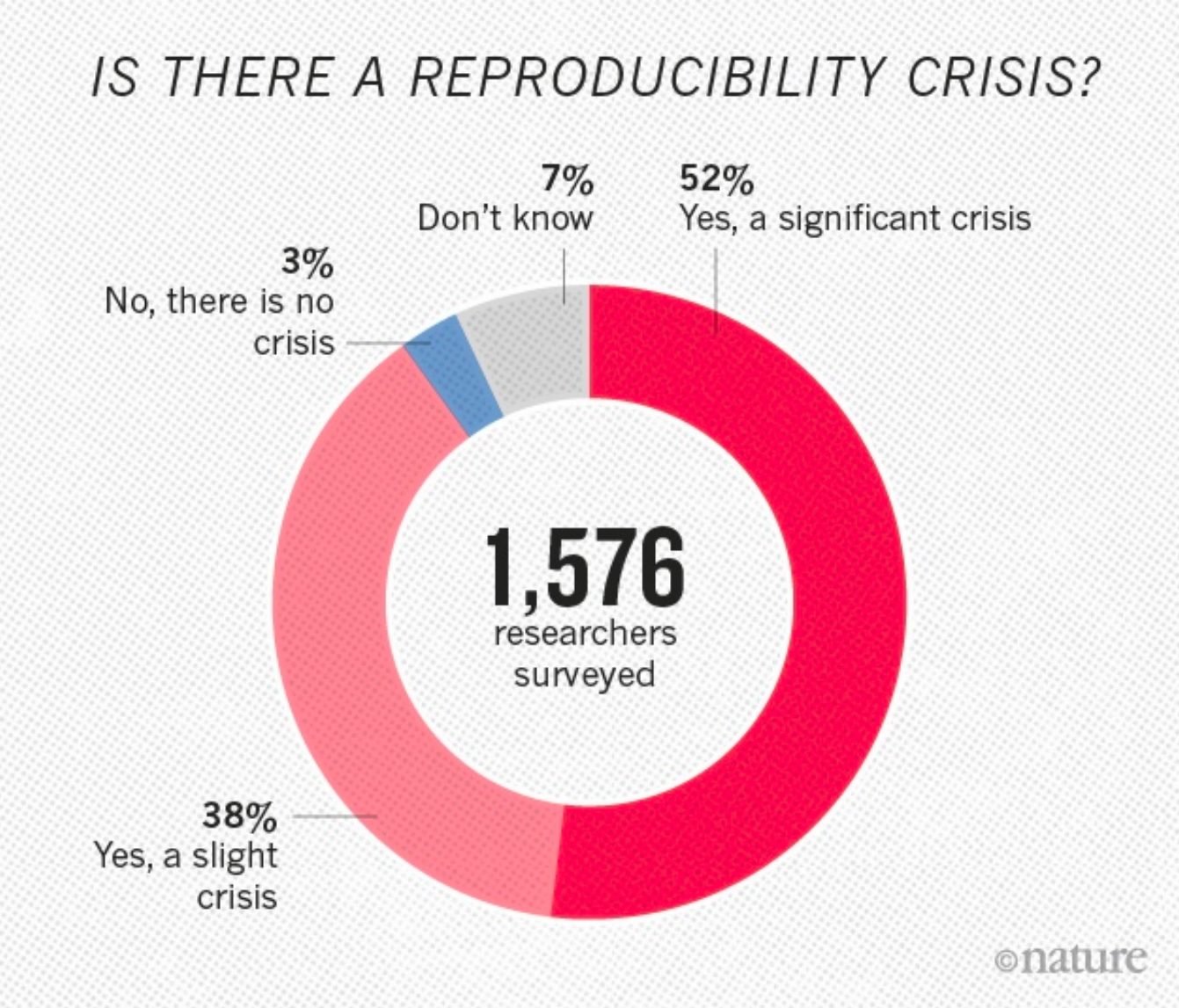
Improving the reproducibility of scientific research

Francisco Rodríguez-Sánchez

https://frodriguezsanchez.net


IS THERE A REPRODUCIBILITY CRISIS?
7%
Don't know
52%
Yes, a significant crisis
3%
No, there is no crisis
1,576
researchers surveyed
38%
Yes, a slight crisis
©nature

