CODECHECK: An open-science initiative to facilitate sharing of computer programs and results presented in scientific publications.

Stephen J Eglen
https://sje30.github.io
sje30@cam.ac.uk

Cambridge Computational Biology Institute
University of Cambridge
@StephenEglen

Daniel Nüst
https://nordholmen.net
daniel.nuest@uni-muenster.de

Institute for Geoinformatics
University of Münster
@nordholmen

HTML Slides: http://bit.ly/eglen-orwg2 (CC-BY 4.0 license)

Declarations

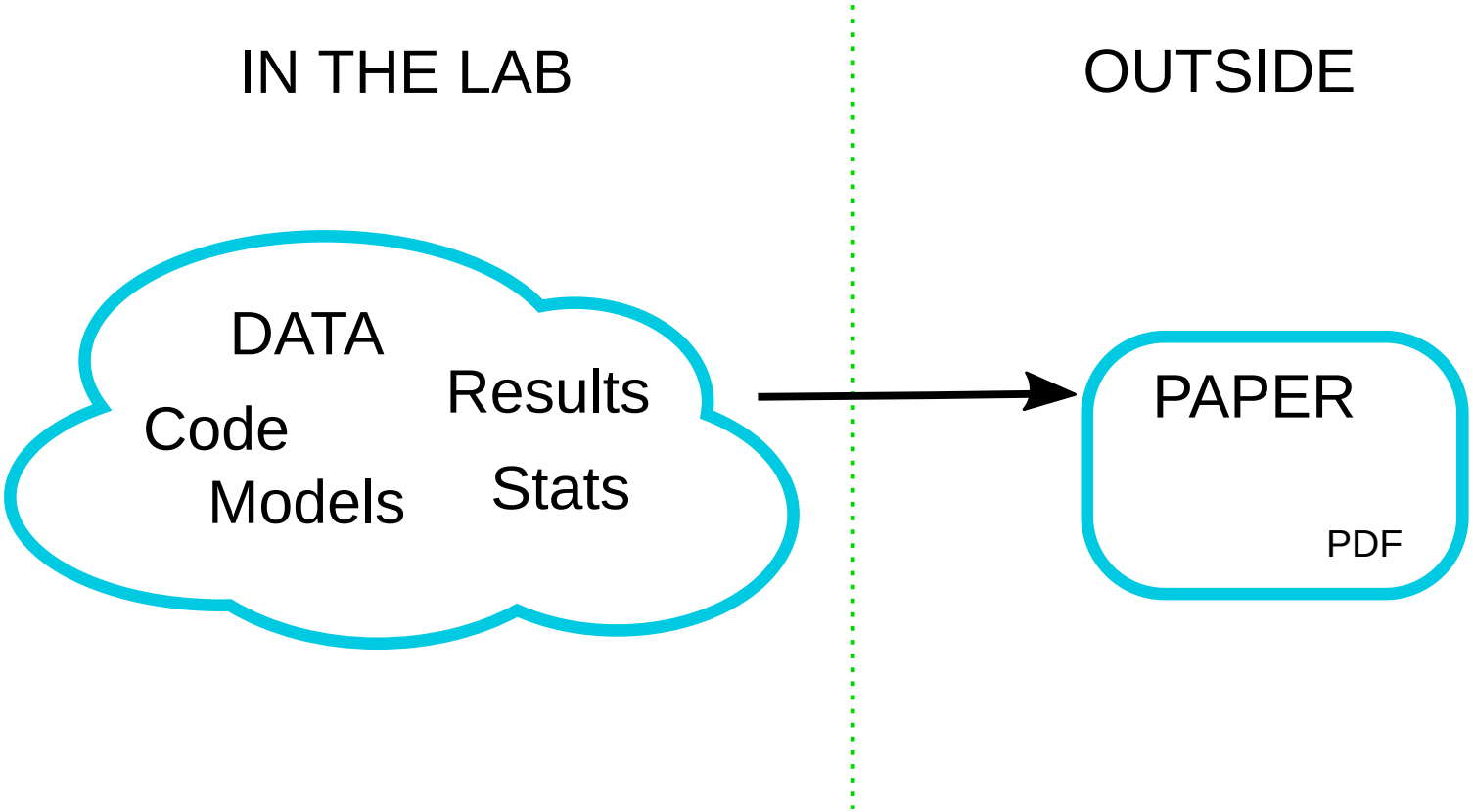
Affiliate editor of *bioRxiv*. Senior editor of *Scientific Data*.

