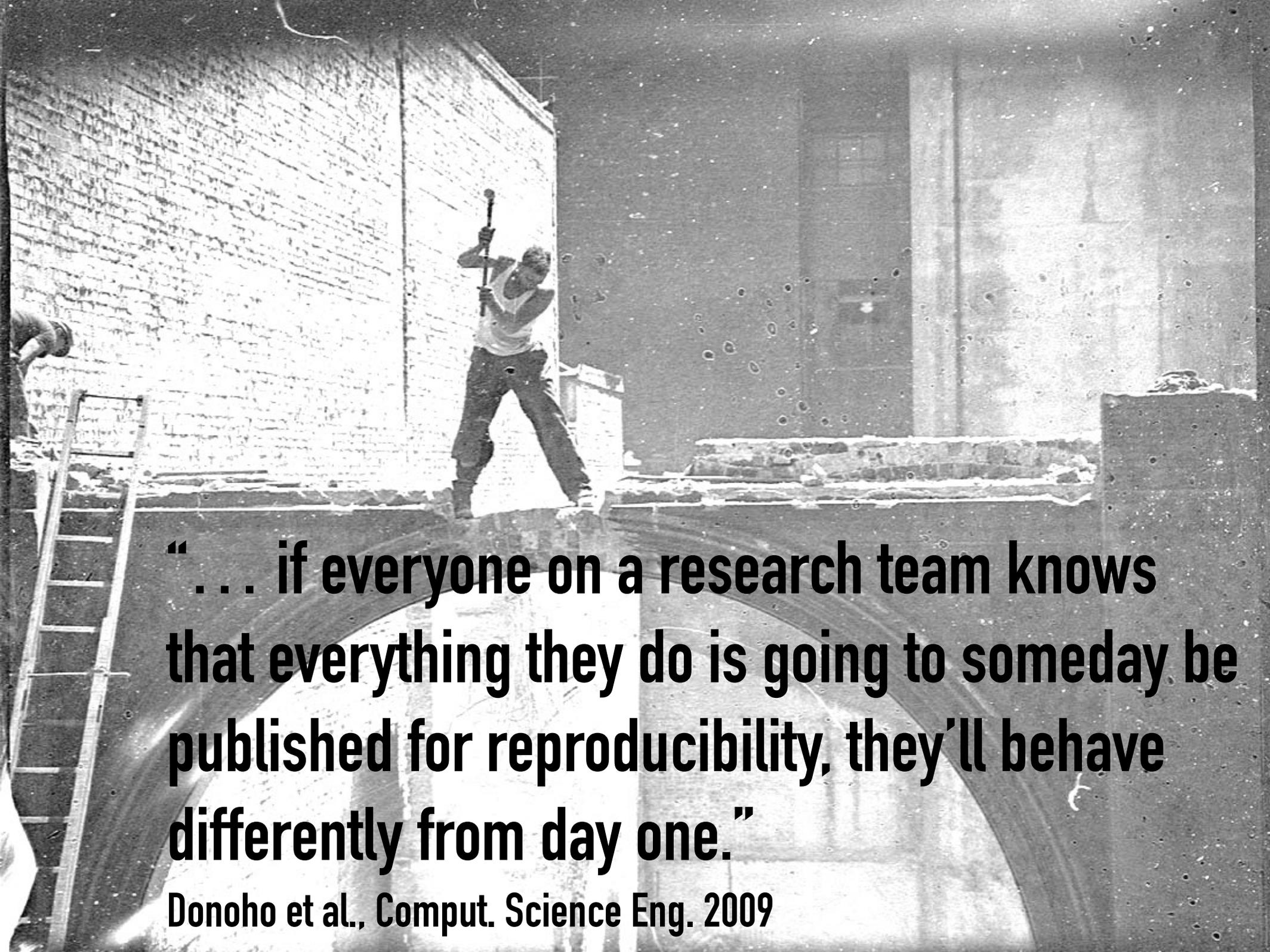


NSF SI2 PI Workshop 2017

How to run a lab for reproducible research

 @LorenaABarba





“... if everyone on a research team knows that everything they do is going to someday be published for reproducibility, they’ll behave differently from day one.”

Donoho et al., Comput. Science Eng. 2009

Action items

1. Commitment
2. Transparency & Open Science
3. Onboarding
4. Collaboration
5. Community & Leadership





Commitment



Jon F. Claerbout

Professor Emeritus of Geophysics
Stanford University

... pioneered the use of computers
in processing and filtering seismic
exploration data [Wikipedia]

... from 1991, he required theses
to conform to a standard of
reproducibility.

Def.— Reproducible research

Authors provide all the necessary data and the computer codes to run the analysis again, re-creating the results.

