

SCIENCE MEETS LIFE

# DO YOU SPEAK OPEN SCIENCE?

*HERE'S WHY YOU SHOULD!*

lennart martens

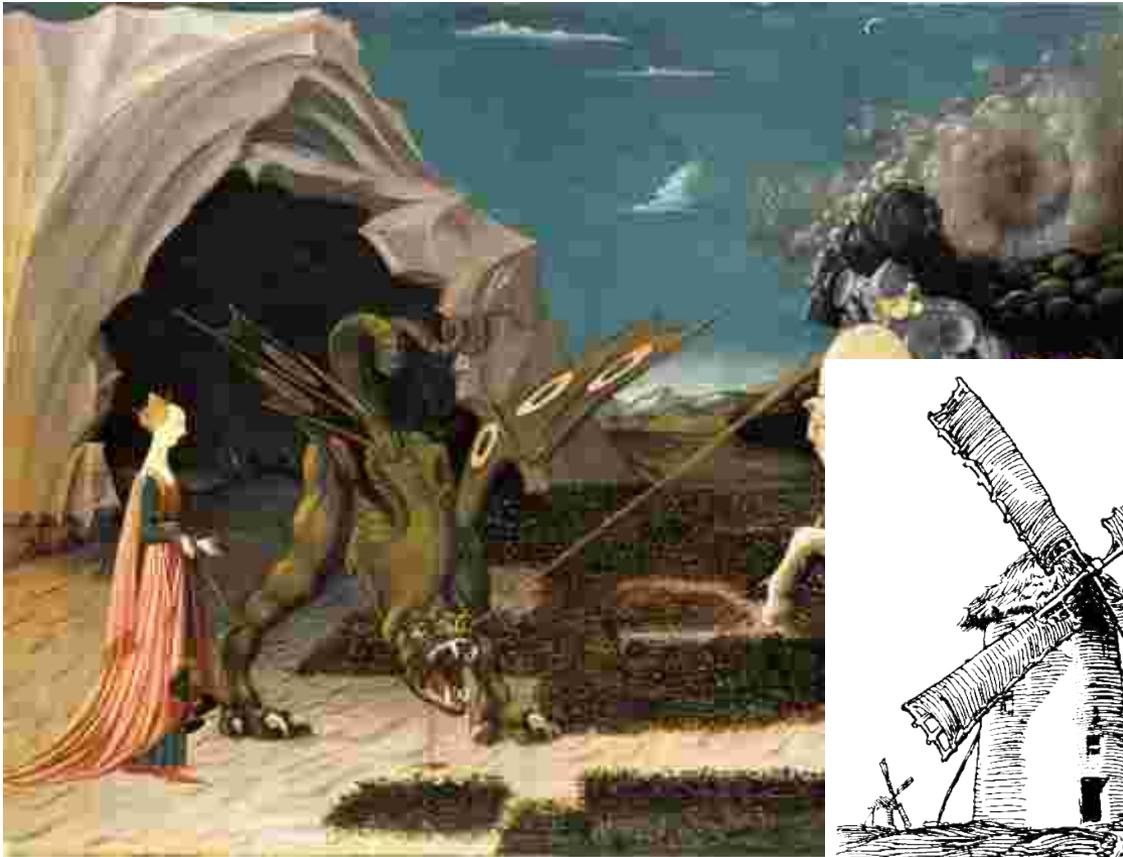
*lennart.martens@ugent.be*

*computational omics and systems biology group*

*Ghent University and VIB, Ghent, Belgium*

 **FACULTEIT GENEESKUNDE EN  
GEZONDHEIDSWETENSCHAPPEN**





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## NOT PEER-REVIEWED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<sup>1,2</sup>, Lennart Martens<sup>1,2</sup>

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 click for details

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/2bf4vq6f>) 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.**



**SHARE** LETTERS

## Editorial expression of concern

Jeremy Berg  
Special Advertising Solutions

View all letters  
All 154 letters (2017-2022)  
Only 107 letters (2017-2022)