Baker 2016

The reproducibility crisis in mass media

≡ EL PAÍS

Materia

La ciencia vive una epidemia de estudios inservibles

Científicos de EE UU, Reino Unido y Holanda denuncian que la investigación está perdiendo parte de su credibilidad

El País

Reproducibility Project: Cancer Biology

NEWS | 09 December 2021

Half of top cancer studies fail high-profile reproducibility effort

- Goal: Replicate 193 experiments from 53 papers

Errington et al 2021

Reproducibility Project: Cancer Biology

NEWS | 09 December 2021

Half of top cancer studies fail high-profile reproducibility effort

- Goal: Replicate 193 experiments from 53 papers
- Finally: 50 experiments from 23 papers

Errington et al 2021

Reproducibility Project: Cancer Biology

NEWS | 09 December 2021

Half of top cancer studies fail high-profile reproducibility effort

- Goal: Replicate 193 experiments from 53 papers
- Finally: 50 experiments from 23 papers
- ~Half **didn't replicate** (much smaller effect sizes)

Errington et al 2021

Reproducibility Project: Cancer Biology

NEWS | 09 December 2021

Half of top cancer studies fail high-profile reproducibility effort

- Goal: Replicate 193 experiments from 53 papers
- Finally: 50 experiments from 23 papers
- ~Half **didn't replicate** (much smaller effect sizes)
- **No paper reported** all required data

Errington et al 2021

Reproducibility Project: Cancer Biology

NEWS | 09 December 2021

Half of top cancer studies fail high-profile reproducibility effort

- Goal: Replicate 193 experiments from 53 papers
- Finally: 50 experiments from 23 papers
- ~Half **didn't replicate** (much smaller effect sizes)
- **No paper reported** all required data
- **Impossible to repeat** experiments w/o contacting authors

Errington et al 2021

Reproducibility Project: Cancer Biology

NEWS | 09 December 2021

Half of top cancer studies fail high-profile reproducibility effort

- Goal: Replicate 193 experiments from 53 papers
- Finally: 50 experiments from 23 papers
- ~Half **didn't replicate** (much smaller effect sizes)
- **No paper reported** all required data
- **Impossible to repeat** experiments w/o contacting authors
- 1/3 authors **didn't respond or help**

Errington et al 2021

Sylvain Deville ❄️👨‍🎓
@DevilleSy

Trying to reproduce the results of a paper using only what's in the Methods section

