

# Data and the scientific literature

## **new directions in** **what data gets published** **how it happens** **& why it matters**

**Todd Vision**

University of North at Chapel Hill

@tjvision

<http://orcid.org/0000-0002-6133-2581>



Associate Director for Informatics, National Evolutionary Synthesis Center

Associate Professor Department of Biology

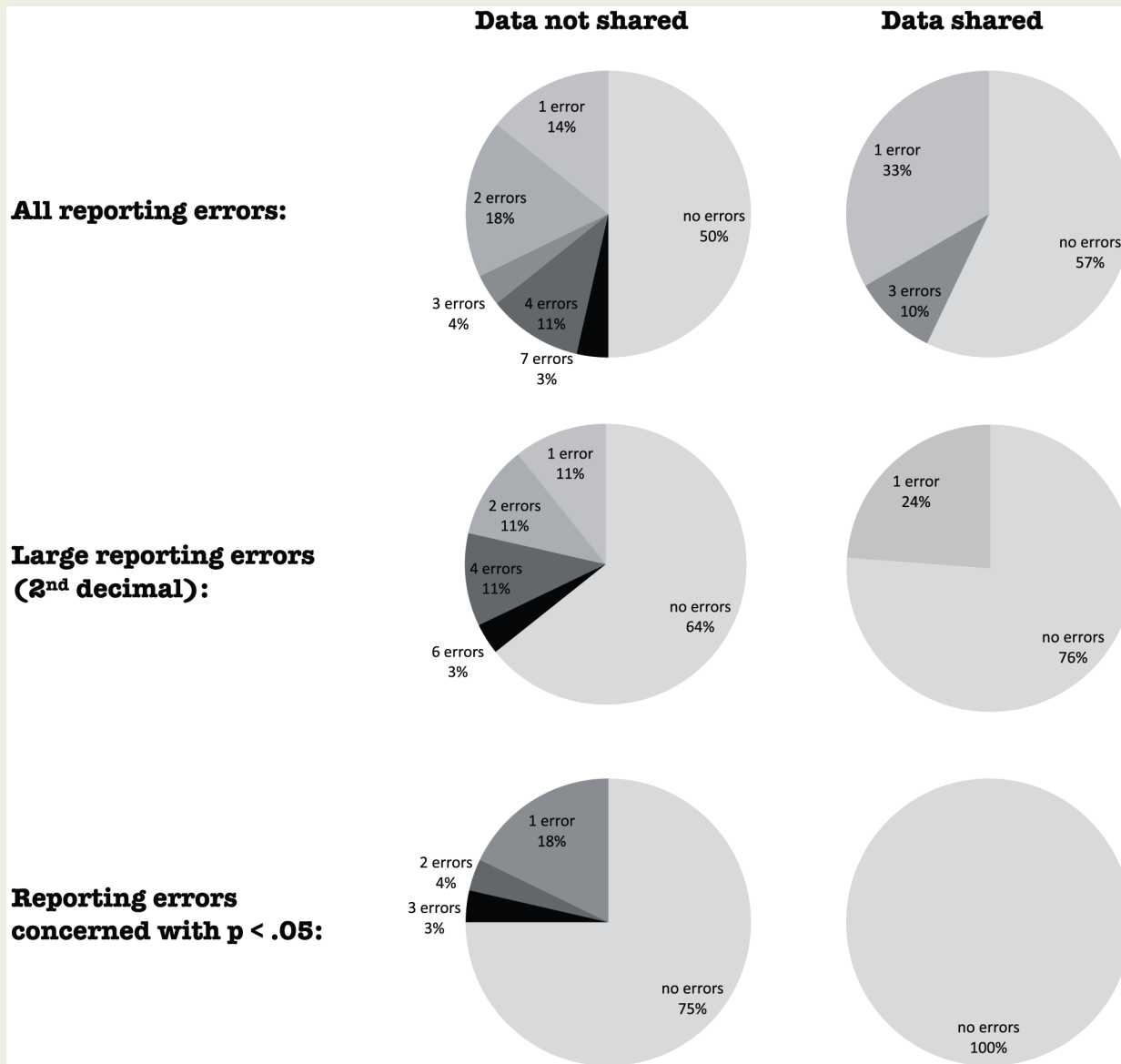
Adjunct, School of Information and Library Sciences



You may reuse any of the **original** content in these slides as you wish, provided you attribute the source





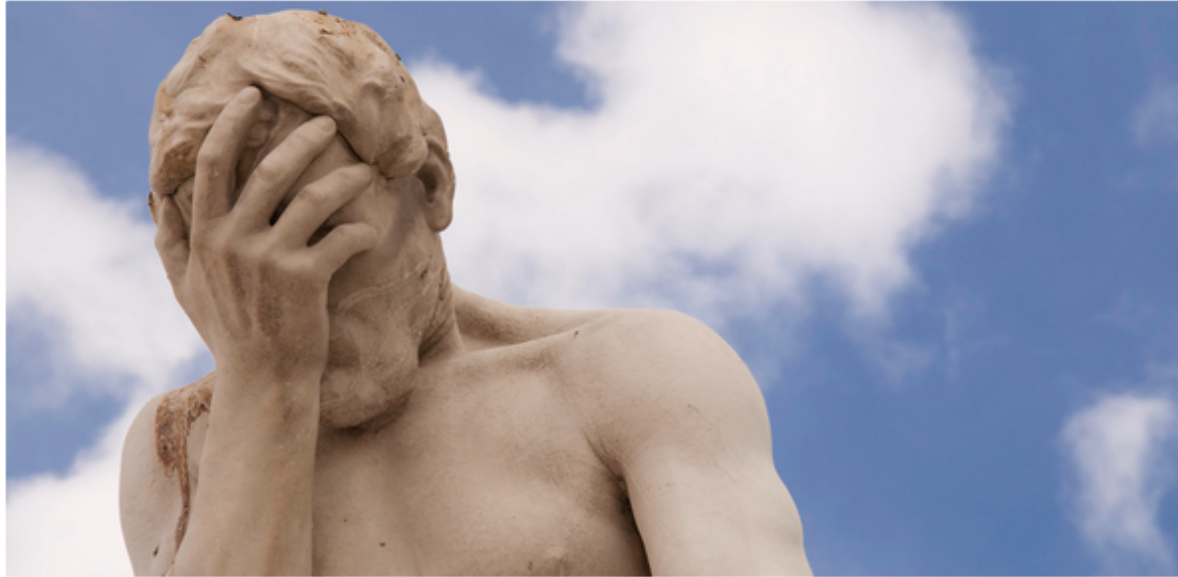


Wicherts JM, Bakker M, Molenaar D (2011) Willingness to Share Research Data Is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Results. PLoS ONE 6(11):e26828



April 22 2013, 4.40pm EDT

## The Reinhart-Rogoff error – or how not to Excel at economics



Data and computer code should be made publicly available at an early stage – or else ... esarastudio

Last week we learned a famous **2010 academic paper**, relied on by political big-hitters to bolster arguments for austerity cuts, contained significant errors; and that those errors came down to misuse of an Excel spreadsheet.

Sadly, these are not the first mistakes of this size and nature when handling data. So what on Earth went wrong, and can we fix it?

