

Uppsala university Junior Faculty Seminar

Leadership in Open Science:

from publishing reproducible research to
opening up educational resources

 @LorenaABarba <http://lorenabarba.com>



About me

- ▶ Reproducibility PI Manifesto
figshare, 2012
- ▶ “The hard road to reproducibility”
Science, Oct. 2016
- ▶ “Repro Packs,” *Nature blogs*, Apr. 2017
- ▶ CiSE editor for Reproducible Research
- ▶ SC19 Reproducibility Chair
- ▶ NASEM Committee member

Lorena A. Barba group

Reproducibility PI Manifesto



<http://lorenabarba.com>

Reproducibility PI Manifesto (2012)

- I teach my graduate students about reproducibility
- All our research code (and writing) is under version control
- We always carry out verification & validation (and make them public)
- For main results, we share data, plotting script & figure under CC-BY
- We upload preprint to arXiv at the time of submission to a journal
- We release code at the time of submission of a paper to a journal
- We add a “Reproducibility” declaration at the end of each paper
- I develop a consistent open-science policy & keep an up-to-date web presence

WORKING LIFE

By Lorena A. Barba

The hard road to reproducibility

Early in my Ph.D. studies, my supervisor assigned me the task of running computer code written by a previous student who was graduated and gone. It was hell.



“My students and I continuously discuss and perfect our standards.”

NATUREJOBS | NATUREJOBS BLOG

TechBlog: My digital toolbox: Lorena Barba

17 Apr 2017 | 12:00 BST | Posted by [Jeffrey Perkel](#) | Category: [Blog](#), [Technology](#)

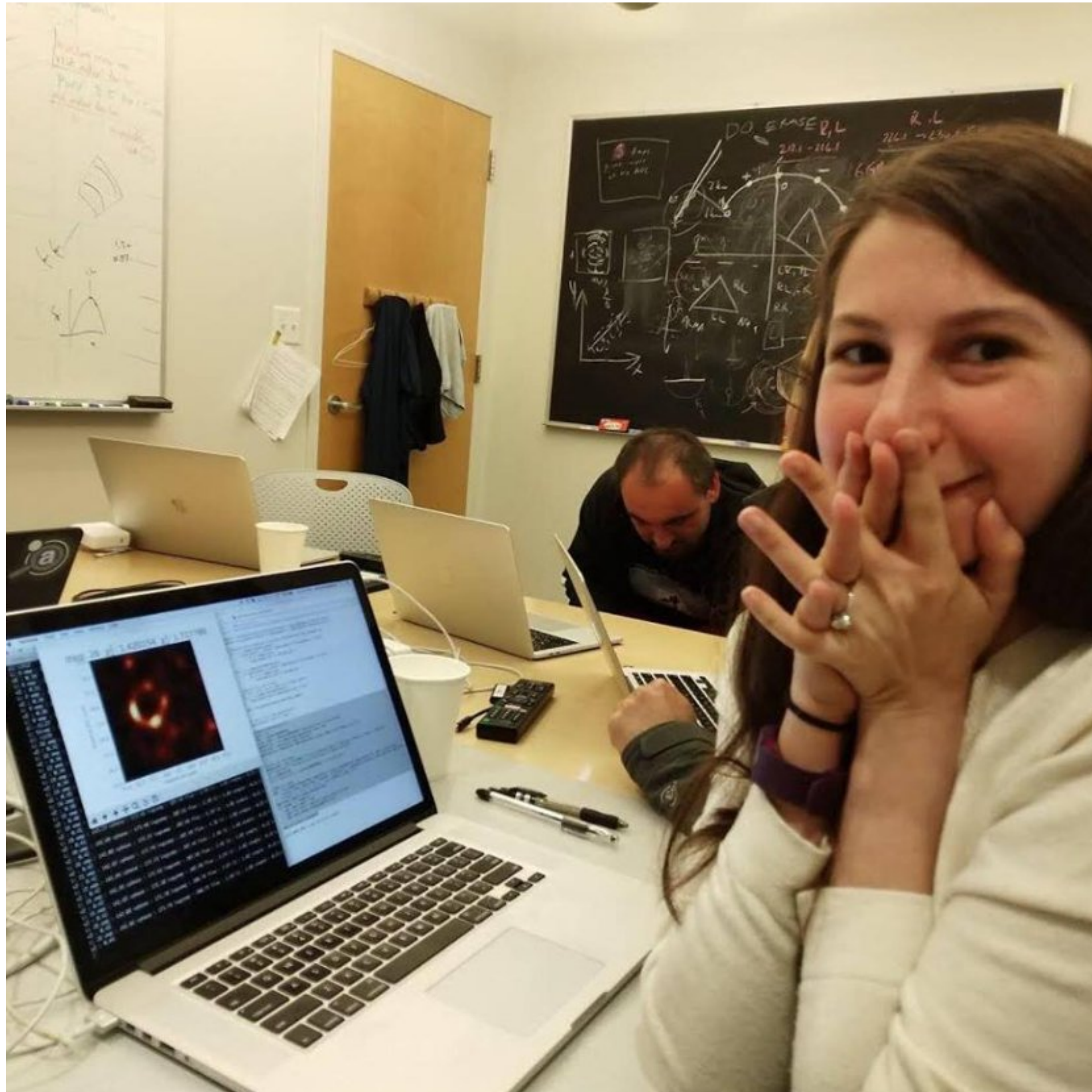
Repro-Packs: our signature open-science practice

