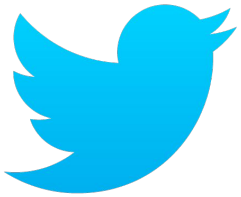
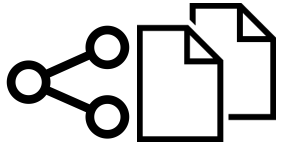
Do you speak Open Science?



paola masuzzo
@pcmasuzzo

You are **free** to:



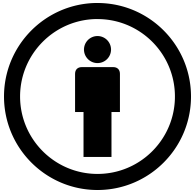
share, adapt, or re-mix

photograph, video or broad-cast

blog, live-blog or post-video

this presentation

Provided that:



you **attribute** the work to its author and **respect** the rights and licenses associated with its components

Do you speak open science? Resources and tips to learn the language

Science and Medical Education

Paola Masuzzo[1,2], Lennart Martens[1,2]

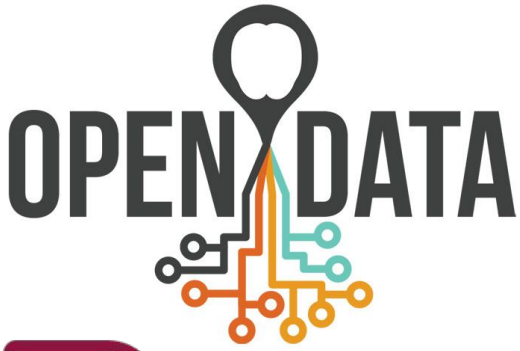
January 3, 2017

Author and article information

Abstract

The internet era, large-scale computing and storage resources, mobile devices, social media, and their high uptake among different groups of people, have all deeply changed the way knowledge is created, communicated, and further deployed. These advances have enabled a radical transformation of the practice of science, which is now more open, more global and collaborative, and closer to society than ever. Open science has therefore become an increasingly important topic. Moreover, as open science is actively pursued by several high-profile funders and institutions, it has fast become a crucial matter to all researchers. However, because this widespread interest in open science has emerged relatively recently, its definition and implementation are constantly shifting and evolving, sometimes leaving researchers in doubt about how to adopt open science, and which are the best practices to follow.

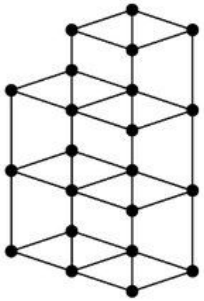
What is open science exactly?



open notebook

open source

open peer review



doing it together science

open source hardware

OPEN ACCESS

OPEN DATA

OPEN ACCESS

OPEN CODE

OPEN PEER REVIEW

OPEN DATA

Open data implies freedom to access, use and re-use for any purpose

Open Definition 2.1

Summary: *Knowledge is open if anyone is free to access, use, modify, and share it — subject, at most, to measures that preserve provenance and openness.*

OPEN KNOWLEDGE

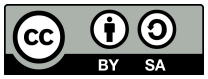
open knowledge definition conformant licenses

CC0 waiver

CC-BY (Attribution only)

CC-BY (Attribution ShareAlike)

http://opendefinition.org/od/ - http://opendefinition.org/licenses/



Research data are first-class citizens in science

REPORT

The Availability of Research Data Declines Rapidly with Article Age

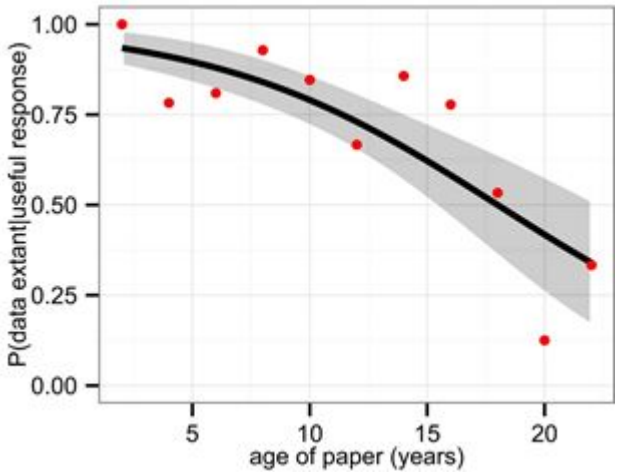
Timothy H. Vines, Arianne Y.K. Albert, Rose L. Andrew, Florence Débarre, Dan G. Bock, Michelle T. Franklin, Kimberly J. Gilbert, Jean-Sébastien Moore, Sébastien Renaut, Diana J. Rennison

Published Online: December 19, 2013

Open Archive DOI: http://dx.doi.org/10.1016/j.cub.2013.11.014 | CrossMark

Article Info

Altmetric 656



Vines et al., Current Biology, 2014; image courtesy Auke Herrema

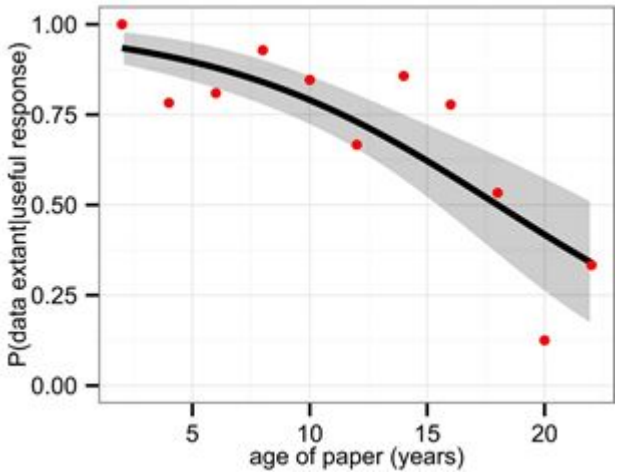
Research data are first-class citizens in science

REPORT

The Availability of Research Data Declines Rapidly with Article Age

Timothy H. Vines, Arianne Y.K. Albert, Rose L. Andrew, Florence Débarre, Dan G. Bock, Michelle T. Franklin, Kimberly J. Gilbert, Jean-Sébastien Moore, Sébastien Renaut, Diana J. Rennison

Published Online: December 19, 2013
Open Archive DOI: http://dx.doi.org/10.1016/j.cub.2013.11.014 | CrossMark
Article Info
Altmetric 656



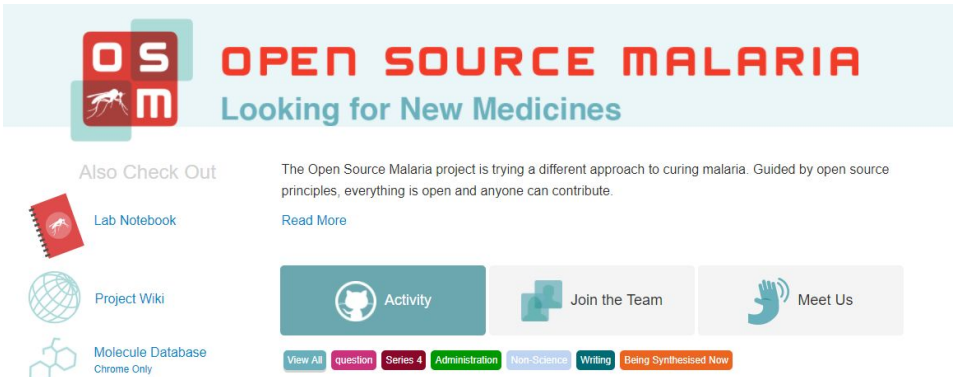
Data should be considered the **primary output of research**

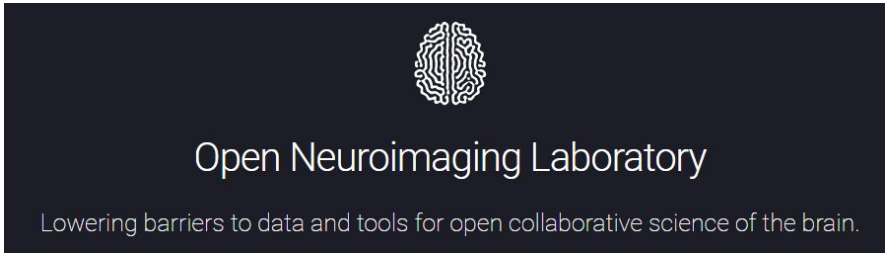
Vines et al., Current Biology, 2014; image courtesy Auke Herrema

Data produced with public funds belong to the public

Image courtesy Auke Herrema

Open data means more hands at work, more brain power and faster innovations





nextstrain

This work is made possible by the **open sharing of genetic data by research groups from all over the world**.

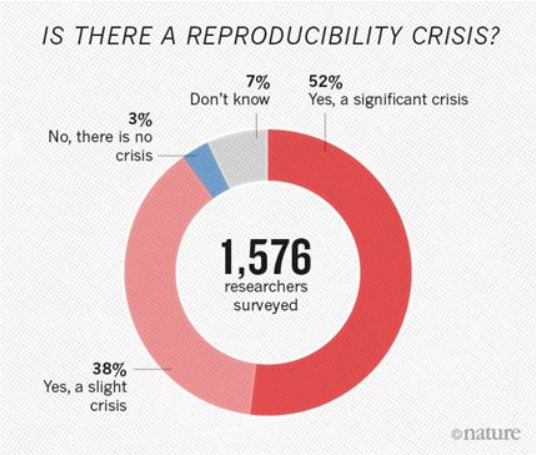
http://opensourcemalaria.org/; http://openneu.ro/start/; http://www.nextstrain.org/

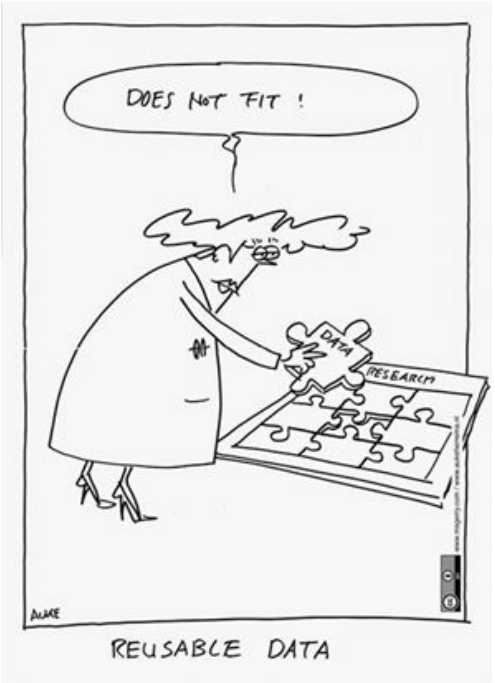
Open data creates a culture of transparency and reproducibility

Sluggish data sharing hampers reproducibility effort

Initiative trying to validate 50 cancer papers finds difficulty in accessing original study data.