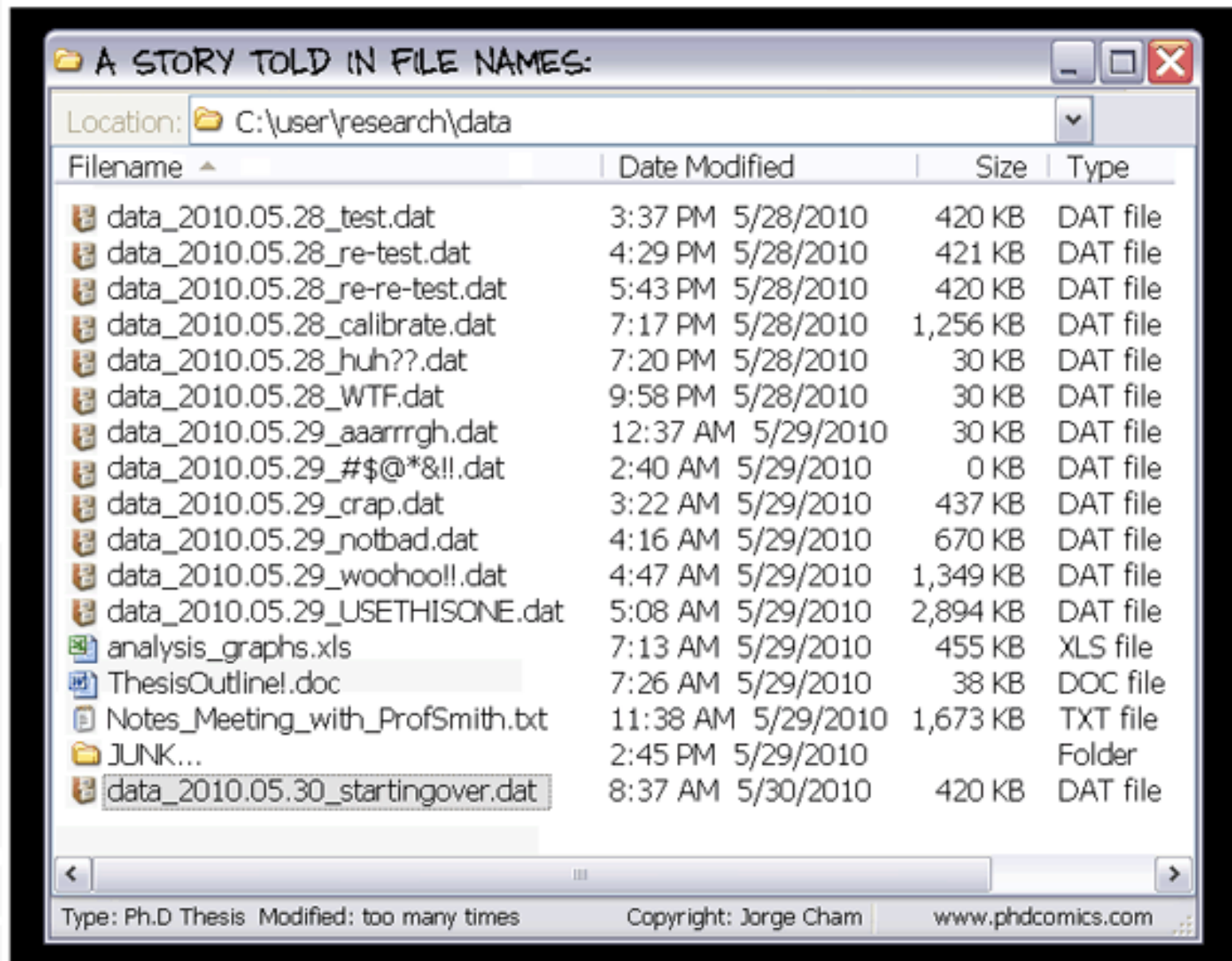
### REPUBLISH THIS ARTICLE

---

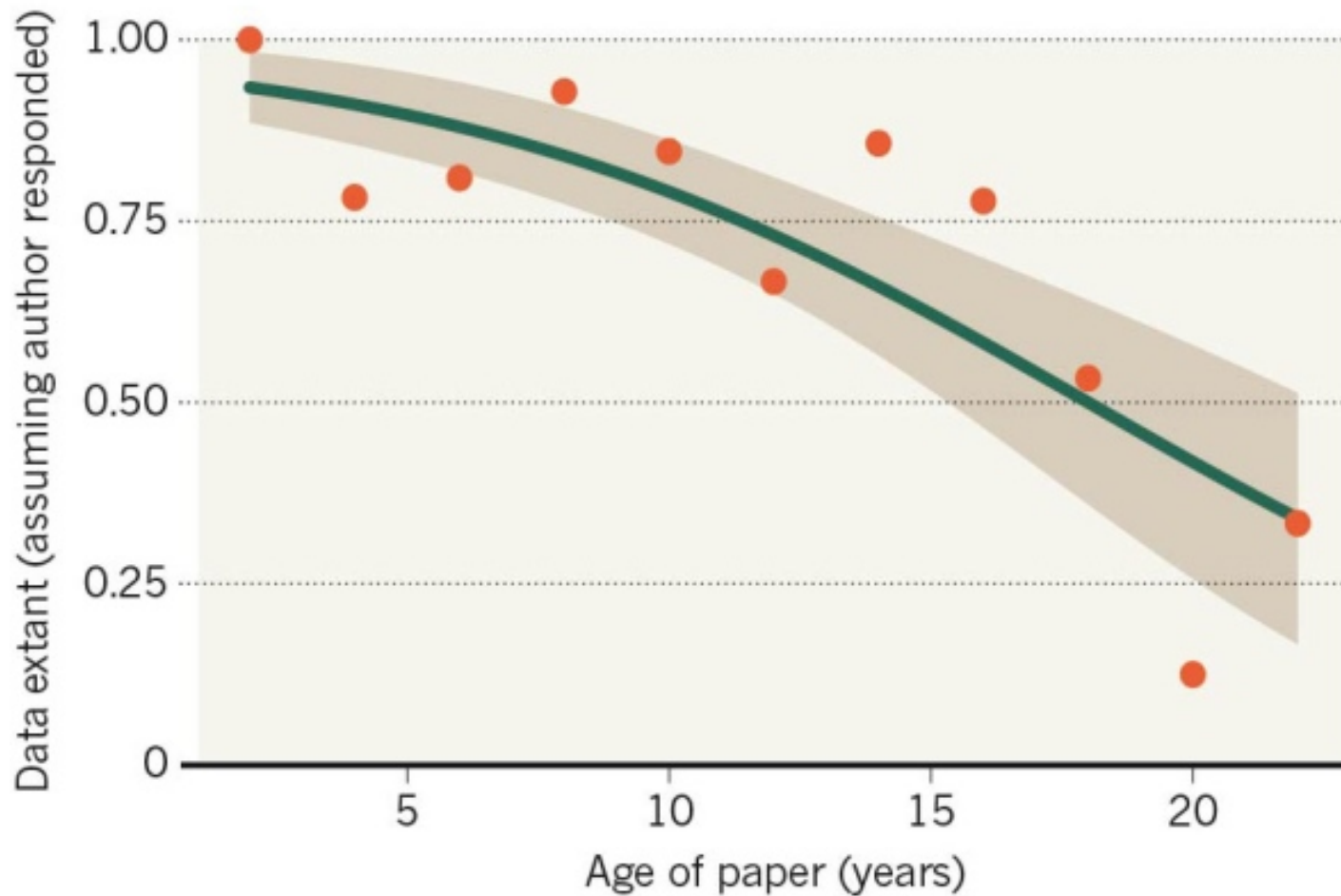
*We believe in the free flow of information. We use a **Creative Commons Attribution NoDerivatives** license, so you can republish our articles for free, online or in print.*

---

<http://theconversation.com/the-reinhart-rogooff-error-or-how-not-to-excel-at-economics-13646>

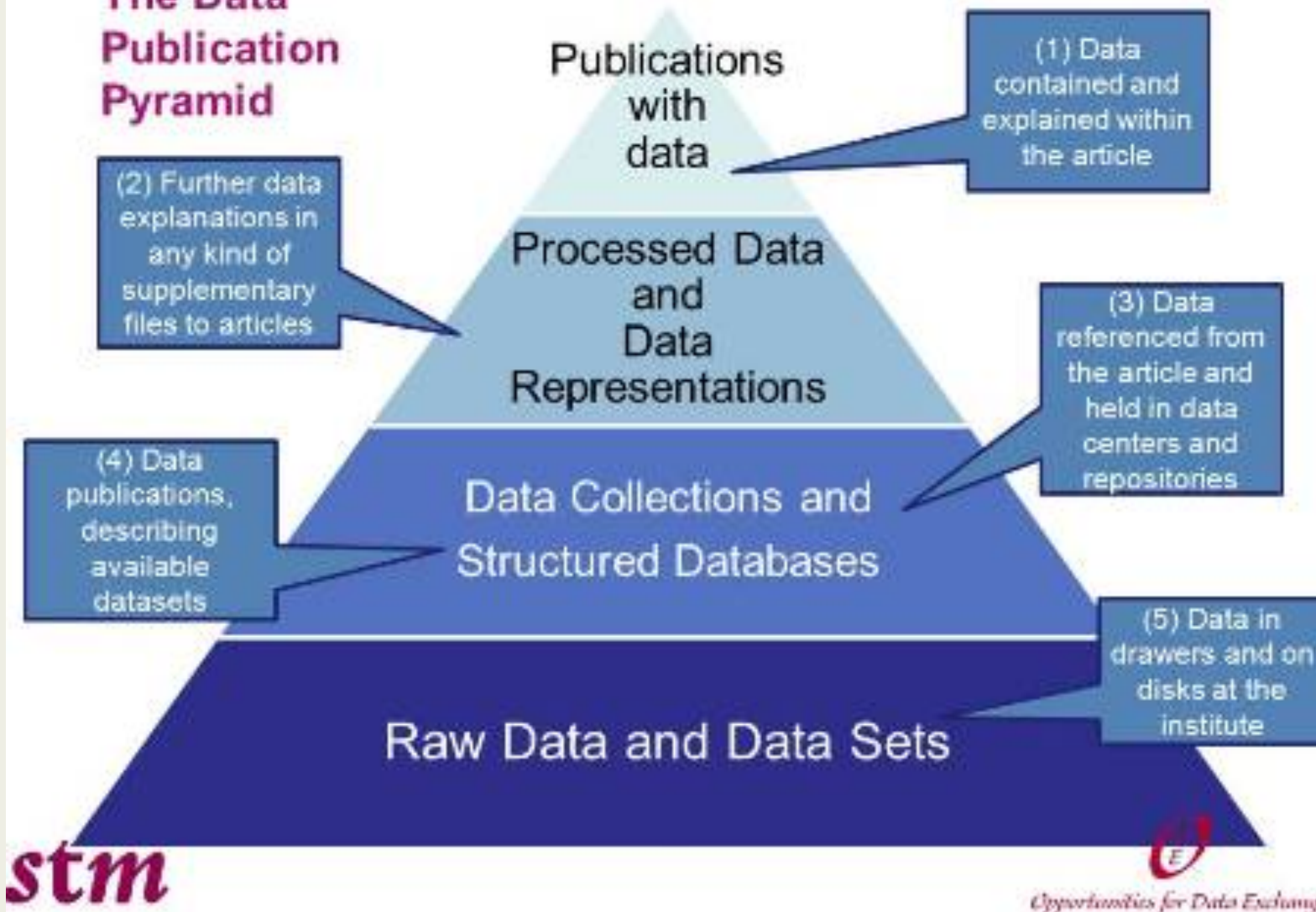


# Data availability must be dealt with no later than the point of publication

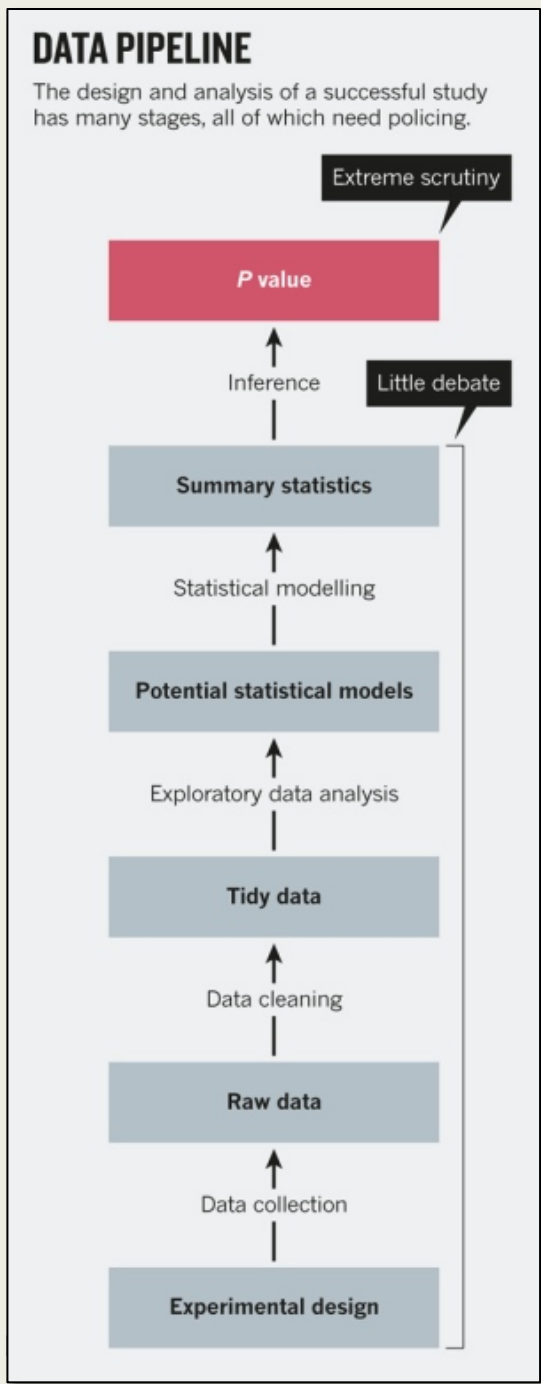


Vines TH *et al.* (2013) *Current Biology* DOI:10.1016/j.cub.2013.11.014

## The Data Publication Pyramid



Leek and Peng (2015)



# Joint Data Archiving Policy (JDAP)

Data are important products of the scientific enterprise, and they should be **preserved** and **usable** for decades in the future.

As a condition for publication, data supporting the results in the article should be deposited in an **appropriate public archive**.

Authors may elect to **embargo** access to the data for a period up to a year after publication.

**Exceptions** may be granted at the discretion of the editor, especially for sensitive information.

<http://datadryad.org/pages/jdap>

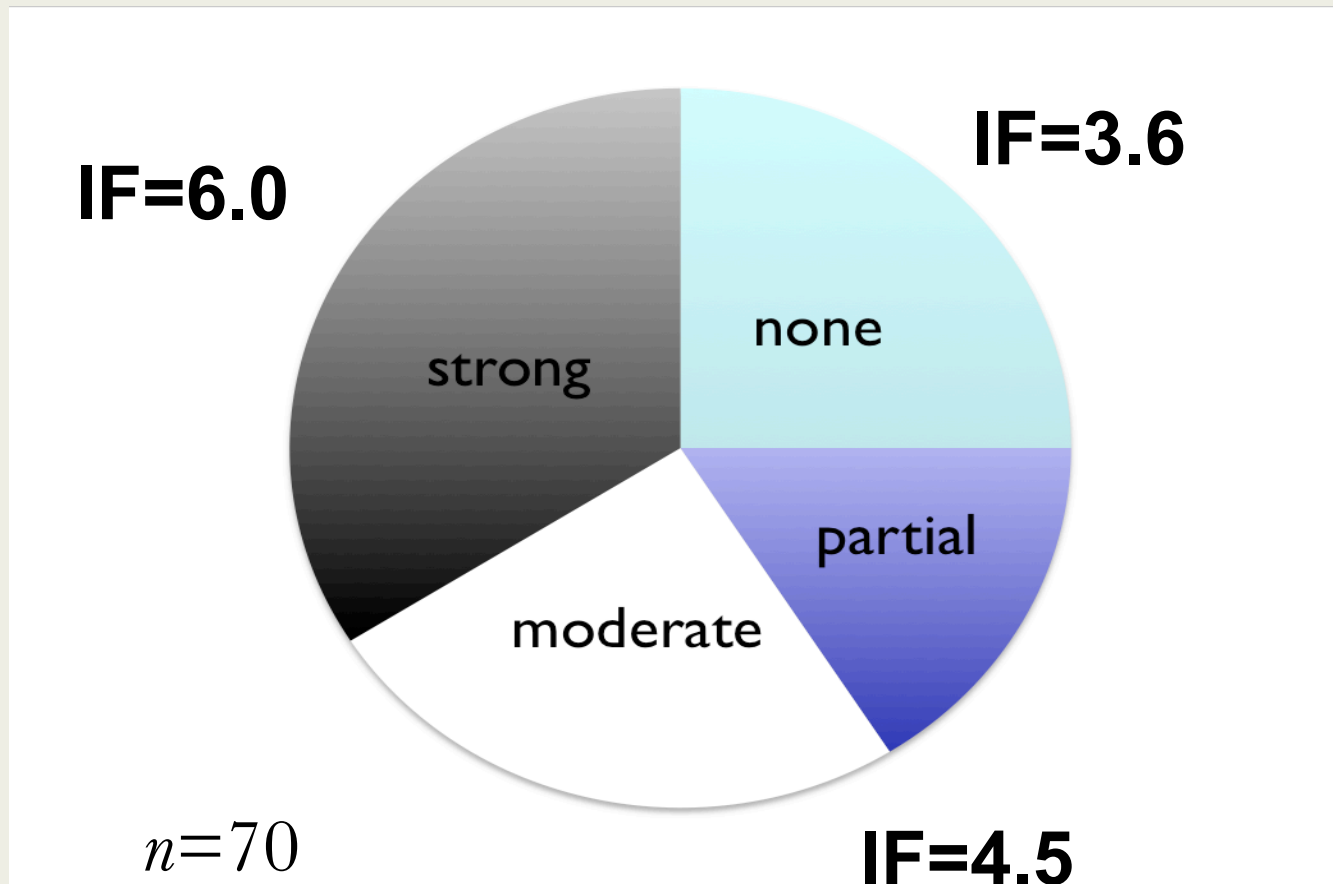


## Data Access for the Open Access Literature: PLOS's Data Policy

Posted on [December 12, 2013](#) by [Theo Bloom](#)

Data are any and all of the digital materials that are collected and analyzed in the pursuit of scientific advances. In line with Open Access to research articles themselves, PLOS strongly believes that to best foster scientific progress, the underlying data should be made freely available for researchers to use, wherever this is legal and ethical. Data availability allows replication, reanalysis, new analysis, interpretation, or inclusion into meta-analyses, and [facilitates reproducibility of research](#), all providing a better 'bang for the buck' out of scientific research, much of which is funded from public or nonprofit sources. Ultimately, all of these considerations aside, our viewpoint is quite simple: ensuring access to the underlying data should be an intrinsic part of the scientific publishing process.

