

# Minor adjustments, major benefits:

## Assessing reproducibility in hydrology research


**Jim Stagge**

**The Ohio State University** Department of Civil,  
Environmental and Geodetic Engineering



# Some relevant motivation





← **Thread**


 **neil\_ferguson** @neil\_ferguson

I'm conscious that lots of people would like to see and run the pandemic simulation code we are using to model control measures against COVID-19. To explain the background - I wrote the code (thousands of lines of undocumented C) 13+ years ago to model flu pandemics...

5:13 PM · Mar 22, 2020 · [Twitter for iPhone](#)





1.2K Retweets 4.3K Likes


   

 **neil\_ferguson** @neil\_ferguson · Mar 22





Replying to @neil\_ferguson


I am happy to say that @Microsoft and @GitHub are working with @Imperial\_JIDEA and @MRC\_Outbreak to document, refactor and extend the code to allow others to use without the multiple days training it would currently require (and which we don't have time to give)...

 27  193  1K 





 **neil\_ferguson** @neil\_ferguson · Mar 22


They are also working with us to develop a web-based front end to allow public health policy makers from around the world to make use of the model in planning. We hope to make v1 releases of both the source and front end in the next 7-10 days...

 18  110  949 

 **neil\_ferguson** @neil\_ferguson · Mar 22

That timescale reflects the balancing of those priorities with the multitude of other urgent policy-relevant COVID-19 questions we are addressing...

 8  36  398 

 **neil\_ferguson** @neil\_ferguson · Mar 22

As well as the partners listed above, I would also like to thank all the other individuals and groups who have offered to help. It is appreciated, but not...

# Purpose of study

Significant effort has been put into improving reproducible research.

- How are we doing (hydrology and water resources)?
- How can we realistically assess and identify bottlenecks?

Stagge, J., Rosenberg, D., Abdallah, A. et al. Assessing data availability and research reproducibility in hydrology and water resources. *Sci Data* 6, 190030 (2019). <https://doi.org/10.1038/sdata.2019.30>

OPEN Assessing data availability and research reproducibility in hydrology and water resources

James H. Stagge<sup>1,2</sup>, David E. Rosenberg<sup>1</sup>, Adel M. Abdallah<sup>1,3</sup>, Hadia Akbar<sup>1</sup>, Nour A. Attallah<sup>1</sup> & Ryan James<sup>1</sup>

Received: 22 October 2018  
Accepted: 23 January 2019  
Published: 26 February 2019

There is broad interest to improve the reproducibility of published research. We developed a survey tool to assess the availability of digital research artifacts published alongside peer-reviewed journal articles (e.g. data, models, code, directions for use) and reproducibility of article results. We used the tool to assess 360 of the 1,989 articles published by six hydrology and water resources journals in 2017. Like studies from other fields, we reproduced results for only a small fraction of articles (1.6% of tested articles) using their available artifacts. We estimated, with 95% confidence, that results might be reproduced for only 0.6% to 6.8% of all 1,989 articles. Unlike prior studies, the survey tool identified key bottlenecks to making work more reproducible. Bottlenecks include: only some digital artifacts available (44% of articles), no directions (89%), or all artifacts available but results not reproducible (5%). The tool (or extensions) can help authors, journals, funders, and institutions to self-assess manuscripts, provide feedback to improve reproducibility, and recognize and reward reproducible articles as examples for others.

# Reproducibility as a Continuum

## Availability

All digital artifacts are available for others



## Reproducibility

Artifacts can be used to exactly reproduce published results.  
“Bit reproducibility”



## Replicability

Existing and new datasets can be used to replicate the published conclusions.

Institute of Education Sciences, US Dept of Education, and NSF. (2018) Companion Guidelines on Replication and Reproducibility in Education Research

# Survey of 360 Articles

- Created 15 question survey instrument:
  - Availability** - Self-evaluation & unfamiliar reviewer
  - Reproducibility**: Subject reviewer
- 6 hydrology/water resources journals
- 360 of 1,989 articles published in 2017
- 60 articles per reviewer
- 5-20 min for avail, 1-3 hours for reproducibility

## Availability

Q5. How accessible to users?

Some or all applicable    Not specified where    Not applicable

Q6. Where available?

All online    Third party    Author    In article

Q7. What is present?

Required

Optional

Input Data    Code /    License    Metadata    Identifiers  
Directions    Software    Hardware / software requirements    File format

Q8. Comments on availability [open response].

Q9. Do you estimate you and readers could use the available artifacts to generate results?

Yes    Not sure    Not familiar with resources    No

Q10. Continue to reproduce results?

Yes    No

## Reproducibility

Q11. Do the outputs verify published results (in text, figures, and tables)?

Yes (explain in Q12)    No (explain in Q13 and Q14)

Q12. If yes, explain what made the work reproducible and other comments [open response].

Q13. If no, why did reproducing the work fail?

Hardware / software errors    Did not generate results    Results differed  
Unclear directions    Other

Q14. Other comments on why reproducing the work failed [open response].

## Time to Complete

Q15: How many minutes did the survey take?