Most scientific articles

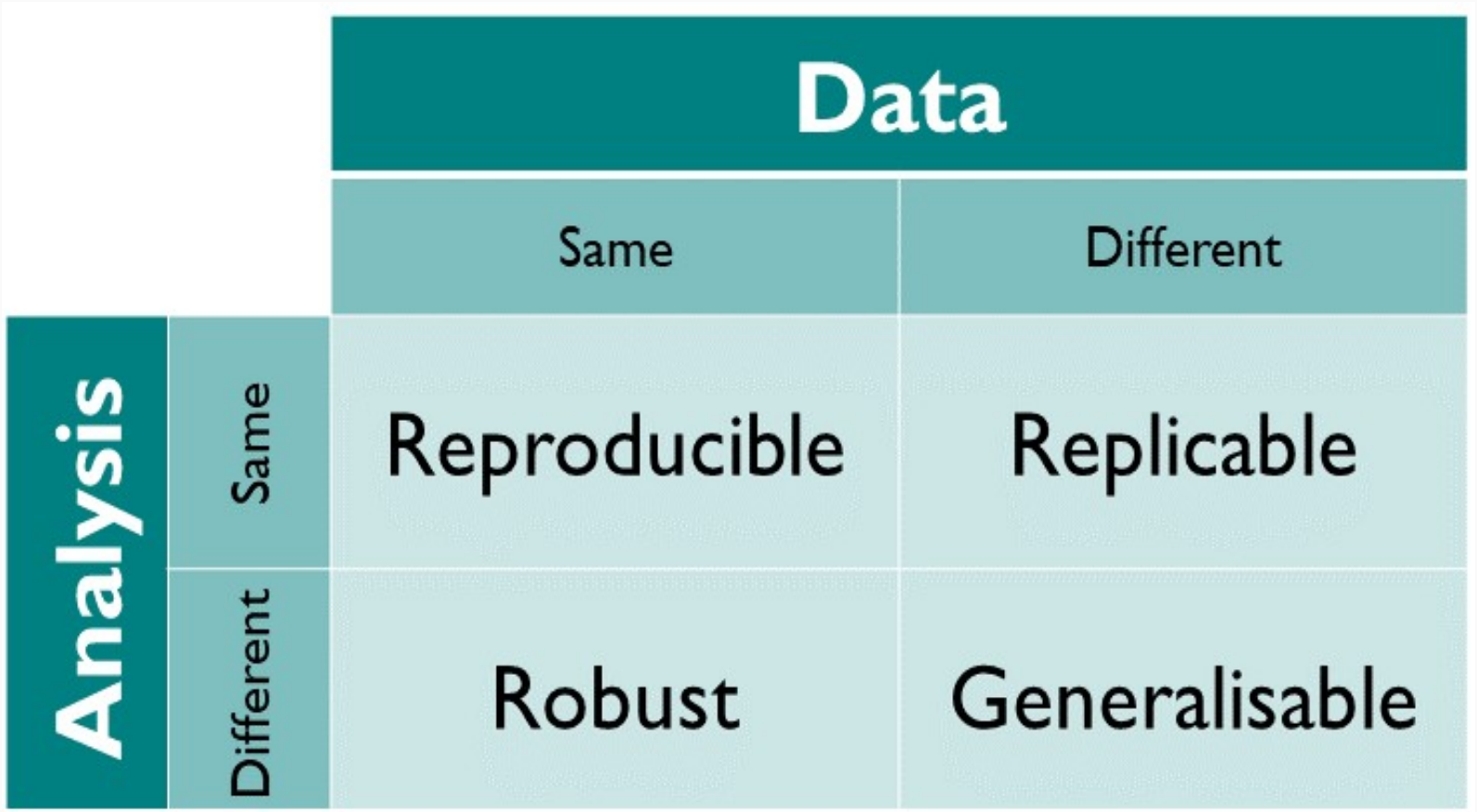
are NOT reproducible

Reproducibility

crisis —> **REVOLUTION**

What is reproducibility?

Reproducibility vs Replicability



The Turing Way

We can’t guarantee that

our results are **REPLICABLE**.

But at least

they should be **REPRODUCIBLE**.

