Pre-reproducibility assessment

Based on AGILE PhD school 2022

Frank Ostermann, ITC, University of Twente

Reproducible research and GIScience: an evaluation using AGILE conference papers

Daniel Nüst[1], Carlos Granell[2], Barbara Hofer[3], Markus Konkol[1], Frank O. Ostermann[4], Rusne Sileryte[5] and Valentina Cerutti[4]

[1] Institute for Geoinformatics, University of Münster, Münster, Germany
[2] Institute of New Imaging Technologies, Universitat Jaume I de Castellón, Castellón, Spain
[3] Interfaculty Department of Geoinformatics - Z_GIS, University of Salzburg, Salzburg, Austria
[4] Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands
[5] Faculty of Architecture and the Built Environment, Delft University of Technology, Delft, The Netherlands

ABSTRACT

The demand for reproducible research is on the rise in disciplines concerned with data analysis and computational methods. Therefore, we reviewed current recommendations for reproducible research and translated them into criteria for assessing the reproducibility of articles in the field of geographic information science (GIScience). Using this criteria, we assessed a sample of GIScience studies from the Association of Geographic Information Laboratories in Europe (AGILE) conference series, and we collected feedback about the assessment from the study authors. Results from the author feedback indicate that although authors support the concept of performing reproducible research, the incentives for doing this in practice are too small. Therefore, we propose concrete actions for individual researchers and the GIScience conference series to improve transparency and reproducibility. For example, to support researchers in producing reproducible work, the GIScience conference series could offer awards and paper badges, provide author guidelines for computational research, and publish articles in Open Access formats.

Author Contributions

- Daniel Nüst, Barbara Hofer and Rusne Sileryte analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Carlos Granell analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Markus Konkol and Frank O. Ostermann analyzed the data, authored or reviewed drafts of the paper, approved the final draft.
- Valentina Cerutti analyzed the data, approved the final draft.

Data Availability

The following information was supplied regarding data availability:

GitHub: https://github.com/nuest/reproducible-research-and-giscience/
Zenodo: https://doi.org/10.5281/zenodo.1227260.

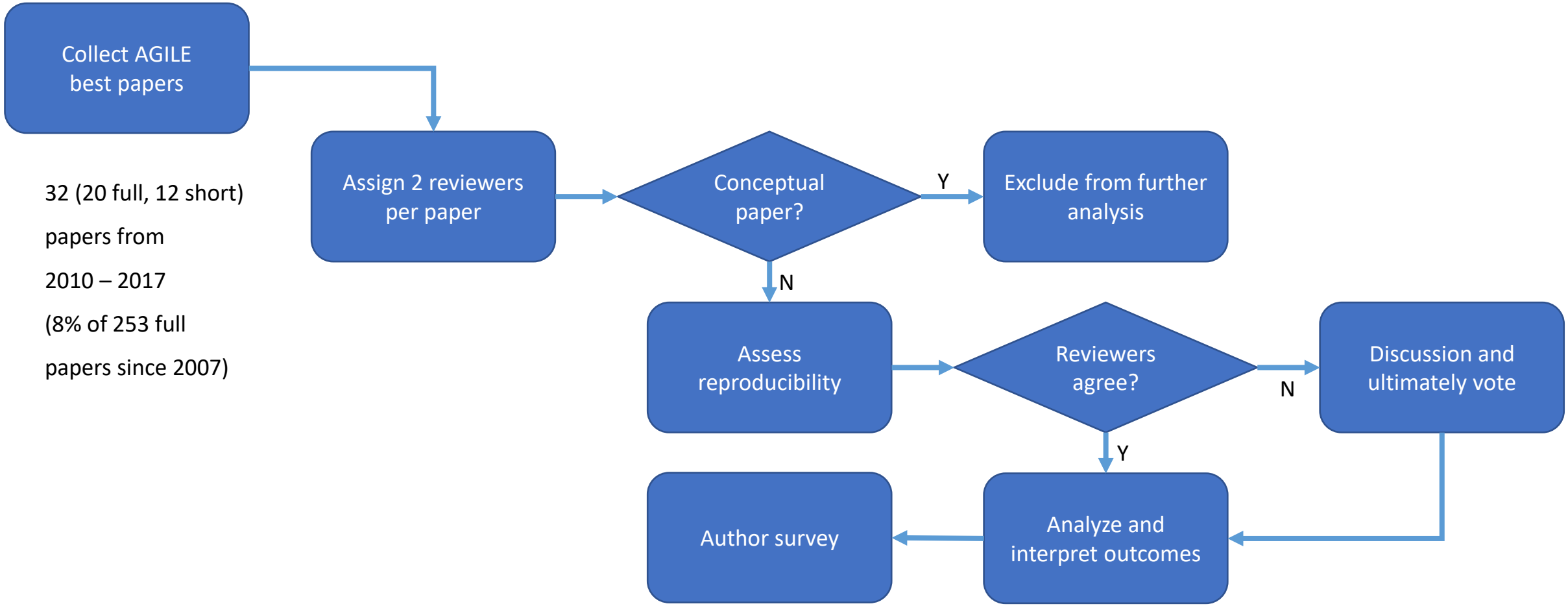
Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.5072#supplemental-information.

