

# Code execution during peer review

with CODECHECK and at the AGILE conference

<https://codecheck.org.uk/> | <https://reproducible-agile.github.io/>

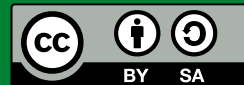
Daniel Nüst @ Collaborations Workshop 2022 (CW22), 2022-04-04

Institute for Geoinformatics, University of Münster | <http://nuest.de> | @nordholmen

Slides: <https://bit.ly/cw22-keynote-daniel>

DOI:10.6084/m9.figshare.19487573

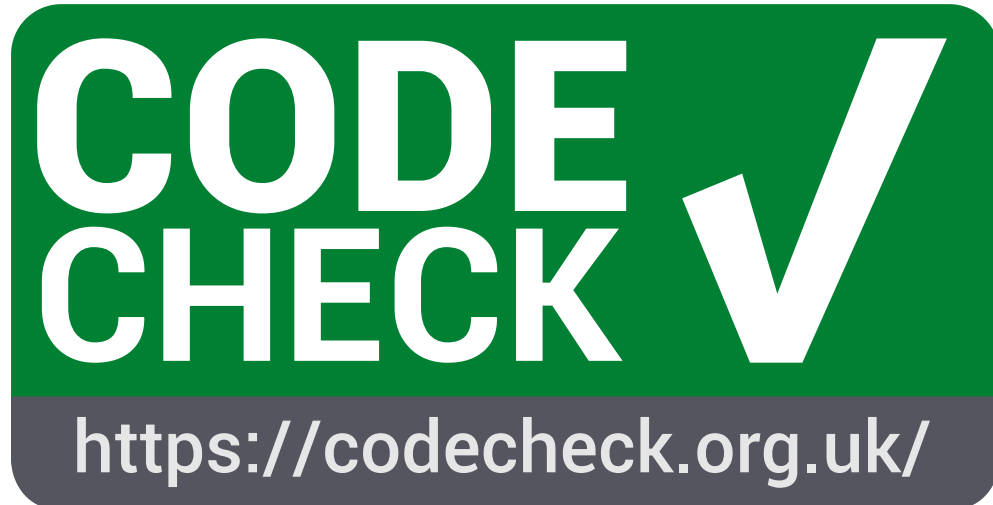
CC-BY-SA 4.0



# Agenda

HTML slides: <https://bit.ly/cw22-keynote-daniel>

PDF slides: <https://doi.org/10.6084/m9.figshare.19487573>



<https://reproducible-agile.github.io>

# Declarations and acknowledgements

## Declarations

Reproducibility Chair AGILE conference

CODECHECK paper: <https://f1000research.com/articles/10-253/v2>

## Acknowledgements

CODECHECK: Mozilla mini science grant, UK SSI; editors @ *Gigascience*, *eLife*, *Scientific Data*

Reproducible AGILE received funding as an [AGILE Initiative](#)

All work was supported by the project Opening Reproducible Research ([o2r](#)) with funding by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) under project numbers PE 1632/10-1 and PE 1632/17-1.

# CODECHECK: Evaluating the reproducibility of computational results reported in scientific journals

Stephen J Eglén  
<https://sje30.github.io>  
sje30@cam.ac.uk

Cambridge Computational Biology Institute  
University of Cambridge  
@StephenEglén

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

Institute for Geoinformatics  
University of Münster  
@nordholmen

<https://codecheck.org.uk/>

# CODECHECK in one slide

Premise: paper submitted to peer review.

1. We take your paper, code and datasets.
1. We run your code on your data.
1. If our results match your results, go to step 5.
1. Else we talk to you to find out where code broke. If you fix your code or data, we return to step 2 and try again.
1. We write a report summarising that we could reproduce your outputs (document error messages, possibly mismatches we see)
1. We work with you to freely share your paper, code, data and our reproduction.

# Premise

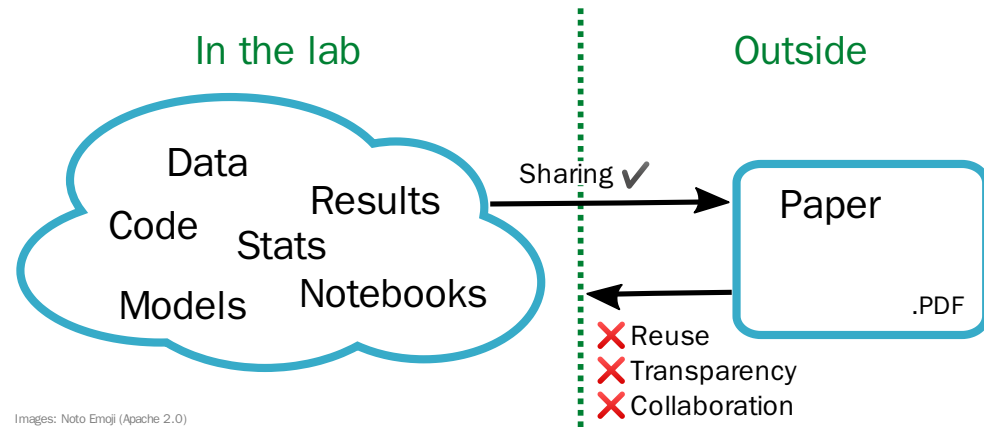


Figure 1 of

<https://doi.org/10.12688/f1000research.51738.2>

We should be sharing material on the left, not the right.