Article Info & Metrics eLetters PDF

In the 3 June issue, *Science* published the Report “Environmentally relevant concentrations of microplastic particles influence larval fish ecology” by Oona M. Lönnstedt and Peter Eklöv (1). The authors have notified *Science* of the fact that the original data for the paper were withheld. These data were not backed up on any other device nor deposited in an appropriate repository. *Science* is publishing this Editorial Expression of Concern to alert our readers to the fact that no further data can be made available beyond those already presented in the paper and its supplement. To enable readers to understand, assess, reproduce or extend the conclusions of the paper:

### Reference

1. Lönnstedt O, Eklöv P. Environ Sci Technol 2016; 50(6):2700-6. doi:10.1021/acs.est.5b02700

# 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

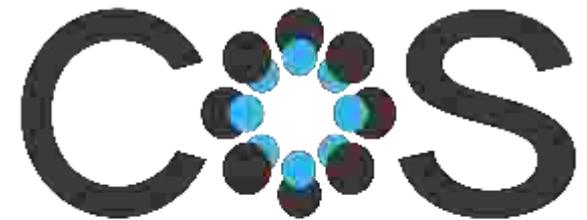
Open code

Open peer review

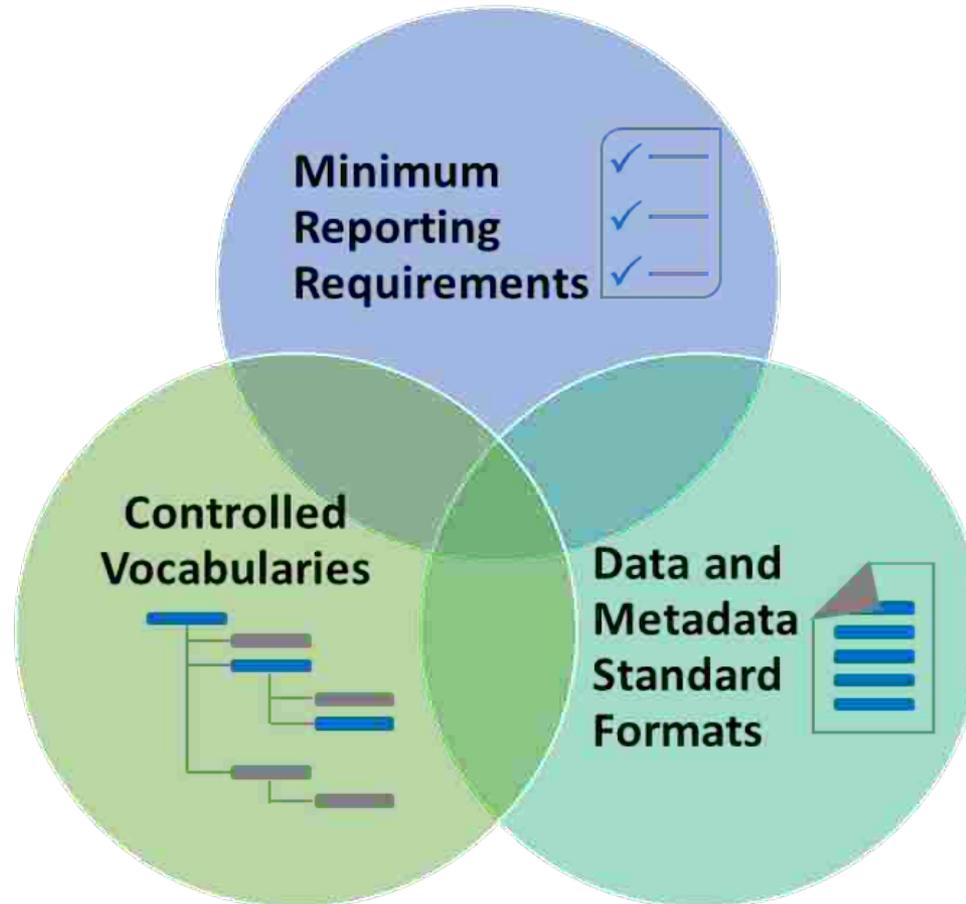
A final call to action

# FAIR principles are commonly adopted, and compliant systems exist

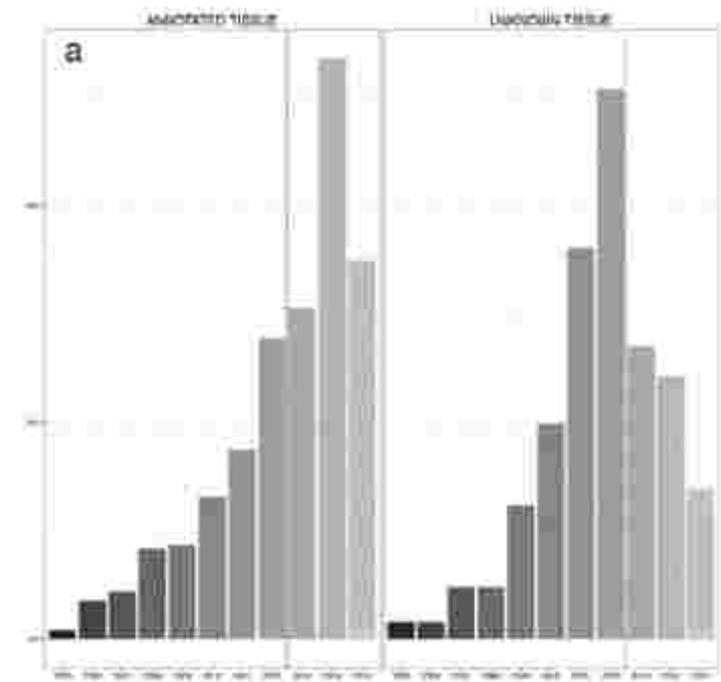
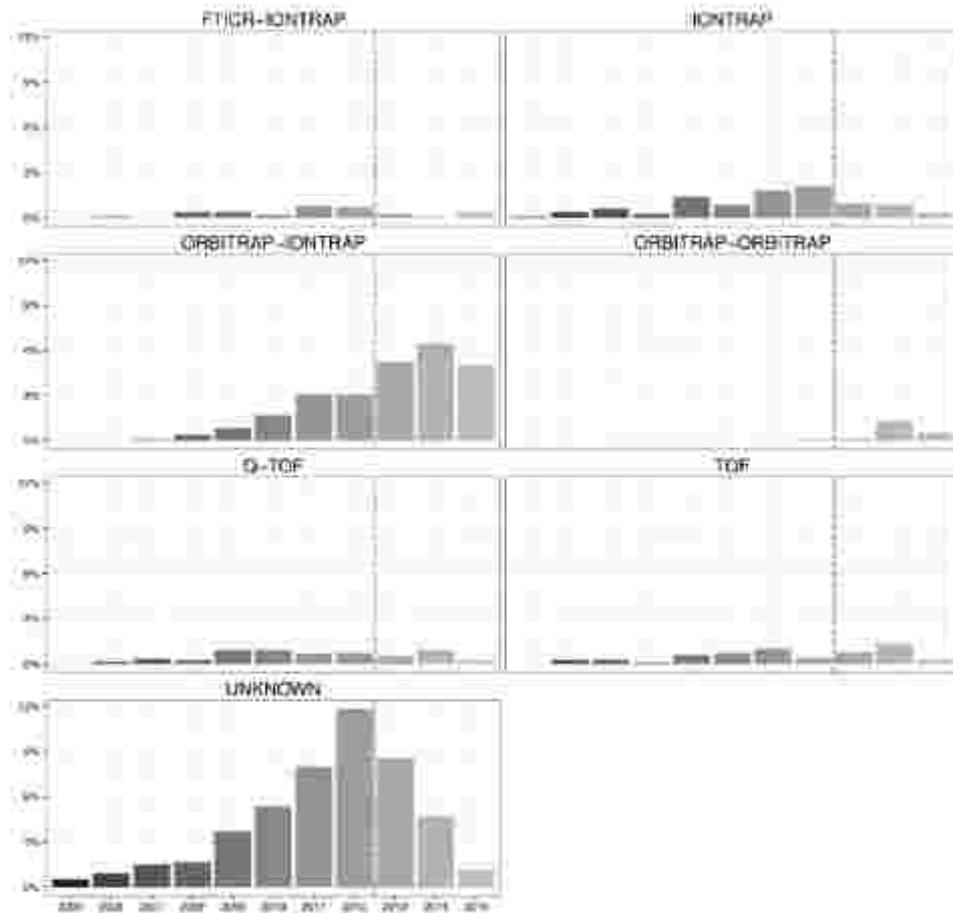
F<sub>indable</sub> A<sub>ccessible</sub> I<sub>nteroperable</sub> R<sub>eusable</sub>



