Assessing the Availability of Source Code in Computational Physics

Extended Abstract

Matthew Scott Krafczyk
National Center for Supercomputing Applications
Champaign, Illinois
krafczyk.matthew@gmail.com

Victoria Stodden
University of Illinois at Urbana-Champaign
Champaign, Illinois
vcs@stodden.net

Yantong Zheng
University of Illinois at Urbana-Champaign
Champaign, Illinois
yzheng56@illinois.edu

David Wong
University of Illinois at Urbana-Champaign
Champaign, Illinois
dawong2@illinois.edu

ABSTRACT

Replicability of scientific work based on computation is a subject which has been receiving increased scrutiny recently. One approach to replicating a computational finding is to run the original source code. Availability of source code however is not routine; Only 3/33 computationally based article authors released source code from JASA in 2006, and a 2015 study showed that only 44% of computer science article authors released their source code. The field of Computational Physics has yet to be examined in this way, nor has the effect of author knowledge of such a study been measured.

We present our findings regarding the availability of source code in recent articles of the Journal of Computational Physics as well as how author knowledge of the study affects their willingness to make source code available.

This work extends current reproducibility efforts being explored by the ACM, SIGHPC, and the SC conference community.

CCS CONCEPTS

• **General and reference** → **Validation**; **Verification**; • **Software and its engineering** → *Software creation and management*; *Empirical software validation*;

KEYWORDS

Code Availability, Data Availability, Reproducibility, Replicability

ACM Reference format:
Matthew Scott Krafczyk, Victoria Stodden, Yantong Zheng, and David Wong. 2017. Assessing the Availability of Source Code in Computational Physics. In *Proceedings of Supercomputing 2017 Conference, Denver, Colorado USA, November 2017 (SC'17),* 3 pages.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

SC'17, November 2017, Denver, Colorado USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

In recent years a few fields of science have been addressing new reproducibility issues. Precipitated by a few pioneering studies [6, 12, 13], efforts and awareness have expanded within fields such as psychological science [7], cancer research [2], geophysics [14], epidemiology [17], economics [5], astronomy [10], and density-functional theory [15].

Experimental sciences are often time and resource intensive and so replicating them is as well. Computational sciences are different from purely experimental sciences in that the craft produces an inherently replicable record: the code. Results dependent on this record should in principle be verifiable on nearly any computer (barring special hardware constraints)[9]. While it's true that a skilled practitioner may be capable of reimplementing an article's methods and produce the same results, this should not be necessary as a first step, nor does it permit the reconciliation of differences between independent implementations[4, 6, 9].

Availability of source code and other related information such as input data or configuration information, in contrast, is a new issue, with only 3/33 computationally based article authors releasing source code from the Journal of the American Statistical Association (JASA) in 2006[21], and a 2015 study showed that only about 44% of computer science article authors consented to releasing their source code to an interested researcher [8]. Differing reasons are given when code is not shared, including but not limited to: messy code, monolithic code-base, no documentation, too much time, and restrictive licensing [22].

To gain insight into this phenomena, we conducted a study of source code availability within the Journal of Computational Physics (JCP), which some HPC projects use to publish their work and one of the leading journals in the field. As of this writing, JCP journal policy encourages authors to share data and code.

Throughout this abstract and poster, when 'source code/data' is mentioned it is meant to encompass not only source code and input data, but other related or necessary information to replicate the work such as hardware information, input configuration information, library versions, processing scripts, or meshes.

This work extends current reproducibility efforts being explored by the ACM[3], SIGHPC, and SC conference community[23–25].

Table 1: Article Classification in order of increasing code/data information disclosed in the article before emailing authors.