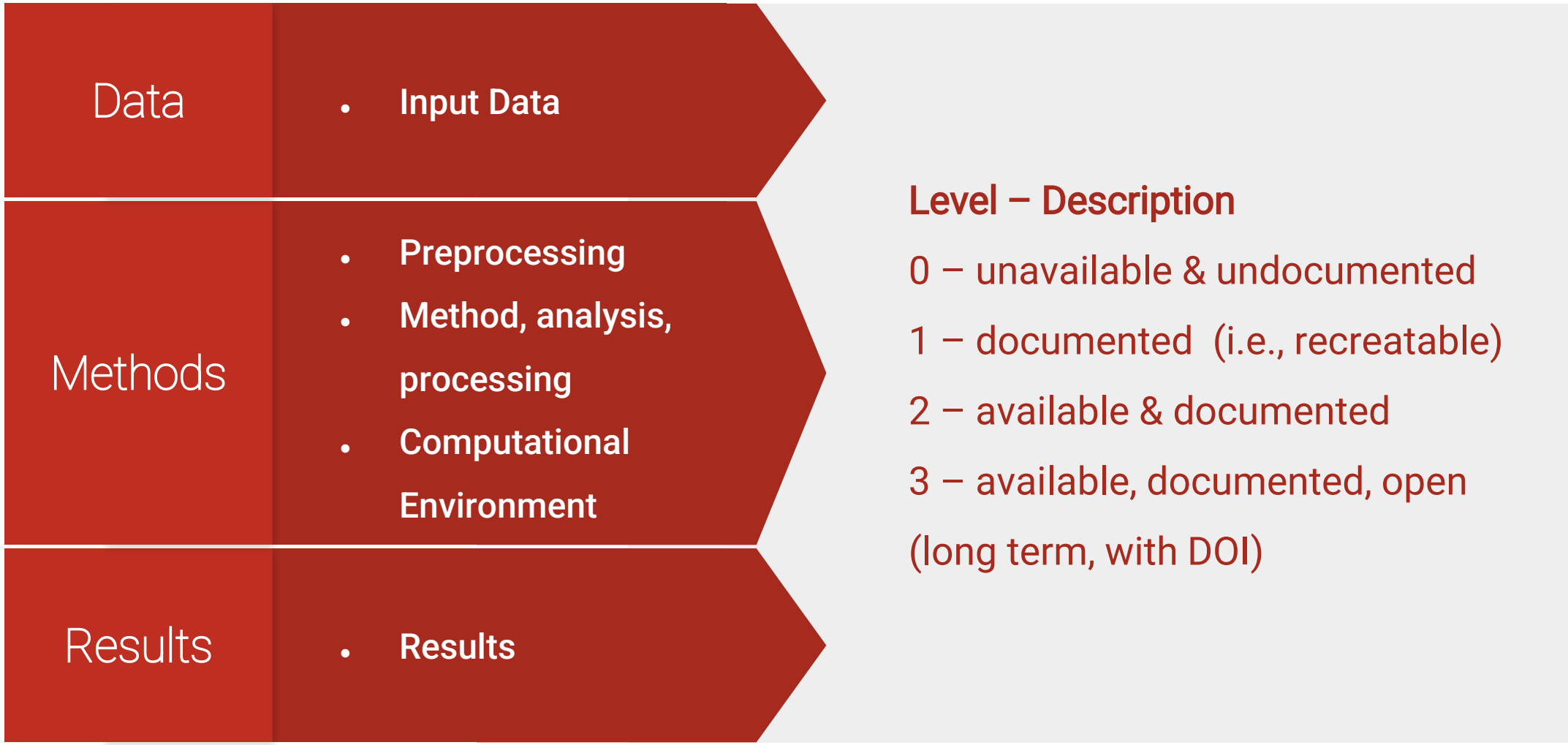
Submitted 13 March 2018

How did we examine AGILE papers' reproducibility?

32 (20 full, 12 short) papers from 2010 – 2017 (8% of 253 full papers since 2007)