Reproducibility and Replicability in Science



- ▶ Study mandated by public law 114-329 (Jan. 2017)
- ▶ commissioned by the National Science Foundation (NSF) to The National Academies of Sciences, Engineering and Medicine (NASEM)
- ▶ 15 experts convened
- ▶ 18 months of in-person meetings, teleconferences, commissioned papers, deliberations, writing
- ▶ report released 7 May 2019

Widespread use of computation & data in science



- ▶ As important as the telescopes were the software libraries and data products needed to create the first image of a black hole

(now iconic photo of Dr. Katie Bouman)

“reproducibility . . . requires having the complete software environment [...] and the full source code available for inspection, modification, and application under varied parameter settings.”

—Buckheit and Donoho (1995)

Top challenges of reproducible research

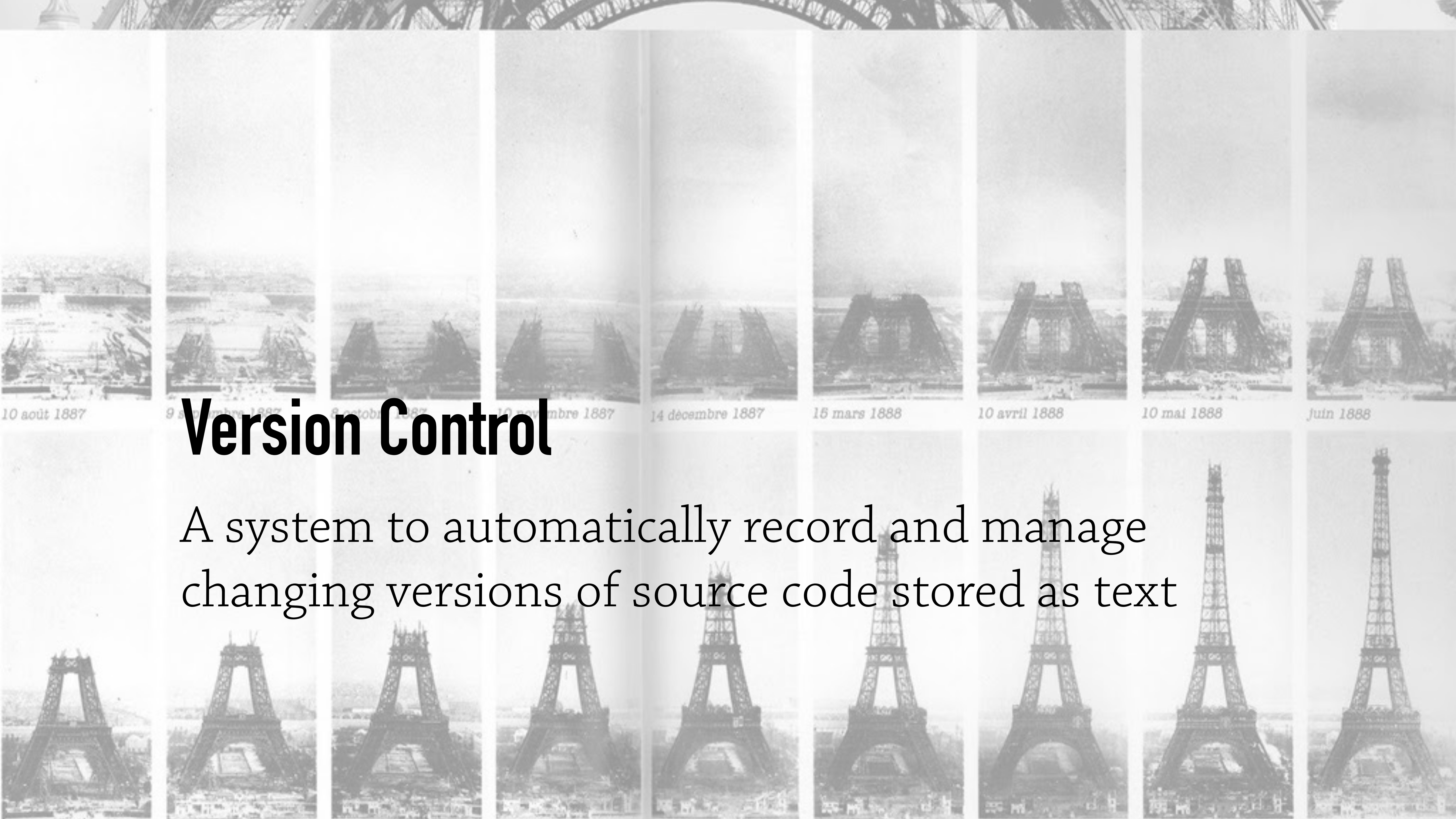
- ▶ creation, curation, usage and publication of research software
- ▶ acceptance, adoption and standardization of open-science practices;
- ▶ misalignment with academic incentive structures and institutional processes for career progression

Improving reproducibility

- ▶ Automatic capture of computational details; workflow management systems
- ▶ Source code and data version control
- ▶ Tools for reproducing results via virtualization, cloud computing, packaging, containers (e.g., Docker, Singularity)
- ▶ Interactive computational notebooks (e.g., Jupyter)

Step 1: Reproducible workflows

- ▶ Version control
- ▶ Script, automate, document
- ▶ Avoid GUIs for manipulating figures



