SCIENCE MEETS LIFE

DO YOU SPEAK OPEN SCIENCE?

HERE'S WHY YOU SHOULD!

lennart martens

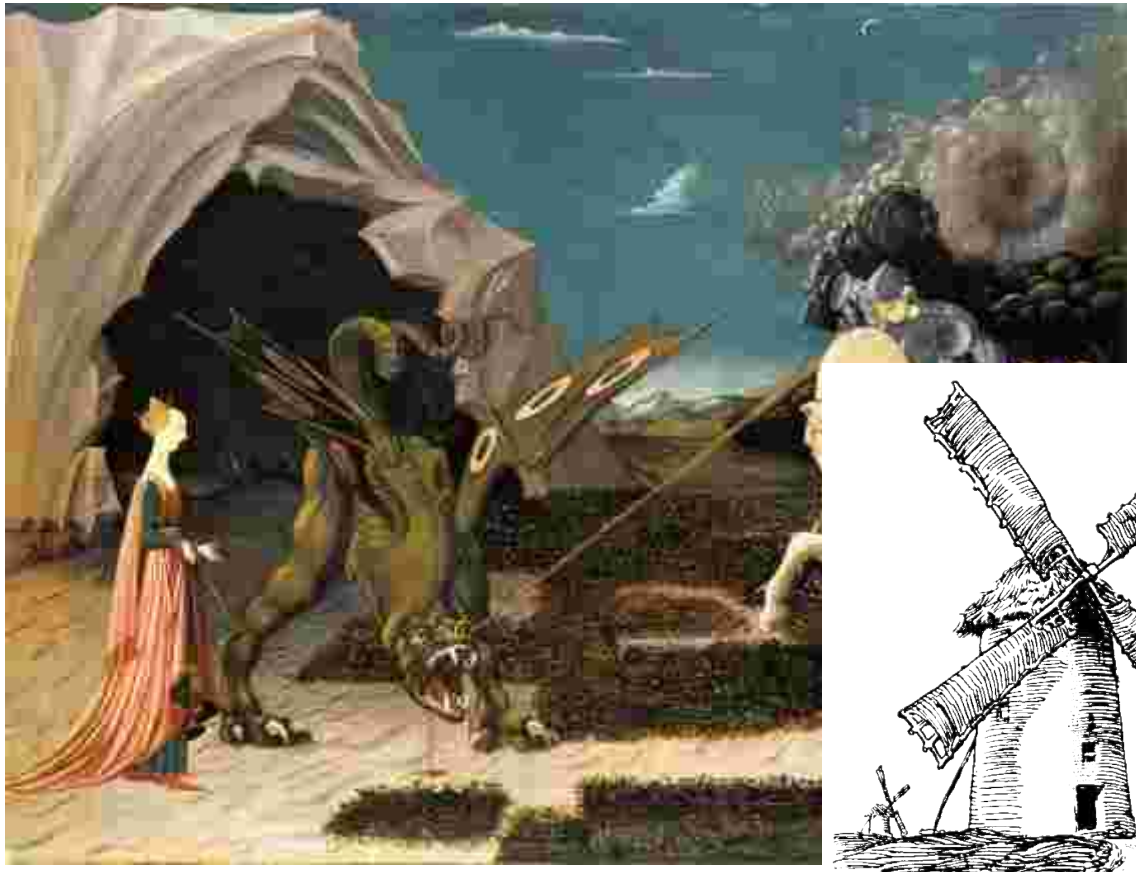
lennart.martens@ugent.be
computational omics and systems biology group
Ghent University and VIB, Ghent, Belgium

FACULTEIT GENEESKUNDE EN GEZONDHEIDSWETENSCHAPPEN

VIB

GHENT UNIVERSITY

Comp omics



Saint George and the Dragon
about 1470, Paolo Uccello

https://www.flickr.com/photos/scriptingnews/3503448168
(CC BY-SA 2.0)

We should ask ourselves a simple question: where does the real revolution lie?

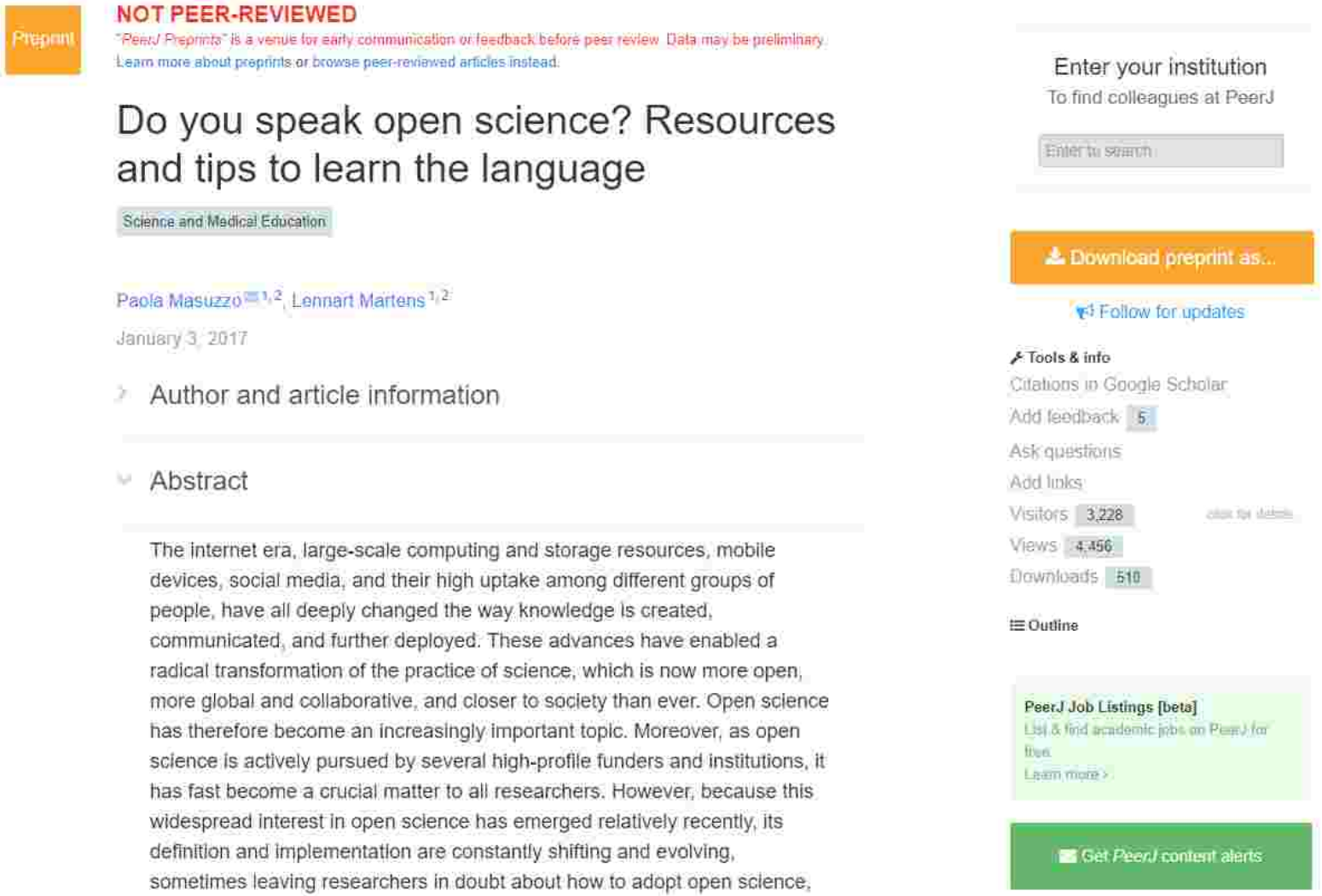
Many people (myself included) have already fought long and hard for open science

But we've really been fighting windmills, even though some of these looked a lot like dragons

The fight may [be | have been] glorious, but in all honesty, the real challenges are:

- to make it all work in practice
- to do amazing things with open science

A field guide to open science for the newly initiated



Preprint

NOT PEER-REVIEWED

"PeerJ Preprints" is a venue for early communication or feedback before peer review. Data may be preliminary.
Learn more about preprints or browse peer-reviewed articles instead.