Schwab, M., Karrenbach, N., Claerbout, J. (2000) “Making scientific computations reproducible,” *Computing in Science and Engineering* Vol. 2(6):61–67



Lorena A. Barba group

Reproducibility PI Manifesto



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

Not everyone agrees

Two points of
contention:

- scripted figures (vs. GUI-based tools)
- version control



“I’ve learned that interactive programs are slavery
(unless they include the ability to arrive in any
previous state by means of a script).”

— Jon Claerbout





Shocking Paper Claims That Microsoft Excel Coding Error Is Behind The Reinhart-Rogoff Study On Debt

Mike Konczal, NewDeal2.0 

🕒 Apr. 16, 2013, 12:40 PM 🔥 92,101

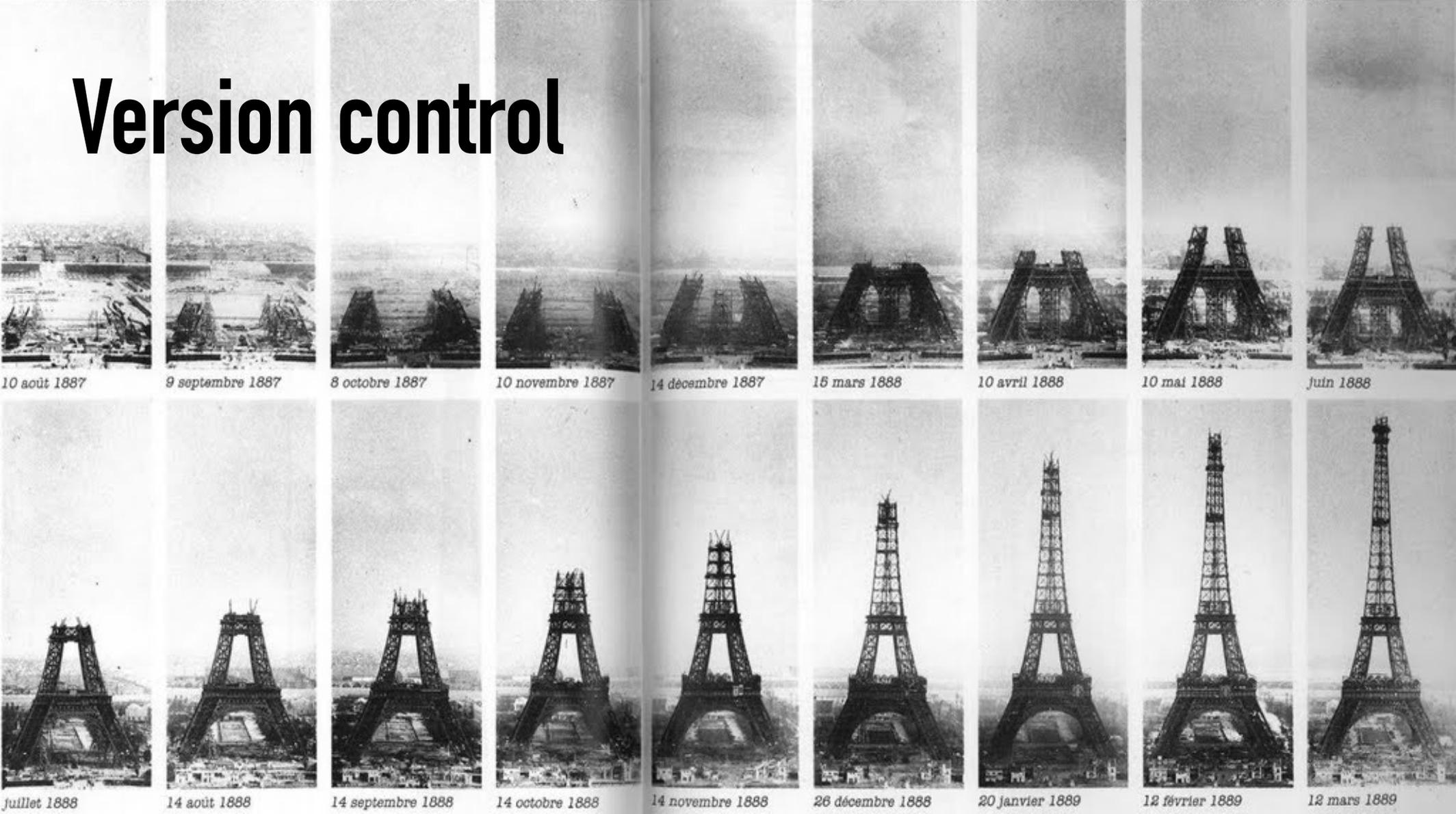
THE WALL STREET JOURNAL.

REAL TIME ECONOMICS

Reinhart, Rogoff Admit Excel Mistake, Rebut Other Critiques



Version control



“private reproducibility”

...we can rebuild our own past research results from the precise version of the code that was used to create them.

What is Science?