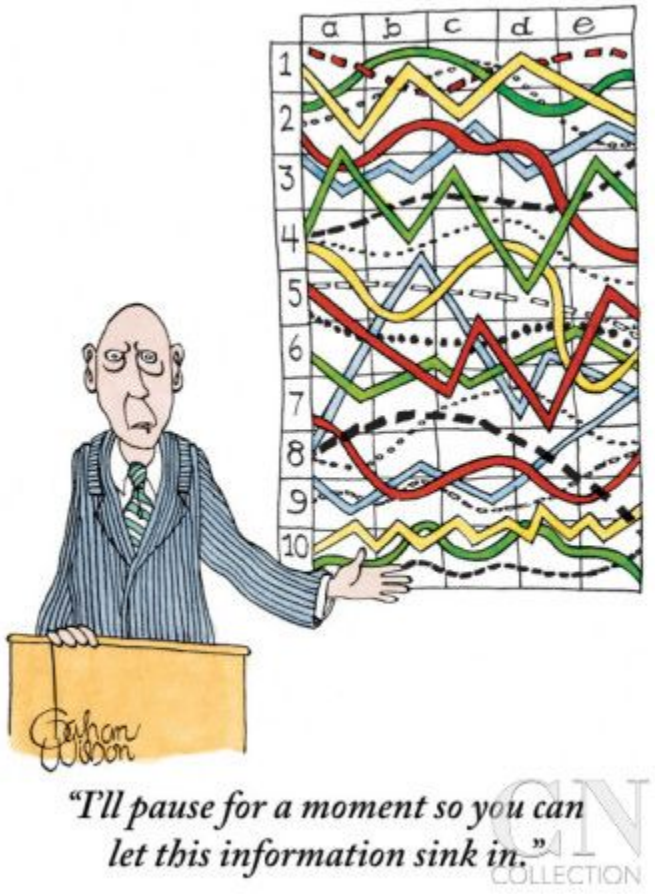
Richard Van Noorden





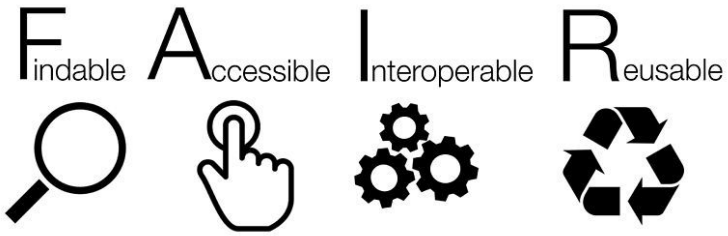
Monya Baker, Nature, 2016; image courtesy Auke Herrema

Dumping a spreadsheet online is not open data



Gahan Wilson, http://gahanwilson.net/, adapted from Erin C. McKiernan, 2016

Intelligent access and interoperability are crucial



What do **abbreviations** mean?

Are **metadata** properly annotated?

Am I using standard **formats**?

Am I using controlled **terminologies**?

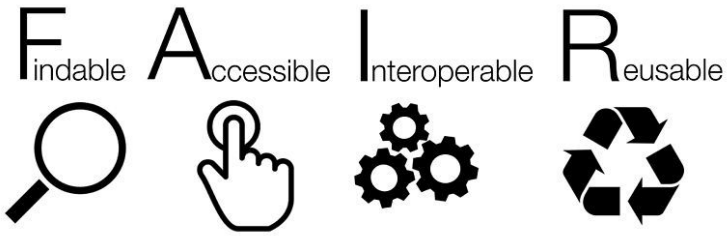
Are **missing data** explained?

Is the **license** clearly mentioned and described?

Am I using a **trusted repository**?

FAIR data principles, WikiMedia Commons, CC BY-SA 4.0

Intelligent access and interoperability are crucial



What do **abbreviations** mean?

Are **metadata** properly annotated?

Am I using standard **formats**?

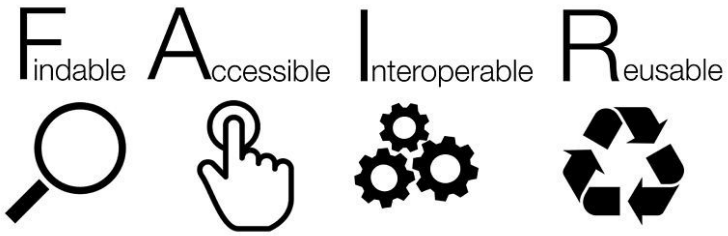
Am I using controlled **terminologies**?

Are **missing data** explained?

Is the **license** clearly mentioned and described?

FAIR data principles, WikiMedia Commons, CC BY-SA 4.0

Intelligent access and interoperability are crucial



RCSB PDB PROTEIN DATA BANK

IDR

NEURO VAULT

CXIDB Coherent X-ray Imaging Data Bank

OpenfMRI

OpenTrials
All the Data, on All the Trials, Linked

What do **abbreviations** mean?

Are **metadata** properly annotated?

Am I using standard **formats**?

Am I using controlled **terminologies**?

Are **missing data** explained?

Is the **license** clearly mentioned and described?

Am I using a **trusted repository**?

zenodo

figshare

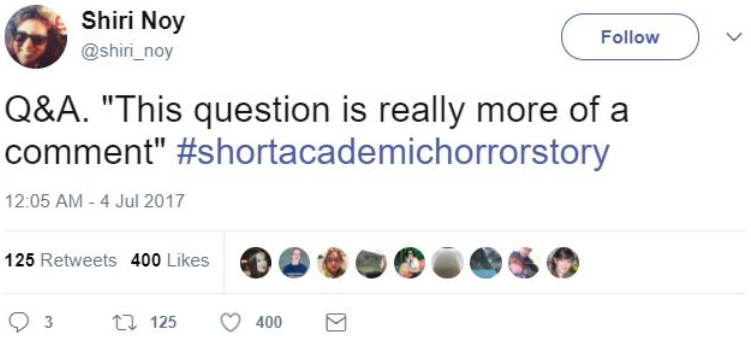
DRYAD

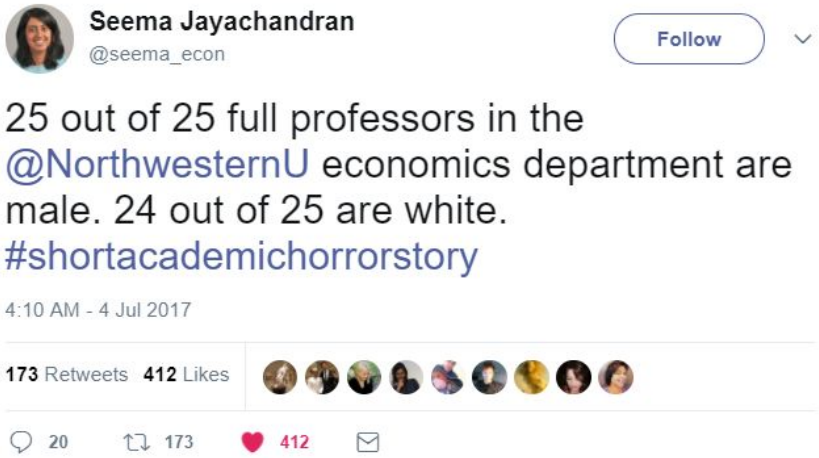
COS CENTER FOR OPEN SCIENCE

FAIR data principles, WikiMedia Commons, CC BY-SA 4.0

OPEN ACCESS

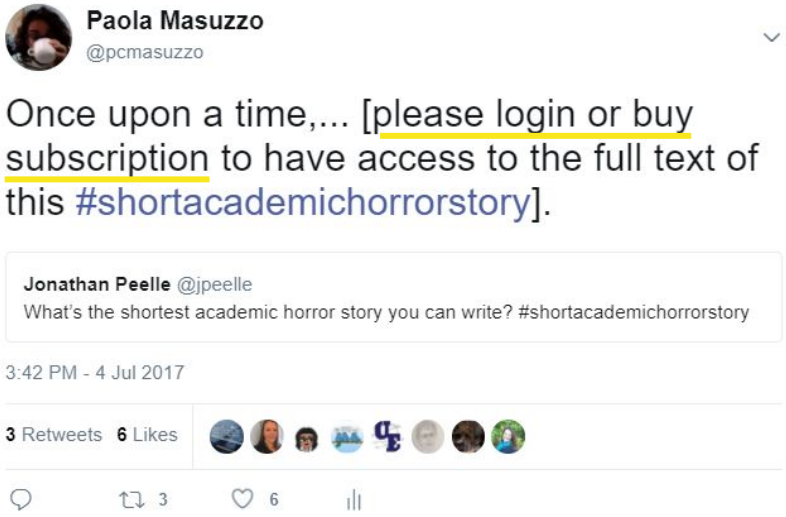
There are academic horror stories everywhere, everyday





