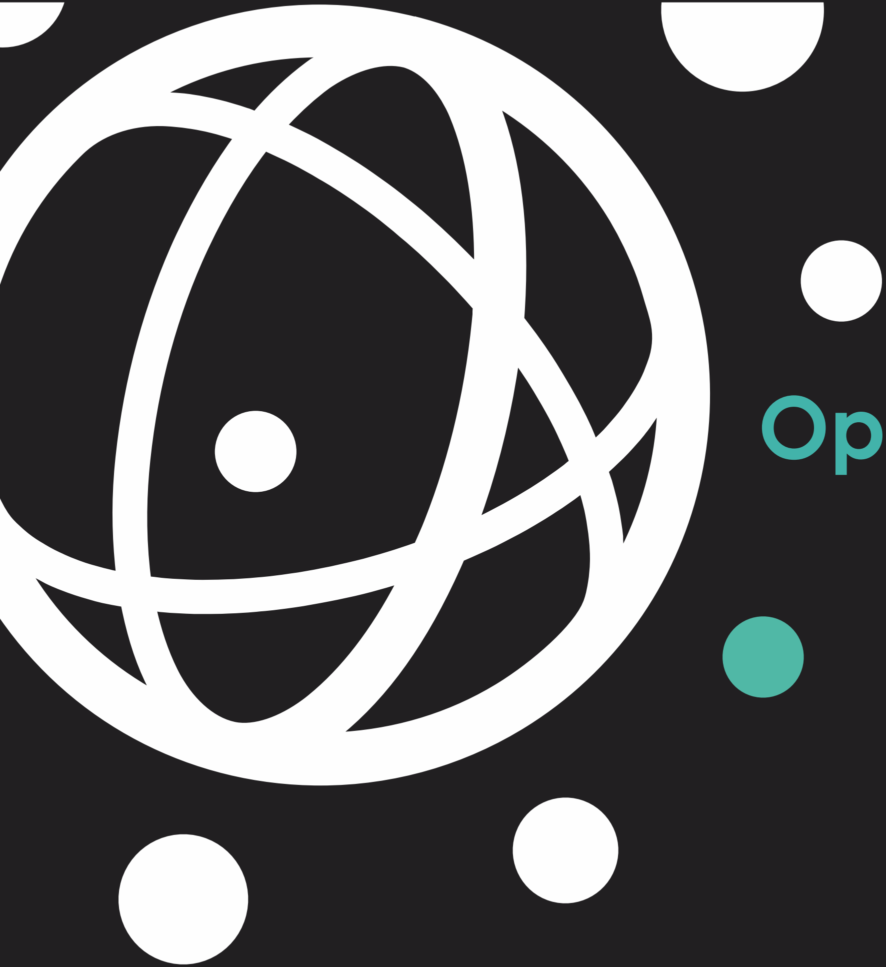


Open Data & Code Sharing
A practical guide

Sensors CDT

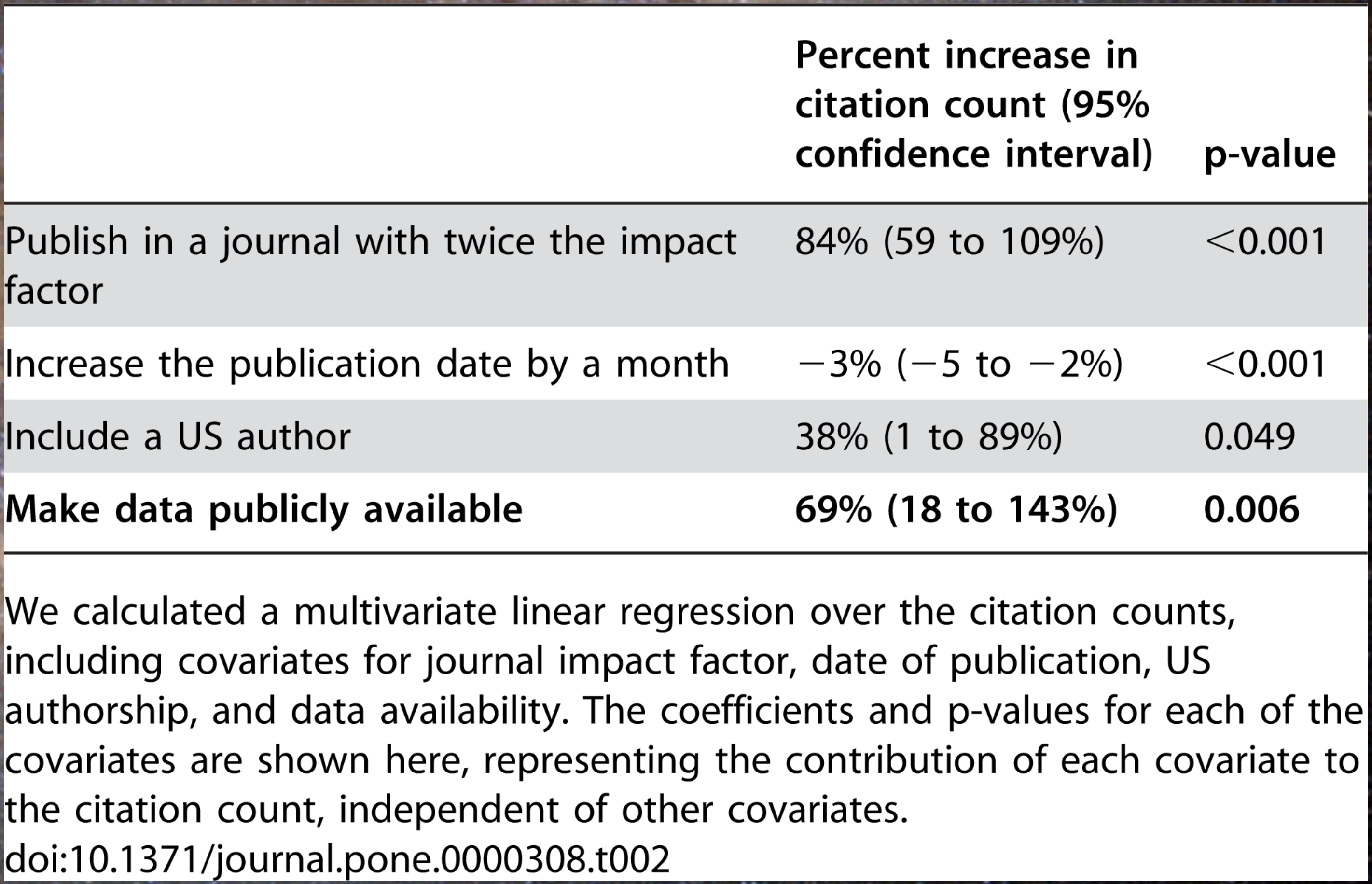
23 November 2018

Dr Leonie Mueck

Division Editor

Physical Sciences and Engineering

PLOS ONE



	Percent increase in citation count (95% confidence interval)	p-value
Publish in a journal with twice the impact factor	84% (59 to 109%)	<0.001
Increase the publication date by a month	−3% (−5 to −2%)	<0.001
Include a US author	38% (1 to 89%)	0.049
Make data publicly available	**69% (18 to 143%)**	**0.006**

We calculated a multivariate linear regression over the citation counts, including covariates for journal impact factor, date of publication, US authorship, and data availability. The coefficients and p-values for each of the covariates are shown here, representing the contribution of each covariate to the citation count, independent of other covariates.
doi:10.1371/journal.pone.0000308.t002

Piwowar *et al*. 2007 **Sharing Detailed Research Data Is Associated with Increased Citation Rate**
pone.0000308