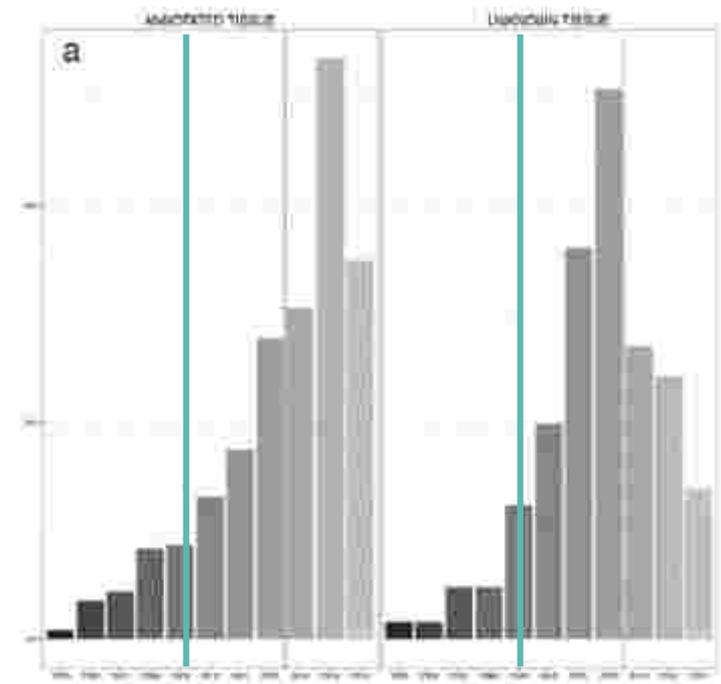
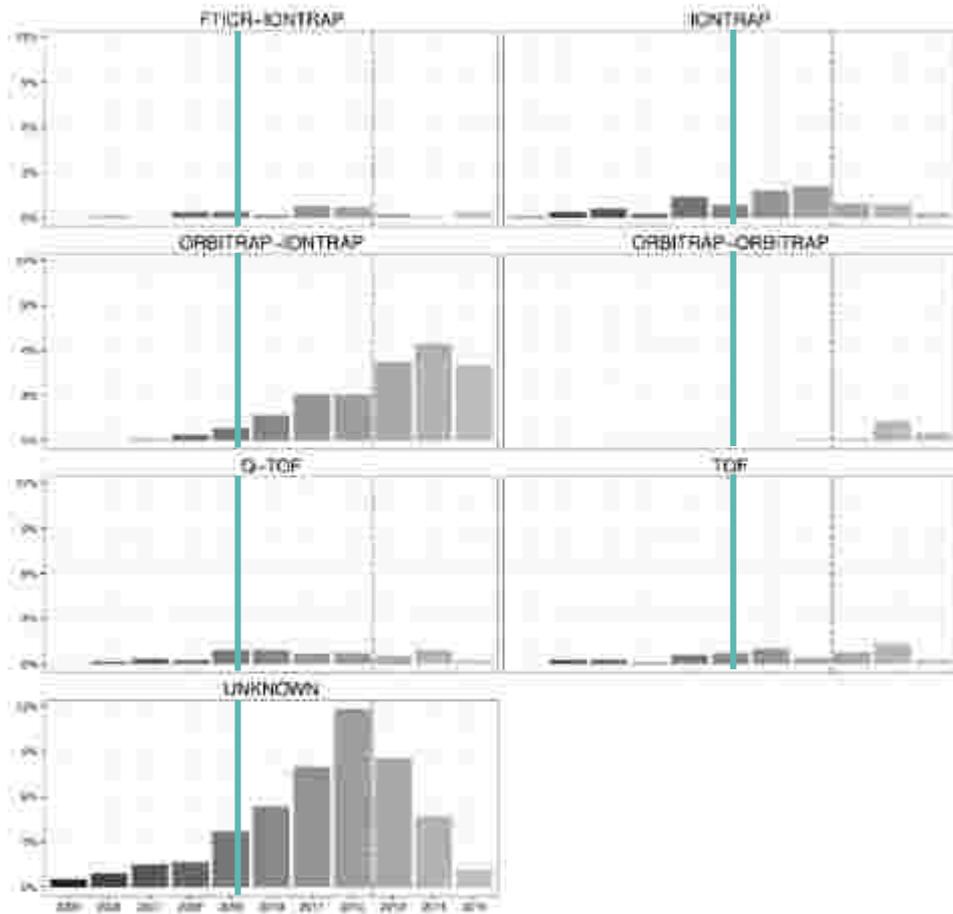
# Data sharing requires 3 building blocks: minimal requirements, CVs, and formats