▶ American Physical Society:

- Ethics and Values, 1999

"The success and credibility of science are anchored in the willingness of scientists to [...] Expose their ideas and results to **independent testing and replication by others**. This requires the open exchange of data, procedures and materials."



Transparency & Open Science



Donoho, D. et al. (2009) “Reproducible research in computational harmonic analysis,” *Computing in Science and Engineering* Vol. 11(1):8–18.

Data and Code Sharing **Recommendations**

- ▶ assign a unique identifier to every version of the data and code
- ▶ describe in each publication the computing environment used
- ▶ use open licenses and non-proprietary formats
- ▶ publish under open-access conditions (and/or post pre-prints)



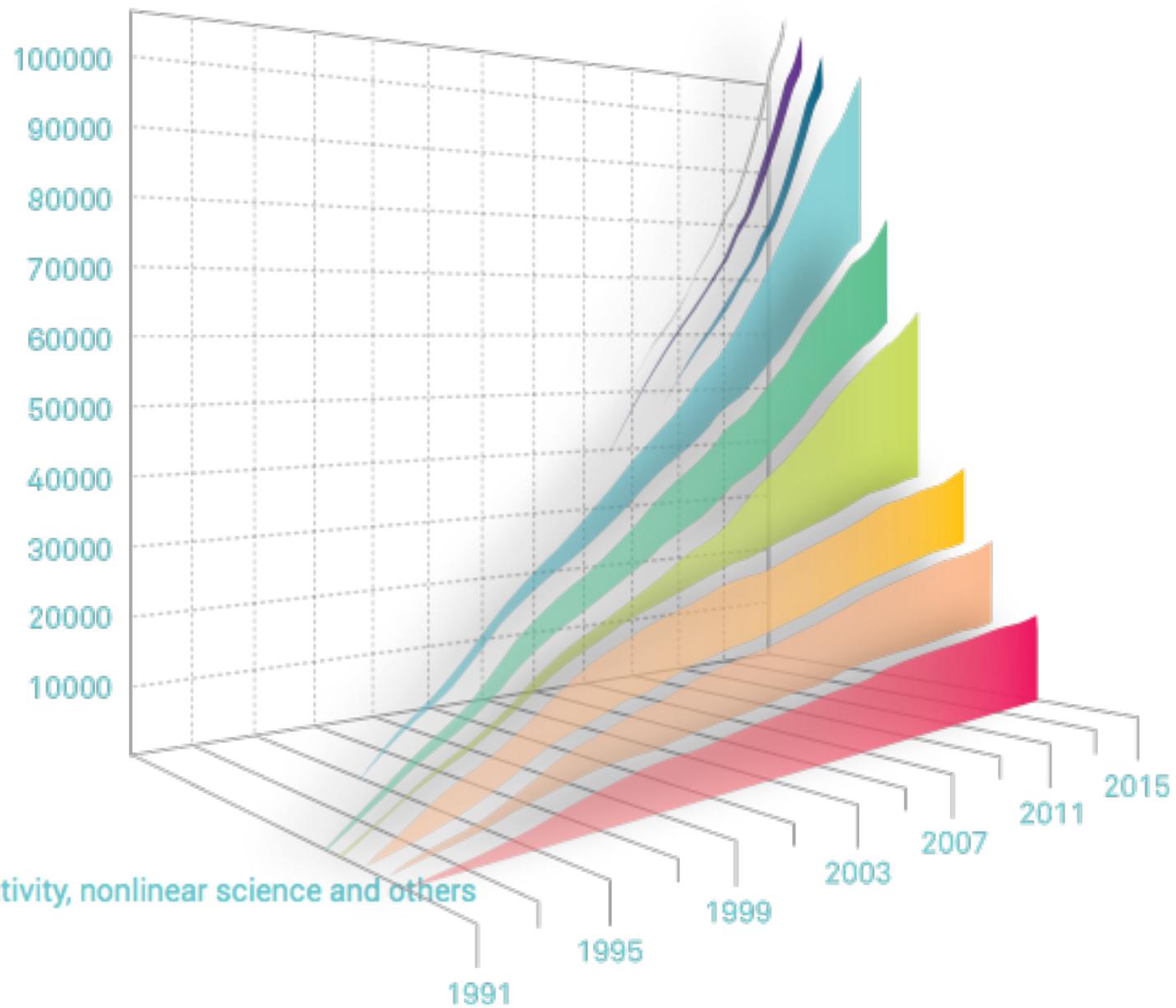
Open-source licenses:

People can coordinate their work freely, within the confines of copyright law, while making access and wide distribution a priority.



arXiv

- finance
- biology
- statistics
- computer science
- physics, including general relativity, nonlinear science and others
- mathematics
- high energy physics
- condensed matter physics
- astrophysics





Opinion: Reproducible research can still be wrong: Adopting a prevention approach

Jeffrey T. Leek^{a,1} and Roger D. Peng^b

^aAssociate Professor of Biostatistics and Oncology and ^bAssociate Professor of Biostatistics,
Johns Hopkins University, Baltimore, MD

PNAS | February 10, 2015 | vol. 112 | no. 6 | 1645–1646

<http://dx.doi.org/10.1073/pnas.1421412111>

“The key is prevention via the training of more people on techniques for data analysis and reproducible research.”