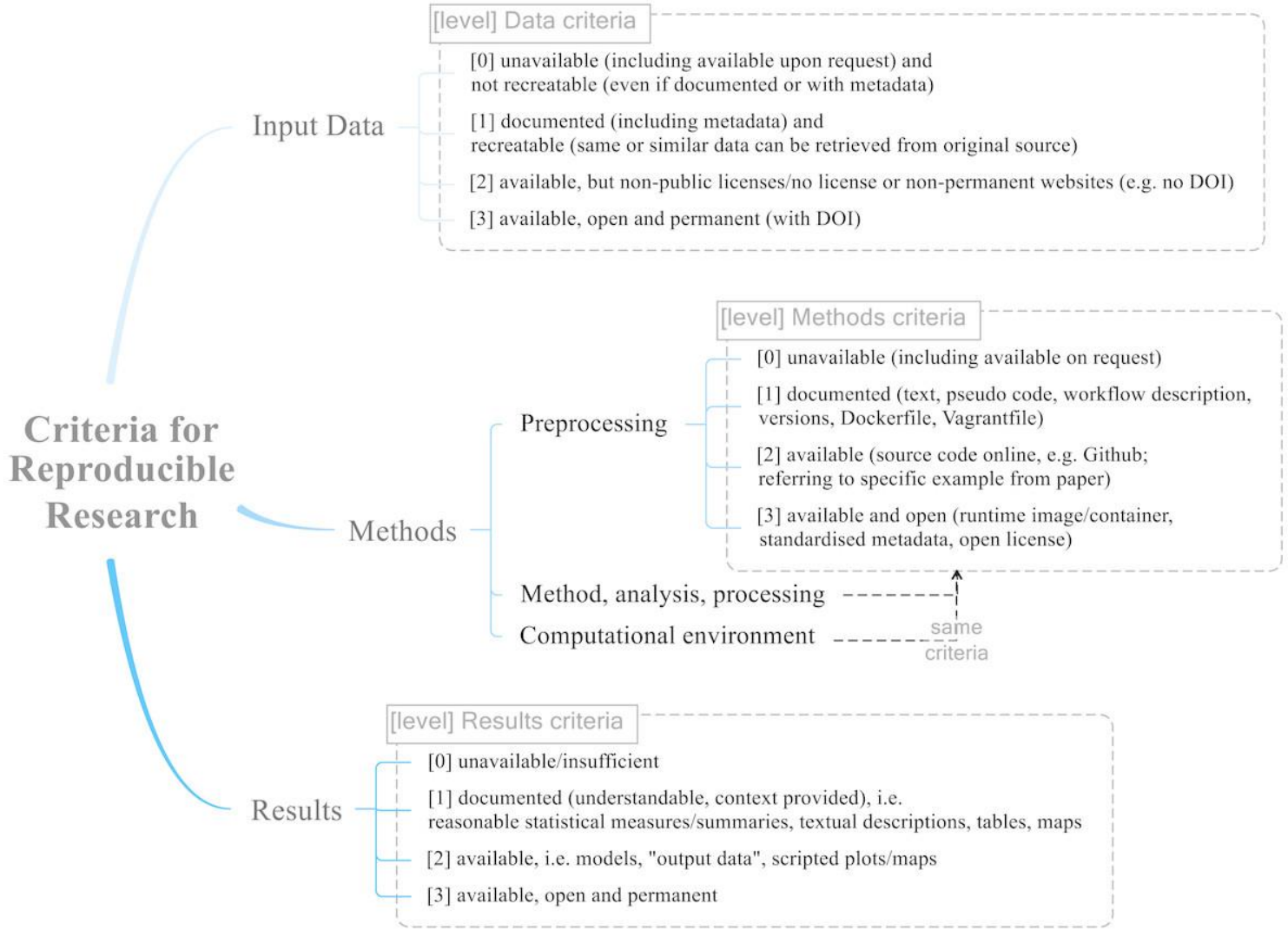


How to assess reproducibility?



Nüst D, Granell C, Hofer B, Konkol M, Ostermann FO, Sileryte R, Cerutti V. (2018)
Reproducible research and GIScience: an evaluation using AGILE conference papers.
PeerJ 6:e5072 https://peerj.com/articles/5072

How to assess reproducibility?



Level – Description

0 – unavailable & undocumented

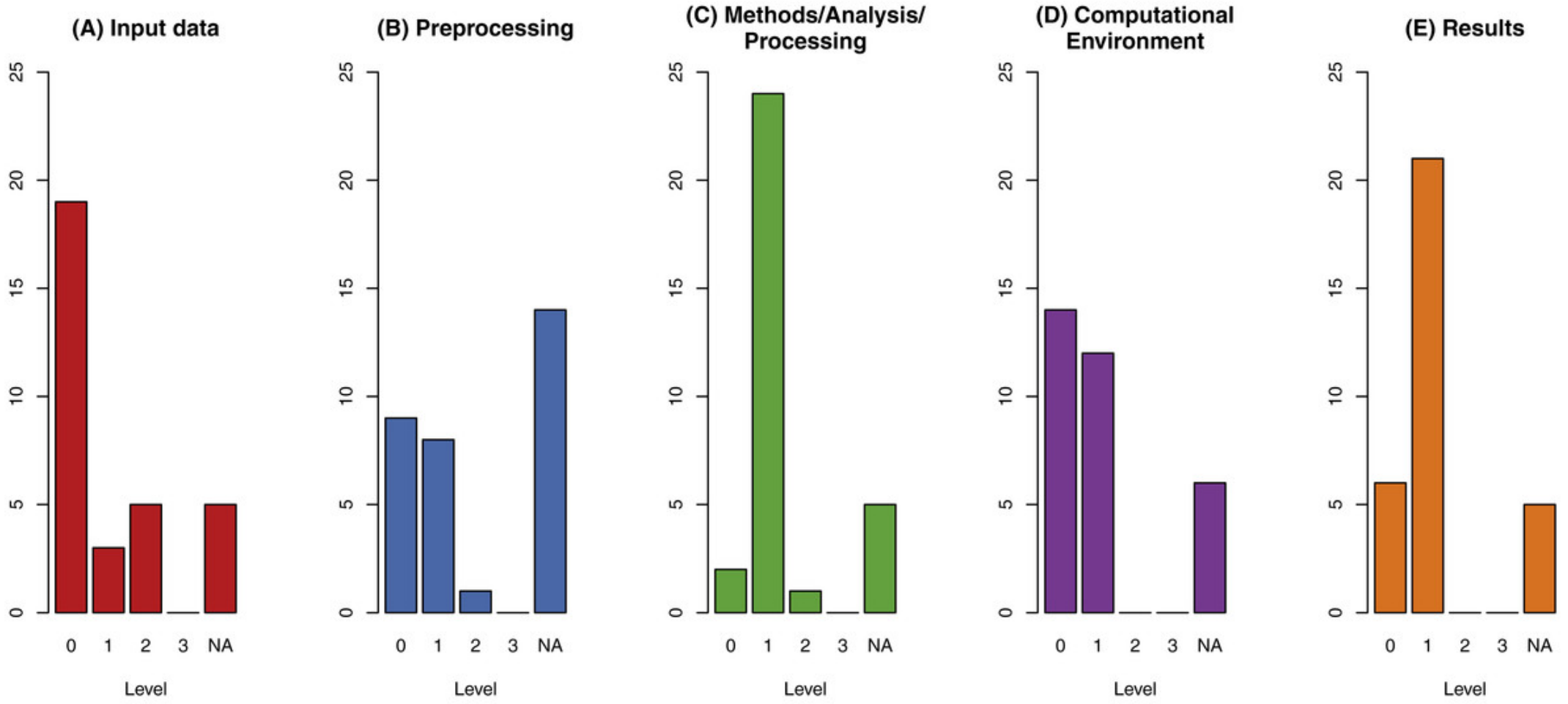
1 – documented (i.e., recreatable)

2 – available & documented

3 – available, documented, open (long term, with DOI)