# Availability

# Reproducibility

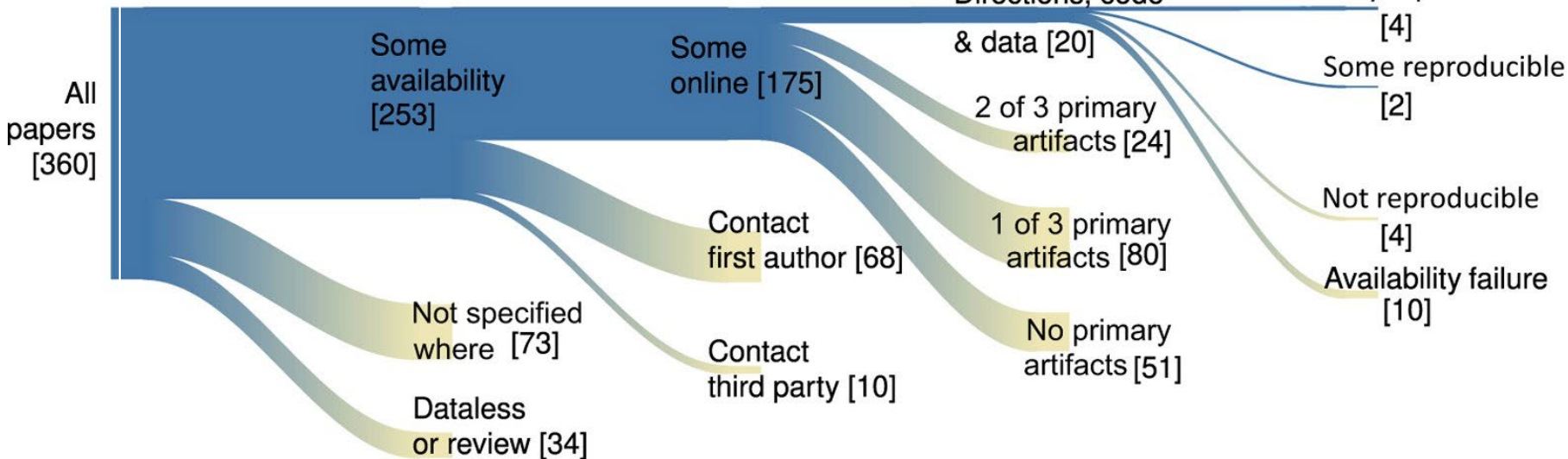


**Q5.** How accessible to users?

**Q6.** Where available?

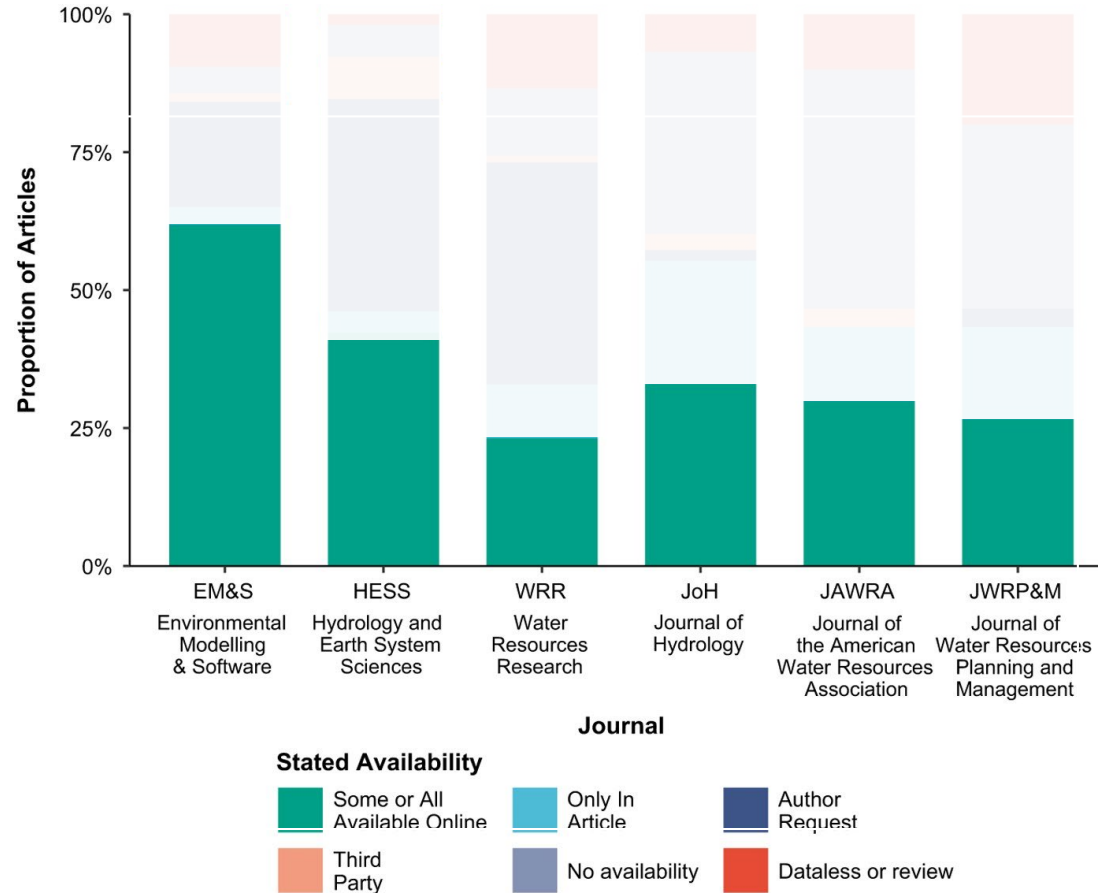
**Q7.** What is present?