Onboarding





Lorena A Barba

Engineering professor, computational scientist, jazz buff, techie, mac fan, academic writer and...

Oct 30, 2016 · 10 min read



Lockheed P-80A airplane (1946). Credit: NASA Commons. — A reminder to test your code.

Barba-group reproducibility syllabus

<https://medium.com/@Lorenaabarba>

A syllabus for research computing

1. command line utilities in Unix/Linux
2. an open-source scientific software ecosystem (our favorite is Python's)
3. software version control (we like the distributed kind: our favorite is git / GitHub)
4. good practices for scientific software development: code hygiene and testing
5. knowledge of licensing options for sharing software

https://barbagroup.github.io/essential_skills_RRC/



Collaboration

<https://github.com/barbagroup>



This repository

Search

Pull requests

Issues

Gist



barbagroup / essential_skills_RRC

Unwatch 5

Star 9

Fork 2

Code

Issues 3

Pull requests 1

Projects 0

Wiki

Pulse

Graphs

Settings

Branch: master

Commits on Jan 16, 2017



Update markdown version of notebooks

gforsyth committed on Jan 16 ✓



534878c



Merge branch 'master' of github.com:barbagroup/essential_skills_RRC

ncclementi committed on Jan 16 ✓



f90cab7



Change keyword arg in notebook plots

ncclementi committed on Jan 16



31ec399



Commits on Jan 6, 2017



Merge pull request #47 from barbagroup/github2 ...

ncclementi committed on GitHub on Jan 6 ✓



30707fd



Commits on Jan 5, 2017



Add course links

labarba committed on GitHub on Jan 5 ✓



f7d7c93



Add instructions on pushing new branches to GitHub

gforsyth committed on Jan 5 ✓



22b700b



Merge pull request #45 from barbagroup/github2 ...

gforsyth committed on GitHub on Jan 5 ✓



bfa9599



Community & Leadership

... clustering of similar areas of interest that allows for interaction, sharing, dialoguing, and thinking together

—George Siemens, 2004



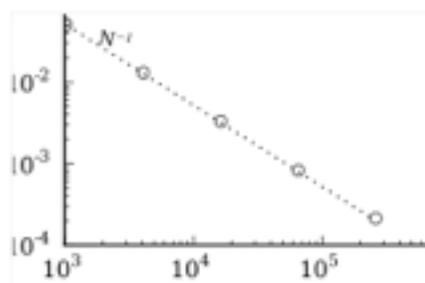


Figure5.pdf (15.72 kB)

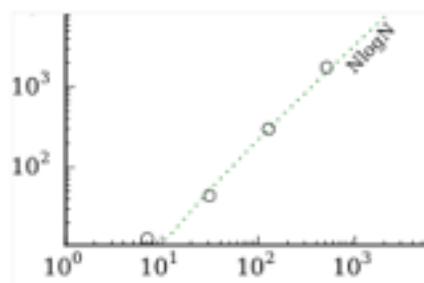


Figure6.pdf (11.83 kB)

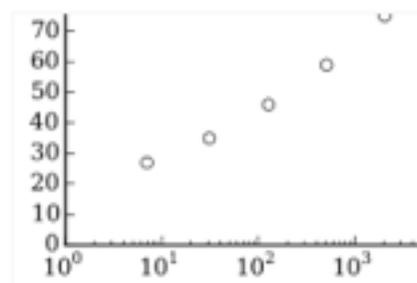


Figure7.pdf (11.43 kB)



PyGBeOrientatio... .zip (12.06 MB)



README (1.39 kB)

[Download all \(24.18 MB\)](#)
[Share](#)
[Cite](#)
[Embed](#)
[+ Collect](#)

5 files



Grid convergence of PyGBe with a spherical molecule near spherical surface

Version 4 ▼ 29.01.2016, 08:56 (GMT) by Christopher D. Cooper, [Lorena A. Barba](#)

 314
views

 40
downloads


Reproducibility package containing data, running script, plotting script and final plot of grid-convergence study for a spherical molecule near a charged surface.

The running script invokes the open-source bioelectrostatics solver PyGBe with the

ReproPacks

For main results in a paper, we share data, plotting script & figure under CC-BY.

File bundle with input data, running scripts, plotting scripts, and figure.

We cite our own figure in the caption!



NSF SI2 PI Workshop 2017

How to run a lab for reproducible research

 @LorenaABarba