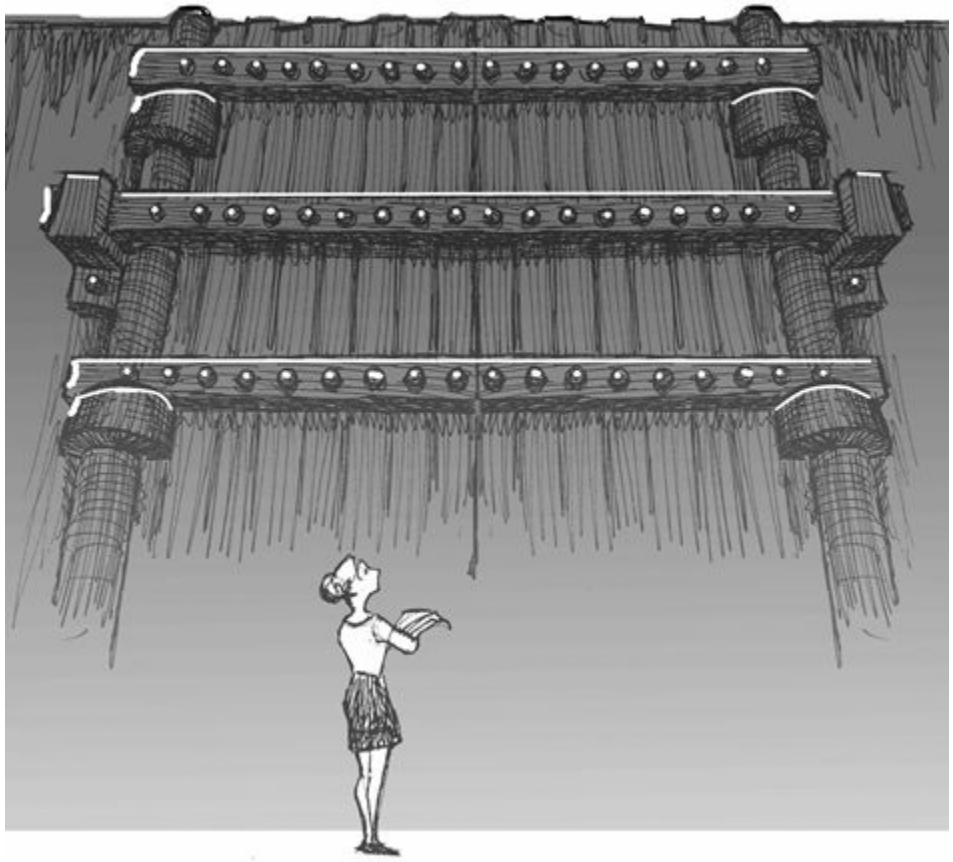
#shortacademichorrorstory

But the real academic monsters are paywalls and the impact factor



#shortacademichorrorstory

Putting research articles behind paywalls is bad, immoral, evil



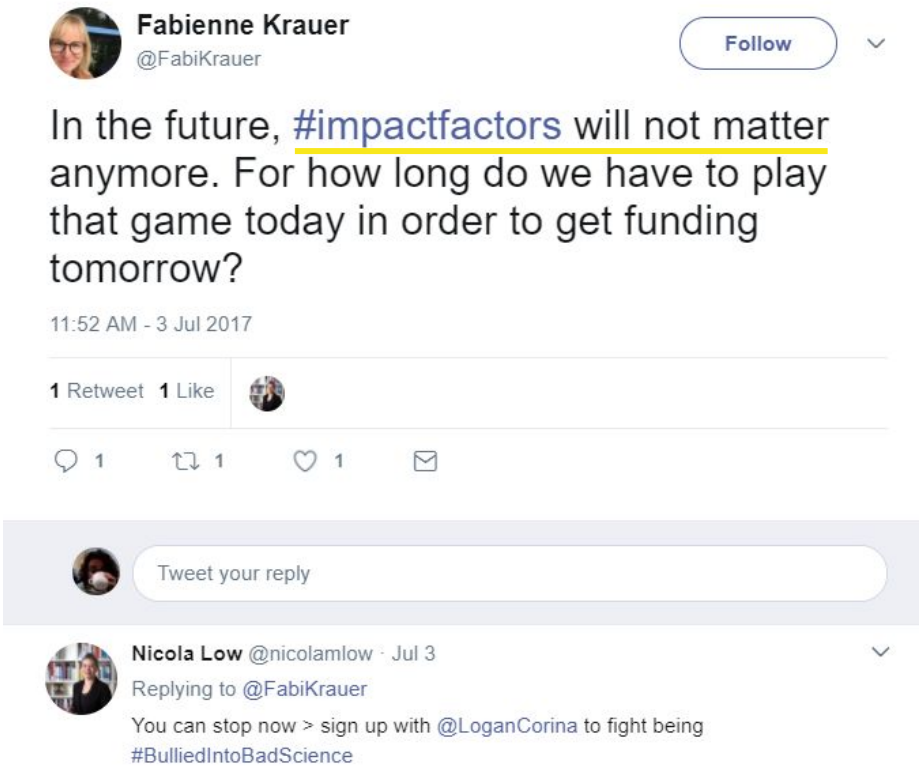
Cartoon by John R. McKiernan http://whyopenresearch.org, CC BY-NC-ND

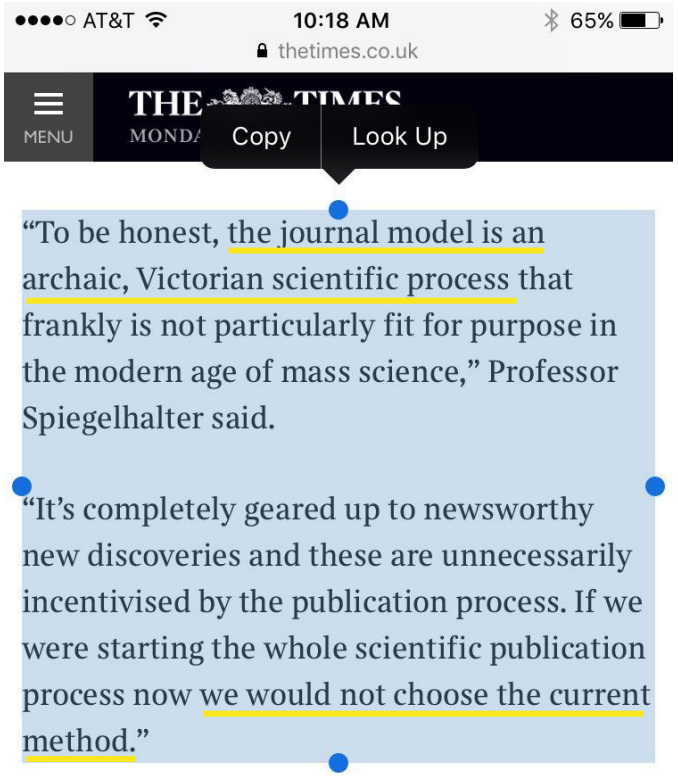
In 2010, Elsevier's reported a 36% profit margin – higher than Apple, Google, or Amazon posted that year

https://www.theguardian.com/science/2017/jun/27/profitable-business-scientific-publishing-bad-for-science

One of the big problems is that too many institutions still worship the impact factor

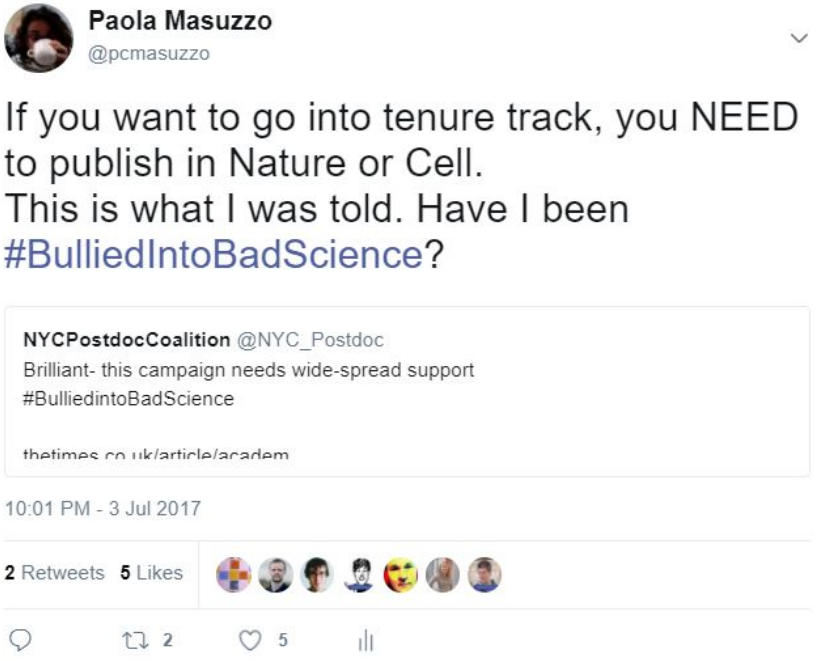
The current journal model is archaic and not fit for purpose in the modern scientific era





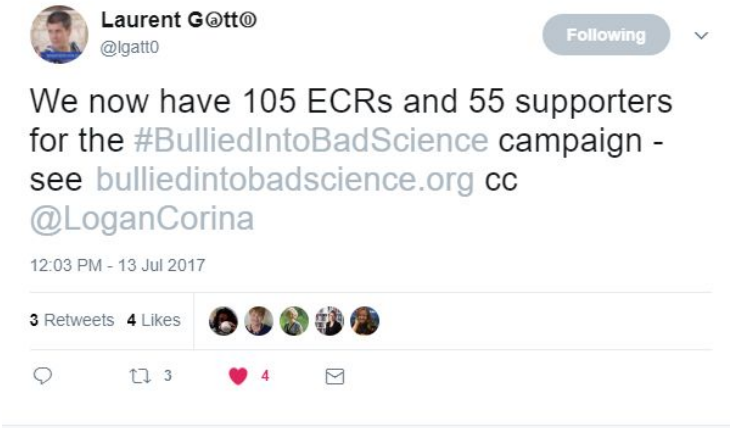
https://www.thetimes.co.uk/article/academics-strike-back-against-bad-science-t5mzcdxv8

I have been bullied into bad science



#BulliedIntoBadScience - http://bulliedintobadscience.org/

Don’t wait on others to change things for you, be the change yourself



Do you feel bullied into bad science?

Dear Early Career Researcher (ECR),

We are postdocs and a reader in the humanities and sciences at the University of Cambridge. **We are concerned about the desperate need for publishing reform** to increase transparency, reproducibility, timeliness, and academic rigour of the production and dissemination of scholarly outputs (see Young et al. 2016, Smaldino & McElreath 2016).