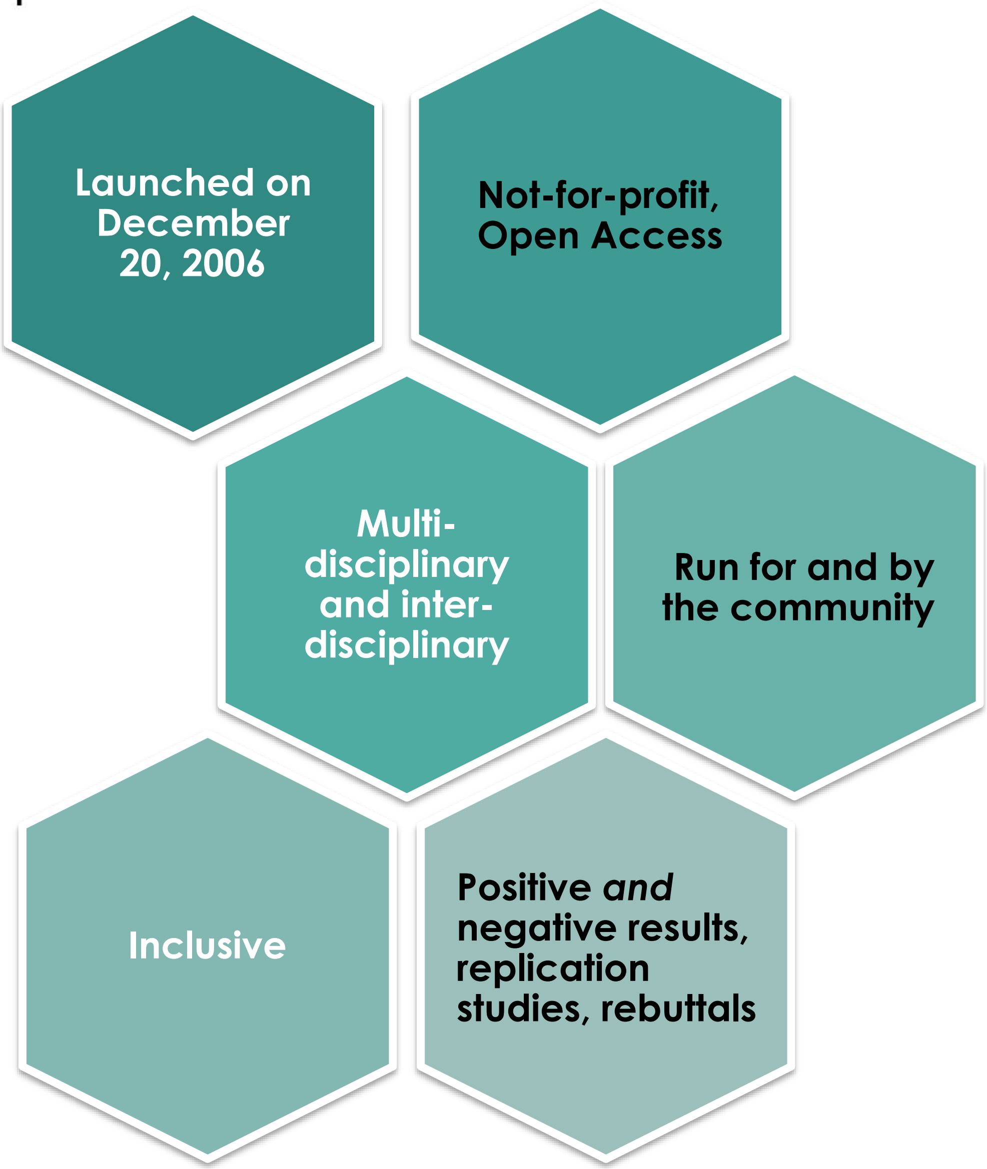
Outline

- Introducing PLOS and *PLOS ONE*
- Our Data Policy
  - What does it say?
  - How does it work in practice?
- Practical Data & Code Sharing
  - Data Repositories
  - Protocols.io
  - Code Sharing
- Open Science Innovations at PLOS

Public Library of Science (PLOS)

- Est. 2001 as a non-profit publisher and advocacy organisation with a mission to accelerate progress in science & medicine by leading a transformation in research communication
- Seven Open Access online journals covering the breadth of science, medicine, engineering and related fields

PLOS | ONE



Launched on December 20, 2006

Not-for-profit, Open Access

Multi-disciplinary and inter-disciplinary

Run for and by the community

Inclusive

Positive *and* negative results, replication studies, rebuttals

PLOS ONE

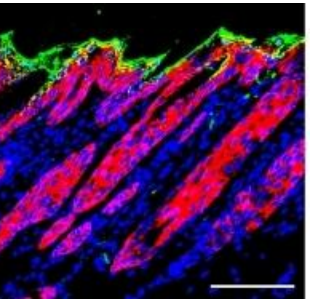
Considerations for policy implementations

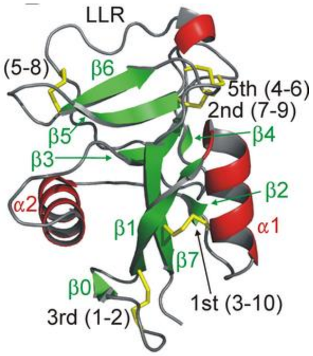
- Scale of PLOS ONE
  - 21,000 publications in 2017
  - 7500 Academic Editors
- 28 staff editors with PhD-level research experience to lead policy discussions
- Multi- and interdisciplinarity: Vastly different communities with different needs, from Social Sciences to Clinical Sciences, from Molecular Biology to Electrical Engineering

PLOS

10.1371/journal.pone.0107610

10.1371/journal.pone.0012292


10.1371/journal.pone.0017867



10.1371/journal.pone.0002616

Outline

- Introducing PLOS and *PLOS ONE*
- Our Data Policy
  - What does it say?
  - How does it work in practice?
- Practical Data & Code Sharing
  - Data Repositories
  - Protocols.io
  - Code Sharing
- Open Science Innovations at PLOS

Making data available fosters scientific progress

Data availability allows:

- Validation, replication, reanalysis, new analysis
- Reproducibility
- Increased value of research through re-use
- Easier citation of data
- Evidence that sharing data increases impact of work

journals.plos.org/plosone/s/data-availability

PLOS Data Policy – what does it say?

Since March 2014…

PLOS journals require authors to make **all data underlying the findings** described in their manuscript fully available without restriction, with rare exception.

When submitting a manuscript online, authors must provide a **Data Availability Statement** describing compliance with PLOS's policy.

Guidance for researchers on which repositories are suitable and how to share data.

PLOS

What Data?

Data underlying the findings

- Dataset used to reach the conclusions, incl. related metadata and methods, and any additional data required for replication

Where?

Preferred: Community repository

- PLOS provides list of acceptable repositories
- Authors must provide dois or accession numbers