**Q11.** Do outputs verify results?



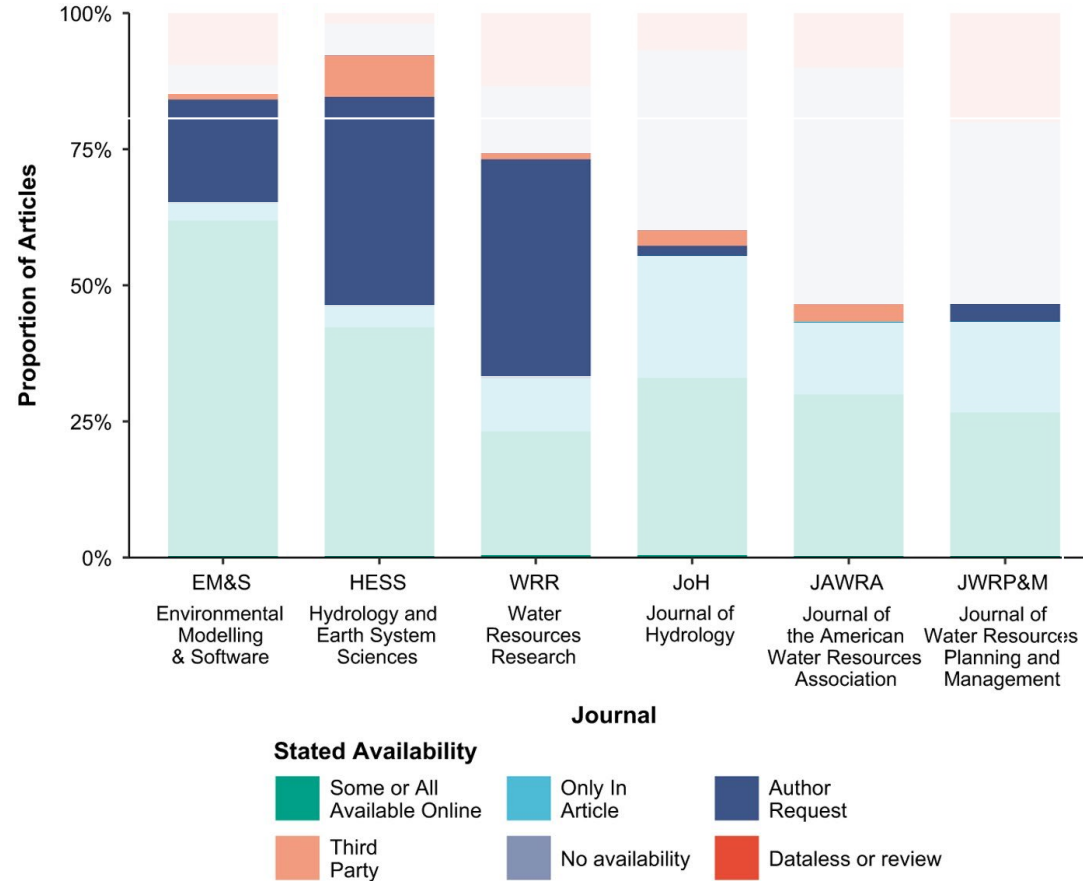
# Stated Availability by Journal

- **70.3%** stated some materials online
- Data availability statement required vs encouraged



