid decrease in tropical tree longevity with temperature

Giuliano Maselli Locosselli^{a,b,1}, Roel J. W. Brienen^c, Melina de Souza Leite^a, Manuel Gloor^c, Stefan Krottenthaler^d, Alexandre A. de Oliveira^a, Jonathan Barichivich^a, Dieter Anhuf^d, Gregorio Ceccantini^a, Jochen Schöngart^e, and Marcos Buckeridge^{a,1}



Browse

Search or

melina-leite-locosselli_et_al_PNAS-84918f2
model_output
data
.gitattributes
.gitignore
DATA_Locosselli_et_al_2020_PNAS.xlsx
README.md
REFERENCES_DATA_Locosselli_et_al_2020_PNAS.docx
dataAnalysis.Rmd
dataAnalysis.html

[Link](#)

DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)

Data Availability Statement

The dataset of tropical trees is available at Figshare (DOI: [10.6084/m9.figshare.13119842](https://doi.org/10.6084/m9.figshare.13119842)) together with the R codes for the two models and their respective diagnostics and results. The dataset from extropical trees is available at the International Tree-Ring Data Bank (ITRDB) (<https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets/tree-ring>). Refer to the *SI Appendix* for detailed information on ITRDB data used in this study.

PrePrints e ligação com repositórios de dados/análises

EcoEvoRxiv Preprints

Submit a Preprint Search Donate Sign Up

Matrix quality determines the strength of habitat loss filtering on bird communities at the landscape scale

AUTHORS
Melina de Souza Leite, Andrea Larissa Boesing, Jean Paul Metzger, Paulo Inácio Prado

AUTHOR ASSERTIONS
Conflict of Interest: No Public Data: Available Preregistered

[Link aqui](https://zenodo.org/record/6511911)

zenodo

May 2, 2022

melina-leite/analysis_matrixBirds: revised version after paper review

Melina de Souza Leite

Revised version with more analysis suggested by reviewers of the manuscript. Main changes:

- Descriptive analysis of the matrix land uses (matrix composition) for the landscapes in high- and low-quality matrix regions.
- Including trait Hand-wing index as an important trait for dispersal ability.
- Robustness analysis of the combined models with only the species that occur at both high- and low-quality matrix landscapes. New appendix 4 with the results.

76 views 4 downloads

Available in

Indexed in

GitHub

OpenAIRE

Publication date: May 2, 2022

DOI: [10.5281/zenodo.6511911](https://doi.org/10.5281/zenodo.6511911)

Related identifiers: https://github.com/melina-leite/analysis_matrixBirds/tree/v2.0

License (for files): Other (Open)

Preview

analysis_matrixBirds-v2.0.zip

- melina-leite-analysis_matrixBirds-94cac55
 - Rprofile 26 Bytes
 - gitattributes 66 Bytes
 - gitignore 41 Bytes
 - 0_data_preparation.html 2.5 MB
 - 1_forest_cover_appendix2.html 4.1 MB
 - 2_modeling_traits_separate.html 3.9 MB
 - 3_modeling_traits_combined.html 2.2 MB
 - 4_models_diagnostic_appendix3.html 5.3 MB
 - 5_sameSpecies_appendix4.html 3.5 MB
 - README.md 2.1 kB
 - appendices_supplinfo
 - 1_appendix1 docx 170.2 kB
 - 1_appendix1 pdf 272.1 kB
 - 2_appendix2 pdf 492.2 kB
 - 3_appendix3 pdf 1.1 MB

DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)

Repositório USP

- Professores podem solicitar acesso para projetos
- Qualquer um pode buscar dados



Universidade de São Paulo
Brasil

Página inicial

Repositório USP

Na ciência da atualidade, a gestão adequada de dados científicos é fundamental para manter a integridade, eficiência e reprodutibilidade das pesquisas.

O Repositório de Dados Científicos da Universidade de São Paulo é um serviço oferecido aos docentes e pesquisadores para que seus dados de pesquisa possam ser armazenados, organizados e se tornem acessíveis ao público.