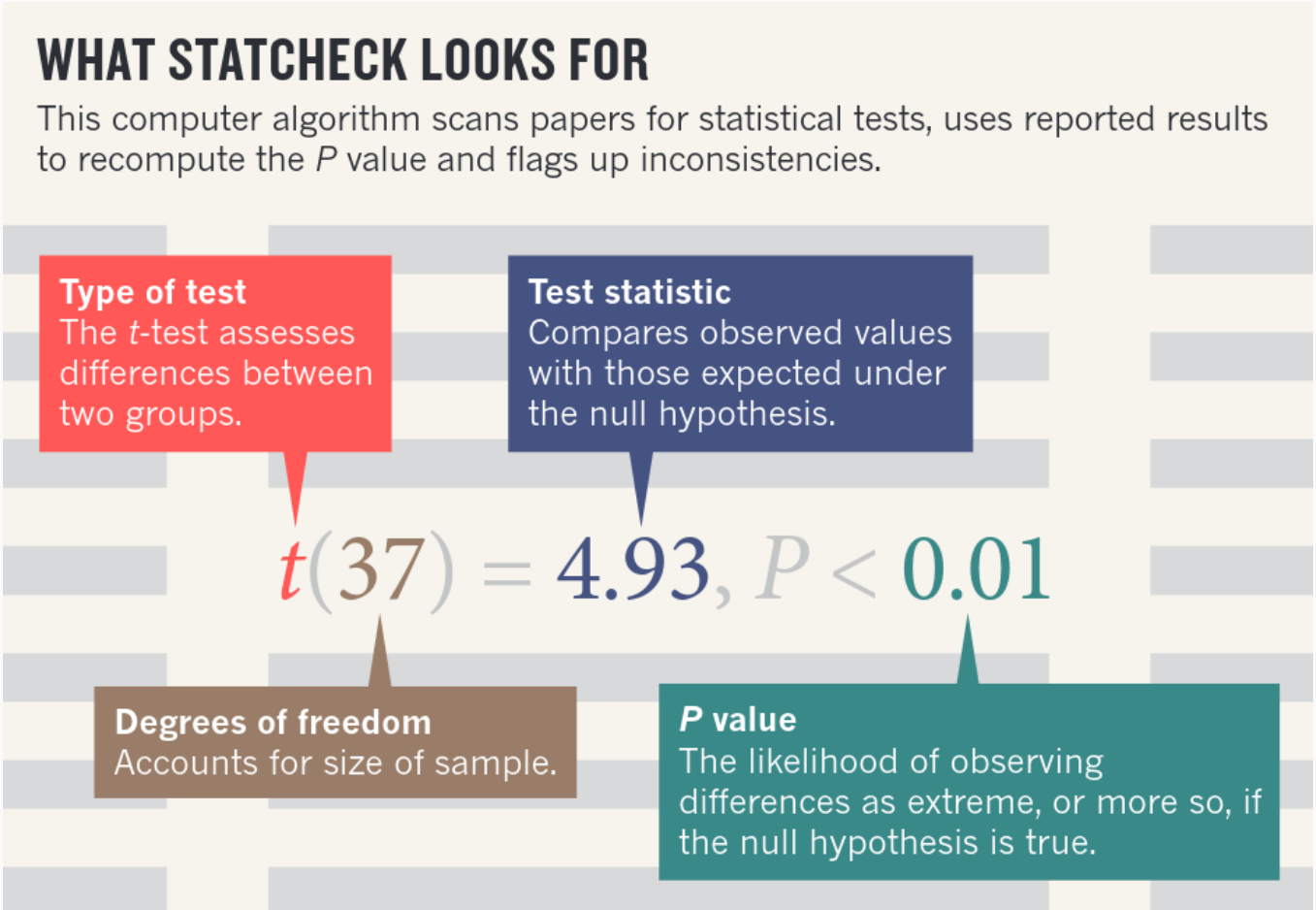
Most scientific articles

are NOT reproducible

Behav Res (2016) 48:1205–1226
DOI 10.3758/s13428-015-0664-2

The prevalence of statistical reporting errors in psychology (1985–2013)

Michèle B. Nuijten[1] · Chris H. J. Hartgerink[1] · Marcel A. L. M. van Assen[1] · Sacha Epskamp[2] · Jelte M. Wicherts[1]



Behav Res (2016) 48:1205–1226
DOI 10.3758/s13428-015-0664-2

The prevalence of statistical reporting errors in psychology (1985–2013)

Michèle B. Nuijten[1] · Chris H. J. Hartgerink[1] · Marcel A. L. M. van Assen[1] · Sacha Epskamp[2] · Jelte M. Wicherts[1]

1/2 articles: **inconsistencies** in p-values

1/8 articles: **grossly inconsistent** p-values

(affecting conclusions -> significance)