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.

Enter your institution
To find colleagues at PeerJ

Enter to search

Download preprint as...

Follow for updates

Tools & info
Citations in Google Scholar
Add feedback 5
Ask questions
Add links
Visitors 3,228
Views 4,456
Downloads 510

Outline

PeerJ Job Listings [beta]
List & find academic jobs on PeerJ for free.
Learn more >

Get PeerJ content alerts

Many reasons for open science

Open data

Open access

Open code

Open peer review

A final call to action

Many reasons for open science

Open data

Open access

Open code

Open peer review

A final call to action

Open science is frequently required by an important third party

Where are the data?

***Nature Biotechnology* now requires data availability statements to be supplied with research papers.**

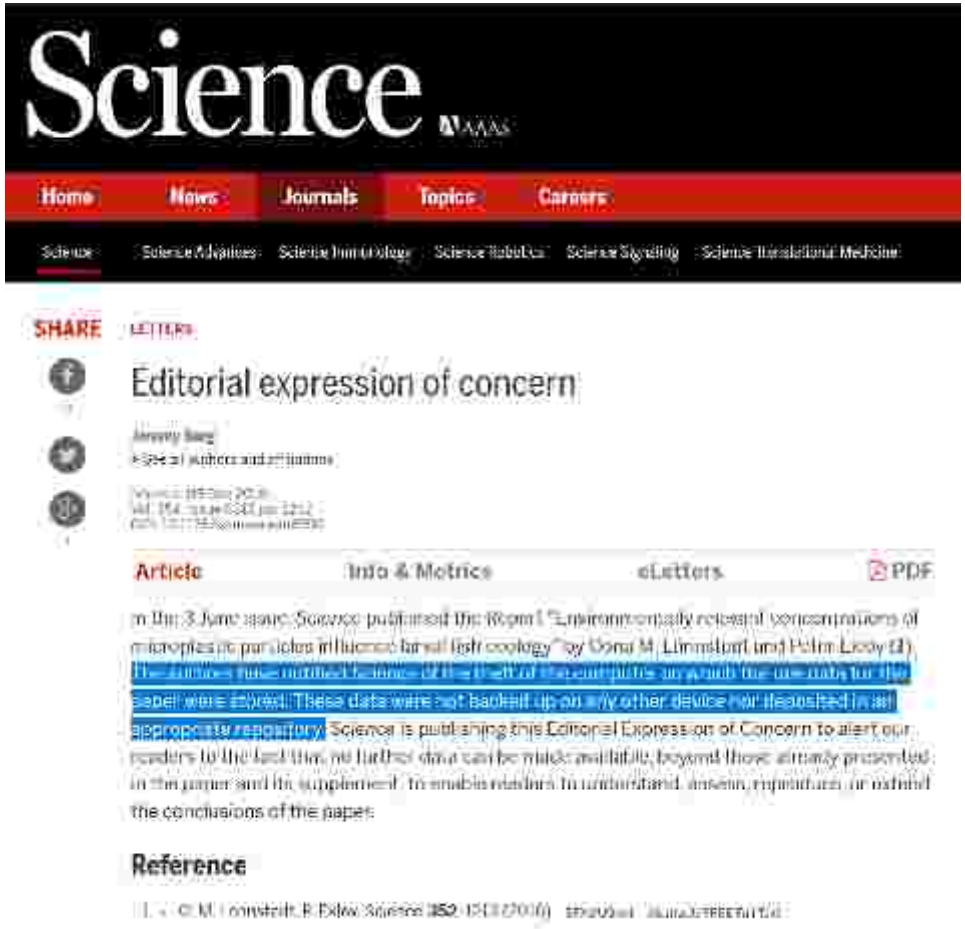
As the research community embraces data sharing, academic journals can do their bit to help. Starting this month, all research papers published in *Nature Biotechnology*, *Nature* and 11 other Nature titles will include information on whether and how others can access the underlying data.

These statements will report the availability of the 'minimal data set' necessary to interpret, replicate and build on the findings reported in the paper. Where applicable, they will include details about publicly archived data sets that have been analyzed or generated during the study. Where restrictions on access are in place — for example, in the case of privacy limitations or third-party control — authors will be expected to make this clear.

The new policy (http://go.nature.com/2bf4vq0) builds on our long-standing support for data availability as a condition of publication. It also extends our support for data citation, the practice of citing data sets in reference lists via digital object identifiers (DOIs).

The absence of openness is increasingly considered a red flag for scientific fraud

After an investigation, the Central Ethical Review Board in Sweden has recommended the retraction of the Report "Environmentally relevant concentrations of microplastic particles influence larval fish ecology," by Oona M. Lönnstedt and Peter Eklöv, published in Science on 3 June 2016 (1). Science ran an Editorial Expression of Concern regarding the Report on 1 December 2016 (2). **The Review Board's report, dated 21 April 2017, cited the following reasons for their recommendation: (i) lack of ethical approval for the experiments; (ii) absence of original data for the experiments reported in the paper; (iii) widespread lack of clarity concerning how the experiments were conducted.** Although the authors have told Science that they disagree with elements of the Board's report, and although Uppsala University has not yet concluded its own investigation, the weight of evidence is that the paper should now be retracted. In light of the Board's recommendation and a 28 April 2017 request from the authors to retract the paper, **Science is retracting the paper in full.**



We usually think we need open science to prevent bad things from happening

While open science helps prevent some cases of fraud or low quality work being published, it is certainly not a panacea (cfr. peer review)

Simultaneously, fraud is regularly detected:

- in the absence of the source data
- from papers published in closed access journals
- without any of the code or metadata available

Why should we define the use of open science through an application with negative connotation?