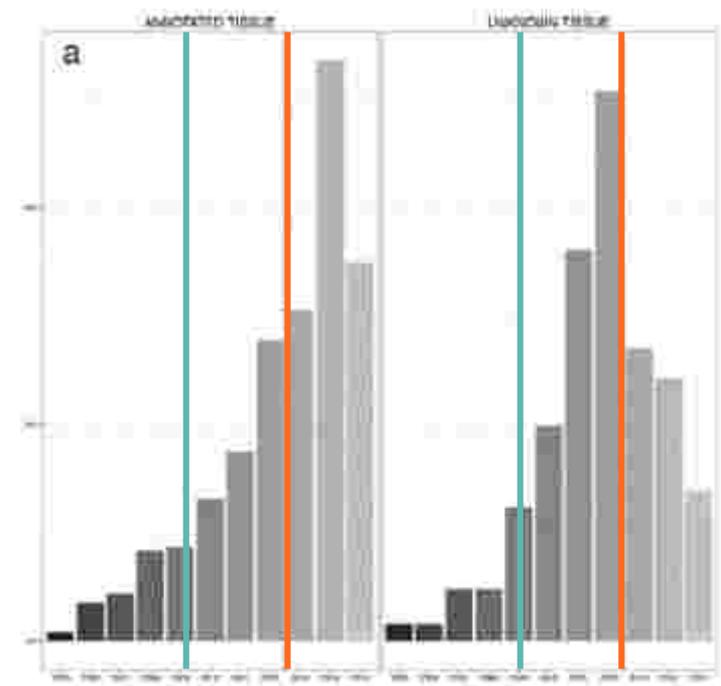
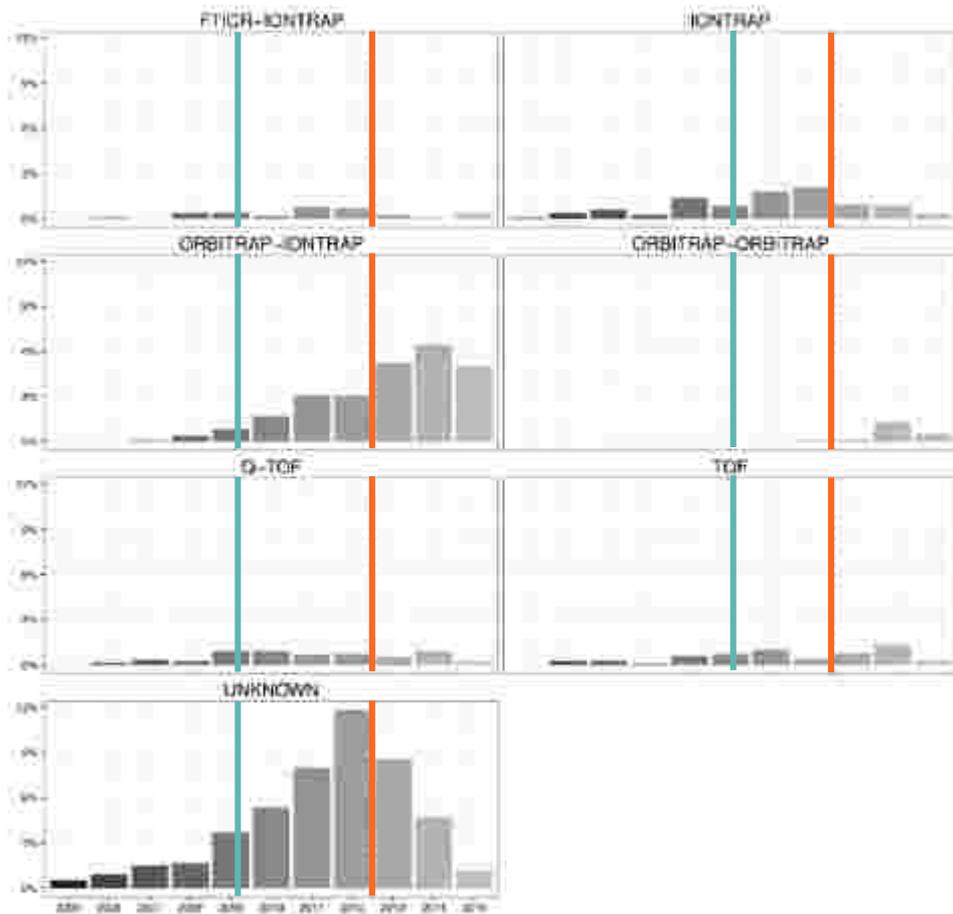
# Metadata is often the key issue, as it requires the most manual work



