Reproducibility and Reuse of Scientific Code

Evolving the Role and Capabilities of Publishers

Michael Forster
Managing Director, Publications
m.b.forster@ieee.org

Pierre Montagano
Director of Business Development
pierre@codeocean.com

IEEE

CODE OCEAN

Agenda

- Redefining publishing
- Open science and why code is important
- Reproducibility, transparency and reuse
- IEEE and Code Ocean partnership
- Technology as an enabler => cultural change

Redefining publishing

Support Increasingly Complex Infrastructure

Providing Services and Tools to Authors and Users

Move from PDF to XML/HTML/Mobile

Plagiarism screening

Create, adopt, and manage identifiers: ORCID, CHORUS, FunderID, OrgID etc.

Provide greater visibility, impact, and recognition for authors (and reviewers)

Redefining what “Publishing” is and how we deliver (and earn) value

~~**Registration**~~ **: Certification : ~~Dissemination~~ : ~~Preservation~~**

Preprint servers

Scholarly Collaboration Networks

Discovery Services (new and old)

Pirate sites (but widely used)

Open science and why code is important

Open Science, not just Open Access

What might a new world of Certification (peer review) look like?

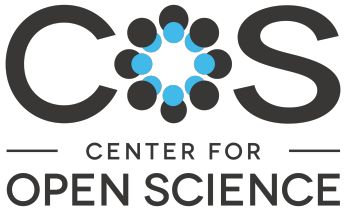
Open Access mandates worldwide

Open Science

- Data sets
- Algorithms, Models, Code

Reproducible Research

- Badging
- Pre-registration?



More complex / different review process? Role of reviewers? Fewer 'published' articles?

Open Science, not just Open Access

- Researchers, funders, Publishers and governments
  - Increasingly expected to go beyond the Article
  - Open science (open data, open code, open workflows) initiatives are taking center stage within the scientific community
  - Government / Agency funding mandates
- Data and code are becoming citable research output
- Reuse drives innovation

Reproducibility, transparency and reuse

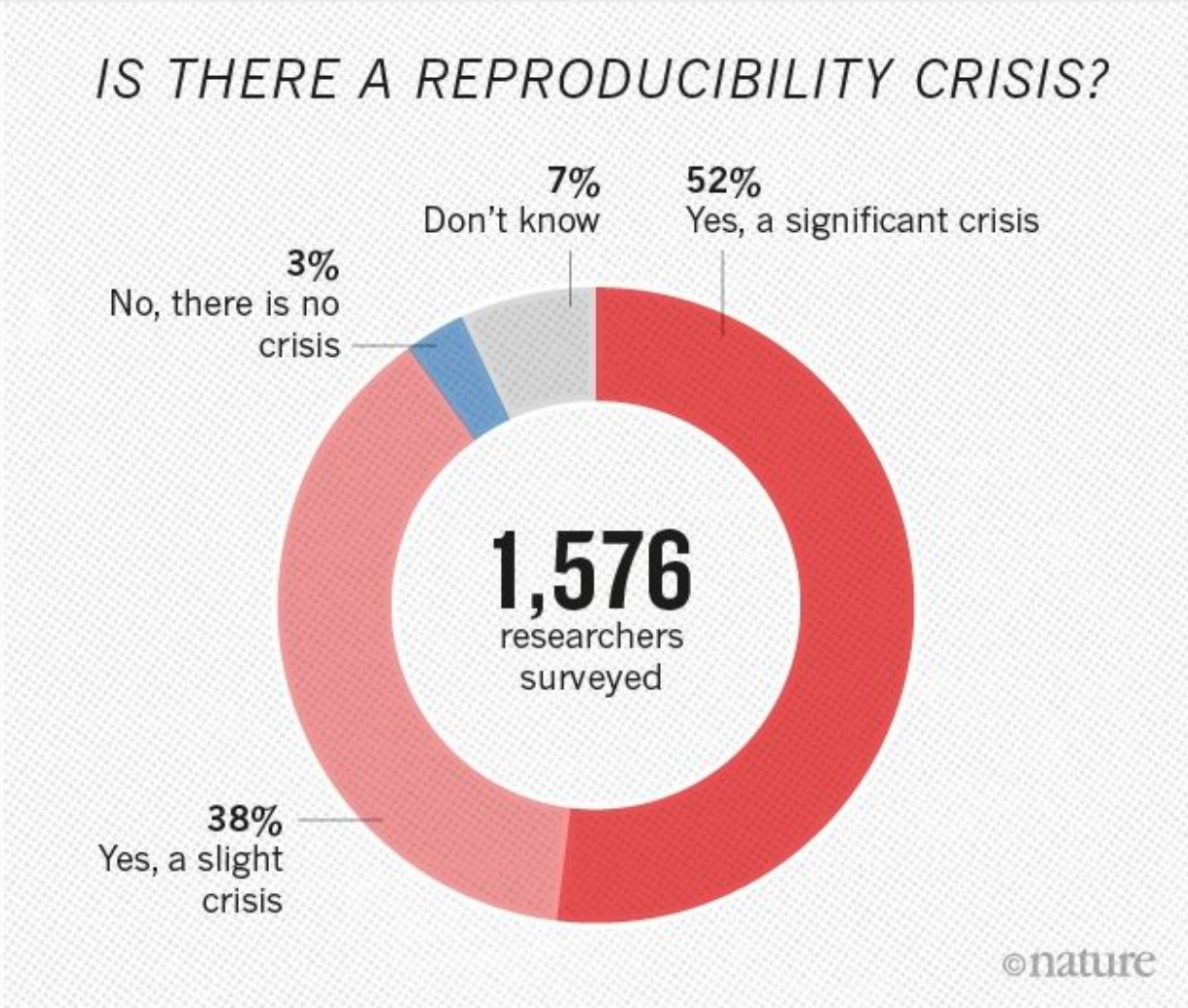
Code Ocean was created to maximize reproducibility of research