Instead, we should rather focus on the good that comes from open science

Open science makes the work accessible to anyone

Open science allows people to build much more efficiently on previous work

Open science helps maximize the usefulness of each individual research effort

Data tend to have a (much!) longer shelf life than our (limited) interpretations

Open science fosters creativity, and stimulates revolutionary research

Many reasons for open science

Open data

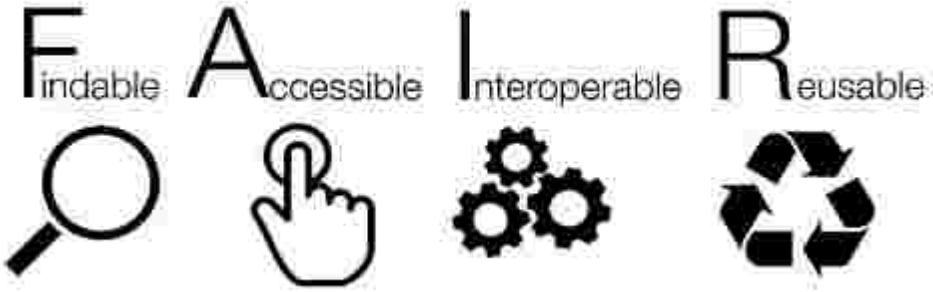
Open access

Open code

Open peer review

A final call to action

FAIR principles are commonly adopted, and compliant systems exist



RCSB PDB PROTEIN DATA BANK

IDR

zenodo

figshare

NEURO VAULT

CXIDB
Coherent X-ray Imaging Data Bank

OpenfMRI

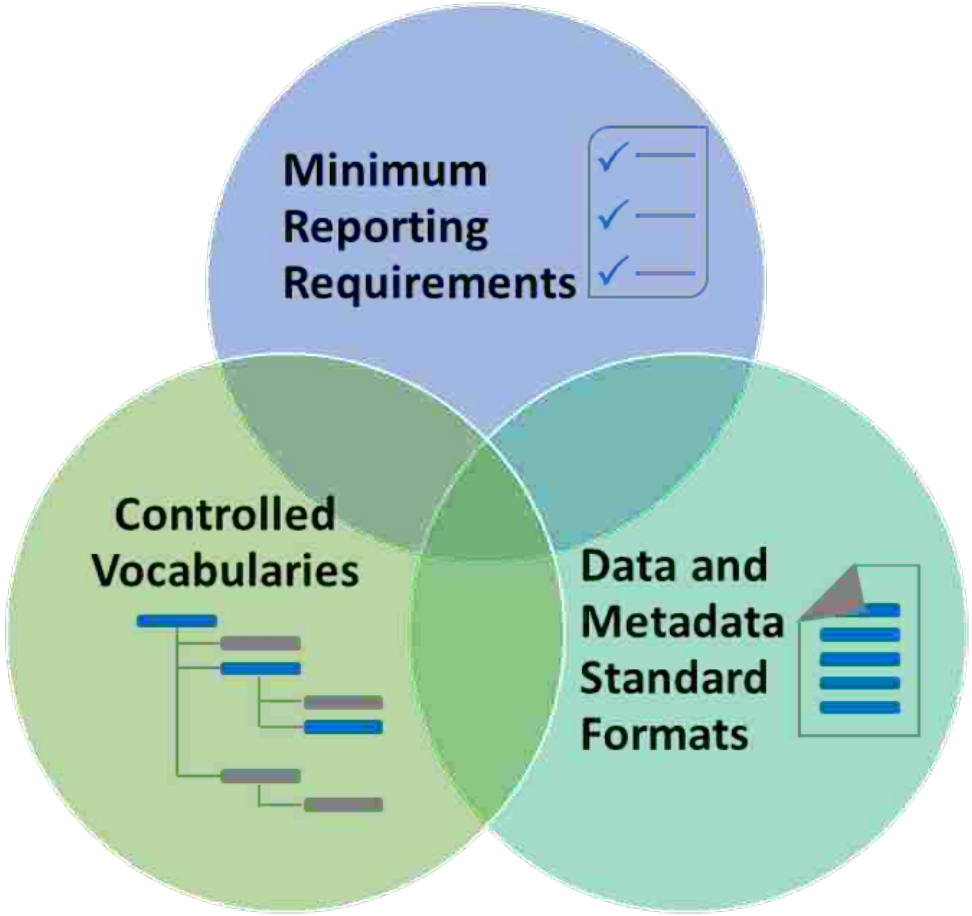
OpenTrials
All the Data, on All the Trials, Linked