Para mais informações sobre gestão de dados científicos, acesse: <http://prp.usp.br/gestao-de-dados-cientificos/>

Navegar no Repositório USP

- [USP](#)

Buscar no Repositório USP

Ir

Navegar

Todo o repositório

[Comunidades e Coleções](#)

[Por data do documento](#)

[Autores](#)

[Títulos](#)

[Assuntos](#)

O Que é Gestão de Dados Científicos? +

Quais São os Princípios FAIR? +

Por que disponibilizar publicamente seus dados científicos? +

O Que é um Plano de Gestão de Dados? +

Como fazer um Plano de Gestão de Dados? +

Como acessar a infraestrutura disponibilizada pela USP para publicação dos dados científicos? +

O que a USP tem feito em relação à gestão de dados científicos? +

DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)

Data journals and Data papers

SOBREVIVENDO NA CIÊNCIA

Um pequeno manual para a jornada do cientista

2017

11/09/2017 POR MARCO

O que é um data paper?

ECOLOGY
ECOLOGICAL SOCIETY OF AMERICA

ATLANTIC: Data Papers from a biodiversity hotspot

[Collections](#) | First published: 30 April 2018 | Last updated: 23 November 2021

ORGANIZERS:

Mauro Galetti and Milton Cezar Ribeiro, Universidade Estadual Paulista (UNESP), Brazil

[Data Papers da Mata Atlântica](#)

DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)

Dados em material suplementar de artigos publicados?

- Também possível, mas não é mais aceito por muitas revistas
- Pouco útil para formatos de dados não tabulares (fotos, vídeo, áudio) ou bancos de dados grandes.

IS *OECOLOGIA AUSTRALIS* PROMOTING GENDER EQUALITY IN ITS REVIEW PROCESS?

Camila dos Santos de Barros^{1}, Nuria Pistón¹, Ana Cláudia Delciellos² & Melina de Souza Leite³*

Oecologia Australis

25(3):642–647, 2021

<https://doi.org/10.4257/oeco.2021.2503.01>

2021



material suplementar simples:

- metadados .txt
- dados .csv
- código .R

DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)

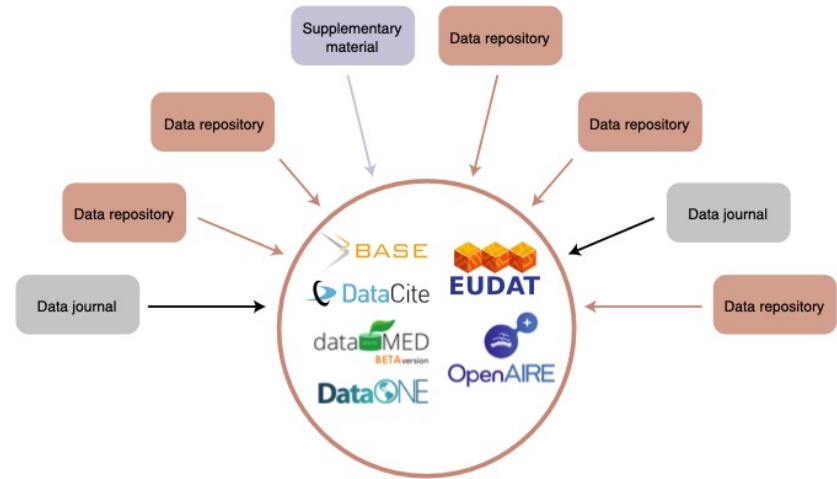
Ferramentas de busca de dados

nature
ecology & evolution 2018

Navigating the unfolding open data landscape in ecology and evolution

Antica Culina^{1*}, Miriam Baglioni², Tom W. Crowther^{1,3}, Marcel E. Visser¹,
Saskia Woutersen-Windhauer¹ and Paolo Manghi²

- [Agregadores](#) de repositórios de dados
- [Buscadores](#) de dados



Google
Dataset Search

Pesquisar conjuntos de dados



DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)

O que comem?



