
# Reproducible research in GIScience

**7th AGILE Doctoral School**

29 November, 2024

J Rafael Verduzco-Torres (University of Glasgow)

With work from: F Osternam (UT), C Granell (UJ), D Nüst (TUD), more…

University of Glasgow

Urban Big Data Centre

UNIVERSITAT JAUME I

AGILE Association of Geographic Information Laboratories in Europe

# Contents

# What is it?

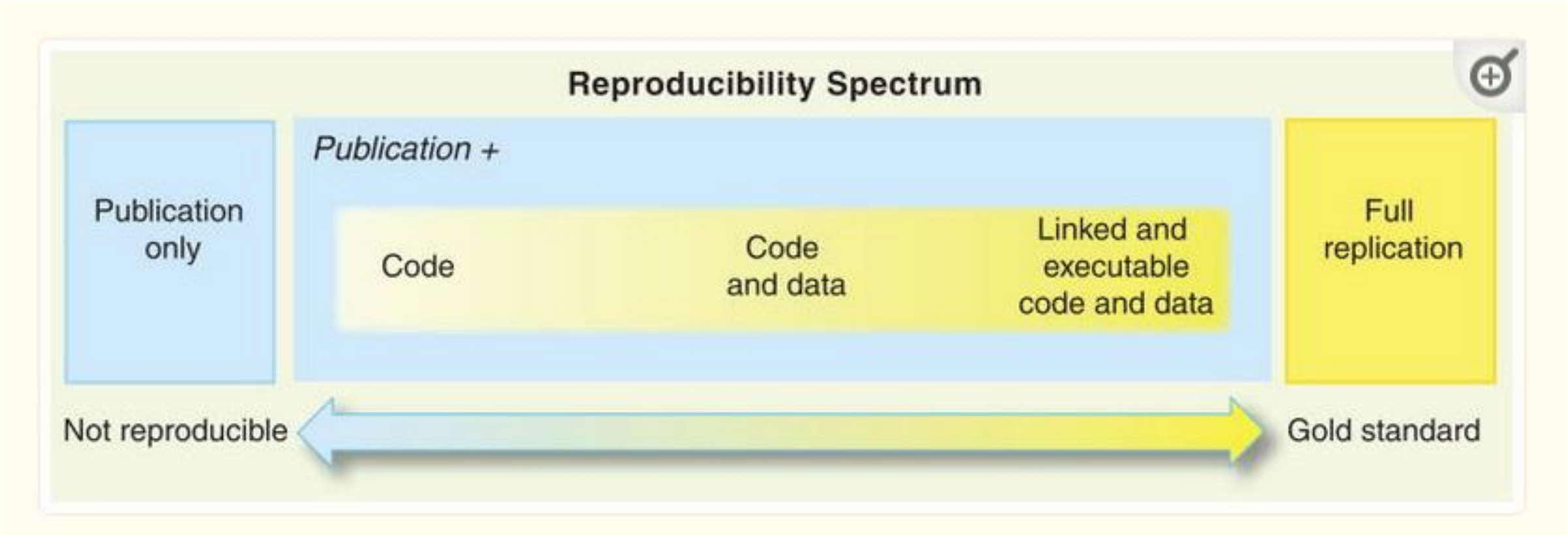
| | | Data | |
|---|---|---|---|
| | | Same | Different |
| Analysis | Same | Reproducible | Replicable |
| | Different | Robust | Generalisable |

The Turing Way Community. (2022). The Turing Way: A handbook for reproducible, ethical and collaborative research (1.0.2). Zenodo. https://doi.org/10.5281/zenodo.7625728

# Coupled with:

- **Open-access data**: Data that is freely available to *use and share*
- **Open-source software**: Software that is free to *use* and *modify*
- **Open-access tools**: Web applications that are based on open source software, that *anyone can use*

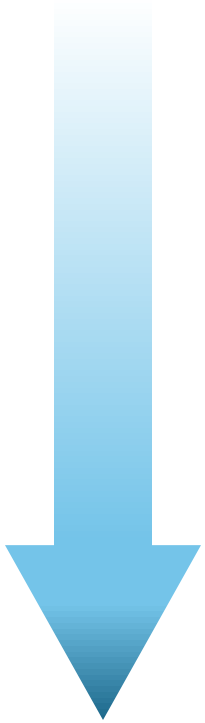
Lovelace (2024). https://www.robinlovelace.net/




Peng, Roger D. 2011. “Reproducible Research in Computational Science.” *Science (New York, N.y.)* 334 (6060): 1226–27. https://doi.org/10.1126/science.1213847.

# Criteria for reproducible research in GIScience

**0 (Low)**



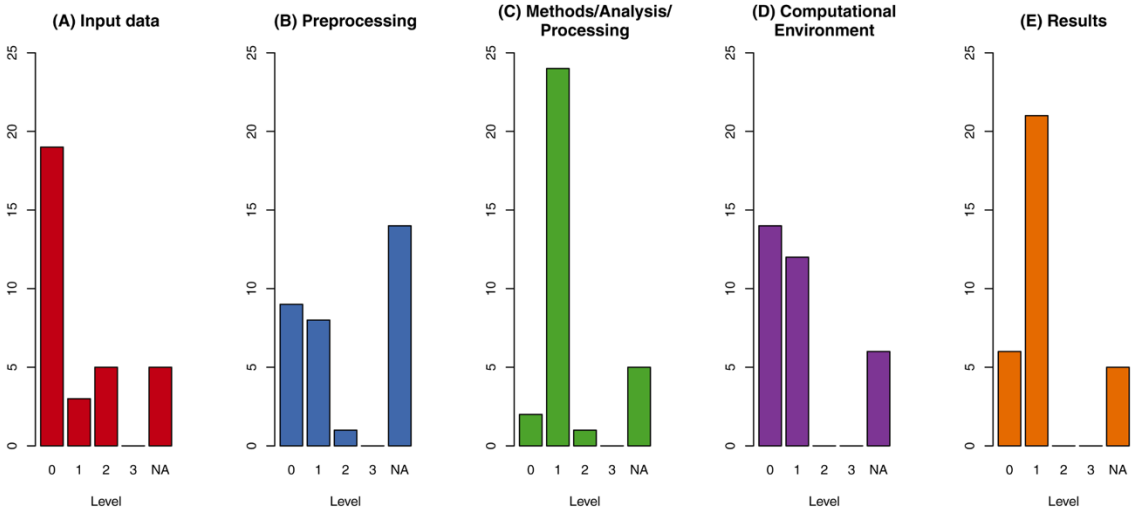
**3 (High)**

| **Data** | **Methods**<br>Preprocessing, analysis, computational environment | **Results** |
|---|---|---|
| [**0**] Unavailable and not recreatetable<br>[**1**] Documented and recreatetable<br>[**2**] Available, but not public licence or permanent<br>[**3**] Available, open and permanent | [**0**] Unavailable<br>[**1**] Documented (text, pseudo code, workflow descriptions)<br>[**2**] Available, e.g. source code<br>[**3**] Available and open, e.g. runtime, image/container | [**0**] Unavailable or insufficient<br>[**1**] Documented, i.e. understandable, context<br>[**2**] Available, e.g. models, scripted plots, output data<br>[**3**] Available, open, and permanent |