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



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



# 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 Flickr uses CC, amongst many others



# 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



# 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.

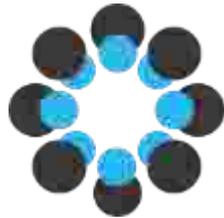
Delayed open access means that your paper becomes open access after an embargo period (often a year)

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



The screenshot shows the "Journal Selector" tool interface. At the top, there is a teal header with the text "Journal Selector". Below the header, there is a white area with instructional text: "Use the options below to find journals that match your requirements. The journals included in the tool are listed at the bottom of the page. Change the options for any of the questions and click Search. The journals list will update to give only the journals that fit your criteria. Click on a journal name for more information about it." Below this text, there is another line of text: "An explanation of the various options and abbreviations is [here](#), or click on the question marks on the right." The main part of the interface consists of five teal horizontal bars, each representing a filter option: "Subject", "Peer review", "Open access", "Speed", and "Other". Each bar has a white plus sign on the left and a white question mark on the right. At the bottom right of the interface, there is a link that says "Reset to default options" and a teal button with the word "Search" in white.

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



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



CC BY-SA 4.0

# 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...

The image shows a screenshot of a web browser displaying a research article on the Nature Genetics website. The article title is "Multi-tiered genomic analysis of head and neck cancer ties TP53 mutation to 3p loss". The authors listed are Ajilvaganan R. Gross, Hyun K. Cho, Joon P. Suh, Ashi Marie Eijffers, Marissa Carter, Mabel Andrea Michel-Dopoulos, Charles S. Coffey, Scott M. Lippman, David H. Hayes, Erik R. Cohen, Jennifer H. Grimes, Geyan T. Nguyen, and Troy Skalak. The article is published in Nature Genetics, volume 46, pages 842-847, in 2014. The abstract is partially visible, starting with "Head and neck squamous cell carcinoma (HNSCC) is characterized by aggressive behavior with a propensity for metastases and recurrence...". On the right side of the page, there is an advertisement for naturejobs.com, which includes a "naturejobs.com" logo and text about "Recruiting Assistant / Publishing Assistant" and "Springer Nature".

# ... but the code on GitHub represents the actual research performed

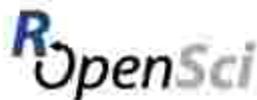
The screenshot shows the GitHub repository page for `theandygross/TCGA`. The repository has 1271 commits, 13 branches, and 2 releases. A table lists files and folders, including `Workflow`, `data`, `src`, `utils`, and `TCGAData`. A section titled "Workflow Overview" contains a paragraph of text:

This repository contains instructions for reproduction and extension of Myrka's *in-silico* genomic analysis of head/neck TCGA *metabarcodes* (i.e. *metabarcodes* by Gens et al.) in minimal code for data processing and visualization is provided in python modules, while *in-silico* level analysis was reproduced in Jupyter Notebooks. The analysis for this project was done and has been split into a number of notebooks as described in Analysis Overview. The results can be replicated by running these notebooks.