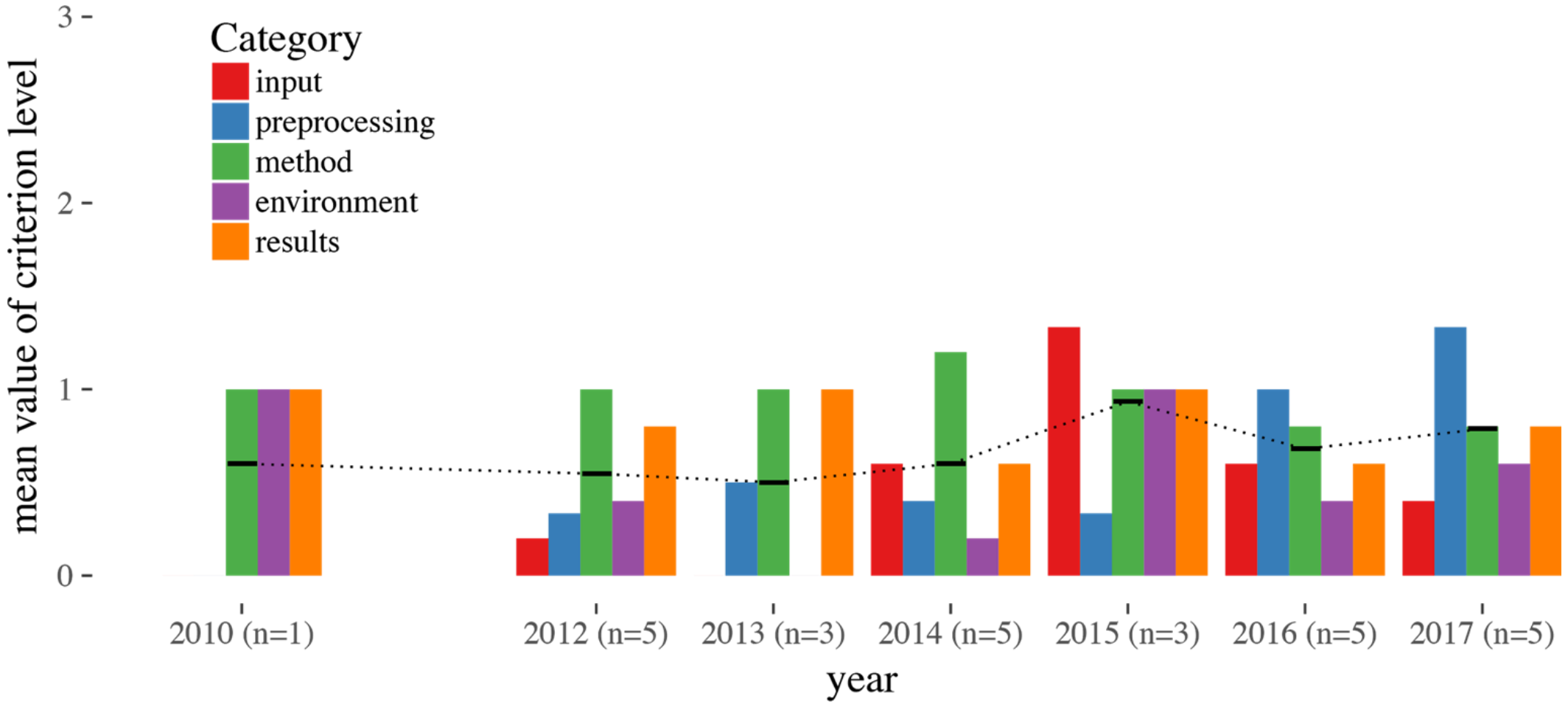
How reproducible were AGILE papers?

Level – Description
0 – unavailable & undocumented
1 – documented (i.e., recreatable)
2 – available & documented
3 – available, documented, open (long term, with DOI)



Does it at least improve over time? (no)

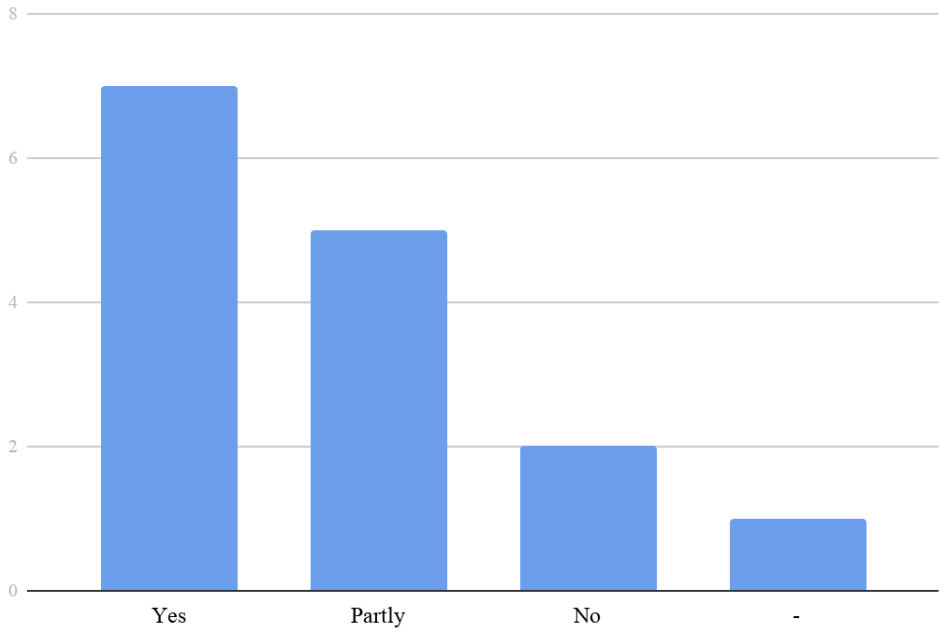
Level – Description
0 – unavailable & undocumented
1 – documented (i.e., recreatable)
2 – available & documented
3 – available, documented, open (long term, with DOI)



What were the authors' views?

- authors were provided with our evaluation of their paper
- 22 / 82 authors filled in the survey for 17/32 papers
- authors were asked to give consent to use their answers in the publications

Do you agree with our assessment?