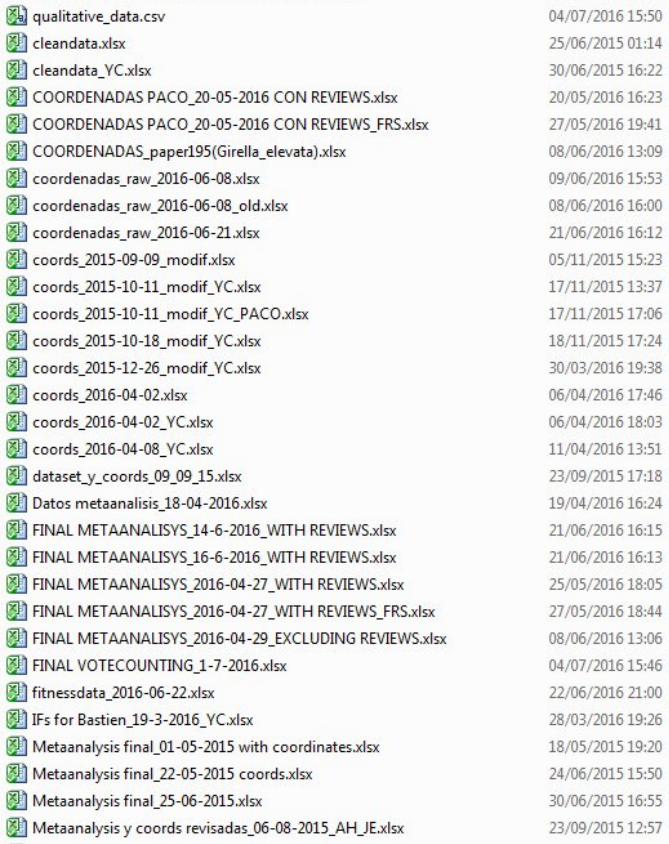
In ecology

< 20% articles are **reproducible**

Culina et al 2020

We can’t even reproduce our own work

Data/Code lost or unusable



qualitative_data.csv	04/07/2016 15:50
cleandata.xlsx	25/06/2015 01:14
cleandata_YC.xlsx	30/06/2015 16:22
COORDENADAS PACO_20-05-2016 CON REVIEWS.xlsx	20/05/2016 16:23
COORDENADAS PACO_20-05-2016 CON REVIEWS_FRS.xlsx	27/05/2016 19:41
COORDENADAS_paper195(Girella_elevata).xlsx	08/06/2016 13:09
coordenadas_raw_2016-06-08.xlsx	09/06/2016 15:53
coordenadas_raw_2016-06-08_old.xlsx	08/06/2016 16:00
coordenadas_raw_2016-06-21.xlsx	21/06/2016 16:12
coords_2015-09-09_modif.xlsx	05/11/2015 15:23
coords_2015-10-11_modif_YC.xlsx	17/11/2015 13:37
coords_2015-10-11_modif_YC_PACO.xlsx	17/11/2015 17:06
coords_2015-10-18_modif_YC.xlsx	18/11/2015 17:24
coords_2015-12-26_modif_YC.xlsx	30/03/2016 19:38
coords_2016-04-02.xlsx	06/04/2016 17:46
coords_2016-04-02_YC.xlsx	06/04/2016 18:03
coords_2016-04-08_YC.xlsx	11/04/2016 13:51
dataset_y_coords_09_09_15.xlsx	23/09/2015 17:18
Datos metaanalisis_18-04-2016.xlsx	19/04/2016 16:24
FINAL METAANALISYS_14-6-2016_WITH REVIEWS.xlsx	21/06/2016 16:15
FINAL METAANALISYS_16-6-2016_WITH REVIEWS.xlsx	21/06/2016 16:13
FINAL METAANALISYS_2016-04-27_WITH REVIEWS.xlsx	25/05/2016 18:05
FINAL METAANALISYS_2016-04-27_WITH REVIEWS_FRS.xlsx	27/05/2016 18:44
FINAL METAANALISYS_2016-04-29_EXCLUDING REVIEWS.xlsx	08/06/2016 13:06
FINAL VOTECOUNTING_1-7-2016.xlsx	04/07/2016 15:46
fitnessdata_2016-06-22.xlsx	22/06/2016 21:00
IFs for Bastien_19-3-2016_YC.xlsx	28/03/2016 19:26
Metaanalysis final_01-05-2015 with coordinates.xlsx	18/05/2015 19:20
Metaanalysis final_22-05-2015 coords.xlsx	24/06/2015 15:50
Metaanalysis final_25-06-2015.xlsx	30/06/2015 16:55
Metaanalysis y coords revisadas_06-08-2015_AH_JE.xlsx	23/09/2015 12:57

What's a reproducible manuscript?

DATA + CODE

- analysis fully **traceable**
- results can be **regenerated**

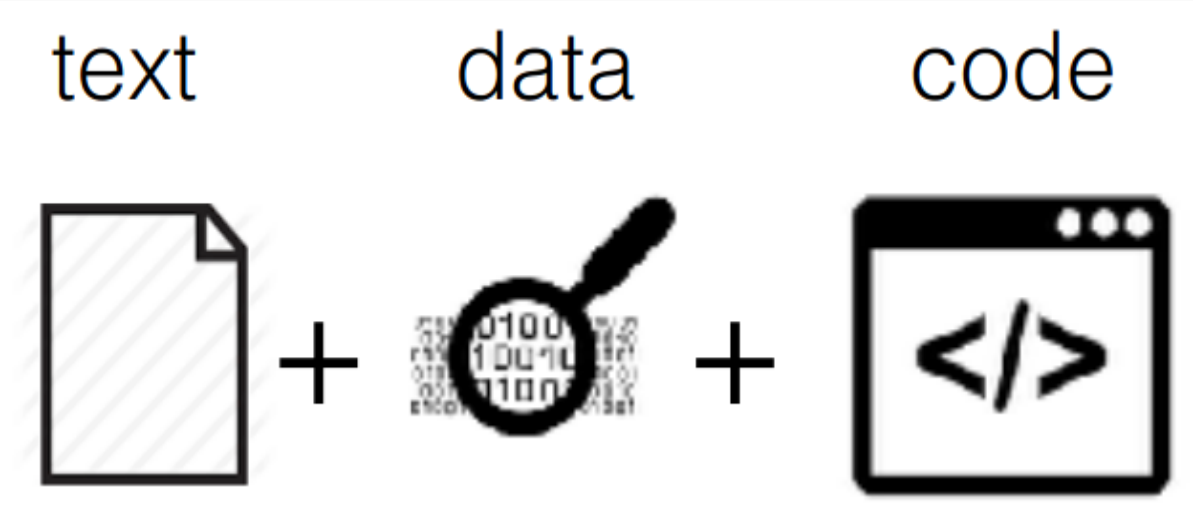
A scientific article is **advertising**, not scholarship.