Possible: Supplementary information and paper itself

- All supplementary information files have doi and are uploaded to figshare

PLOS

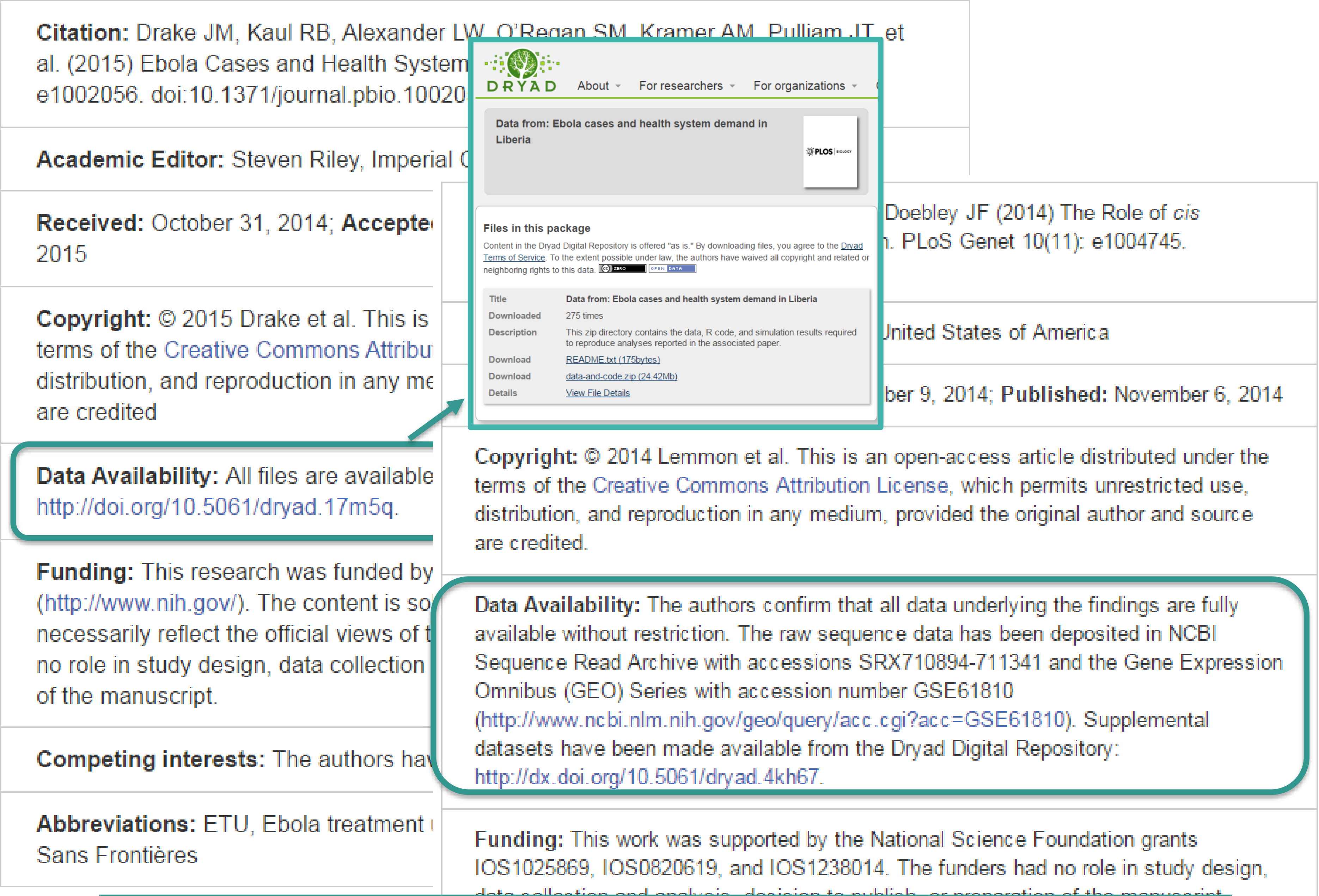
Exceptions

- **Ethical or legal reasons**, e.g., compromising patient confidentiality or participant privacy
- Data deposition could present some **other threat**, e.g., revealing the locations of fossil deposits

Examples of non-compliance

- “Available upon request” from author **without giving valid reason**
- **Proprietary data** that other researchers cannot obtain

PLOS



Citation: Drake JM, Kaul RB, Alexander LW, O'Regan SM, Kramer AM, Pulliam JT, et al. (2015) Ebola Cases and Health System
e1002056. doi:10.1371/journal.pbio.10020

Academic Editor: Steven Riley, Imperial C

Received: October 31, 2014; **Accepte**
2015

Copyright: © 2015 Drake et al. This is
terms of the Creative Commons Attribu
distribution, and reproduction in any me
are credited

Data Availability: All files are available
http://doi.org/10.5061/dryad.17m5q.

Funding: This research was funded by
(http://www.nih.gov/). The content is so
necessarily reflect the official views of t
no role in study design, data collection
of the manuscript.

Competing interests: The authors hav

Abbreviations: ETU, Ebola treatment
Sans Frontières

Data from: Ebola cases and health system demand in Liberia

Files in this package

Content in the Dryad Digital Repository is offered "as is." By downloading files, you agree to the Dryad Terms of Service. To the extent possible under law, the authors have waived all copyright and related or neighboring rights to this data.

Title	**Data from: Ebola cases and health system demand in Liberia**
Downloaded	275 times
Description	This zip directory contains the data, R code, and simulation results required to reproduce analyses reported in the associated paper.
Download	README.txt (175bytes)
Download	data-and-code.zip (24.42Mb)
Details	View File Details

Doebley JF (2014) The Role of *cis*
n. PLoS Genet 10(11): e1004745.

United States of America

ber 9, 2014; **Published:** November 6, 2014

Copyright: © 2014 Lemmon et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

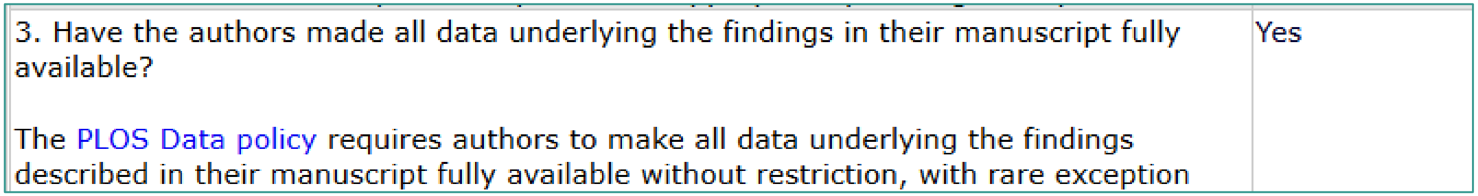
Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. The raw sequence data has been deposited in NCBI Sequence Read Archive with accessions SRX710894-711341 and the Gene Expression Omnibus (GEO) Series with accession number GSE61810 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE61810). Supplemental datasets have been made available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.4kh67.

Funding: This work was supported by the National Science Foundation grants IOS1025869, IOS0820619, and IOS1238014. The funders had no role in study design,

The data availability statement is openly available, and machine-readable as part of the PLOS search API

Data availability checks

- **In-house checks on basic compliance**:
  - Does data availability statement comply with policy?
  - Are there some files available?
- **Academic Editors and Referees:**
  - What constitutes a "data underlying the findings" in any given case?



3. Have the authors made all data underlying the findings in their manuscript fully available? The PLOS Data policy requires authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception	Yes