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

# What is the landscape in GIScience?

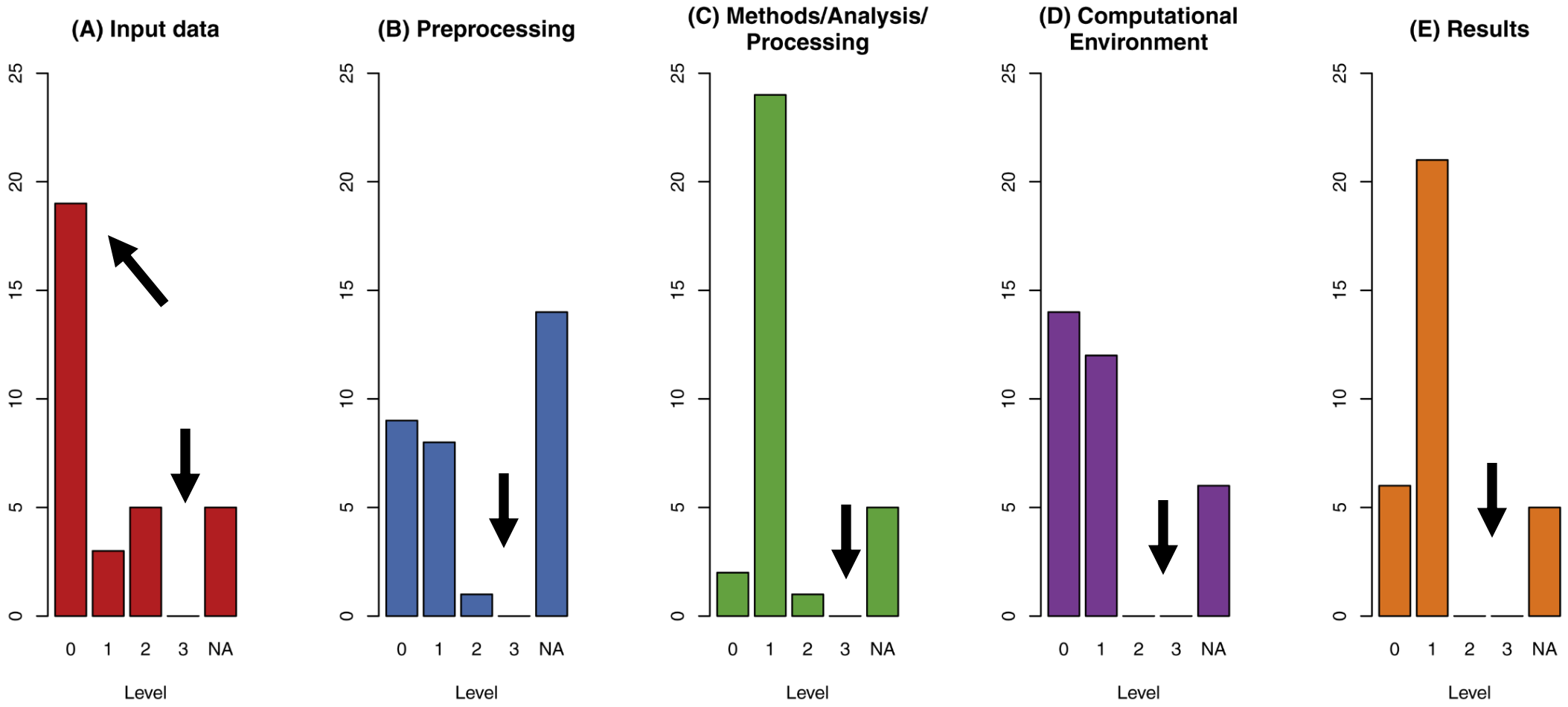
A sample (N=32) of nominated submissions for best paper to AGILE Conference (2010-2017)




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

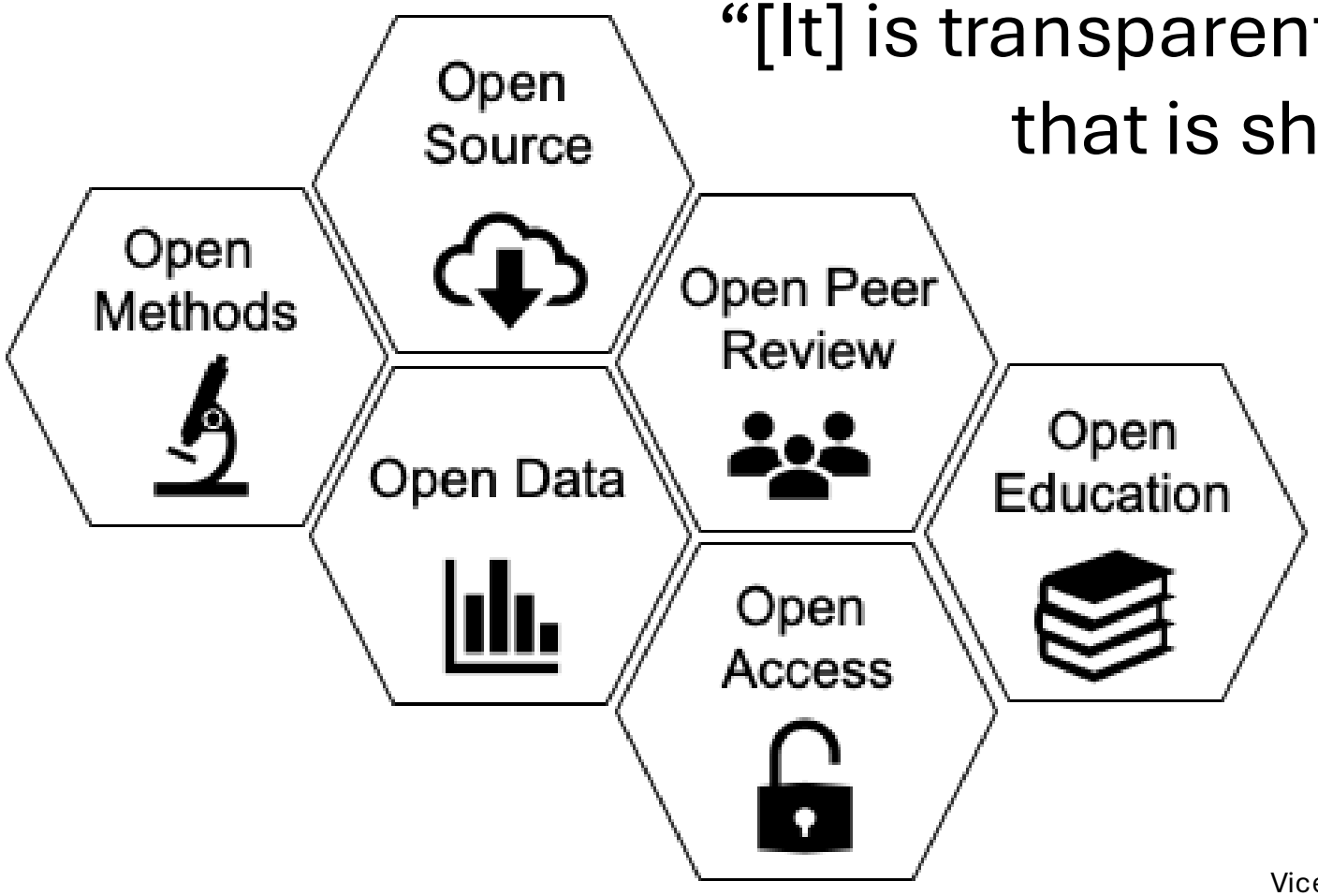
# What is the landscape in GIScience?

