CODE CHECK: independent reproduction of computations underlying biomedical research

```
Stephen J Eglen                  Cambridge Computational Biology Institute
https://sje30.github.io          University of Cambridge
sje30@cam.ac.uk                  @StephenEglen
```

Slides: http://bit.ly/eglen2019-ati (CC-BY license)

Acknowledgements

Joint work with Daniel Nüst (Muenster).

(Last update Sat Feb 9 21:40:26 2019)

The reproducibility crisis

Many key findings in publications are either not independently verified, or fail verification when it is attempted Baker, 2016.

Never mind the headline fraud cases...

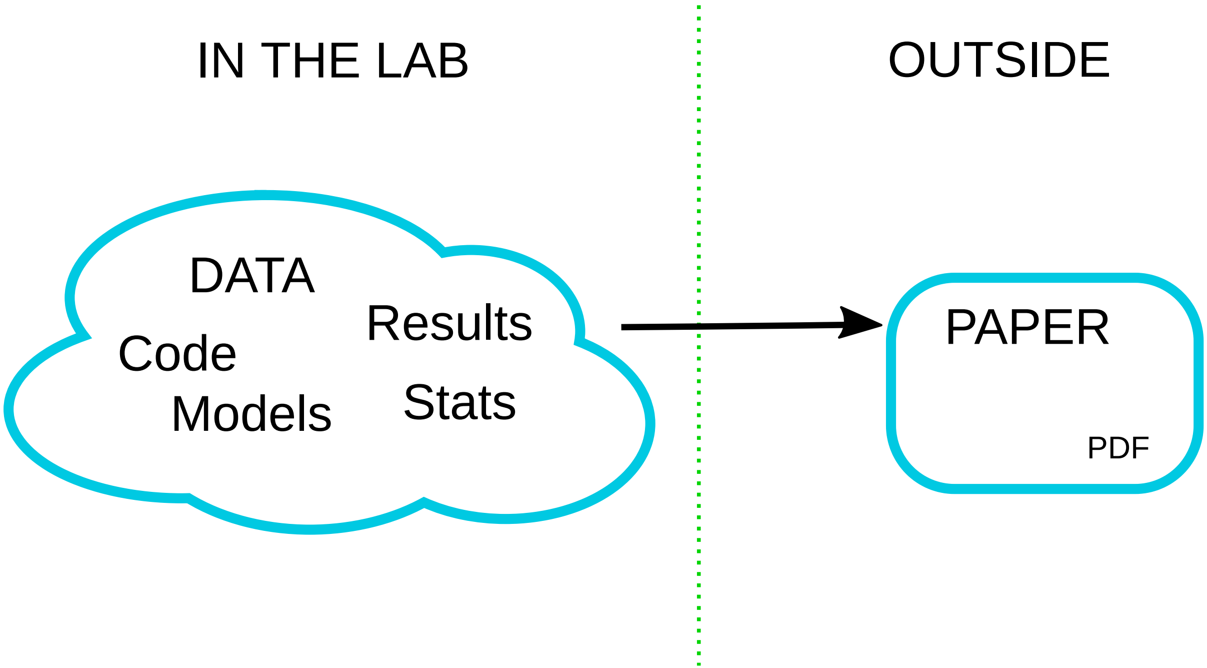
Data goes missing quite quickly (Vines et al 2014)] (http://www.ncbi.nlm.nih.gov/pubmed/24361065)

A systematic study examining 300 papers published in 2016/17 in Journal of Computational Physics, *a journal that promotes sharing of digital artefacts*, found that only 5.6% made artefacts available Stodden et al 2018.

Who's got the time to verify results?

Disclaimer: do I mean "reproducibility" or "replicability"? (Barba 2018) https://arxiv.org/pdf/1802.03311.pdf

Science as an inverse problem



Reproducible papers

- "Reproducible papers" are slowly gaining in popularity. People providing data and code.
- Journals are exploring systems like Code Ocean and BinderHubs.
- These systems are good, and when they work are super. BUT...

Difficulties with current approaches.