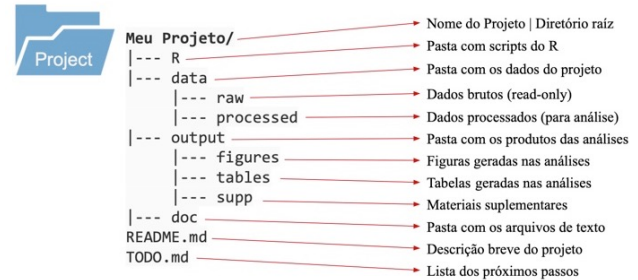
Boas práticas em organização de dados

- Estrutura de pastas/diretórios e arquivos no computador
- Controle de versão dos seus dados (GitHub, Bitbucket, GitLab)
- Estrutura de tabelas/planilhas de dados

Table 1: Examples of good and bad variable names.

good name	good alternative	avoid
Max-temp.C	MaxTemp	Maximum Temp (°C)
Precipitation_mm	Precipitation	precmm
Mean_year_growth	MeanYearGrowth	Mean growth/year
sex	sex	M/F
weight	weight	w.
cell-type	CellType	Cell type
Observation_01	first_observation	1st Obs.

[Broman & Woo 2018](#). Data organization in spreadsheets.



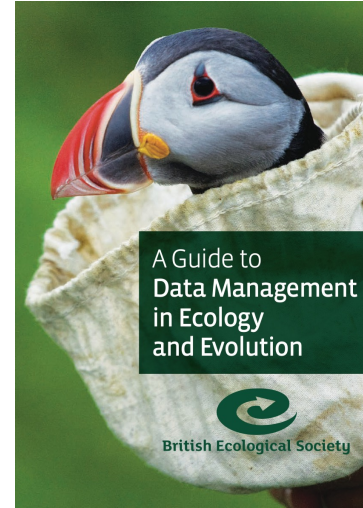
[Nakamura 2021](#) – curso GitHub pages

Ciclo de vida dos dados



Fig. 1 The data lifecycle

BES [2014](#)



DataONE

www.dataone.org

Primer on Data Management: What you always wanted to know*

* but were afraid to ask

DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)

Padronização de dados

- Diferentes áreas e tradições em ecologia -> diferentes formas de coletar e guardar dados.
- Níveis de padronização
- [Darwin core Standards \(DwC\)](#)
 - Padrão de dados de biodiversidade
 - Usado pelo GBIF

Biodiversity
Information
Standards
T D W G



DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)



“

*Muita atenção aos seus **metadados**, eles serão a primeira coisa que alguém verá ao procurar e se interessar pelos seus **dados**.*

Dmitry Schigel

Metadados – os dados sobre seus dados

- Conteúdo,
- Forma de coleta,
- Proprietários, Instituições
- Licenças,
- Financiamento
- Definição dos parâmetros/variáveis

“Curate your master data file while you create it and document it with metadata at the same time. Creating metadata afterwards is way more painful and you risk misinterpreting your own data, your memory is not usually as good as you think!”

- Ignasi Bartomeus, Swedish University of Agricultural Sciences, Sweden

[BES guide 2014](#)

DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)

Como se reproduzem?



Dados abertos – compartilhamento e publicação

PERSPECTIVE

Public Data Archiving in Ecology and Evolution: How Well Are We Doing?

Dominique G. Roche^{1,2*}, Loeske E. B. Kruuk^{1,3}, Robert Lanfear^{1,4}, Sandra A. Binning^{1,2}

[PLOS Comp. Biol. 2015](#)

- Estudos arquivados de forma que impede reuso parcial ou total
 - 2015: 64%
 - 2022: 54%

Slow improvement to the archiving quality of open datasets shared by researchers in ecology and evolution

Dominique G. Roche^{1,2,3}, Ilias Berberi¹, Fares Dhane², Félix Lauzon^{2,4},
[Proc. R. Soc. B. 2022](#)



DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)

Licenças de uso de dados

- Licenças são textos definindo legalmente como conteúdos podem ser usados, modificados e distribuídos.
- [Exemplo de licenças não restritivas](#)
 - CC0
 - CC-BY
 - MIT
- Repositórios podem ter diferentes políticas de licenças para dados, verifique-as.