A sample (N=32) of nominated submissions for best paper to AGILE Conference (2010-2017)




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

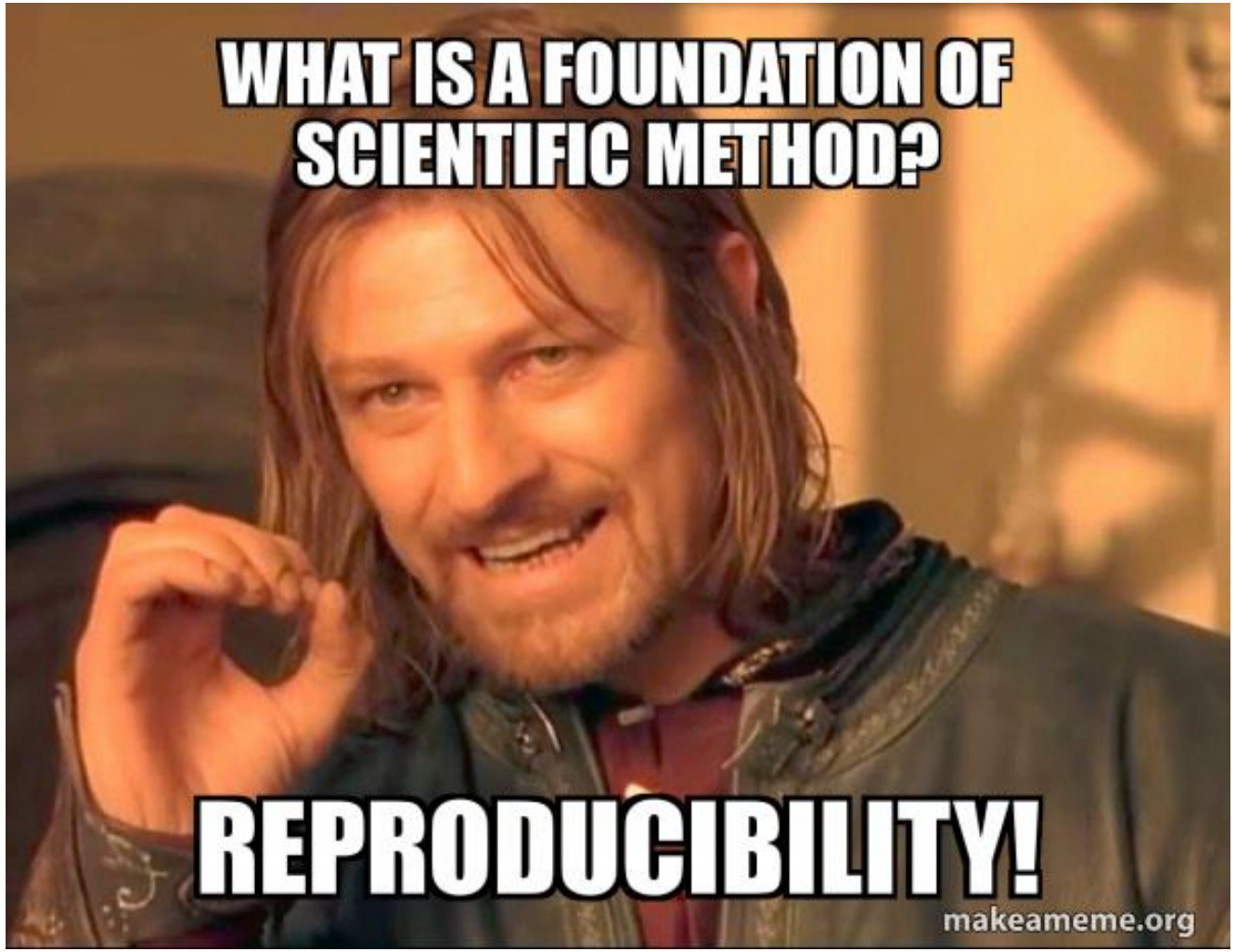
# Fits in the broader context of ‘Open science’



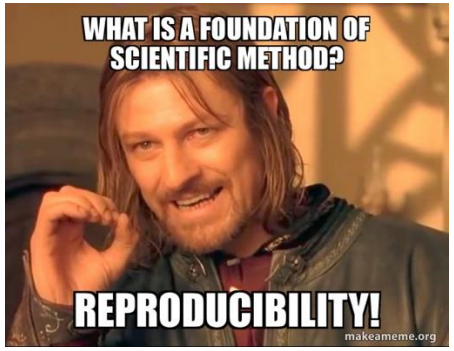

“[It] is transparent and accessible **knowledge** that is shared and developed through collaborative network”

Image: https://opensocialwork.org/research/open-science/

Vicente-Saez, R., & Martinez-Fuentes, C. (2018). Open Science now: A systematic literature review for an integrated definition. *Journal of Business Research*, *88*, 428–436. https://doi.org/10.1016/j.jbusres.2017.12.043