# Stated Availability by Journal

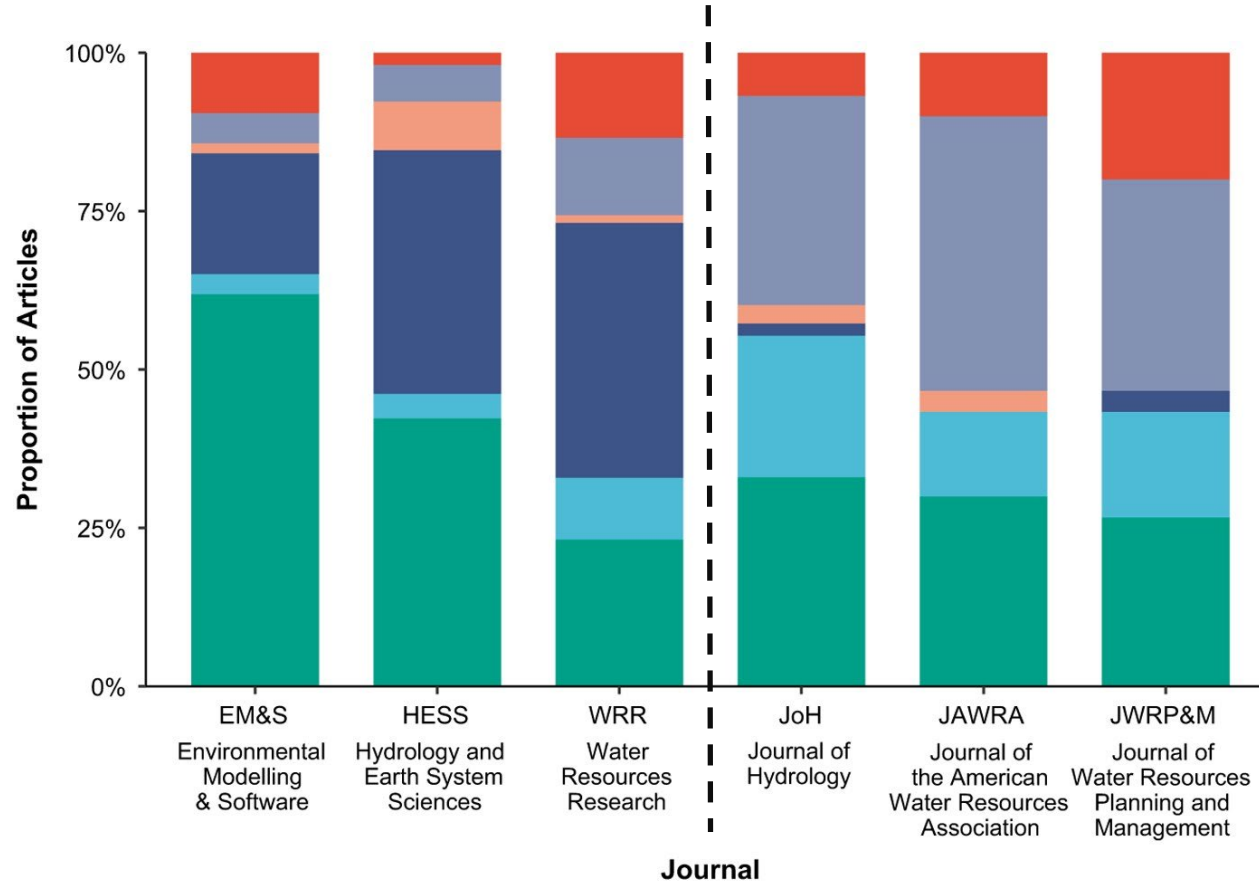
- Data availability statement increased the author/third party requests
- Did not test





# Stated Availability by Journal

## Stated Availability



# Availability

# Reproducibility



**Q5. How accessible to users?**

**Q6. Where available?**

**Q7. What is present?**

**Q11. Do outputs verify results?**

All papers [360]

Some availability [253]

Some online [175]

Directions, code & data [20]

Fully reproducible [4]

Some reproducible [2]

2 of 3 primary artifacts [24]

Not reproducible [4]

Availability failure [10]

Contact first author [68]

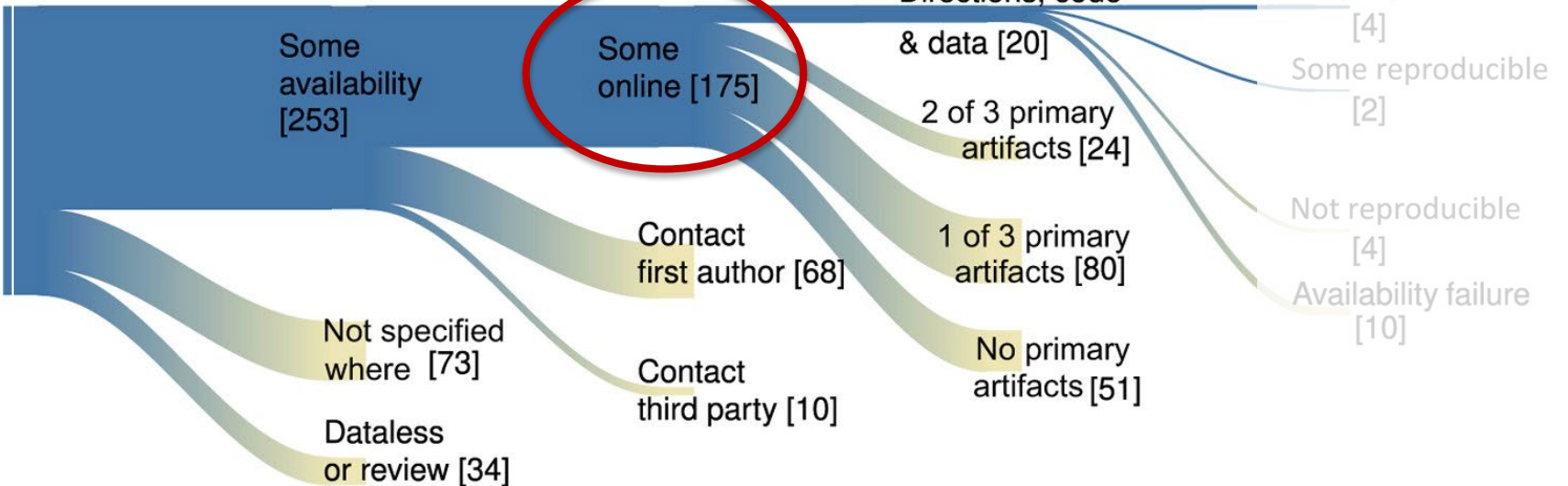
1 of 3 primary artifacts [80]