# High impact factor journals have stronger data archiving policies



Piwowar HA, Chapman WW (2008) [hdl:10101/npre.2008.1700.1](https://doi.org/10.1011/npre.2008.1700.1)

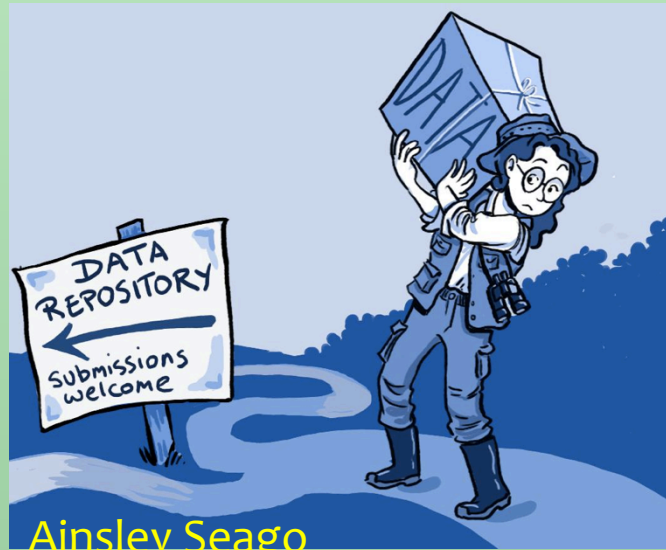


# Roles of a journal

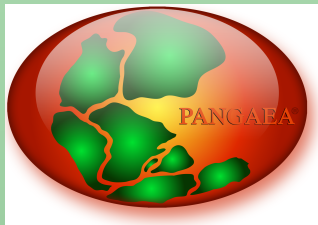
Function	Role in data
Registration	✓
Certification/ validation	✓ for methods
Dissemination	✓ limited
Archiving	✓ limited



zenodo



Ainsley Seago



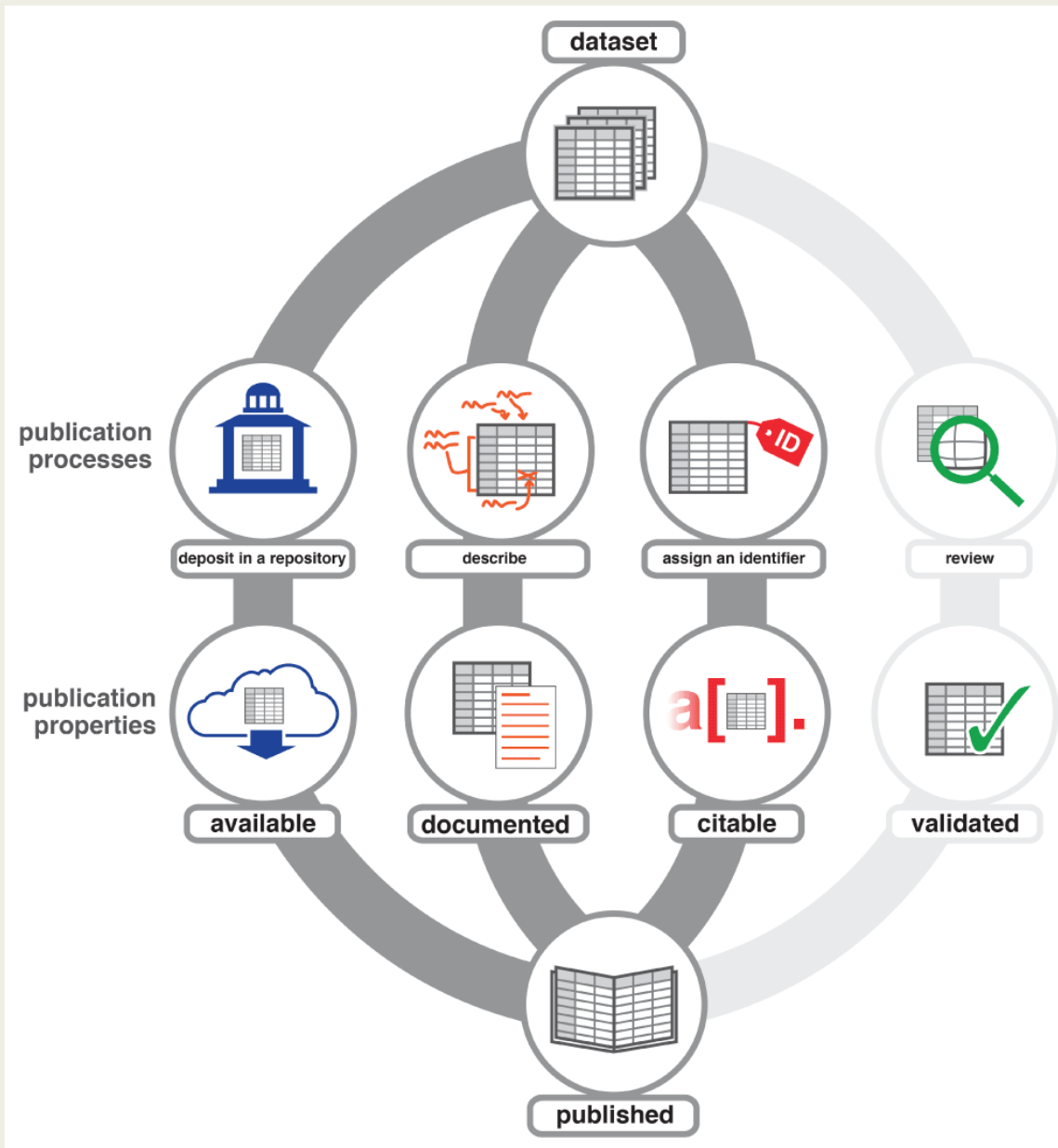
And many others

# Joint Declaration of Data Citation Principles

## #3: Evidence

“In scholarly literature, whenever and wherever a claim relies upon data, the corresponding data should be cited.”

<http://force11.org/datacitation>



Kratz J and Strasser C (2014) Data publication consensus and controversies. doi: 10.12688/f1000research.3979.3

Many open issues drift  
around data publication, but  
validation is both the biggest  
and the haziest.

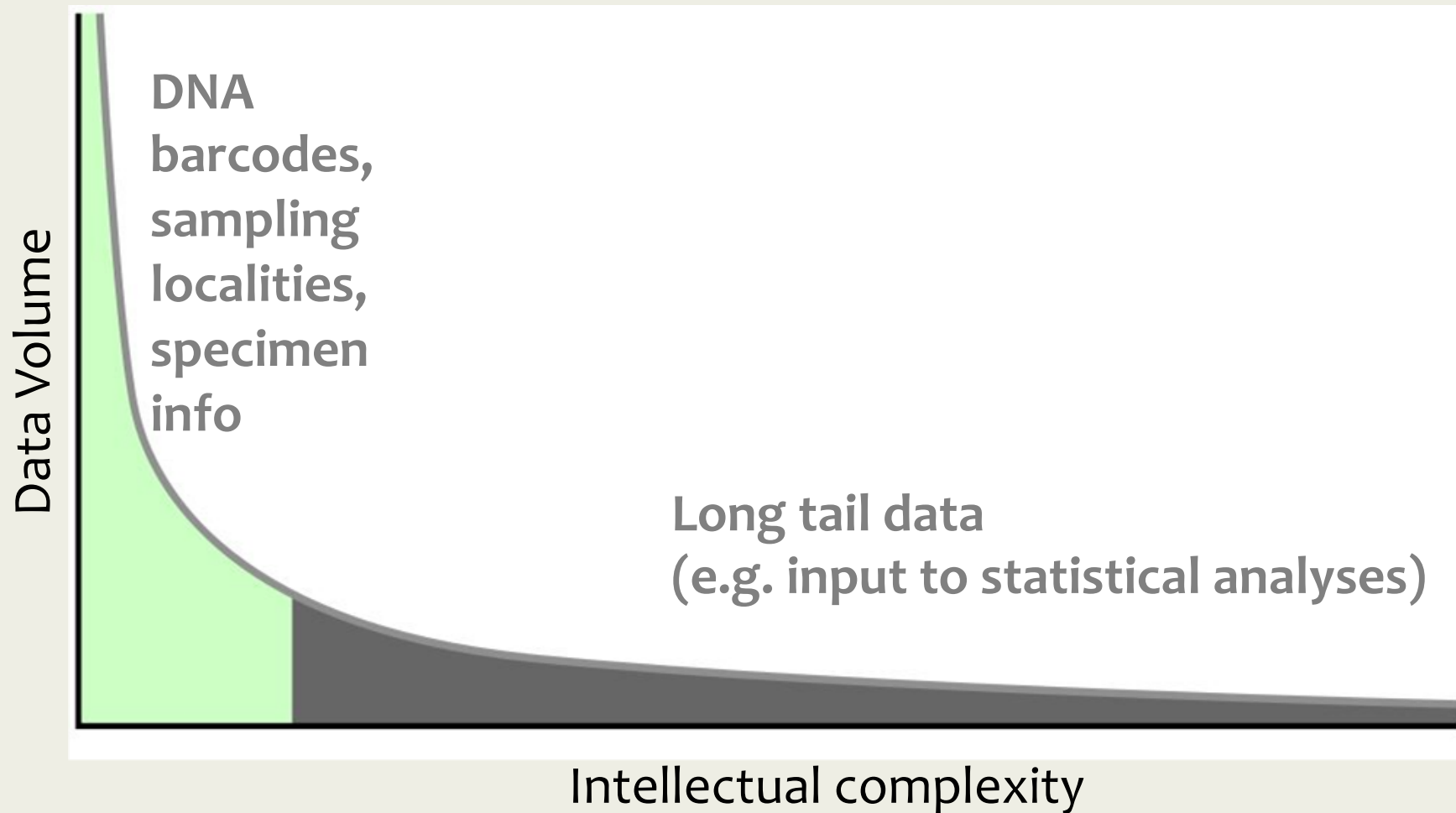
John Kratz

<http://datapub.cdlib.org/2014/05/08/fifteen-ideas-about-data-validation-and-peer-review/>

# Issues in (non-data) peer review

- Reviewer capacity
- Quality of reviews
  - Ability to judge quality of science
  - Ability to provide useful feedback
- Rejection cascades
  - Peer review as independent service
  - Soundness versus fit and importance
  - Rise of recommendation services
- Timing
  - Post-publication review and services like PubPeer
  - Preprints as a disruptive technology
- Confidentiality
  - Open-ness of reviews and the review process
  - Blindedness and anonymity
- The diminishing role of the journal
  - Article-level metrics

# Long tail data: includes data behind figures, tables, statistical analyses; files where context matters



After Heidorn (2008) <http://hdl.handle.net/2142/9127>



[Submit data now](#)

[How and why?](#)

### Search for data

Enter keyword, author, title, DOI, etc [Go](#)

[Advanced search](#)

### Browse for data

- Recently published
- Popular
- By Author
- By Journal

#### Recently Published Data

Bryson RW, Savary WE, Prendini L (2013) Data from: Biogeography of scorpions in the *Pseudouroctonus minimus* complex (Vaejovidae) from south-western North America: implications of ecological specialization for pre-Quaternary diversification. *Journal of Biogeography* [doi:10.5061/dryad.q58r0](https://doi.org/10.5061/dryad.q58r0)

Rouger R, Jump AS (2014) Data from: A seascape genetic analysis reveals strong biogeographical structuring driven by contrasting processes in the polyploid saltmarsh species *Puccinellia maritima* and *Triglochin maritima*. *Molecular Ecology* [doi:10.5061/dryad.dc56n](https://doi.org/10.5061/dryad.dc56n)

Sremba AL, Martin AR, Baker CS (2014) Data from: Species identification and likely catch time period of whale bones from South Georgia. *Marine Mammal Science*