We have identified actions that institutions and managers can take to better support ECRs (below). These actions are crucial for our success because we are eager to publish openly and at places that keep profits inside academia in accordance with many modern online publication venues (Logan 2017). However, **ECRs are often pressured into publishing against their ethics** through threats that we would not get a job/grant unless we publish in particular journals (Carter et al. 2014, Who is going to make change happen?, Kent 2016; usually these journals are older and more familiar, have a print version, a high impact factor, and are not 100% open access). These out of date practices and ideas hinder ECRs rather than help us: evidence shows that publishing open access results in increased citations, media attention, and job/funding opportunities (McKiernan et al. 2016). Open dissemination of all research outputs is also a fundamental principle on which ECRs rely to fight the ongoing reproducibility crisis in science and thus improve the quality of their research.

Non-ECRs: support the campaign

You are an established researcher (Professor) or in a position connected with academia and want to show your support: fill out this form.

Show 5 entries Search:

50 ECR campaign supporters as of Wed Jul 05 2017 06:00:13

http://bulliedintobadscience.org/

Research evaluation criteria need to change

800+ organizations and 12,000+ individuals have committed:

- **not** to consider JIF in hiring, promotion or funding decisions
- the scientific content of an article, not the JIF of the journal where it was published, is **what matters**
- to consider the value and impact of **all the research outputs**

http://www.ascb.org/dora/ - www.ascb.org/SFdeclaration.html

Research evaluation criteria need to change

San Francisco DORA
Declaration on Research Assessment

800+ organizations and 12,000+ individuals have committed:

- **not** to consider JIF in hiring, promotion or funding decisions
- the scientific content of an article, not the JIF of the journal where it was published, is **what matters**
- to consider the value and impact of **all the research outputs**

HAVE YOU SIGNED YET?

http://www.ascb.org/dora/ - www.ascb.org/SFdeclaration.html

But if you really, really care, many open access journals have moderate to high IF

Cell Reports

IF 7.87

PLOS MEDICINE

IF 13.58

BMC Biology

IF 6.78

nature COMMUNICATIONS

IF 12.12

PLOS | BIOLOGY

IF 9.80

PeerJ

IF 2.18

SCIENTIFIC REPORTS

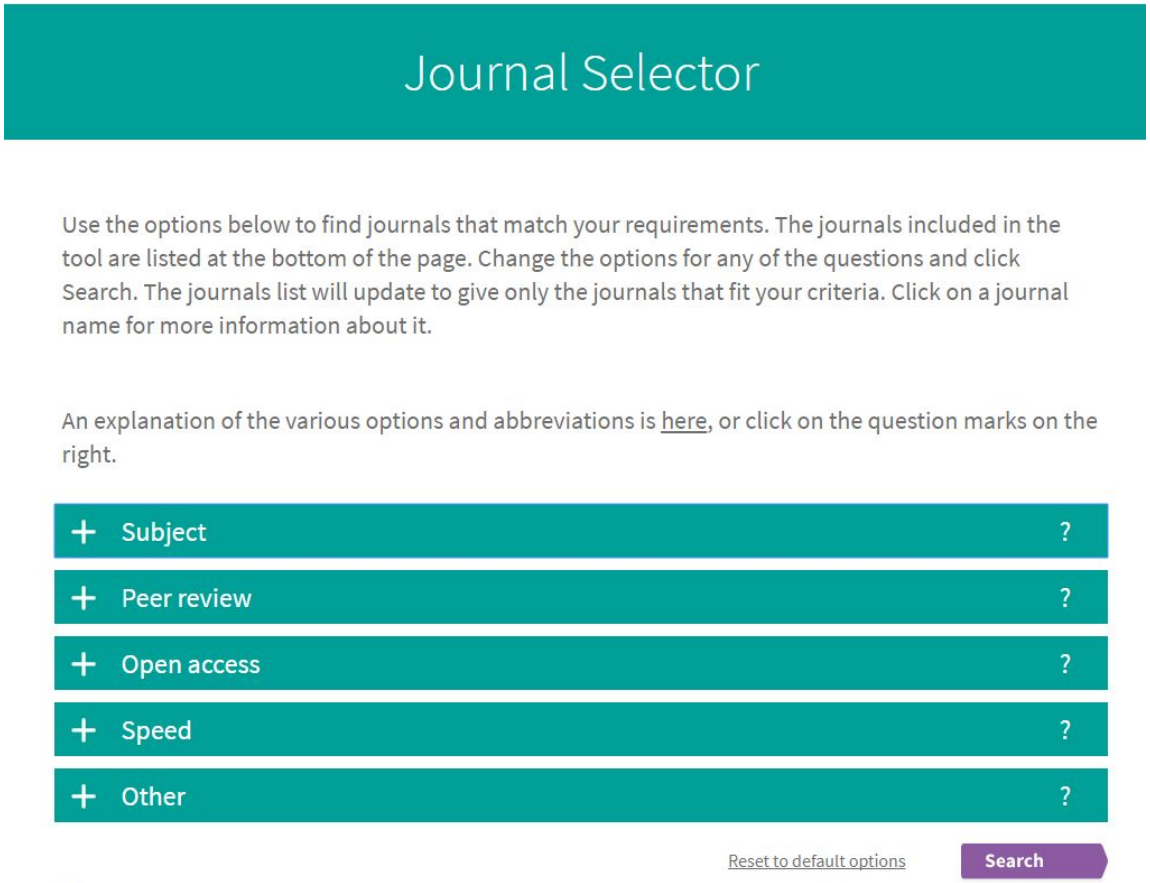
IF 4.25

eLIFE

IF 7.72

Adapted from Erin C. McKiernan, 2015

You can use the Cofactor Journal Selector tool to find open access journals



Cofactor

http://cofactorscience.com/journal-selector

The green route: self-archive and pre-prints

PeerJ Preprints

zenodo

figshare

arXiv.org

bioRxiv beta

THE PREPRINT SERVER FOR BIOLOGY

ASAPbio

Innovations in peer review

JUNE 26, 2017 PROTOHEDGEHOG OPEN SCIENCE

Peer review is broken, but we hold the technological and social solutions to fix it. Right now.

We just submitted a monster paper on the history and present diversity of peer review practices to F1000 Research. It's available in advance here, and soon will be open to public commenting from anyone as it undergoes formal peer review. We wrote it in a similar manner to another paper published on Open Access last year.

INSTITUTIONAL REPOSITORIES

The impacts of open access are really numerous

Check for updates

REVIEW

REVISED **The academic, economic and societal impacts of Open Access: an evidence-based review [version 3; referees: 3 approved, 2 approved with reservations]**

Jonathan P. Tennant [1], François Waldner[2], Damien C. Jacques[2], Paola Masuzzo[3,4], Lauren B. Collister[5], Chris. H. J. Hartgerink[6]

Author details

Grant information

This article is included in the The Future of Scholarly Publishing collection.

Abstract

Ongoing debates surrounding Open Access to the scholarly literature are multifaceted and complicated by disparate and often polarised viewpoints from engaged stakeholders. At the current stage, Open Access has become such a global issue that it is critical for all involved in scholarly publishing, including policymakers, publishers, research funders, governments, learned societies,

METRICS

13163 VIEWS

1826 DOWNLOADS

Get PDF

Get XML

Cite

Export

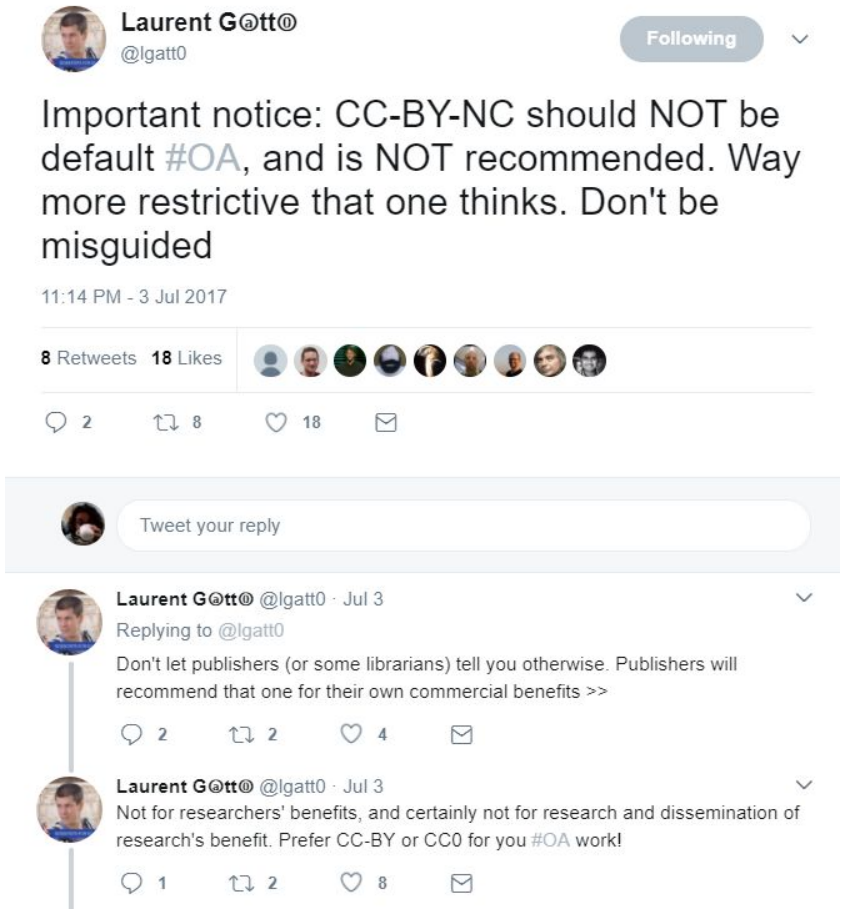
Track

Email

Share

https://f1000research.com/articles/5-632/v3

Important: CC-BY-NC should *not* be the default open access license



- many people identify "non-commercial" with "non-profit"
- significant differences exist and the license chosen may impose some unexpected re-use limitations on works
- NC restricts the value of text and data mining outputs
- commercial use is awesome!
- it means that someone has found a way to add enough value to your work that others will pay for it
- unless you plan on selling your paper, there's zero opportunity cost

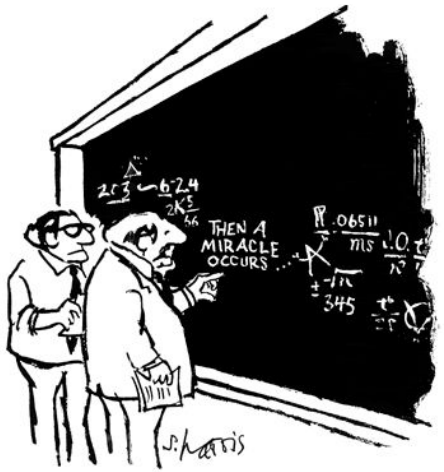
Laurent Gatto, @lgatt0

OPEN CODE

A published article is only the tip of the iceberg of the research process

"An article about computational result is advertising, not scholarship. The actual scholarship is the **full software environment, code and data**, that produced the result."