DRYAD

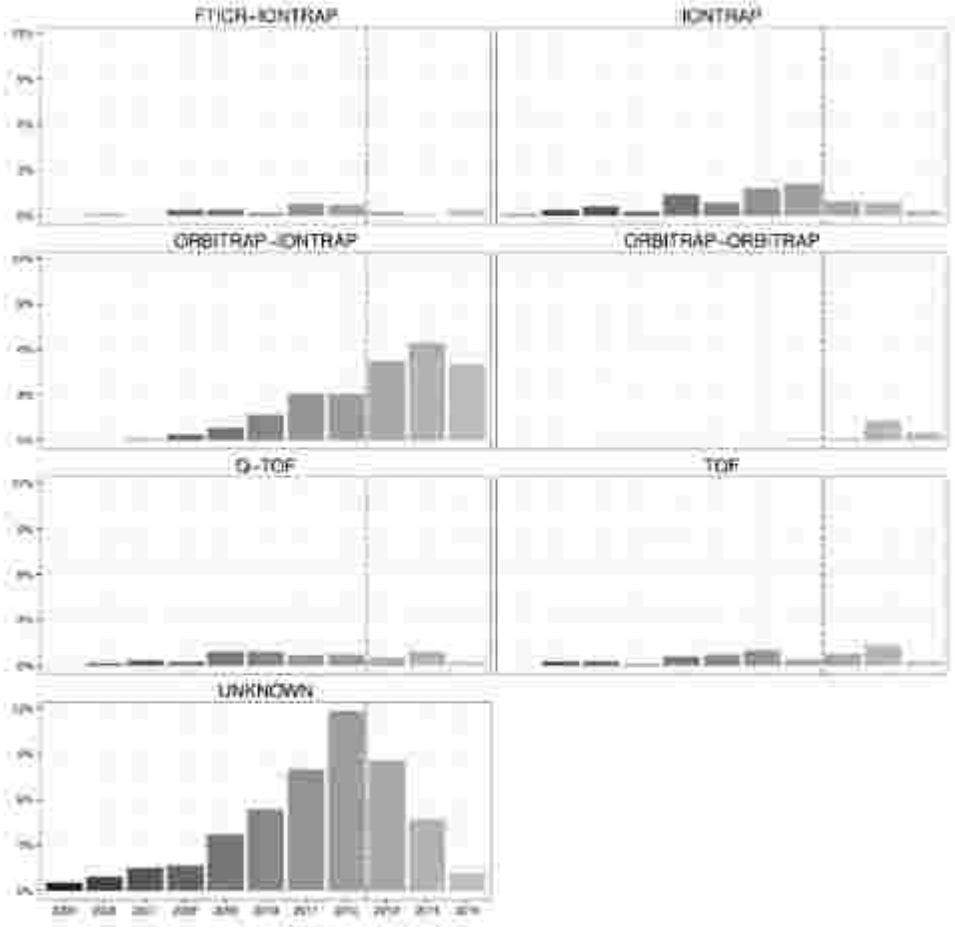
CENTER FOR OPEN SCIENCE

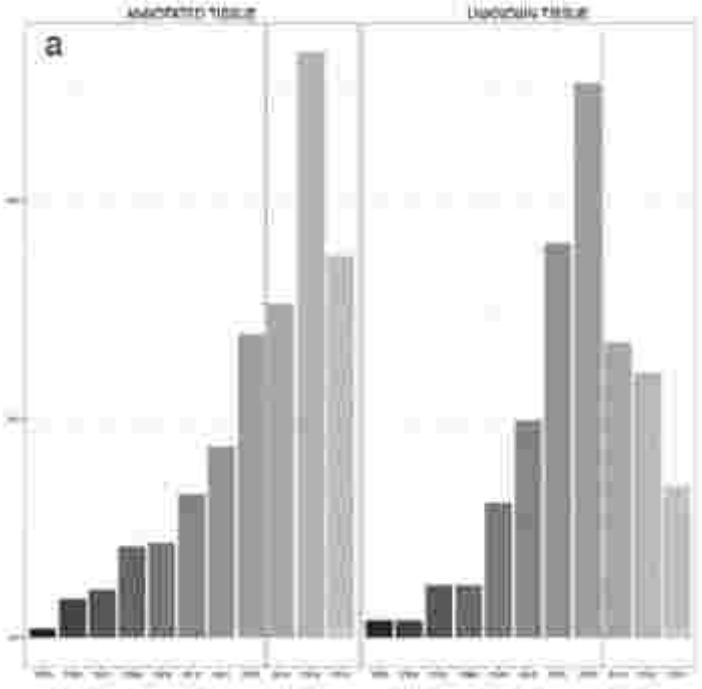
FAIR data principles, WikiMedia Commons, CC BY-SA 4.0

Data sharing requires 3 building blocks: minimal requirements, CVs, and formats



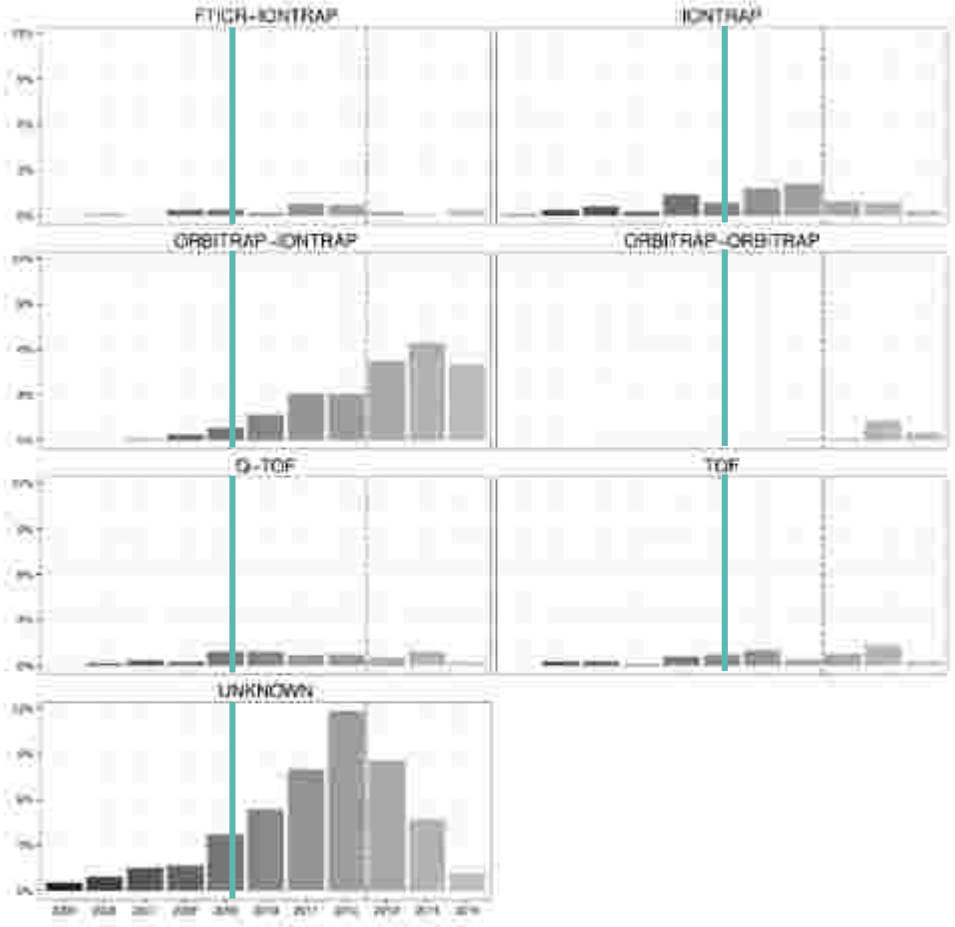
Metadata is often the key issue, as it requires the most manual work