PLOS

>100,000

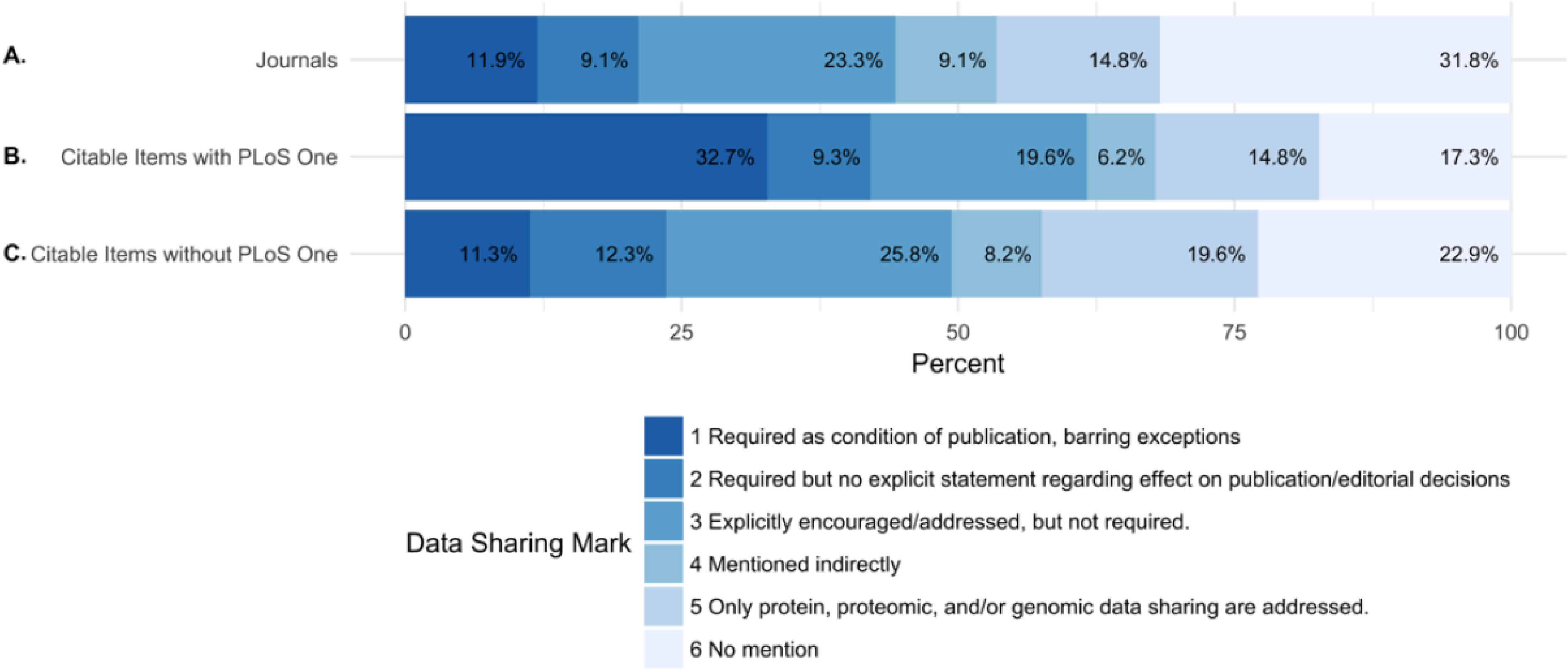
papers published with a data statement at PLOS

<0.1%

of submissions rejected due to authors' unwillingness or inability to share data

PLOS

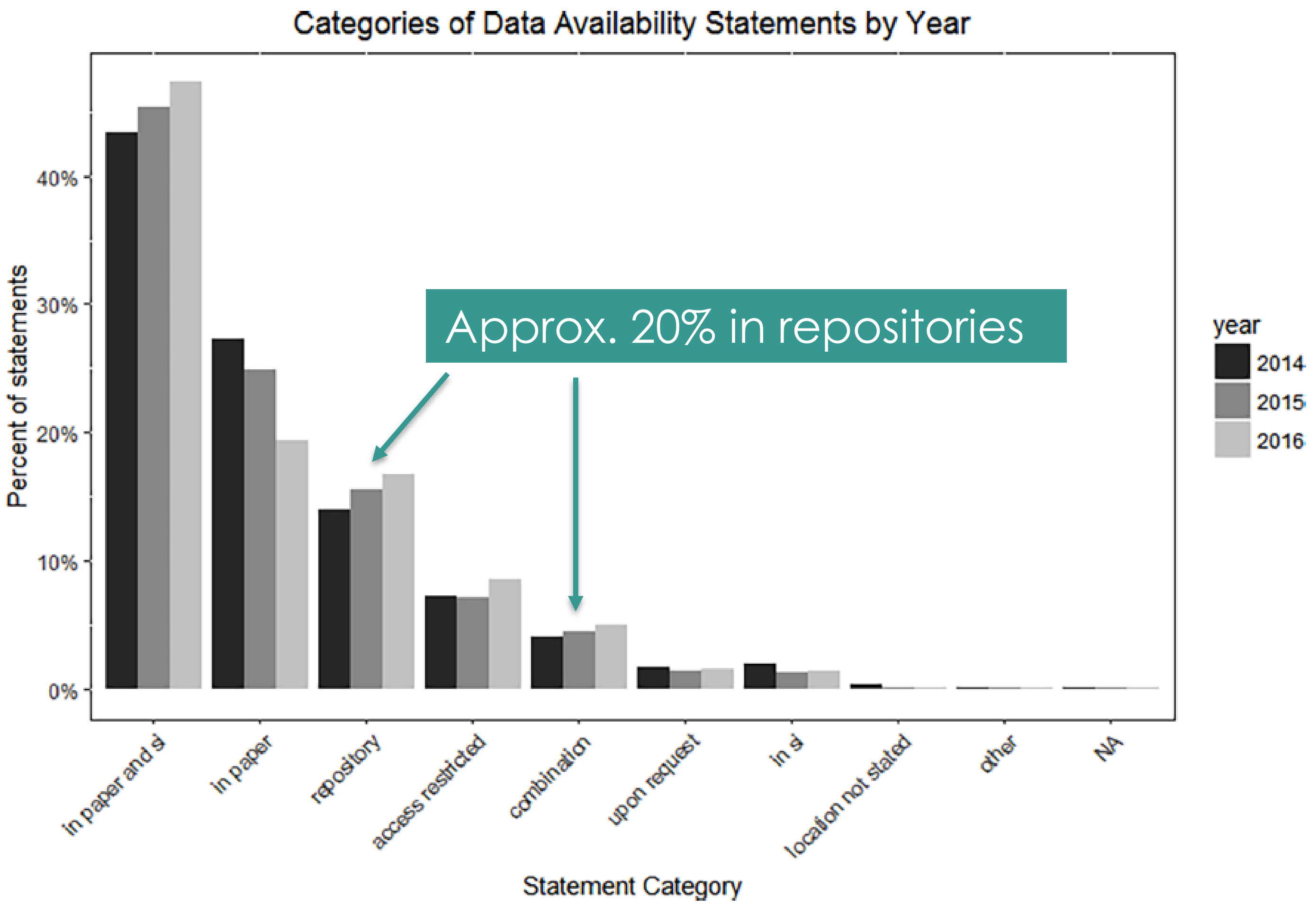
Data sharing policies make a difference



PLOS

Vasilevsky *et al*. *PeerJ*, DOI 10.7717/peerj.3208 (2017)