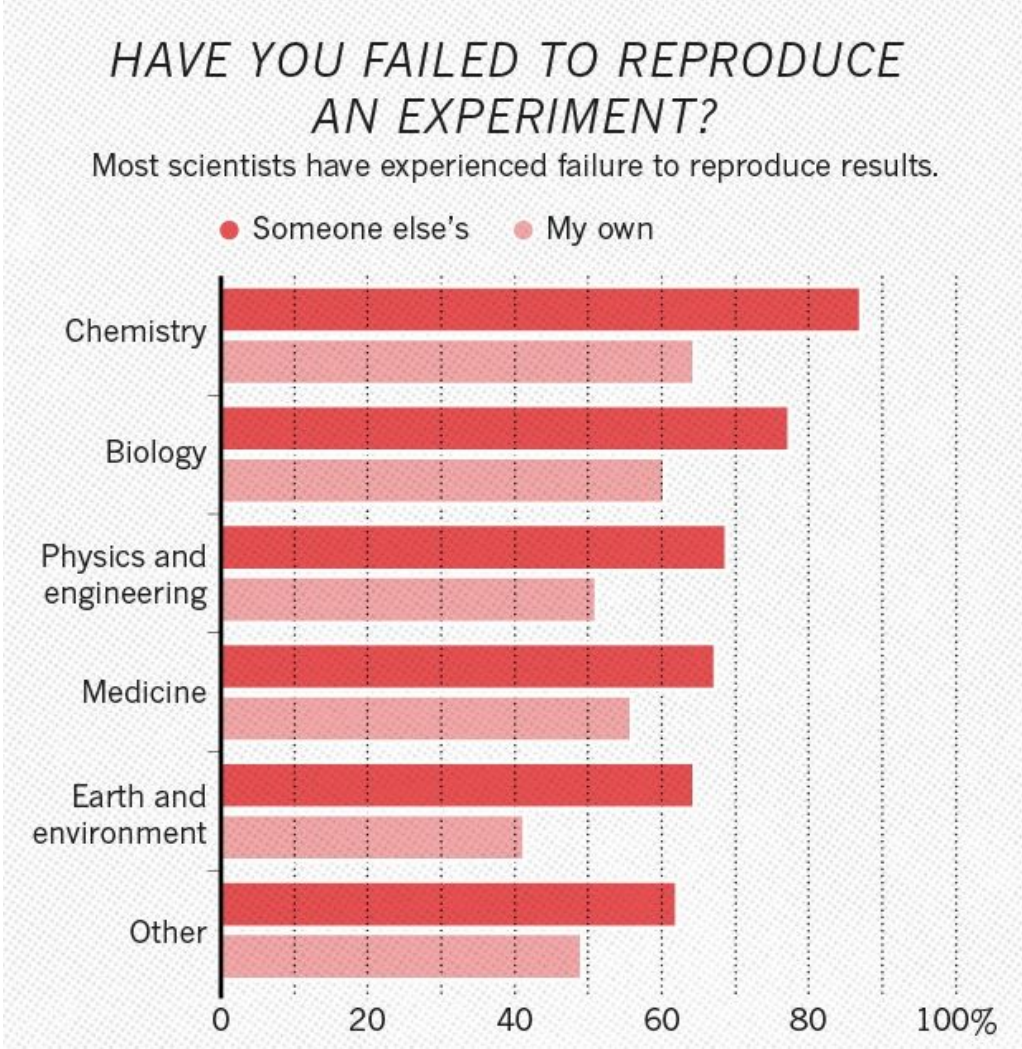
People trying to repeat others' research often do not have the time, funding or resources to gain the same expertise with the experimental protocol as the original authors, who were perhaps operating under a multi-year federal grant and aiming for a high-profile publication. If a researcher spends six months, say, trying to replicate such work and reports that it is irreproducible, that can deter other scientists from pursuing a promising line of research, jeopardize the original scientists' chances of obtaining funding to continue it themselves, and potentially damage their reputations.

THE SCHOLARLY kitchen

Reproducible Research, Just Not Reproducible By You

By DAVID CROTTY | MAY 24, 2017

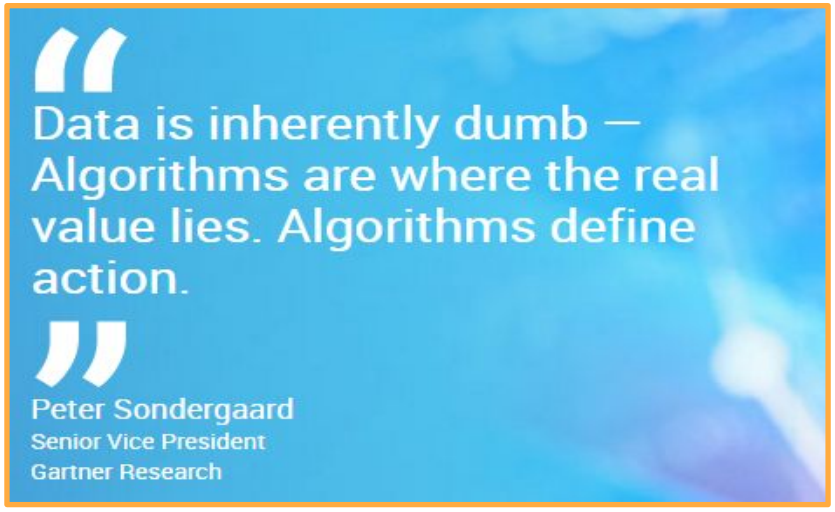




Nature **533,** 452–454 (26 May 2016) doi:10.1038/533452a

IEEE and Code Ocean partnership

Algorithms will continue to grow



Code v. Algorithm

1. **Algorithm** is an idea, a concept. **Code** is a practical realization of the algorithm.
2. **Algorithm** is the abstract recipe for the calculation, independent of implementation.
3. **Code** is written in a particular language. **Algorithms** are language independent.

Seven major themes about the algorithm era

INEVITABLE ALGORITHMS

Theme 1 **Algorithms will continue to spread everywhere**

- The benefits will be visible and invisible and can lead to greater human insight into the world
- The many upsides of algorithms are accompanied by challenges

Theme 2 **Good things lie ahead**

- Data-driven approaches to problem-solving will expand
- Code processes will be refined and improved; ethical issues are being worked out
- "Algorithms don't have to be perfect; they just have to be better than people"

Source – Pew Research Center Feb 8, 2017.

Code Ocean provides IEEE with a platform for executable research

- Every month at least 3,000 [1] IEEE papers released on *Xplore* contain at least one published algorithm.
- Code Ocean enables modifications, experimentation / enhancements, and execution of code to reproduce results.
- Hence **Code Ocean is a platform to:** *host software, so one can execute the authors' code published in an IEEE paper*.