Buckheit and Donoho (1995)

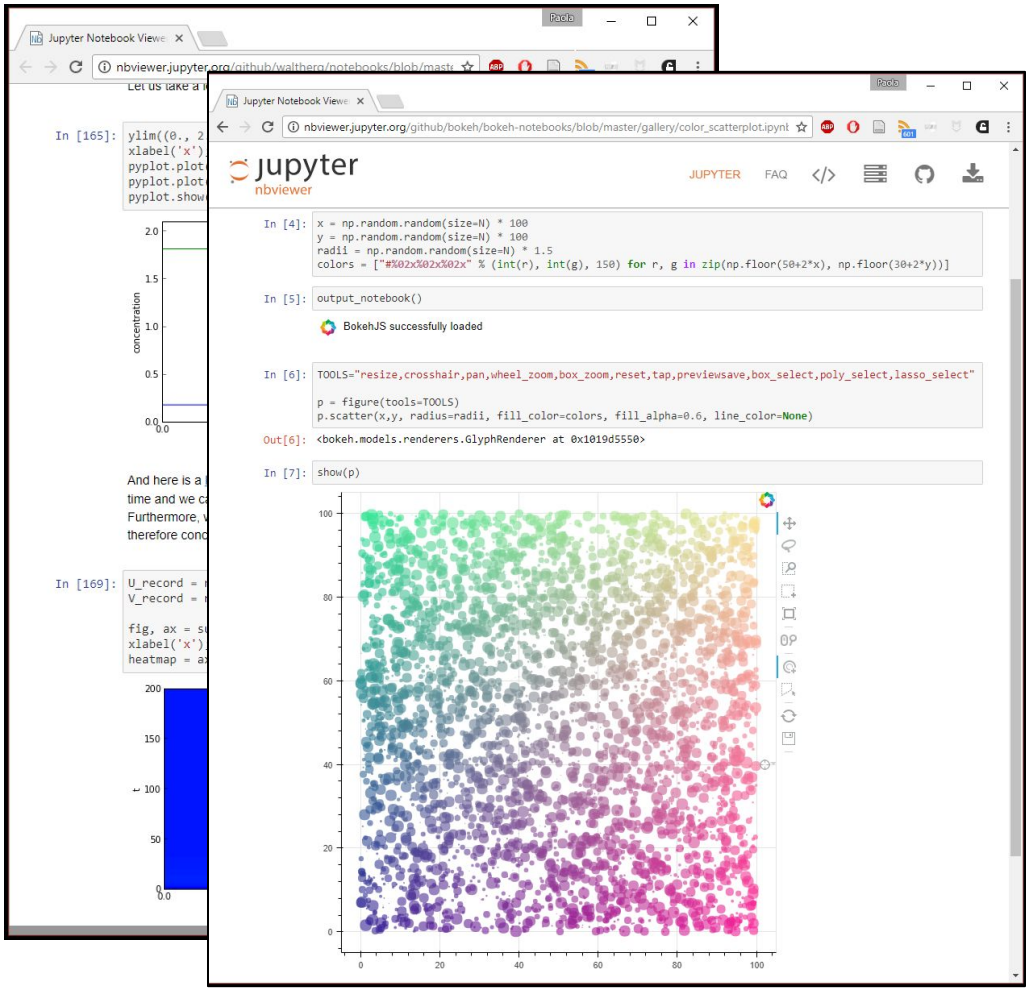


"I think you should be more explicit here in step two."

NumFOCUS supports and promotes world-class, innovative, open source scientific software

https://www.numfocus.org/

Literate programming: sharing the code through interactive notebooks

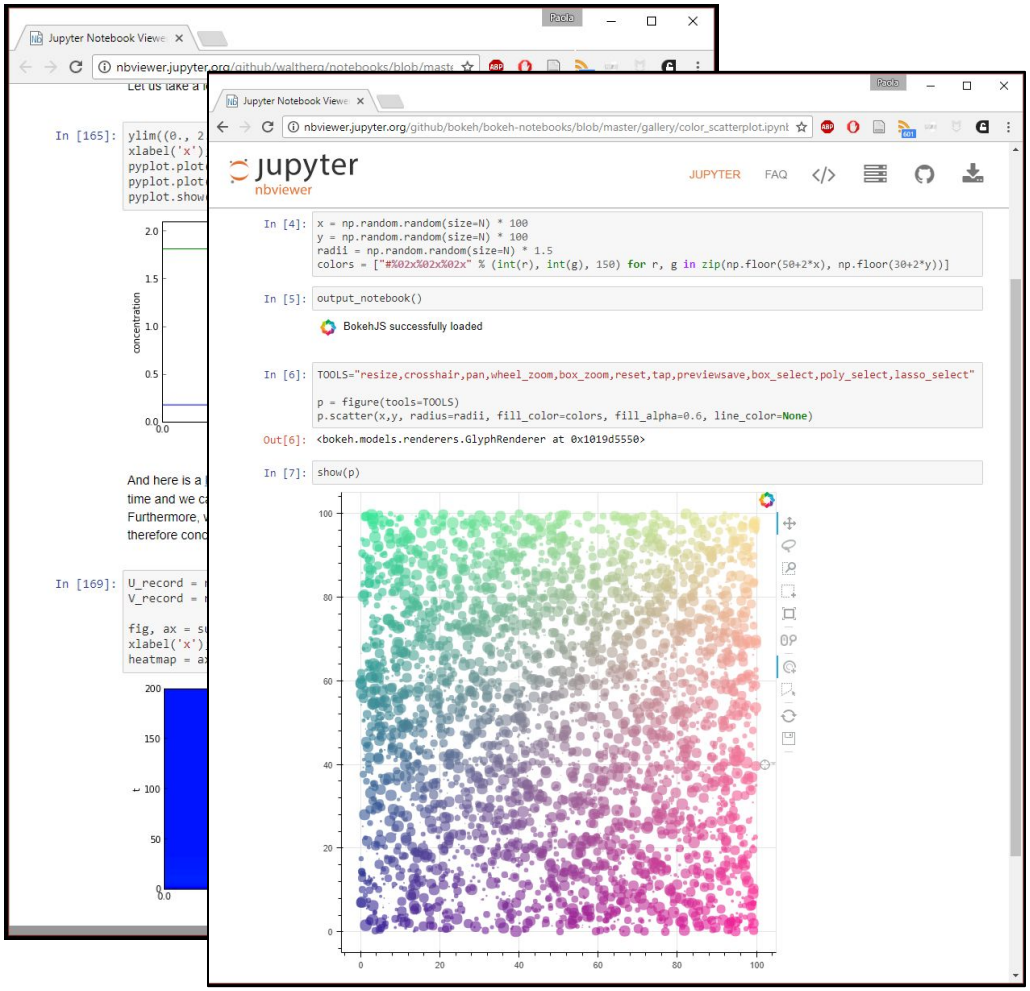


a browser-based and interactive notebook with support for code, rich text, mathematical expressions, inline plots and other rich media

an ideal platform to support **open** and **reproducible** research

https://jupyter.org/

Literate programming: sharing the code through interactive notebooks



a browser-based and interactive notebook with support for code, rich text, mathematical expressions, inline plots and other rich media

an ideal platform to support **open** and **reproducible** research

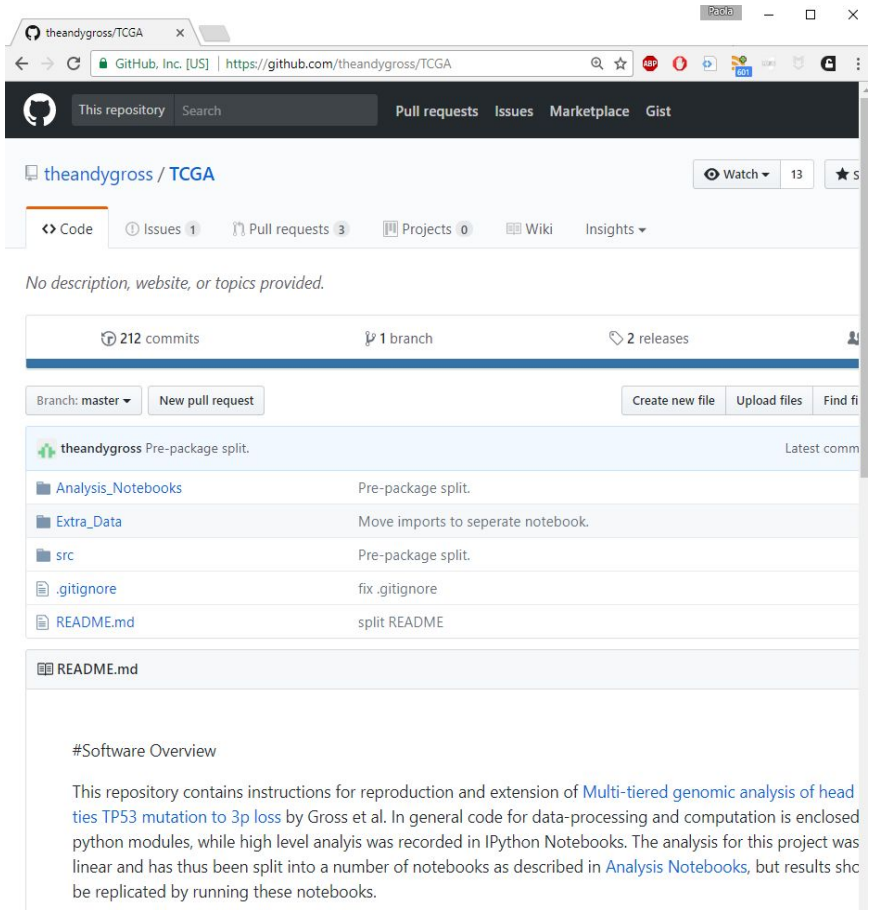
will Jupyter notebook change our publication model???