### Be part of Dryad



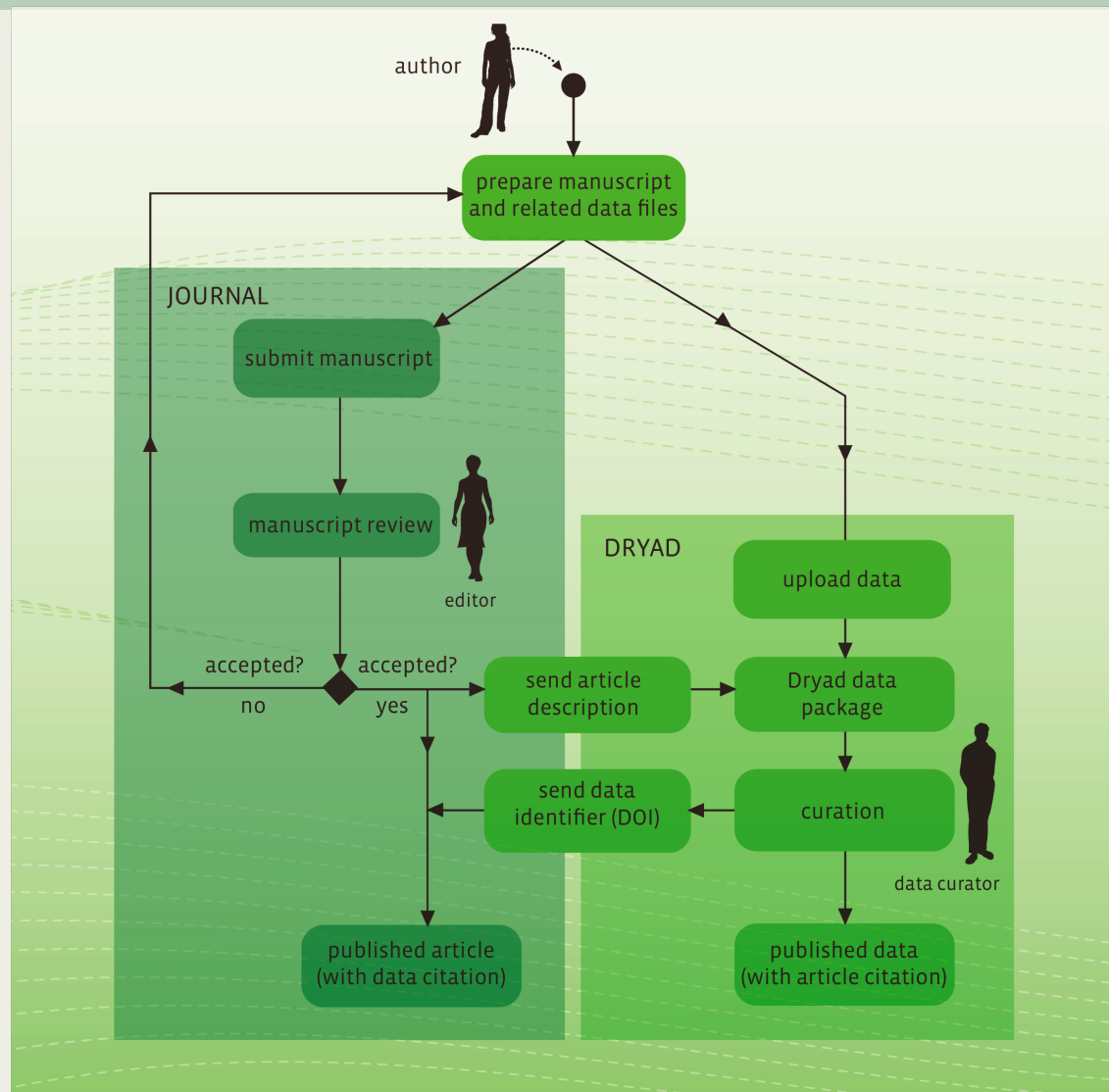
Publishers, societies, universities, libraries, funders, and other stakeholder organizations are invited to become [members](#). Tap into an active knowledge-sharing network, receive discounts on submission fees, and help shape Dryad's future.

[Submission integration](#) is a free service that allows publishers to coordinate manuscript and data submissions. It makes submitting data easy for researchers; makes linking articles and data easy for journals; and enables confidential review of

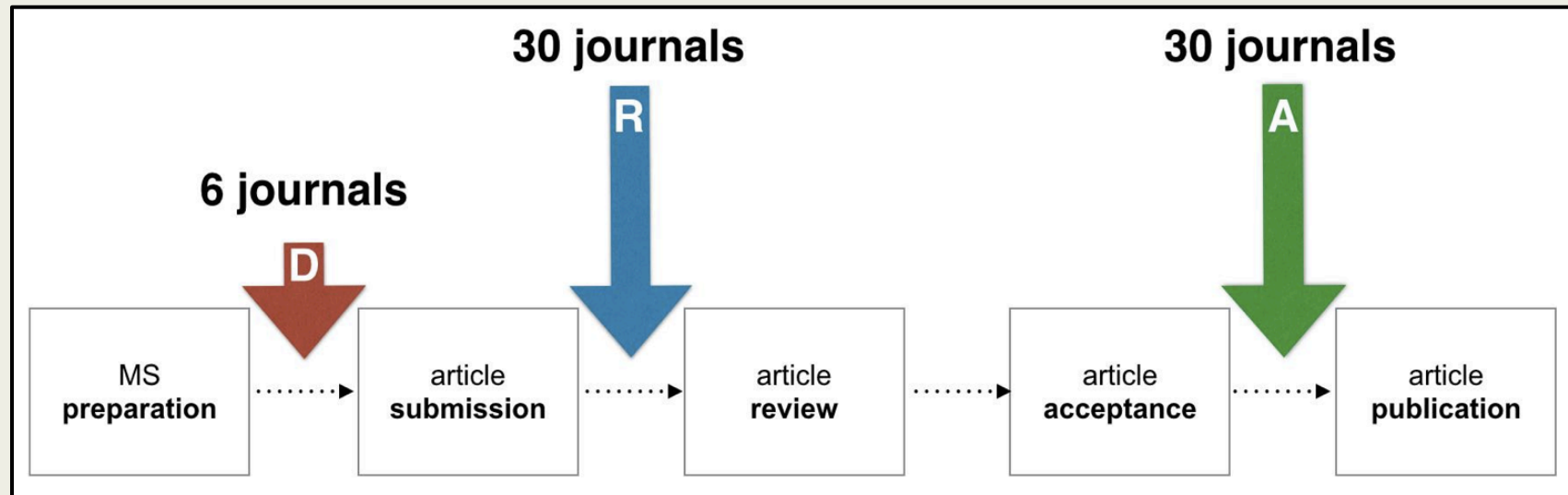




# Submission integration of publications & data



# When to submit data?



# Bidirectional linkage with publications

The screenshot shows the ScienceDirect interface for the journal *Molecular Phylogenetics and Evolution*, Volume 28, Issue 2, August 2003, Pages 261–275. The article title is "Molecular systematics of armadillos (*Xenarthra*, *Dasypodidae*): contribution of maximum likelihood and Bayesian analyses of mitochondrial and nuclear genes" by Frédéric Delsuc<sup>a</sup>, Michael J Stanhope<sup>b</sup>, and Emmanuel J.P Douzery<sup>a</sup>. The article is available for purchase at \$39.95. A yellow box highlights the "Data for this Article" section, which states: "Data for this article is available at the following data repositories:" and features a "Data in DRYAD" button. The interface also includes a navigation menu, a search bar, and a sidebar with a table of contents and a thumbnail of a phylogenetic tree.

# One component of a larger research object

The image shows a composite screenshot of a web interface. The top portion is a PubMed page for a research article. The bottom portion is a GenBank page for a specific nucleotide sequence. Both pages have yellow boxes highlighting external resource links.

**PubMed Article:**

- Journal: *Mol. Ecol.*, 2011 Feb;20(3):584-600. doi: 10.1111/j.1365-294X.2010.04953.x. Epub 2010 Dec 16.
- Title: **Comparative phylogeography, genetic differentiation and contrasting reproductive modes in three fungal symbionts of a multipartite bark beetle symbiosis.**
- Authors: Roe AD, Rice AV, Coltman DW, Cooke JE, Sperling FA.
- Abstract: Multipartite symbioses are complex symbiotic relationships involving multiple interacting partners. These types of partnerships provide excellent opportunities in which to apply a comparative approach to identify common historical patterns of population differentiation and species-specific life history traits. Using three symbiotic blue-stain fungal species (Ophiostomataceae) associated with outbreaking populations of the mountain pine beetle (*Dendroctonus ponderosae* Hopkins) in western Canada, we applied phylogenetic, population genetic and demographic approaches to clarify phylogeographic patterns among the three fungal species in northern and southern populations, despite dramatic differences in their life history traits. The patterns were generally consistent, showing some interspecific incongruence in their recombination rate and ecological traits that could be related to their approach to partners of a multipartite symbiosis. These results help us to understand the complexity and evolution of multipartite symbioses.

**GenBank Sequence:**

- Sequence ID: Ophiostoma montium isolate ss547 5.8S ribosomal RNA gene, partial sequence; internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence.
- Accession: HQ413650
- Length: 918 bp
- Definition: Ophiostoma montium isolate ss547 5.8S ribosomal RNA gene, partial sequence; internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence.

**Highlighted Links:**

- PubMed Article:** Other Literature Sources, Dryad Digital Repository, Labome Researcher Resources - ExactAntigen
- GenBank Page:** LinkOut to external resources, SILVA LSU Database, Dryad Digital Repository

## Integrated Journals and Costs for Submitters

Show all amounts in USDS



All journals with either [integrated data submission](#) or sponsored [Data Publishing Charges](#) are listed below. Note that for large data packages, submitters will be asked to pay an additional \$15 for the first GB beyond 10GB and \$10 for each GB thereafter.

Journal	Submission				Payment
	Integrated	Submit: R: Prior to Review A: On Acceptance	Data embargo allowed	Metadata hidden until publication	Sponsor/Cost to submitter
<i>If your journal is not listed below...</i>	<i>N</i>	<i>A</i>	<i>Y</i>	<i>Y</i>	<i>\$90</i>
American Naturalist, The	Y	A	Y	N	American Society of Naturalists
Biodiversity Data Journal	N	A	Y	Y	Pensoft
Biological Journal of the Linnean Society	Y	A	Y	N	\$80
Biology Letters	Y	A	Y	Y	The Royal Society
BioRisk	Y	R	Y	N	Pensoft
BMJ Open	Y	R	N	N	\$80
Comparative Cytogenetics	Y	R	Y	N	Pensoft

## Data from: Towards a worldwide wood economics spectrum



### 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	Global Wood Density Database
Downloaded	7318 times
Description	Please direct all correspondence to G. Lopez-Gonzalez <G.Lopez-Gonzalez@leeds.ac.uk>
Download	<a href="#">GlobalWoodDensityDatabase.xls (2.047Mb)</a>
Details	<a href="#">View File Details</a>

When using this data, please cite the original article:

Chave J, Coomes DA, Jansen S, Lewis SL, Swenson NG, Zanne AE (2009) Towards a worldwide wood economics spectrum. *Ecology Letters* 12(4): 351-366. [doi:10.1111/j.1461-0248.2009.01285.x](https://doi.org/10.1111/j.1461-0248.2009.01285.x)

Additionally, please cite the Dryad data package:

Zanne AE, Lopez-Gonzalez G, Coomes DA, Ilic J, Jansen S, Lewis SL, Miller RB, Swenson NG, Wiemann MC, Chave J (2009) Data from: Towards a worldwide wood economics spectrum. Dryad Digital Repository. [doi:10.5061/dryad.234](https://doi.org/10.5061/dryad.234)



# Dryad is not just for tabular data I

Tamiflu, Relenza, and influenza: what the data do (or don't) tell us

2014/04/17 by tjvision | Edit

Business Pharmaceuticals industry

## What the Tamiflu saga tells us about drug trials and big pharma

We now know the government's Tamiflu stockpile wouldn't have done us much good in the event of a flu epidemic. But the secrecy surrounding clinical trials means there's a lot we don't know about other medicines we take

## Neuraminidase inhibitors for preventing and treating influenza in healthy adults and children

Tom Jefferson<sup>1</sup>, Mark A Jones<sup>2</sup>, Peter Doshi<sup>3</sup>, Chris B Del Mar<sup>4</sup>, Rokuro Hama<sup>5</sup>, Matthew J Thompson<sup>6</sup>, Elizabeth A Spencer<sup>7</sup>, Igho Onakpoya<sup>8</sup>, Kamal R Mahtani<sup>8</sup>, David Nunan<sup>8</sup>, Jeremy Howick<sup>9</sup>, Carl J Heneghan<sup>8</sup>

<sup>1</sup>The Cochrane Collaboration, Roma, Italy. <sup>2</sup>School of Population Health, The University of Queensland, Brisbane, Australia. <sup>3</sup>Department of Pharmaceutical Health Services Research, University of Maryland School of Pharmacy, Baltimore, Maryland, USA. <sup>4</sup>Centre for Research in Evidence-Based Practice (CREBP), Bond University, Gold Coast, Australia. <sup>5</sup>Japan Institute of Pharmacovigilance, Osaka, Japan. <sup>6</sup>Department of Family Medicine, University of Washington, Seattle, WA, USA. <sup>7</sup>Department of Primary Care Health Sciences, University of Oxford, Oxford, UK. <sup>8</sup>Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK. <sup>9</sup>Centre for Evidence-Based Medicine, University of Oxford, Oxford, UK

Contact address: Tom Jefferson, The Cochrane Collaboration, Via Puglie 23, Roma, 00187, Italy. [jefferson.tom@gmail.com](mailto:jefferson.tom@gmail.com). [jefferson@assr.it](mailto:jefferson@assr.it).