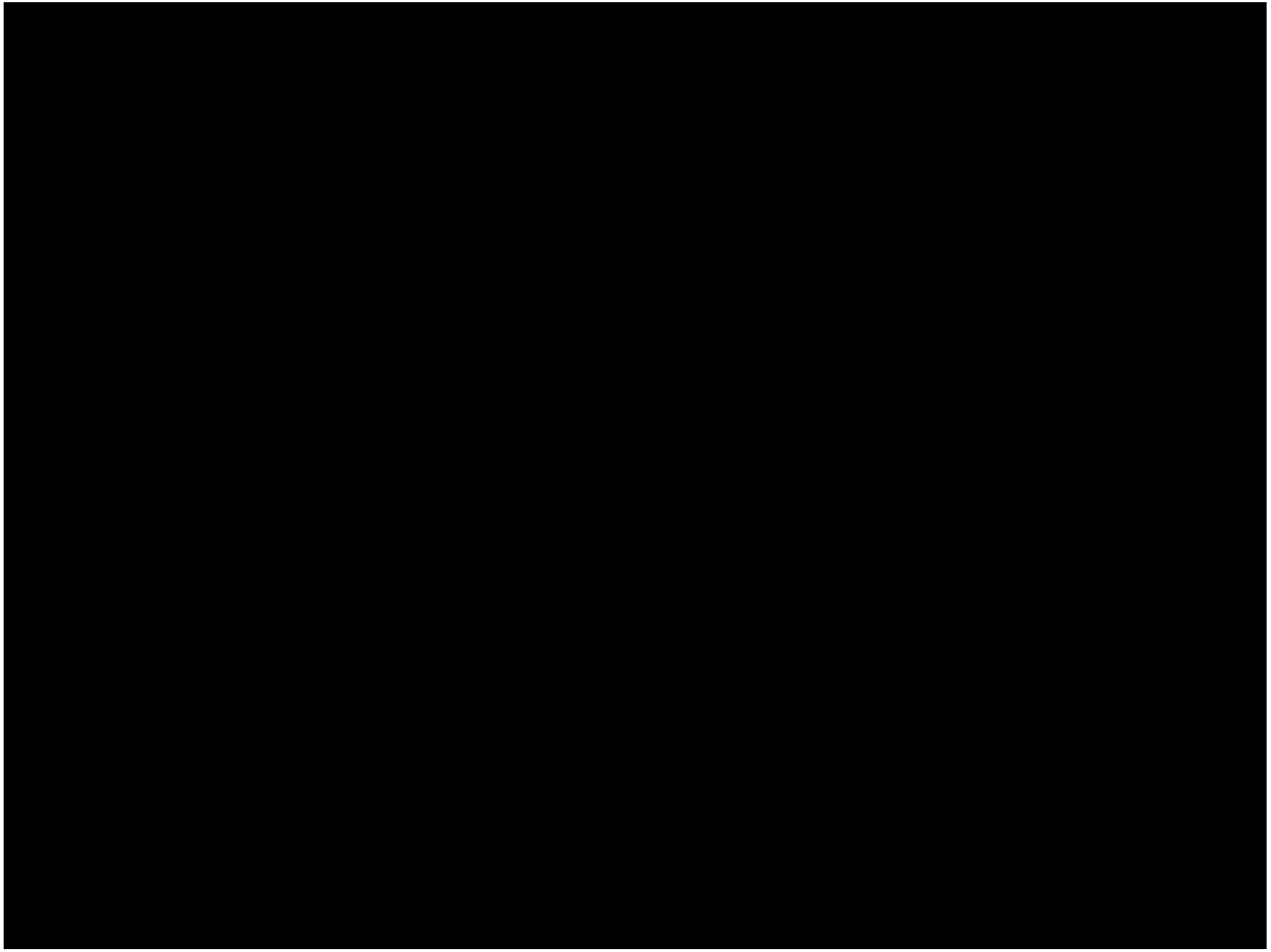
Code Ocean is an easy-to-use scalable platform to share and run code in the cloud

✓ Upload Software Implementations of Code

✓ Run Code in a cloud platform and via IEEE Xplore

✓ Modify and experiment with Code

[1] Based on items tagged as "Algorithm" in IEEE XML articles since 2001.



http://ieeexplore.ieee.org/search/searchresult.jsp?refinements=4219406591

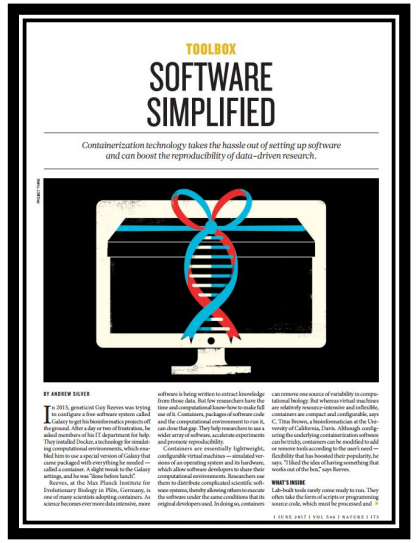
The Code Ocean platform is built on open, standard technologies

- **Code Upload Process**
  - Organized folder structure designed for ease of use.
  - The approach is similar to other code repositories, such as GitHub
- **Software application**
  - Code Ocean uses industry standard programming tools
  - Code Ocean application does not determine how code is executed. It simply provides an environment to find, compile, and run code in the cloud
- **Execution**
  - Code is executed on servers hosted on Amazon Web Services (AWS)
  - Code can be easily downloaded & run on local system
  - Author can add details about the execution environment to the metadata
- **Non Proprietary**
  - All code execution tools are either commercially available or open source.

Nature - May 29 2017

Software Simplified

"*Containerization technology takes the hassle out of setting up software and can boost reproducibility of data driven research"*



TOOLBOX

SOFTWARE SIMPLIFIED

Containerization technology takes the hassle out of setting up software and can boost the reproducibility of data-driven research.