Data availability statements 2014-16

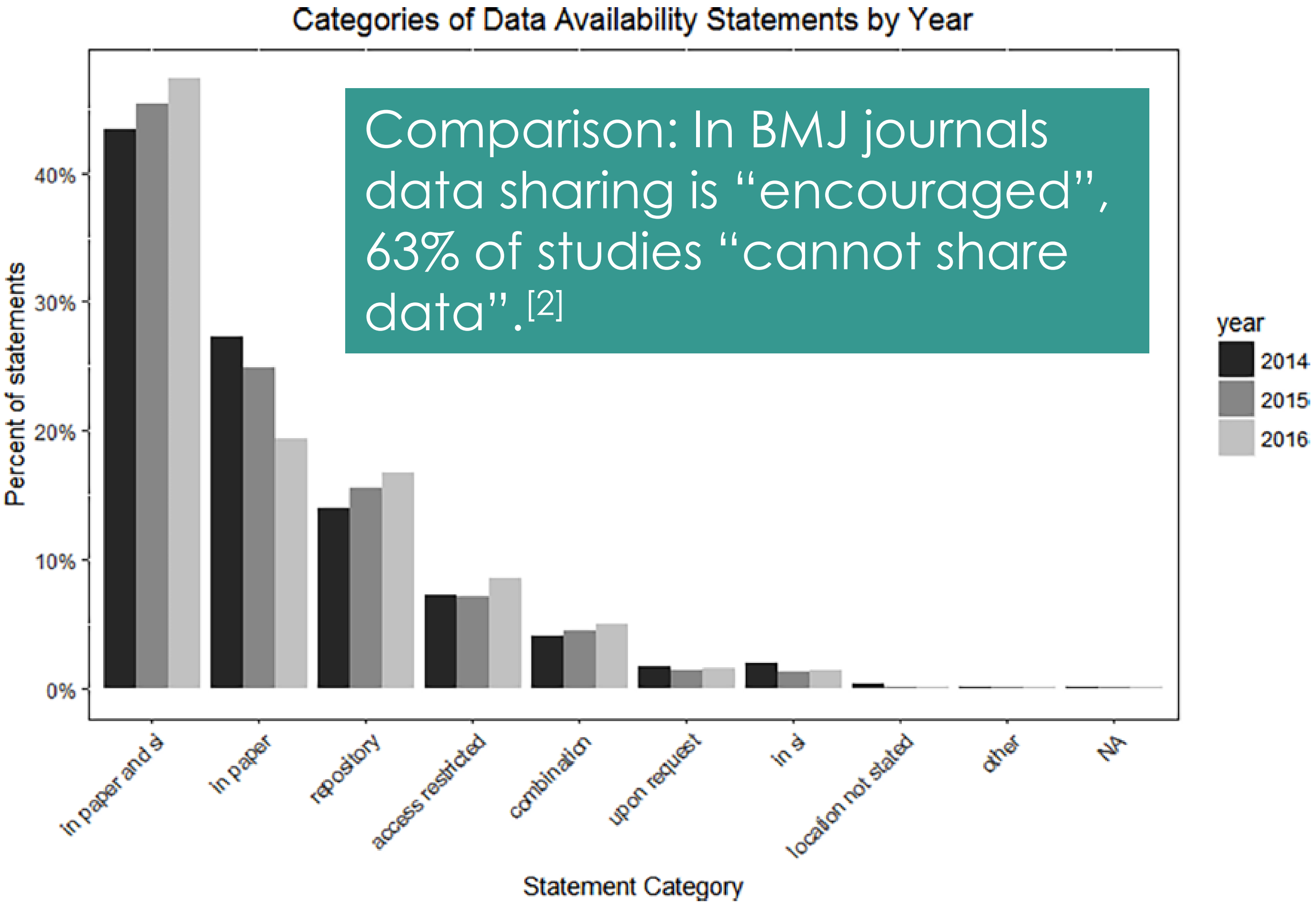


Approx. 20% in repositories

Federer LM, Belter CW, Joubert DJ, Livinski A, Lu Y-L, Snyders LN, et al. (2018) Data sharing in PLOS ONE: An analysis of Data Availability Statements. PLoS ONE 13(5): e0194768. https://doi.org/10.1371/journal.pone.0194768

PLOS

Data availability statements 2014-16



Comparison: In BMJ journals data sharing is “encouraged”, 63% of studies “cannot share data”.[2]

[1] Federer LM et al. (2018) Data sharing in PLOS ONE: An analysis of Data Availability Statements. PLoS ONE 13(5): e0194768.
[2] McDonald L et al. (2017) A review of data sharing statements in observational studies published in the BMJ: A cross-sectional study. F1000Research 2017, 6: 1708

Popular repositories 2014-16

Repository	Count of mentions
Figshare	1,446
Gene Expression Omnibus (GEO)	1,001
Genbank	999
Dryad	987
Non-repository website	329
Institutional repository	317
Zenodo	100

Federer LM, Belter CW, Joubert DJ, Livinski A, Lu Y-L, Snyders LN, et al. (2018) Data sharing in PLOS ONE: An analysis of Data Availability Statements. PLoS ONE 13(5): e0194768. https://doi.org/10.1371/journal.pone.0194768

PLOS

Outline

- Introducing PLOS and *PLOS ONE*
- Our Data Policy
  - What does it say?
  - How does it work in practice?
- Practical Data & Code Sharing
  - Data Repositories
  - Protocols.io
  - Code Sharing
- Open Science Innovations at PLOS

Practical Data Sharing – general tips

Build open data sharing into everything you do

Prepare all data sets that you use and produce in the knowledge that they will be shared -- or share openly as you create them.

Consider

What are community standards around presentation of this data?

Which metadata is necessary to make this data useful?

How to document processing steps?

Standards around data and metadata

re3data.org

Some general purpose data repositories: Dryad, Harvard Dataverse, Zenodo, Open Science Framework

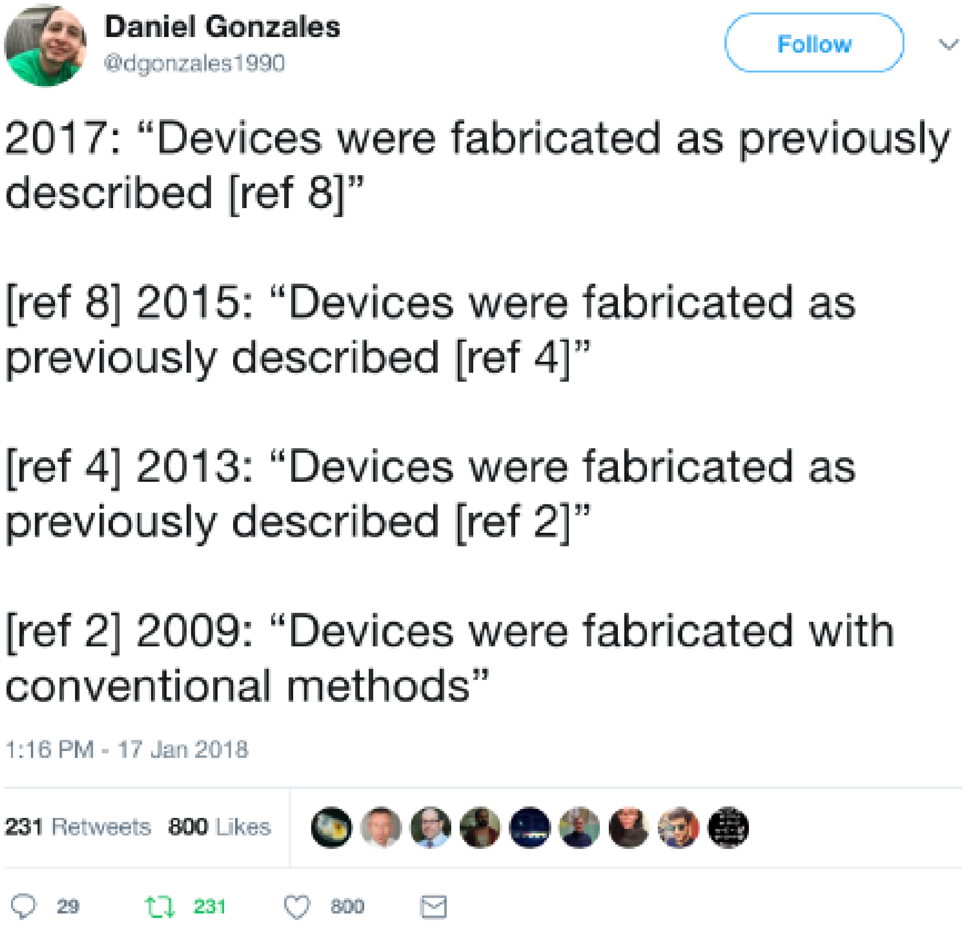
FAIRsharing.org
standards, databases, policies

If you can't find standards around data sharing and metadata for your specific method, get together with your colleagues and mentors and start the discussion!

