Authors & Algorithms

IEEE Xplore®
Digital Library

CODE OCEAN
Building the world's algorithmic library

IEEE

Algorithms will continue to grow

> "Data is inherently dumm — Algorithms are where the real value lies. Algorithms define action."
>
> Peter Sondergaard
> Senior Vice President
> Gartner Research

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

Source – Pew Research Center Feb 8, 2017.

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"
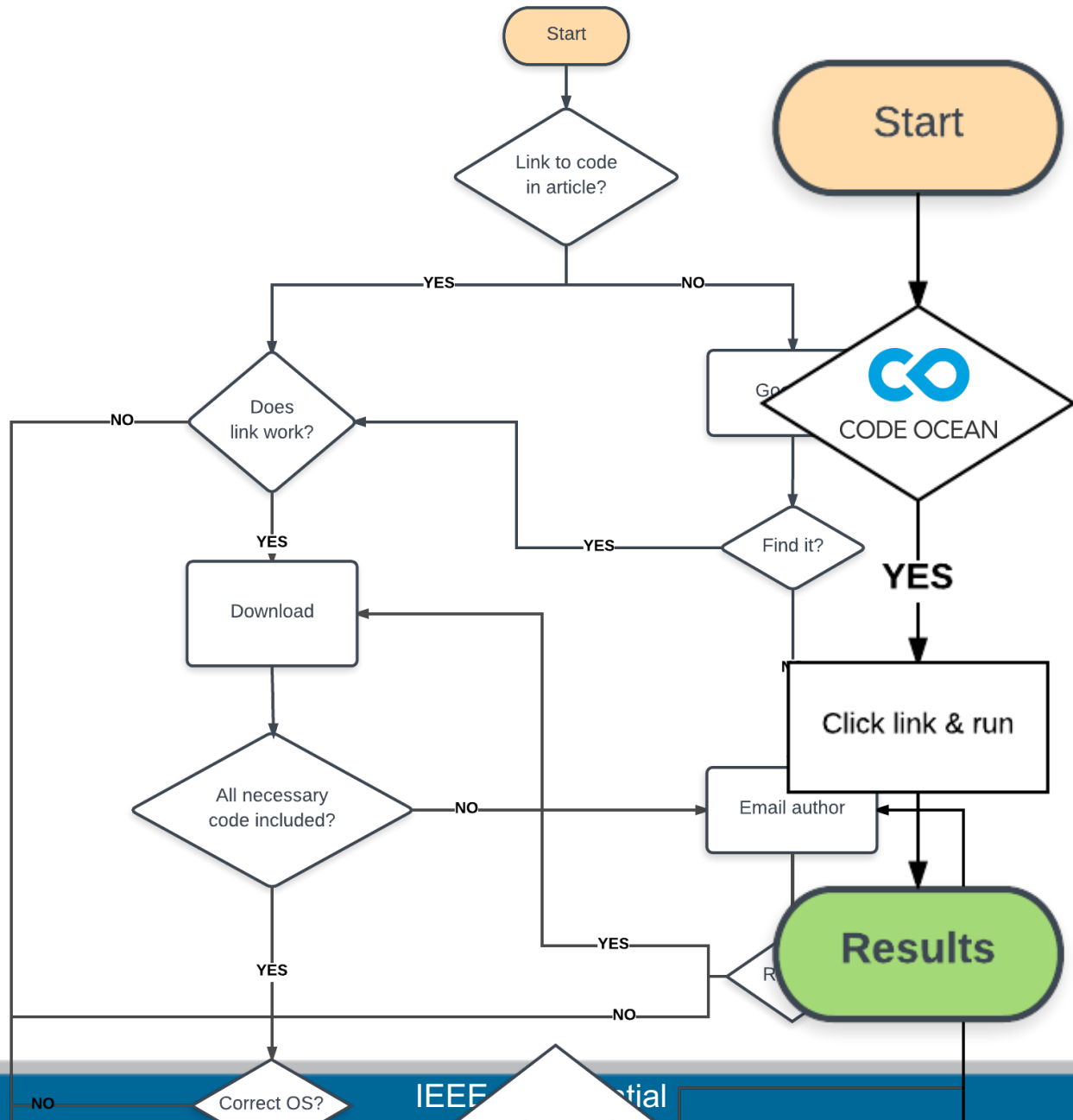- In the future, the world may be governed by benevolent AI

Code Ocean was created to maximize reproducibility of research

THE SCHOLARLY
kitchen

Reproducible Research, Just Not Reproducible By You

By DAVID CROTTY | MAY 24, 2017

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.

IEEE
Advancing Technology for Humanity

Start

Link to code in article?

YES   NO

Does link work?

NO

Go

YES

Find it?

YES

Download

All necessary code included?

NO   Email author

YES

YES   NO

Correct OS?

NO

Start

CODE OCEAN

YES

Click link & run

Results

IEEE ... tial

# Code Ocean provide IEEE with a platform for executable research

- IEEE Papers contain code.
- 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

C/C++
Fortran
Java
Julia
Lua
Matlab
Octave
Perl
Python
R
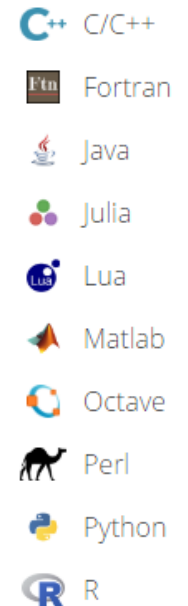
[1] Based on items tagged as "Algorithm" in IEEE XML articles since 2001.