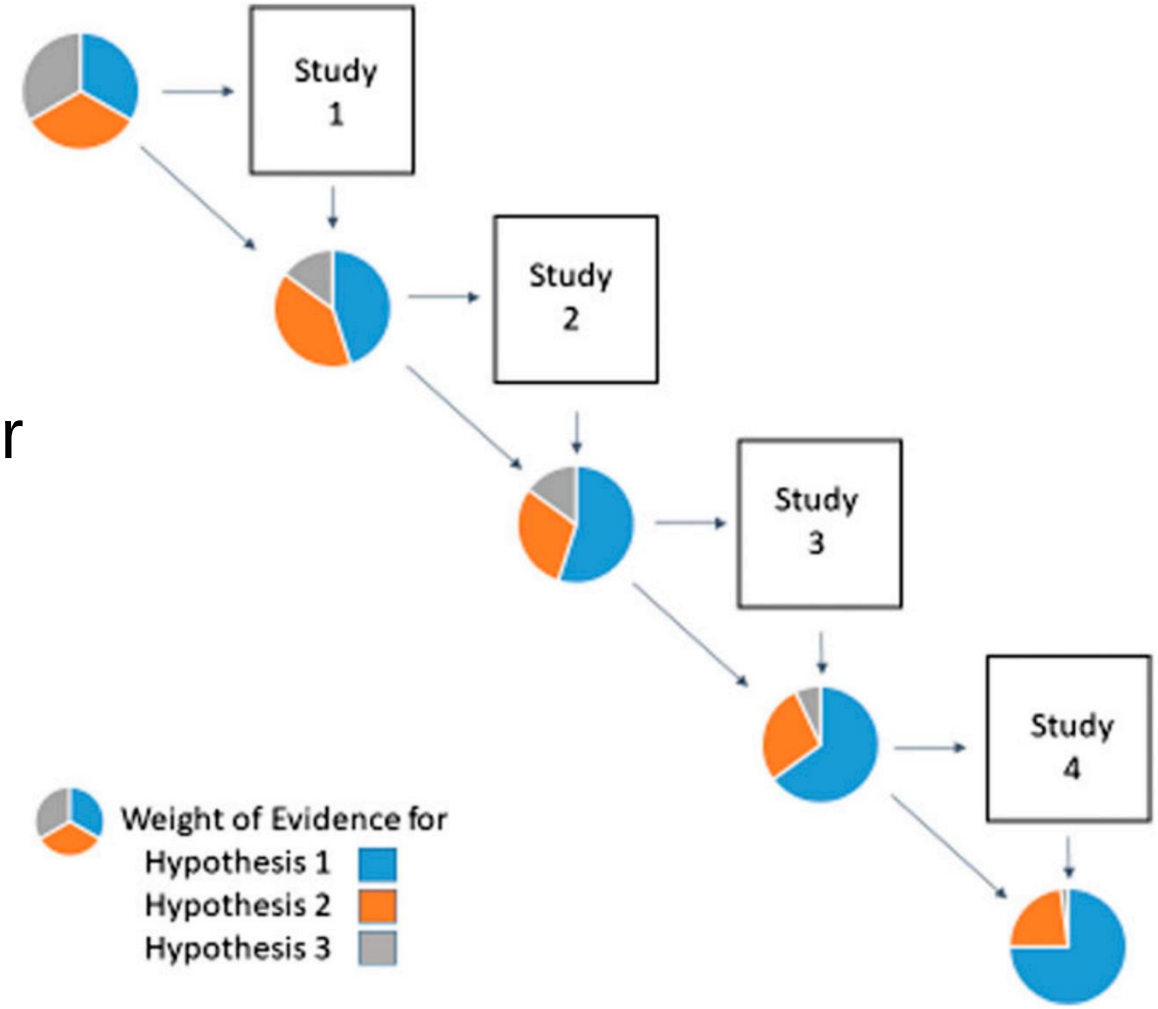

WHAT IS A FOUNDATION OF SCIENTIFIC METHOD?
REPRODUCIBILITY!
makeameme.org

# Why does matter for (GIS) science?




- Science: Discover laws, axioms, rules, etc. and describe them and under which conditions they occur

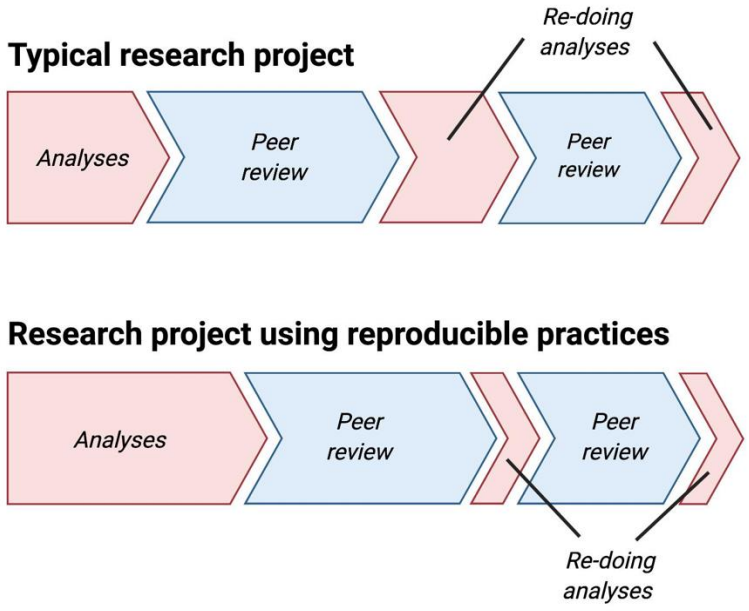
- Without reproducibility, replication is difficult
- Without replication, new knowledge is limited/slow




Evolving information state

Nichols et al (2021).
https://pnas.org/doi/full/10.1073/pnas.2100769118

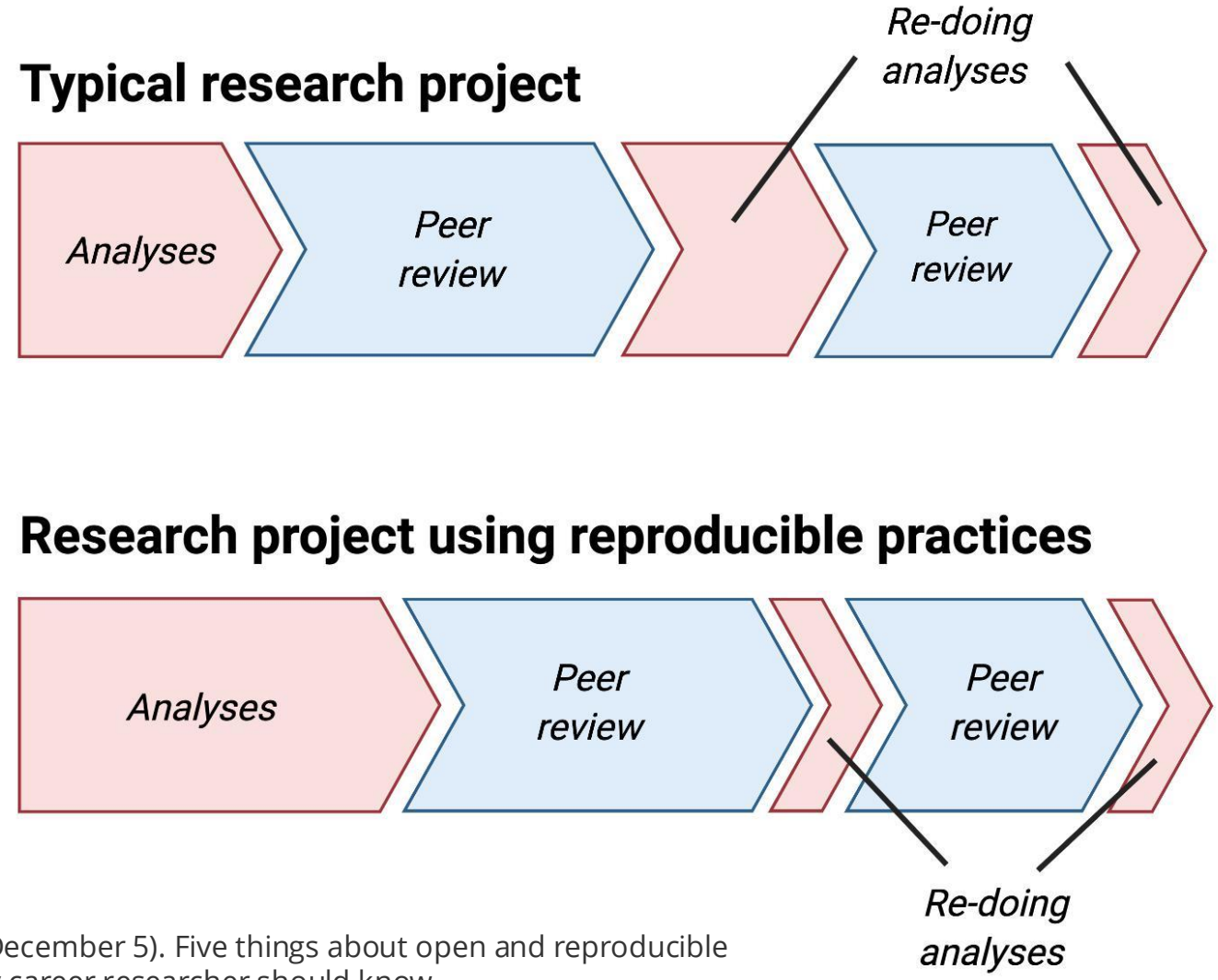
# Why does it matter for researchers?