The actual scholarship is the **full software environment**,

code and data, that produced the result.

Claerbout & Karrenback 1992

Reproducible article: text + data + code



Are we sharing the data?

Are we sharing data?

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]

1 Division of Evolution, Ecology and Genetics, Research School of Biology, The Australian National University, Canberra, Australian Capital Territory, Australia, **2** Éco-Éthologie, Institut de Biologie, Université de Neuchâtel, Neuchâtel, Switzerland, **3** Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, United Kingdom, **4** Department of Biological Sciences, Macquarie University, Sydney, Australia

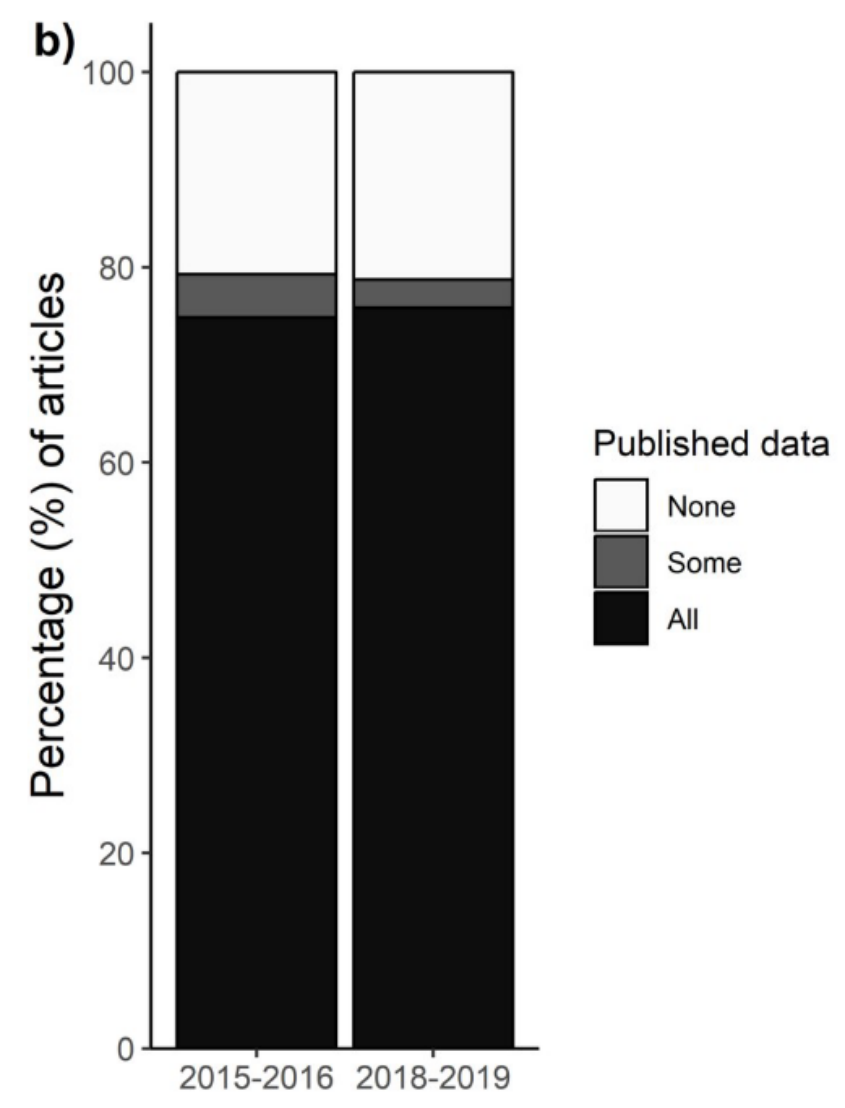
* dominique.roche@mail.mcgill.ca

Abstract

Policies that mandate public data archiving (PDA) successfully increase accessibility to data underlying scientific publications. However, is the data quality sufficient to allow reuse and reanalysis? We surveyed 100 datasets associated with nonmolecular studies in journals that commonly publish ecological and evolutionary research and have a strong PDA policy. Out of these datasets, 56% were incomplete, and 64% were archived in a way that partially or entirely prevented reuse. We suggest that cultural shifts facilitating clearer benefits to authors are necessary to achieve high-quality PDA and highlight key guidelines to help authors increase their data's reuse potential and compliance with journal data policies.

Roche et al. 2015

Are we sharing data?



Culina et al. 2020

Are we sharing data?

Quickly getting better

Scientific Life

Early Career Researchers Embrace Data Sharing

Hamish A. Campbell,[1,*] Mariana A. Micheli-Campbell,[1] and Vinay Udyawer[2]

Campbell et al. 2019

Are we sharing the code?