Acknowledgements

Mozilla mini science grant, UK Software Sustainability Institute.

Editors @ *Gigascience*, *eLife*, *Scientific Data*.

Premise



We should be sharing material on the left, not the right. "Paper as advert for Scholarship" Buckheit & Donoho (1995)

Approaches to code sharing

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

Column: World View

Publish your computer code: it is good enough

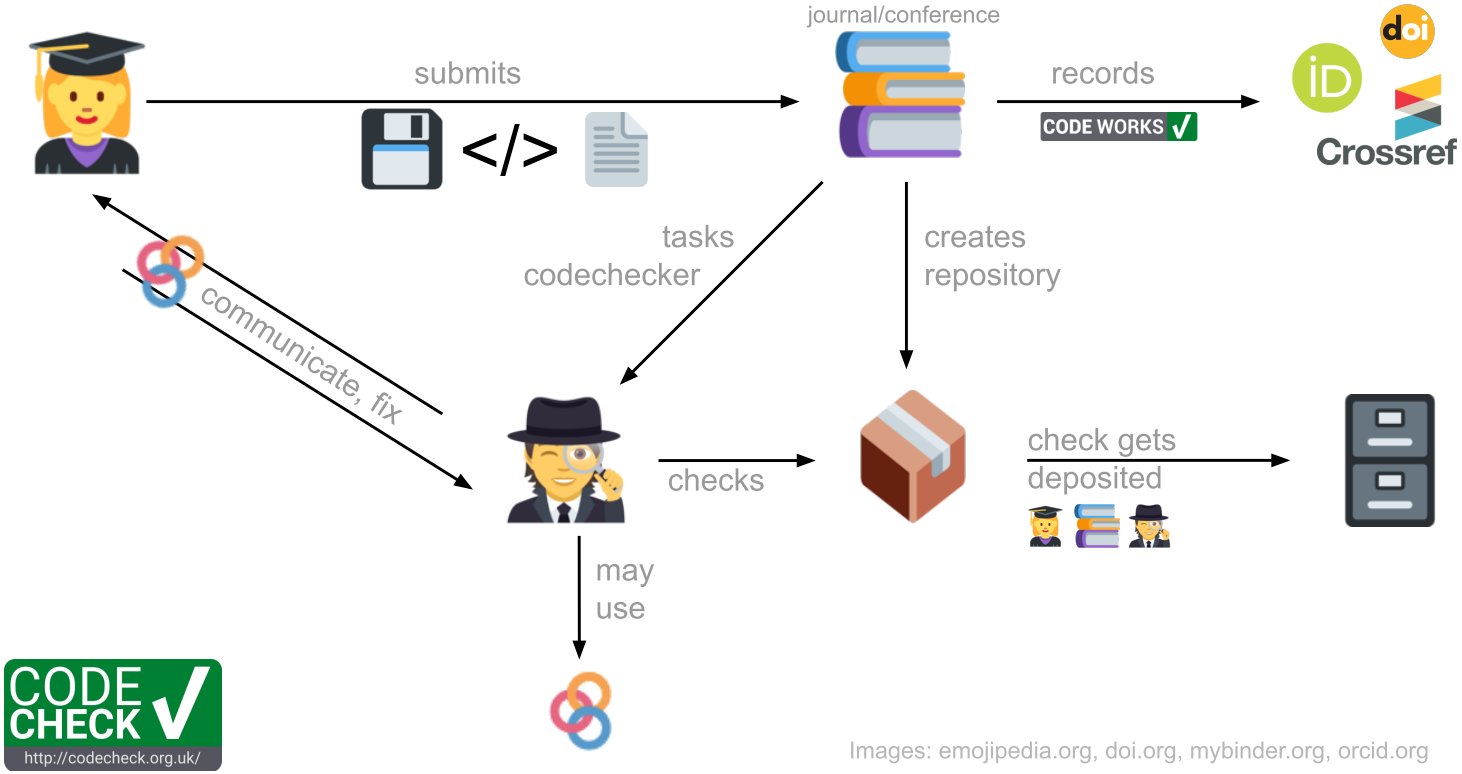
Freely provided working code — whatever its quality — improves programming and enables others to engage with your research, says Nick Barnes.