Quintana, D. S. (2020, December 5). Five things about open and reproducible science that every early career researcher should know. https://doi.org/10.17605/OSF.IO/DZTVQ

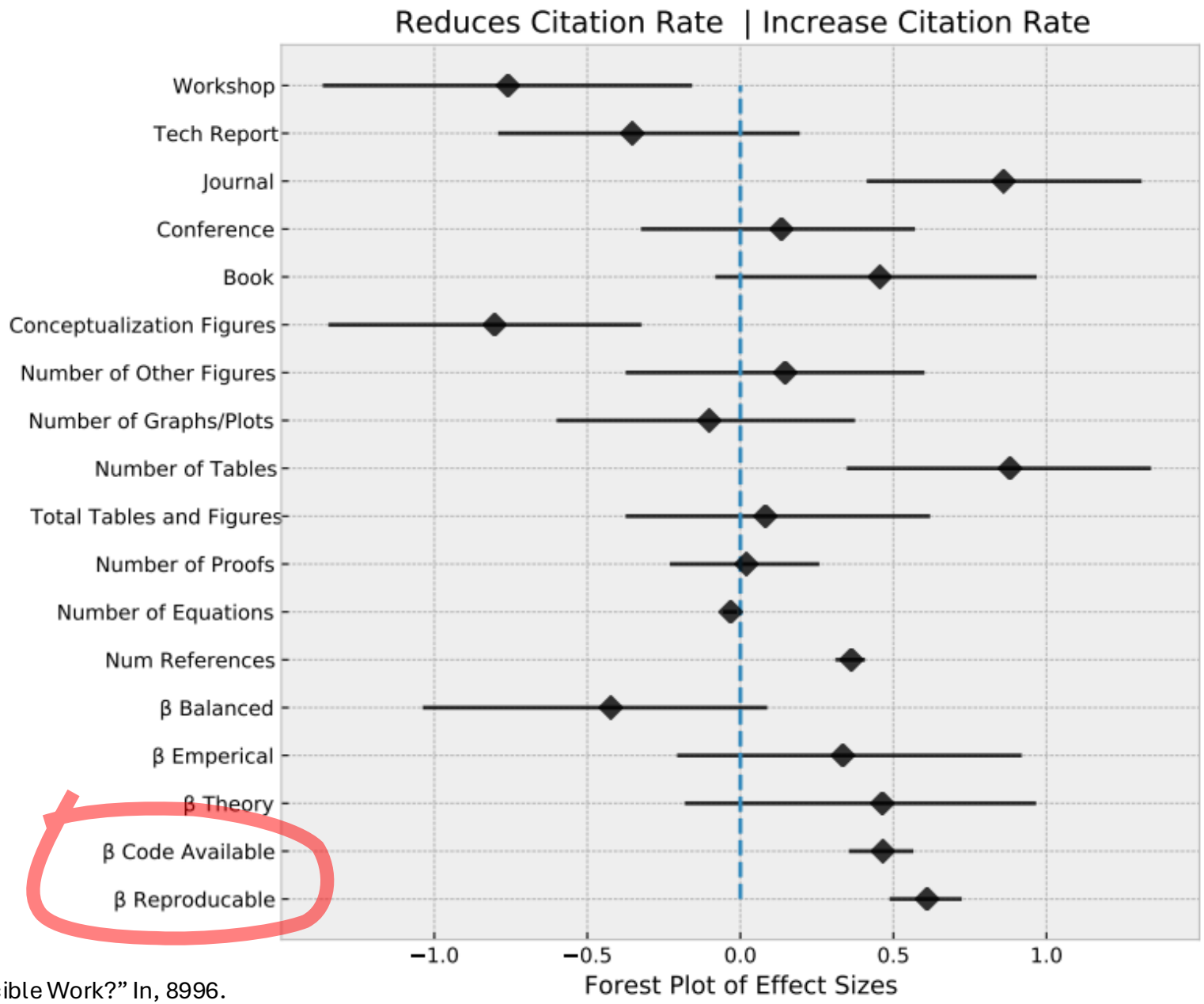
# Why does it matter for researchers?





Quintana, D. S. (2020, December 5). Five things about open and reproducible science that every early career researcher should know. https://doi.org/10.17605/OSF.IO/DZTVQ

- Benefits to your future self
- Benefits to others
- Scientific rigour
- Further potential for impact

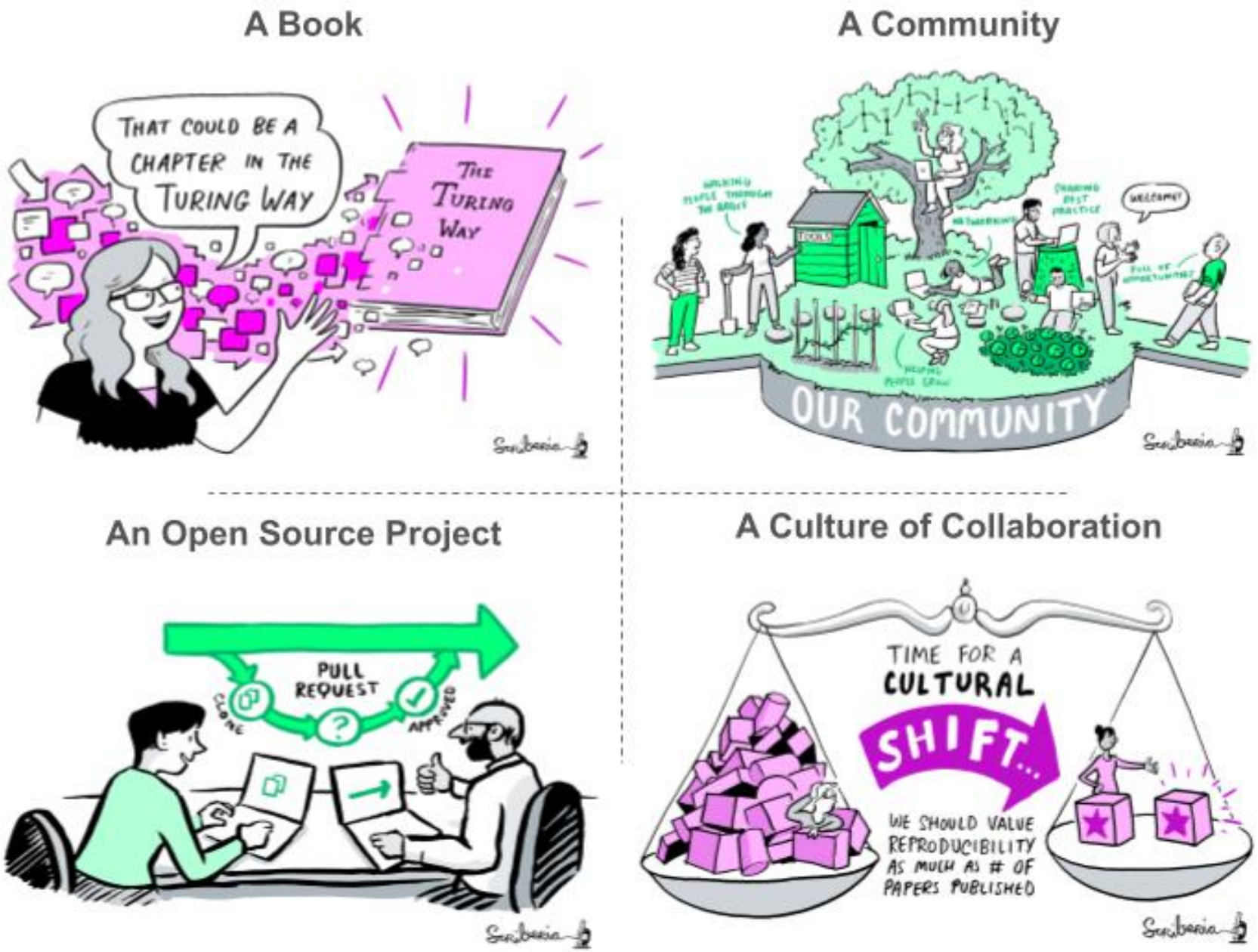



Raff, Edward. 2023. "Does the Market of Citations Reward Reproducible Work?" In, 8996. ACM REP '23. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3589806.3600041 in Lovelace (2024). https://www.robinlovelace.net/

What can we do and how?