Not specified where [73]

Contact third party [10]

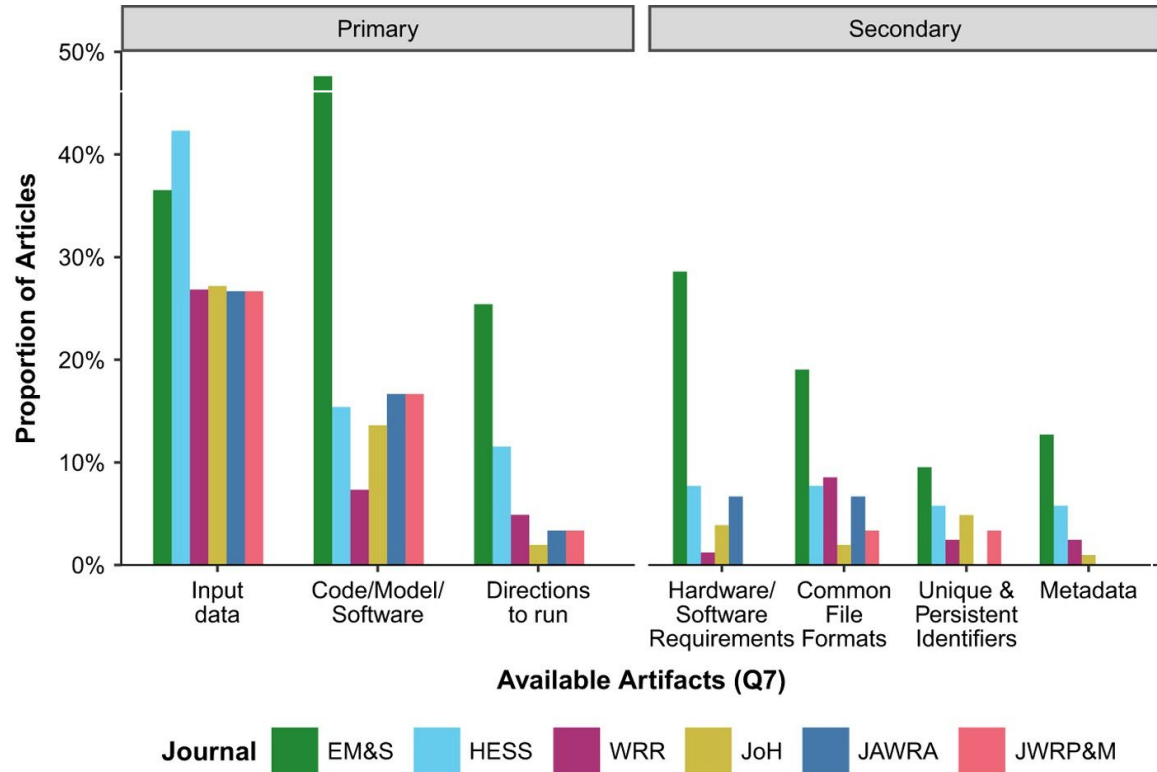
No primary artifacts [51]

Dataless or review [34]