Embargo

- ⦿ Até a data da publicação ou
- ⦿ Por algum tempo depois da publicação
- ⦿ Descrição dos dados/metadados ficam abertos para interessados

DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)

O que eu tenho a ver com isso?



Meu projeto e o Plano de Gestão de Dados

- ⦿ Exigências de agências de fomento
- ⦿ Planejamento de várias etapas de gestão dos dados



Gestão de Dados

Home > Financiamento à pesquisa > Gestão de Dados

DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)

Data availability statements para publicação de artigos

- Exigências de revistas
- Associação entre revistas e repositórios:
 - Dryad
 - Zenodo
 - FigshareOSF (Open Society Foundation)
- Associação entre GitHub e repositórios
 - Zenodo
 - Figshare



Living database of journal data policies in E&E

Contributors: [Ilias Berberi](#), [Dominique Roche](#)

Date created: 2021-11-15 02:15 PM | Last Updated: 2021-11-16 05:37 PM

Identifier: DOI [10.17605/OSF.IO/D6SP3](https://doi.org/10.17605/OSF.IO/D6SP3)

Category:  Data

License: [CC-BY Attribution 4.0 International](#) 

116 das 199 revistas em E&E (58%)
recomenda Data Availability Statement

DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)

Dados abertos = mais citação, mais colaboração

PLOS ONE 2020

RESEARCH ARTICLE

The citation advantage of linking publications to research data

Giovanni Colavizza^{1,2}, Iain Hrynaszkiewicz^{3,4}, Isla Staden^{1,5}, Kirstie Whitaker^{1,6},
Barbara McGillivray^{1,6*}

+ 25% em citações de artigos com Data Statement e link para dados

PeerJ 2013

Data reuse and the open data citation advantage

Heather A. Piwowar^{1,2} and Todd J. Vision^{1,2,3}

Citação aumenta com o tempo, até 30% após 5 anos

DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)

Compartilhamento e publicação dos dados

Making forest data fair and open

Data on tropical forests are in high demand. But ground forest measurements are hard to sustain and the people who make them are extremely disadvantaged compared to those who use them. We propose a new approach to forest data that **focuses on the needs of data originators, and ensures users and funders contribute properly.**

Renato A. F. de Lima, Oliver L. Phillips, Alvaro Duque, J. Sebastian Tello, Stuart J. Davies,
Alexandre Adalardo de Oliveira, Sandra Muller, Euridice N. Honorio Coronado, Emilio Vilanova,

[Nat. Eco. Evo 2022](#)

Open data: Empowering the empowered or effective data use for everyone?

[Michael B. Gurstein](#) 2011

DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)

Publicação de dados: uma nova moeda na carreira científica?

Motivating data contributions via a distinct career currency

Mark Westoby¹, Daniel S. Falster² and Julian Schrader^{1,3}

[Proc. R. Soc. B 2021](#)