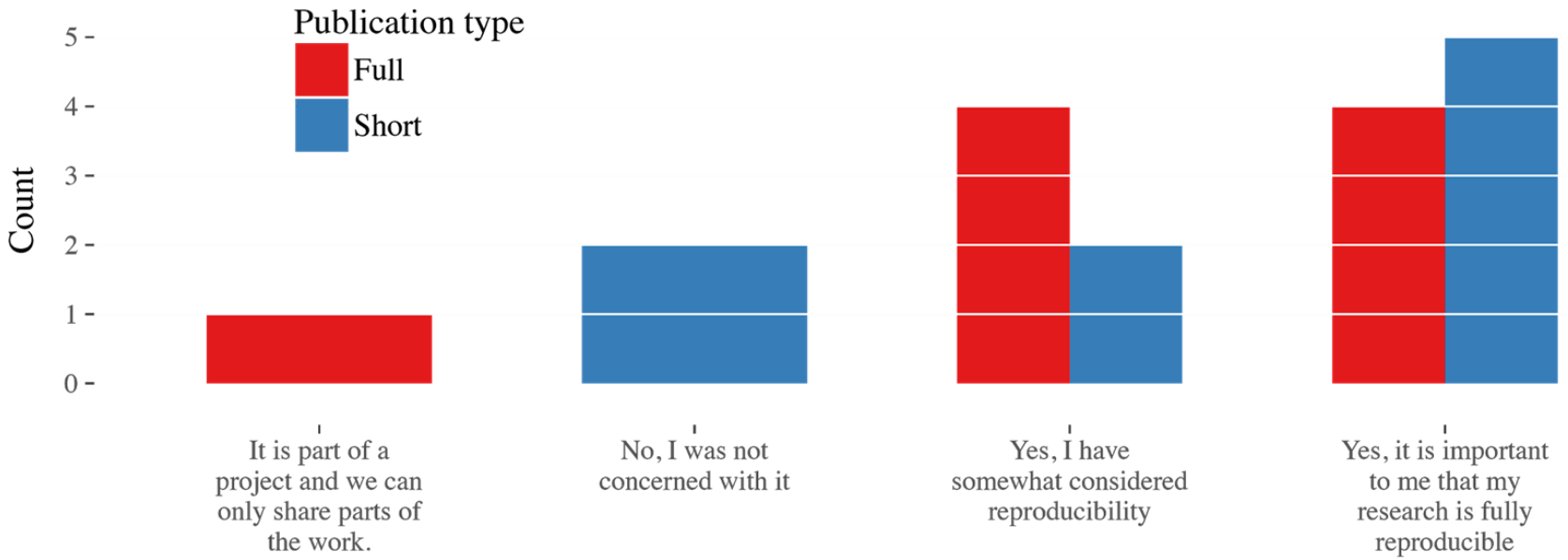


Reasons for disagreement:

- Requirements should not be applicable for short paper
- Specific data is not always necessary for reproducibility
- "Availability upon request" means "available"
- OSM data is by default "open and permanent"

Did they consider reproducibility? Why not?

Have you considered the reproducibility of research published in your nominated paper?



Reasons for lack of reproducibility

- Legal restrictions
- Not enough time
- Inadequate tools
- Lack of knowledge or skills
- Insufficient incentives

Next step: Replicate with another conference

Reproducible Research and GIScience: An Evaluation Using GIScience Conference Papers

Frank O. Ostermann [1]
Faculty of Geo-Information Science and Earth Observation (ITC),
University of Twente, Enschede, The Netherlands

Daniel Nüst
Institute for Geoinformatics, University of Münster, Germany

Carlos Granell
Institute of New Imaging Technologies, Universitat Jaume I de Castellón, Spain

Barbara Hofer
Christian Doppler Laboratory GEOHUM and Department of Geoinformatics - Z_GIS,
University of Salzburg, Austria

Markus Konkol
Faculty of Geo-Information Science and Earth Observation (ITC),
University of Twente, Enschede, The Netherlands

Abstract

GIScience conference authors and researchers face the same computational reproducibility challenges as authors and researchers from other disciplines who use computers to analyse data. Here, to assess the reproducibility of GIScience research, we apply a rubric for assessing the reproducibility of 75 conference papers published at the GIScience conference series in the years 2012-2018. Since the rubric and process were previously applied to the publications of the AGILE conference series, this paper itself is an attempt to replicate that analysis, however going beyond the previous work by evaluating and discussing proposed measures to improve reproducibility in the specific context of the GIScience conference series. The results of the GIScience paper assessment are in line with previous findings: although descriptions of workflows and the inclusion of the data and software suffice to explain the presented work, in most published papers they do not allow a third party to reproduce the results and findings with a reasonable effort. We summarise and adapt previous recommendations for improving this situation and propose the GIScience community to start a broad discussion on the reusability, quality, and openness of its research. Further, we critically reflect on the process of assessing paper reproducibility, and provide suggestions for improving future assessments. The code and data for this article are published at `https://doi.org/10.5281/zenodo.4032875`.