PLOS

The importance of sharing protocols

protocols.io



Daniel Gonzales
@dgonzales1990

2017: “Devices were fabricated as previously described [ref 8]”

[ref 8] 2015: “Devices were fabricated as previously described [ref 4]”

[ref 4] 2013: “Devices were fabricated as previously described [ref 2]”

[ref 2] 2009: “Devices were fabricated with conventional methods”

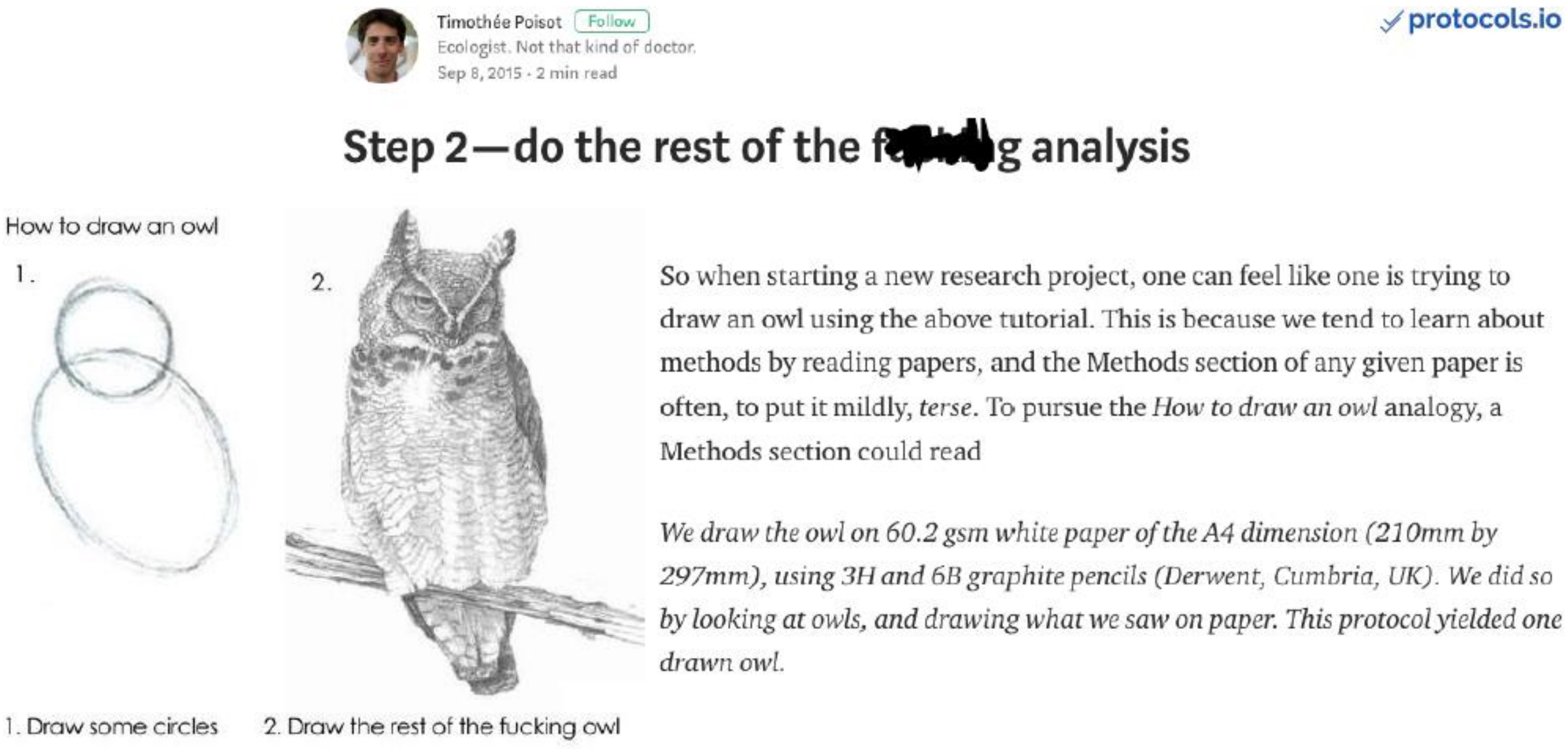
1:16 PM - 17 Jan 2018

231 Retweets 800 Likes

Adapted from Lenny Teytelman

PLOS

The importance of sharing protocols



Timothée Poisot Follow
Ecologist. Not that kind of doctor.
Sep 8, 2015 · 2 min read

protocols.io

Step 2—do the rest of the f[redacted]g analysis

How to draw an owl

1.

2.

1. Draw some circles

2. Draw the rest of the fucking owl

So when starting a new research project, one can feel like one is trying to draw an owl using the above tutorial. This is because we tend to learn about methods by reading papers, and the Methods section of any given paper is often, to put it mildly, *terse*. To pursue the *How to draw an owl* analogy, a Methods section could read

We draw the owl on 60.2 gsm white paper of the A4 dimension (210mm by 297mm), using 3H and 6B graphite pencils (Derwent, Cumbria, UK). We did so by looking at owls, and drawing what we saw on paper. This protocol yielded one drawn owl.

https://medium.com/@tpoi/do-the-rest-of-the-fucking-analysis-8fcef22fd991

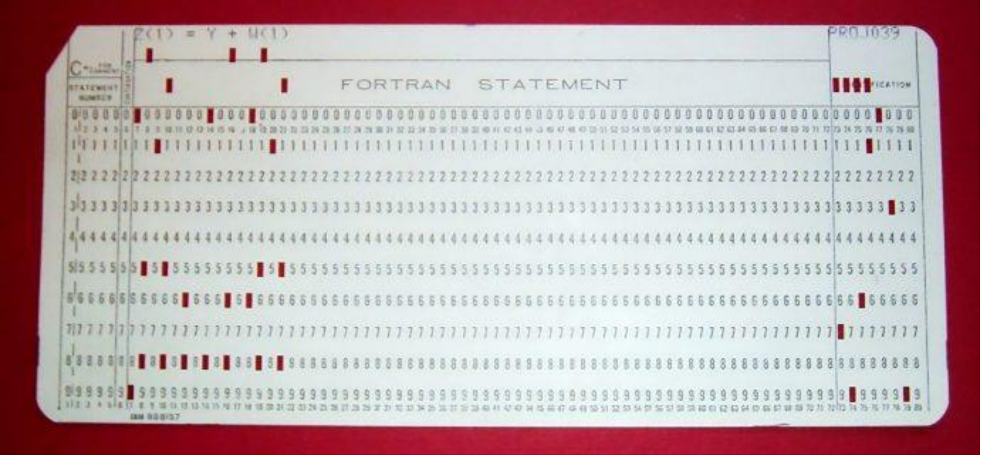
PLOS

Adapted from Lenny Teytelman

protocols.io

PLOS

Source Code Sharing: Opportunities



Sharing Code in the 21st century does not require snail mail!