**Editorial group:** Cochrane Acute Respiratory Infections Group.

**Publication status and date:** New search for studies and content updated (conclusions changed), published in Issue 4, 2014.

**Review content assessed as up-to-date:** 22 July 2013.

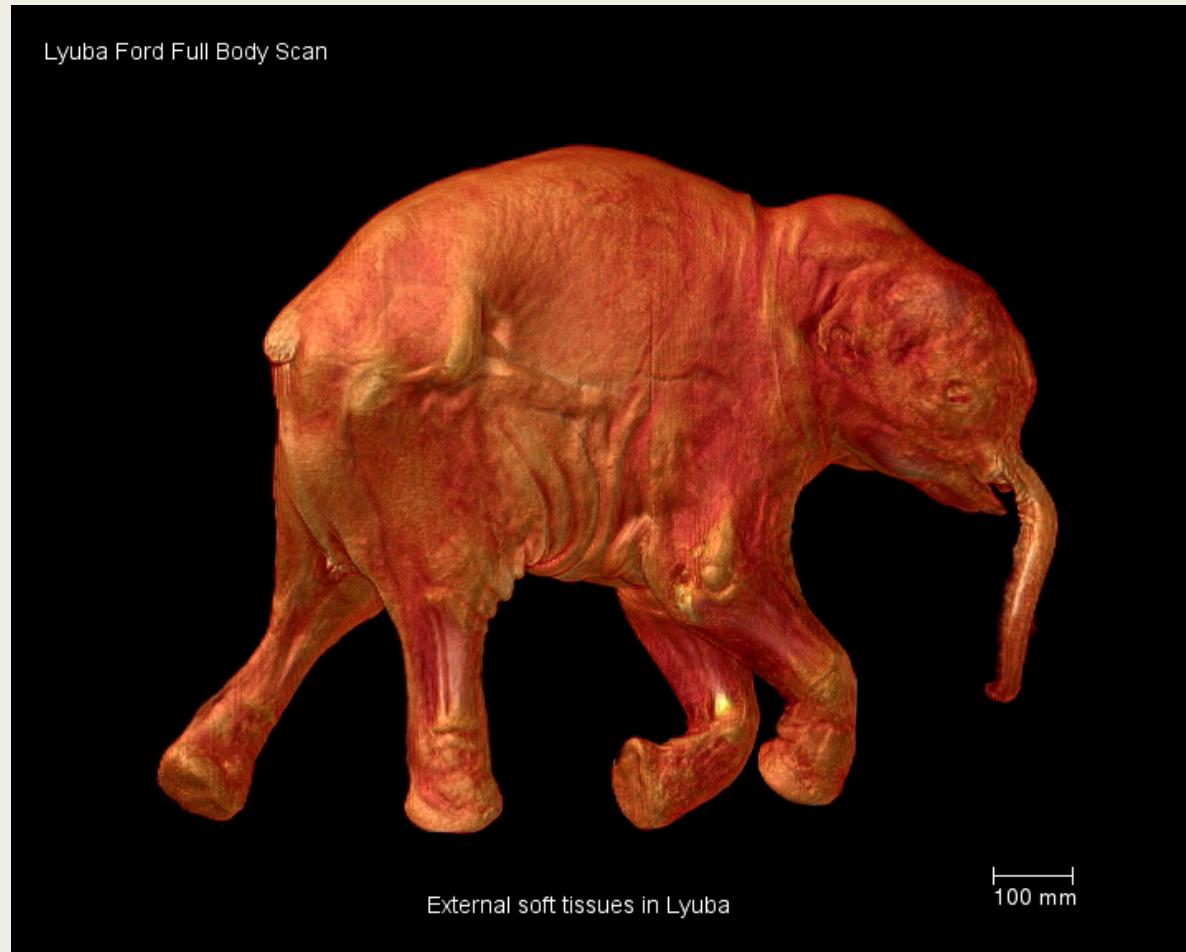
**Citation:** Jefferson T, Jones MA, Doshi P, Del Mar CB, Hama R, Thompson MJ, Spencer EA, Onakpoya I, Mahtani KR, Nunan D, Howick J, Heneghan CJ. Neuraminidase inhibitors for preventing and treating influenza in healthy adults and children. *Cochrane Database of Systematic Reviews* 2014, Issue 4. Art. No.: CD008965. DOI: 10.1002/14651858.CD008965.pub4.

...on of The Cochrane Collaboration,  
Carl Heneghan from the University  
...and their recent Cochrane  
...which holds valuable lessons  
...c health recommendations are



...y down viral spread giving time for

# Dryad is not just for tabular data II





# Also books, theses, other vetted publications

## 40 Years of Data

2014/04/03 by Peggy Schaeffer | Edit

### Odd couples in the animal kingdom, but not in a data repository

2013/04/30 by Peggy Schaeffer | Edit

We are celebrating the recent publication in Dryad of the first data to accompany a book [1, 2]. [Odd Couples: Extraordinary Differences Between the Sexes in the Animal Kingdom](#), from Princeton University Press, examines the occasionally surprising gender differences in animals, and what it means to be male or female in the animal kingdom. It is intended for both general and scientific readers.



A dominant male northern elephant seal attempts to copulate with a female. Photo by Derek Roff, courtesy Princeton Univ. Press.



A mature female *Argiope aurantia* (left) hanging at the hub of her orb web, with a mature male (right). Photo by Troy Bartlett, <http://www.naturecloseups.com>, courtesy Princeton Univ Press.

availability of the data underlying the book "40

Primary Grant.

account of

in's finches

from the

Press.

d

logical,

including

uses, and

rior-to

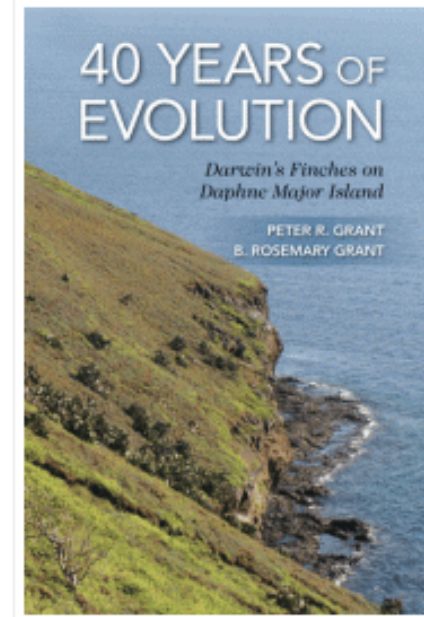
populations

the Major in

They find

ens

ridges and



# ReadMe files (aka poor man's codebook)

- One per data package or one per file
- Recommendations
  - Plain text
  - For each filename, a short description of contents
  - For tabular data: definitions of column headings and row labels; data codes (including missing data); and measurement units
  - Data processing steps not described elsewhere that may affect interpretation of results
  - What if any associated datasets are stored elsewhere
  - Whom to contact with questions

# What could data review mean?

# Data journals and data papers



# Biodiversity Data Journal review criteria



- Quality of the data
- Quality of the description
- Consistency between manuscript and data

Process includes multiple categories of reviewer & optional public comment

# Biodiversity Data Journal review criteria:

## 1. & 2. Quality of the data and description

- Are the data correct, given the protocols? Authors are encouraged to report any tests undertaken to address this point
- Are the data completely and consistently recorded within the dataset(s)?
- Does the data resource cover scientifically important and sufficiently large region(s), time period(s) and/or group(s) of taxa to be worthy of publication?
- Are the data consistent internally and described using applicable standards (e.g. in terms of file formats, file names, file size, units and metadata)?
- Are the methods used to process and analyses the raw data, thereby creating processed data or analytical results, sufficiently well documented that they could be repeated by third parties?
- Is the repository to which the data are submitted appropriate for the nature of the data?

# Biodiversity Data Journal review criteria:

## 3. Consistency between manuscript and data

- Does the manuscript provide an accurate description of the data?
- Does the manuscript properly describe how to access the data?
- Are the methods used to generate the data (including calibration, code and suitable controls) described in sufficient detail?
- Is the dataset sufficiently novel to merit publication?
- Have possible sources of error been appropriately addressed in the protocols and/ or the paper?
- Is anything missing in the manuscript or the data resource itself that would prevent replication of the measurements, or reproduction of the figures or other representations of the data?
- Are all claims made in the manuscript substantiated by the underlying data?

# NPG Scientific Data review criteria

## **Experimental Rigour and Technical Data Quality**

- Were the data produced in a rigorous and methodologically sound manner?
- Was the technical quality of the data supported convincingly with technical validation experiments and statistical analyses of data quality or error, as needed?
- Are the depth, coverage, size and/or completeness of these data sufficient for the types of applications or research questions outlined by the authors?

## **Completeness of the Description**

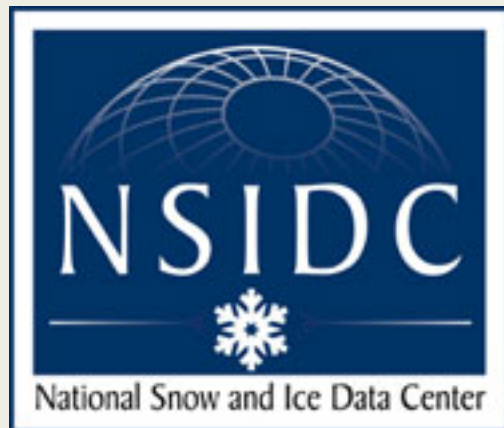
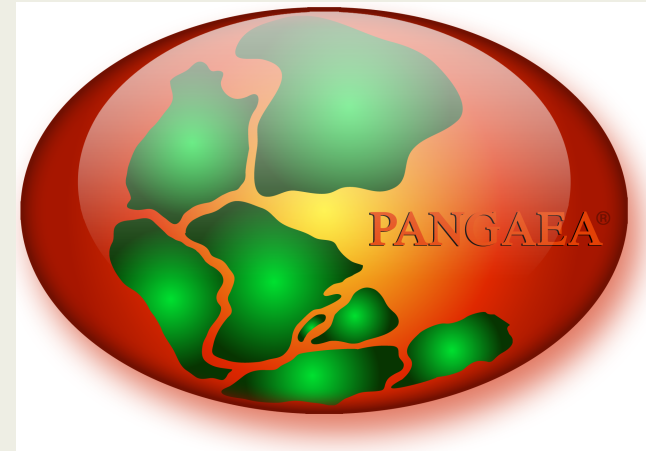
- Are the methods and any data-processing steps described in sufficient detail to allow others to reproduce these steps?
- Did the authors provide all the information needed for others to reuse this dataset or integrate it with other data?
- Is this Data Descriptor, in combination with any repository metadata, consistent with relevant minimum information or reporting standards?

## **Integrity of the Data Files and Repository Record**

- Have you confirmed that the data files deposited by the authors are complete and match the descriptions in the Data Descriptor?
- Have these data files been deposited in the most appropriate available data repository?



# Models of extensive review



# Automating technical review

## details of checkCIF/PLATON tests

This page provides a full listing of the tests that are carried out by a [prefilter](#) which checks for the data necessary to carry out the tests.

The tests check whether the data in a CIF are internally consistent. They were designed by crystallographers who have a particular knowledge of the CIF file format. Details of how to use checkCIF/PLATON reports are given in the [manual](#).