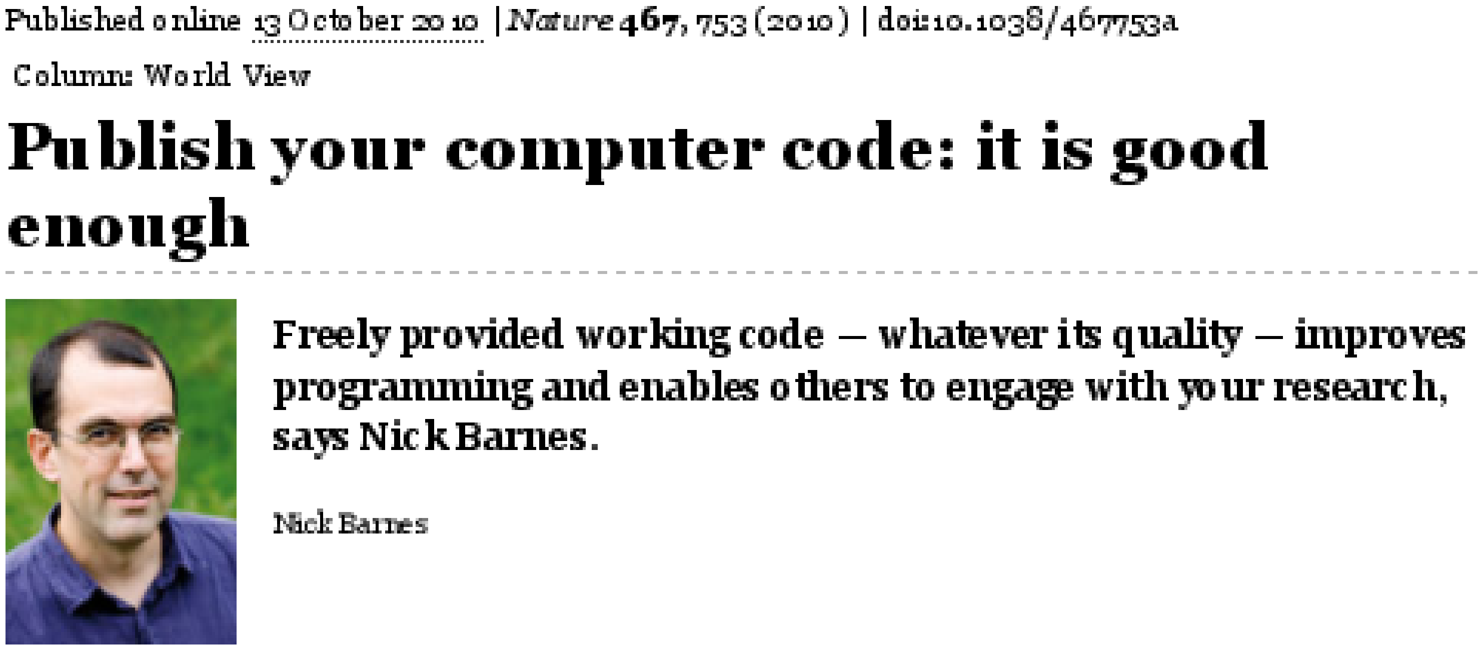
- Who pays for mybinder instances?

- "Reproducible forever" is a BIG challenge. My first attempt lasted about 6 months. Dockerhub for waverepo. Is forever useful?

- Interactive approach suitable for shorter jobs. But how about for long jobs (days--months?)

- Do editors / reviewers have time to check that code and data work? I can recall only one comment (positive!) about running my code by a reviewer.

Barnes 2010 convinced me that something is better than nothing.



Published online 13 October 2010 | *Nature* **467**, 753 (2010) | doi:10.1038/467753a

Column: World View

Publish your computer code: it is good enough

Freely provided working code — whatever its quality — improves programming and enables others to engage with your research, says Nick Barnes.

Nick Barnes

Our thoughts

- Having code and data archived w/ paper already a win.
- Dockerfile helps you understand the environment, even if it's broke.
- Having an independent assessment that the code ran *once* in an indepedent environment is useful knowledge for a reviewer.
- Proposal to Wellcome Open Research Fund Summer 2018 failed, but brought Daniel and Stephen together.

CODE CHECK

- http://www.codecheck.org.uk
- CODE CHECK will be a system for generating "Certificates of Computation"
- These state that code was run on a given date on a given platform and generated particular results (figures, tables).
- Output log files are archived with the results.
- Relevant outputs archived on zenodo with permanent URL.

Example certificate

https://sje30.github.io/codecheck/eglen2016/eglen2016-crc.html

How much should be shared?

- What is practical to ask people to share?
- We follow the suggestion from modelDB that at least one key figure / table / result should be replicable.
- Toy examples welcome; long jobs might be too difficult.
- Who checks the figures? No-one. It stands alone. Reviewers could scan them.

Challenges to tackle

- How automated can we get? repo2docker tool could be adapted.
- Provide training examples for people to copy.
- Compute infrastructure for long jobs.

Target journals to work with

- Neurons, Behaviour, Data, Theory
- Gigascience
- ELife

Next steps

- Code papers? "Scientific Code"?
- Work with mybinder for archival?
- Relationship to British Library