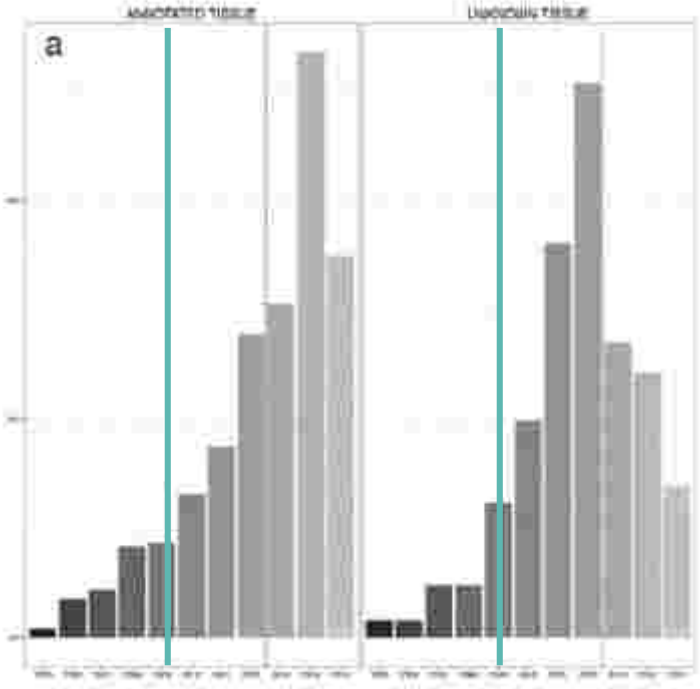




Verheggen, *EuPA Open Proteomics*, 2015

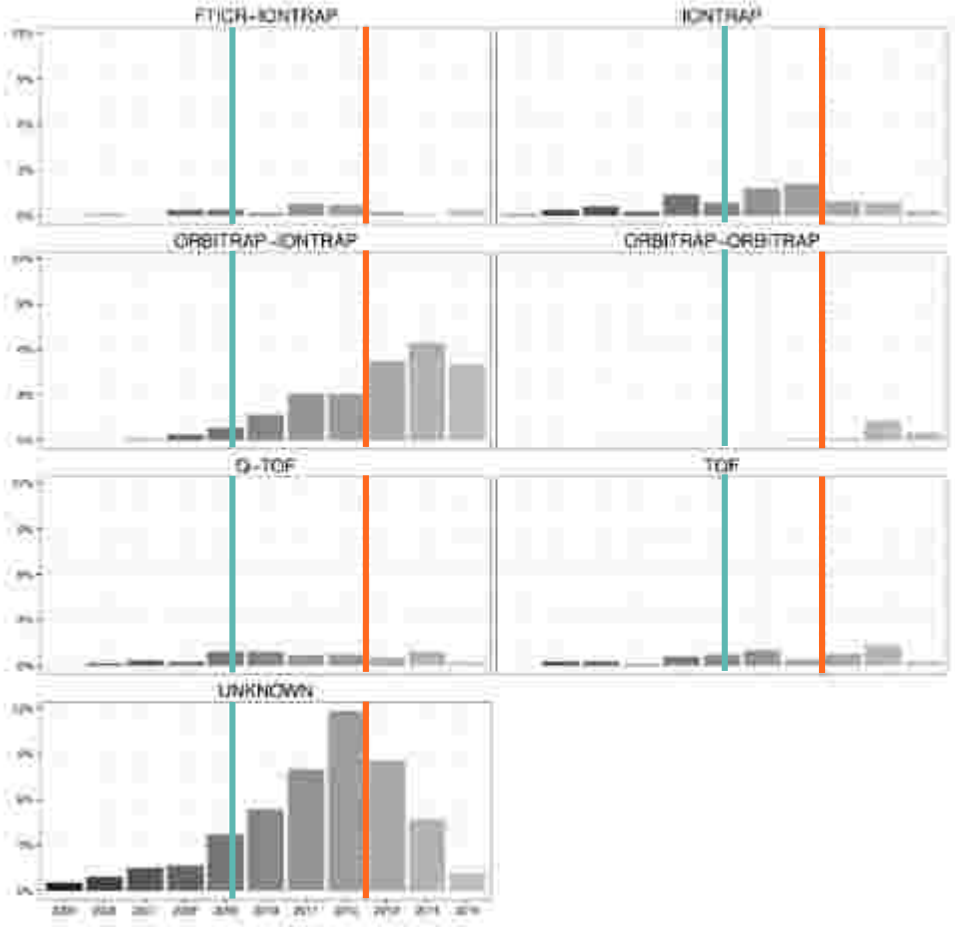
Even user-friendly submission tools cannot correct for a lack of elementary motivation

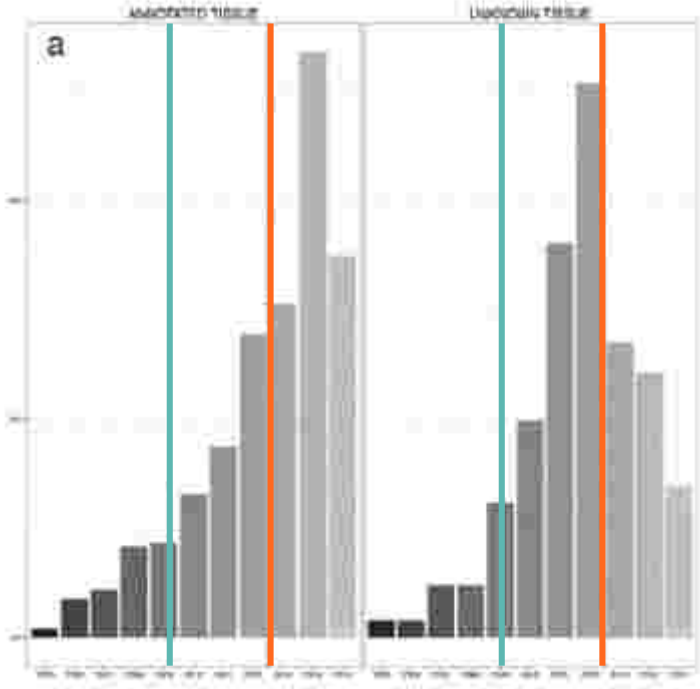




Verheggen, *EuPA Open Proteomics*, 2015

Manual curation of submissions, equivalent to restrictive policing, does help





Verheggen, *EuPA Open Proteomics*, 2015

A non-exhaustive checklist to ensure suitability of your data archival efforts

Are all abbreviations explained?

Are all metadata properly annotated?

Are standard formats used to represent the data in?

Are all terms derived from a controlled vocabulary?

Are missing data explained?

Am I using a trusted, long-lived, third-party repository?

Non-exhaustive checklist for sharing (sensitive) data with the world

Do I have the proper clearances from study participants and/or other data originators?

Double-check clearances for: sharing, reuse, combination

Is the data properly anonymized/de-identified?

These things should really have already been sorted up front, with the ethical committee and other relevant experts!

Does the data have a well-considered license?

Creative Commons (CC) licenses for any content that is not software code

Data without license may NOT be shared at all

Two Creative Commons licenses should be your top choices:

- Attribution (CC BY), only mandates recognition of the author
- Attribution-ShareAlike (CC BY-SA), as CC BY above, but all derived works need to be licensed CC-BY-SA as well (*infectious licence*)

Note that CC licenses are **not** meant for software code; see earlier for examples of suitable open source licenses!

Wikipedia and a lot of Flicr uses CC, amongst many others

https://creativecommons.org ; https://creativecommons.org/licenses

Many reasons for open science

Open data

Open access

Open code

Open peer review

A final call to action

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

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

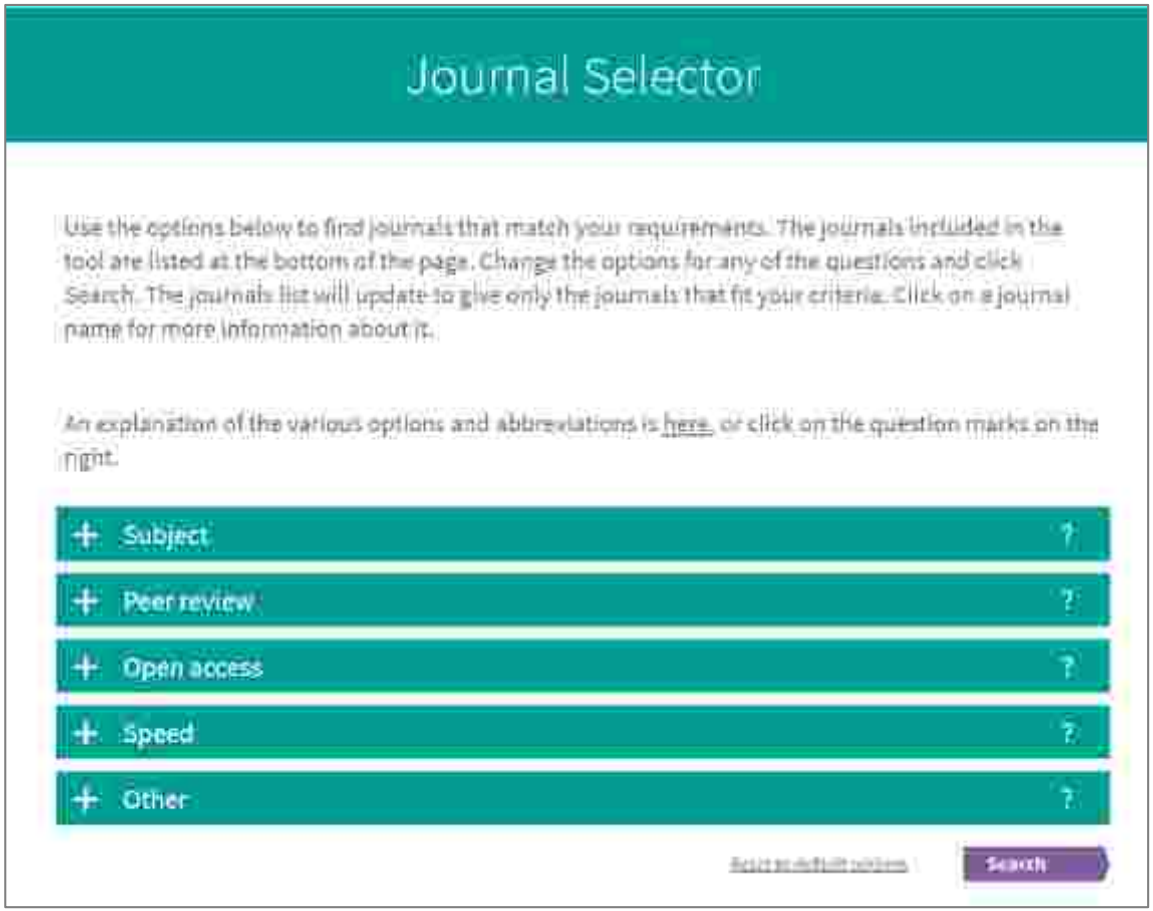
Three types of open access: gold, green, and delayed