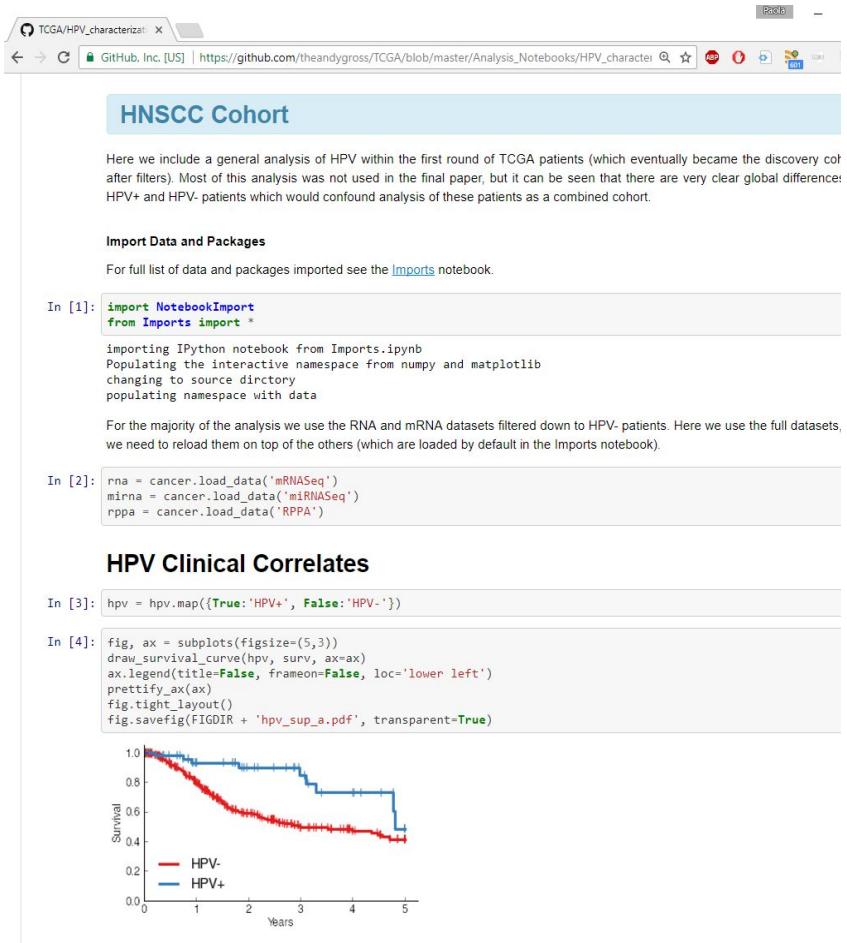
https://jupyter.org/

The paper does the ‘advertisement’

http://www.nature.com/ng/journal/v46/n9/full/ng.3051.html

The notebooks on GitHub are the actual 'scholarship'





https://github.com/theandygross/TCGA

Make sure your project is *actually* open source

Choose an open source license

Which of the following best describes your situation?

I want it simple and permissive.

The **MIT License** is a permissive license that is short and to the point. It lets people do anything they want with your code as long as they provide attribution back to you and don't hold you liable.

jQuery, **.NET Core**, and **Rails** use the MIT License.

I'm concerned about patents.

The **Apache License 2.0** is a permissive license similar to the MIT License, but also provides an express grant of patent rights from contributors to users.

Android, **Apache**, and **Swift** use the Apache License 2.0.

I care about sharing improvements.

The **GNU GPLv3** is a copyleft license that requires anyone who distributes your code or a derivative work to make the source available under the same terms, and also provides an express grant of patent rights from contributors to users.

Bash, **GIMP**, and **Privacy Badger** use the GNU GPLv3.

What if none of these work for me?

My project isn't software.

There are licenses for that.

I want more choices.

More licenses are available.

I don't want to choose a license.

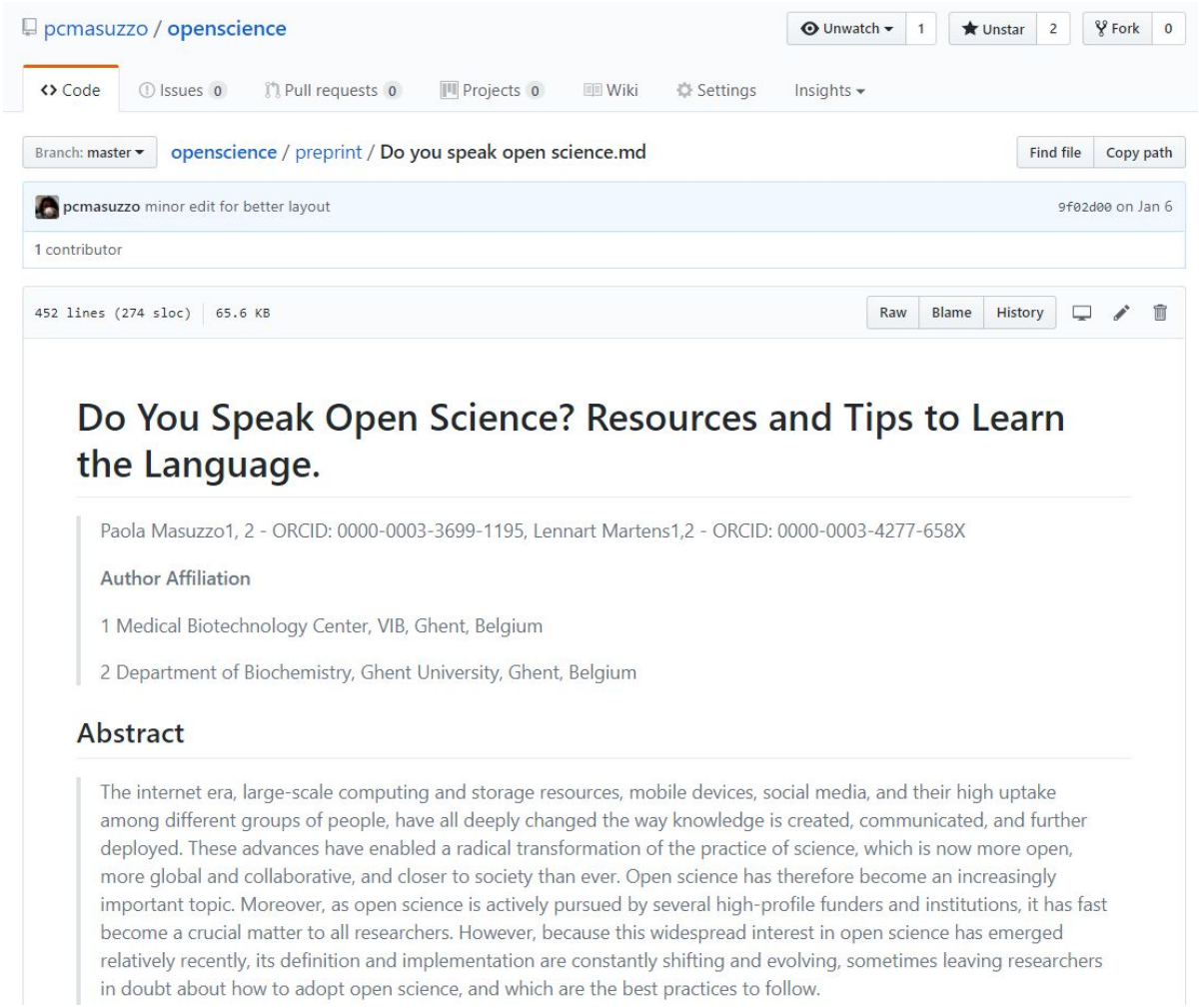
You don't have to.

About Terms of Service

Curated with <3 by GitHub, Inc. and You!