10 août 1887

9 septembre 1887

8 octobre 1887

10 novembre 1887

14 décembre 1887

15 mars 1888

10 avril 1888

10 mai 1888

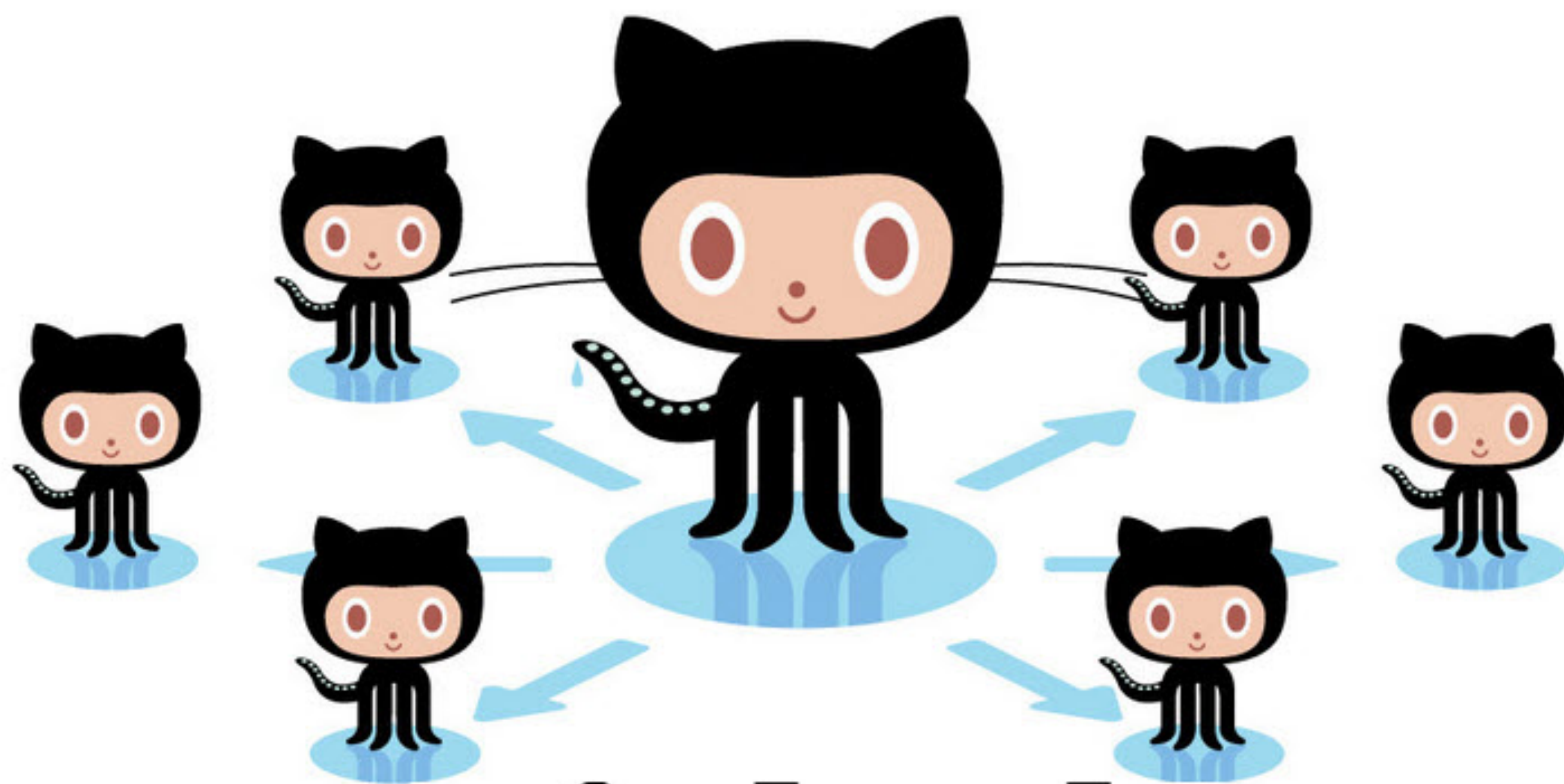
juin 1888

Version Control

A system to automatically record and manage changing versions of source code stored as text

We use version control:

- ▶ internal reports on Markdown or Jupyter
- ▶ manuscripts in LaTeX



github
SOCIAL CODING

Open source as a development model

Linus's Law — “Given enough eyeballs, all bugs are shallow.”



A set of open-source tools
for **interactive** and
exploratory computing.

Step 2: Publish the software

- ▶ *“We’re not a discipline, until we value software”*

—L. Barba at 2015 SIAM Conference on Computational Science and Engineering (CSE) panel “The Future of CSE as a Discipline”



The Journal of Open Source Software is a **developer friendly**, open access journal for research software packages.