"Paper as advert for Scholarship" (Buckheit & Donoho, 1995)



<https://media0.giphy.com/media/mBpVZelPdn269zc07N/giphy.gif>

# Approaches to code sharing

- Barnes (2010): "Publish your computer code: it is good enough"
- Informal 'code buddy' system
- Community-led *research compedia* - <https://research-compendium.science/>
- 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 executable *once* for *someone* else?"
- We check the code runs and generates the expected number of output files
- The contents of those output files must not strictly be checked, though in practice until today they are; in any case outputs available for others (authors) to see
- The validity of the code is *not* checked; complement to scientific peer review

More details see [paper](#) and [CODECHECK principles](#).



# The CODECHECK example process implementation

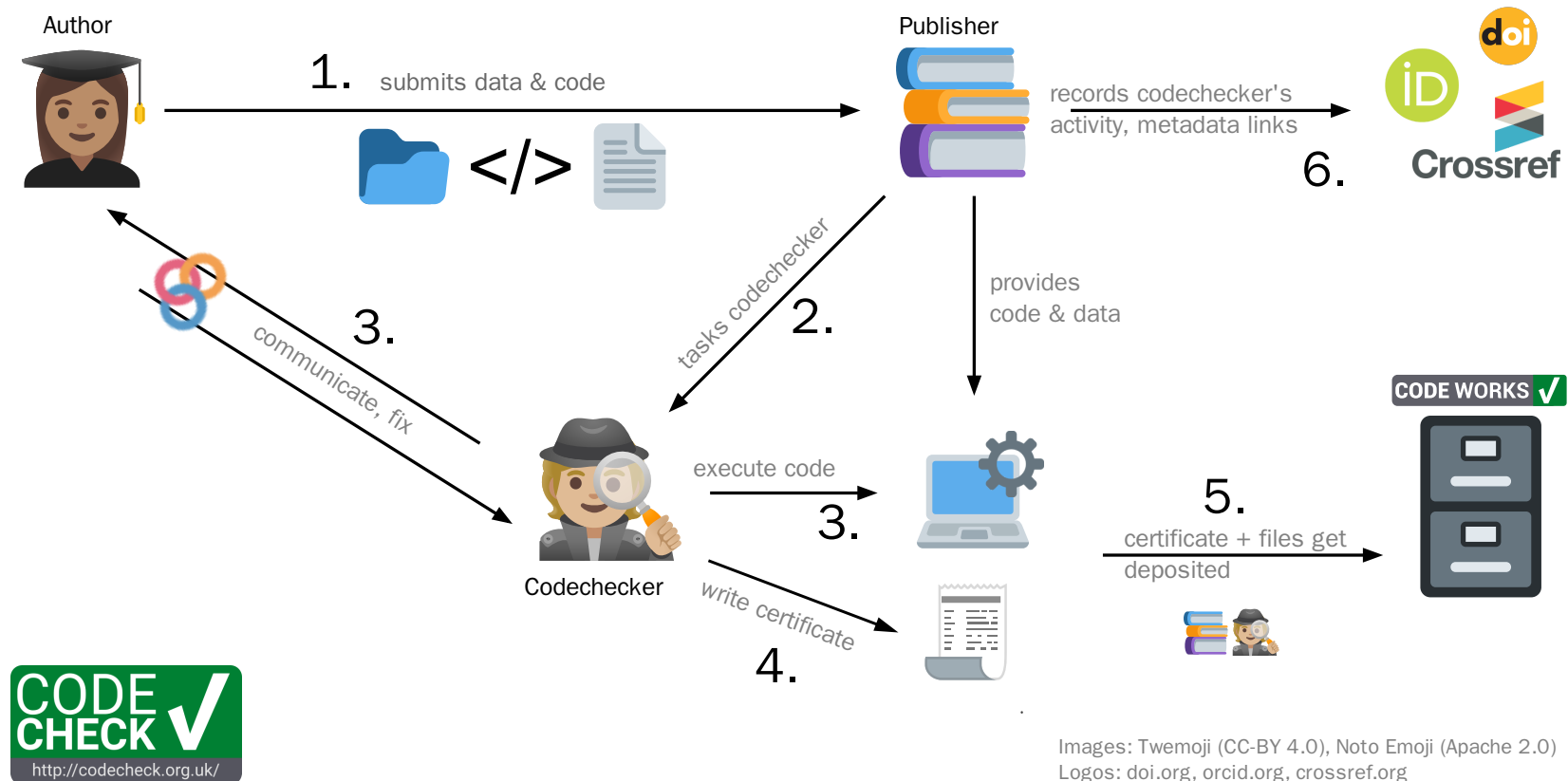


Figure 2 of <https://doi.org/10.12688/f1000research.51738.2>

# Variations in a codecheck

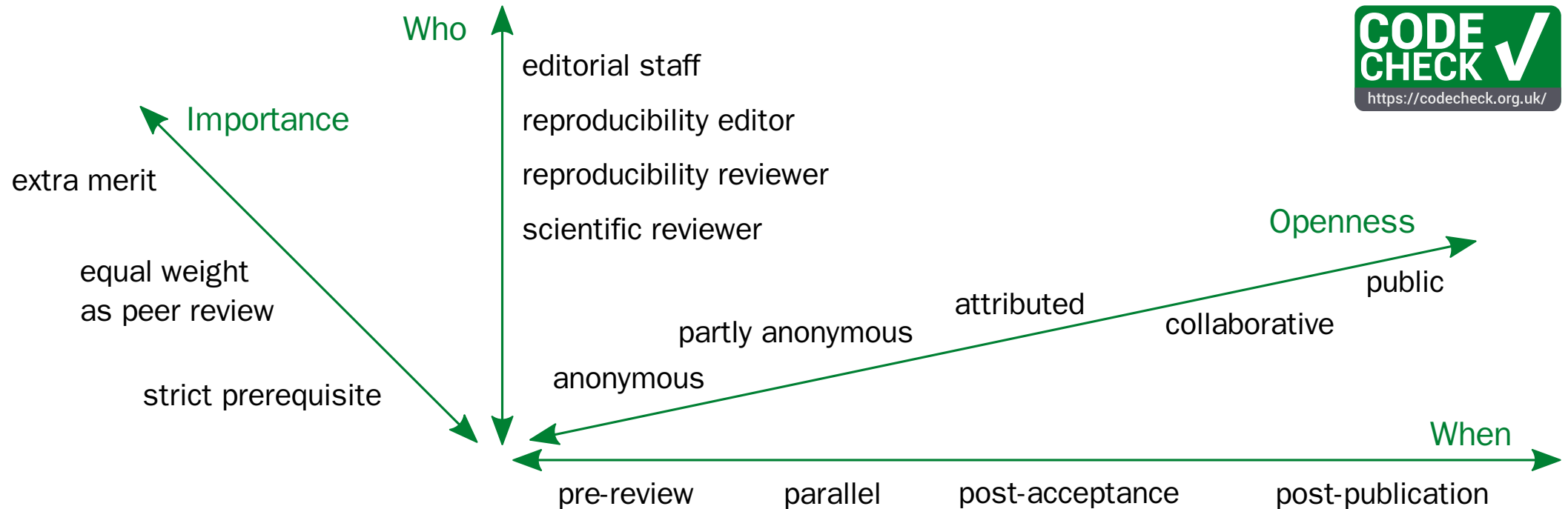


















Figure 3 of <https://doi.org/10.12688/f1000research.51738.2>

# Core principles

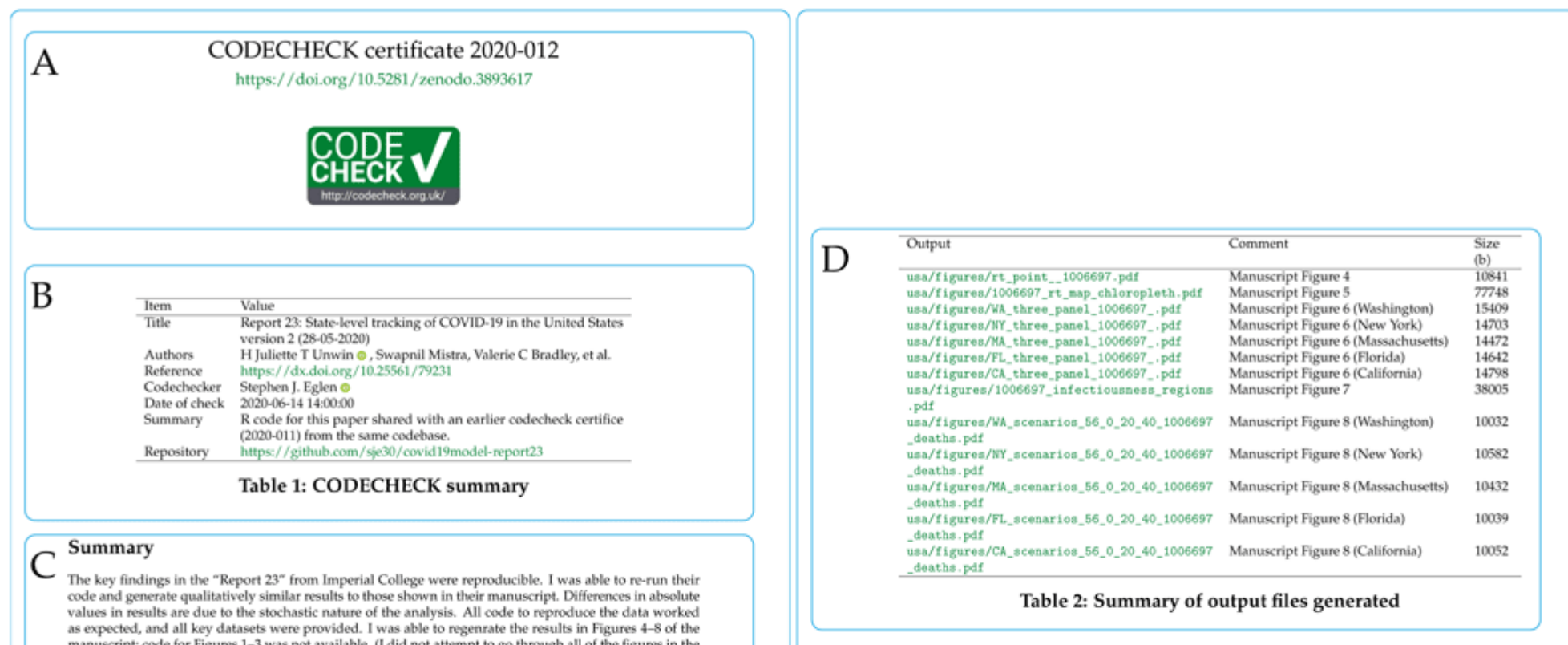
1. Codecheckers record but don't investigate or fix.
2. Communication between humans is key.
3. Credit is given to codecheckers.
4. Workflows must be auditable.
5. Open by default and transitional by disposition.