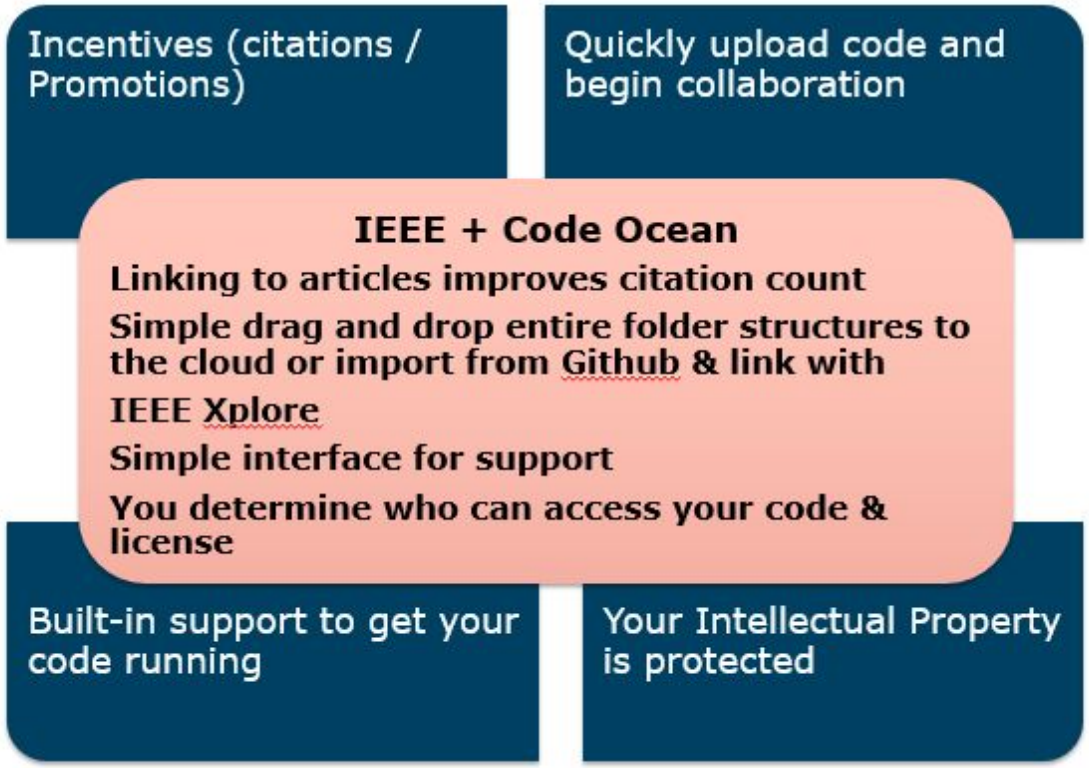
BY ANDREW SILVER

IEEE

CODE OCEAN

Author Demand: Why IEEE authors should share code on Xplore and Code Ocean

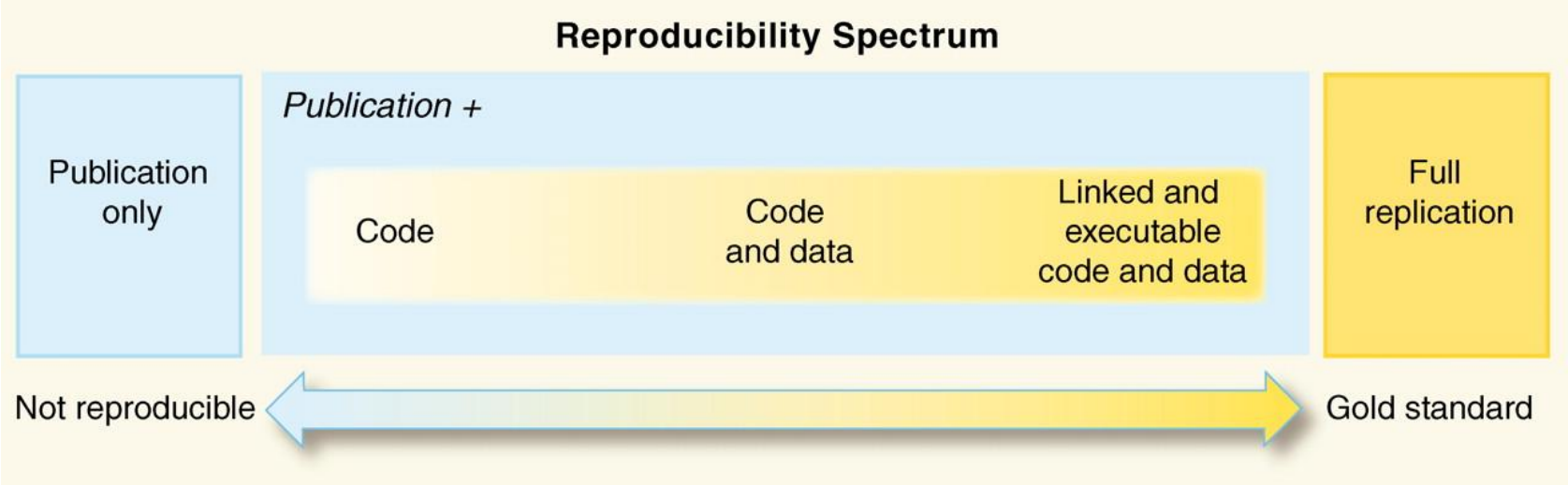
- Supports Reproducible Research
- Enhances Author Visibility
- Reduces Customer Support burden on an Author
- *Author owns the algorithm and assigns license.*
- Collaboration



source: "Implementing Reproducible Research, V. Stodden, Friedrich Leisch, Roger D. Peng, 2013

Levels of computational reproducibility

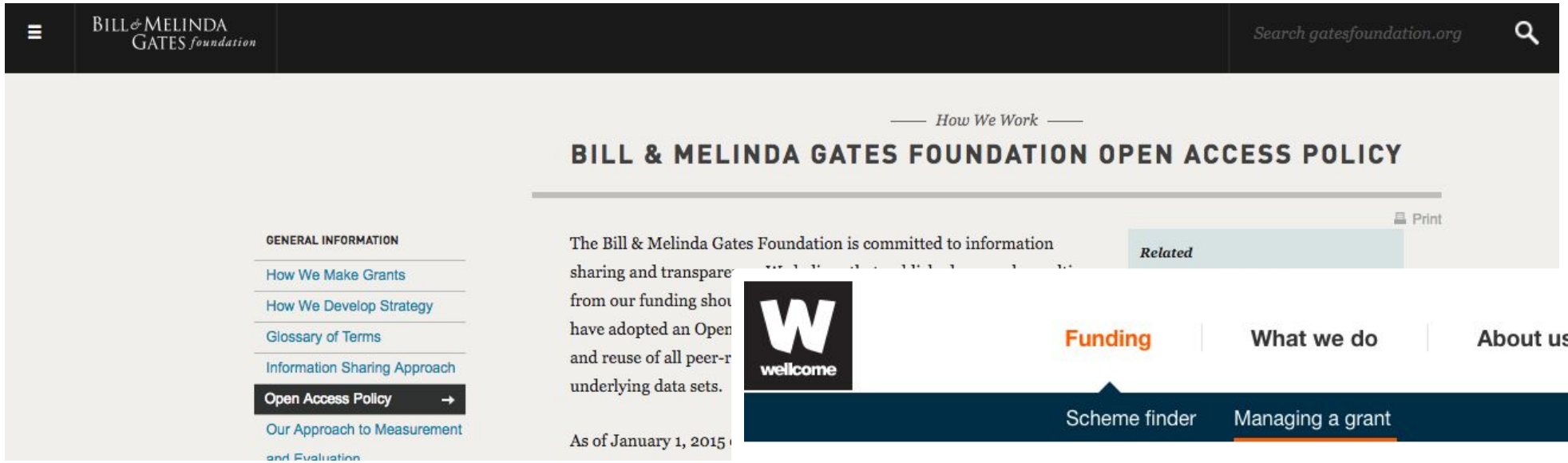
Software today is an integral part of research across most scientific disciplines.