Committed to publishing quality research software with zero article processing charges or subscription fees.

[Submit a paper to JOSS](#)[🎉 Volunteer to review](#)[📖 Explore Papers](#)[📄 Documentation](#)[ℹ️ Learn More](#)

Recently Published Papers 690

PUBLISHED Published 2 days ago

GoFish: Fishing for Line Observations in Protoplanetary Disks

Python

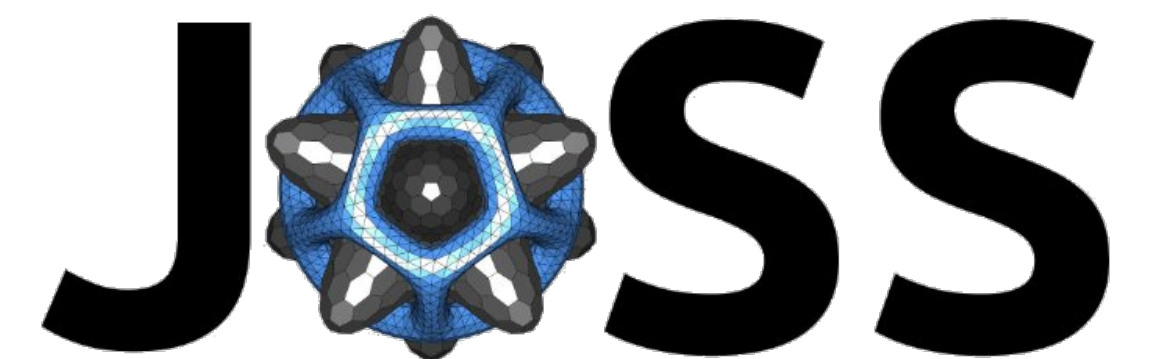


[@richteague](#)

DOI [10.21105/joss.01632](https://doi.org/10.21105/joss.01632)

JOSS infrastructure

- ▶ GitHub—open-source software hosting & collaboration
- ▶ Zenodo—data repository by CERN
- ▶ ORCID—author identification
- ▶ CrossRef—DOI minting
- ▶ custom web app and Ruby bot





Open-Source Software (OSS)

Reproducible research is vitally connected to open-source software, open data and open science.

Be aware:

... just because the source code is available on a website, doesn't mean it is open source!



Standard public licenses

It's not sufficient to make the source public to read. We must attach a **license** that allows others to modify and distribute the code.



Open-source licenses:

Anyone developing software in an academic setting should have working knowledge of software licenses.

Education

A Quick Guide to Software Licensing for the Scientist-Programmer

Andrew Morin¹, Jennifer Urban², Piotr Sliz^{1*}

1 Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts, United States of America, **2** Samuelson Law, Technology & Public Policy Clinic, School of Law, University of California Berkeley, Berkeley, California, United States of America

How to choose?

For academic work: simple & permissive is best.

—BSD3 for code; CC-BY for content

<http://choosealicense.com/>

Step 3: Open data / Open science

- ▶ Archive interim data products (e.g., meshes)
- ▶ Share input files, configuration, parameter lists, runtime options
- ▶ Archive secondary data, figures, and plotting scripts (“repro-packs”)

Good data management

FAIR Principles: digital artifacts of research should be Findable, Accessible, Interoperable and Reusable for machines and for people

—Wilkinson et al., 2016.

Data repositories

- ▶ must provide a unique global identifier for your data (typically a digital object identifier, DOI)
- ▶ must offer long-term preservation guarantees (at least 10 years)

Free data repositories:





- ▶ general-purpose repository for all kinds of digital artifacts of research
- ▶ any file format, up to 5GB in size
- ▶ free and unlimited for public items

zenodo

- ▶ created by CERN and OpenAIRE
- ▶ free and non-commercial
- ▶ log in with your ORCID
- ▶ deposit large files: up to 50GB by default

<https://zenodo.org/communities/barbagroup/>

Open-access publishing

- ▶ Yale Law School Roundtable on Data and Code Sharing (2009) recommended publishing under open-access conditions (or post pre-prints).