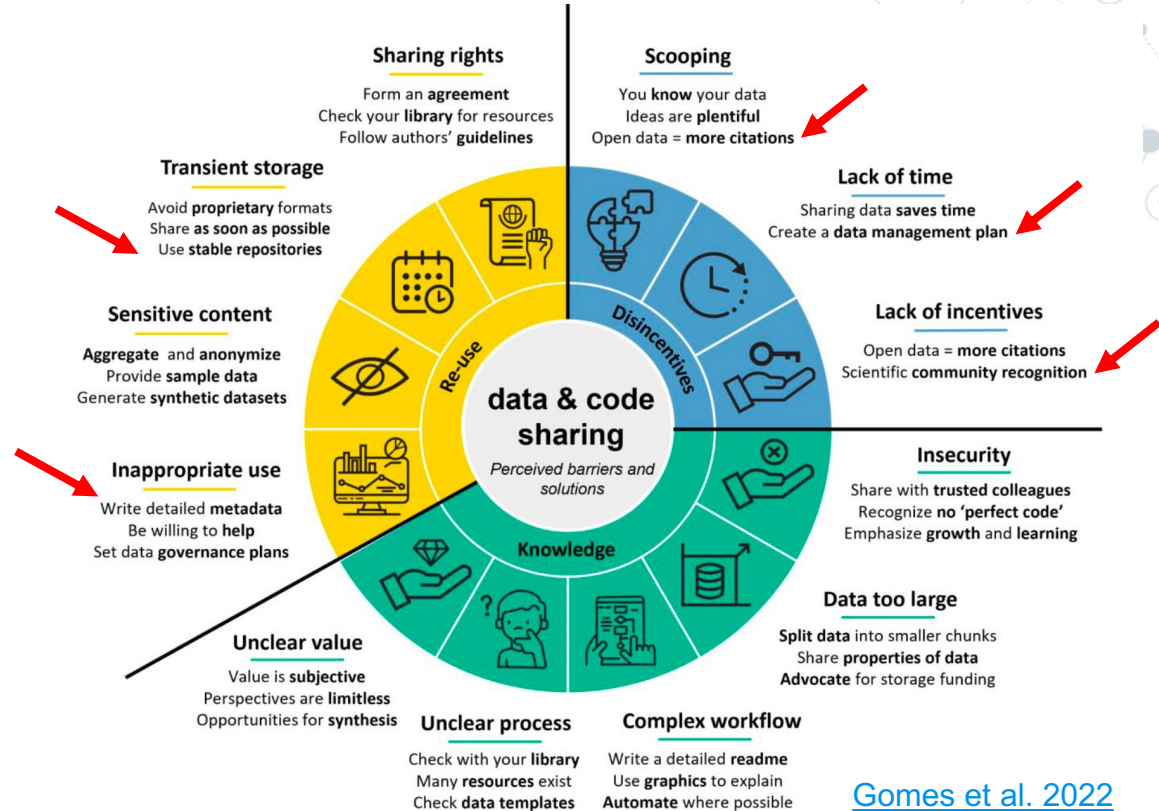
- Valoração do pesquisador em sua contribuição em coleta e publicação de dados

DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)

Por que não compartilhamos dados?

Barreiras

- falta de conhecimento
- receios a mau uso
- falta de incentivo
- falta de tempo
- falta de infra-estrutura
- falta de treinamento



Gomes et al. 2022

Enfim...



OPEN DATA IN RESEARCH



WHAT?

OPEN DATA is data that may be **used, re-used,** and **redistributed** freely by anyone

FAIR

- Findable
- Accessible
- Interoperable
- Reusable



WHY?

Increase your:

- Citations by 25%¹
- Public trust
- Exposure
- Transparency
- Reproducibility



HOW?

- 1 Prepare your data
Tidy data + meta-data
Editable and open fileformats
- 2 Pick platform
OSF, GitHub, Zenodo, Dryad, Figshare, PANGAEA and more!
- 3 Make it public and citeable!



1. Colavizza, G. et al. (2020). The citation advantage of linking publications to research data. *PLoS ONE*, 15(4), e0230416.

O que eu aprendi com o movimento da ciência/dados abertos

- ◎ A pensar com carinho sobre os dados como **produto** de um grande esforço.
- ◎ Dicas de **boa prática** em organização/manejo dos dados
- ◎ Diferentes formas de **publicar/compartilhar** dados
- ◎ Pensar em novos hábitos e pequenas mudanças que podemos fazer para tornar nossa **ciência mais aberta!**

DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)



Agradecimentos



Júlia Barreto



Paulo Inácio Prado



DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)

Muito obrigada!

melinasouzaleite.weebly.com

melina.leite@ib.usp.br

@melina_sleite



DOI: [10.5281/zenodo.6861029](https://doi.org/10.5281/zenodo.6861029)