# CODECHECK Register

Certificate	Repository	Type	Issue	Report	Check date
2020-001	 <a href="#">codecheckers/Piccolo-2020</a>	journal (GigaScience)	NA	<a href="http://doi.org/10.5281/zenodo.3674056">http://doi.org/10.5281/zenodo.3674056</a>	2019-02-14
2020-002	 <a href="#">codecheckers/Reproduction-Hancock</a>	community	2	<a href="http://doi.org/10.5281/zenodo.3750741">http://doi.org/10.5281/zenodo.3750741</a>	2020-04-13
2020-003	 <a href="#">codecheckers/Hopfield-1982</a>	community	1	<a href="https://doi.org/10.5281/zenodo.3741797">https://doi.org/10.5281/zenodo.3741797</a>	2020-04-06
2020-004	 <a href="#">codecheckers/Barto-Sutton-Anderson-1983</a>	community	4	<a href="https://doi.org/10.5281/zenodo.3827371">https://doi.org/10.5281/zenodo.3827371</a>	2020-05-14
2020-005	 <a href="#">codecheckers/Larisch-reproduction</a>	community	5	<a href="https://doi.org/10.5281/zenodo.3959175">https://doi.org/10.5281/zenodo.3959175</a>	2020-07-23
2020-006	 <a href="#">codecheckers/Detorakis-reproduction</a>	community	6	<a href="https://doi.org/10.5281/zenodo.3948353">https://doi.org/10.5281/zenodo.3948353</a>	2020-07-16
2020-008	 <a href="#">codecheckers/covid-uk</a>	community (preprint)	8	<a href="http://doi.org/10.5281/zenodo.3746024">http://doi.org/10.5281/zenodo.3746024</a>	2020-04-09
2020-009	 <a href="#">codecheckers/2020-cov-tracing</a>	community (preprint)	9	<a href="http://doi.org/10.5281/zenodo.3767060">http://doi.org/10.5281/zenodo.3767060</a>	2020-04-26
2020-010	 <a href="#">codecheckers/covid-report9</a>	community (preprint)	14	<a href="https://doi.org/10.5281/zenodo.3865491">https://doi.org/10.5281/zenodo.3865491</a>	2020-05-29
2020-011	 <a href="#">codecheckers/covid19model-nature</a>	community (in press)	18	<a href="https://doi.org/10.5281/zenodo.3893138">https://doi.org/10.5281/zenodo.3893138</a>	2020-06-13
2020-012	 <a href="#">codecheckers/covid19model-report23</a>	community (preprint)	19	<a href="https://doi.org/10.5281/zenodo.3893617">https://doi.org/10.5281/zenodo.3893617</a>	2020-06-14
2020-013	 <a href="#">codecheckers/Spitschan2020_bioRxiv</a>	community (preprint)	20	<a href="https://doi.org/10.5281/zenodo.3947959">https://doi.org/10.5281/zenodo.3947959</a>	2020-07-14
2020-014	 <a href="#">codecheckers/Sadeh-and-Clopath</a>	community	21	<a href="https://doi.org/10.5281/zenodo.3967326">https://doi.org/10.5281/zenodo.3967326</a>	2020-07-28
2020-015	 <a href="#">codecheckers/Liou-and-Bateman</a>	community	22	<a href="https://doi.org/10.5281/zenodo.3978402">https://doi.org/10.5281/zenodo.3978402</a>	2020-08-04
2020-016	 <a href="#">codecheckers/OpeningPractice</a>	journal (J Geogr Syst)	15	<a href="https://doi.org/10.5281/zenodo.3981253">https://doi.org/10.5281/zenodo.3981253</a>	2020-06-02
2020-017	 <a href="#">codecheckers/JGSY-D-19-00087</a>	journal (J Geogr Syst)	24	<a href="https://doi.org/10.5281/zenodo.4003848">https://doi.org/10.5281/zenodo.4003848</a>	2020-08-27

# Example certificate

Figure 4 of <https://doi.org/10.12688/f1000research.51738.2> (click image to scroll)



# Limitations

1. CODECHECKER time is valuable, so needs credit.
2. Very easy to cheat the system, but who cares?
3. Authors' 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 journals' workflows.
2. Training a community of codecheckers (❤️ [ReproHack](#)).
3. Funding for a codecheck editor.
4. Come and [get involved](#)

For more information please see: <http://codecheck.org.uk> and [#CODECHECK](#)





# Reproducible AGILE

<https://reproducible-agile.github.io/>

2017, '18 & '19: Workshops on reproducibility

2019: Reproducible publications at AGILE conferences (initiative)

2020: AGILE Reproducible Paper Guidelines v1

2020: First AGILE reproducibility review; guidelines v2

2021: Guidelines mandatory; repro reviews linked from papers: <https://agile-giss.copernicus.org/articles/2/index.html>

# AGILE Reproducible Paper Guidelines

<https://doi.org/10.17605/OSF.IO/CB7Z8>

- Promotion, not exclusion
- Data and software availability section
- Author & reviewer guidelines
- Reproducibility checklist

14 successful reproductions in 2020 & '21








Website: <https://osf.io/phmoe/>  
Version: December 2020  
DOI: 10.17605/OSF.IO/CB7Z8



## REPRODUCIBLE PAPER GUIDELINES

Full and short papers submitted to the AGILE conference **have** to include a **Data and Software Availability** section which documents data, software, and computational infrastructure to support reproduction, or mentions reasons for not publishing them.

The above requirement is the only one to comply with the AGILE Reproducible Paper Guidelines. The remainder of the document provides concrete recommendations for all involved stakeholders to increase transparency, reproducibility, and openness of computational GIScience research. The following table of contents shows the recommended parts for different readers. Familiarity with all sections is, of course, beneficial.

Author	Reproducibility Reviewer	Scientific Reviewer
		
<b>Reproducibility Checklist</b>		
Helps to ensure authors and reviewers do not miss anything important.		
2		
		
<b>Author Guidelines</b>		
Show how to write the Data and Software Availability Section and give practical recommendations to make data and computational workflows reproducible.		
Writing the Data and Software Availability Section		
Including Data in Research Papers		
Including Computational Workflows in Research Papers		
		
<b>Scientific Reviewer Guidelines</b>		
Describe role in evaluating plausibility and completeness of the data and software availability documentation.		
7		
		
<b>Reproducibility Reviewer Guidelines</b>		
Describe role and approach to execute workflows and clarify efforts.		
8		
<b>Background</b>		
10		

### Further resources

These guidelines can not cover all details of the reproducibility review at AGILE conferences. For more information for authors, translations, and practical examples see the [guidelines wiki](#). For more information about the review process and deadlines, see the [process description](#). For any questions, please visit the AGILE Discourse server's [forum for the Reproducible Paper Guidelines](#).

# Review process

Reproducibility review after accept/reject decisions

Reproducibility review & communication

Community conference & volunteers

Badges on proceedings website, article website with link, and first article page (💖 Copernicus!)

The screenshot shows the article page for 'Window Operators for Processing Spatio-Temporal Data Streams on Unmanned Vehicles' by Tobias Werner and Thomas Brinkhoff. The page includes a sidebar with navigation links (About, Articles, Review process, Licence & copyright, Publication ethics, Article level metrics), a search bar, and a download section. The main content area displays the article title, authors, keywords, abstract, and a reproducibility badge. The abstract discusses the challenges of processing spatio-temporal data streams on unmanned vehicles and proposes a set of six new operators. The reproducibility badge is a yellow rectangle with the AGILE logo and the word 'reproducible'.

The screenshot shows the article page for 'Vehicles' by Dominik Bucher, Henry Martin, Jannik Hamper, Atefeh Jaleh, Henrik Becker, Pengxiang Zhao, and Martin Raubal. The page includes a sidebar with navigation links, a search bar, and a download section. The main content area displays the article title, authors, keywords, abstract, and a reproducibility badge. The abstract discusses the challenges of processing spatio-temporal data streams on unmanned vehicles and proposes a set of six new operators. The reproducibility badge is a yellow rectangle with the AGILE logo and the word 'reproducible'.

The screenshot shows the article page for 'Reproducibility review: https://doi.org/10.17605/osf.io/anv9r'. The page includes a sidebar with navigation links, a search bar, and a download section. The main content area displays the article title, authors, keywords, abstract, and a reproducibility badge. The abstract discusses the challenges of processing spatio-temporal data streams on unmanned vehicles and proposes a set of six new operators. The reproducibility badge is a yellow rectangle with the AGILE logo and the word 'reproducible'.



How to put your community on a path towards more reproducibility in 5 ~~easy~~ hard steps

- 1 Build a team of enthusiasts (workshop, social events) 💪🧠
- 2 Assess the current state and raise awareness (workshop, paper) 🔬
- 3 Institutional support (🙏 AGILE Council 🙏 + committee chairs)
- 4 Positive encouragement (no reproduction != bad science) ✨
- 5 Keep at it! 😊

---

(Next) steps

Reproducibility reviews 2022+

Grow reproducibility reviewer team

Continue community discourse

Re-assess new papers > impact?

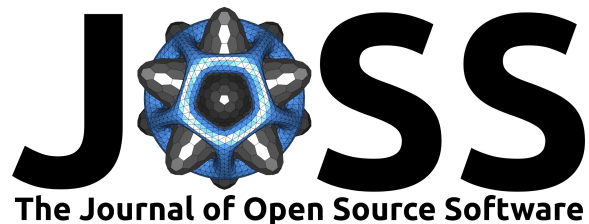
Towards open scholarship: Open review if tenured?

Format-free first submission? CRediT?

# What is RSE(ng) about this?



# Code review != reproducibility review



Code Review Community  
working group

Code Review During Development session tomorrow!



# Cultural Change



One thing on more reproducible research publications:

# Have a README: all else is details.

Inspired by Greg Wilson's Teaching Tech Together Rule 1 - <http://teachtogether.tech/en/index.html>

---

Thank you! Questions?

HTML slides: <https://bit.ly/cw22-keynote-daniel> | PDF slides: <https://doi.org/10.6084/m9.figshare.19487573>



Encore

# "It ain't pretty, but it works" (H. Bastian)

(The most prominent check until today!)



Sabine L. van Elstrand  
@SabineLvE



Independent review [@StephenEglen](#) confirmed that [@MRC\\_Outbreak](#) team's [#COVID19](#) simulation is reproducible: thumbs up from code-checking efforts [@nature](#) [#COVID19](#) [#covid19science](#)



Thomas Angus/Imperial College London

[nature.com](#)

Critiqued coronavirus simulation gets thumbs up from code-che...

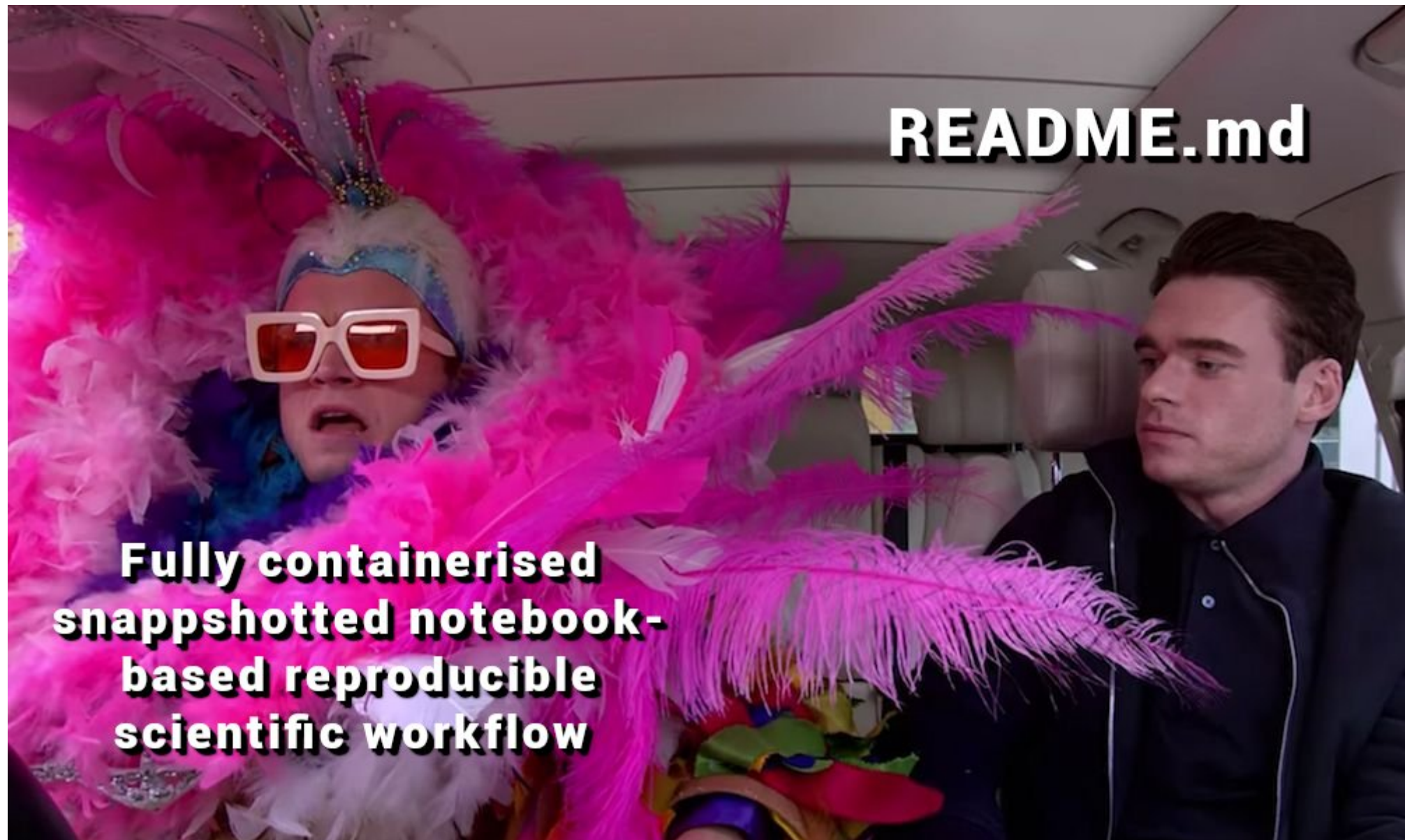
Nature - Influential model judged reproducible — although software engineers called its code 'horrible' and 'a buggy mess'.

# 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.



**README.md**

**Fully containerised  
snappshotted notebook-  
based reproducible  
scientific workflow**

# Definition

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

# Learn more about code execution practices at journals and conferences

<https://osf.io/x32nc>

Daniel Nüst, Heidi Seibold, Stephen Eglen, Lea Schulz-Vanheyden, Limor Peer, Josef Spillner

# Deep dive



Chiarelli, Andrea, Loffreda, Lucia, & Johnson, Rob.  
(2021). The Art of Publishing Reproducible Research  
Outputs: Supporting emerging practices through cultural  
and technological innovation. Zenodo.

<https://doi.org/10.5281/zenodo.5521077>

Chiarelli, Andrea, Loffreda, Lucia, & Johnson, Rob.  
(2021). Executive Summary: The Art of Publishing  
Reproducible Research Outputs: Supporting emerging  
practices through cultural and technological innovation.  
Zenodo. <https://doi.org/10.5281/zenodo.5639384>

<https://knowledge-exchange.info/event/publishing-reproducible-research-output>



# Reproducible AGILE and CODECHECK: Highlights of Lessons learned

Spectrum or layers of reproducibility very apparent

Effect of guidelines at AGILE: improved reproducibility, community discourse

Reproducibility reports/CODECHECK certificates full of recommendations for improvement, often well received by authors, many included in revised submission

Good practices spread slowly, establishing a process is tedious, needs time until familiarity

Challenges for reproducibility reviewer: Inconsistencies and disconnects (figures), lack of documentation, unknown runtimes vs. no subsets of data, lack of reprod. guidance

Reproductions are rewarding and educational, matching expertises tricky

Communication is without alternative

Safety net (👁👁), not security



# What can communities & institutions do?

Introduce reproducibility reviews - CODECHECK (or not) - at your journals, labs, collaborations!

Workshops on RCR, ReproHacks

Provide support (R2S2, PhD edu.)

Rewards and incentives

Community discourse

Awareness > Change

Throw technology at it

*Digital information lasts forever, or five  
years - whichever comes first.*

Rothenberg, Jeff. 1995. "Ensuring the Longevity of Digital Documents." Scientific American 272 (1): 42–47. JSTOR via  
[https://twitter.com/snet\\_jklump/status/1141934045820887040?s=09](https://twitter.com/snet_jklump/status/1141934045820887040?s=09)

# "Preproducibility" - Philip Stark

*"Science should be 'show me', not 'trust me'; it should be 'help me if you can', not 'catch me if you can'."*

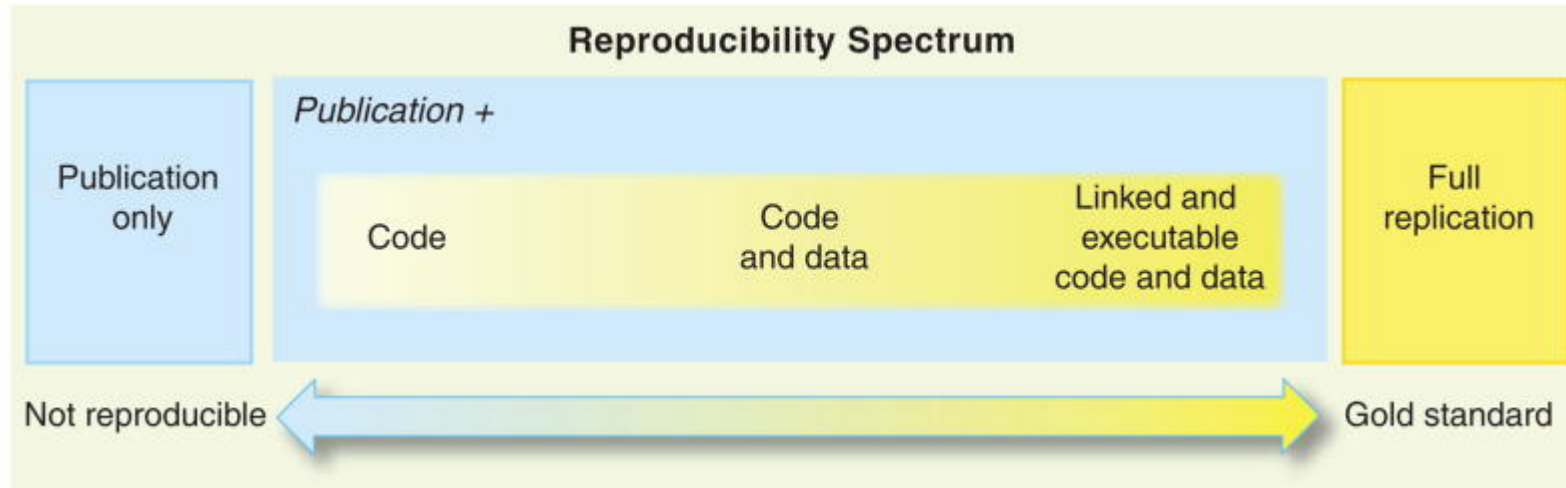
Nature 557, 613 (2018).

<https://doi.org/10.1038/d41586-018-05256-0>

[...]

*"If you and I get different results, preproducibility can help us to identify why — and the answer might be fascinating."*

# Reproducibility spectrum



Peng R. D. (2011). Reproducible research in computational science. *Science* (New York, N.Y.), 334(6060), 1226–1227. <https://doi.org/10.1126/science.1213847>

# Five selfish reasons to work reproducibly

1. reproducibility helps to avoid disaster
2. reproducibility makes it easier to write papers
3. reproducibility helps reviewers see it your way
4. reproducibility enables continuity of your work
5. reproducibility helps to build your reputation

Markowetz, F. Five selfish reasons to work reproducibly. *Genome Biol* 16, 274 (2015).

<https://doi.org/10.1186/s13059-015-0850-7>

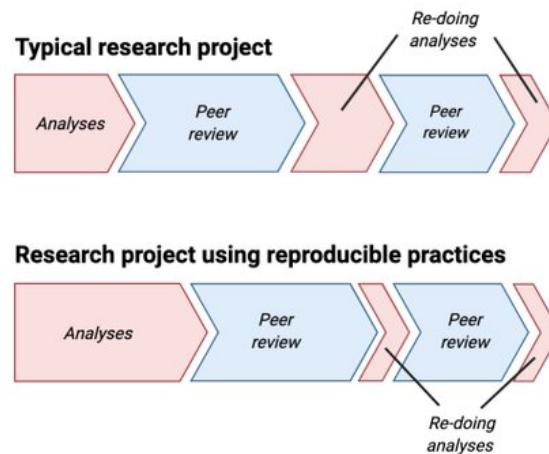
# Reproducibility is "more work"



Dan Quintana  
@dsquintana



In my experience, you don't lose time doing reproducible science—you just \*relocate\* how you're spending it



3:13 PM · Nov 26, 2020



[Read the full conversation on Twitter](#)

# GIScience assessment

Nüst, Daniel. 2021. Infrastructures and Practices for Reproducible Research in Geography, Geosciences, and GIScience. Doctoral dissertation, University of Münster, Germany. <https://doi.org/10.5281/zenodo.4768096>

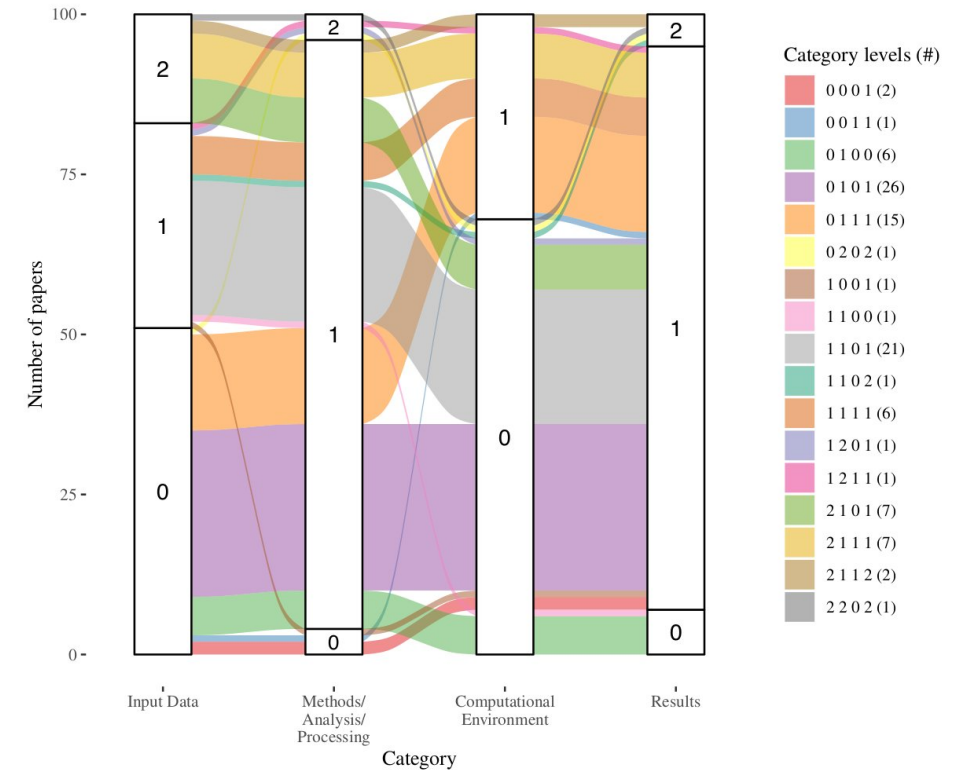


Figure 4: Combined alluvial diagram. Includes groups of papers across four categories for the merged AGILE (Chapter 11) and GIScience (Chapter 12) datasets; the category *Preprocessing* was dropped because of difficulties to clearly assess it; included are 100 papers without any “not applicable” value from 2010 to 2018; see analysis notebook at [https://zivgitlab.uni-muenster.de/d\\_nues01/phd-package/-/blob/master/combined-giscience-assessment-figure.Rmd](https://zivgitlab.uni-muenster.de/d_nues01/phd-package/-/blob/master/combined-giscience-assessment-figure.Rmd).