Code Ocean is now integrated with IEEE *Xplore*

Code File

Code

Code Execution

Code Data

Code Output

Code Tab

Code – CO link

The Code Ocean platform is built on open, standard technologies

Code Upload Process
- Folder structure and organization is completely determined by the author.
- 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 browse, compile, and run code in the cloud
- Same environment is also available easily on desktops

Algorithm 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
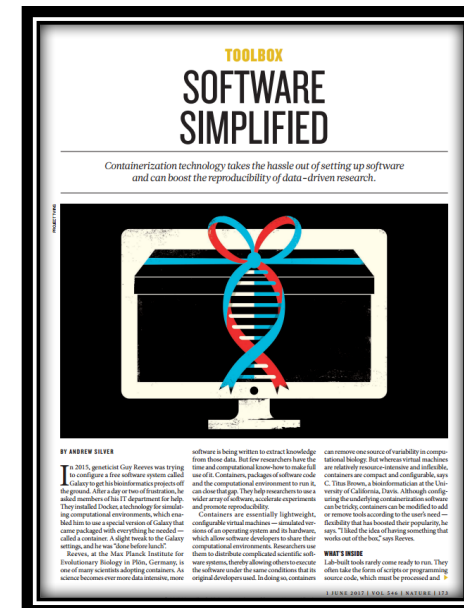
Non Proprietary
- All code execution tools are either commercially available or open source.
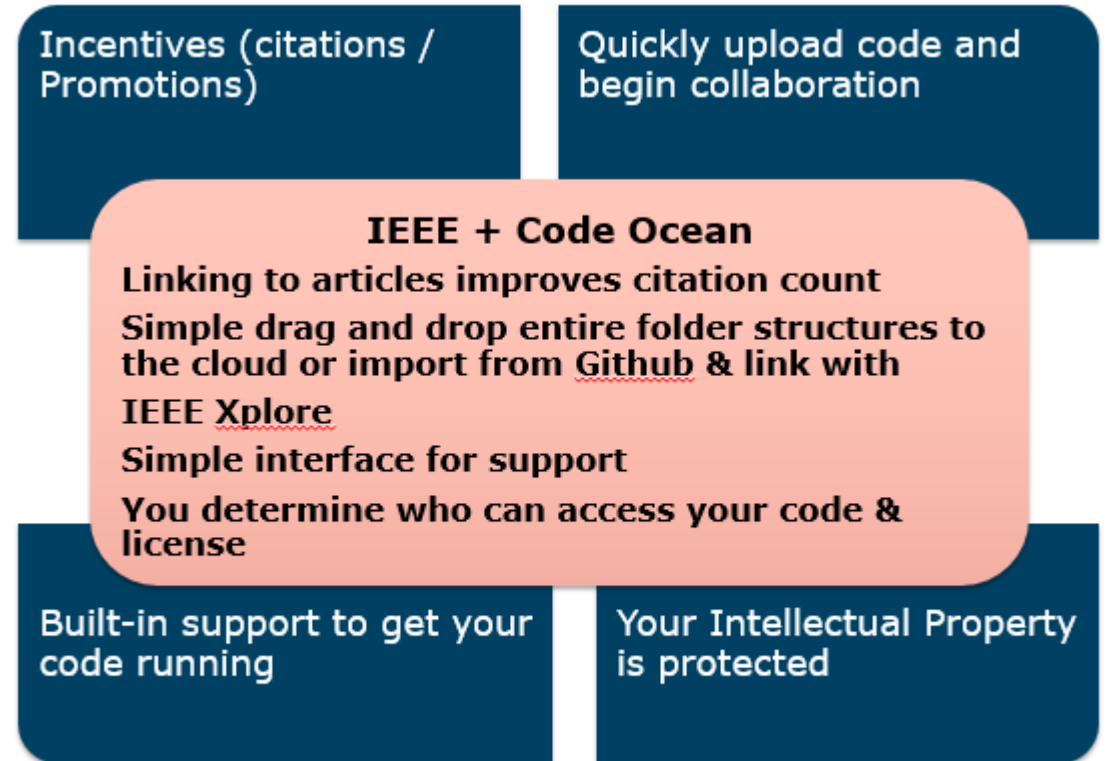
Nature - May 29 2017

Software Simplified

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

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

- Supports Reproducible Research
  - *"A minimal standard for data analysis and other scientific computations is that they be **reproducible**: that the code and data are assembled in a way so that another group can <u>re-create all of the results (e.g., the figures in a paper)</u>"*
    - *Source: Prof Karl Broman, University of Wisconsin- Madison*
- Enhances Author Visibility
  - *Algorithm can be cited and referenced just like papers – get their own DoI*
  - *Enhances Impact Factor for a Journal*
- Reduces Customer Support burden on an Author
  - *Eliminates need for Author to provide support to get the code running on another platform*
- Terms of Use
  - *Author owns the algorithm and assigns license.*
- Collaboration
  - *Allows the author to collaborate with his students or colleagues in developing the algorithm along with the paper.*

Incentives (citations / Promotions)

Quickly upload code and begin collaboration

IEEE + Code Ocean
Linking to articles improves citation count
Simple drag and drop entire folder structures to the cloud or import from Github & link with IEEE Xplore
Simple interface for support
You determine who can access your code & license

Built-in support to get your code running

Your Intellectual Property is protected

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

Reproducible Research
Code Sharing

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?

COPUBLISHED BY THE IEEE CS AND THE AIP

PATRICK VANDEWALLE

COMPUTING IN SCIENCE & ENGINEERING

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

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

It is important to IEEE authors

- Increases:
    - Citations by 3X
    - Downloading
    - Collaboration

Demo

For Questions, Comments and Feedback please contact:

David J Goldstein
Lead Director
D.Goldstein@IEEE.Org

Dr. Sunil Gupta
Director
S.Gupta@IEEE.Org

Naveen Maddali
Senior Project Manager
N.Maddali@IEEE.Org