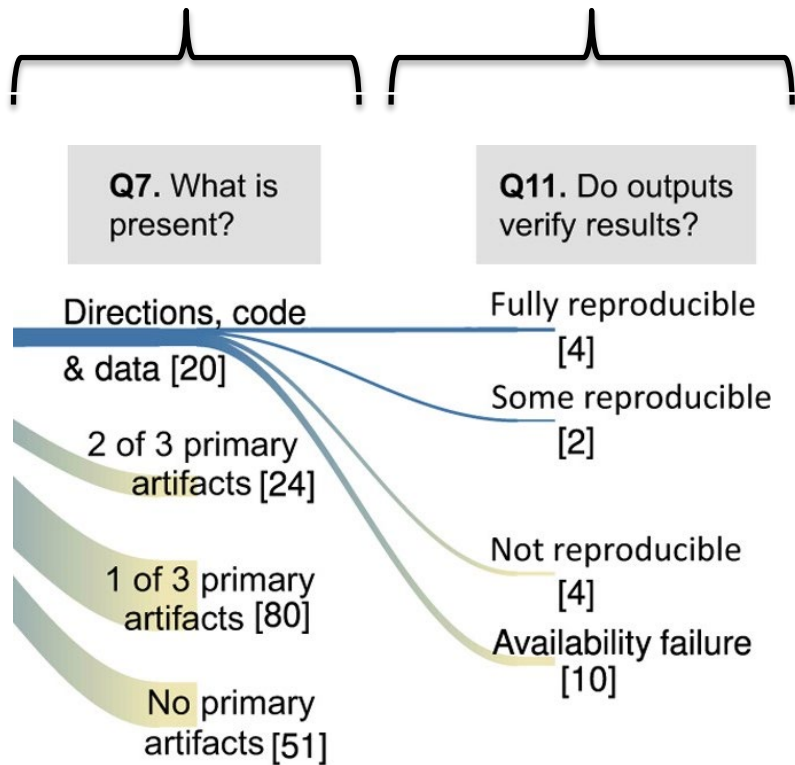


# Available digital artifacts

- Primary (necessary) & Secondary (useful)
- Data most common artifact by far (25-40%)
- Followed by Code (15%) and Directions (5%)

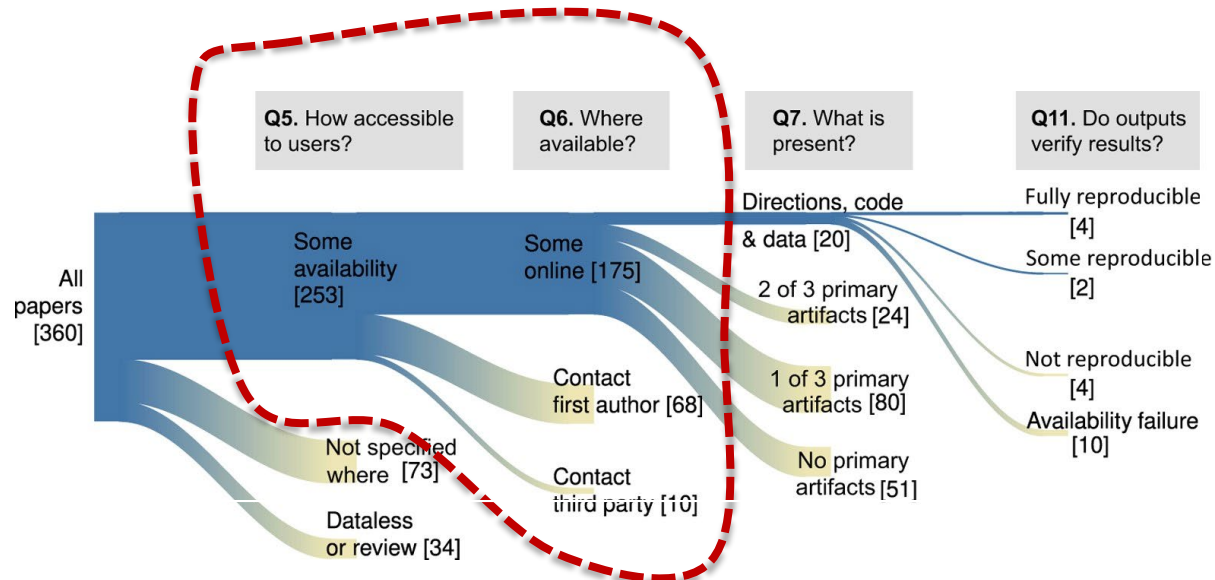


# Availability    Reproducibility



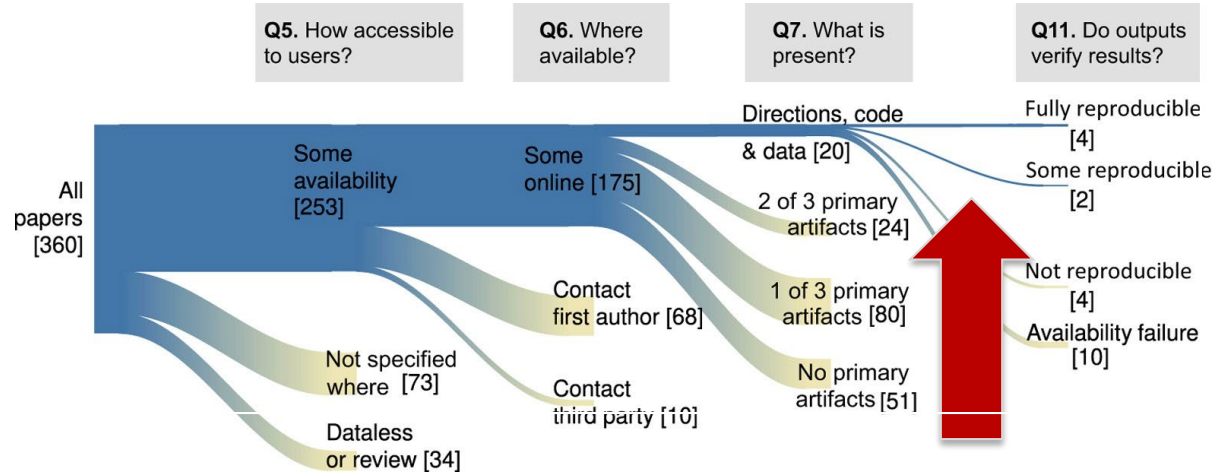
# Conclusions: Bottlenecks or Opportunities