# The Turing Way

https://book.the-turing-way.org/




The Turing Way Community, & Scriberia. (2021, May 29). Illustrations from the Turing Way book dashes. Zenodo. https://doi.org/10.5281/zenodo.4906004

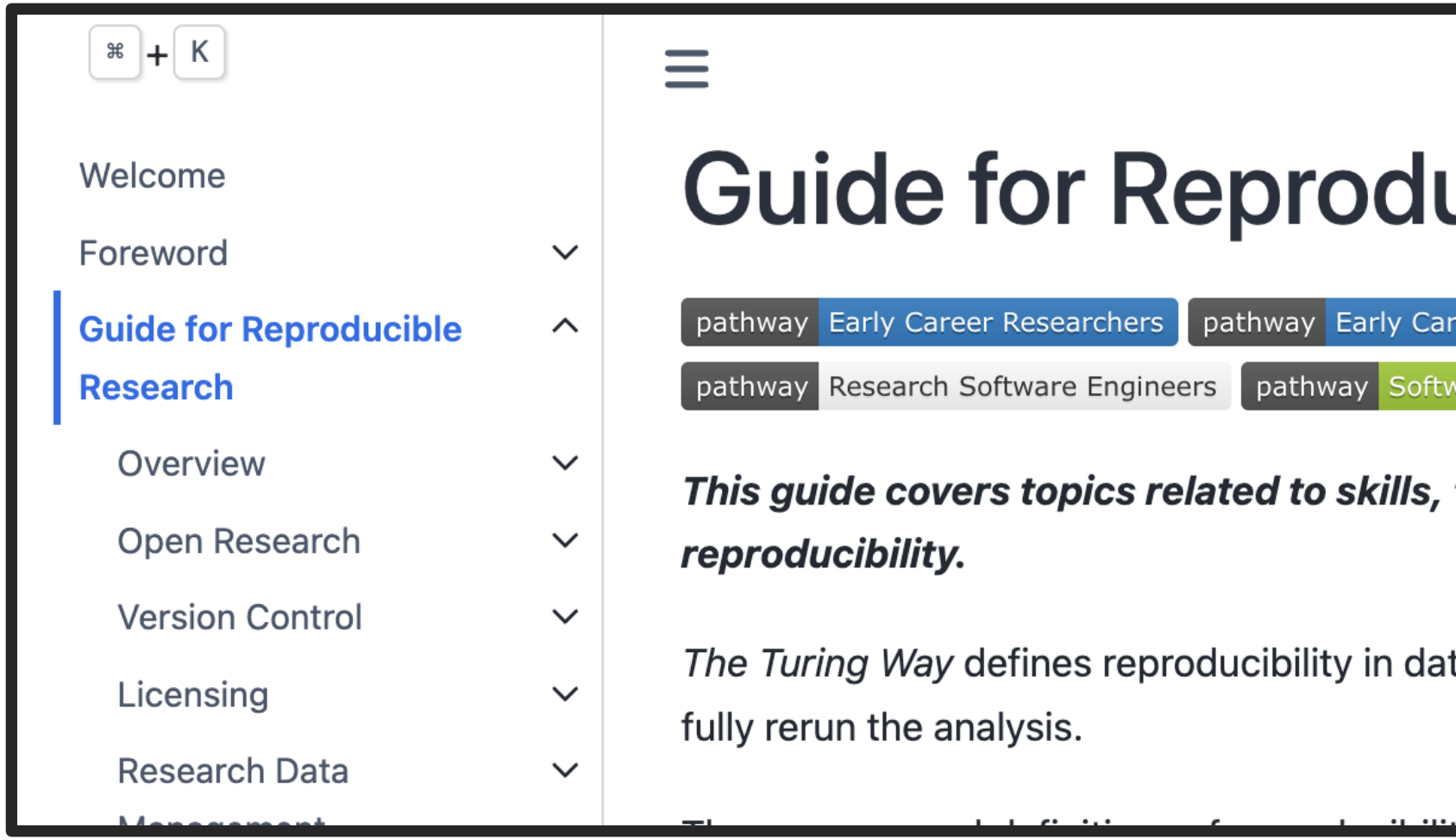
# The Turing Way

Covers:

- Skills
- Tools
- Best practices

for research reproducibility

It's free!