Peng (2011), Science

Technology as an enabler ➡ cultural change

https://brucegerencser.net/2015/01/heaven-hell-carrot-stick/



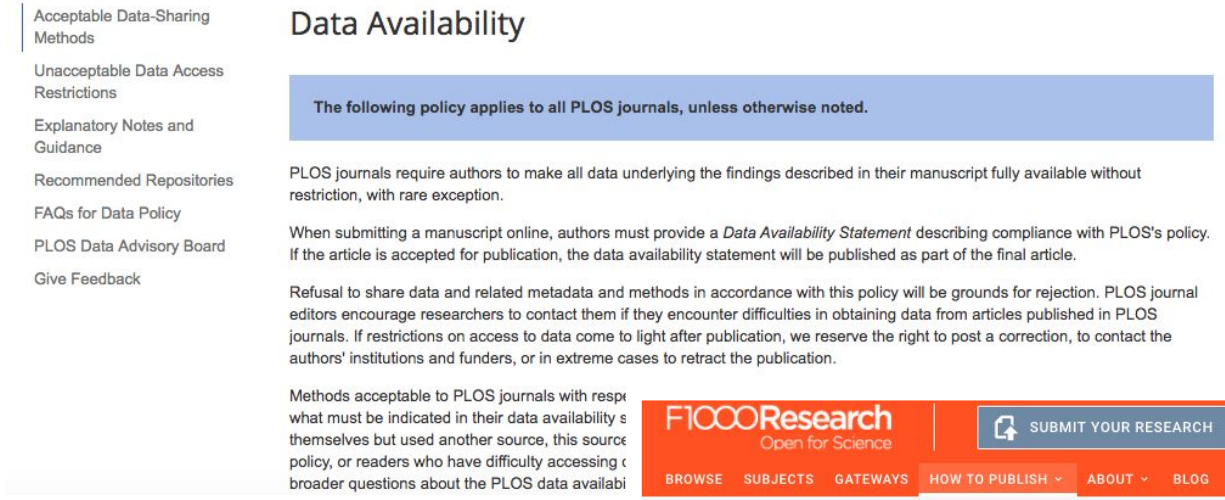
Open access policy

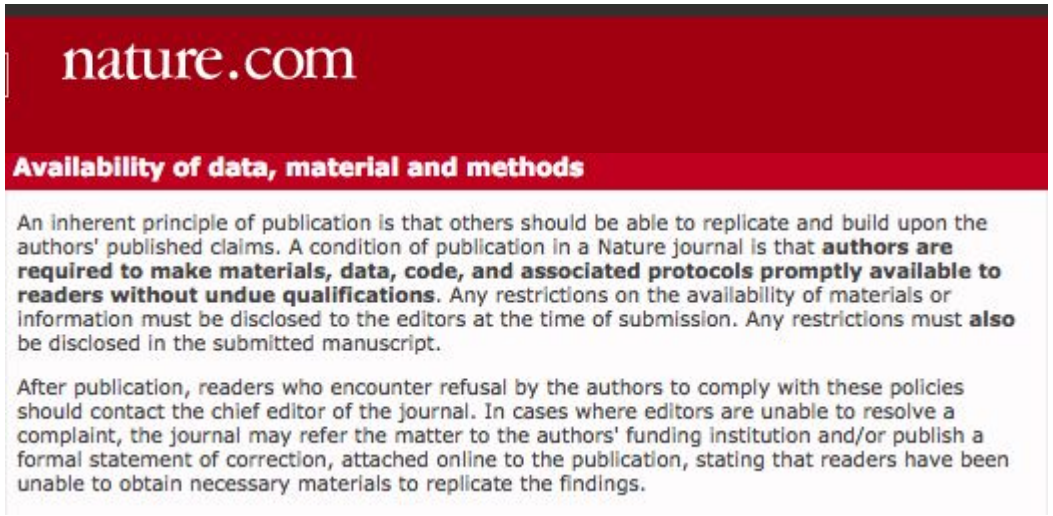
Our mission is to improve health by supporting bright minds in science, the humanities and social sciences, and public engagement.

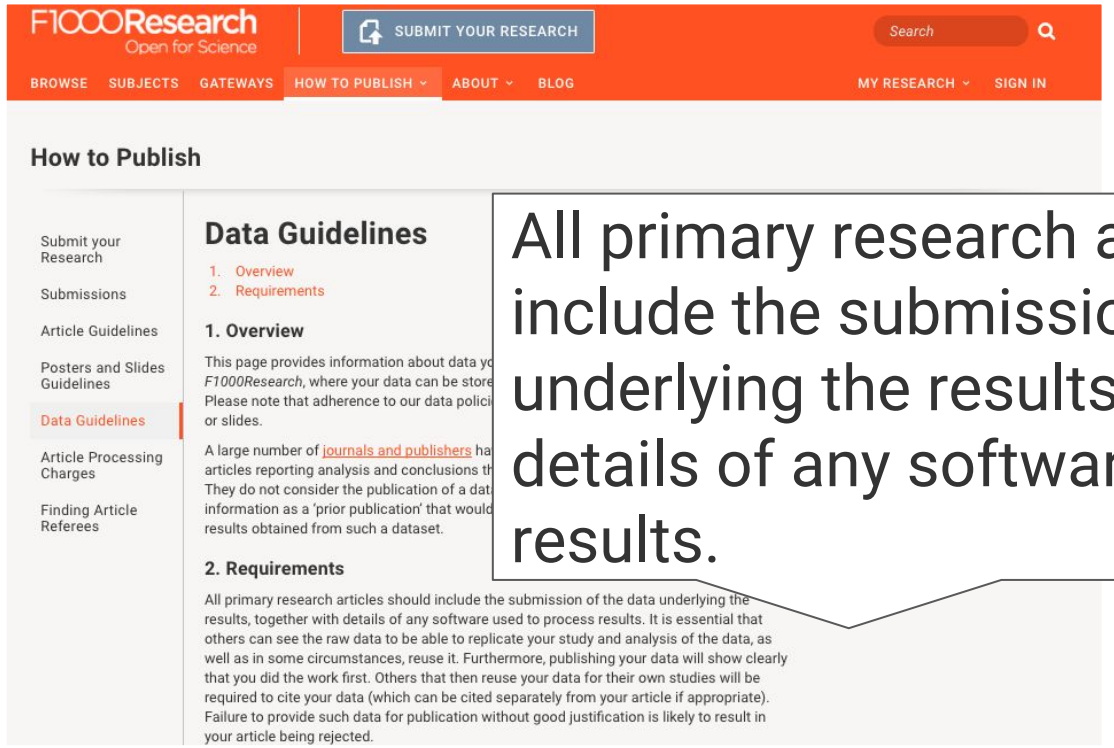
The main output of this research is new ideas and knowledge, which we expect our researchers to publish as high-quality, peer-reviewed research articles, monographs and book chapters.

We believe that maximising the distribution of these publications – by providing free, online access – is the most effective way of ensuring that the research we fund can be accessed, read and built upon. In turn, this will foster a richer research culture.

We therefore support unrestricted access to the published output of research as a fundamental part of its charitable mission and a public benefit to be encouraged wherever possible.







All primary research articles should include the submission of the data underlying the results, together with details of any software used to process results.

IEEE

CODE OCEAN

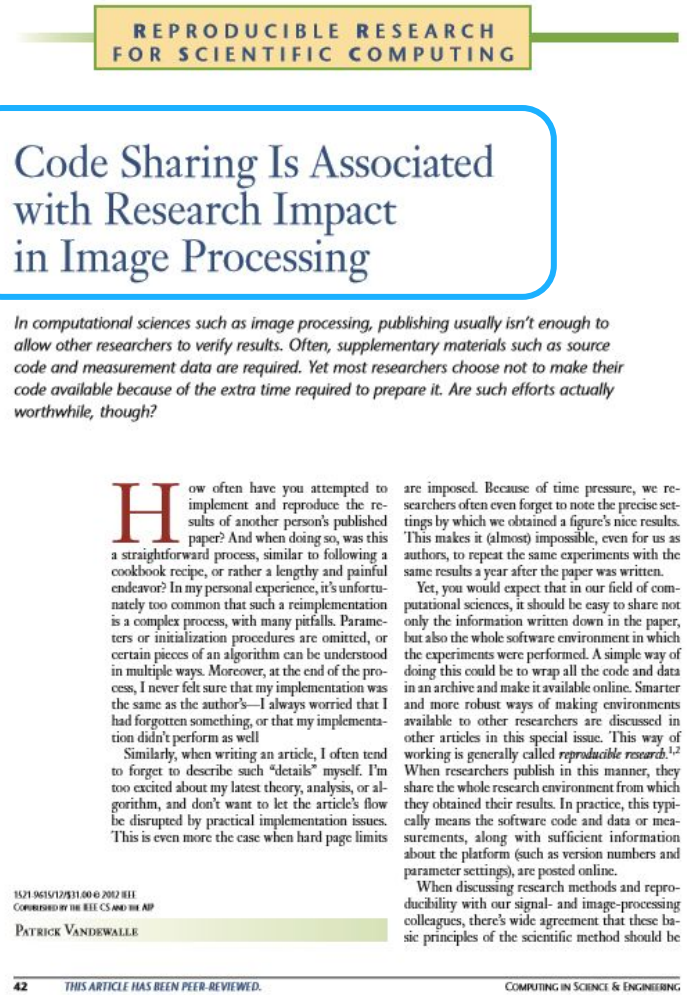
http://vegnerds.co.uk/index.php/2015/06/08/bunch-of-carrots/

Incensed citations and impact

COPUBLISHED BY THE IEEE CS AND THE AIP

PATRICK VANDEWALLE

COMPUTING IN SCIENCE & ENGINEERING



REPRODUCIBLE RESEARCH FOR SCIENTIFIC COMPUTING

Code Sharing Is Associated with Research Impact in Image Processing

In computational sciences such as image processing, publishing usually isn't enough to allow other researchers to verify results. Often, supplementary materials such as source code and measurement data are required. Yet most researchers choose not to make their code available because of the extra time required to prepare it. Are such efforts actually worthwhile, though?