Classification	Count	
None	180	58.8%
Some Info, No Code or Data	108	35.3%
At Least Some Code or Data	18	5.9%
Total Articles	306	100.0%

Table 2: Email Classification in order of increasing code and data disclosure

Classification	Informed Group		Uninformed Group		Total Emails	
No Response	69	45.7%	41	27.9%	110	36.9%
No Code Or Data	60	39.7%	77	52.4%	137	46.0%
Some Code Or Data	22	14.6%	29	19.7%	51	17.1%
Total Emails	151	100.0%	147	100.0%	298	100.0%

2 METHODS

This study concerns articles whose main conclusions were arrived at by running code in volumes 322–331[1] of JCP and proceeded in two phases. First each article was classified according to the level of information it contained about the code/data used. Authors who did not include or release enough information to replicate the article results were then emailed with a customized request for the relevant code/data. The articles were randomly divided into two groups: One group was informed of this replication study, and the other remained uninformed. Since we planned to collect data by contacting authors we sought and received Institutional Review Board approval (IRB #17329) prior to commencing the study. There were 153 articles in the informed group and 147 articles in the uninformed group. Author responses to our email request were then classified according to how much code or information they disclosed.

2.1 Scanning and classification of articles

306 articles across 10 volumes of JCP published in 2016 and 2017 were examined by 3 coders to determine what kind of information they contained about the code/data written for the article. Table 1 contains a breakdown of the article classifications.

2.2 Emailing of corresponding authors

Three hundred articles were determined to have not shared enough code/data to completely replicate their results. The corresponding author on each of these studies was sent an email request for this code/data. Two pairs of articles shared the same corresponding author; We sent 298 initial emails as these corresponding authors were contacted about both articles in a single email. The emails were sent from one of two undergraduate students (one for each group) to minimize bias through potential name recognition and to test the ability of junior scholars to obtain responses. A follow-up email was sent after two weeks if there was no response.

We classified the author replies based on how much information regarding the article's code/data we received. These classifications are shown in Table 2.

3 RESULTS

Using Tables 1 and 2 we pose and answer three questions regarding the informed and uninformed groups:

(1) **Does email response rate differ significantly**
Grouping the email responses into two super-classes 'No Response' and 'Response' and applying the Wilcoxon Ranked Sum Test [1, 20] reveals that the response rates between the informed and uninformed groups are significantly different.

(2) **Is disclosure behavior significantly different?**
Grouping the email responses into three super-classes 'No Response', 'Replied but no code/data', and 'Partial to full code/data disclosure' and applying the Pearson Chi-squared [11, 19] reveals that the code disclosure behavior between informed and uninformed groups is significantly different.

(3) **Does the code/data disclosure rate differ significantly?**
Grouping the email responses into two super-classes 'Code/data disclosed' and 'Code/data not disclosed' and applying the Wilcoxon Ranked Sum Test [1, 20] reveals that code/data disclosure rate between the two groups is not significantly different.

4 DISCUSSION

Most striking to us was the remarkably low code disclosure rate in the articles themselves. Only 18 articles either made available or indicated that they would make available some of the code/data backing their article.

In total, 188 out of 298 authors responded to our requests for source code. Eighty-two out of 151 authors responded in the Informed group while 110 out of 147 authors responded in the Uninformed group. This is a 34% response rate increase when we did not tell the authors that they were part of a replicability study.

This response rate difference is statistically significant for our sample size, however given that total code disclosure was not significantly different, we can also conclude that the difference is attributable to additional people indicating that they would not be sharing their code.

5 CONCLUSION

It is currently difficult to obtain source code/data from researchers. We found an overall disclosure rate of 17%, and a significant difference in response rates if informed that the request is part of a study. Statistical analysis of the data in this poster was performed using the R language [18]. You can download the scripts used to perform this analysis as well as the email templates at [16].

6 ACKNOWLEDGEMENTS

We thank Adhithya Baskar, Alex Dickinson, and Geraldine Padilla for excellent research assistance and diligent review of the source code we received from authors and the author demographic data.

[1]These volumes represent the published volumes from the start of the study until we had accumulated about 300 articles.