### Browse tests

Test name	Type	Purpose
<a href="#">ABSMU01</a>	1	To check that <code>_exptl_absorpt_coefficient</code> is a positive value.
<a href="#">ABSTY01</a>	1	To check that <code>_exptl_absorpt_correction</code> is a positive value.
<a href="#">ABSTY02</a>	1	To check that <code>_exptl_absorpt_correction</code> is a positive value.
<a href="#">ABSTY03</a>	1	To check that <code>_exptl_absorpt_correction</code> is a positive value.
<a href="#">CELLK01</a>	1	To check that <code>_cell_measurement_temperature</code> has the correct units ie. Kelvin rather than Celsius.
<a href="#">CELLT01</a>	1	To check that <code>_cell_measurement_theta_min</code> is less than <code>_cell_measurement_theta_max</code> .
<a href="#">CELLV01</a>	1	To check that the <code>_cell_volume</code> matches <code>_cell_length_</code> and <code>_cell_angle_</code> values.
<a href="#">CELLV02</a>	1	To check that the <code>_cell_volume</code> su matches <code>_cell_length_</code> and <code>_cell_angle_</code> su values.
<a href="#">CELLZ01</a>	1	To check that the cell contents calculated from <code>_chemical_formula_sum</code> and <code>_cell_formula_units_Z</code> matches that from the atom sites present in the <code>_atom_site_</code> list and the <code>_symmetry_</code> information.
<a href="#">CHEMS01</a>	1	To check that the <code>_chemical_formula_sum</code> is properly constructed. i.e. C H followed by alphabetic.
<a href="#">CHEMS02</a>	1	To check that the stated category <code>_publ_requested_category</code> is consistent with the chemical formula of the compound.
<a href="#">CHEMW01</a>	1	To check that the <code>_chemical_formula_weight</code> matches that calculated for the <code>_chemical_formula_sum</code> .
<a href="#">CHEMW03</a>	2	To check that the <code>_chemical_formula_weight</code> matches that calculated for the atomic content data from the two lists, <code>_atom_site_</code> and <code>_atom_type_</code> , using the number of formula units per cell and the number of symmetry equivalent positions. (NSYM from procedure SM3)

## research papers

Acta Crystallographica Section D  
**Biological  
Crystallography**  
ISSN 0907-4449

**Anthony L. Spek**

Utrecht University, Bijvoet Center for  
Biomolecular Research, Padualaan 8,  
3584 CH Utrecht, The Netherlands

Correspondence e-mail: a.l.spek@uu.nl

## Structure validation in chemical crystallography

Automated structure validation was introduced in chemical crystallography about 12 years ago as a tool to assist practitioners with the exponential growth in crystal structure analyses. Validation has since evolved into an easy-to-use *checkCIF/PLATON* web-based IUCr service. The result of a crystal structure determination has to be supplied as a CIF-formatted computer-readable file. The checking software tests the data in the CIF for completeness, quality and consistency.

Received 28 July 2008  
Accepted 22 December 2008

## COPE Ethical Guidelines for Peer Reviewers

Irene Hames on behalf of COPE Council  
March 2013, v.1

Peer review in all its form plays an important role in ensuring the integrity of the scholarly record. The process depends to a large extent on trust, and requires that everyone involved behaves responsibly and ethically. Peer reviewers play a central and critical part in the peer-review process, but too often come to the role without any guidance and may be unaware of their ethical obligations. The COPE Ethical Guidelines for Peer Reviewers set out the basic principles and standards to which all peer reviewers should adhere during the peer-review process. It is hoped they will provide helpful guidance to researchers, be a reference for journals and editors in guiding their reviewers, and act as an educational resource for institutions in training their students and researchers.

### Basic principles to which peer reviewers should adhere

Peer reviewers should:

- only agree to review manuscripts if they can carry out a proper assessment
- respect the confidentiality of the peer review, during or after the review process, and not to disclose information to other people, including colleagues, during or after the review process

*Peer reviewers should:*  
“not use information obtained during the peer-review process for their own or any other person’s or organization’s advantage, or to disadvantage or discredit others”

# DANS data review pilot

Below you can see how users have responded to the dataset 'De steentijd van Nederland'. The legend to the right explains the ratings.

### Ratings:

- 5: very good
- 4: good
- 3: neither good nor bad
- 2: insufficient
- 1: bad

Aspect	Rating						Average rating
	(5)	(4)	(3)	(2)	(1)	(n/a)	
data quality	4	7	1	0	0	0	★★★★★ (4.25/5)
quality of the documentation	4	6	2	0	0	0	★★★★☆ (4.17/5)
completeness of the data	4	5	2	0	0	1	★★★★☆ (4.18/5)
consistency of the dataset (if applicable)	1	4	2	1	0	4	★★★★☆ (3.63/5)
structure of the dataset (if applicable)	2	4	2	1	0	3	★★★★☆ (3.78/5)
usefulness of the file formats	3	5	4	0	0	0	★★★★☆ (3.92/5)

9 out of 12 reviewers of this dataset recommend the use of it.

2 out of 12 reviewers of this dataset have published using this dataset.

5 out of 12 reviewers of this dataset intend to use this dataset for a publication.

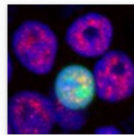
Figure 2. Screenshot with ratings, based on 12 reviews of the same data set.

Grootveld M, van Egmon J (2012) Peer-Reviewed Open Research Data: Results of a Pilot. IJDC doi:10.2218/ijdc.v7i2.231

# Post-publication review: images



Published  
27 Apr 2015



## Splicing function of mitotic regulators links R-loop-mediated DNA damage to tumor cell killing

Yihan Wan, Xiaobin Zheng, Haiyang Chen, Yuxuan Guo, Hao Jiang, Xiaonan He, Xueliang Zhu, Yixian Zheng

JCB vol. 209 no. 2 235-246 Article DOI: [10.1083/jcb.201409073](https://doi.org/10.1083/jcb.201409073) DataViewer DOI: [10.1083/jcb.201409073.dv](https://doi.org/10.1083/jcb.201409073.dv)

### Viewing Options

- Normal
- Max Intensity
- Split Channel

Quality

Zoom

Line Plot

### Rendering Details

Edit Channels

BuGZ  GFP-RNase H1  DAPI

Color

### Current Image

Z: 1/1 | T: 1/1

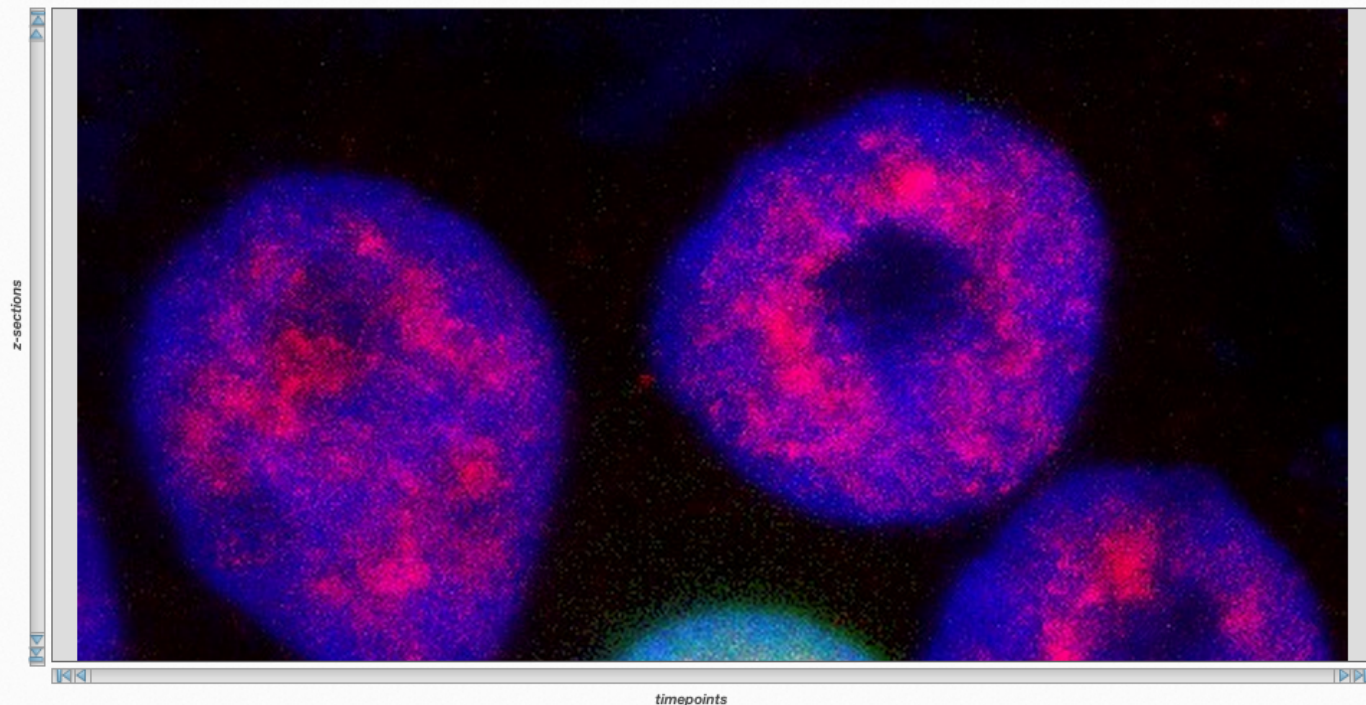
[Image Information](#)

[Image Legend](#)

[Image Link](#)

[Make Movie](#)

[Download as OME-TIFF](#)





# Post-publication review: citation

*Ecological Monographs*, 82(2), 2012, pp. 221–228  
© 2012 by the Ecological Society of America

## Novel forests maintain ecosystem processes after the decline of native tree species

JOSEPH MASCARO,<sup>1,4</sup> R. FLINT HUGHES,<sup>2</sup> AND STEFAN A. SCHNITZER<sup>1,3</sup>

<sup>1</sup>*Department of Biological Sciences, University of Wisconsin, Milwaukee, Wisconsin 53211 USA*

<sup>2</sup>*Institute for Pacific Islands Forestry, USDA Forest Service, Hilo, Hawaii 96720 USA*

<sup>3</sup>*Smithsonian Tropical Research Institute, Apartado 2072, Balboa, Republic of Panama*

**Abstract.** The positive relationship between species diversity (richness and evenness) and critical ecosystem functions, such as productivity, carbon storage, and nutrient cycling, is often used to predict the consequences of extinction. At regional scales, however, plant species richness is mostly increasing rather than decreasing because successful plant species introductions far outnumber extinctions. If these regional increases in richness lead to local

Dombois, and P. A. Matson. 1987. Biological invasion by *Myrica faya* alters ecosystem development in Hawaii. *Science* 238:802–804.

Wagner, W. L., D. R. Herbst, and S. H. Sohmer. 1999. Manual of the flowering plants of Hawai'i. University of Hawai'i Press/Bishop Museum Press, Honolulu, Hawaii, USA.

Walker, L. R., and R. del Moral. 2003. Primary succession and ecosystem rehabilitation. Cambridge University Press, Cambridge, UK.

Wardle, D. A. 2002. Communities and ecosystems: linking the aboveground and belowground components. Princeton University Press, Princeton, New Jersey, USA.

Wardle, D. A., R. D. Bargett, R. M. Callaway, and W. H. Van der Putten. 2011. Terrestrial ecosystem responses to species gains and losses. *Science* 332:1273–1277.

and J. Pastor. 1993. Nitrogen mineralization in grass monocultures. *Oecologia* 96:186–192.

M. 2004. The parable of Green Mountain: land, ecosystem construction, and ecological biogeography. *Biogeography* 31:1–4.

P. Daneshgar, and H. W. Polley. 2011. Phenology and temporal niche differences between native and novel exotic-dominated grasslands. *Plant Ecology, Evolution and Systematics*

B. Teaschner, P. P. Daneshgar, F. I. Isbell, and J. M. 2009. Biodiversity maintenance mechanisms in native and novel exotic-dominated communities. *Letters* 12:432–442.

and J. Morris. 1996. Geologic map of the Island of Hawaii. I-2524-A. USGS, Denver, Colorado, USA.