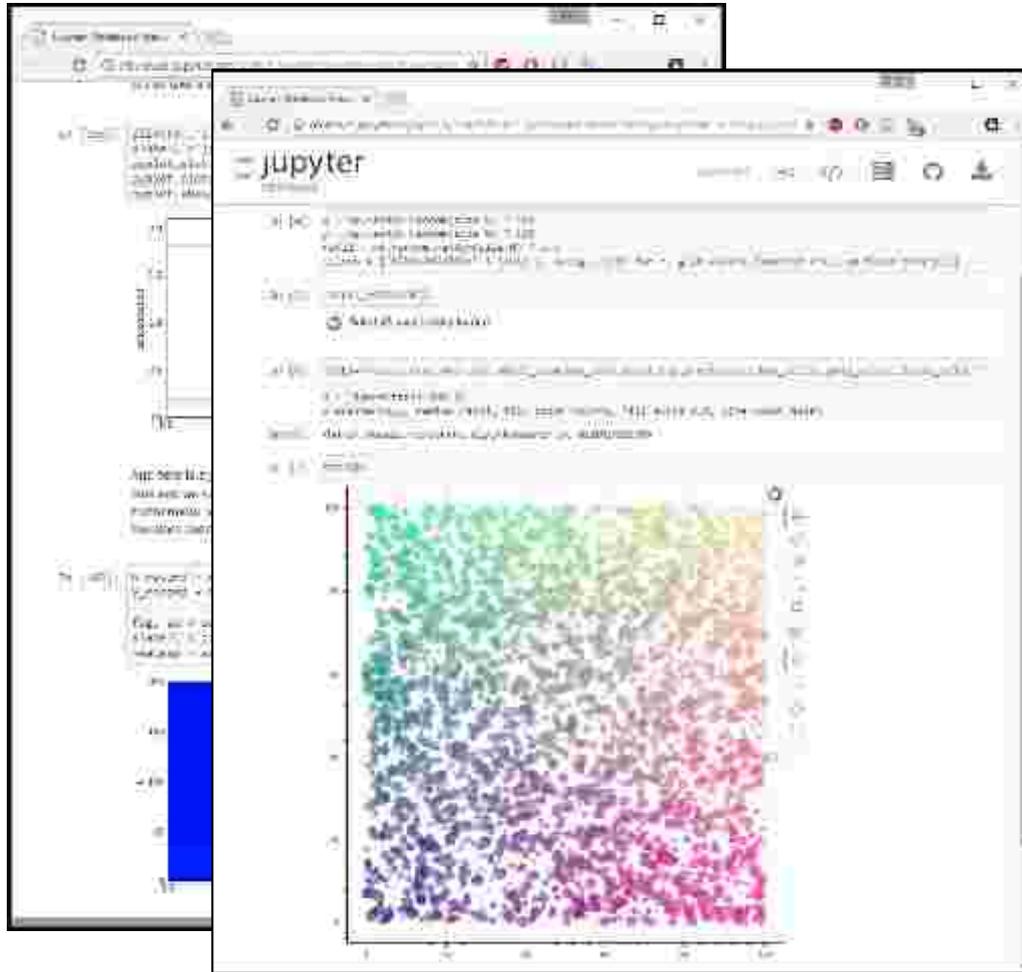
The screenshot shows a Jupyter Notebook titled "HNSCC Cohort". It contains text and code for data processing. A section titled "HPV Clinical Correlates" includes a plot showing the relationship between HPV status and a clinical correlate. The plot has a y-axis from 0.0 to 1.0 and an x-axis with categories 1, 2, 3, 4, 5. Two lines are plotted: a red line for "HPV+" and a black line for "HPV-". The red line starts at 1.0 and drops to approximately 0.5, while the black line starts at 1.0 and drops to approximately 0.8.

# NumFOCUS supports and promotes innovative, open source scientific software

NUMFOCUS  
OPEN CODE = BETTER SCIENCE



# 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!

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



The screenshot shows a GitHub repository page for 'pcmasuzzo / openscience'. The repository name is 'openscience' and the file path is '/README / Do you speak open science.md'. The repository has 1 commit and 1 file. The file name is 'Do you speak open science.md'. The file content is as follows:

**Do You Speak Open Science? Resources and Tips to Learn the Language.**

Paola Masuzzo, Z - ORCID: 0000-0002-7078-1195, Lennart Merxmal, Z - ORCID: 0000-0001-4277-8149