- Data availability statement** appears to nudge towards availability



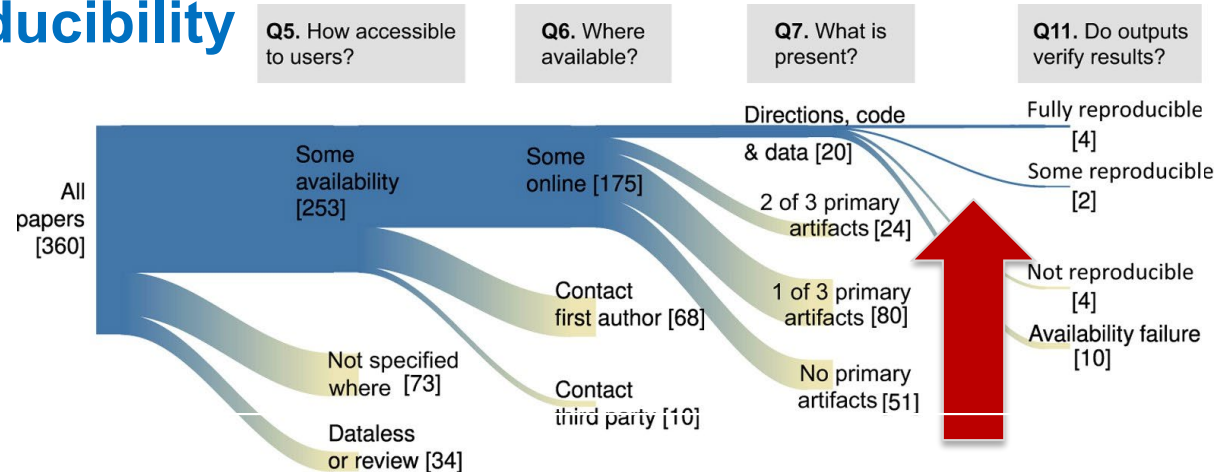
# Conclusions: Bottlenecks or Opportunities

- Data availability statement appears to nudge towards availability
- **Experiment directions/documentation** vs. data dump
- Nudge towards documentation: repository tools, templates



# Conclusions: Bottlenecks or Opportunities

- Data availability statement appears to nudge towards availability
- Directions/documentation vs. data/code dump
- Nudge towards documentation: repository tools, templates
- When artifacts available, **high likelihood of reproducibility**