Gold open access means you pay, as author, to have your paper open access. *Other papers in the same journal may not be, and so subscriptions remain necessary.* Dual profit?

Green open access means that you deposit a preprint (without the 'added value' of peer review and typesetting) to a public repository such arXiv or BioRxiv.

Delayed open access means that your paper becomes open access after an embargo period (often a year)

The Cofactor Journal Selector tool can help you find open access journals



http://cofactorscience.com/journal-selector

Green open access as a means to fast and free open access

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

Many reasons for open science

Open data

Open access

Open code

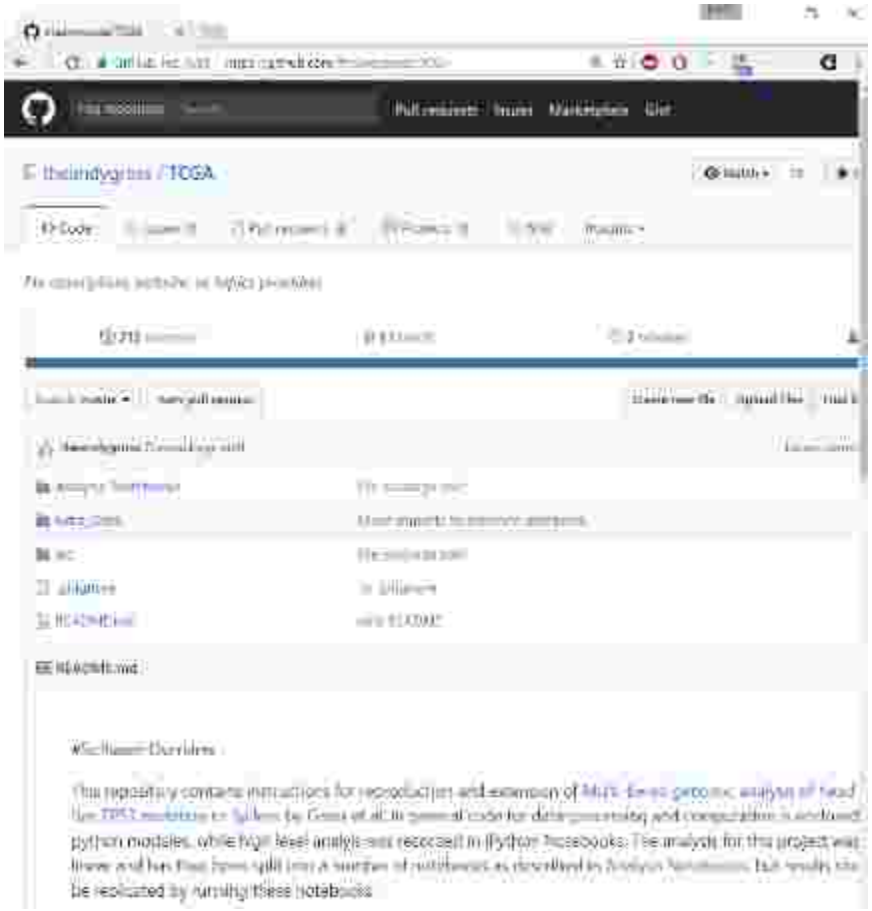
Open peer review

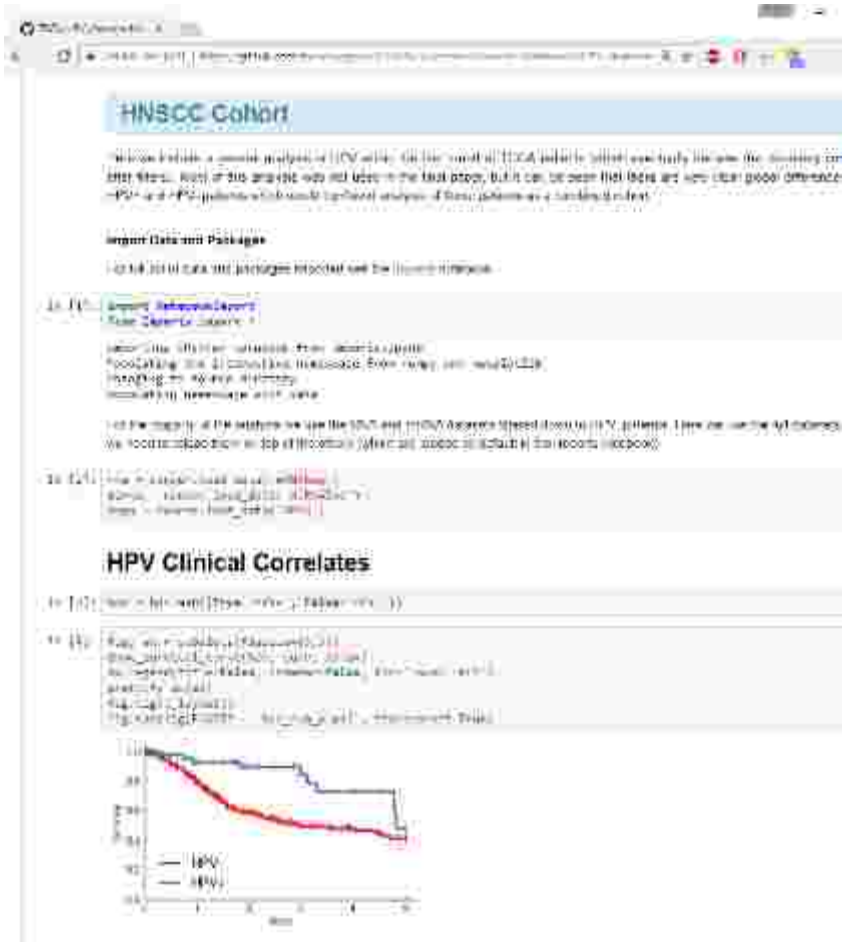
A final call to action

When it comes to the analysis of your data, your paper contains the advertisement…