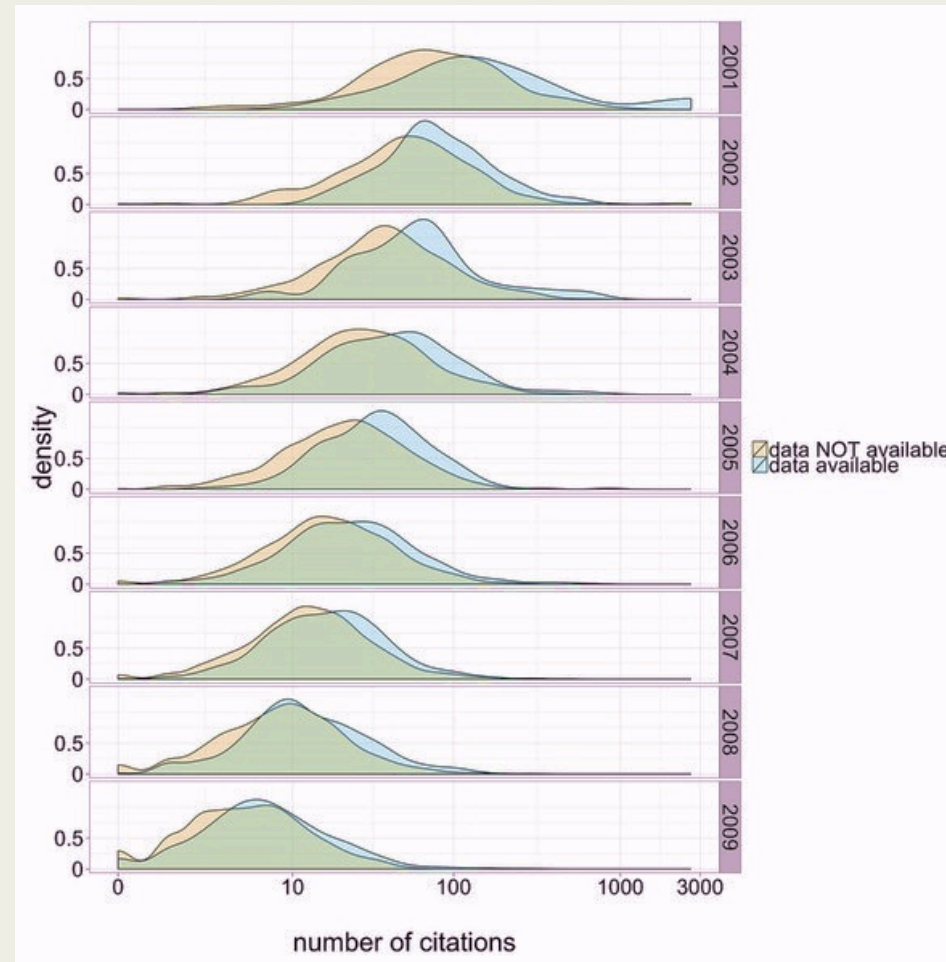
Woodcock, D. 2003. To restore the watersheds: Early twentieth-century tree planting in Hawai'i. *Annals of the Association of American Geographers* 93:624–655.

Zanne, A. E., G. Lopez-Gonzalez, D. A. Coomes, J. Ilic, S. Jansen, S. L. Lewis, R. B. Miller, N. G. Swenson, M. C. Wiemann, and J. Chave. 2009. Global wood density database. Dryad Digital Repository, North Carolina, USA. <http://dx.doi.org/10.5061/dryad.234>

Ziegler, A. C. 2002. Hawaiian natural history and evolution. University of Hawai'i Press, Honolulu, Hawaii, USA.

Zimmerman, N., R. F. Hughes, S. Cordell, P. Hart, H. K. Chang, D. Perez, R. K. Like, and R. Ostertag. 2008. Patterns of primary succession of native and introduced plants in lowland wet forests in Eastern Hawai'i. *Biotropica* 40:277–284.

# Post-publication review: usage



Piwowar and Vision (2013) doi:10.7717/peerj.175

# Post-publication review: usage

## ImpactStory.

**Emilio M. Bruna**

Overview

Map

Fans

articles (83)

books (1)

**datasets (8)**

software products (8)

8 Datasets

4.6k views

6 tweets

6 shares

Show tweets (6)

Sorting by default

**Data from: Resilient networks of ant-plant mutualists in Amazonian forest fragments**

(2012) Passmore, Bruna, Heredia et al.. *Dryad Digital Repository*.

view dataset

highly viewed +3

cited

discussed

viewed

**Data from: Asymmetric dispersal and colonization success of Amazonian plant-ants queens**

(2011) Bruna, Izzo, Inouye et al.. *Dryad Digital Repository*.

view dataset

highly viewed +4

cited

viewed

**Data from: Edge effects on growth and biomass partitioning of an Amazonian understory herb (*Heliconia acuminata*; Heliconiaceae)**

(2012) Bruna, Segalin de Andrade. *Dryad Digital Repository*.

view dataset

highly viewed +3

discussed

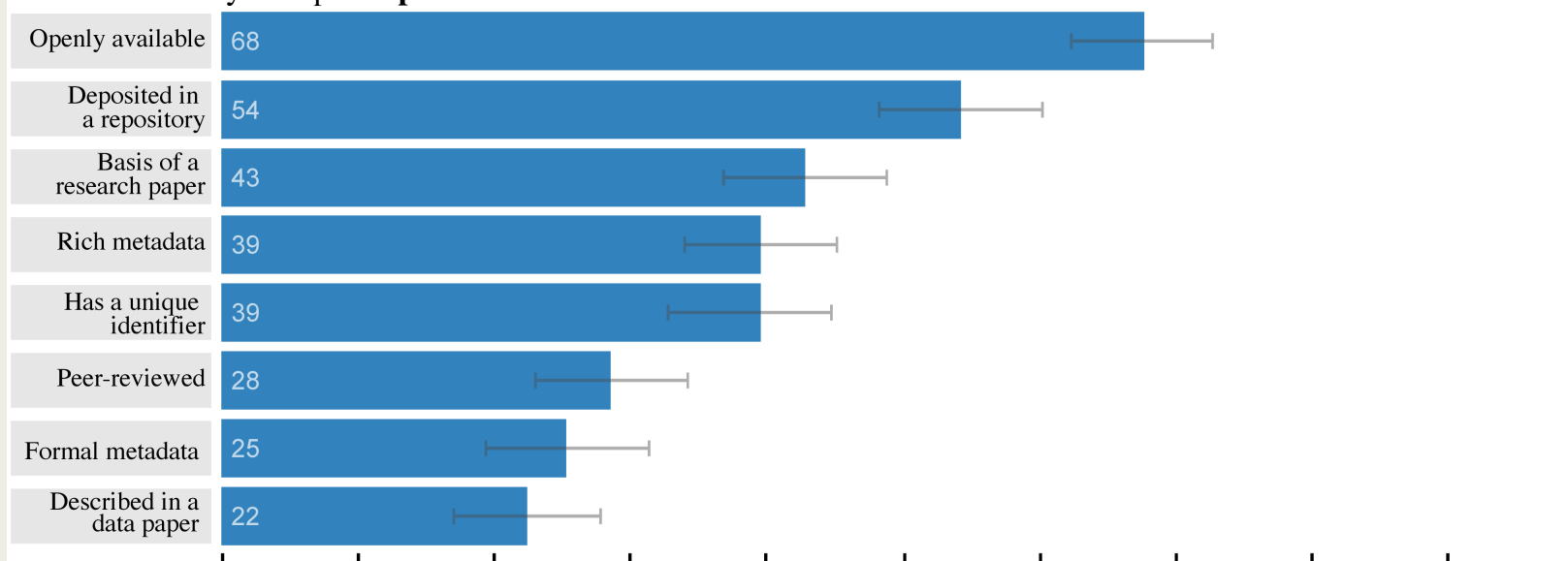
viewed



# What do researchers think data review means?

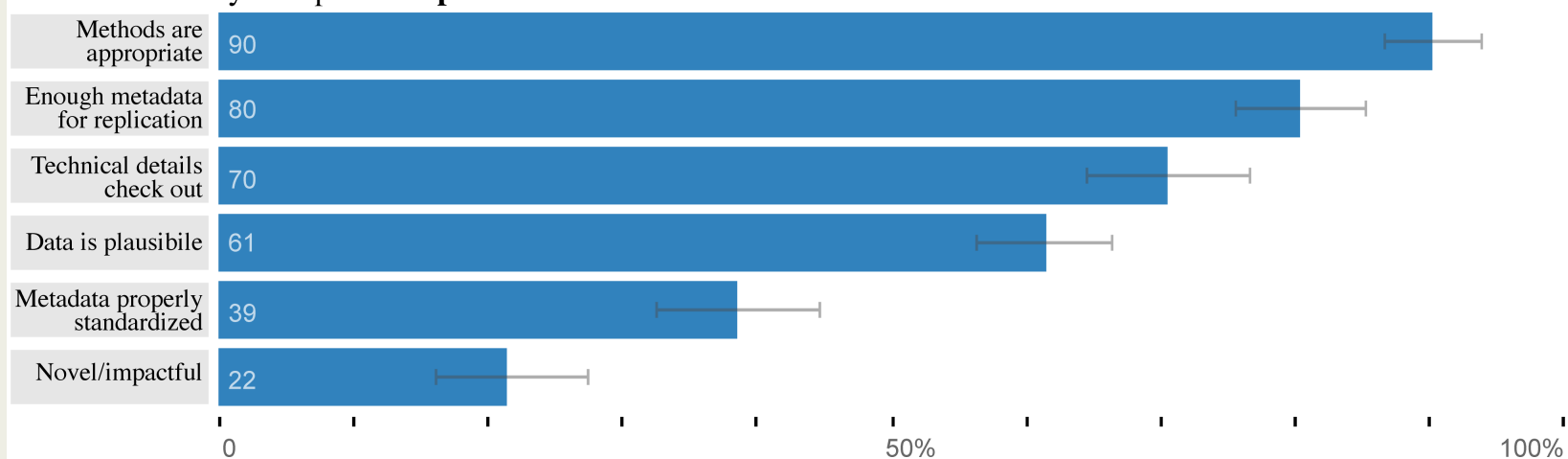
**A. How would you expect a **published** dataset to differ from a **shared** one?**

n = 246



**B. What would you expect data **peer review** to consider?**

n = 244



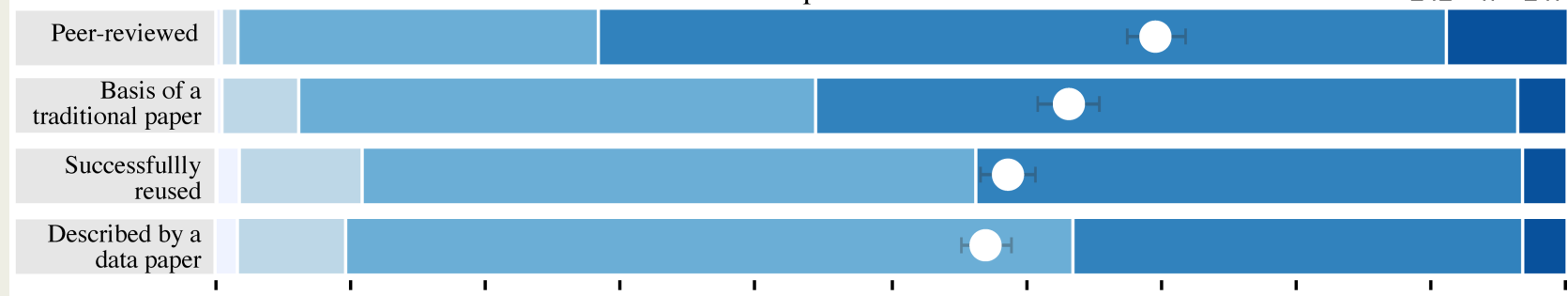
Kratz JE, Strasser C (2015) Researcher Perspectives on Publication and Peer Review of Data. PLoS ONE doi:10.1371/journal.pone.0117619

7-May-2015

National Data Integrity Conference

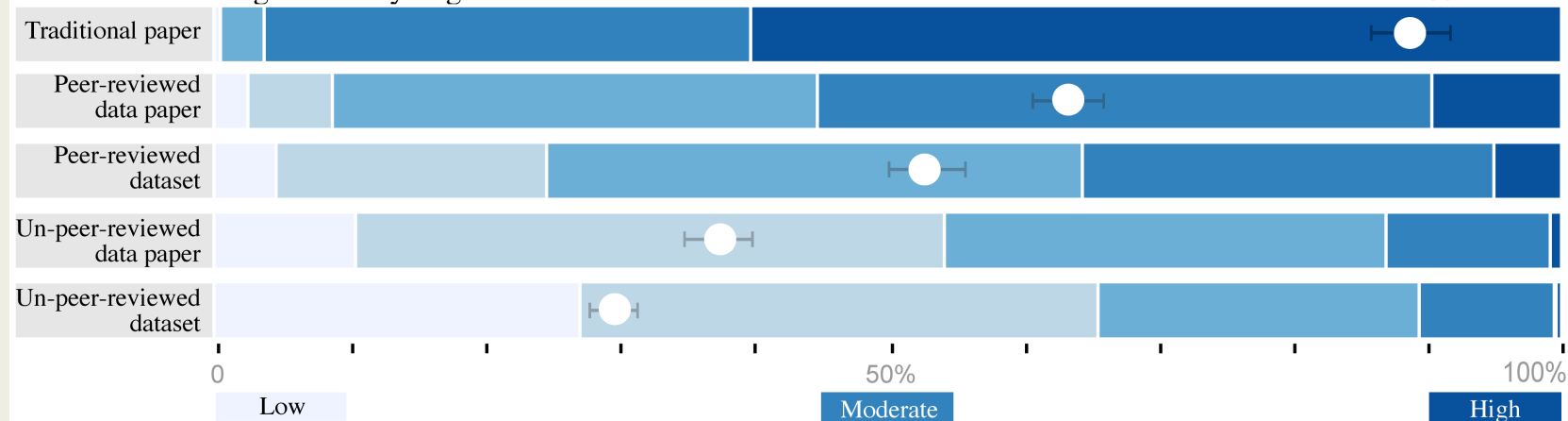
**B. How much confidence in a dataset does each attribute inspire?**