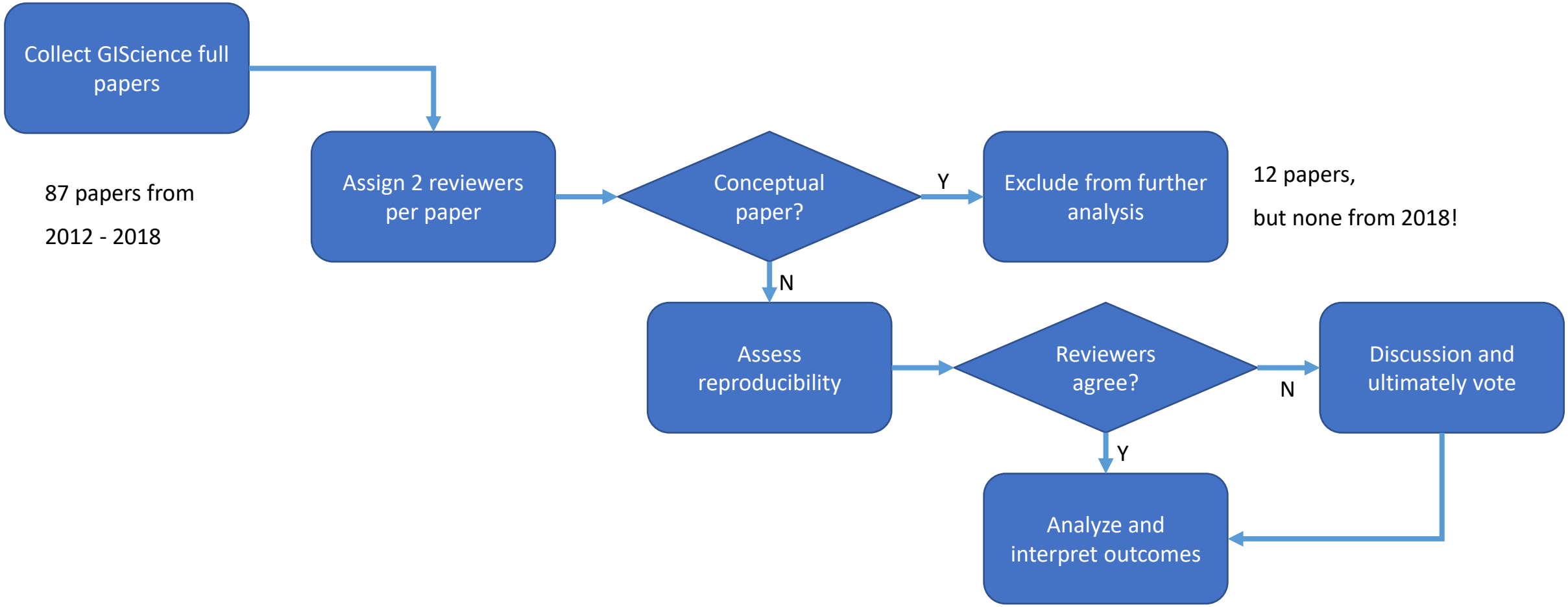
Supplementary Material The input data for this work are the full texts of GIScience conference proceedings from the years 2012 to 2018 [35, 7, 20, 34]. The paper assessment results and source code of figures are published at `https://github.com/nuest/reproducible-research-at-giscience` and archived on Zenodo [27]. The used computing environment is containerised with Docker pinning the R version to 3.6.3 and R packages to the MRAN snapshot of July 5th 2019.

Next step: Replicate with another conference

What did we want to do?

1. Investigate the state of reproducibility at GIScience conference series
2. Replicate an earlier assessment for AGILE conference series:
    - Is the method generalizable?
    - How do AGILE and GIScience compare?
3. Discuss strategies for improving reproducibility

How did we go about it?



Was our approach replicable?

Short answer: yes

But:

- labor-intensive, thus difficult to scale up
- *Preprocessing* not too helpful criterion (overlap with *Analysis*)
- *Computational environment* of limited use because relates mostly to processing time

Future replications should drop *preprocessing* and could drop *computational environment* criteria

Try it out!

https://github.com/nuest/reproducible-research-at-giscience

Level – Description

0 – unavailable & undocumented

1 – documented (i.e., recreatable)

2 – available & documented

3 – available, documented, open (long term, with DOI)

What's the outcome for GIScience?