https://book.the-turing-way.org/

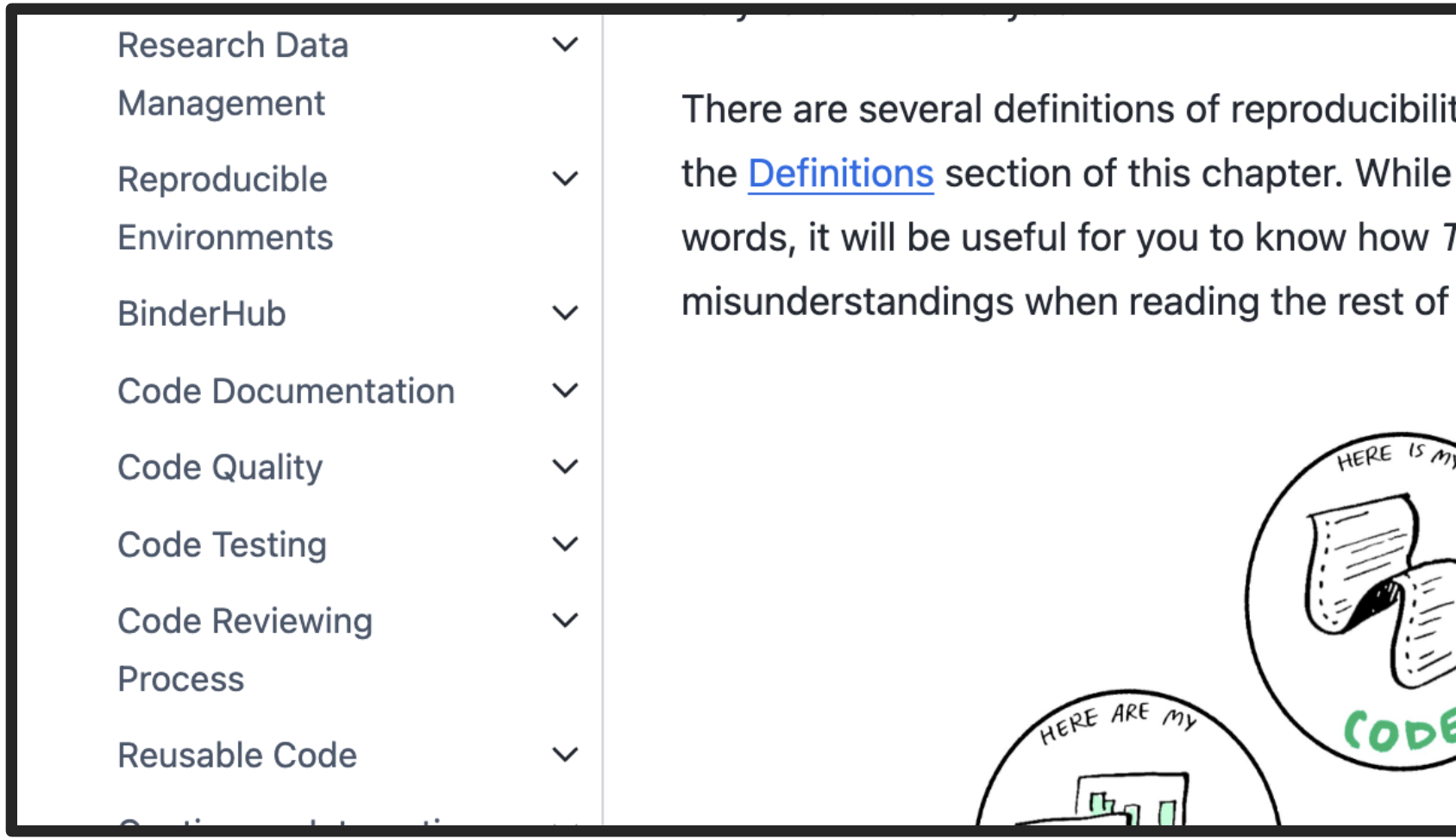
# The Turing Way

Covers:

- Skills
- Tools
- Best practices

for research reproducibility

It's free!




https://book.the-turing-way.org/

# AGILE Reproducibility Initiative

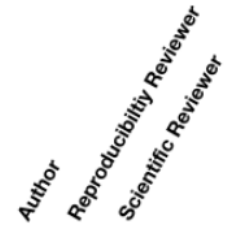
A reproducibility review is conducted with all accepted full papers based on *Reproducible paper guidelines*.

Website: https://osf.io/phmce/
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.



# AGILE Reproducibility Initiative

A reproducibility review is conducted with all accepted full papers based on *Reproducible paper guidelines*.

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

**Author Guidelines**
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**
Describe role in evaluating plausibility and completeness of the data and software availability documentation.

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

**Background**

# AGILE Reproducibility Initiative

## Author Guidelines

Show how to write the Data and Software Availability Section and give practical recommendations to make data and computational workflows reproducible.

### What if...

- **the datasets are openly available?** Cite the dataset[11] and clearly indicate which subset (if any) has been used.
- **the dataset is not openly available, is only temporarily available or is difficult to recreate?** Upload the dataset into a public repository if the original dataset license permits.
- **the licence or privacy considerations do not permit public re-sharing of the (part of) dataset?** Document the dataset and explain the procedures and conditions needed to access it. Provide a synthetic dataset to demonstrate your workflow and ideally a script for downloading.
- **you are the creator of the dataset?**: Select a license that allows the maximum reuse.
- **your data is published under your name in a public repository?** You can use **anonymised links**[12] to support anonymous review; mention the date and version of the record in the text.

# AGILE Reproducibility Initiative

## Author Guidelines

Show how to write the Data and Software Availability Section and give practical recommendations to make data and computational workflows reproducible.

### Examples

- **Social media data**: If the platform's terms of service do not allow for sharing all the data in a repository provide unique identifiers of the posts used[13].
- **OpenStreetMap data**: Provide feature type(s) used, geographic coverage, and the date of extraction or usage, ideally upload the extract to a data repository.
- **Framework data, socio-demographic and statistical data** (e.g administrative or natural boundaries, elevation data, 3D city models): Use the appropriate unique identifier to cite the dataset, e.g URI, DOI, POI, and describe the exact data source and the timestamp.
- **Personal data** (data containing information which can lead to the identification of individuals) should be shared after anonymisation / sufficient aggregation. If this is not possible, a dataset can be uploaded to a restricted access repository (e.g., DANS Easy) and metadata can be made public.

“Reproducible research is like riding a bike”

“Every step toward higher reproducibility counts” (AGILE, 2020)

Have a plan!

Granell et al (2018). Reproducible Research is like riding a bike (e27216v1). PeerJ Inc. https://doi.org/10.7287/peerj.preprints.27216v1

# Have a plan!

"Every step toward higher reproducibility counts" (AGILE, 2020)

## Reproducibility plan

Reproducibility levels: as per Nüst et al 2018

**Data**

| Dataset<br>(add more rows as needed) | Current reproducibility level and reasoning why | Planned measures for improvement and target reproducibility level |
|---|---|---|
| | | |
| | | |
| | | |

**Methods**

| Method<br>(add more rows as needed) | Current reproducibility level and reasoning why | Planned measures for improvement and target reproducibility level |
|---|---|---|
| | | |
| | | |

**Results**

| Results<br>(add more rows as needed) | Current reproducibility level and reasoning why | Planned measures for improvement and target reproducibility level |
|---|---|---|
| | | |
| | | |

# Research compendium

“[Provides] all the building blocks available and give a description of how the user can execute the contained code.”

The Turing Way Community, & Scriberia. (2021, May 29). Illustrations from the Turing Way book dashes. Zenodo. https://doi.org/10.5281/zenodo.4906004