Code exists but rarely shared

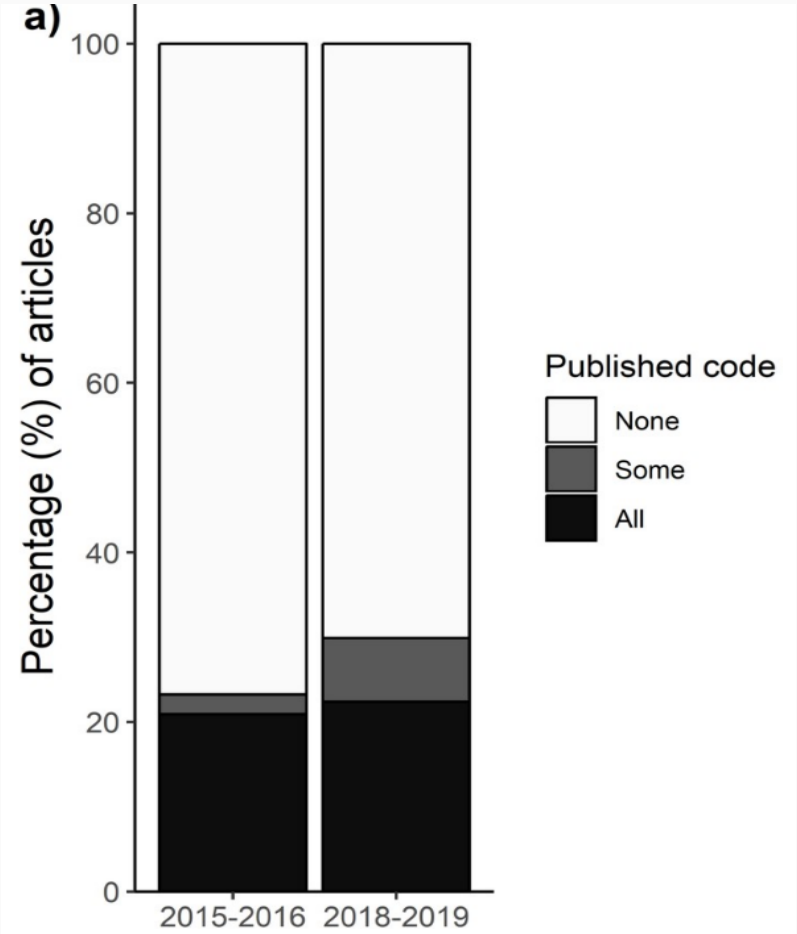
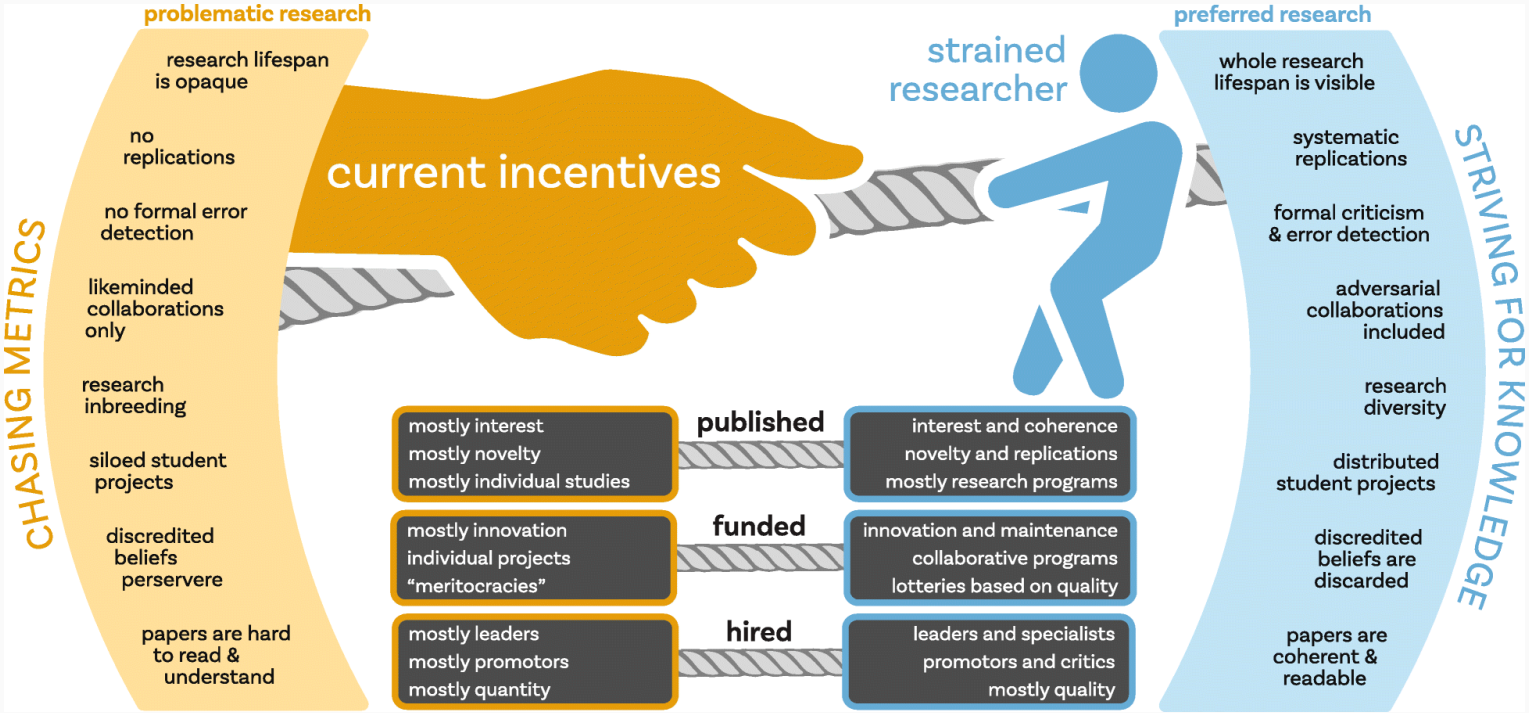


Fig 1. Code-sharing is at its infancy in ecology, wher

Culina et al. 2020

WHY?

Poor incentives



O'Dea et al 2021

Doing reproducible research can be costly

The Costs of Reproducibility

Russell A. Poldrack[1,*]
[1]Department of Psychology, Stanford University, Stanford, CA, USA
*Correspondence: poldrack@stanford.edu
https://doi.org/10.1016/j.neuron.2018.11.030

PERSPECTIVE

Open science challenges, benefits and tips in early career and beyond

Christopher Allen[1]*, **David M. A. Mehler**[1,2]*

Must value diverse contributions to reproducible research

Credit data generators for data reuse

To promote effective sharing, we must create an enduring link between the people who generate data and its future uses, urge **Heather H. Pierce** and colleagues.

Pierce et al 2019

Let’s knock down psychological barriers

Publish your computer code: it is good enough

Freely provided working code — whatever its quality — improves programming and enables others to engage with your research, says **Nick Barnes**.

Barnes 2010

- Improve training
- Code review, preprints…
- Avoid shaming -> constructive critique
- Ugly code better than no code