**Author Affiliation**

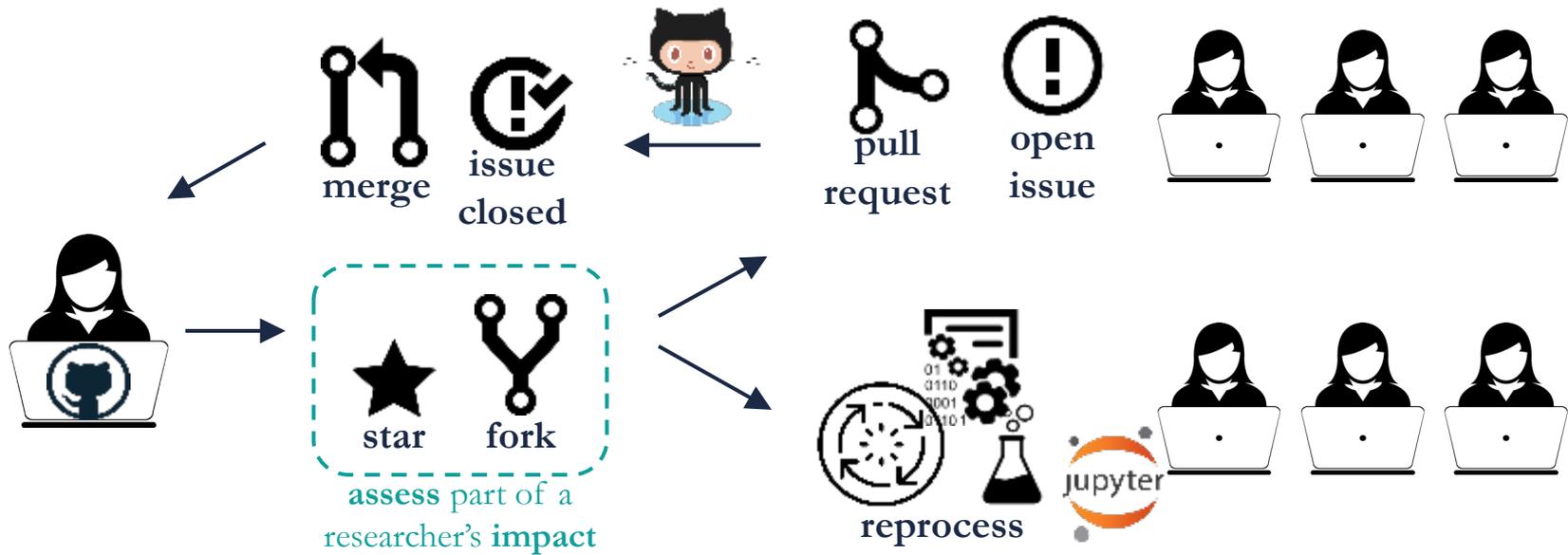
† Medical Biotechnology Center, Vrije Universiteit Brussel

> Department of Biochemistry, Ghent University, Ghent, Belgium

**Abstract**

The internet, i.e., large scale computing and storage resources, mobile devices, social media, and their tight integration among different groups of people, have all deeply changed the way knowledge is created, communicated, and further developed. These advances have enabled a radical transformation of the practice of science which is now more public, 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 funded and institutions, it has fast become a central 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.

# Open code allows collaboration as well as reproduction



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

## 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 allows you to do whatever you want with your code as long as they provide attribution back to you and don't hold you liable.

jQuery, AngularJS, and jQuery Mobile 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 GPL v3 is a copyleft license that requires anyone who modifies your code to make those changes available under the same terms, and guarantees its access (with appropriate social connections).

Drupal, Joomla!, and WordPress use the GNU GPL v3.

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.

The Content of this site is licensed under the Creative Commons Attribution 3.0 License. © 2014 ChooseALicense.com

MIT License: <https://opensource.org/licenses/MIT>  
Apache License 2.0: <https://www.apache.org/licenses/LICENSE-2.0>  
GNU GPL v3: <https://www.gnu.org/licenses/gpl-3.0.html>

# 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!

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

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

# 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)



# Comp omics



FACULTEIT GENEESKUNDE EN  
GEZONDHEIDSWETENSCHAPPEN

