|  |  |
| --- | --- |
|  | