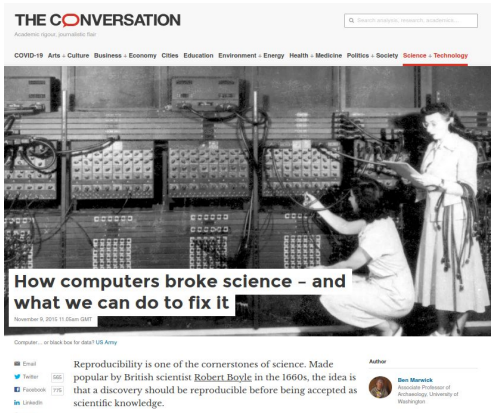
Executing workflows during peer review for transparency, reproducibility, and reusability

Thuringian RDM-Days 2021

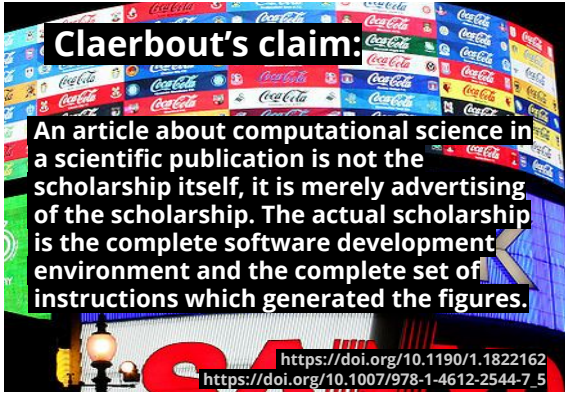
Daniel Nüst
Institute for Geoinformatics, Uni Münster
d.n@wwu.de | @nordholmen
0000-0002-0024-5046

Slides: https://doi.org/10.5281/zenodo.5006379

https://giphy.com/gifs/usnationalarchives-nasa-scientist-scientists-1F1JGyGZhiSAA8Vuhn



https://theconversation.com/how-computers-broke-science-and-what-we-can-do-to-fix-it-49938



CC-BY 3.0, Sebastian Bertalan, Wikimedia Commons

https://giphy.com/gifs/with-computers-fascination-PxSFAnuubLkSA

https://giphy.com/gifs/david-hasselhoff-M3o3fL9nnxG4o

https://twitter.com/BioMickWatson/status/1402157533897318402?s=09

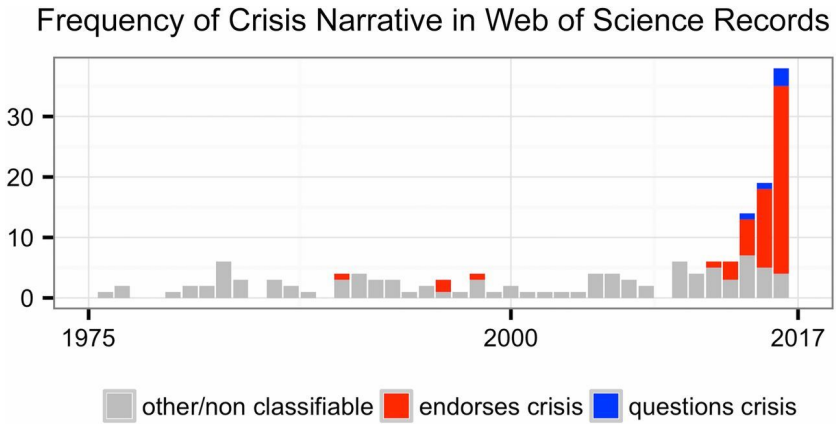
Crisis? Crisis of what?

COLLOQUIUM OPINION

Opinion: Is science really facing a reproducibility crisis, and do we need it to?

Daniele Fanelli

PNAS March 13, 2018 115 (11) 2628-2631; first published March 12, 2018; https://doi.org/10.1073/pnas.1708272114



https://doi.org/10.1073/pnas.1708272114

Credibility crisis?

Replicability crisis?

Reproducibility crisis?

Robustness crisis?

Generalisability crisis?

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

https://the-turing-way.netlify.app/reproducible-research/overview/overview-definitions.html

Reproducible Research

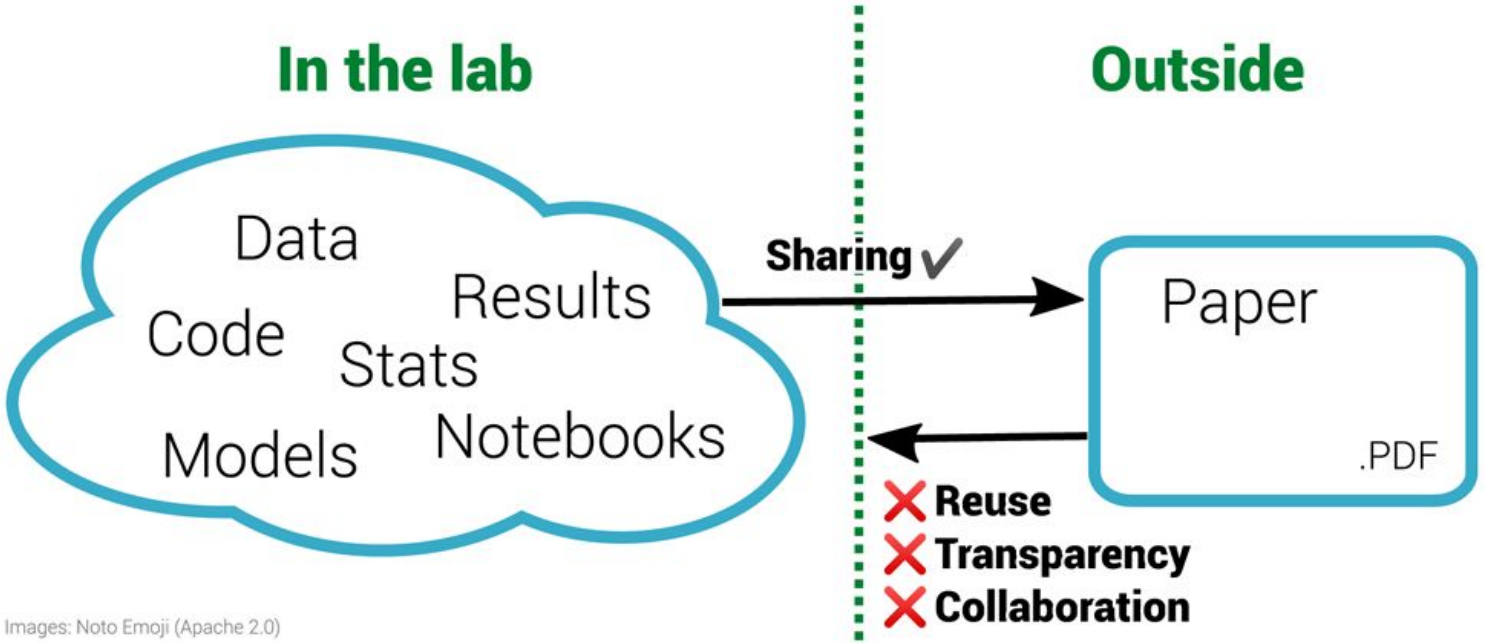
Peer Review

Reproducible research and peer review are cornerstones of science. But are they getting along?

https://giphy.com/gifs/suspicious-YNEpBZC0Ly08M https://giphy.com/gifs/lol-futurama-humor-cFgb5p5e1My3K

CODECHECK

https://codecheck.org.uk/



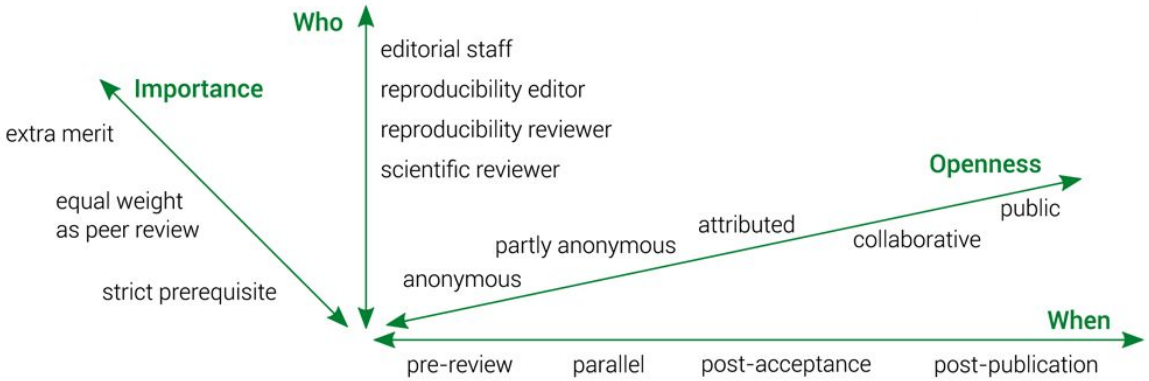
The inverse problem in reproducible research. Figure 1 of https://doi.org/10.12688/f1000research.51738.1

The left half of the diagram shows a diverse range of materials used within a laboratory. These materials are often then condensed for sharing with the outside world via the research paper, a static PDF document. Working backwards from the PDF to the underlying materials is impossible. This prohibits reuse and is not only non-transparent for a specific paper but is also ineffective for science as a whole. By sharing the materials on the left, others outside the lab can enhance this work.

One re-execution of computational workflow by codechecker during peer review

CODE CHECK ✓
https://codecheck.org.uk/

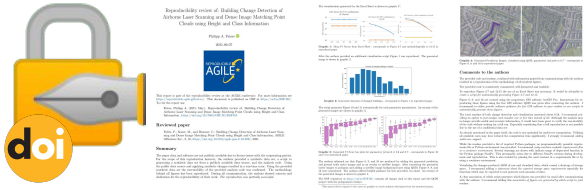
Independent execution of computations underlying research articles.



The CODECHECK variations. Figure 1 of https://doi.org/10.12688/f1000research.51738.1

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.

CODE WORKS ✓



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

METHOD ARTICLE

CODECHECK: an Open Science initiative for the independent execution of computations underlying research articles during peer review to improve reproducibility [version 1; peer review: 1 approved, 1 approved with reservations]

Daniel Nüst [1], Stephen J. Eglen [2]

Nüst D and Eglen SJ. **CODECHECK: an Open Science initiative for the independent execution of computations underlying research articles during peer review to improve reproducibility** [version 1; peer review: awaiting peer review]. F1000Research 2021, 10:253 (https://doi.org/10.12688/f1000research.51738.1)

Next steps

https://codecheck.org.uk/get-involved/

CODECHECK paper ✔️

Build journal partnerships through community checks 😓

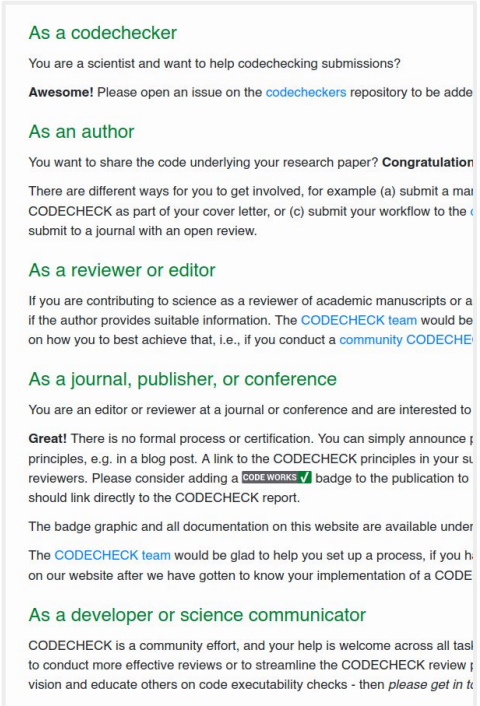
Grow codechecker community 🔥

Collaborate on education (ReproHack, PhD schools) 👩‍🎓

CODECHECK editor for 💎 OA journals

CODE CHECK ✓
https://codecheck.org.uk/

Independent execution of computations underlying research articles.



As a codechecker

You are a scientist and want to help codechecking submissions?

Awesome! Please open an issue on the codecheckers repository to be adde

As an author

You want to share the code underlying your research paper? **Congratulation**

There are different ways for you to get involved, for example (a) submit a ma CODECHECK as part of your cover letter, or (c) submit your workflow to the submit to a journal with an open review.

As a reviewer or editor

If you are contributing to science as a reviewer of academic manuscripts or a if the author provides suitable information. The CODECHECK team would be on how you to best achieve that, i.e., if you conduct a community CODECHE

As a journal, publisher, or conference

You are an editor or reviewer at a journal or conference and are interested to

Great! There is no formal process or certification. You can simply announce principles, e.g. in a blog post. A link to the CODECHECK principles in your su reviewers. Please consider adding a CODE WORKS ✓ badge to the publication to should link directly to the CODECHECK report.

The badge graphic and all documentation on this website are available under

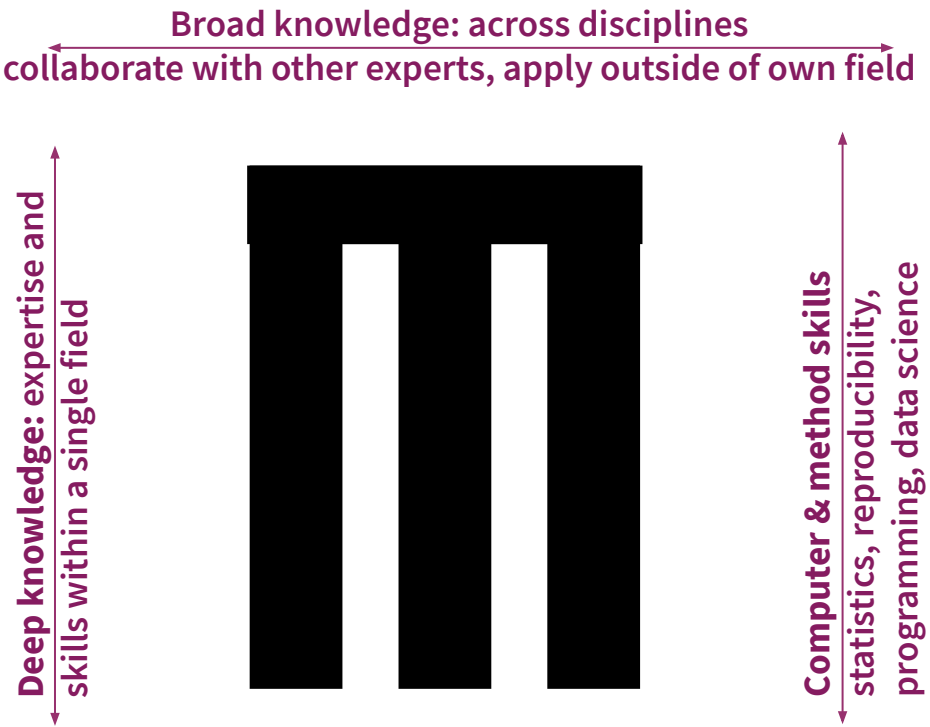
The CODECHECK team would be glad to help you set up a process, if you h on our website after we have gotten to know your implementation of a CODE

As a developer or science communicator

CODECHECK is a community effort, and your help is welcome across all tasl to conduct more effective reviews or to streamline the CODECHECK review vision and educate others on code executability checks - then *please get in t*

<Excursion>

Traditional and modern scientists



US RSE
US RESEARCH SOFTWARE ENGINEER ASSOCIATION

de RSE

RSE
RESEARCH SOFTWARE ENGINEERS ASSOCIATION

<ReSA>
Research Software Alliance

https://en.wikipedia.org/wiki/T-shaped_skills

https://doi.org/10.1007/s10816-015-9272-9

https://jakevdp.github.io/blog/2014/08/22/hacking-academia/
https://www.sciencemag.org/careers/2013/05/when-all-science-becomes-data-science
https://escience.washington.edu/community-level-data-science-and-its-spheres-of-influence-beyond-novelty-squared/

Who are you?

researcher

hacker

BETTER SOFTWARE

BETTER RESEARCH

scientist

developer

T

≠

https://en.wikipedia.org/wiki/T-shaped_skills

https://doi.org/10.1007/s10816-015-9272-9

https://jakevdp.github.io/blog/2014/08/22/hacking-academia/

https://www.sciencemag.org/careers/2013/05/when-all-science-becomes-data-science

https://escience.washington.edu/community-level-data-science-and-its-spheres-of-influence-beyond-novelty-squared/

https://www.software.ac.uk/resources/publications/better-software-better-research

Where is the NFSI?

Nationale Forschungssoftwareinfrastruktur!

F1000Research

OPINION ARTICLE

REVISED **An environment for sustainable research software in Germany and beyond: current state, open challenges, and call for action [version 2; peer review: 2 approved]**

Hartwig Anzt[1,2*], Felix Bach [1*], Stephan Druskat [3-5*], Frank Löffler[3,6*], Axel Loewe [1*], Bernhard Y. Renard[7*], Gunnar Seemann [8*], Alexander Struck [5*], Elke Achhammer[9], Piush Aggarwal [10], Franziska Appel[11], Michael Bader[9], Lutz Brusch [12], Christian Busse [13], Gerasimos Chourdakis [9], Piotr Wojciech Dabrowski [14], Peter Ebert[15], Bernd Flemisch[16], Sven Friedl [17], Bernadette Fritzsch[18], Maximilian D. Funk[19], Volker Gast[3], Florian Goth[20], Jean-Noël Grad [16], Jan Hegewald [18], Sibylle Hermann[16], Florian Hohmann[21], Stephan Janosch[22], Dominik Kutra [23], Jan Linxweiler [24], Thilo Muth [25], Wolfgang Peters-Kottig [26], Fabian Rack[27], Fabian H.C. Raters [28], Stephan Rave [29], Guido Reina [16], Malte Reißig [30], Timo Ropinski[31,32], Joerg Schaarschmidt[1], Heidi Seibold [33], Jan P. Thiele [34], Benjamin Uekermann [35], Stefan Unger[36], Rudolf Weeber[16]

* Equal contributors

Anzt H, Bach F, Druskat S *et al.* An environment for sustainable research software in Germany and beyond: current state, open challenges, and call for action [version 2; peer review: 2 approved]. *F1000Research* 2021, 9:295 (https://doi.org/10.12688/f1000research.23224.2)

Software = Research output!

F1000Research

METHOD ARTICLE

REVISED **Recognizing the value of software: a software citation guide [version 2; peer review: 2 approved]**

Previously titled: "The importance of software citation"

Daniel S. Katz [1], Neil P. Chue Hong [2], Tim Clark[3], August Muench [4], Shelley Stall [5], Daina Bouquin[6], Matthew Cannon [7], Scott Edmunds[8], Telli Faez[9], Patricia Feeney[10], Martin Fenner[11], Michael Friedman [12], Gerry Grenier [13], Melissa Harrison [14], Joerg Heber[15], Adam Leary [16], Catriona MacCallum [17], Hollydawn Murray[18], Erika Pastrana[19], Katherine Perry [20], Douglas Schuster[21], Martina Stockhause [22], Jake Yeston[23]

Katz DS, Chue Hong NP, Clark T *et al.* Recognizing the value of software: a software citation guide [version 2; peer review: 2 approved]. *F1000Research* 2021, 9:1257 (https://doi.org/10.12688/f1000research.26932.2)

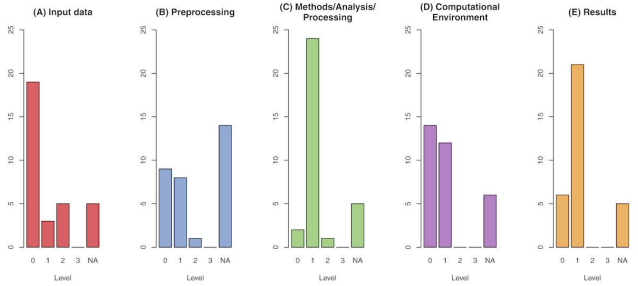
</Excursion>

Reproducible AGILE

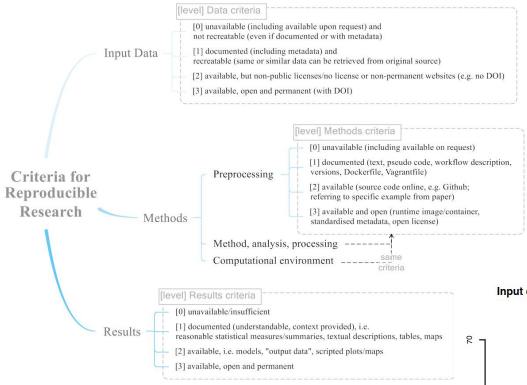
https://reproducible-agile.github.io/

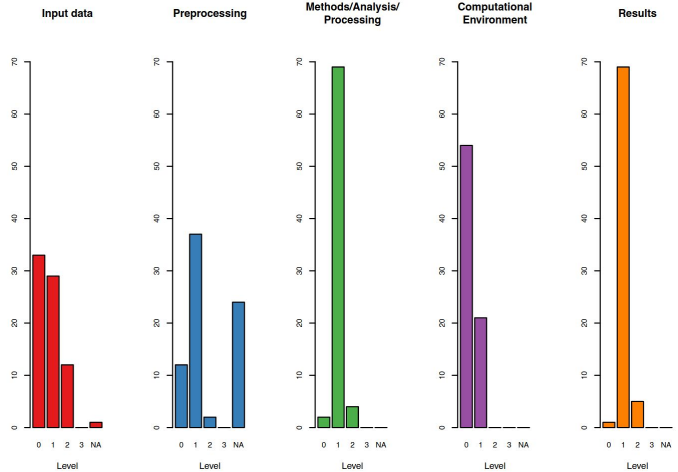
2017, ‘18 & ‘19: Workshops on reproducibility
2019: Reproducible publications at AGILE conferences (initiative)
2020: First AGILE reproducibility review
2021: Second AGILE reproducibility review

Assessment of GIScience papers



Legal restrictions	Lack of time	Lack of tools	Lack of knowledge	Lack of incentiv
Main reason	Strongly hindered	Not at all	Not at all	Strongly hindere
Main reason	Not at all	Not at all	Not at all	Moderately hinde
Main reason	Slightly hindered	Strongly hindered	Moderately hindered	Strongly hindere
Main reason	Not at all	Slightly hindered	Not at all	Not at all
Strongly hindered	Strongly hindered	Strongly hindered	Moderately hindered	Strongly hindere
Moderately hindered	Main reason	Not at all	Not at all	Not at all
Slightly hindered	Moderately hindered	Slightly hindered	Slightly hindered	Moderately hinde
Slightly hindered	Not at all	Main reason	Strongly hindered	Not at all
Not at all	Moderately hindered	Not at all	Moderately hindered	Not at all
Not at all	Strongly hindered	Strongly hindered	Strongly hindered	Slightly hindered
Not at all	Moderately hindered	Not at all	Not at all	Not at all
Not at all	Slightly hindered	Main reason	Not at all	Strongly hindere
Not at all	Main reason	Not at all	Not at all	Not at all
Not at all	Main reason	Not at all	Not at all	Not at all
Not at all	Moderately hindered	Moderately hindered	Not at all	Strongly hindere
Not at all	Not at all	Not at all	Not at all	Not at all
Not at all	Slightly hindered	Not at all	Slightly hindered	Not at all





Nüst, D., Granell, C., Hofer, B., Konkol, M., Ostermann, F. O., Sileryte, R., & Cerutti, V. (2018). *Reproducible research and GIScience: an evaluation using AGILE conference papers*. PeerJ, 6, e5072. **https://doi.org/10.7717/peerj.5072**

Ostermann, F., Nüst, D., Granell, C., Hofer, B., & Konkol, M. (2020). *Reproducible Research and GIScience: an evaluation using GIScience conference papers*. EarthArXiv. **https://doi.org/10.31223/x5zk5v** | *pub. pending at GIScience conf.*

AGILE Reproducible Paper Guidelines

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

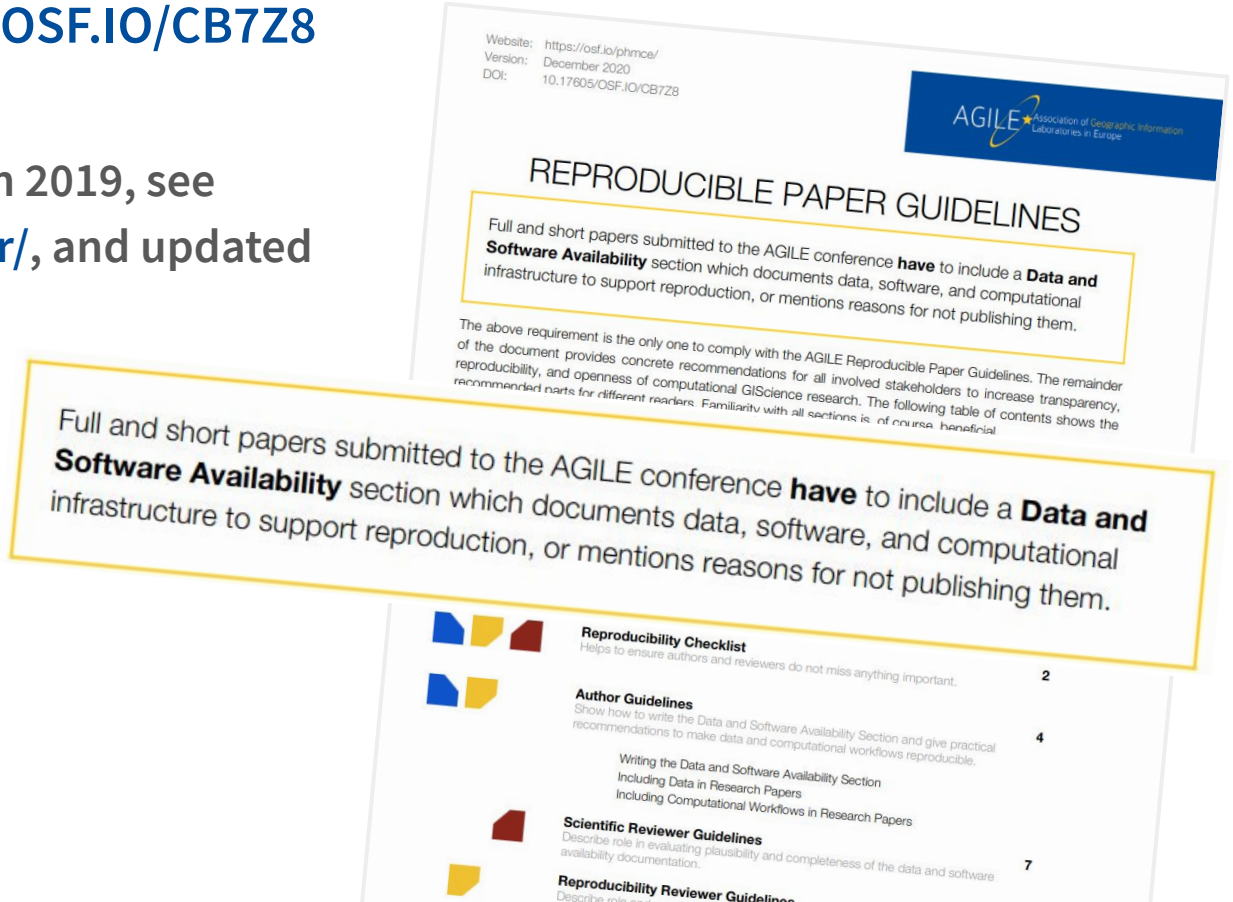
Created by AGILE Initiative in 2019, see report at https://osf.io/hupxr/, and updated in 2020

Transparency

Promotion

Acknowledge spectrum

GIScience



The guidelines

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

Reproducibility checklist

Author guidelines

Writing DASA section

Data in Research Papers

Computational workflows in Research Papers

Reviewer guidelines

Reproducibility reviewer guidelines

Background

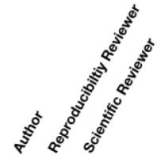
Website: https://osf.io/phmce/
Version: December 2020
DOI: 10.17605/OSF.IO/CB7Z8

AGILE Association of Geographic Information Laboratories in Europe

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.



Reproducibility Checklist 2
Helps to ensure authors and reviewers do not miss anything important.

Author Guidelines 4
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

Scientific Reviewer Guidelines 7
Describe role in evaluating plausibility and completeness of the data and software availability documentation.

Reproducibility Reviewer Guidelines 8
Describe role and approach to execute workflows and clarify efforts.

Background 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.

 CC-BY 4.0

Checklist and writing the DASA section

REPRODUCIBILITY CHECKLIST

For all **datasets** included/produced in the paper, check if data:

- ❑ Is provided in a non-proprietary format
- ❑ Is documented for third parties to reuse
- ❑ Is accessible in a public repository and has an open data licence

For all **software tools/libraries/packages** and **computational workflows** included/produced, check if:

- ❑ Reproduction steps are explained in a README (plain text file), flowchart, or script
- ❑ Computational environments (including hardware) are documented or provided
- ❑ Versions of relevant software components (libraries, packages) are provided
- ❑ All parameters and expected execution times for the computational workflow are provided
- ❑ Software developed by the authors is available in a public repository and has an open licence
- ❑ There is a clear connection between **tables, figures, maps, and statistical values** and the data and code that they are based on, e.g., using file names or documentation in the README

In the **Data and Software Availability section**, check if you include:

- ❑ Data and software statements (see examples below)
- ❑ The reasons, if any, for not being able to share (parts of) data or code

For all **data and software** check that:

- ❑ All datasets and code (used or mentioned) are assigned DOIs
- ❑ Datasets and code are cited throughout the paper

After acceptance in the **camera-ready paper** check that:

- ❑ If data has been shared privately or anonymously for peer review, they are updated with all metadata and accessible via a DOI and referenced from the paper
- ❑ If a reproducibility review report will be published for your paper, a DOI URL in the Data and Software Availability section is included using the following template:
  A reproducibility report for this paper is available confirming that [considerable parts of the computational workflow / all results / Figures 1 and 4] could be independently reproduced, see https://doi.org/link_to_report.

WRITING THE DATA AND SOFTWARE AVAILABILITY SECTION

The DASA section provides references to where data, software and documentation is available (e.g., paper section or README file) and under what conditions (e.g., copyright, licenses or access procedures for protected data). It should be concise and contain persistent links to repositories using Digital Object Identifiers[7] (DOI). You may remove links for anonymity during peer review ("xxx"), or share anonymized links[8] if your repository supports them. Data, software and (third-party) tools should be cited following recommended citation or standard citation guidelines. Possible statements for the DASA section are provided below. You may include one of these statements or draft your own.

Statements for non-computational or conceptual work

No data or code was collected, developed, or used in this work.

The full list of reviewed literature is available at [link to attachment or citable deposit of bibliography].

The full concept maps are available at [link] *and the ideas were first sketched in a blog post at* [link].

Research data/code supporting this publication ...

… is available in [name of the repository(-ies)] *and is accessible via the following DOI* [DOI link(s)]

… was accessed on [date of dataset access/download] *with the following* [query parameters, if applicable] *under the license* [dataset license].

… was downloaded manually using the services at [name of organisation] *(using a departmental subscription for costs) and* [name of organisation]. *The compiled dataset cannot be redistributed due to licensing restrictions.*

...is not available due to [indicate reasons, e.g., licenses, sensitive data on human subjects, privacy statements; if there are processes to obtain the data, describe them].

The computational workflow supporting this publication …

… is executed via [choose, e.g., a single command/file, a workflow management software, a set of numbered scripts] *published under license* [the license] *at* [DOI of repository].

… is published in a [language] *module/package at* [link of software project]. *The used version is archived at* [DOI of repository].

… is provided as a [container/VM] *published at* [DOI of repository] *with instructions included in the file README.md in the repository.*

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

The guidelines for data

REPRODUCIBLE AGILE

"What if…"

Examples

INCLUDING DATA IN RESEARCH PAPERS

	Minimum requirements	Recommended practices
What?	• All input data and configuration • Data description/documentation, including provenance, field or column types, etc. • If data is retrieved from an external source, documentation on collection queries and download steps	• Standardised, discipline-specific metadata[8] and ontologies to describe your data • Data download scripts
Where?	• Publish data in a public repository providing a DOI • Cite data (including date and version) in the paper	• Discipline- or data type-specific repository[9] • Include recommended citation in dataset description (unless already provided by repository) • Create a registration for OSF projects[10] and use the DOI to cite it
How?	• Use open data formats; export from proprietary format for publication • Specify the license	• Use plain text-based file formats

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

The guidelines for computational workflows

REPRODUCIBLE AGILE

INCLUDING COMPUTATIONAL WORKFLOWS IN RESEARCH PAPERS

	Minimum requirements	Recommended practices
What? Computational environment	• Describe the used environment and computational infrastructure, e.g., hardware specs, operating system • List software versions • Cite used software[14]	• Provide the actual environment, e.g., a Dockerfile + container[15] or a Virtual Machine (e.g., using OSGeo-Live) • Provide a pinned freeze of your dependencies (structured configuration files with dependency information) • Add a colophon or "reproducibility receipt"[16] to your notebooks • Installation and execution instructions for different operating systems
Computation steps	• Document the detailed steps in a text file and/or flowchart (every action/click) • Document expected execution times given computing power unless negligible • Ask a colleague to try out the instructions	• Scripts/models and a README file that explains their use • All figures are fully scripted and a peer has read your README's instructions (incl. interactive visualisations and interactive adjustments • Multi-panel plots are composited with scripts[17] • Software package with structured metadata[18], tests/CI[19], and a pipeline framework[20] or workflow language[21] • Live documents for analyses, e.g., Binder[22] • Live demo of APIs/online applications (e.g., anonymous cloud resources, such as Google Cloud Run or AWS) • Subset or a synthetic dataset for quick evaluation
Where?	• Repository providing a persistent identifier, e.g., a DOI or SWHID[23]	• Versioned code repository, such as GitHub or GitLab, and ongoing open development
How? Tools used	• Use generally available tools (avoid proprietary tools that are not available to reviewers and other researchers)	• Use and create Open Source tools • Cite core modules/tools/language used
Development practices	• Use clear licenses[24] that fit your environment • Follow one of "Good enough practices in scientific computing"[25]	• Follow all "Good enough practices.." Use development guidelines for your environment / language of choice (e.g., for R[26])

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

Scientific reviewer guidelines… concerning the reproducibility review only!

REPRODUCIBLE AGILE

SCIENTIFIC REVIEWER GUIDELINES

This section clarifies the expectations and role of the scientific reviewer with respect to the reproducible paper guidelines. For information for the Reproducibility Reviewer, please see the following section.

Reproducibility is considered good scientific practice that provides input for the quality assessment of a paper. Therefore, reviewers of AGILE papers should be aware of the **author guidelines on reproducibility** and be familiar with the **reproducibility checklist**, as well as the expected content of the **mandatory data and software availability section.** Using this information, reviewers should evaluate the plausibility and completeness of the data and software availability documentation, and whenever possible and readily available **include feedback on reproducibility aspects** in their comments. Scientific reviewers are free to but **are not expected to attempt reproductions of computations**.

Data and software availability documentation provide an additional set of information for assessing the quality of research presented in a manuscript. Reviewers are asked to know about the AGILE reproducible paper guidelines and to consider the level of reproducibility reached in a manuscript. To do so, they shall assume the position of someone who would like to reproduce the submitted work to assess whether the provided material is likely to allow reproduction of the submitted work. Based on this impression, reviewers may challenge authors regarding the level of reproducibility reached, if any statements are made regarding reproducibility in a manuscript.

Scientific reviewers are not required to actually reproduce a manuscript, but, if the data and code are provided in an anonymous format, and if a reviewer attempts to reproduce all or parts of the submitted work, then they are asked to document the process and outcomes (see Reproducibility Reviewer Guidelines below). Please reach out to the reproducibility chair if you are keen on conducting a reproducibility review for a paper you are reviewing.

The peer review of AGILE papers is a fully anonymous peer review, i.e. authors and reviewers do not know each other's identity. Reviewers should be supportive to authors and consider potential limitations in access to resources due to anonymisation. Since the provision of information to help reproduction of a paper can accidentally lead to disclosure of an author's identity, the reviewers should not use any such additional information to the disadvantage of the authors. The reviewers' comments provided to the authors are expected to be neutral[28] and contribute to improved reproducibility of the reported findings.

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

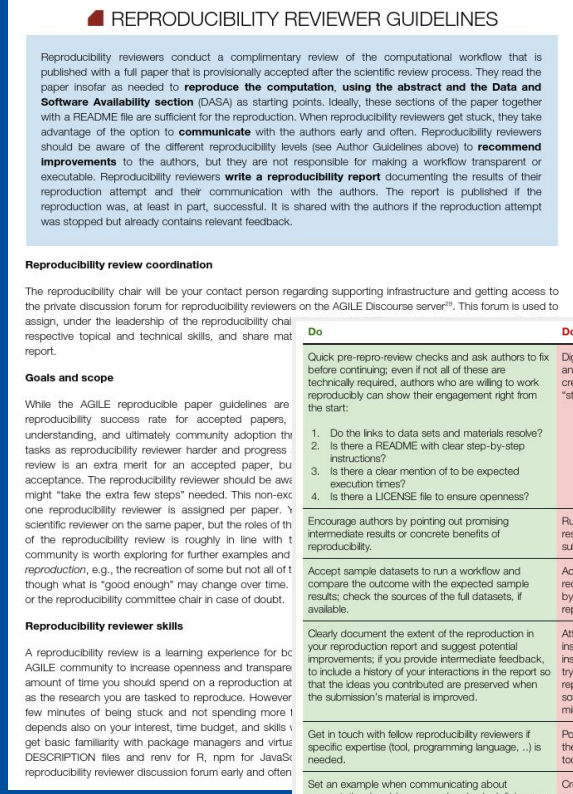
The guidelines for reproducibility reviewers

Ideal vs. realistic

Role & skills

Examples for “Do’s and Don’ts”:

- Do shift burden to author
- Do encourage and set examples
- Private data/code sharing last resort
- Document your work in report (impact)
- Be kind (career stage, knowledge, privileges)
- No rummaging



REPRODUCIBILITY REVIEWER GUIDELINES

Reproducibility reviewers conduct a complimentary review of the computational workflow that is published with a full paper that is provisionally accepted after the scientific review process. They read the paper insofar as needed to **reproduce the computation**, **using the abstract and the Data and Software Availability section** (DASA) as starting points. Ideally, these sections of the paper together with a README file are sufficient for the reproduction. When reproducibility reviewers get stuck, they take advantage of the option to **communicate** with the authors early and often. Reproducibility reviewers should be aware of the different reproducibility levels (see Author Guidelines above) to **recommend improvements** to the authors, but they are not responsible for making a workflow transparent or executable. Reproducibility reviewers **write a reproducibility report** documenting the results of their reproduction attempt and their communication with the authors. The report is published if the reproduction was, at least in part, successful. It is shared with the authors if the reproduction attempt was stopped but already contains relevant feedback.

Reproducibility review coordination

The reproducibility chair will be your contact person regarding supporting infrastructure and getting access to the private discussion forum for reproducibility reviewers on the AGILE Discourse server[29]. This forum is used to assign, under the leadership of the reproducibility chai
respective topical and technical skills, and share mat
report.

Goals and scope

While the AGILE reproducible paper guidelines are
reproducibility success rate for accepted papers,
understanding, and ultimately community adoption th
tasks as reproducibility reviewer harder and progress
review is an extra merit for an accepted paper, bu
acceptance. The reproducibility reviewer should be awa
might “take the extra few steps” needed. This non-exc
one reproducibility reviewer is assigned per paper.
scientific reviewer on the same paper, but the roles of th
of the reproducibility review is roughly in line with t
community is worth exploring for further examples and
reproduction, e.g., the recreation of some but not all of t
though what is “good enough” may change over time.
or the reproducibility committee chair in case of doubt.

Reproducibility reviewer skills

A reproducibility review is a learning experience for bo
AGILE community to increase openness and transpare
amount of time you should spend on a reproduction at
as the research you are tasked to reproduce. However
few minutes of being stuck and not spending more t
depends also on your interest, time budget, and skills
get basic familiarity with package managers and virtu
DESCRIPTION files and renv for R, npm for JavaS
reproducibility reviewer discussion forum early and often

Do	Don’t
Quick pre-repro-review checks and ask authors to fix before continuing; even if not all of these are technically required, authors who are willing to work reproducibly can show their engagement right from the start: 1. Do the links to data sets and materials resolve? 2. Is there a README with clear step-by-step instructions? 3. Is there a clear mention of to be expected execution times? 4. Is there a LICENSE file to ensure openness?	Dig across badly or un-documented collections of files and functions to identify which part of the code/data creates which figure/table/output; find or build the “start button” yourself.
Encourage authors by pointing out promising intermediate results or concrete benefits of reproducibility.	Run workflows requiring considerable computational resources (unless interesting for you) but ask for data subsets for demonstration purposes.
Accept sample datasets to run a workflow and compare the outcome with the expected sample results; check the sources of the full datasets, if available.	Accept private sharing of data or code, unless strictly required for protection of sensitive data. All changes by the author should update to the public reproduction material.
Clearly document the extent of the reproduction in your reproduction report and suggest potential improvements; if you provide intermediate feedback, to include a history of your interactions in the report so that the ideas you contributed are preserved when the submission’s material is improved.	Attempt to install software without any instructions, install binary software of unknown origin, or try to fix installation problems you encounter on your machine; try to install without (a) asking for help from a fellow reproducibility reviewer who is familiar with the software, or (b) asking the author to help, providing a minimal reproducible example of your problem.
Get in touch with fellow reproducibility reviewers if specific expertise (tool, programming language, ..) is needed.	Point out or even fix problems that are not specific to the submission, e.g., general problems in a software tool.
Set an example when communicating about computational problems, e.g., by clearly defining your system (OS version, language version, etc.)	Create accounts on any service or platform to access code, data, or other resources.
Ask specific questions or point out concrete problems that may lead authors to improve their material, including referencing these guidelines or concrete tools/methods that you already (!) know about, especially if you suspect that the author might now be familiar with them (e.g., version pinning/dependency management, absolute paths).	Fix anything (unless you really enjoy doing so), e.g., • compiler problems, • outdated libraries, • broken paths, or • Incomplete computing environment specifications, especially if the author can fix them even quicker.
Make sure that you are aware of any templates or specific resources provided for reproducibility reviewers from the reproducibility committee chair before starting your review.	
Consider the author’s background, career stage, and position to be aware of (a lack of) privileges or institutional power to decide how much support you provide and how you communicate; your reproducibility review can be a contribution to improve equity and inclusion in academia.	Be a bro.

Review process

Proceedings:
https://www.agile-giscience-series.net/review_process.html

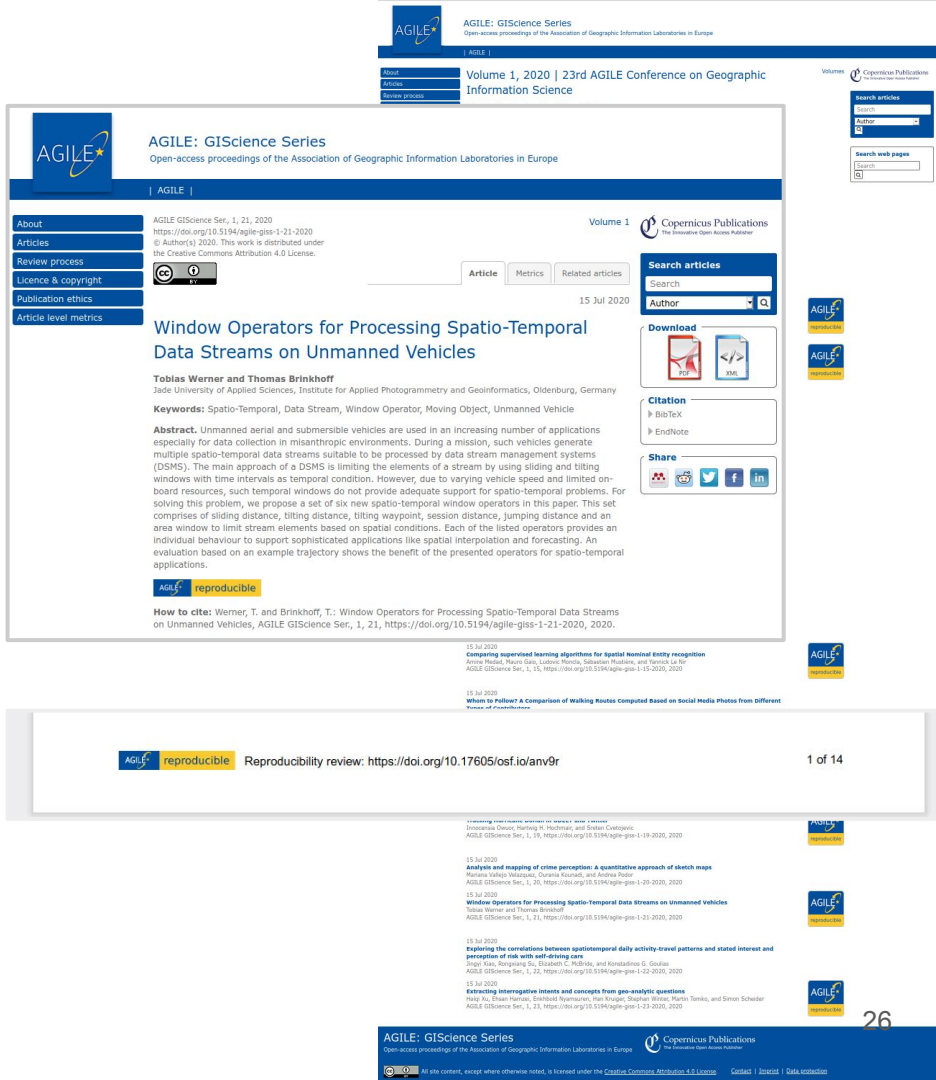
Process documentation:
https://osf.io/7rjpe/

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!)



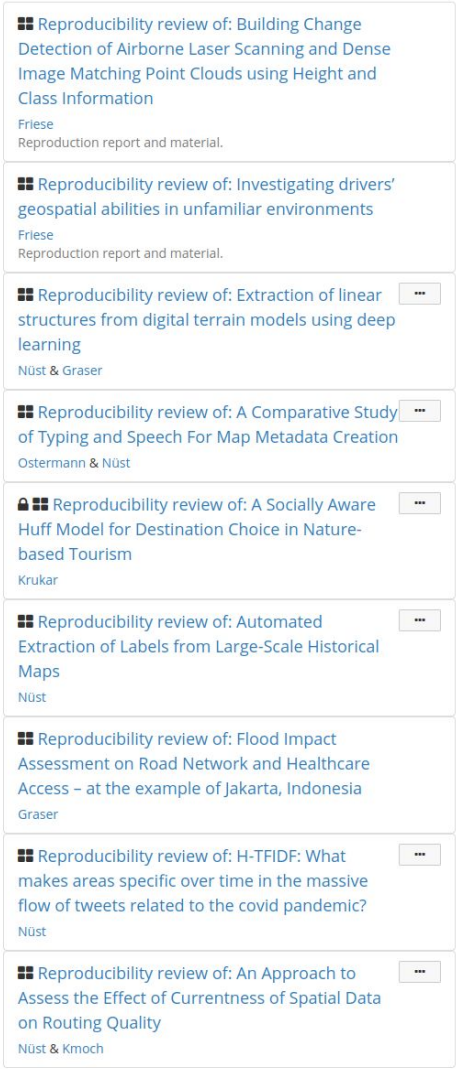
Reproducibility review results 2021

9 reproducibility reports published (2020: 6)

- no starting point in the paper
- documentation insufficient for third party

8 not reproducible:

- conceptual papers
- data not shared (choice, licence)
- code not shared (choice) or proprietary software (repro reviewer matching failed)



https://osf.io/h64sd/

Reproducibility Reports

Published on OSF with a DOI
Title page, cites the paper

Paper links to report via URL
(no citation)

Automatically added to ORCID profile

Eventually indexed in GS

Reproducibility review of: Investigating drivers' geospatial abilities in unfamiliar environments

Philipp A. Friese

2021-06-07

REPRODUCIBLE AGILE

This report is part of the reproducibility review at the AGILE conference. For more information see https://reproducible-agile.github.io/. This document is published on OSF at https://osf.io/dx92a. To cite the report use

Friese, Philipp A. (2021, May). Reproducibility review of: Investigating drivers' geospatial abilities in unfamiliar environments. https://doi.org/10.17605/OSF.IO/DX92A

Reviewed paper

Karkasina, D., Kokla, M., and Tomai, E.: Investigating drivers' geospatial abilities in unfamiliar environments, AGILE GIScience Ser., 2, 3, https://doi.org/10.5194/agile-giss-2-3-2021, 2021.

Summary

The updated submiss... tionnaires. The provi... , dataset and ques... paper and generates ...n...

2.4 Data and Software Availability

Questionnaires and sketches were collected anonymously. All statistical analyses, which results are detailed in the following section, have been performed in R (R Core Team, 2021) using the tidyverse package (Wickham et al., 2019). Driving directions given to participants, an Exemplary Questionnaire in English, the collected survey data in tabular form, the R code of the statistical analysis workflow, and all necessary metadata supporting this publication, are available on figshare and are accessible via the following DOI: https://doi.org/10.6084/m9.figshare.14460102.v4. The workflow underlying this paper was successfully reproduced by an independent reviewer during the AGILE reproducibility review and a reproducibility report was published at https://doi.org/10.17605/OSF.IO/DX92A.

ORCID Connecting Research and Researchers

ABOUT FOR RESEARCHERS MEMBERSHIP DOCUMENTATION

Daniel Nüst

Biography

Daniel is a research software engineer and PhD student at th... producible geoscientific research in the project Opening Repr... producibility.

› Employment (6)
› Education and qualifications (2)
› Invited positions and distinctions (1)
› Membership and service (5)
› Funding (3)
› Works (50 of 74)

Items per page: 50 1 – 50 of 74

Reproducibility review of: A Comparative Study of Typing ... Creation
Open Science Framework
2021 | other
DOI: 10.17605/osf.io/7fqtm
Source: DataCite

Reproducibility review of: An Approach to Assess ... Data on Routing Quality
Open Science Framework
2021 | other
DOI: 10.17605/osf.io/bdu28
Source: DataCite ★ Preferred source

Reproducibility review of: Automated Extraction of Labels from Large-Scale Historical Maps
Open Science Framework
2021 | other
DOI: 10.17605/osf.io/anv9r

Your new notifications

YOUR RECORD

DataCite has made changes to your ORCID record

Showing **5** *out of* **5** *changes made by this client*

WORKS

Added

- Reproducibility review of: A Comparative Study of Typing ... (2021-06-08)
- Reproducibility review of: An Approach to Assess the Effec... (2021-06-08)
- Reproducibility review of: Automated Extraction of Labels f...
- Reproducibility review of: Extraction of linear structures fro... (2021-06-08)
- Reproducibility review of: H-TFIDF: What makes areas spe... to the covid pandemic? (2021-06-08)

...cholar agile "reproducibility review of"

4 Ergebnisse (0,08 Sek.)

[PDF] **Reproducibility review of**: Window operators for processing spatio-temporal data streams on unmanned vehicles
D Nüst, F Ostermann - 2020 - ris.utwente.nl
Page 1. **Reproducibility review of**: Window Operators for Processing Spatio-Temporal Data Streams on Unmanned Vehicles Daniel Nüst , Frank O. Ostermann 2020-07-13 This report is part of the reproducibility review at the **AGILE** conference ...
☆ 99 Zitiert von: 1 Alle 4 Versionen »

Reproducibility review:" Comparing supervised learning algorithms for Spatial Nominal Entity recognition
A Medad, M Gaio, L Moncla, S Mustière, Y Le Nir - research.utwente.nl
... For more information see https://reproducible-**agile**.github.io/ This document is published on OSF at https://osf.io/suwpj/ To cite this report use Ostermann, FO, and Nüst, D. (2020, July). **Reproducibility review of**: Comparing supervised learning algorithms for Spatial Nominal ...
☆ 99 Alle 2 Versionen »

Reproducibility review:" Tracking Hurricane Dorian in GDELT and Twitter
I Owuor, H Hochmair, S Cvetojevic - research.utwente.nl
... **Reproducibility review of**: Tracking Hurricane Dorian in GDELT and Twitter. https://doi.org/10.17605/OSF.IO/XS5YR Reviewed paper Owuor, Innocensia, Hochmair, Hartwig and Cvetojevic, Sreten: Tracking Hurricane Dorian in GDELT and Twitter. **AGILE** GIScience Ser., 1, 19 ...
☆ 99 »

Nach Relevanz sortieren
Nach Datum sortieren
Beliebige Sprache
Seiten auf Deutsch
Patente einschließen
Zitate einschließen
Alert erstellen

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

Do it again in 2022 🎉

Grow reproducibility reviewer team

Opportunity ECRs (mentoring/workshops/…)

Continue discourse (meaning of rprdcblty)

Re-assess new papers > impact?

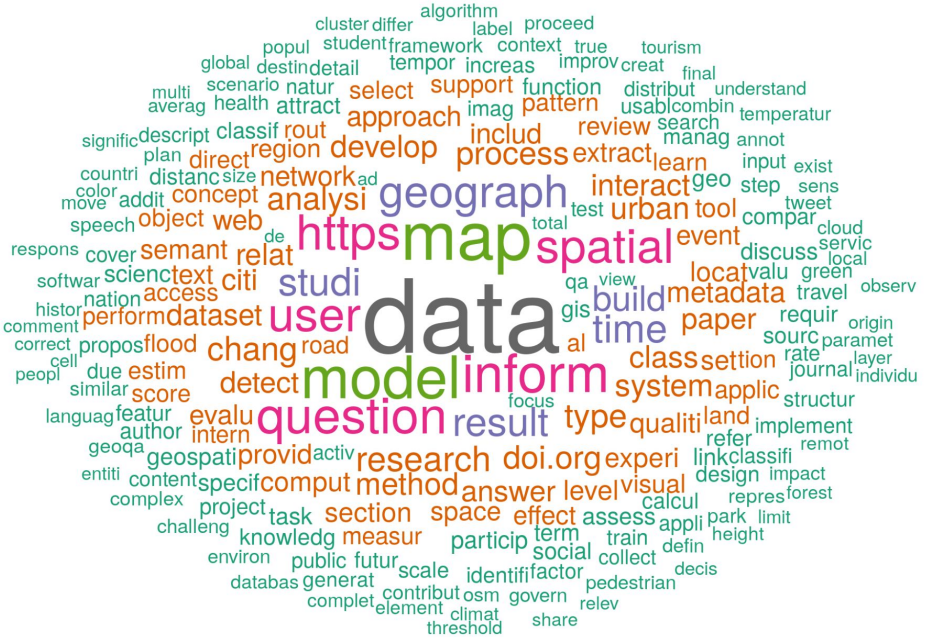
Towards opening scholarship

Scope, requirements, acceptance condition?

Open review if tenured? Format-free first submission

CRediT

Phase out when standard practice...



Word-stem cloud of all AGILE 2021 submissions (full/short/poster & accepted/rejected)

What did we learn?

https://giphy.com/gifs/foxhomeent-l44QgV6RFRGgjDsDC/media

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 you do today?

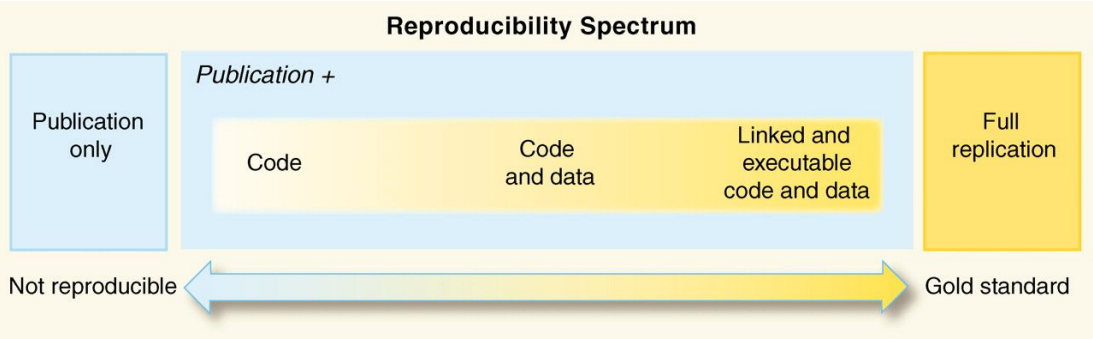
THÜRINGER FDM-TAGE 2021

REPRODUCIBLE RESEARCH IS THE BEST DATA DOCUMENTATION

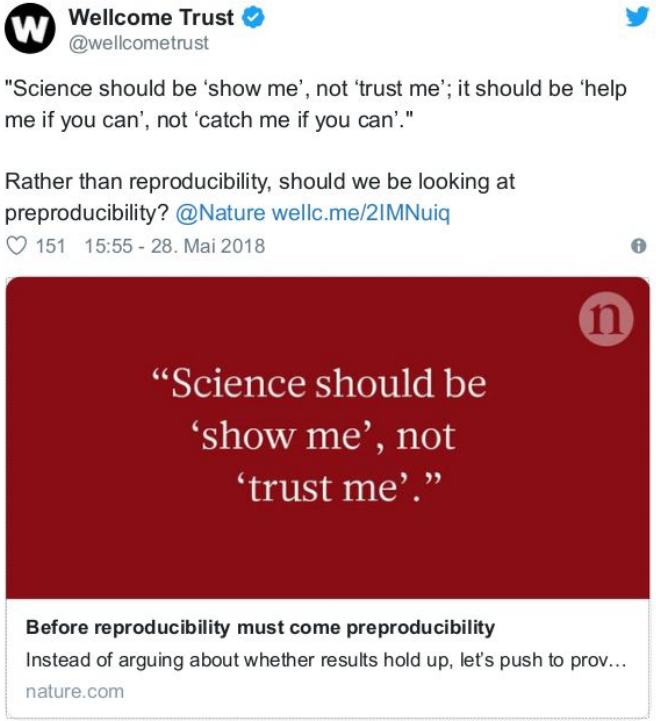
```
<title>
  A love note to the future!
</title>
<meta name="description" content="event">
<meta name="dates" content="21.06.21-25.06.21">
<meta name="author" content="Thueringer
Kompetenznetzwerk Forschungsdatenmanagement">
```

FDM THÜRINGEN

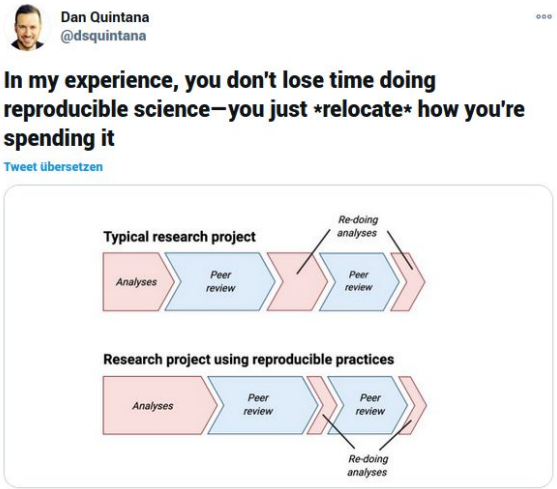
Reproducible Research & Open Science

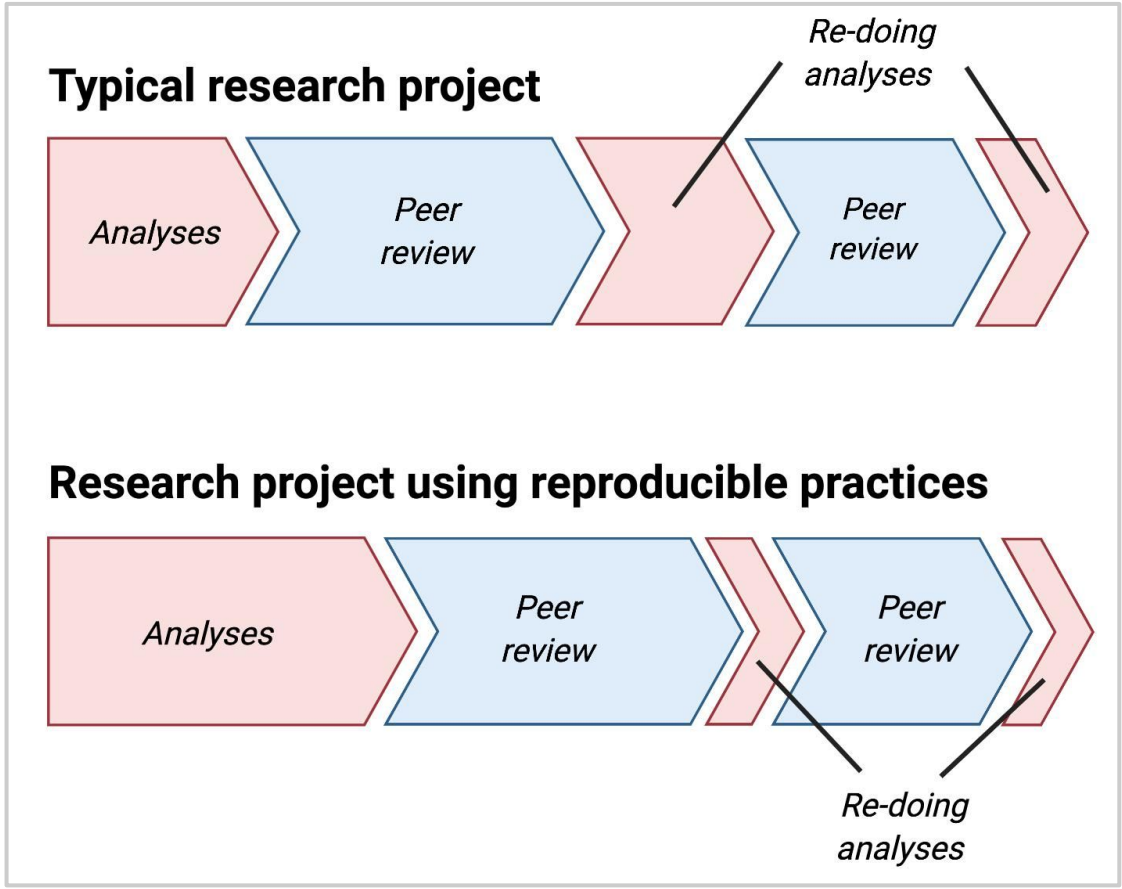


https://doi.org/10.1126/science.1213847



https://www.nature.com/articles/d41586-018-05256-0





Quintana, D. S. (2020, November 28). Five things about open and reproducible science that every early career researcher should know. https://doi.org/10.17605/OSF.IO/DZTVQ

Work more reproducibly… by finding the champions in your domain!

🌍🗺️🌳🌊⛏️😔😂🖥️⚙️👫

Full list of resources for many disciplines at http://bit.ly/rcr-in

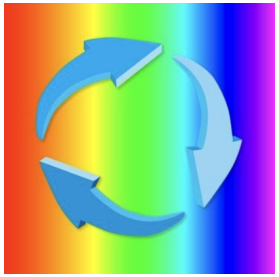
Creating reproducible workflows

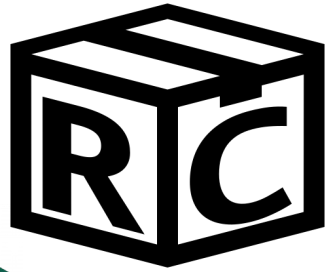
Computing environment: hardware + software, containers/virtualisation (Binder), freezing/pinning
Script-based workflows: no point-and-click GIS, notebooks (Jupyter, R Markdown)
> Research compendium > https://research-compendium.science/

Challenges (for geography, geosciences, GIScience)

Education, publishing practices, SDIs, GIS, proprietary software,
lack of rewards/pressure, sensitive data, time, ...
> all solvable

What can scientists do?





Take one step at a time.

Create and publish **Research Compendia** (**Your code is good enough**!): **https://research-compendium.science/**

Become a **codechecker** or **reprohacker**.
Join a **Reproducibility 4 Everyone** workshop.

Strive to be an open science champion **especially** if you're junior in your field. [**RIOT talk by Gavin Buckinham**; **preprint by Sam Westwood**]

Be the change, find communities, do not rely on those in power - they don't know!

What can communities and 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

https://giphy.com/gifs/chicagodancecrash-KCqjrcPfL55q3MkgHZ

Vision

o2r opening reproducible research

https://o2r.info

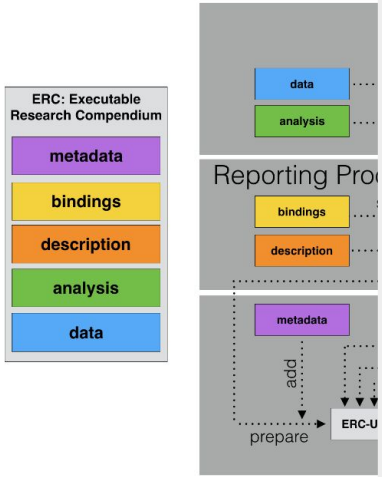


Figure 1 Executable Research Compe
ERCs can be integrated into the researc
stands for an unvalidated ERC, ERC-V fo
for a published one. Processes are sequen

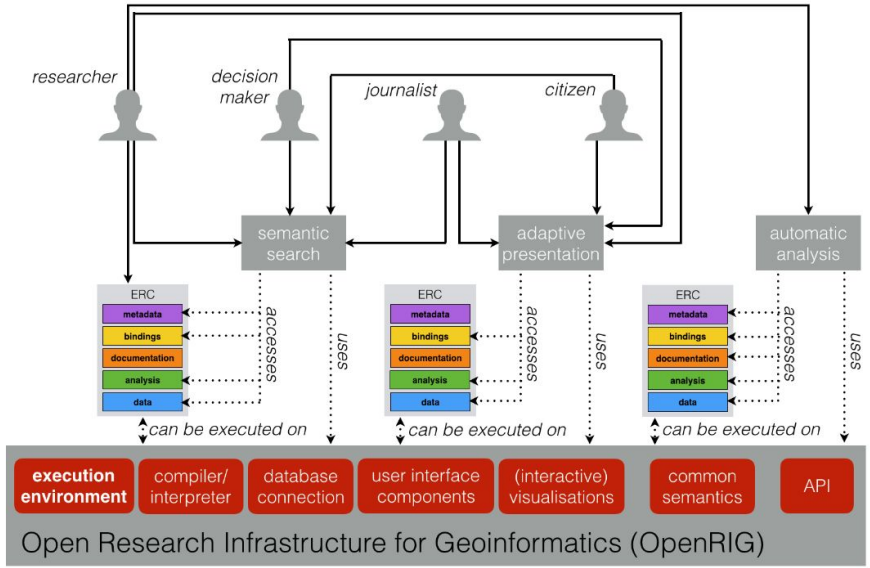


Figure 2 Open Research Infrastructure for Geoinformatics (OpenRIG): key components (red), essential functionalities enabled by it (grey boxes) and different stakeholders wanting to access them.

Reproducible Research in Geoinformatics: Concepts, Challenges and Benefits (Vision Paper) *Kray C, Pebesma E, Konkol M, Nüst D.* doi:10.4230/LIPIcs.COSIT.2019.8
GenR blog: https://genr.eu/wp/a-vision-for-reproducible-research-in-geoinformatics-geography-and-geosciences/

Thanks!

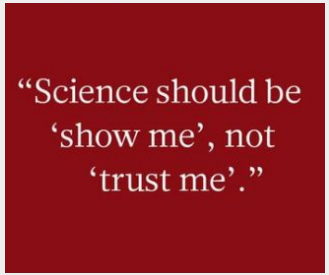
Daniel Nüst

Institute for Geoinformatics, University of Münster

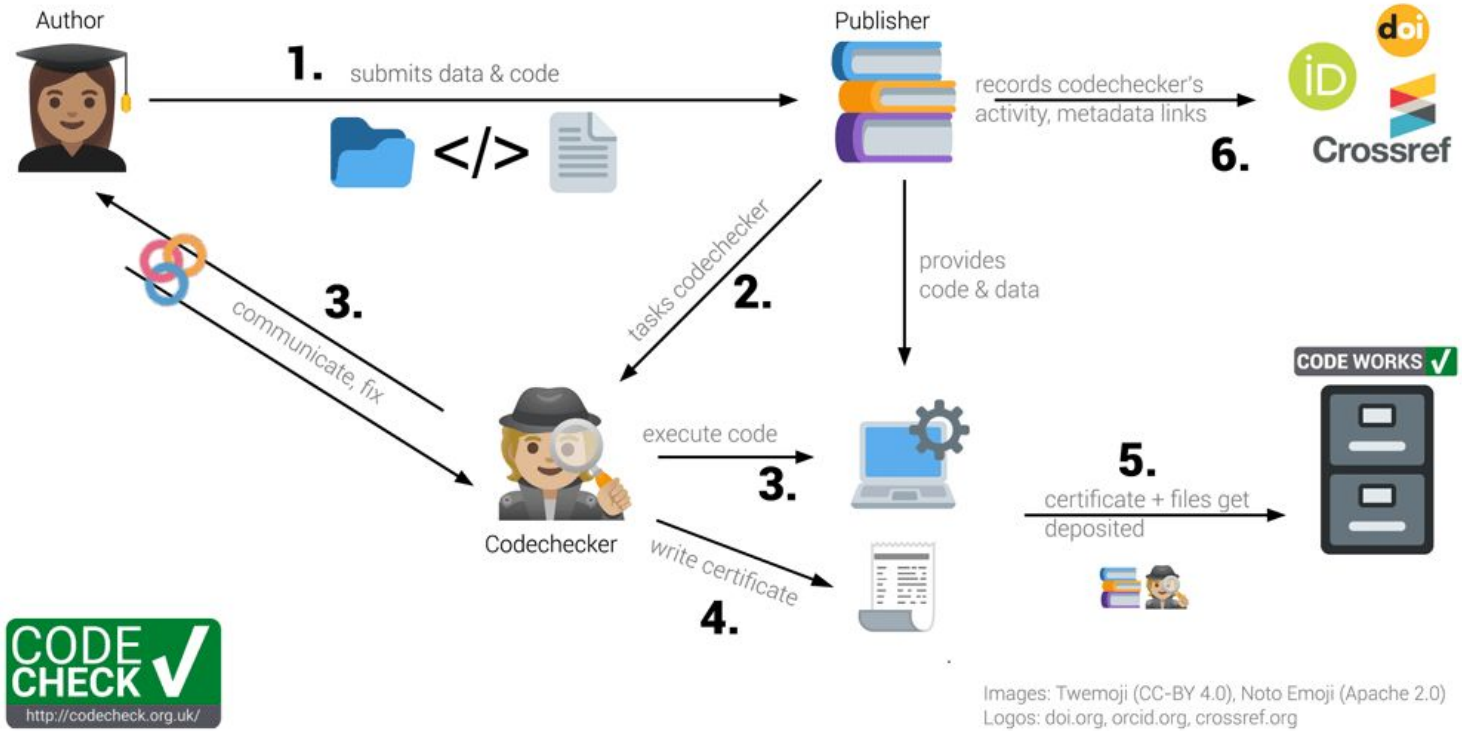
d.n@wwu.de | @nordholmen

0000-0002-0024-5046

Slides: https://doi.org/10.5281/zenodo.5006379



Bonus slides for discussion



The CODECHECK example process implementation. Figure 2 of https://doi.org/10.12688/f1000research.51738.1

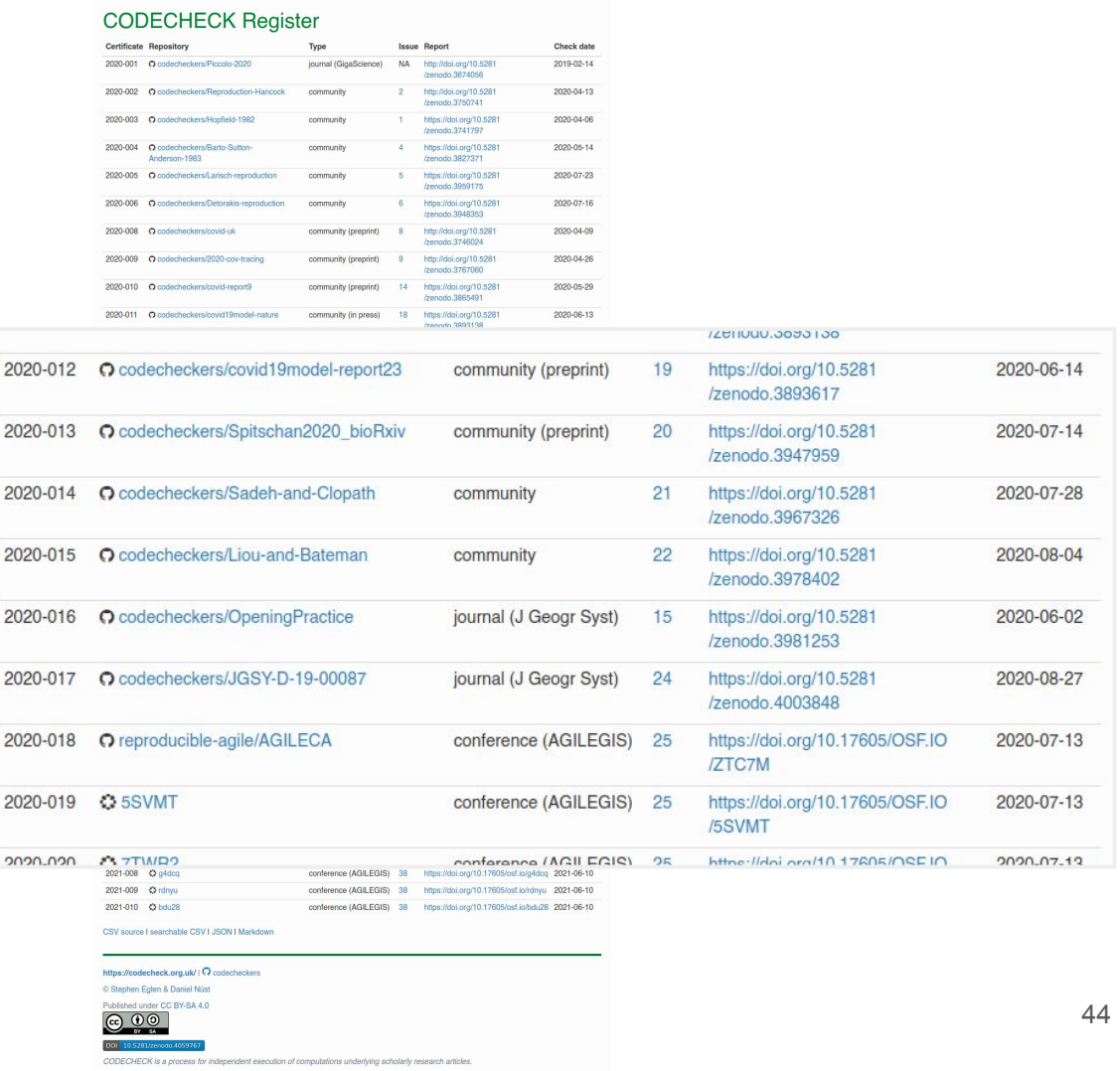
The left half of the diagram shows a diverse range of materials used within a laboratory. These materials are often then condensed for sharing with the outside world via the research paper, a static PDF document. Working backwards from the PDF to the underlying materials is impossible. This prohibits reuse and is not only non-transparent for a specific paper but is also ineffective for science as a whole. By sharing the materials on the left, others outside the lab can enhance this work.

https://codecheck.org.uk/process/

CODE CHECK

https://codecheck.org.uk/

Independent execution of computations underlying research articles.



CODECHECK Register

Certificate	Repository	Type	Issue	Report	Check date
2020-001	codecheckers/Piccolo-2020	journal (GigaScience)	NA	http://doi.org/10.5281/zenodo.3674056	2019-02-14
2020-002	codecheckers/Reproduction-Hancock	community	2	http://doi.org/10.5281/zenodo.3750741	2020-04-13
2020-003	codecheckers/Hopfield-1982	community	1	https://doi.org/10.5281/zenodo.3741797	2020-04-06
2020-004	codecheckers/Barto-Sutton-Anderson-1983	community	4	https://doi.org/10.5281/zenodo.3827371	2020-05-14
2020-005	codecheckers/Larisch-reproduction	community	5	https://doi.org/10.5281/zenodo.3959175	2020-07-23
2020-006	codecheckers/Detorakis-reproduction	community	6	https://doi.org/10.5281/zenodo.3948353	2020-07-16
2020-008	codecheckers/covid-uk	community (preprint)	8	http://doi.org/10.5281/zenodo.3746024	2020-04-09
2020-009	codecheckers/2020-cov-tracing	community (preprint)	9	http://doi.org/10.5281/zenodo.3767060	2020-04-26
2020-010	codecheckers/covid-report9	community (preprint)	14	https://doi.org/10.5281/zenodo.3865491	2020-05-29
2020-011	codecheckers/covid19model-nature	community (in press)	18	https://doi.org/10.5281/zenodo.3893138	2020-06-13
2020-012	codecheckers/covid19model-report23	community (preprint)	19	https://doi.org/10.5281/zenodo.3893617	2020-06-14
2020-013	codecheckers/Spitschan2020_bioRxiv	community (preprint)	20	https://doi.org/10.5281/zenodo.3947959	2020-07-14
2020-014	codecheckers/Sadeh-and-Clopath	community	21	https://doi.org/10.5281/zenodo.3967326	2020-07-28
2020-015	codecheckers/Liou-and-Bateman	community	22	https://doi.org/10.5281/zenodo.3978402	2020-08-04
2020-016	codecheckers/OpeningPractice	journal (J Geogr Syst)	15	https://doi.org/10.5281/zenodo.3981253	2020-06-02
2020-017	codecheckers/JGSY-D-19-00087	journal (J Geogr Syst)	24	https://doi.org/10.5281/zenodo.4003848	2020-08-27
2020-018	reproducible-agile/AGILECA	conference (AGILEGIS)	25	https://doi.org/10.17605/OSF.IO/ZTC7M	2020-07-13
2020-019	5SVMT	conference (AGILEGIS)	25	https://doi.org/10.17605/OSF.IO/5SVMT	2020-07-13
2020-020	7TWB2	conference (AGILEGIS)	25	https://doi.org/10.17605/OSF.IO	2020-07-13
2021-008	g4dcq	conference (AGILEGIS)	38	https://doi.org/10.17605/osf.io/g4dcq	2021-06-10
2021-009	rdnyu	conference (AGILEGIS)	38	https://doi.org/10.17605/osf.io/rdnyu	2021-06-10
2021-010	bdu28	conference (AGILEGIS)	38	https://doi.org/10.17605/osf.io/bdu28	2021-06-10

CSV source | searchable CSV | JSON | Markdown

https://codecheck.org.uk/ / codecheckers

© Stephen Eglen & Daniel Nüst

Published under CC BY-SA 4.0

DOI 10.5281/zenodo.4059767

CODECHECK is a process for independent execution of computations underlying scholarly research articles.

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

The many problems of science

Publish or perish
Broken metrics (citations, JIF)
Structural change not considering senior academics
Publication bias
Long-term funding for tools & infrastructure
HARKing
p-Hacking
Scholarly communication 1.0
Lack of reusability
Lack of transparency
Lack of reproducibility
Reinventing the wheel
Retraction practices
Not invented here syndrome
Fraud
Imposter syndrome
No “negative” citation
…

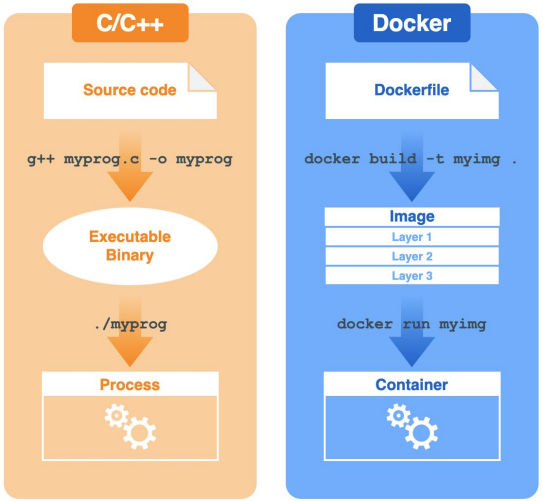
https://giphy.com/gifs/bbcamerica-cute-animals-lifestory-Ze3RpHue7qkwvcYOOf

Open Science (OER, OA, OS, OPR)
Registered reports/preregistration
Altmetrics
Preprints
Leiden Manifesto
DORA
Vienna Principles
Citing data and software
Software papers
Data and software as products of research
RSEng & RSEs (software sustainability)
CRediT
Research Compendia
Ten Hot Topics Around Scholarly Publishing
Code review (PyOpenSci, ROpenSci, JOSS)
…

Ten simple rules for writing Dockerfiles for reproducible data science

Daniel Nüst, Vanessa Sochat, Ben Marwick, Stephen J. Eglen, Tim Head, Tony Hirst, Benjamin D. Evans

Published: November 10, 2020 • https://doi.org/10.1371/journal.pcbi.1008316



© Benjamin D. Evans, 2020. This work is released under the CC-BY 4.0 license https://creativecommons.org/licenses/by/4.0/

Ten Simple Rules for Writing Dockerfiles for Reproducible Data Science

1. Use available tools
2. Build upon existing images
3. Format for clarity
4. Document within the Dockerfile
5. Specify software versions
6. Use version control
7. Mount datasets at run time
8. Make the image one-click runnable
9. Order the instructions
10. Regularly use and rebuild containers

© Benjamin D. Evans, 2020. This work is released under the CC-BY 4.0 license https://creativecommons.org/licenses/by/4.0/

Comment | Open Access | Published: 08 December 2015

Five selfish reasons to work reproducibly

Florian Markowetz

Genome Biology 16, Article number: 274 (2015) | Cite this article

15k Accesses | 28 Citations | 443 Altmetric | Metrics

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

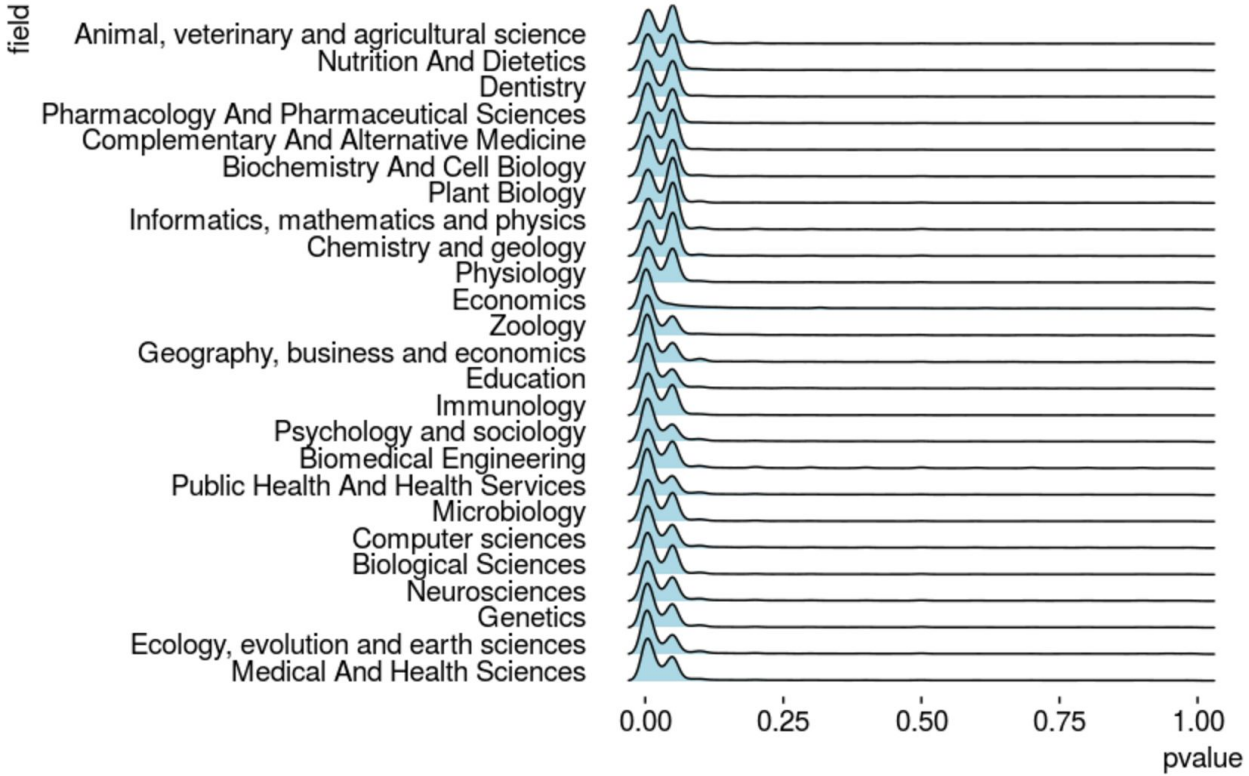
1. reproducibility helps to avoid disaster
2. reproducibility makes it easier to write papers
3. **reproducibility helps reviewers see it your way**
4. reproducibility enables continuity of your work
5. reproducibility helps to build your reputation

General observations and lessons learned

- **Further improvement over last years submissions - better prepared workflows!** Biggest hurdles remain: insufficient documentation, no “quick” variant or lack of expected data size/runtime, links Figures < > Scripts
- **Community understanding better, but needs time**: Had to remind authors to add DASA section - how can we be clearer in the communication? Camera-ready papers by authors possible, but exhausting.
- Additional **reproducibility questions for scientific reviewers worked better**, but triggering only by regular reviewers doesn’t work well - fortunately not too many submission to check for repro chair
- Repro reviews **were less strict than original ideal but on par with last year**
  > promote positive examples and don’t expect perfection
- **Non-blindness** served its purpose, but unblinding also delayed procedures
- **Schedule** still very much a challenge, partly because infrastructure (EasyChair) does not enable reviewer roles and communication > working around that with scripts and scraping
- **Improvements to process were good**: clarity in communication for authors that **DASA section is mandatory**, not attempting short papers, do not offer authors to object to report publications (no problems!)
- **Reproduction not attempted != bad science**, reproducibility is not binary but a ***spectrum***
  > continue education on reproducibility, increase requirements while practices spread in community

J. Leek's `tidypvals`

The tidypvals package is an effort to find previous collections of published p-values, synthesize them, and tidy them into one analyzable data set. The currently available p-value data sets in this package are:



"Notice Anything funny?"

https://**simplystatistics.org**/2017/07/26/announcing-the-tidypvals-package/

Code Review

Boettiger, C., Chamberlain, S., Hart, E., & Ram, K. (2015). Building Software, Building Community: Lessons from the rOpenSci Project. Journal of Open Research Software, 3(1), e8. doi:10.5334/jors.bu

Code Review Community Working Group

Reproducible computational research in journals & conferences

ACM Transactions on Mathematical Software

Journal of Statistical Software

ReproZip

Biostatistics

SC19 PROGRAM

Reproducibility Initiative

NEURAL INFORMATION PROCESSING SYSTEMS

Reproducibility review reports


Reproducibility review of: Integrating cellular automata and discrete global grid systems: a case study into wildfire modelling

Daniel Nüst

2020-06-20

REPRODUCIBLE AGILE

This report is part of the reproducibility review at the AGILE conference. For more information see https://reproducible-agile.gitlab.io/. This document is published on OSF at https://osf.io/ztc7m/. To cite the report use:

Nüst, D. (2020, June 5): Reproducibility review of: Integrating cellular automata and discrete global grid systems: a case study into wildfire modelling. https://doi.org/10.17605/OSF.IO/ZTC7M

Reviewed paper

Hojati, Majid and Roberts a case study into wildfire yse-1-5-2020, 2020.

Summary

The paper code and a sample da working Binder link. The workf the workflow does not create a the repository while further da *successful*, the authors demonst therefore include some commen authors after sending them a fi

wind_0.3_50.png

model-output-example.png

Reproducibility review of: Extracting interrogative intents and concepts from geo-analytic questions

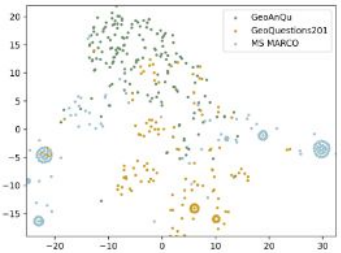
Daniel Nüst

19 Juni, 2020

REPRODU AGI

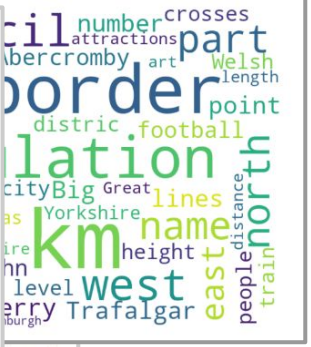
Reproduced Figure 6 (a)

Figure 10



Evaluation script

```
$ cd evaluation
$ python evaluation.py
[...]
$ ll -h
total 180K
drwxrwxr-x 2 daniel daniel 4,0K Jun  5 09:05 ./
drwxrwxr-x 9 daniel daniel 4,0K Dez 23 00:47 ../
-rw-rw-r-- 1 daniel daniel  22K Dez 23 00:47 1.csv
-rw-rw-r-- 1 daniel daniel  21K Dez 23 00:47 2.csv
-rw-rw-r-- 1 daniel daniel  22K Dez 23 00:47 3.csv
-rw-rw-r-- 1 daniel daniel  21K Dez 23 00:47 4.csv
```



https://doi.org/10.17605/OSF.IO/ZTC7M

https://doi.org/10.17605/OSF.IO/7XRQG

Reproducibility review reports

Reproducibility review:
"Tracking Hurricane Dorian in GDELT and Twitter"

This report is part of the reproducibility review at the AGILE conference.

For more information see https://reproducible-agile.github.io/.

This document is published on OSF at https://osf.io/xs5yr/.

To cite this report use

Ostermann
Hurricane

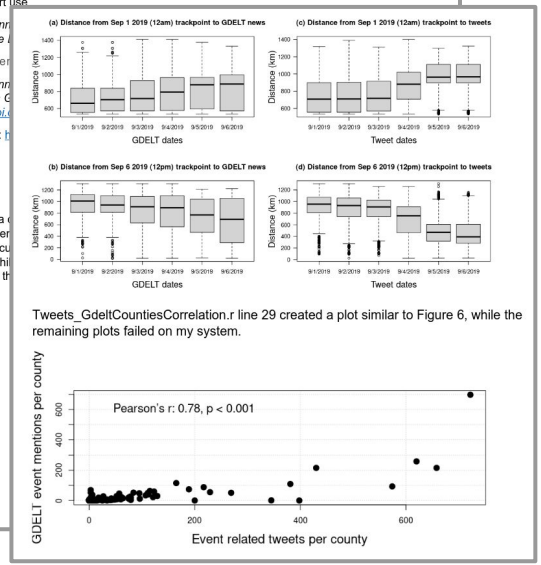
Reviewed paper

Owuor, Inn
Dorian in G
https://doi.

Code repository: h
Twitter

Summary

The authors do a
(given platform ter
absence of a docu
are to be run. Whi
reproduced with th
successful.



Tweets_GdeltCountiesCorrelation.r line 29 created a plot similar to Figure 6, while the remaining plots failed on my system.

https://doi.org/10.17605/OSF.IO/XS5YR

Reproducibility review:
"Comparing supervised learning algorithms for Spatial Nominal Entity recognition"

This report is part of the reproducibility review at the AG

For more information see https://reproducible-agile.githu

This document is published on OSF at https://osf.io/suw

To cite this report use

Ostermann, F. O., and Nüst, D. (2020, July). Re
supervised learning algorithms for Spatial Nomin
https://doi.org/10.17605/OSF.IO/SUWPJ

Reviewed paper

Amine Medad, Mauro Gaio, Ludovic Moncla, Sé
Comparing supervised learning algorithms for Sp
AGILE GiScience Ser., 1, 15. https://doi.org/10.5

Source code: https://github.com/MedadAmine/Spatial-n

Summary

The authors have done a commendable job at providing
documentation to run the analysis. The reproduction wa
computational environment required some initially undo
libraries used, which have now been documented. It sh
requires substantial downloads, disk space, and proces
reproduction was mostly successful.

Reproducibility reviewer notes

The materials on GitHub have an MIT license.

Data

Original hiking texts: not available, although there is a li

Lexicon: FastText freely available online

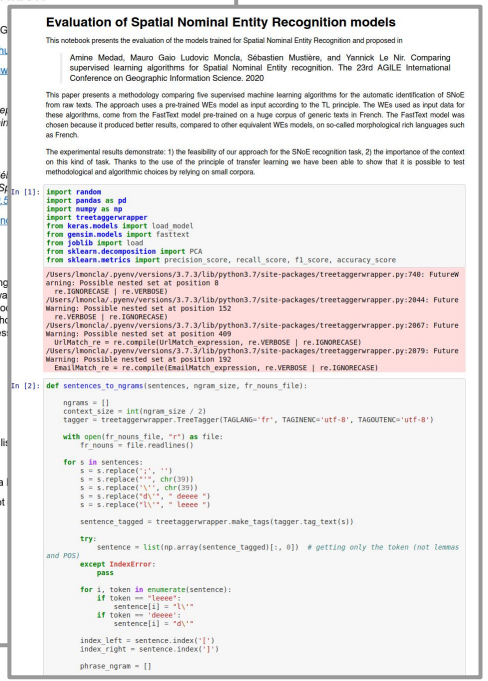
Corpus: entire corpus not available, although there is a

Samples for analysis available (named corpus), but not

Processing

- uses open source libraries

- Scripts and hyper-parameters are available



Evaluation of Spatial Nominal Entity Recognition models

This notebook presents the evaluation of the models trained for Spatial Nominal Entity Recognition and proposed in

> Amine Medad, Mauro Gaio Ludovic Moncla, Sébastien Mustière, and Yannick Le Nir. Comparing supervised learning algorithms for Spatial Nominal Entity recognition. The 23rd AGILE International Conference on Geographic Information Science. 2020

This paper presents a methodology comparing five supervised machine learning algorithms for the automatic identification of SNoE from raw texts. The approach uses a pre-trained WEs model as input according to the TL principle. The WEs used as input data for these algorithms, come from the FastText model pre-trained on a huge corpus of generic texts in French. The FastText model was chosen because it produced better results, compared to other equivalent WEs models, on so-called morphological rich languages such as French.

The experimental results demonstrate: 1) the feasibility of our approach for the SNoE recognition task, 2) the importance of the context on this kind of task. Thanks to the use of the principle of transfer learning we have been able to show that it is possible to test methodological and algorithmic choices by relying on small corpora.

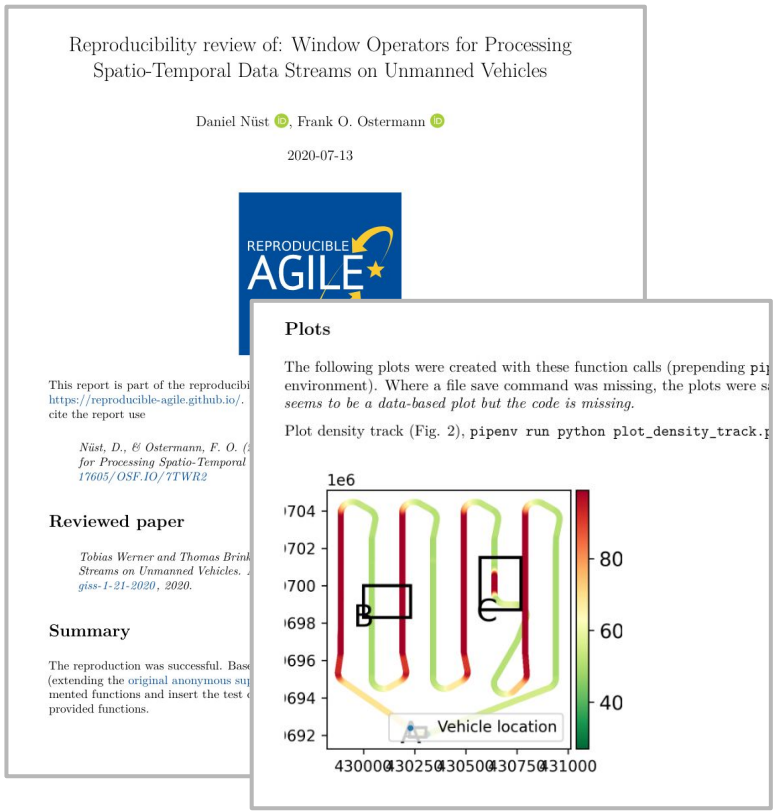
```
In [1]: import random
        import pandas as pd
        import numpy as np
        import treetaggerwrapper
        from keras.models import load_model
        from gensim.models import fasttext
        from joblib import load
        from sklearn.decomposition import PCA
        from sklearn.metrics import precision_score, recall_score, f1_score, accuracy_score
```

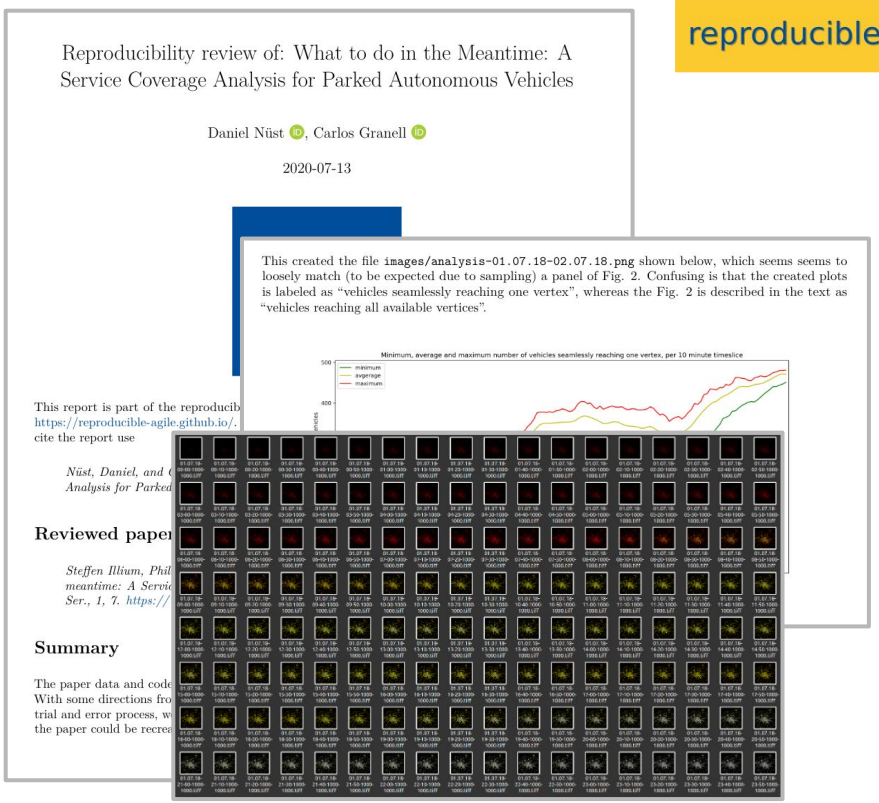
```
/Users/lmoncla/.pyenv/versions/3.7.3/lib/python3.7/site-packages/treetaggerwrapper.py:740: FutureW
arning: Possible nested set at position 8
  re.IGNORECASE | re.VERBOSE)
/Users/lmoncla/.pyenv/versions/3.7.3/lib/python3.7/site-packages/treetaggerwrapper.py:2044: Future
Warning: Possible nested set at position 152
  re.VERBOSE | re.IGNORECASE)
/Users/lmoncla/.pyenv/versions/3.7.3/lib/python3.7/site-packages/treetaggerwrapper.py:2067: Future
Warning: Possible nested set at position 409
  UrlMatch_re = re.compile(UrlMatch_expression, re.VERBOSE | re.IGNORECASE)
/Users/lmoncla/.pyenv/versions/3.7.3/lib/python3.7/site-packages/treetaggerwrapper.py:2079: Future
Warning: Possible nested set at position 192
  EmailMatch_re = re.compile(EmailMatch_expression, re.VERBOSE | re.IGNORECASE)
```

```
In [2]: def sentences_to_ngrams(sentences, ngram_size, fr_nouns_file):

            ngrams = []
            context_size = int(ngram_size / 2)
            tagger = treetaggerwrapper.TreeTagger(TAGLANG='fr', TAGINENC='utf-8', TAGOUTENC='utf-8')

            with open(fr_nouns_file, "r") as file:
                fr_nouns = file.readlines()

            for s in sentences:
                s = s.replace(';', '')
                s = s.replace('"', chr(39))
                s = s.replace('\'', chr(39))
                s = s.replace("d\'", " deeee ")
                s = s.replace("l\'", " leeee ")

                sentence_tagged = treetaggerwrapper.make_tags(tagger.tag_text(s))

                try:
                    sentence = list(np.array(sentence_tagged)[:, 0])  # getting only the token (not lemmas
and POS)
                except IndexError:
                    pass

                for i, token in enumerate(sentence):
                    if token == "leeee":
                        sentence[i] = "l\'"
                    if token == 'deeee':
                        sentence[i] = "d\'"

                index_left = sentence.index('[')
                index_right = sentence.index(']')

                phrase_ngram = []
```

https://doi.org/10.17605/OSF.IO/SUWPJ

Reproducibility review reports





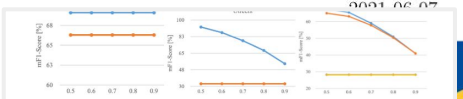
https://doi.org/10.17605/OSF.IO/7TWR2

https://doi.org/10.17605/OSF.IO/5SVMT

Reproducibility review reports AGILE 2021

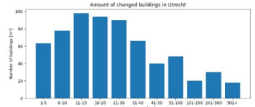
Reproducibility review of: Building Change Detection of Airborne Laser Scanning and Dense Image Matching Point Clouds using Height and Class Information

Philipp A. Friese



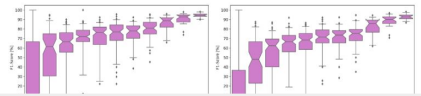
Graphic 1: Mean F1 Scores from Excel Sheet - corresponds to Figure 2-7 and methodologically to 12-15 in reproduced paper

After the authors provided an additional visualization script Figure 1 was reproduced. The generated image is shown in graphic 2.



Graphic 2: Generated Amounts of changed buildings - corresponds to Figure 1 in reproduced paper

The script generates Figure 10 and 11 automatically for each parameter permutation. An excerpt of the generated images are shown in graphic 3.



ILE conference. For more information see
lished on OSF at https://osf.io/RSF4M/.

Reviewed paper

Politz, F., Sester, M., and Bre
ning and Dense Image Matchi
GIScience Ser., 2, 10, https://

Graphic 4: Generated Prediction Images, visualised using QGIS, parameters 'jsd prob ct 0.7' - corresponds to Figure 8, 9, and 16 in reproduced paper

https://doi.org/10.17605/OSF.IO/RSF4M

Reproducibility review of: Investigating drivers' geospatial abilities in unfamiliar environments

Initial execution of the analysis script raised errors while generating Fig. 4. This was however resolved after contacting the authors and Fig. 4 was reproduced. The generated image is shown in Figure 1.

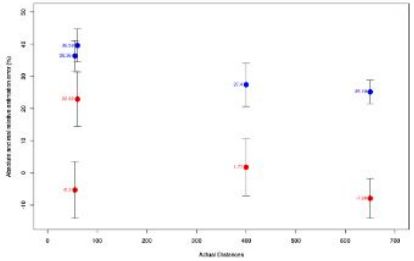


Figure 1: Average absolute (blue) and real relative (red) estimation error - corresponds to Figure 1 in reproduced paper

Table 1 and 2 were not generated automatically but instead the corresponding values where printed to the console. After extracting these values and exporting them to .csv files, both tables were reproduced. Refer to Table 1 and 2 in this report, which contain the raw values generated by the script.

Table 1: Spearman correlations Group 1 - corresponds to Table 1 in reproduced paper

type	S statistic	rho	p-value
SBSOD - Map Errors	463.24	-0.6197337	0.03160
SBSOD – Landmarks omitted	364.51	-0.2745132	0.38790
SBSOD – Road Segments mistakes	471.27	-0.6477894	0.02274
Landmarks omitted– Road Segments mistakes	160.86	0.4375473	0.15490
SBSOD – Direction estimates	233.45	0.1837559	0.56750
SBSOD – Distance estimates	342.00	-0.5545455	0.07665
Map Errors – Direction estimates	278.88	0.0249112	0.93870
Map Errors – Distance estimates	102.19	0.5354817	0.08958
Distance estimates – Direction estimates	205.81	0.0645223	0.85050

Table 2: Spearman correlations Group 2 - corresponds to Table 2 in reproduced paper

type	S statistic	rho	p-value
SBSOD - Map Errors	578.43	-0.5890887	0.03414
SBSOD – Landmarks omitted	520.67	-0.4304142	0.14210

This report is part of the reprodu
https://reproducible-agile.github.io
cite the report use

Friese, Philipp A. (2021, May
abilities in unfamiliar environ

Reviewed paper

Karkasina, D., Kokla, M., and
iar environments, AGILE GI
2021.

https://doi.org/10.17605/OSF.IO/DX92A

Reproducibility review reports AGILE 2021

Reproducibility review of: A Comparative Study of Typing and Speech For Map Metadata Creation

F.O. Ostermann and Daniel Nüst

Reproducibility reviewer notes

This review focuses on the reproduction of the analysis results. No in-depth examination
code was conducted, but it was confirmed that the provided code and be run and see
application used in the study. Using

```
# with npm version 6.14.8 and node version 14.13.0
npm install
npm start
```

we could run the application on http://localhost:8080, as shown in the screenshot bel

Figure 1: Screenshot of application executed locally

A clear license is missing in the repository. The most important information (soft
overview, exact questionnaire, maps used in the experiment) is also provided as supple

This report is part of the reproducibili
https://reproducible-agile.github.io/. T
cite the report use

Ostermann, F. O., & Nüst, D. (202
of Typing and Speech For Map M

Reviewed paper

Lai, P.-C. and Degbelo, A.: A Comparative Study of Typing and Speech For Map Metadata
Creation, AGILE GIScience Ser., 2, 7, https://doi.org/10.5194/agile-giss-2-7-2021, 2021.

Summary

The paper presents the results of a user experiment to improve GI-metadata using speech. A complete
reproduction is practically impossible to achieve. This reproducibility report therefore investigated two
components: First, whether sufficient information is provided to replicate the experiment elsewhere with

https://doi.org/10.17605/osf.io/7fqtm

Reproducibility review of: A Socially Aware Huff Model for Destination Choice in Nature-based Tourism

Jakub Krukar

Construction.ipynb and can be verified by comparing the values from the output of chunk [12] to
Table 8 in the paper. I expected that the values in the `Number of photos` column would stay the same
(because the number of photos does not change), but the values in `outgoing/incoming trips` columns
would change.

The location of the threshold variable is marked with an in-code comment in the file `Trip Construction.ipynb`:

```
if length.days > 4: #time threshold: average length of stay in both NPs
```

Table 8 Summary of attractions in Acadia National Park

Attraction	Number of photos	Outgoing trips	Incoming trips
Schoodic Institute	1119	53	64
Bass Harbor	2298	260	288
Southwest Harbor	723	109	111
Northeast Harbor	605	67	76
Bar Harbor	6259	433	357
Wild Gardens of Acadia	550	60	66
Cadillac Mountain	3285	349	345
Penobscot Peak	776	16	15
Bubble Rock	703	83	89
Jordan Pond	1250	227	250
Boulder Beach	536	85	102
Thunder Hole	977	167	185
Sand Beach	1253	216	177

Figure 1: Original Table 8 from the paper.

Places	a	b	c	d	e	f	g	h	i	j	k	l	m	total_out	total_in	cross_boundary	photos
Schoodic Institute	0	13	7	1	12	1	8	0	0	4	3	2	6	57	66	123	1119
Bass Harbor	12	0	34	9	64	13	53	4	8	25	12	15	21	268	295	563	2298
Southwest Harbor	3	44	0	6	30	3	15	4	1	4	1	2	2	115	117	232	723
Northeast Harbor	5	16	8	0	13	1	7	0	2	10	1	2	3	68	78	146	605
Bar Harbor	20	60	25	21	0	17	118	3	12	50	15	40	56	437	367	804	6259
Wild Gardens of Acadia	1	3	1	2	10	0	6	1	1	6	4	11	15	61	67	128	550
Cadillac Mountain	8	57	12	13	102	16	0	0	14	51	12	24	45	354	350	704	3285
Penobscot Peak	2	3	3	2	2	0	0	0	0	2	0	1	1	16	15	31	776

This report is part of the reproducil
https://reproducible-agile.github.io/.
the report use

Krukar, J. (2021, May 7). Reproducil
in Nature-based Tourism. https://dc

Reviewed paper

Shi, M., Janowicz, K., Cai, L., Mai, G
in Nature-based Tourism, AGILE G
2021.

Summary

The code, sample API query, and dov
working Binder link. All files containi

https://doi.org/10.17605/OSF.IO/4CPM3

Reproducibility review reports AGILE 2021

Reproducibility review of: Automated Extraction of Labels from Large-Scale Historical Maps

Daniel Nüst

2021-06-07

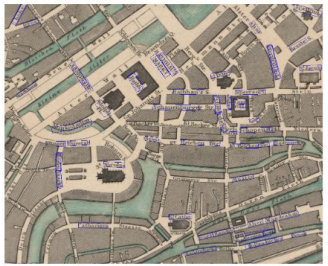


Figure 1: One output file from text recognition run.

I could run all cells in `String_Similarity_by_Levenshtein_Distance.ipynb`, and found the documentation to be extensive and a little bit raw, though very transparent, including tests by the author while developing the workflow etc. At first I got an error reading the `OCR_results.xlsx` file: `XLRDError: Excel xlsx file; not supported`, so I changed the data loading to use `openpyxl`.

Figure 6 seems to be created using QGIS, but no project file or georeferenced version of the base map was included in the repository.

I did not run the final notebook of step "06 Approximate georeferencing", because of the advertised run

This report is part of the reproduc
https://reproducible-agile.github.io
cite the report use

Nüst, D. (2021, May 6). Rep
Large-Scale Historical Maps.

Reviewed paper

Schlegel, I.: Automated Ext
GIScience Ser., 2, 12, https:/

Summary

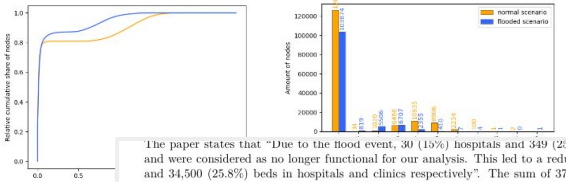
The provided workflow could be p
calculations could be executed and
providing all input and output data for each step, however, some manual steps were included which

Reproducibility review of: Flood Impact Assessment on Road Network and Healthcare Access – at the example of Jakarta, Indonesia

Anita Graser

2021-06-07

```
CDF results saved.
Histogram results saved.
```



This report is part
https://reproducibl
cite the report use

Graser, A. (20
Network and H
17605/OSF.IO

The resulting CDF plots
Therefore, it is not strai

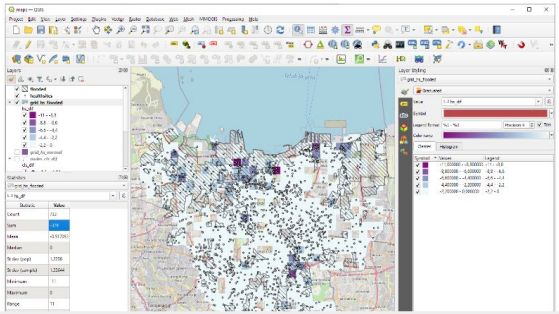
Reviewed paper

Klipper, I. G., Zipf, A., and Lautenbach, S.: Flood
Access at the example of Jakarta, Indonesia, AGI
giss-2-4-2021, 2021.

Summary

The provided workflow was **partially reproduc**
multiple sources (a Github.com repo, a GitLab r

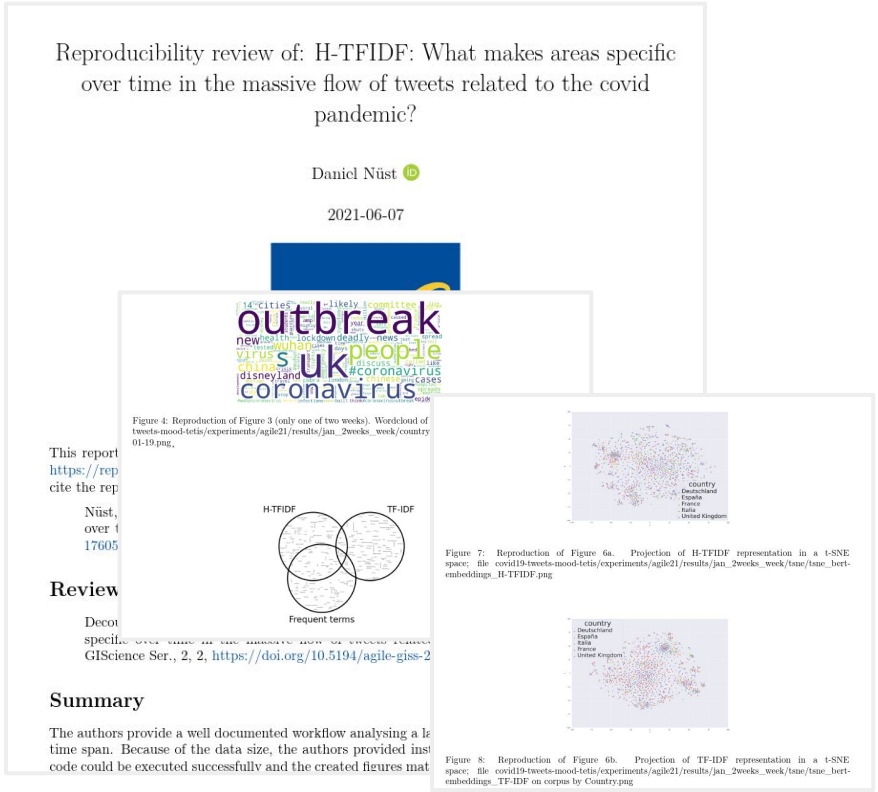
The paper states that "Due to the flood event, 30 (15%) hospitals and 349 (25%) clinics were affected and were considered as no longer functional for our analysis. This led to a reduction of 12,000 (16.6%) and 34,500 (25.8%) beds in hospitals and clinics respectively". The sum of 379 affected heath service locations could be confirmed, as shown in the following screenshot:



https://doi.org/10.17605/osf.io/anv9r

https://doi.org/10.17605/OSF.IO/G4DCQ

Reproducibility review reports AGILE 2021



Reproducibility review of: H-TFIDF: What makes areas specific over time in the massive flow of tweets related to the covid pandemic?

Daniel Nüst

2021-06-07

Figure 4: Reproduction of Figure 3 (only one of two weeks). Wordcloud of tweets-mood-tetis/experiments/agile21/results/jan_2weeks_week/country 01-19.png.

Figure 7: Reproduction of Figure 6a. Projection of H-TFIDF representation in a t-SNE space; file covid19-tweets-mood-tetis/experiments/agile21/results/jan_2weeks_week/tsne/tsne_bert-embeddings_H-TFIDF.png

Figure 8: Reproduction of Figure 6b. Projection of TF-IDF representation in a t-SNE space; file covid19-tweets-mood-tetis/experiments/agile21/results/jan_2weeks_week/tsne/tsne_bert-embeddings_TF-IDF on corpus by Country.png

This report
https://rep
cite the rep

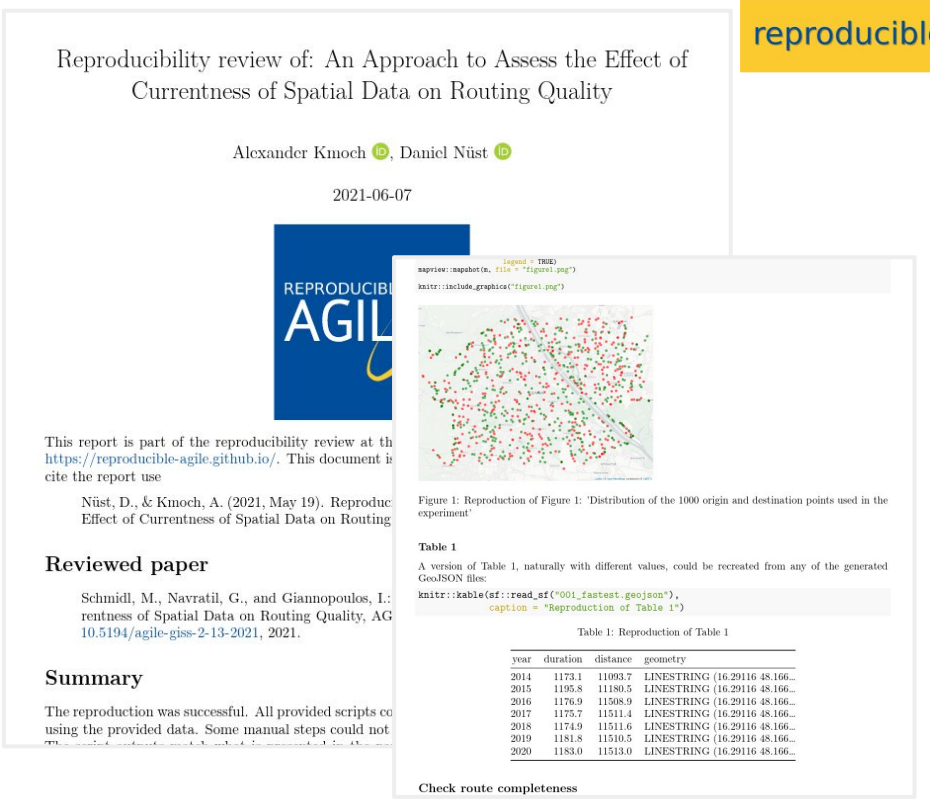
Nüst,
over t
17605

Review

Decou
specia
GIScience Ser., 2, 2, https://doi.org/10.5194/agile-giss-2

Summary

The authors provide a well documented workflow analysing a la
time span. Because of the data size, the authors provided inst
code could be executed successfully and the created figures mat



Reproducibility review of: An Approach to Assess the Effect of Currentness of Spatial Data on Routing Quality

Alexander Kmoch, Daniel Nüst

2021-06-07

This report is part of the reproducibility review at th
https://reproducible-agile.github.io/. This document i
cite the report use

Nüst, D., & Kmoch, A. (2021, May 19). Reproduc
Effect of Currentness of Spatial Data on Routing

Reviewed paper

Schmidl, M., Navratil, G., and Giannopoulos, I.:
rentness of Spatial Data on Routing Quality, AG
10.5194/agile-giss-2-13-2021, 2021.

Summary

The reproduction was successful. All provided scripts co
using the provided data. Some manual steps could not

```
legend = TRUE)
mapview::mapshot(m, file = "figure1.png")
knitr::include_graphics("figure1.png")
```


Figure 1: Reproduction of Figure 1: 'Distribution of the 1000 origin and destination points used in the experiment'

Table 1

A version of Table 1, naturally with different values, could be recreated from any of the generated GeoJSON files:

```
knitr::kable(sf::read_sf("001_fastest.geojson"),
             caption = "Reproduction of Table 1")
```

Table 1: Reproduction of Table 1

year	duration	distance	geometry
2014	1173.1	11093.7	LINESTRING (16.29116 48.166...
2015	1195.8	11180.5	LINESTRING (16.29116 48.166...
2016	1176.9	11508.9	LINESTRING (16.29116 48.166...
2017	1175.7	11511.4	LINESTRING (16.29116 48.166...
2018	1174.9	11511.6	LINESTRING (16.29116 48.166...
2019	1181.8	11510.5	LINESTRING (16.29116 48.166...
2020	1183.0	11513.0	LINESTRING (16.29116 48.166...

Check route completeness

https://doi.org/10.17605/rdnyu

https://doi.org/10.17605/bdu28

Reproducibility review reports AGILE 2021

Reproducibility review of: Extraction of linear structures from digital terrain models using deep learning

Daniel Nüst and Anita Graser

2021-06-07

This report is part of the reproducibility review at the AGILE conference. For more inf
https://reproducible-agile.github.io/. This document is published on OSF at https://osf.i
cite the report use

> Nüst, D., & Graser, A. (2021, April 30). Reproducibility review of: Extraction o
> structures from digital terrain models using deep learning. https://doi.org/10.17605
> 2sc7g

Reviewed paper

> Satari, R., Kazimi, B., and Sester, M.: Extraction of linear structures from digital
> models using deep learning, AGILE GIScience Ser., 2, 11, https://doi.org/10.5194/ag
> 2-11-2021, 2021.

Summary

The provided workflow was **partially reproduced**. Based on the provided test file and i
was able to recreate the computing environment and run the segmentation models. Relevan
the paper could be recreated. The training and validation part of the workflow is irreprodu

This finished within a minute! These values match the column `SegNet` of Table 1, within a level of precision to be expected from such a classification.

```
hrnet <- read.csv("agile-submission-2021/HRNetBinarySegmentation/files/evaluation_file.csv")
segnet <- read.csv("agile-submission-2021/SegnetBinarySegmentation//files/evaluation_file.csv")

suppressPackageStartupMessages(library("tidyverse"))
dplyr::full_join(hrnet, segnet) %>%
  knitr::kable()
```

acc	f1_m	loss	model_type	precision_m	recall_m	specific_name
0.9069678	0.7762662	0.2579104	hrnet	0.8187426	0.7393778	simple_binary
0.8787796	0.6977567	0.3094499	segnetCustomized	0.7688990	0.6399856	simple_binary

Run the next segmentation:

```
cd multiclassSegmentation
python3 evaluate.py --evaluation_file=evaluation_file.csv
```

This completes and recreates the data in Table 3 within reasonable numerical precision based on the file `multiClassEvaluation.csv`. It is unclear to me how Table 2 can be constructed from `evaluation_file.csv` of this segmentation, but I assume it can be.

```
multi <- read.csv("agile-submission-2021/multiclassSegmentation/files/multiClassEvaluation.csv")

rows <- lapply(c(0:5), function(class) {
  classValues <- multi %>%
    dplyr::select(dplyr::ends_with(as.character(class)))
  names(classValues) <- c("sparse_iou", "prediction", "recall", "f1.score", "support")
  c('Class label' = as.character(class), classValues)
})

dplyr::bind_rows(rows) %>%
  knitr::kable()
```

Class label	sparse_iou	prediction	recall	f1.score	support
0	0.8952831	0.9270560	0.9630006	0.9446866	12342971
1	0.2073642	0.4009994	0.3745212	0.3873083	247492
2	0.0324831	0.3139039	0.0373107	0.0666941	282037
3	0.7648531	0.8797419	0.8552982	0.8673479	1961849
4	0.4453632	0.6953347	0.5580455	0.6191711	830866
5	0.1593569	0.2697686	0.2787574	0.2741893	145345

https://doi.org/10.17605/osf.io/2sc7g

Structural challenges

Metrics for acknowledging/measuring impact in science **are broken** (impact factor, ..) and they lead to publication bias, HARKing, p-Hacking, intransparency and lack of reproducibility

Leiden Manifesto: **http://www.leidenmanifesto.org**
DORA: **https://sfdora.org**
Vienna Principles: **https://viennaprinciples.org**

Acknowledging data and software as valuable products of research (instead of shoehorning software into papers)

Essays

Data Without Software Are Just Numbers

Authors: James Harold Davenport, James Grant, Catherine Mary Jones

http://doi.org/10.5334/dsj-2020-003

An excess of positive results: Comparing the standard Psychology literature with Registered Reports

Anne M. Scheel[1], Mitchell Schijen[1], & Daniël Lakens[1]

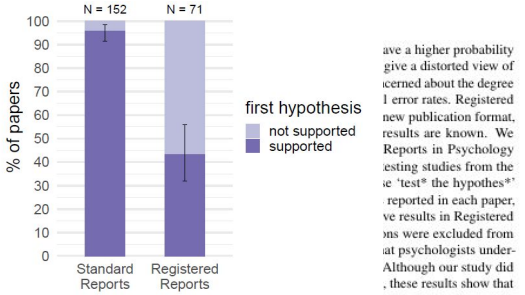
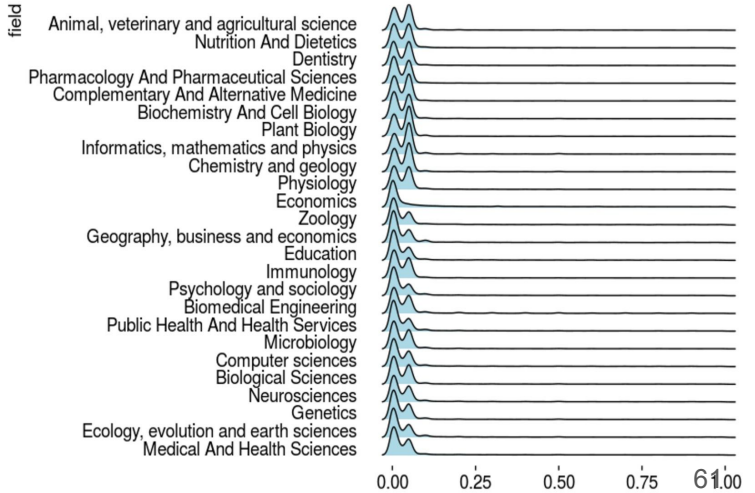


Figure 2. Positive result rates for standard reports and Registered Reports. Error bars indicate 95% confidence intervals around the observed positive result rate.

ave a higher probability
give a distorted view of
icerned about the degree
l error rates. Registered
new publication format,
results are known. We
Reports in Psychology
esting studies from the
e 'test* the hypothes*'
reported in each paper,
ve results in Registered
ons were excluded from
at psychologists under-
Although our study did
, these results show that
tion of negative results

https://doi.org/10.31234/osf.io/p6e9c



https://simplystatistics.org/2017/07/26/announcing-the-tidypvals-package/

nature International weekly journal of science

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

I am a professional software engineer and I want to share a trade secret with scientists: most professional computer software isn't very good. The code inside your laptop, television, phone or car is often badly documented, inconsistent and poorly tested.

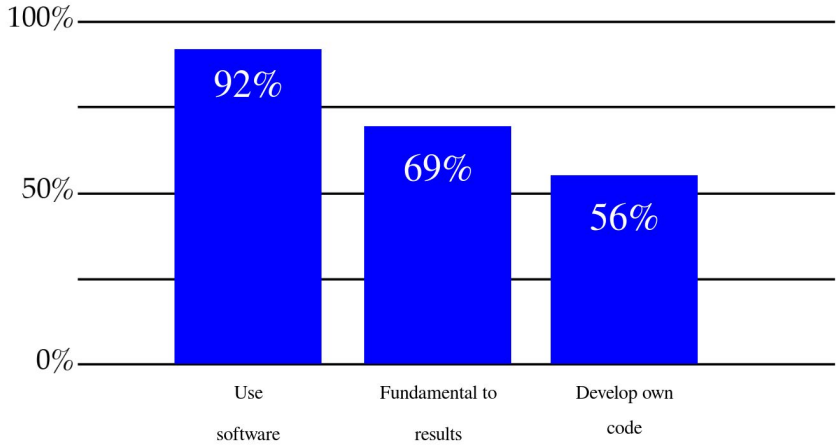
Why does this matter to science? Because to turn raw data into published research papers often requires a little programming, which means that most scientists write software. And you scientists generally think the code you write is poor. It doesn't contain good comments, have sensible variable names or proper indentation. It breaks if you introduce badly formatted data, and you need to edit the output by hand to get the columns to line up. It includes a routine written by a graduate student which you never completely understood, and so on. Sound familiar? Well, those things don't matter.

https://doi.org/10.1038/467753a

Motivation

Back to 2010 The Software Sustainability Institute (SSI, UK) run a study (1000 randomly chosen researchers) …

“It's impossible to conduct research without software, say 7 out of 10 UK researchers”



https://www.software.ac.uk/blog/2014-12-04-its-impossible-conduct-research-without-software-say-7-out-10-uk-researchers

Motivation

A study of Nature papers from Jan-March 2016 reveals that

“32 of the 40 papers examined mention software, and the 32 papers contain 211 mentions of distinct pieces of software, for an average of 6.5 mentions per paper.”

[2] Nangia, Udit; Katz, Daniel S. (2017): Understanding Software in Research: Initial Results from Examining Nature and a Call for Collaboration. doi:**10.1109/eScience.2017.78**

Image: Event Horizon Telescope

RSEng = create research software
RSEs = people behind research software
RSEs ≠ IT !!!

Reproducibility guru dives deeply into manifold software and tools to make his research reproducible and develops his own software in a sustainable way.

Researcher uses scripts for data analysis and needs working stable software for her work. She learns what is necessary to achieve her research goals.

”Software is 95% human and only 5% code” *

Person for tough problems knows how to solve all kinds of computer-related issues; he was not hired for that, but enjoys to help and spends time to get to the bottom of other people’s challenges.

Software developer was hired to implement software for a research project and contributes to large collaborative software projects to realise the next generation of digital infrastructure for science.

Geek writes software as part of her research project and would like to code more, but must keep an eye on her career in science and needs to write papers.

* Eric Albers, CCC2019, https://media.ccc.de/v/thms-49-ber-die-nachhaltigkeit-von-software

Bilder © H. Seibold, S. Janosch, OSD2019