242 ≤ n ≤ 247



**D. How much weight would you give each item on a researcher's CV?**

238 ≤ n ≤ 241



Kratz JE, Strasser C (2015) Researcher Perspectives on Publication and Peer Review of Data. PLoS ONE doi:10.1371/journal.pone.0117619

# Researcher expectations of published data

- Only a minority *expect* peer review
- What peer review of data means
  - Firstly, review of methods & metadata
  - Secondly, quality review of data itself
  - It involves domain expertise (not just technical review)
- Peer review of data inspires more confidence than it being published upon or reused
- But peer reviewed data confers little professional credit

# Is software review a better metaphor?

arXiv.org > cs > arXiv:1407.5648

Search or

Computer Science > Software Engineering

## Code Review For and By Scientists

[Marian Petre](#), [Greg Wilson](#)

*(Submitted on 21 Jul 2014 (v1), last revised 7 Sep 2014 (this version, v2))*

We describe two pilot studies of code review by and for scientists. Our principal findings are that scientists are enthusiastic, but need to be shown code review in action, and that just-in-time review of small code changes is more likely to succeed than large-scale end-of-work reviews.

Comments: 4 pages

Subjects: **Software Engineering (cs.SE)**

Cite as: [arXiv:1407.5648](#) [cs.SE]

(or [arXiv:1407.5648v2](#) [cs.SE] for this version)

# Is software review a better metaphor?

- “...an obvious conclusion here is that software is fully fair game for in depth peer review, right? (Never mind that most scientists probably aren't capable of doing good peer review of code, or that any reasonably strong code review requirements would mean that virtually no more software would be published - an effective but rather punitive way to ensure only good software is published in science”


C. Titus Brown <http://ivory.idyll.org/blog/2015-we-live-in-a-bubble.html>

# What do we need?

- More data on current practice?
- More experimentation?
- Consensus among journals and repositories?
- Training for authors and reviewers?
- Better software tools?
- Business models?
- Funding mandates?

# More information



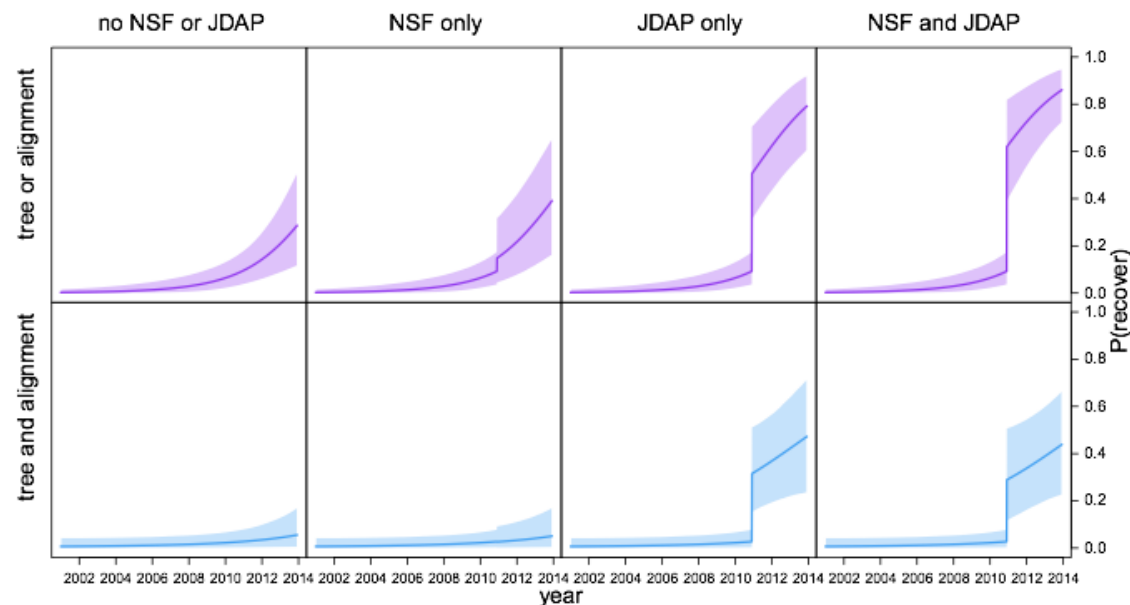
- <http://datadryad.org>
- <http://blog.datadryad.org>
-  @datadryad







# Effects of JDAP since 2011



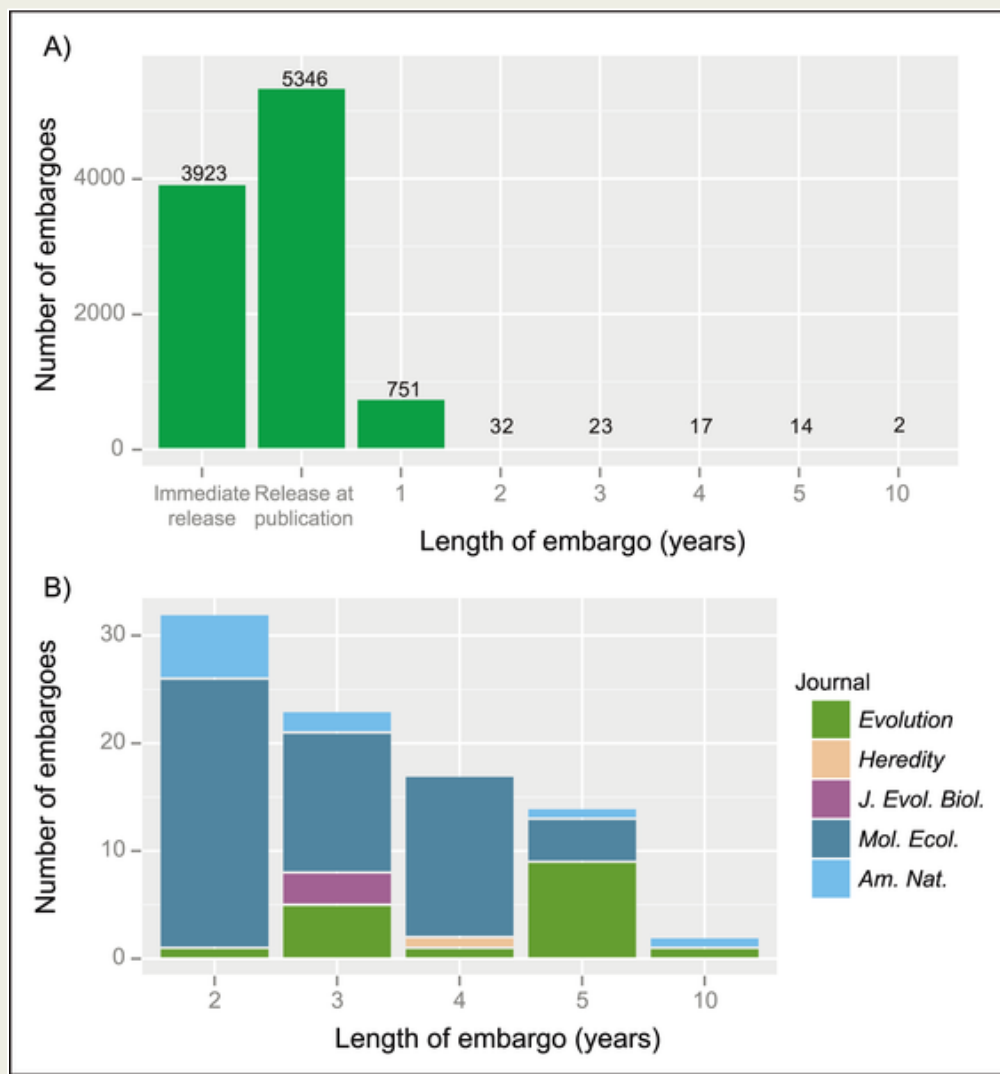
**Figure 5. Availability of archived phylogenetic data as a function of age.** We estimated the effect of publication age on our ability to procure partial (top panels) and complete (bottom panels) phylogenetic datasets from online archives. Overall, the probability of recovering archived phylogenetic data increases toward the present, with a conspicuous recent increase for partial datasets (left panels). The recent surge of archived phylogenetic data likely reflects recent policy changes (middle panels): studies with NSF funding are more likely to archive alignment (but not tree) files (*c.f.*, Table [S.14](#)); whereas studies published in journals with JDAP membership are dramatically more likely to archive both partial and complete phylogenetic datasets. The effects of these policy initiatives are not strictly additive (right panels): the correlation of these predictor variables suggests that studies published in JDAP journals are likely to have NSF funding. Shaded areas reflect the 95% credible intervals.

Magee *et al.* (2014) Dawn of open access to phylogenetic data. arXiv:14.1405.6623.v1

# Data on Data @PLOS

- From Jennifer Lin ACRS
- April 1, 2014 - early Feb 2015)
- # of papers with Data Availability Statements ~11k
- Fully shared
- % data deposited in repositories ~11% % all data included in manuscript/Supplemental Information files ~57%
- Restrictions
- % data held by 3rd parties ~2%
- % data containing sensitive information (available ~2% upon request)

# Embargoes are the exception, not the rule



**A.** Embargo selections of Dryad data authors for the 10,108 files in Dryad deposited from inception to September 20, 2013. Data include only datasets related to articles published in journals for which the authors had the option of selecting an embargo.

**B.** Long-term embargoes (>1 year) by journal that granted them.

Data: Vision TJ, Scherle R, Mannheimer S (2013) Embargo selections of Dryad data authors. FigShare. <http://doi.org/10.6084/m9.figshare.805946>.

Article: Roche DG, Lanfear R, Binning SA, Haff TM, et al. (2014) Troubleshooting Public Data Archiving: Suggestions to Increase Participation. PLoS Biol 12(1): e1001779

# Data publication

## Data Quality:

- • Is the format acceptable? If so, is there an automatic format checker available, and if there is, does the dataset/file pass the automated checks?
- • Are data values internally consistent? Do they fall within an appropriate range for the phenomenon being measured/observed/simulated? (For example, does a temperature dataset with values in degrees Kelvin have negative values?)
- • Does the data represent reality with sufficient accuracy to use? Is the data of tolerable precision? (In the case of simulations, can the simulation be trivially repeated, in which case publication of the data is probably unwarranted.)
- • Does the extent and coverage of the data match expectations? Does the coverage (spatial and/or temporal) add significant value to what is already available? (If not, is there added precision or some other reason for its publication? See also the discussion below on granularity.)
- • Are the data values reported physically possible and plausible? (This requires significant domain knowledge, or a clear definition of what the data values range should be.)
- • Is the data validated against an independent dataset? (Has it been calibrated?)

## Metadata Quality:

- • Is there sufficient quality metadata describing the format and physical content? (See for example, the requirements of the PDS, 2009.)
- • Is there sufficient quality metadata describing provenance and context? Has the data changed in some way since it was measured? Is the processing chain visible and well documented? Have all the human interactions with the data prior to ingest/publication been recorded?
- • Is there existing metadata (or are there references) already making assertions about the quality and usefulness of the data? If so, are these included in the metadata?
- • Is there suitable quality discovery metadata? At a bare minimum, can Dublin Core be constructed?
- • Does the metadata use appropriate, controlled vocabularies?
- • Can all internal references (both electronic, e.g. URL/DOI, and traditional, e.g. to ISO690) be resolved to real entities? Are the external electronic references stored in a trusted repository? If not, can they be cached with the metadata?
- • Is all the available metadata conforming to standards?

## General:

- • Is there an existing user community? Is that community happy that the data is usable? (This can be tracked after publication through citation, or before publication through the use of user surveys or comments in a process of open peer review.)
- • What is the track record of the data source? Are they reliable?
- • Are the intellectual property rights for the data established?
- • Is the data available at the correct network address?

In some cases there will be electronic services, such as visualizations, associated with the data, in which case the reviewer will need to address the service/data compatibility and function:

- • Do the advertised services work with the data? Is it likely that these services can be maintained with time?