Gross, *Nature Genetics*, 2014

… but the code on GitHub represents the actual research performed



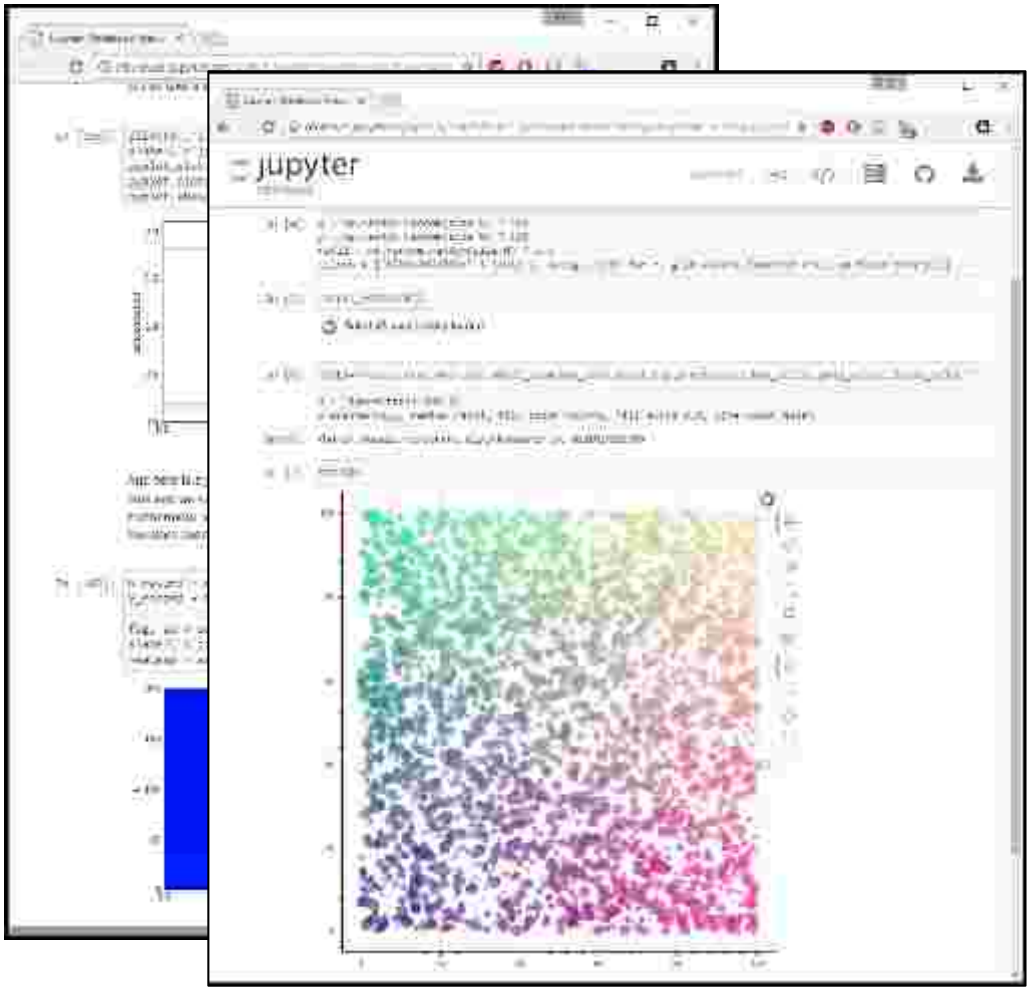


https://github.com/theandygross/TCGA

NumFOCUS supports and promotes innovative, open source scientific software

https://www.numfocus.org

Interactive notebooks enable development, code sharing, and reporting all in one place



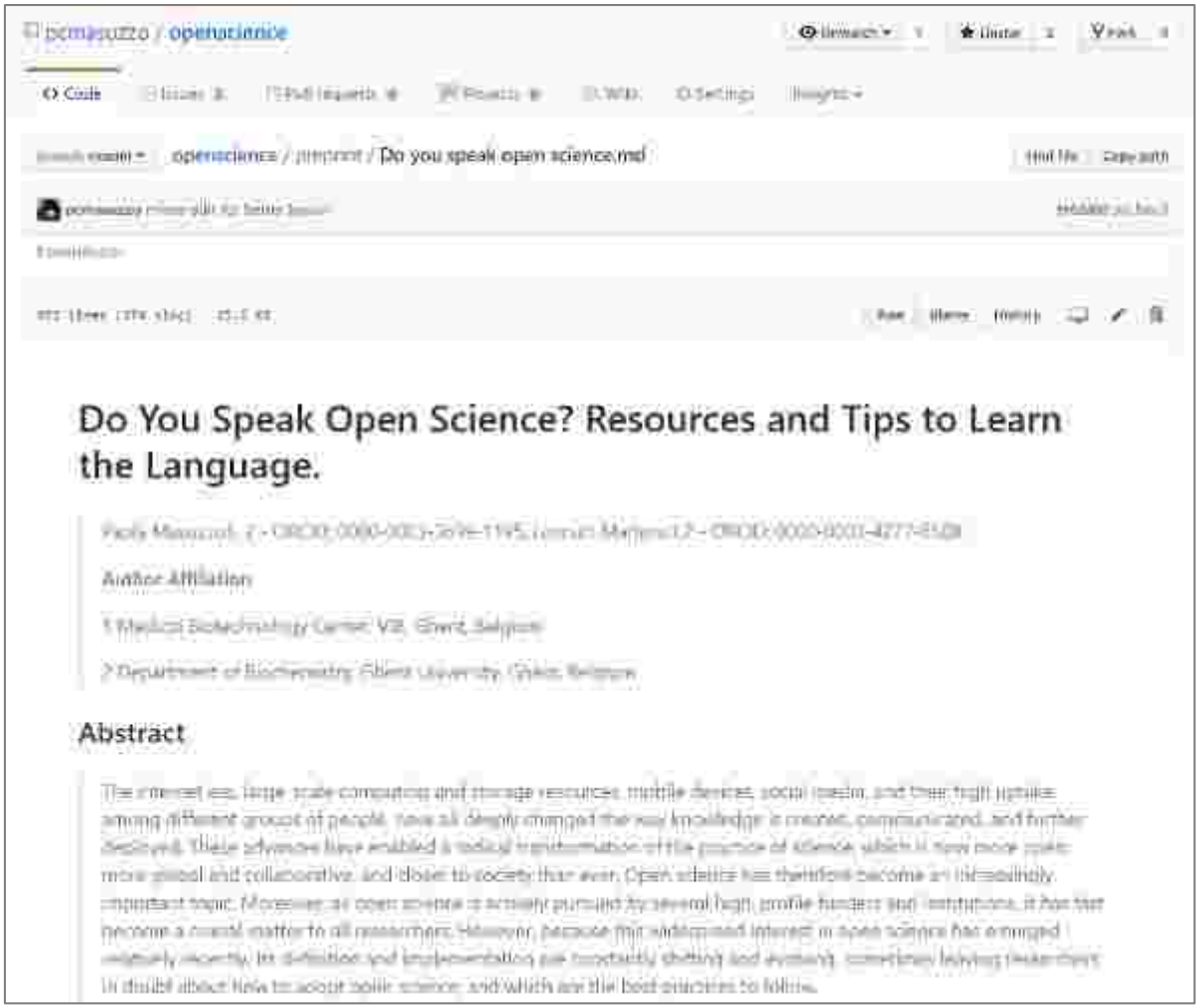
a browser-based and interactive notebook with support for code, rich text, mathematical expressions, inline plots and other rich media

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

Technically, a Jupyter notebook could easily be a publication!