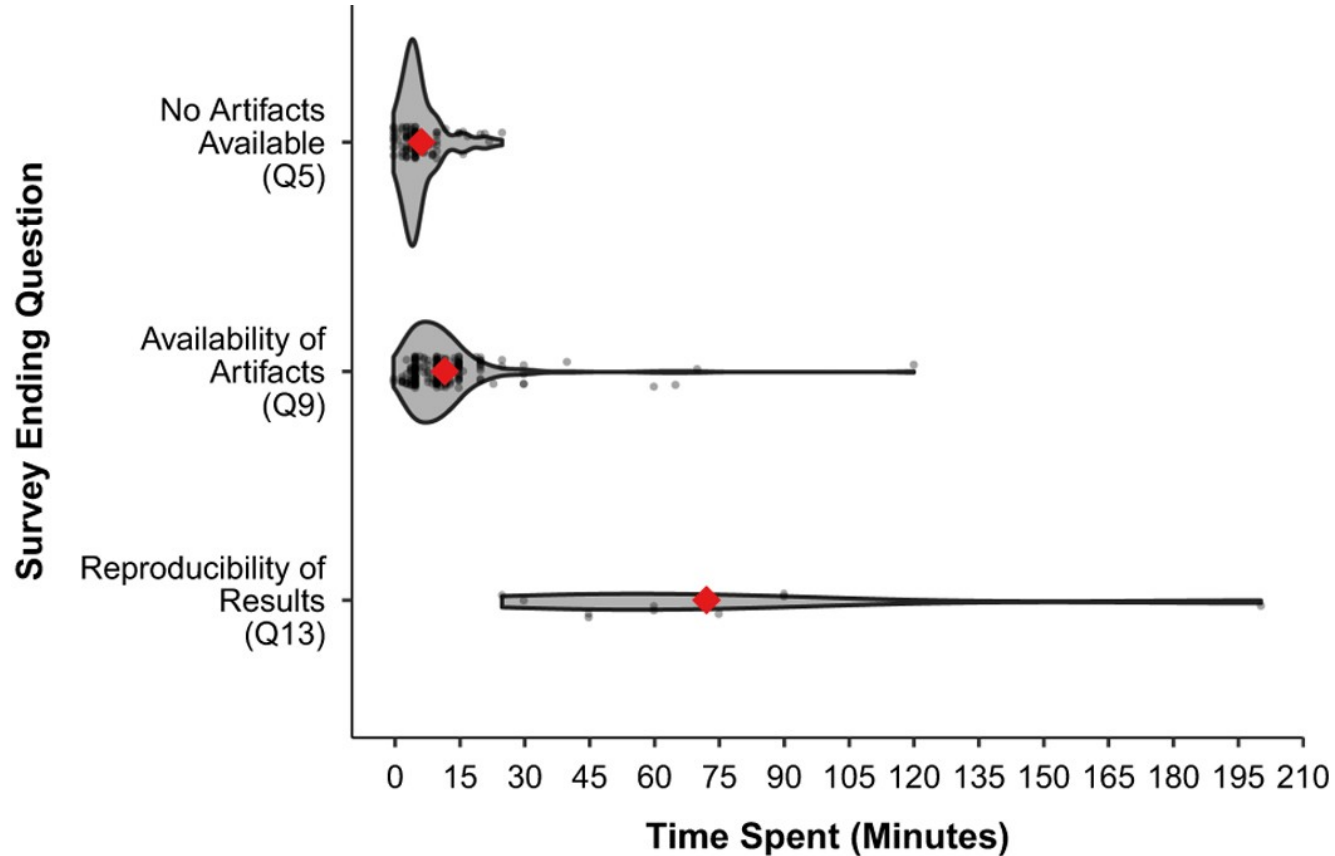
# Acknowledgements

- Co-authors: David Rosenberg, Adel Abdallah, Hadia Akbar, Nour Attallah, and Ryan James
- Journals and Authors for making their work reproducible



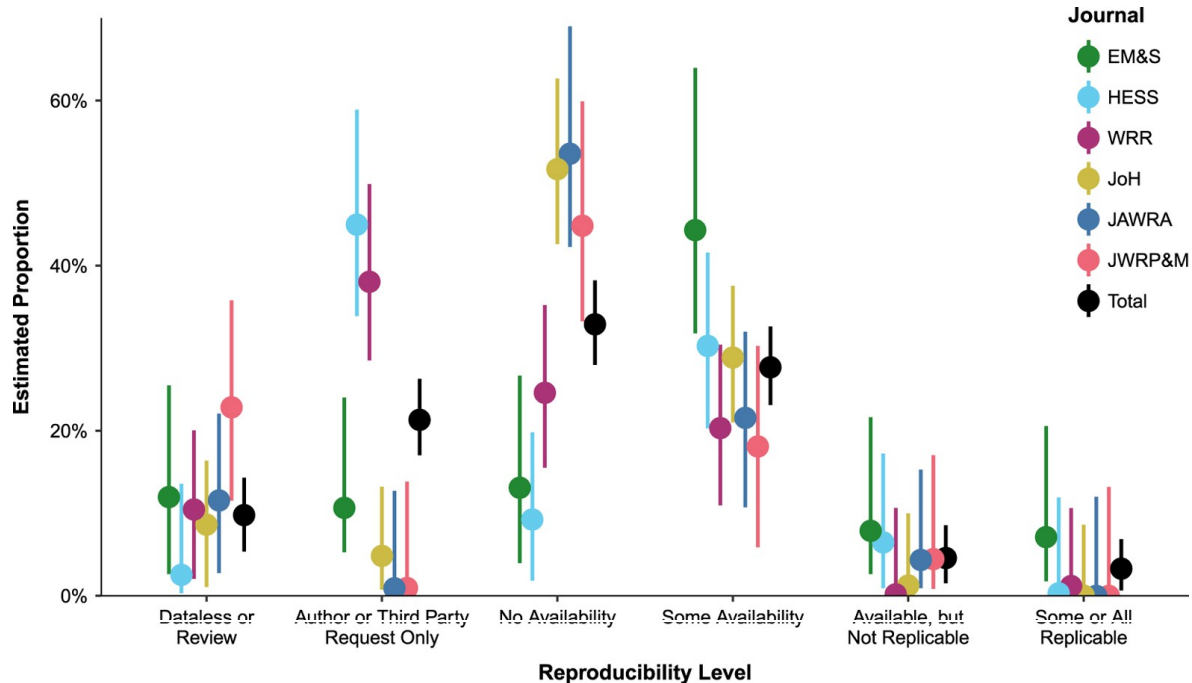


# Time Spent



# Population Estimates

- Oversampled for reproducible keywords
- 1.6% reproduced in sample
- 0.6 – 6.8% population estimate



# Sampled articles

	<b>EM&amp;S</b>		<b>HESS</b>		<b>WRR</b>		<b>JoH</b>		<b>JAWRA</b>		<b>JWRP&amp;M</b>	
	<i>2017</i>	<i>Sample</i>	<i>2017</i>	<i>Sample</i>	<i>2017</i>	<i>Sample</i>	<i>2017</i>	<i>Sample</i>	<i>2017</i>	<i>Sample</i>	<i>2017</i>	<i>Sample</i>
Keyword	49	48	9	9	23	23	24	24	7	7	8	8
Non-keyword	181	15	319	43	511	59	645	79	102	23	111	22
Total	230	63	328	52	534	82	669	103	109	30	119	30