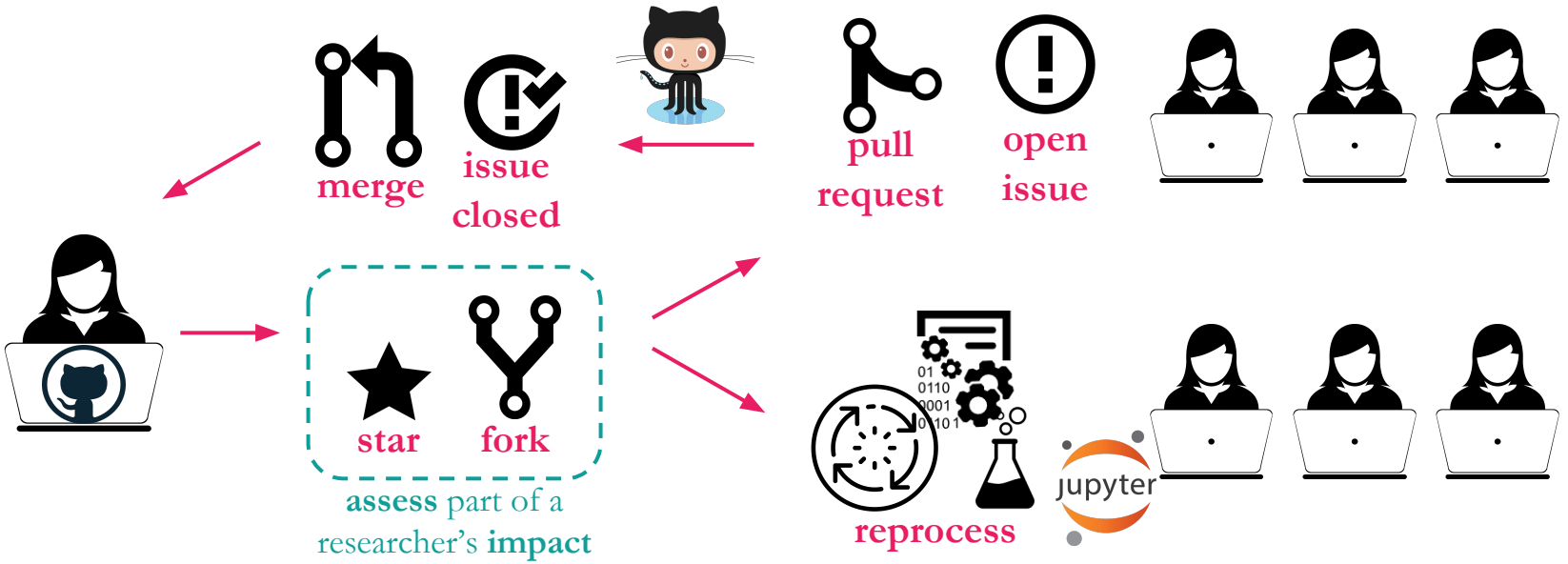
https://choosealicense.com/

GitHub is more than just code



https://github.com/pcmasuzzo/openscience

Open science is by default collaborative and reproducible science



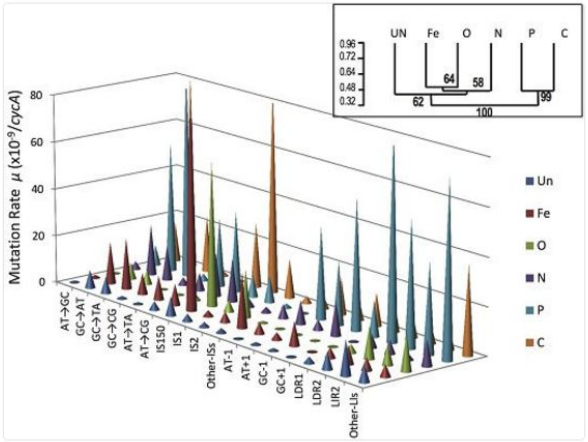
https://github.com/logos; https://octicons.github.com/

An example: improve data visualization with open data and open code

PLOS Biology
@PLOSBiology

Following

5 different environmental stresses cause 5 different patterns of #mutagenesis in #bacteria #PLOSBiology plos.io/2r8ul0u



12:11 PM - 9 Jun 2017

4 Retweets **7** Likes

1 4 7

Tweet your reply

Clement Viguier @ClementViguier · Jun 9
Replying to @PLOSBiology
This could be interesting, but you should not promote it with this kind of badly designed chart. This chart is terrible. #dataviz #wtfviz

1

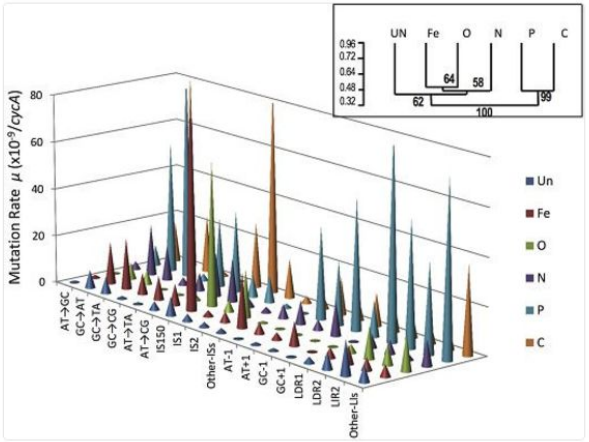
PLOS Biology @PLOSBiology · Jun 9
It's Fig 5 from the paper. Luckily we insist on all underlying data, if you feel like making something prettier ;-) doi.org/10.1371/journa…

1 2

An example: improve data visualization with open data and open code

PLOS Biology
@PLOSBiology
Following

5 different environmental stresses cause 5 different patterns of #mutagenesis in #bacteria #PLOSBiology plos.io/2r8ul0u



12:11 PM - 9 Jun 2017

4 Retweets 7 Likes

1 4 7

Tweet your reply

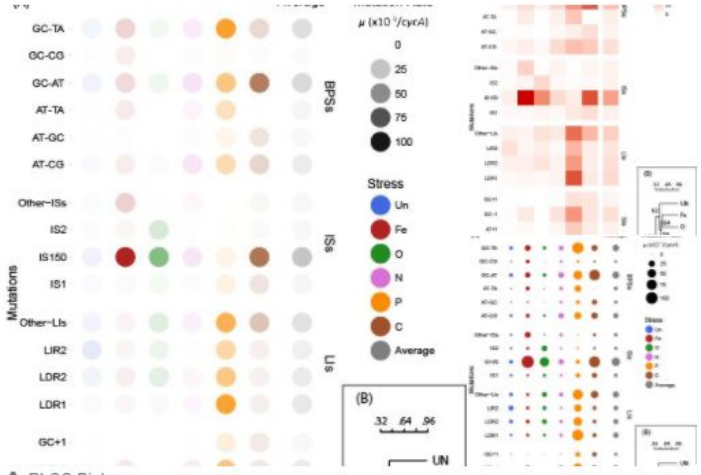
Clement Viguier @ClementViguier · Jun 9
Replying to @PLOSBiology
This could be interesting, but you should not promote it with this kind of badly designed chart. This chart is terrible. #dataviz #wtfviz

1

PLOS Biology @PLOSBiology · Jun 9
It's Fig 5 from the paper. Luckily we insist on all underlying data, if you feel like making something prettier ;-) doi.org/10.1371/journa…

1 2

Clement Viguier @ClementViguier · Jun 21
Better late than never. Here 3 alternative designs. I prefer the heatmap but the alpha plot is more attractive. Could work better alignment.



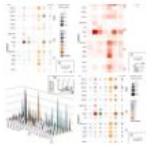
PLOS Biology

1 2

PLOS Biology @PLOSBiology · Jun 21
Fantastic! We'll see if we can figure out a way to link them to the paper...

1 2

Clement Viguier @ClementViguier · Jul 3
...and there is the code to produce them:

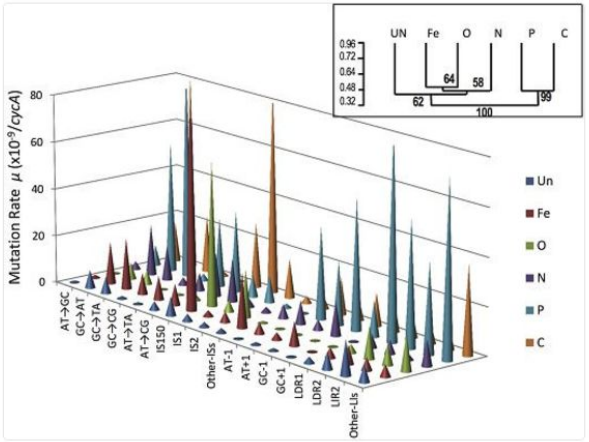


Clement Viguier @ClementViguier
You can read about the story behind my redesigns for @PLOSBiology and code used @ggplot2 on this blog post:medium.com/@clmentviguier… Enjoy!

An example: improve data visualization with open data and open code

PLOS Biology
@PLOSBiology
Following

5 different environmental stresses cause 5 different patterns of #mutagenesis in #bacteria #PLOSBiology plos.io/2r8ul0u



12:11 PM - 9 Jun 2017

4 Retweets 7 Likes

1 4 7

Tweet your reply

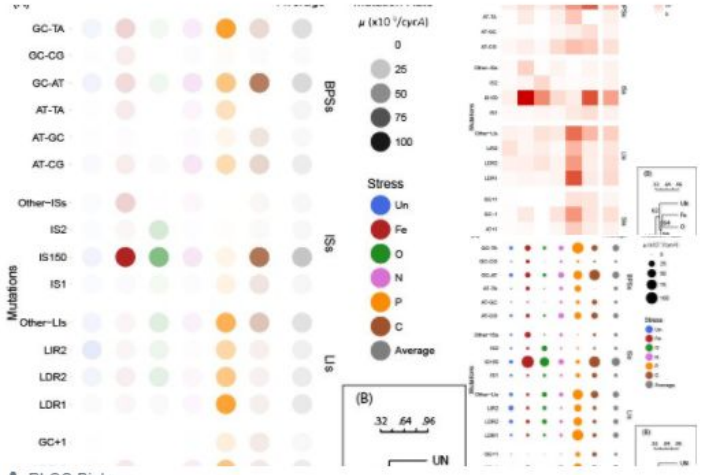
Clement Viguier @ClementViguier · Jun 9
Replying to @PLOSBiology
This could be interesting, but you should not promote it with this kind of badly designed chart. This chart is terrible. #dataviz #wtfviz

1

PLOS Biology @PLOSBiology · Jun 9
It's Fig 5 from the paper. Luckily we insist on all underlying data, if you feel like making something prettier ;-) doi.org/10.1371/journa...

1 2

Clement Viguier @ClementViguier · Jun 21
Better late than never. Here 3 alternative designs. I prefer the heatmap but the alpha plot is more attractive. Could work better alignment.



PLOS Biology

1 2

PLOS Biology @PLOSBiology · Jun 21
Fantastic! We'll see if we can figure out a way to link them to the paper...

1 2

Clement Viguier @ClementViguier · Jul 3

POST-PUBLICATION PEER REVIEW

OPEN PEER REVIEW

Is peer review broken?