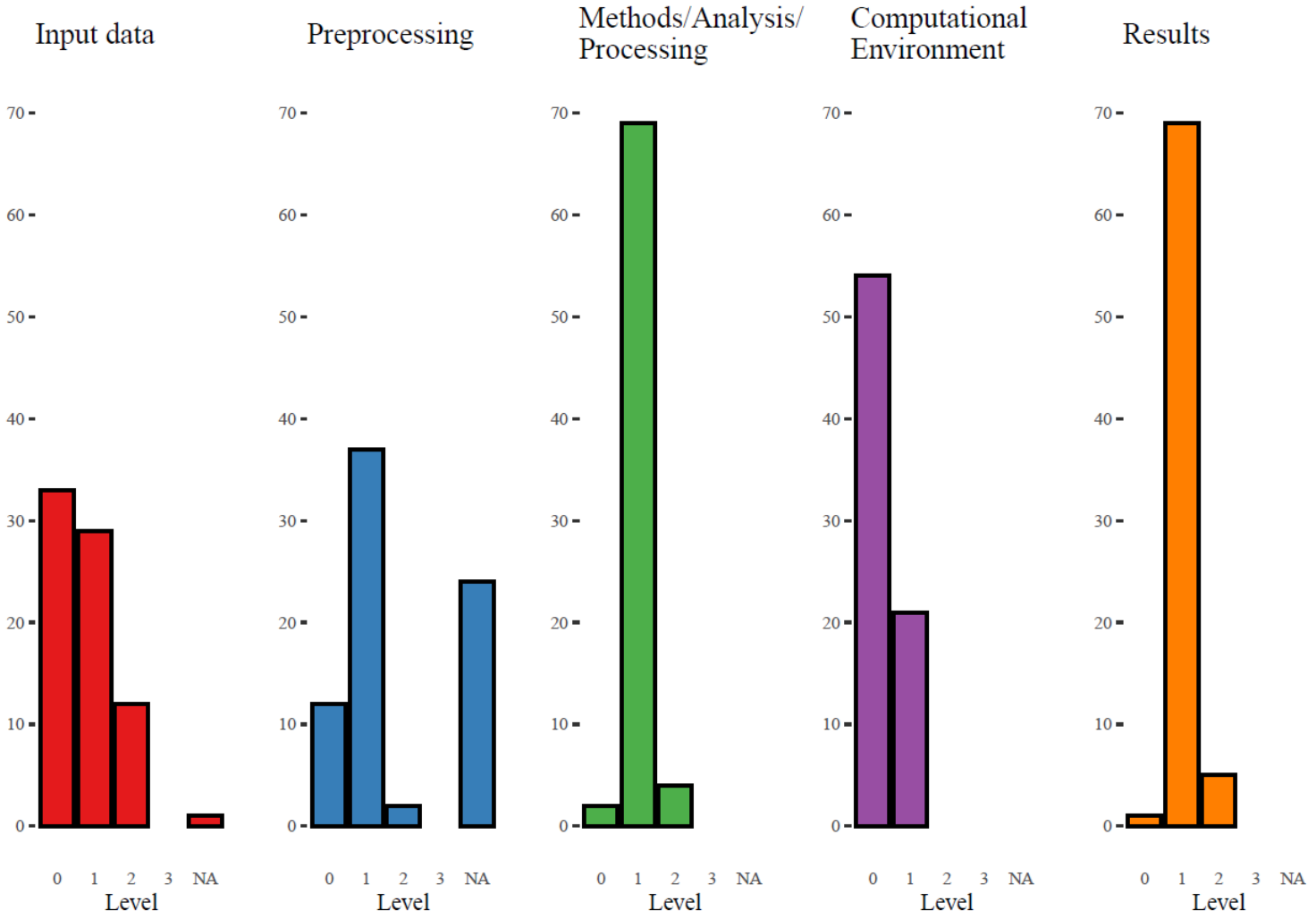


Figure 1 Barplots of reproducibility assessment results; levels range from 0 (leftmost bar) to 'not applicable' (rightmost bar).

Level – Description
0 – unavailable & undocumented
1 – documented (i.e., recreatable)
2 – available & documented
3 – available, documented, open (long term, with DOI)

Any patterns visible?

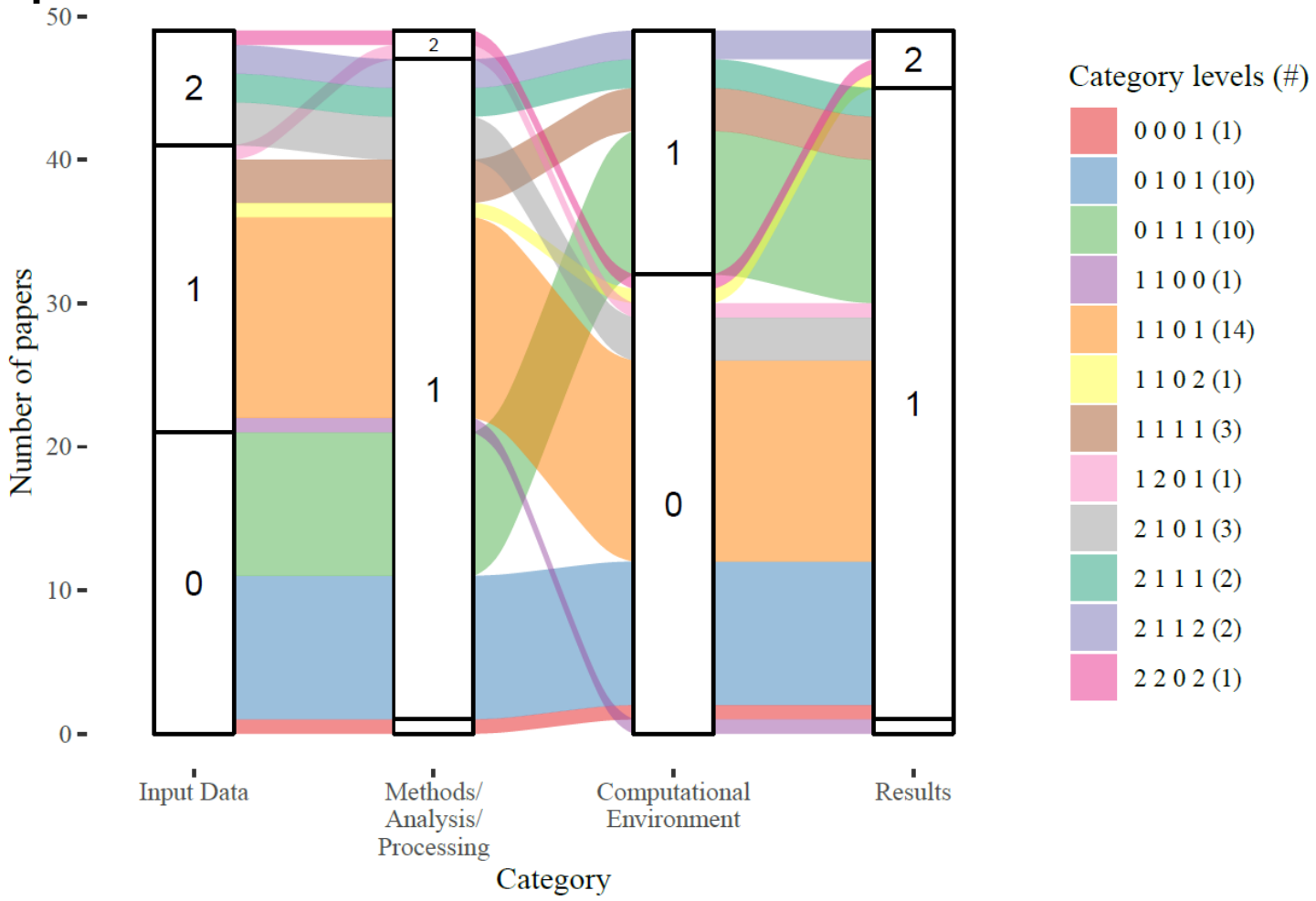
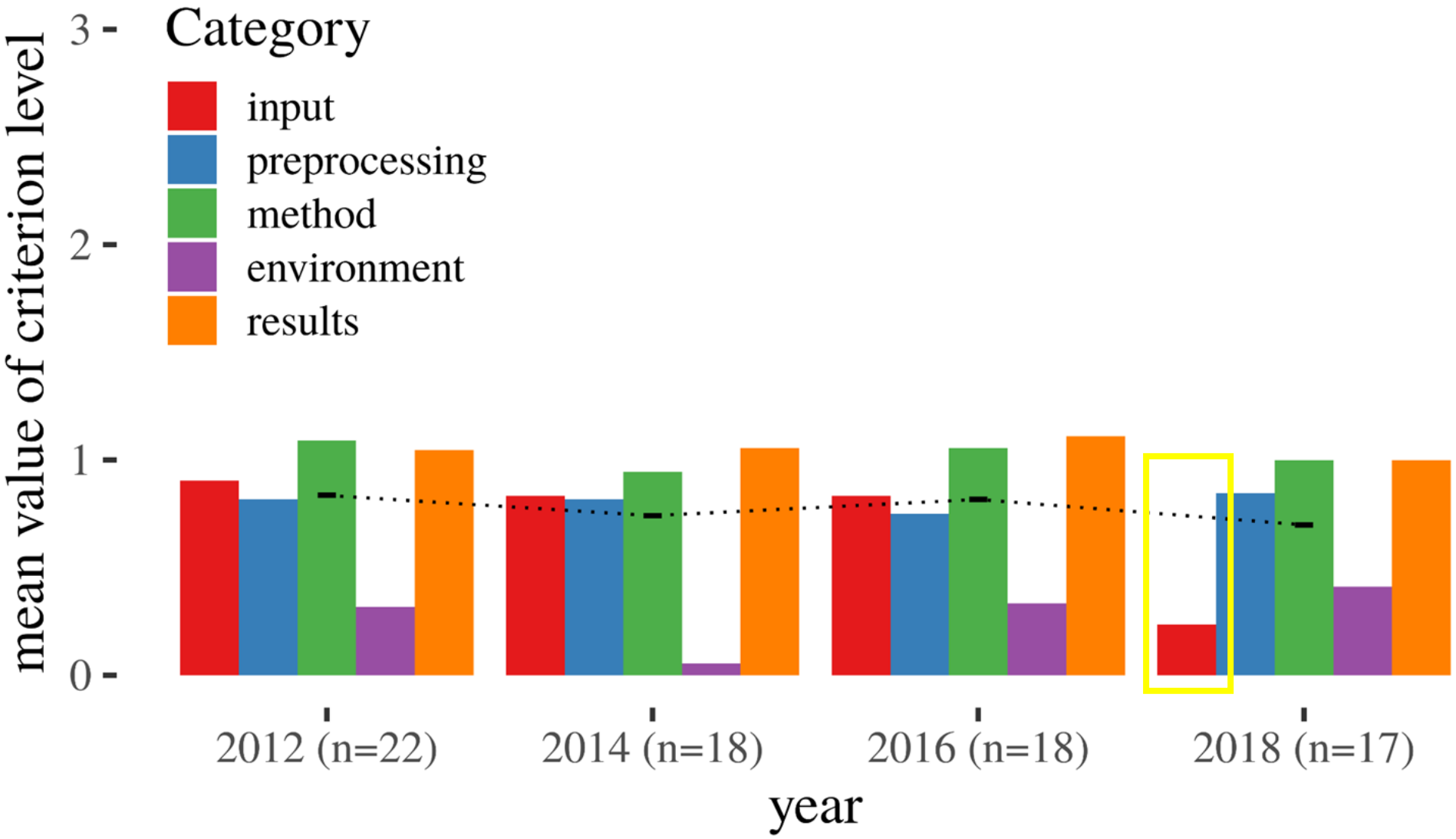