REFERENCES

[1] David F. Bauer. 1972. Constructing Confidence Sets Using Rank Statistics. *J. Amer. Statist. Assoc.* 67, 339 (1972), 687–690. https://doi.org/10.1080/01621459.1972.10481279 arXiv:http://amstat.tandfonline.com/doi/pdf/10.1080/01621459.1972.10481279

[2] C. Glenn "Begley and Lee M." Ellis. "2012". "Drug development: Raise standards for preclinical cancer research". *"Nature"* "483", "7391" ("29" "Mar" "2012"), "531–533". https://doi.org/"10.1038/483531a"

[3] Ronald F. Boisvert. 2016. Incentivizing Reproducibility. *Commun. ACM* 59, 10 (Sept. 2016), 5–5. https://doi.org/10.1145/2994031

[4] Jonathan B. Buckheit and David L. Donoho. 1995. *WaveLab and Reproducible Research.* Springer New York, New York, NY, 55–81. https://doi.org/10.1007/978-1-4612-2544-7_5

[5] Andrew C. Chang and Phillip Li. 2015. Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say âĂIUsually NotâĂİ. (2015). https://doi.org/10.17016/FEDS.2015.083

[6] Jon F. Claerbout and Martin Karrenbach. 1992. Electronic documents give reproducible research a new meaning. In *SEG Technical Program Expanded Abstracts*. 601–604. https://doi.org/10.1190/1.1822162

[7] Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. *Science* 349, 6251 (2015). https://doi.org/10.1126/science.aac4716 arXiv:http://science.sciencemag.org/content/349/6251/aac4716.full.pdf

[8] Christian Collberg and Todd A. Proebsting. 2016. Repeatability in Computer Systems Research. *Commun. ACM* 59, 3 (Feb. 2016), 62–69. https://doi.org/10.1145/2812803

[9] David L. Donoho, Arian Maleki, Inam Ur Rahman, Morteza Shahram, and Victoria Stodden. 2009. Reproducible Research in Computational Harmonic Analysis. *Computing in Science & Engineering* 11, 1 (2009), 8–18. https://doi.org/10.1109/MCSE.2009.15 arXiv:http://aip.scitation.org/doi/pdf/10.1109/MCSE.2009.15

[10] Ji hoon Kim, Oscar Agertz, Romain Teyssier, Michael J. Butler, Daniel Ceverino, Jun-Hwan Choi, Robert Feldmann, Ben W. Keller, Alessandro Lupi, Thomas Quinn, Yves Revaz, Spencer Wallace, Nickolay Y. Gnedin, Samuel N. Leitner, Sijing Shen, Britton D. Smith, Robert Thompson, Matthew J. Turk, Tom Abel, Kenza S. Arraki, Samantha M. Benincasa, Sukanya Chakrabarti, Colin DeGraf, Avishai Dekel, Nathan J. Goldbaum, Philip F. Hopkins, Cameron B. Hummels, Anatoly Klypin, Hui Li, Piero Madau, Nir Mandelker, Lucio Mayer, Kentaro Nagamine, Sarah Nickerson, Brian W. OâĂŹShea, Joel R. Primack, Santi Roca-FÃăbrega, Vadim Semenov, Ikkoh Shimizu, Christine M. Simpson, Keita Todoroki, James W. Wadsley, John H. Wise, and AGORA Collaboration. 2016. The AGORA High-resolution Galaxy Simulations Comparison Project. II. Isolated Disk Test. *The Astrophysical Journal* 833, 2 (2016), 202. http://stacks.iop.org/0004-637X/833/i=2/a=202

[11] Adery C. A. Hope. 1968. A Simplified Monte Carlo Significance Test Procedure. *Journal of the Royal Statistical Society. Series B (Methodological)* 30, 3 (1968), 582–598. http://www.jstor.org/stable/2984263

[12] John P. A. Ioannidis. 2005. Why Most Published Research Findings Are False. *PLOS Medicine* 2, 8 (08 2005). https://doi.org/10.1371/journal.pmed.0020124

[13] Ioannidis JA. 2005. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 294, 2 (2005), 218–228. https://doi.org/10.1001/jama.294.2.218 arXiv:/data/journals/jama/4983/joc50060.pdf

[14] Gary King. 1995. Replication, Replication. *PS: Political Science and Politics* 28 (September 1995), 444–452. See updates to this paper for how I use this paper as a class assignment now.