Cartoon by John R. McKiernan http://whyopenresearch.org, CC BY-NC-ND

Peer review is a complex process

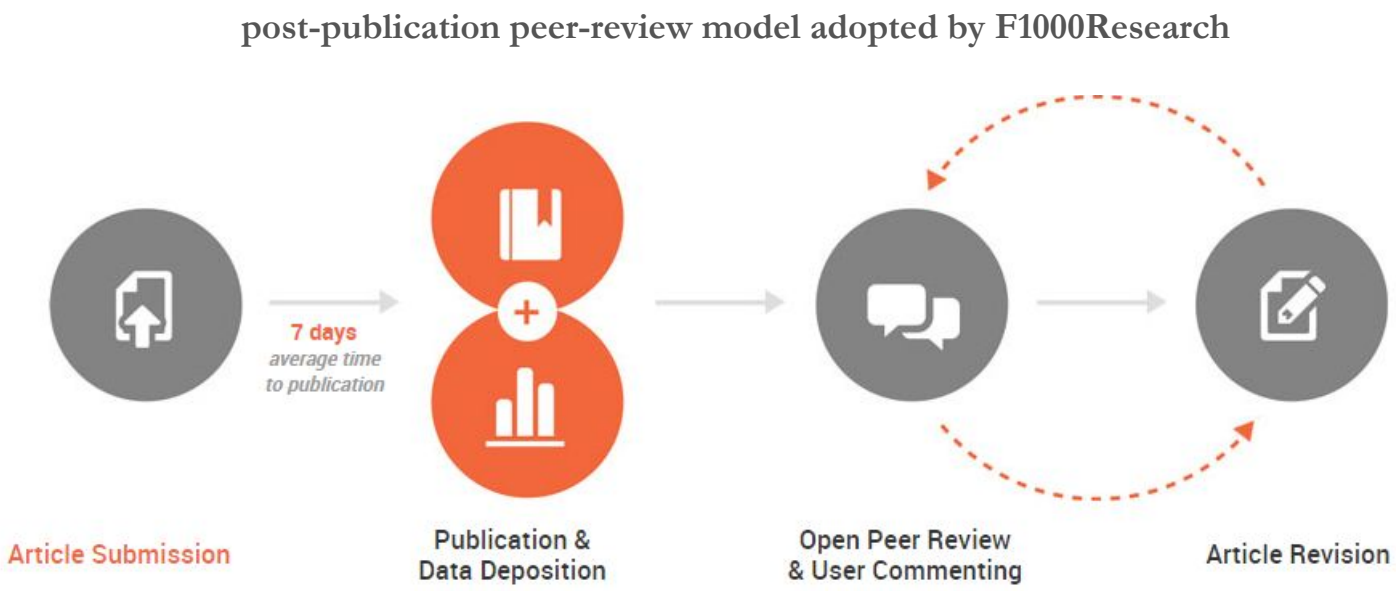
F1000Research

F1000Research 2015 - DRAFT ARTICLE (PRE-SUBMISSION)

A multi-disciplinary perspective on emergent and future innovations in peer review

Jonathan P. Tennant*[1], Jonathan M. Dugan[2], Daniel Graziotin[3], Damien C. Jacques[4], François Waldner[4], Daniel Mietchen[5], Yehia Elkhatib[6], Lauren B. Collister[7], Christina K. Pikas[8], Tom Crick[9], Paola Masuzzo[10], Anthony Caravaggi[11], Devin R. Berg[12], Kyle E. Niemeyer[13], Tony Ross-Hellauer[14], Sara Mannheimer[15], Lillian Rigling[16], Daniel S. Katz[17], Bastian Greshake[18], Josmel Pacheco-Mendoza[19], Nazeefa Fatima[20], Marta Poblet[21], Marios Isaakidis[22], Dasapta Erwin Irawan[23], Sébastien Renaut[24], Christopher R. Madan[25], Lisa Matthias[26], Jesper Nørgaard Kjær[27], Daniel Paul O'Donnell[28], Cameron Neylon[29], Sarah Kearns[30], Manojkumar Selvaraju[31], and Julien Colomb[32]

The main idea behind open peer-review is to enable a fully transparent and author-driven process



https://f1000research.com/about

Researchers can use open reviews to their advantage

REVIEW

REVISED **The academic, economic and societal impacts of Open Access: an evidence-based review [version 3; referees: 3 approved, 2 approved with reservations]**

Jonathan P. Tennant [1], François Waldner[2], Damien C. Jacques[2], Paola Masuzzo[3,4], Lauren B. Collister[5], Chris. H. J. Hartgerink[6]

open feedback

Version 1

Referee Report 28 Apr 2016

Chris Chambers , School of Psychology, Cardiff University, Cardiff, UK

✔ Approved

Views 96 Cite

Tennant *et al* offer a timely and insightful review of the various effects of open access publishing on science and society. The paper is well structured and enjoyable to read. Although I am not an expert on open access publishing, I also found the discussion of the literature quite balanced and evidence-based.

Referee Report 25 Apr 2016

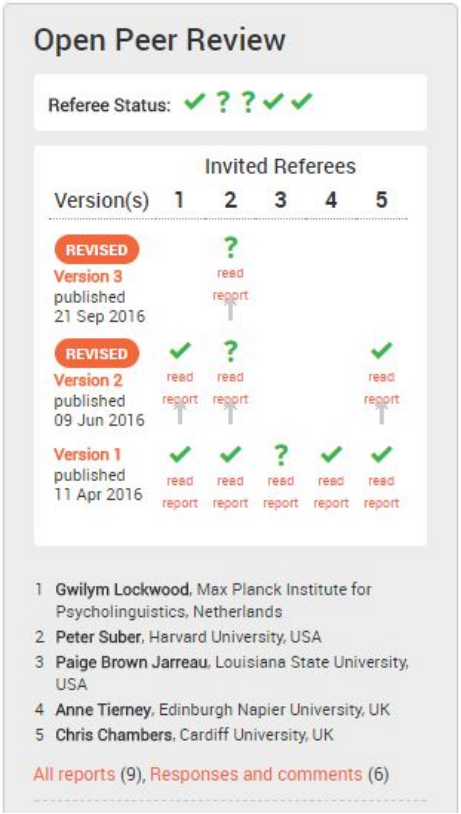
Anne Tierney , Department of Learning and Teaching Enhancement, Edinburgh Napier University, Edinburgh, UK

✔ Approved

Views 69 Cite

This paper is a comprehensive review of the complexities of OA. I have come late to the discussion on this paper, and I find that the previous reviewers have been meticulous in their critique of the paper, to the point I have very little to add. However, there are a couple of points for consideration. What is the effect (if any) of the UK Research Excellence Framework on Open Access? To what extent is disciplinarity a factor in Open Access? I ask this question because of the high impact of the sciences and biomedical research, but wonder about Arts and Humanities (and other areas) as a

transparent submission timeline



https://f1000research.com/articles/5-632/v3

And they can get credit for the reviews they perform

BioMed Central
The Open Access Publisher

ROYAL SOCIETY
OPEN SCIENCE

BMJ

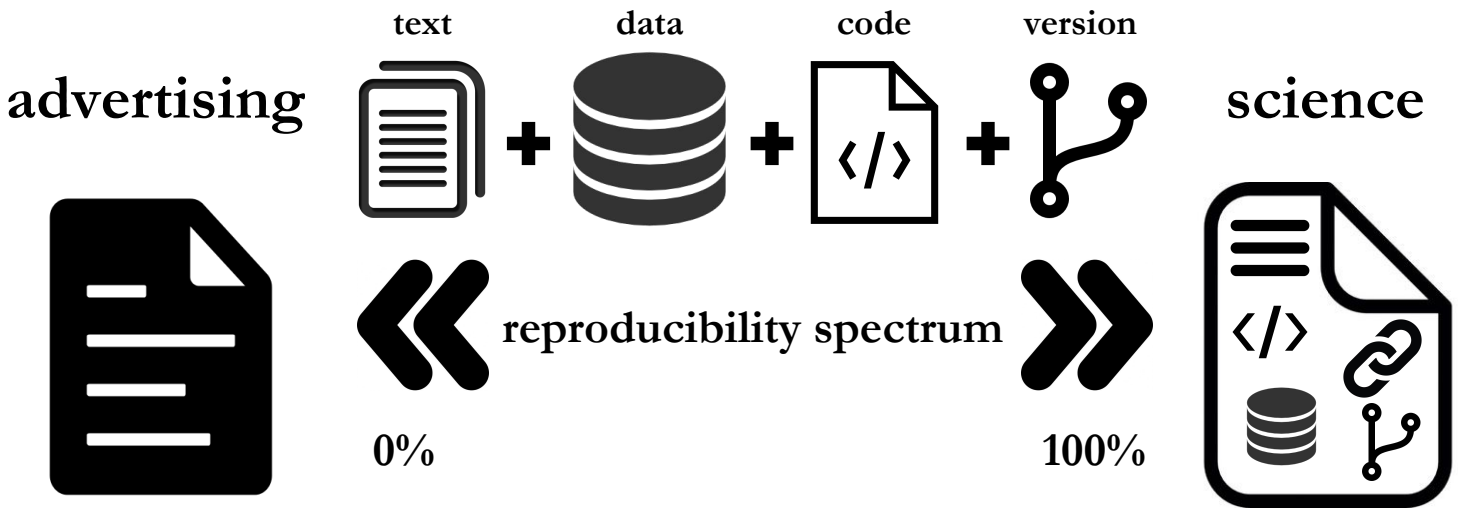
(GIGA)n SCIENCE

F1000Research
Open for Science

Adapted from Erin C. McKiernan, 2015

BRINGING IT ALL TOGETHER

Open science is just science



Adapted from Vaclav (Vashek) Petras, 2017

ALL AMAZING?

Open science: too much talk, too little action

‹ Prev Next ›

Feb 03

Open Science: Too Much Talk, Too Little Action

In: Science Politics • Tags: infrastructure, open science, publishing

Starting this year, I will stop traveling to any speaking engagements on open science (or, more generally, infrastructure reform), as long as these events do not entail a clear goal for action. I have several reasons for this decision, most of them boil down to a cost/benefit estimate. The time spent traveling does not seem worth the hardly noticeable benefits any more.

Blog post Björn Brembs
http://bjoern.brembs.net/2017/02/open-science-too-much-talk-too-little-action/

Open science: too much talk, too little action

‹ Prev Next ›

Feb 03

Open Science: Too Much Talk, Too Little Action

In: Science Politics • Tags: infrastructure, open science, publishing

Starting this year, I will stop traveling to any speaking engagements on open science (or, more generally, infrastructure reform), as long as these events do not entail a clear goal for action. I have several reasons for this decision, most of them boil down to a cost/benefit estimate. The time spent traveling does not seem worth the hardly noticeable benefits any more.

- Publishing is still dominated by the main publishers which keep increasing their **profit margins**
- Most of our work is still behind **paywalls**
- You won't get a job unless you publish in **high-ranking journals**
- Higher ranking journals still publish **less reliable science**
- Libraries are still told by their faculty that **subscriptions** are important
- There are no institutional solutions to **sustainably archive and make accessible** our narratives other than text, or our code or our data

Blog post Björn Brembs
http://bjoern.brembs.net/2017/02/open-science-too-much-talk-too-little-action/

Ideas?
Thoughts?
Wishes?
Actions?

THANKS!

paola masuzzo
@pcmasuzzo