How often have you attempted to implement and reproduce the results of another person's published paper? And when doing so, was this a straightforward process, similar to following a cookbook recipe, or rather a lengthy and painful endeavor? In my personal experience, it's unfortunately too common that such a reimplementation is a complex process, with many pitfalls. Parameters or initialization procedures are omitted, or certain pieces of an algorithm can be understood in multiple ways. Moreover, at the end of the process, I never felt sure that my implementation was the same as the author's—I always worried that I had forgotten something, or that my implementation didn't perform as well

Similarly, when writing an article, I often tend to forget to describe such "details" myself. I'm too excited about my latest theory, analysis, or algorithm, and don't want to let the article's flow be disrupted by practical implementation issues. This is even more the case when hard page limits are imposed. Because of time pressure, we researchers often even forget to note the precise settings by which we obtained a figure's nice results. This makes it (almost) impossible, even for us as authors, to repeat the same experiments with the same results a year after the paper was written.

Yet, you would expect that in our field of computational sciences, it should be easy to share not only the information written down in the paper, but also the whole software environment in which the experiments were performed. A simple way of doing this could be to wrap all the code and data in an archive and make it available online. Smarter and more robust ways of making environments available to other researchers are discussed in other articles in this special issue. This way of working is generally called *reproducible research*.[1,2] When researchers publish in this manner, they share the whole research environment from which they obtained their results. In practice, this typically means the software code and data or measurements, along with sufficient information about the platform (such as version numbers and parameter settings), are posted online.

When discussing research methods and reproducibility with our signal- and image-processing colleagues, there's wide agreement that these basic principles of the scientific method should be

1521-9615/12/$31.00 © 2012 IEEE
COPUBLISHED BY THE IEEE CS AND THE AIP

PATRICK VANDEWALLE

42 THIS ARTICLE HAS BEEN PEER-REVIEWED. COMPUTING IN SCIENCE & ENGINEERING

It is important to IEEE authors

Increases:

+ Citations by 3X
+ Downloading
+ Collaboration

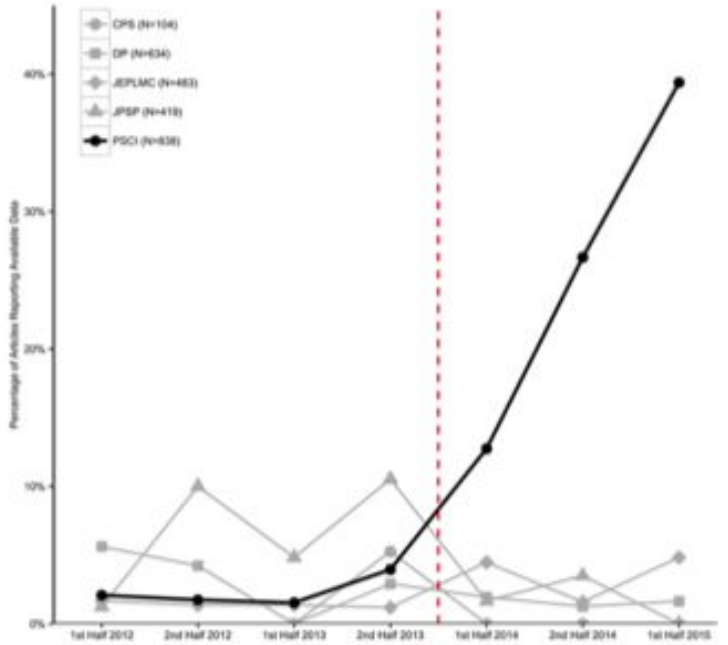
= ...IMPACT!

Badging

Can badging influence and steer user behavior?

- Indicators are promising
- Research outside of scholarly publishing is also promising
  - https://www.cs.cornell.edu/home/kleinber/www13-badges.pdf
  - https://doi.org/10.1016/j.chb.2015.03.036

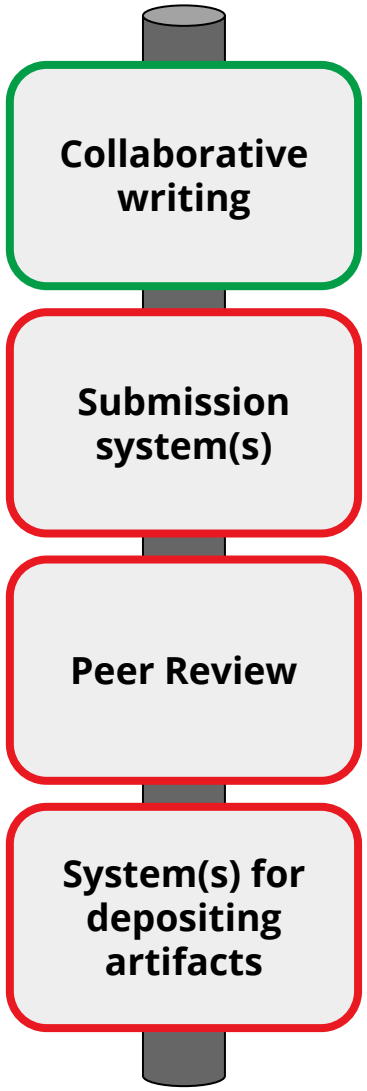
Fig 2. Reportedly available data.



Kidwell MC, Lazarević LB, Baranski E, Hardwicke TE, Piechowski S, et al. (2016) Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. PLOS Biology 14(5): e1002456. https://doi.org/10.1371/journal.pbio.1002456
http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002456