Figure 2 Alluvial diagram of common groups of papers throughout 4 of 5 categories including only papers without any "not applicable" *(Level NA)* value; category *Preprocessing* was dropped because difficulty to clearly assess it lead to many "not applicable" values.

Level – Description
0 – unavailable & undocumented
1 – documented (i.e., recreatable)
2 – available & documented
3 – available, documented, open (long term, with DOI)

Any change over time? (again, no)



But what does this mean for GIScience?

- Overall reproducibility not great but: most papers meet standards for publication (‘*documented*’ in all three main criteria)
- Main problem is *input data* (several score only ‘*unavailable*’)
  - Scores not a result of link rot (although that is a problem!): if there was reason to assume data was available at time of publication, paper received ‘*available*’
  - Worrisome, because of increased focus on data science and need for ML training data

Level – Description
0 – unavailable & undocumented
1 – documented (i.e., recreatable)
2 – available & documented
3 – available, documented, open (long term, with DOI)

How do GIScience and AGILE compare?

Table 3 Mean values per criterion for both conferences (rounded to two decimal places).

Criterion	AGILE full papers	GIScience papers
input data	0.67	0.72
method/analysis/processing	1.00	1.03
computational environment	0.62	0.28
results	0.88	1.05

- Similar in terms of topics
- overlap of authors noticeable but not majority
- different geographic scope
- Biannual vs annual
- AGILE has institutional framework (council) that supported newly implemented guidelines, reproducibility committee, and badges

First a quick assessment …

The first step is to be able to assess a paper's reproducibility quickly. You don't have to understand the paper deeply for doing this!

Look for

- Data availability: What data sources are used, and how could you retrieve or recreate them?
- Methods used: Which methods were used, and could you use the same libraries/software? Are all parameters documented?
- Result details: Are all results presented and properly explained?

… of this paper

As example, let's investigate this paper (or choose your own *if* you have a good contender – don't go searching now):

https://drops.dagstuhl.de/opus/volltexte/2018/9343/pdf/LIPIcs-GISCIENCE-2018-15.pdf

What do you think should be its scores?

Use the simple scoring system or input data, methods, and results

Level – Description
0 – unavailable & undocumented
1 – documented (i.e., recreatable)
2 – available & documented
3 – available, documented, open (long term, with DOI)

Immediate question:

How far do the authors of a paper have to go?

A: “Just” a repository with code and data and documentation

B: Exact reproduction of paper including figures

C: Full container of study (e.g., Docker)

Take-home message

Some reproducibility is always

better than no reproducibility!

If the gold standard seems unreachable,

don’t give up, do what you can!