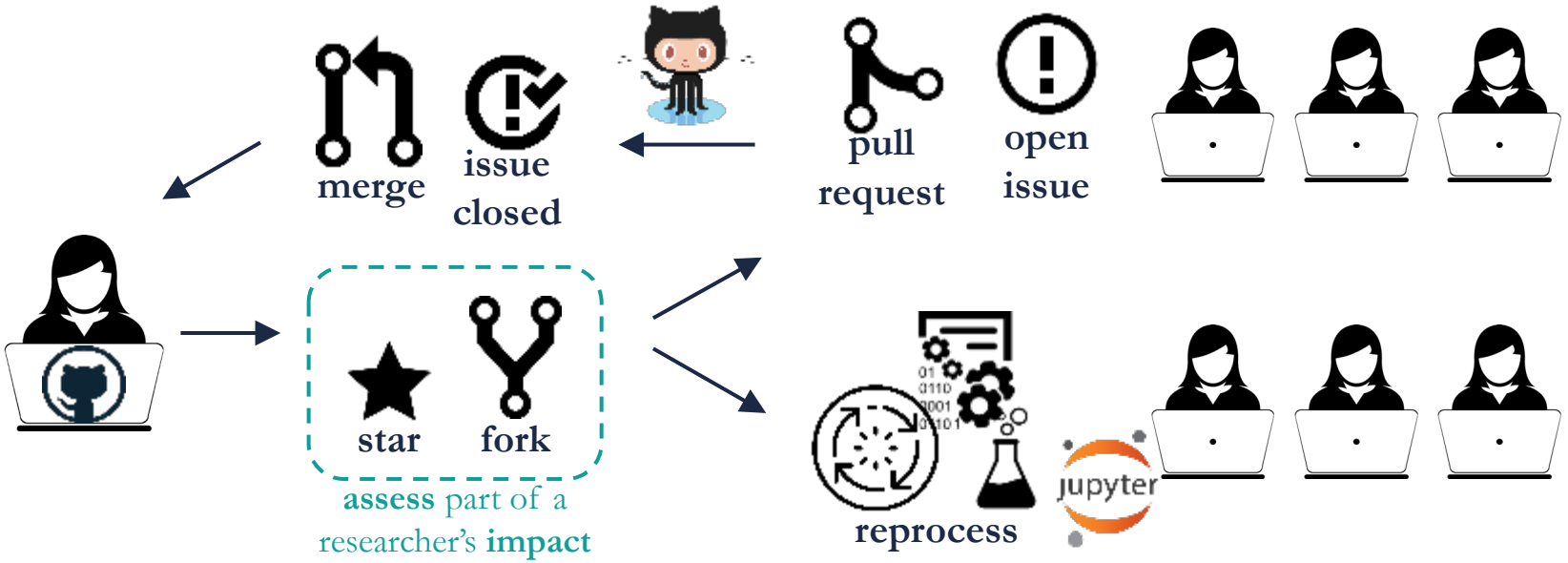
https://jupyter.org

Open code can be shared and managed through freely available third-party systems



https://github.com/pcmasuzzo/openscience

Open code allows collaboration as well as reproduction



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

Code should also always come with an appropriate license, especially open code

https://choosealicense.com

Many reasons for open science

Open data

Open access

Open code

Open peer review

A final call to action

It's fascinating to read about peer review innovations – especially after a reject!

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]

https://f1000research.com/articles/6-1151/v1

Issues and promise of open peer review

Junior researchers (who do most peer reviews) are unlikely to review fully honestly in their own name

Post-publication review is currently predominantly negative

Peer reviewing gets you no real credit today

Even journal editors can't answer a simple question: can I publish (say on a blog) a peer review on one of my manuscripts?

But in the end, a form of open peer review just makes sense; it's closed peer review that is bizarre

Many reasons for open science

Open data

Open access

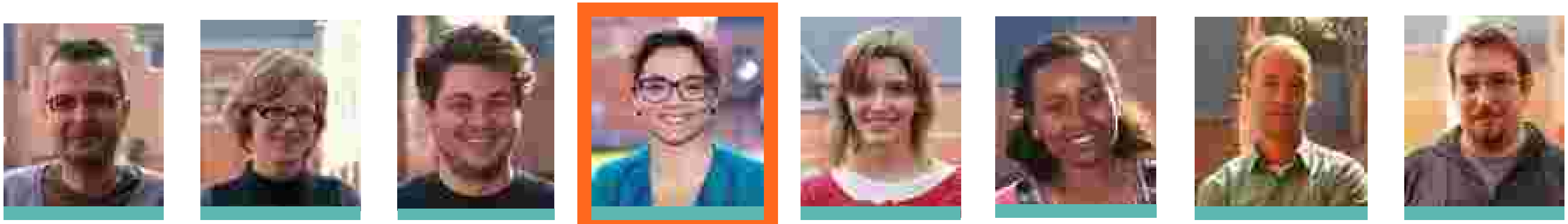
Open code

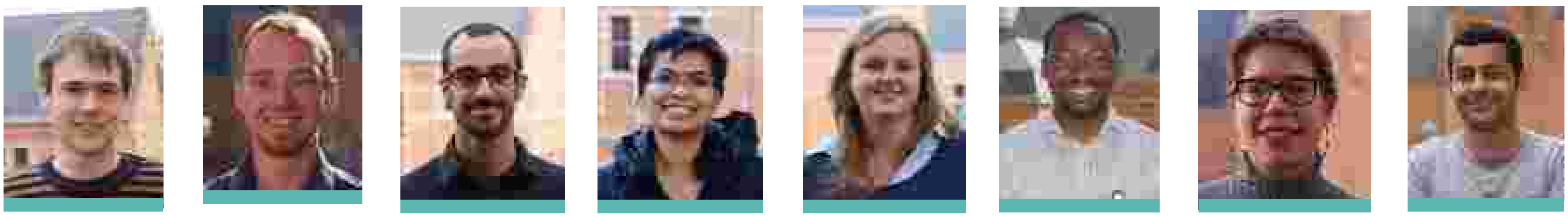
Open peer review

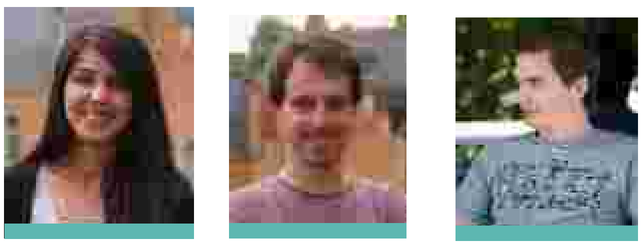
A final call to action

Here is treasure of unlimited size, with all dragons chased away – *what will you do?*

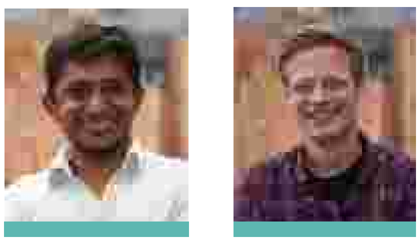
https://www.flickr.com/photos/fantasy-art-and-portraits/2884954207
(CC BY-NC-SA 2.0)







Compomics



VIB

GHENT UNIVERSITY

FACULTEIT GENEESKUNDE EN GEZONDHEIDSWETENSCHAPPEN