PLOS BIOLOGY

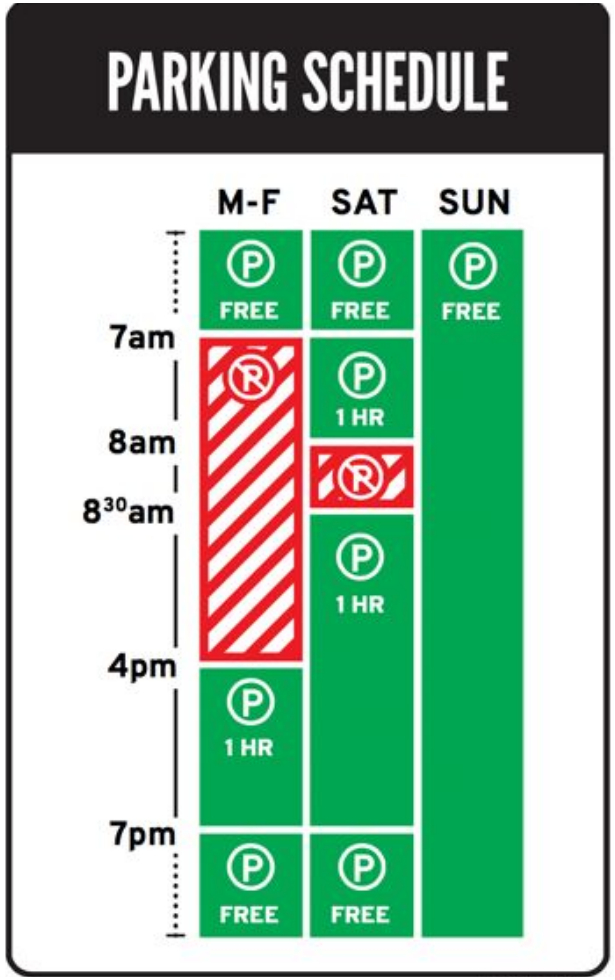
User Experience



LA Parking Signs

Author/Copyright holder: Jorge Gonzalez.
Copyright terms and licence: CC BY-SA 2.0

Nikki Sylianteng's Parking Sign



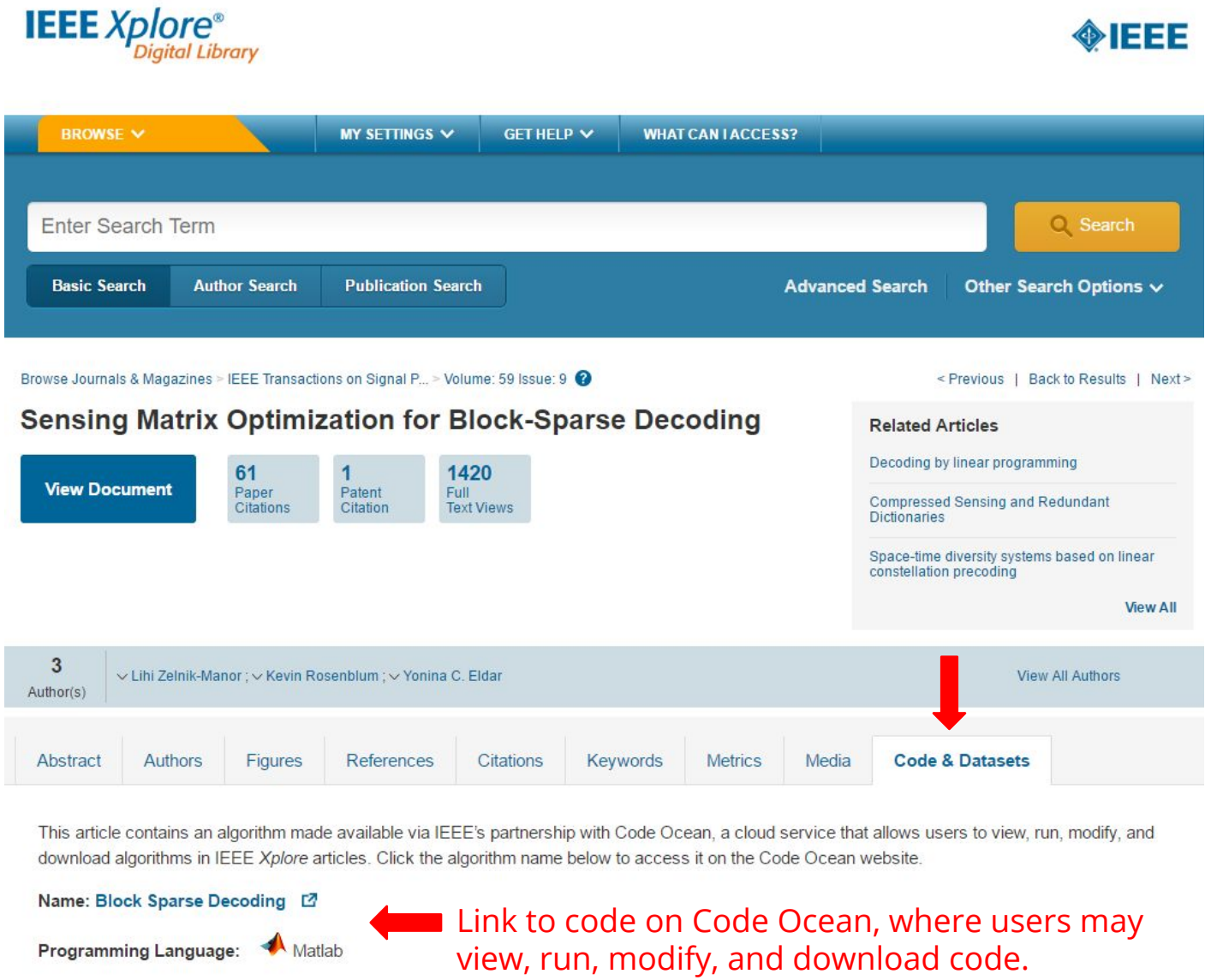
Author/Copyright holder: Nikki Sylianteng.
Copyright terms and licence: CC BY-NC-SA 4.0