Nick Barnes

- Informal 'code buddy' system
- Community-led *research compedia*.
- Code Ocean (Nature trial)
- Certify reproducibility with confidential data (CASCAD) (Pérignon et al 2019)

The CODECHECK philosophy

- Systems like Code Ocean set the bar high by "making code reproducible *forever* for *everyone*".
- CODECHECK simply asks "was the code reproducible *once* for *someone* else?"
- We check the code runs and generates the expected number of output files.
- The contents of those output files are not checked, but are available for others to see.
- The validity of the code is *not* checked.



Four principles underlying CODECHECK

1. CODECHECKERS are **humans** and communication is key.
2. CODECHECKERS **record** but don't investigate or fix.
3. **Credit** is given to CODECHECKERS.
4. Workflows are scripted, **auditable**, and they work (or *worked once*).

Who does the work?

1. **AUTHOR** provides code/data and instructions on how to run.
2. **CODECHECKER** runs code and writes certificate.
3. **PUBLISHER** oversees process, helps depositing artifacts, and persistently publishes certificate.

Who benefits?

1. **AUTHOR** gets early check that "code works"; gets snapshot of code archived and increased trust in stability of results.
2. **CODECHECKER** gets insight in latest research and methods, credit from community, and citable object.
3. **PUBLISHER** Gets citable certificate with code/data bundle to share and increases reputation of published articles.
4. **PEER REVIEWERS** can see certificate rather than check code themselves.
5. **READER** Can check certificate and build upon work immediately.

Our register of certificates

https://codecheck.org.uk/register/

CODECHECK Register

Certificate	Repository	Type	Issue	Report	Check date
2020-001	Piccolo-2020	journal (GigaScience)	NA	http://doi.org/10.5281/zenodo.3674056	2019-02-14
2020-002	Reproduction-Hancock	community	2	http://doi.org/10.5281/zenodo.3750741	2020-04-13
2020-003	Hopfield-1982	community	1	https://doi.org/10.5281/zenodo.3741797	2020-04-06
2020-004	Barto-Sutton-Anderson-1983	community	4	https://doi.org/10.5281/zenodo.3827371	2020-05-14
2020-005	Larisch-reproduction	community	5	https://doi.org/10.5281/zenodo.3959175	2020-07-23
2020-006	Detorakis-reproduction	community	6	https://doi.org/10.5281/zenodo.3948353	2020-07-16
2020-007	Hathway-Goodman-2018	community	7	NA	NA
2020-008	covid-uk	community (preprint)	8	http://doi.org/10.5281/zenodo.3746024	2020-04-09
2020-009	2020-cov-tracing	community (preprint)	9	http://doi.org/10.5281/zenodo.3767060	2020-04-26
2020-010	covid-report9	community (preprint)	14	https://doi.org/10.5281/zenodo.3865491	2020-05-29
2020-011	covid19model-nature	community (in press)	18	https://doi.org/10.5281/zenodo.3893138	2020-06-13
2020-012	covid19model-report23	community (preprint)	19	https://doi.org/10.5281/zenodo.3893617	2020-06-14
2020-013	Spitschan2020_bioRxiv	community (preprint)	20	https://doi.org/10.5281/zenodo.3947959	2020-07-14
2020-014	Sadeh-and-Clopath	community	21	https://doi.org/10.5281/zenodo.3967326	2020-07-28
2020-015	Liou-and-Bateman	community	22	https://doi.org/10.5281/zenodo.3978402	2020-08-04
2020-016	OpeningPractice	community	15	https://doi.org/10.5281/zenodo.3981253	2020-06-02

CSV source | searchable CSV | JSON | Markdown

Example certificate: https://zenodo.org/record/3865491/files/codecheck.pdf

"It ain't pretty, but it works" (Hilda Bastian)

Sabine L.van Elsland
@SabineLvE

Independent review @StephenEglen confirmed that @MRC_Outbreak team's #COVID19 simulation is reproducible: thumbs up from code-checking efforts @nature #COVID19 #covid19science

Critiqued coronavirus simulation gets thumbs up from code-che...
Influential model judged reproducible — although software engineers called its code 'horrible' and 'a buggy mess'.
nature.com

7:47 PM · Jun 10, 2020

Limitations

1. CODECHECKER time is valuable, so needs credit.
2. Very easy to cheat the system, but who cares?
3. Author's code/data must be freely available.
4. Deliberately low threshold for gaining a certificate.
5. High-performance compute is a resource drain.
6. Cannot (yet) support all thinkable/existing workflows and languages.

Next steps

1. Embedding into journal workflows.
2. Training a community of codecheckers.
3. Funding for a codecheck editor …

For more information please see: http://codecheck.org.uk