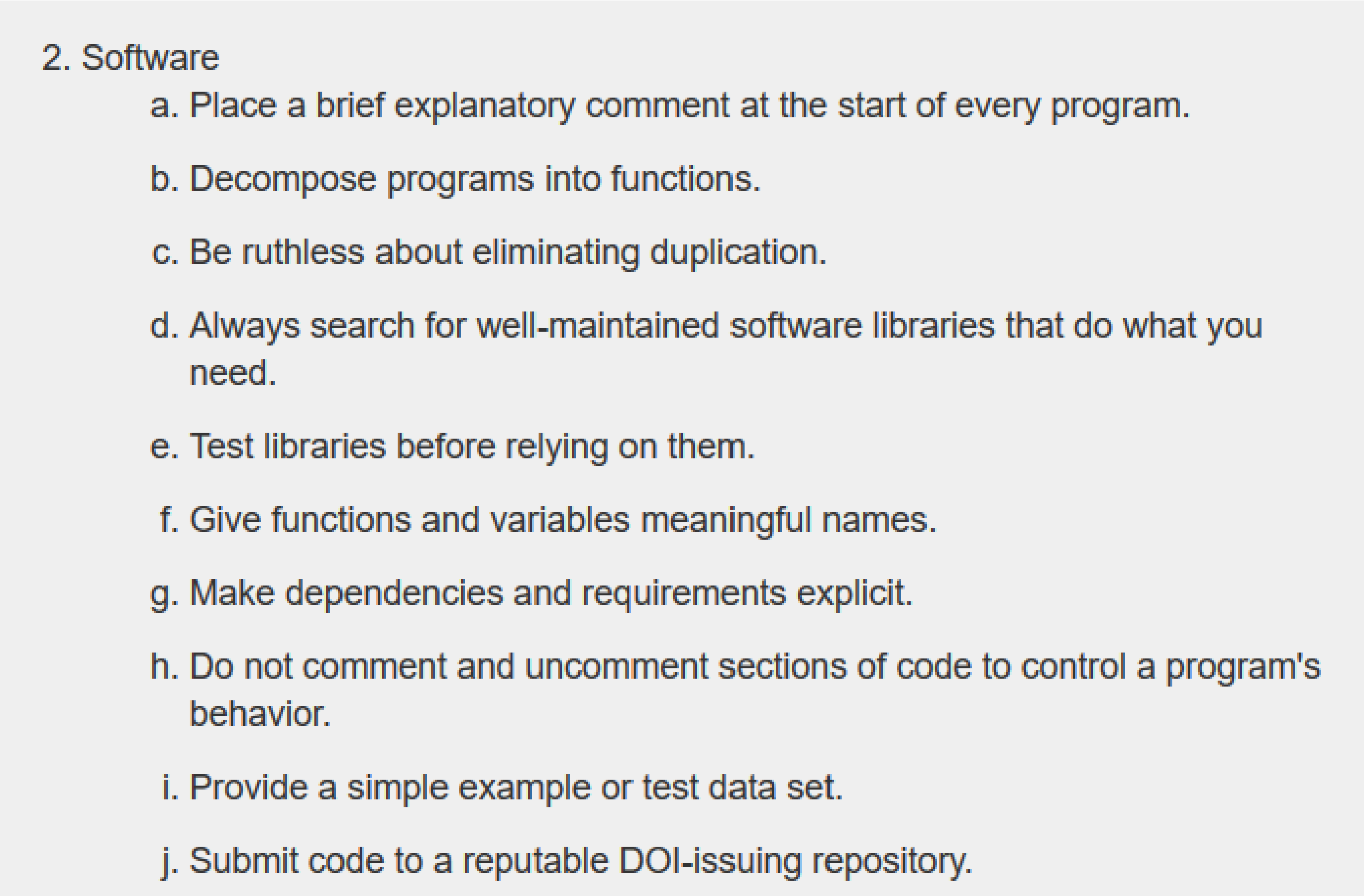
- Sharing will increase impact of the work
- Forces better maintenance and documentation
- Credit for software development
- A great GitHub page is invaluable for students who don’t stay in academia
- For some, code is easier to understand than equations
- Provenance

Peter Wittek: Stop Hiding your Code
https://blogs.plos.org/everyone/2018/04/18/stop-hiding-your-code/

PLOS

Good enough practices in scientific computing

Wilson et al. https://doi.org/10.1371/journal.pcbi.1005510



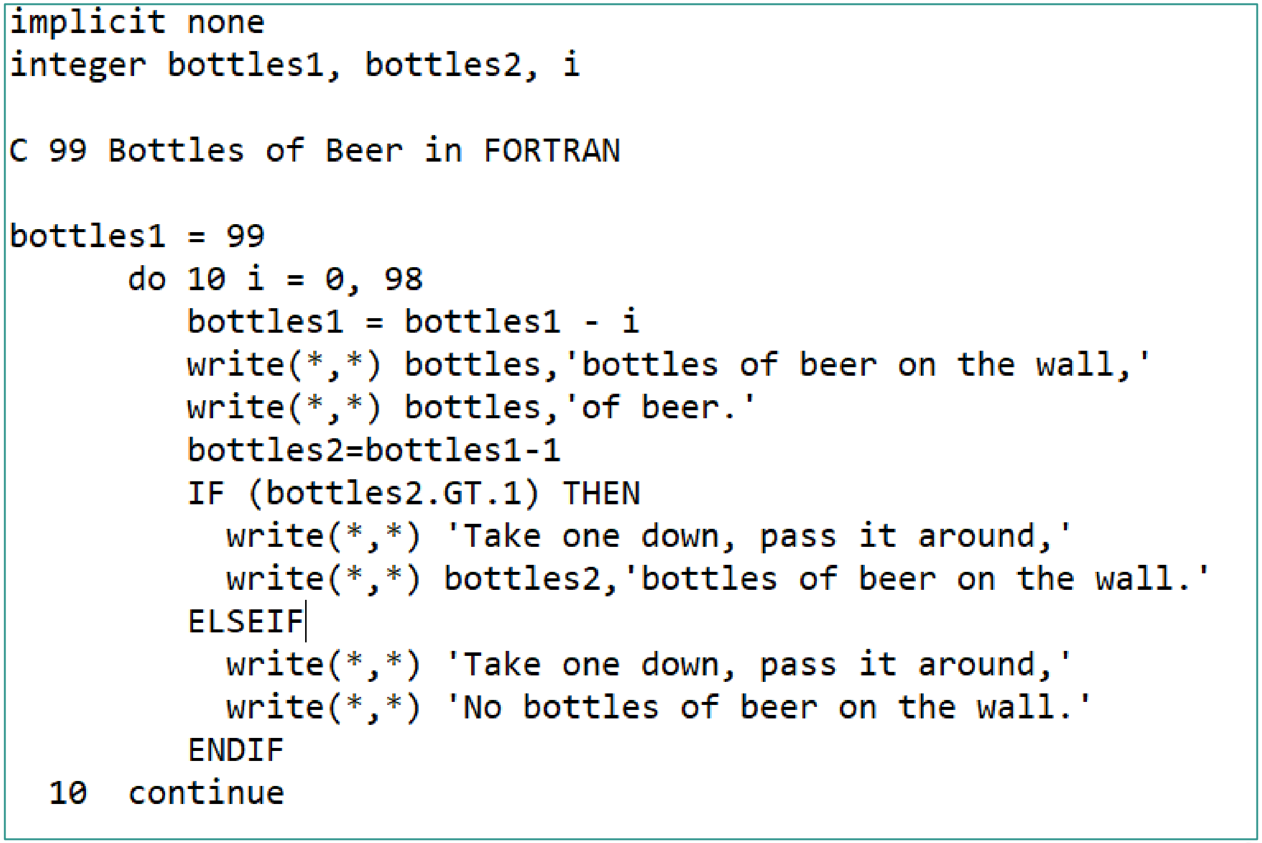
2. Software
    a. Place a brief explanatory comment at the start of every program.
    b. Decompose programs into functions.
    c. Be ruthless about eliminating duplication.
    d. Always search for well-maintained software libraries that do what you need.
    e. Test libraries before relying on them.
    f. Give functions and variables meaningful names.
    g. Make dependencies and requirements explicit.
    h. Do not comment and uncomment sections of code to control a program's behavior.
    i. Provide a simple example or test data set.
    j. Submit code to a reputable DOI-issuing repository.

software carpentry

PLOS

Source Code Sharing: Github

Make it permanent with Zenodo and give it a licence!