[15] Kurt Lejaeghere, Gustav Bihlmayer, Torbjörn Björkman, Peter Blaha, Stefan Blügel, Volker Blum, Damien Caliste, Ivano E. Castelli, Stewart J. Clark, Andrea Dal Corso, Stefano de Gironcoli, Thierry Deutsch, John Kay Dewhurst, Igor Di Marco, Claudia Draxl, Marcin Dułak, Olle Eriksson, José A. Flores-Livas, Kevin F. Garrity, Luigi Genovese, Paolo Giannozzi, Matteo Giantomassi, Stefan Goedecker, Xavier Gonze, Oscar Grånäs, E. K. U. Gross, Andris Gulans, François Gygi, D. R. Hamann, Phil J. Hasnip, N. A. W. Holzwarth, Diana Iuşan, Dominik B. Jochym, François Jollet, Daniel Jones, Georg Kresse, Klaus Koepernik, Emine Küçükbenli, Yaroslav O. Kvashnin, Inka L. M. Locht, Sven Lubeck, Martijn Marsman, Nicola Marzari, Ulrike Nitzsche, Lars Nordström, Taisuke Ozaki, Lorenzo Paulatto, Chris J. Pickard, Ward Poelmans, Matt I. J. Probert, Keith Refson, Manuel Richter, Gian-Marco Rignanese, Santanu Saha, Matthias Scheffler, Martin Schlipf, Karlheinz Schwarz, Sangeeta Sharma, Francesca Tavazza, Patrik Thunström, Alexandre Tkatchenko, Marc Torrent, David Vanderbilt, Michiel J. van Setten, Veronique Van Speybroeck, John M. Wills, Jonathan R. Yates, Guo-Xu Zhang, and Stefaan Cottenier. 2016. Reproducibility in density functional theory calculations of solids. *Science* 351, 6280 (2016). https://doi.org/10.1126/science.aad3000 arXiv:http://science.sciencemag.org/content/351/6280/aad3000.full.pdf

[16] Y.Zheng D.Wong M.S. Krafczyk, V.Stodden. 2017. Digital Materials for the Krafczyk-Stodden poster at SC'17. (2017). https://github.com/ReproducibilityInPublishing/Krafczyk-Stodden-SC17-Poster-Materials

[17] Roger D. Peng, Francesca Dominici, and Scott L. Zeger. 2006. Reproducible Epidemiologic Research. *American Journal of Epidemiology* 163, 9 (2006), 783–789. https://doi.org/10.1093/aje/kwj093 arXiv:/oup/backfile/content_public/journal/aje/163/9/10.1093/aje/kwj093/2/kwj093.pdf

[18] R Development Core Team. 2008. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org ISBN 3-900051-07-0.

[19] R Development Core Team. 2017. Pearson's Chi-squared Test for Count Data. (2017). https://stat.ethz.ch/R-manual/R-devel/library/stats/html/chisq.test.html

[20] R Development Core Team. 2017. Wilcoxon Rank Sum and Signed Rank Tests. (2017). https://stat.ethz.ch/R-manual/R-devel/library/stats/html/wilcox.test.html

[21] Victoria Stodden. 2009. How Computational Science is Changing the Scientific Method. (29 July 2009). https://web.stanford.edu/~vcs/talks/Science20July2009VictoriaStodden.pdf

[22] Victoria Stodden. 2010. The Scientific Method in Practice: Reproducibility in the Computational Sciences. *MIT Sloan Research Paper* 4773-10 (9 February 2010). https://doi.org/10.2139/ssrn.1550193

[23] Michela Taufer. 2016. SC16 Explores Reproducibility for Advanced Computing Through Student Cluster Competition. (16 March 2016). http://sc16.supercomputing.org/2016/03/16/sc16-explores-reproducibility-advanced-computing-student-competition-michela-taufer/index.html

[24] Rich Vuduc (Ed.). 2016. *ACM SIGHPC Connect* 5, 1 (2016).

[25] Rich Vuduc (Ed.). 2017. *ACM SIGHPC Connect* 5, 2 (2017).