Preprints

- ▶ In physics, math, CS... arXiv is a way of life
- ▶ Preprints growing by all metrics
- ▶ Explosion of 'Xiv sites



Nature Publishing Group	Compatible	Communication between researchers includes not only conferences but also preprint servers. The ArXiv preprint server is the medium of choice for (mainly) physicists and astronomers who wish to share drafts of their papers with their colleagues, and with anyone else with sufficient time and knowledge to navigate it. [...] If scientists wish to display drafts of their research papers	[9] and [10]
IOP Publishing	Compatible	You may use an IOP copyright wordmark on IOP Publications. It. The submitted article may be uploaded to the Author Centre	[11]
Oxford Journals	Compatible	"Prior to publication on the subject of this article. This article is available in their journals as follows:	[12]
Elsevier	Compatible	Elsevier anywhere. Identification note the Inform	[13]
Springer, incl. SpringerOpen Journals and BioMed Central (BMC)	Compatible	Postin Centre	[14][15][16]
Taylor & Francis	Compatible	"This is your original manuscript (often called a "preprint"), and you can share this as much as you like. If you do decide to post it anywhere, including onto an academic networking site, we would recommend you use an amended version of the wording below to encourage usage and citation of your final, published article."	[17]

Most journals accept manuscripts previously submitted to preprint servers

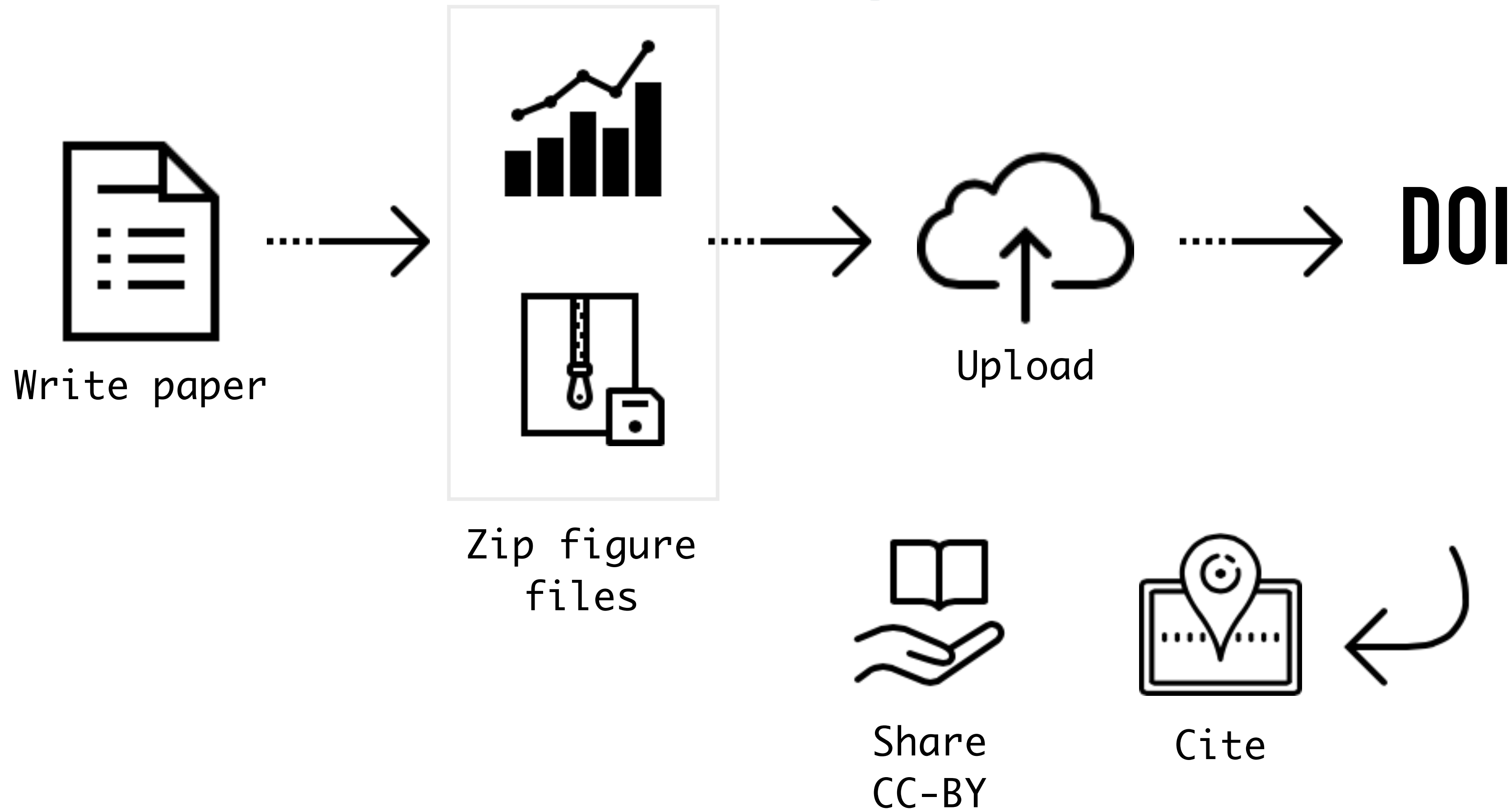
List of academic journals by preprint policy, Wikipedia