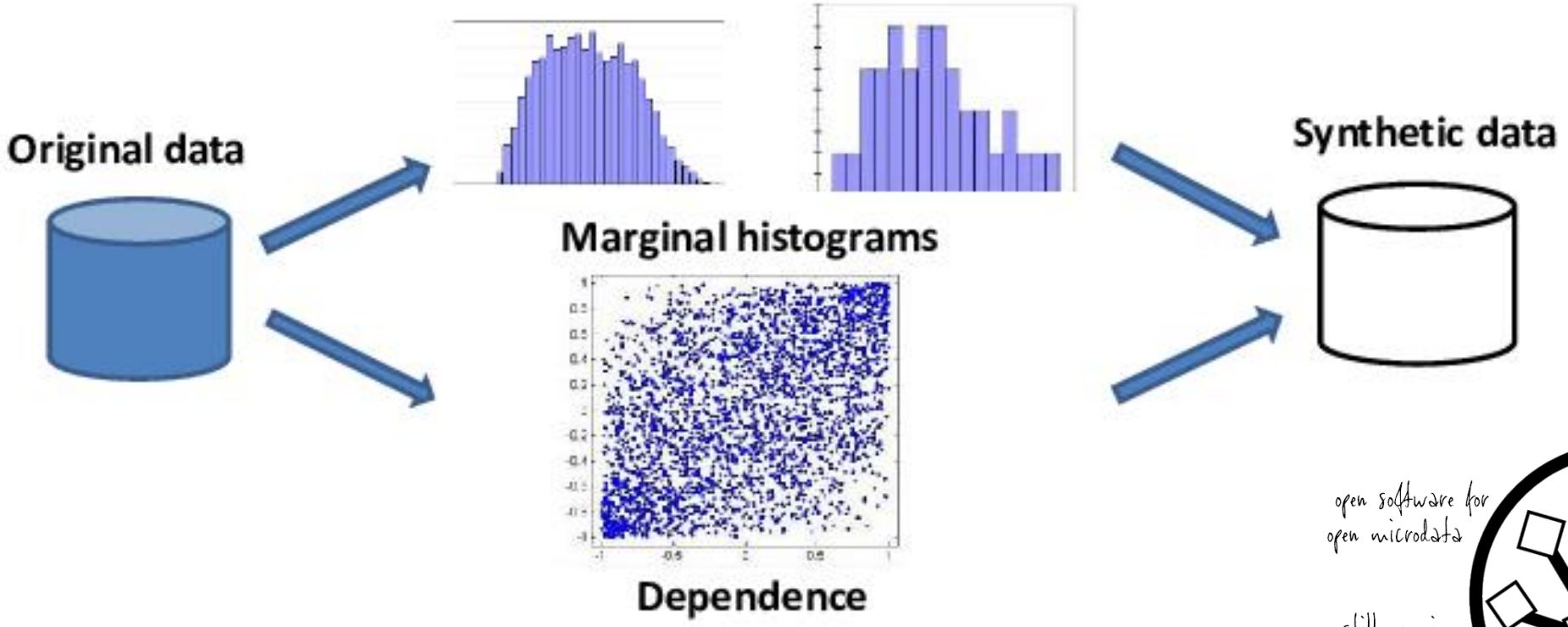
# Research compendium

## Research compendium principles

| | |
|---|---|
| | Stick with the conventions of your peers |
| | Keep data, methods and outputs separate |
| | Specify your computational environment as clearly as you can |

Kartik Ram: rstudio::conf 2019 talk


# Synthetic data




open software for open microdata

synthetic but look and behave like real data

still growing and you can contribute

tree-based and parametric methods

synthpop

R package for generating synthetic versions of sensitive microdata for statistical disclosure control


## Reproducibility committee

‼️ We are looking for more reproducibility reviewers ‼️

If you are interested to learn more about computational reproducibility and contribute to the cultural change in GIScience and in the AGILE community, please nominate yourself to become a reproducibility reviewer by sending an email to Carlos. See below for more information.

What is expected from reproducibility reviewers?

- An **interest** in learning more about computational reproducibility.
- Any **skills** with different software, tools, or programming languages are welcome, but not strictly necessary.
- Some **time** in April/May, it takes between 2-4 hours

carlos.granell@uji.es

# Final thoughts

- Reproducibility is not an all or nothing game
- Start early!
- It becomes easier every time (+skills, +resources)
- Healthier research community
- Increases chances for impact

# Thank you!

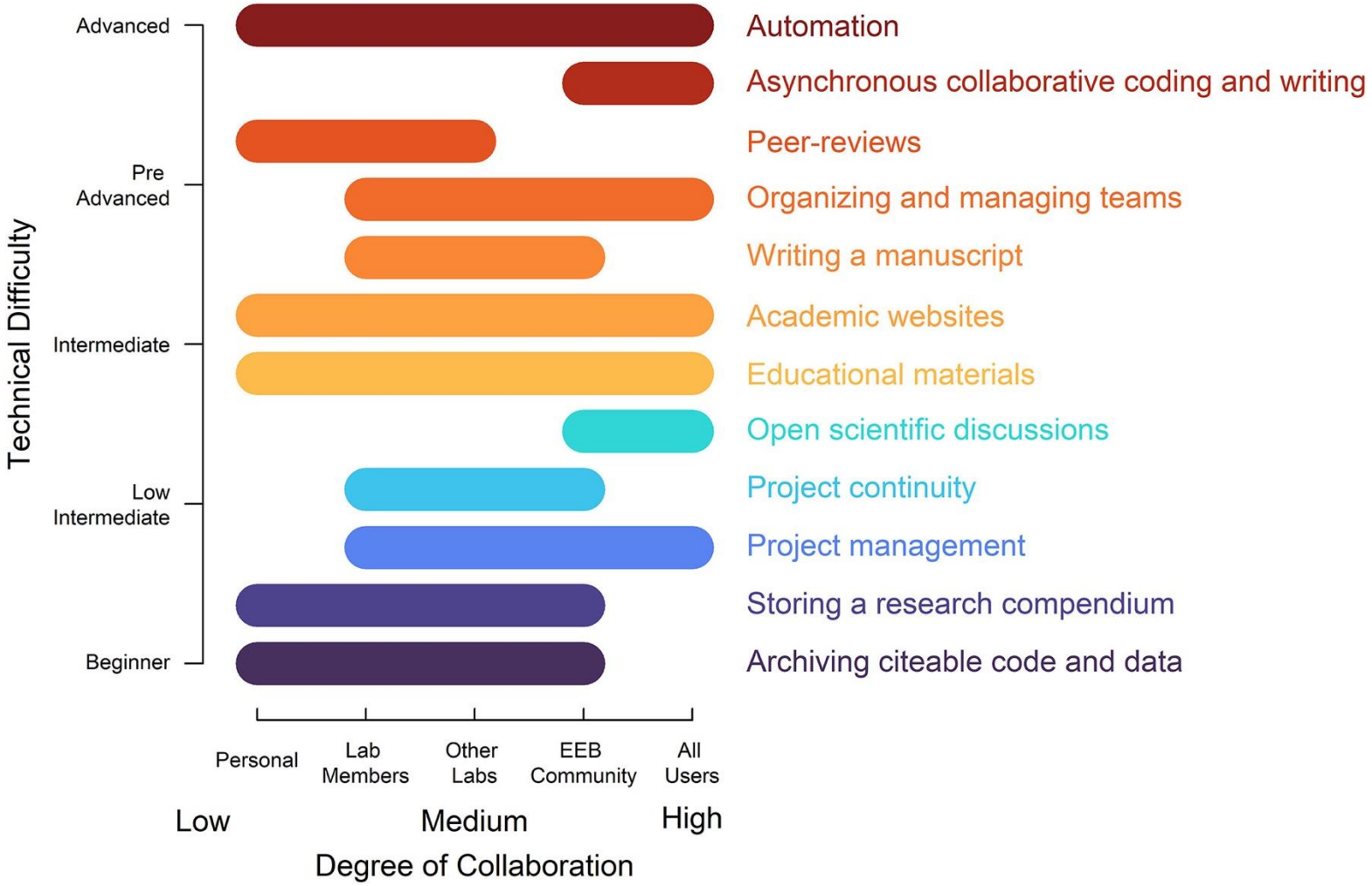
## Let’s stay connected

Bluesky

@rafa-vdz.bsky.social

JoseRafael.Verduzco-Torres@glasgow.ac.uk

# Not just for programmers: How GitHub can accelerate collaborative and reproducible research




Braga, P. H. P. et al (2023). Not just for programmers: How GitHub can accelerate collaborative and reproducible research in ecology and evolution. *Methods in Ecology and Evolution*, *14*(6), 1364–1380. https://doi.org/10.1111/2041-210X.14108

# FAIR principles

- Findable
- Accessible
- Interoperable
- Reusable


The Turing Way Community, & Scriberia. (2021, May 29). Illustrations from the Turing Way book dashes. Zenodo. https://doi.org/10.5281/zenodo.4906004