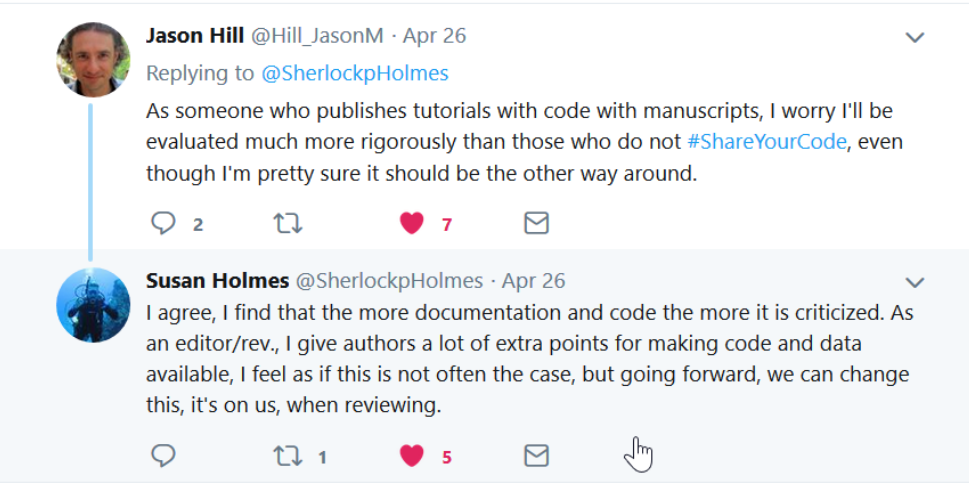
```
implicit none
integer bottles1, bottles2, i

C 99 Bottles of Beer in FORTRAN

bottles1 = 99
      do 10 i = 0, 98
         bottles1 = bottles1 - i
         write(*,*) bottles,'bottles of beer on the wall,'
         write(*,*) bottles,'of beer.'
         bottles2=bottles1-1
         IF (bottles2.GT.1) THEN
           write(*,*) 'Take one down, pass it around,'
           write(*,*) bottles2,'bottles of beer on the wall.'
         ELSEIF
           write(*,*) 'Take one down, pass it around,'
           write(*,*) 'No bottles of beer on the wall.'
         ENDIF
  10  continue
```

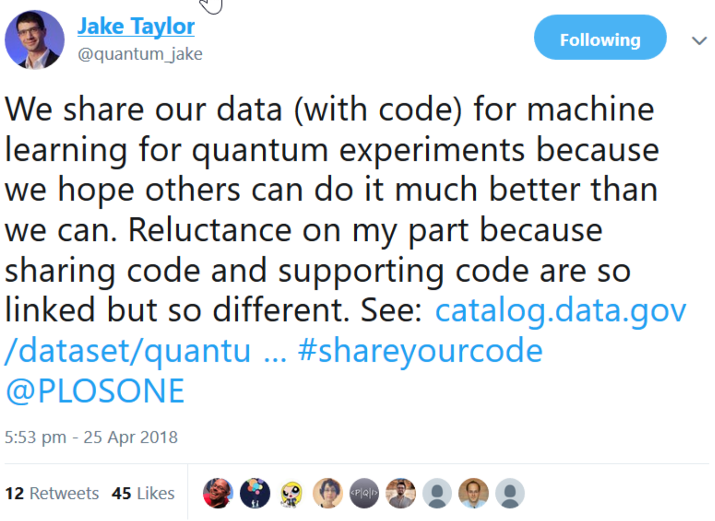
Source Code Sharing: Challenges

"I would like to share my code but I don't know how."
-- *PLOS ONE* Academic Editor



Jason Hill @Hill_JasonM · Apr 26
Replying to @SherlockpHolmes
As someone who publishes tutorials with code with manuscripts, I worry I'll be evaluated much more rigorously than those who do not #ShareYourCode, even though I'm pretty sure it should be the other way around.

Susan Holmes @SherlockpHolmes · Apr 26
I agree, I find that the more documentation and code the more it is criticized. As an editor/rev., I give authors a lot of extra points for making code and data available, I feel as if this is not often the case, but going forward, we can change this, it's on us, when reviewing.



Jake Taylor
@quantum_jake

We share our data (with code) for machine learning for quantum experiments because we hope others can do it much better than we can. Reluctance on my part because sharing code and supporting code are so linked but so different. See: catalog.data.gov /dataset/quantu ... #shareyourcode @PLOSONE

5:53 pm - 25 Apr 2018

12 Retweets 45 Likes

→ Stay tuned for collections with exemplary data sharing on Quantum Computation & Simulation and Machine Learning in Health

PLOS

Outline

- Introducing PLOS and *PLOS ONE*
- Our Data Policy
  - What does it say?
  - How does it work in practice?
- Practical Data & Code Sharing
  - Data Repositories
  - Protocols.io
  - Code Sharing
- Open Science Innovations at PLOS

Preprints – partnership with bioRxiv

Authors can choose to have their work posted to the bioRxiv preprint server upon submission to PLOS journals

bioRxiv

THE PREPRINT SERVER FOR BIOLOGY

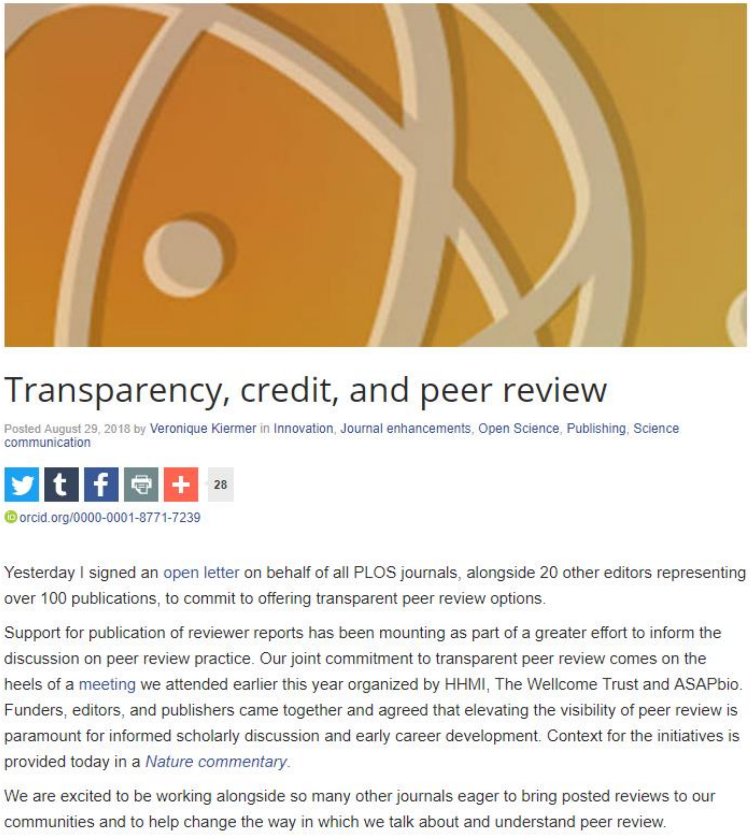
- PLOS staff perform initial screening to determine suitability and match with bioRxiv's scope
- Authors must opt-in at submission
- Editors can consider commentary on the preprint during the peer review process

Launched in May

PLOS

journals.plos.org/plosone/s/preprints

Transparent peer review



Transparency, credit, and peer review

Posted August 29, 2018 by Veronique Kiermer in Innovation, Journal enhancements, Open Science, Publishing, Science communication

28

orcid.org/0000-0001-8771-7239

Yesterday I signed an open letter on behalf of all PLOS journals, alongside 20 other editors representing over 100 publications, to commit to offering transparent peer review options.

Support for publication of reviewer reports has been mounting as part of a greater effort to inform the discussion on peer review practice. Our joint commitment to transparent peer review comes on the heels of a meeting we attended earlier this year organized by HHMI, The Wellcome Trust and ASAPbio. Funders, editors, and publishers came together and agreed that elevating the visibility of peer review is paramount for informed scholarly discussion and early career development. Context for the initiatives is provided today in a *Nature commentary*.

We are excited to be working alongside so many other journals eager to bring posted reviews to our communities and to help change the way in which we talk about and understand peer review.

In August 2018, PLOS joined over 20 publishers in announcing its commitment to offering optional transparent peer review (publication of review reports) across its journal portfolio.

blogs.plos.org/plos/2018/08/transparency-credit-and-peer-review/

Questions?

lmueck@plos.org
@LeonieMueck
@PLOSONE

Backup