ReproPacks

- ▶ For main results in a paper, we share data, plotting script & figure under CC-BY.
- ▶ Deposit the file bundle as a Figshare object and get a DOI
- ▶ We cite this DOI in the figure caption!

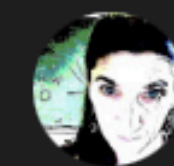


Our workflow:



Step 4: Teach in the open

- ▶ Jupyter for teaching: go.gwu.edu/jupyter4edu
- ▶ Version-control, again!
- ▶ Publish learning objects—digital materials can be made to be *reused*.



ME 702 — Introduction

Computational Fluid Dynamics, CFD

What is it?

Collection of Engineering

0:00 / CC Settings Full Screen Comment Like

ME 702 - Computational Fluid Dynamics

Lorena Barba - 1 / 32



▶ 33:26

ME 702 - Computational Fluid Dynamics - Video Lesson 1

Boston University

2 12:48

ME 702 - Computational Fluid Dynamics (Lecture "zero", par...

Boston University

3 32:57

ME 702 - Computational Fluid Dynamics (Lecture "zero", par...

Boston University

4

ME 702 - Computational Fluid Dynamics (Lecture "zero", par...

Added views ~1,053,921

ME 702 - Computational Fluid Dynamics - Video Lesson 1

163,770 views Jan 22, 2012 NEW! (August 2014) Prof. Barba is teaching a MOOC titled "Practical Numerical Methods with Python." Check it out: ...more

981 Dislike Share Clip Save ...

What did OER miss from FOSS?

- ▶ developing in the open
- ▶ collaborating/contributing
- ▶ community around OS projects
- ▶ culture & value-based framework

Openness is about the possibilities of communicating with other people. It's not about *stuff*, what you do with stuff. It's about what you do with each other

— Stephen Downes, 2017

<https://youtu.be/FPHYAFcUziA>

The Journal of Open Source Education

An **educator-friendly** journal for publishing computational learning modules and educational software.

Notice: We are not yet accepting submissions. Please follow our development on [GitHub](#) or [Twitter](#).

[Volunteer to review for JOSE!](#)

[Learn more »](#)

Uppsala university Junior Faculty Seminar

Leadership in Open Science:

from publishing reproducible research to
opening up educational resources

 @LorenaABarba <http://lorenabarba.com>