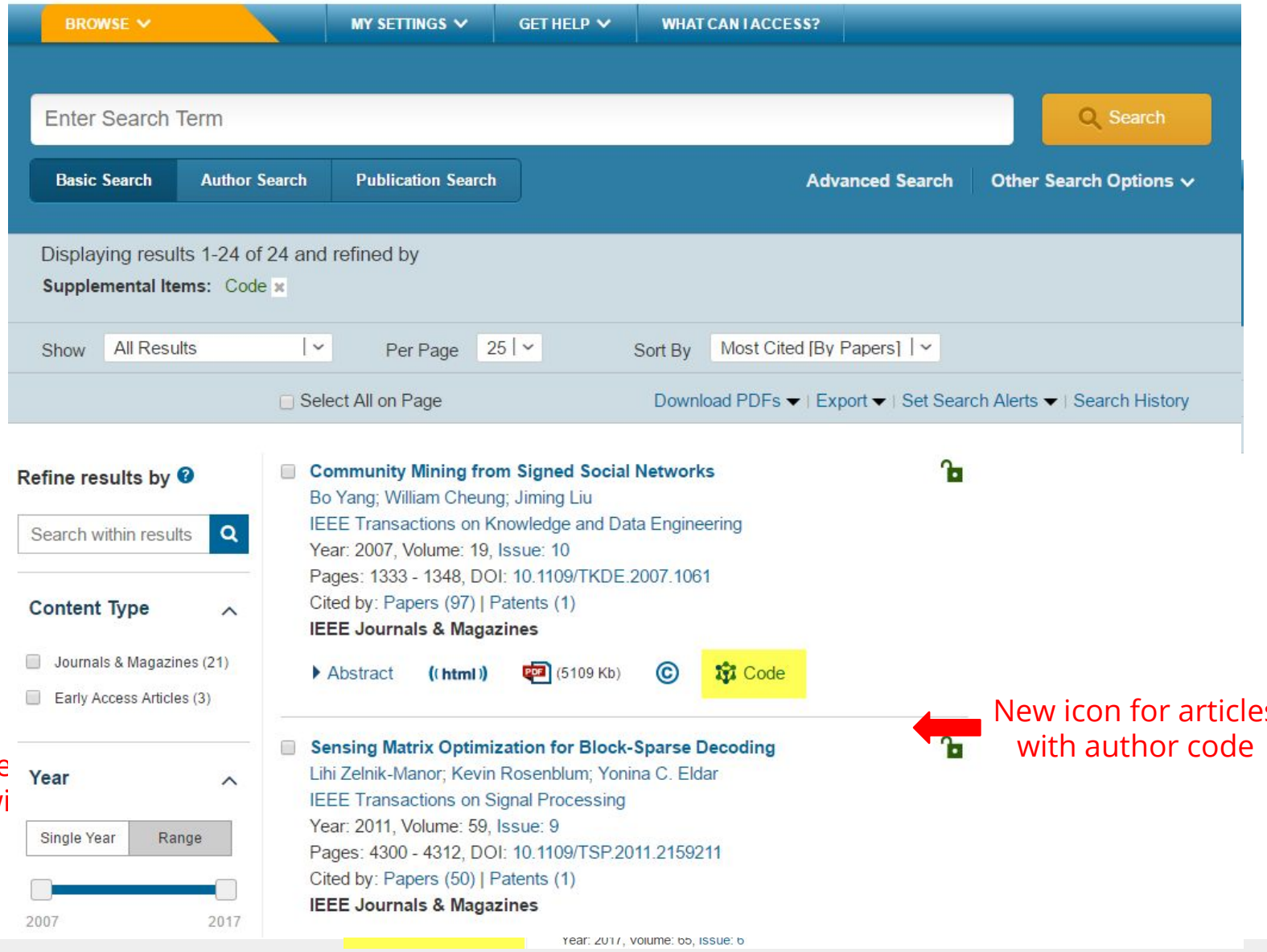
Thank you for your time

Michael Forster
Managing Director, Publications
m.b.forster@ieee.org

Pierre Montagano
Director of Business Development
pierre@codeocean.com

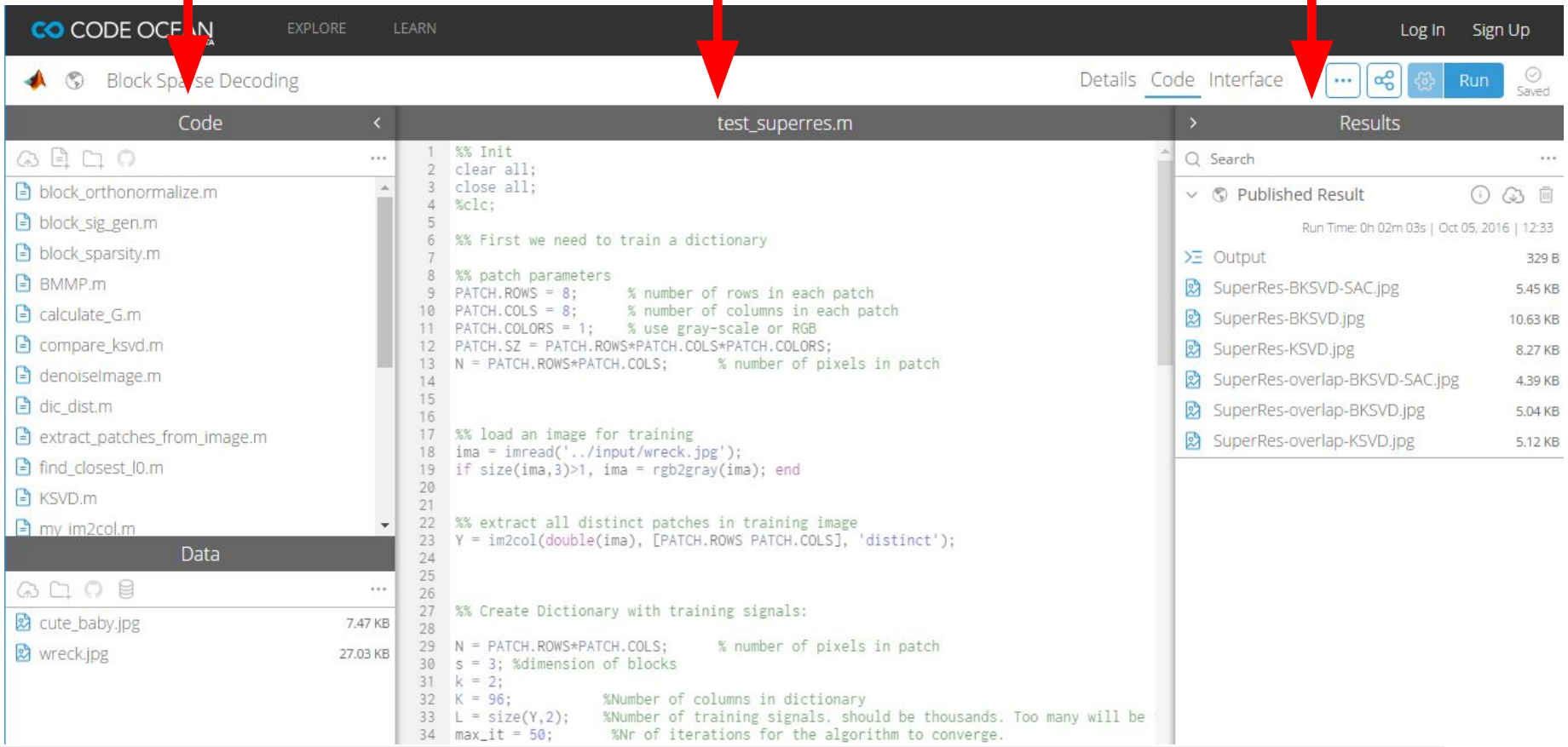
New Code & Datasets Tab within Article




IEEE Xplore® Digital Library
IEEE
BROWSE
MY SETTINGS
GET HELP
WHAT CAN I ACCESS?
Enter Search Term
Search
Basic Search
Author Search
Publication Search
Advanced Search
Other Search Options
Displaying results 1-24 of 24 and refined by
Supplemental Items: Code
Show All Results
Per Page 25
Sort By Most Cited [By Papers]
Select All on Page
Download PDFs | Export | Set Search Alerts | Search History
Refine results by
Search within results
Content Type
Journals & Magazines (21)
Early Access Articles (3)
Year
Single Year
Range
2007
2017
Community Mining from Signed Social Networks
Bo Yang; William Cheung; Jiming Liu
IEEE Transactions on Knowledge and Data Engineering
Year: 2007, Volume: 19, Issue: 10
Pages: 1333 - 1348, DOI: 10.1109/TKDE.2007.1061
Cited by: Papers (97) | Patents (1)
IEEE Journals & Magazines
Abstract
html
(5109 Kb)
Code
Sensing Matrix Optimization for Block-Sparse Decoding
Lihi Zelnik-Manor; Kevin Rosenblum; Yonina C. Eldar
IEEE Transactions on Signal Processing
Year: 2011, Volume: 59, Issue: 9
Pages: 4300 - 4312, DOI: 10.1109/TSP.2011.2159211
Cited by: Papers (50) | Patents (1)
IEEE Journals & Magazines
New icon for articles with author code

Author's Code on Code Ocean

Source Files

Code

Monitor Code Execution/Results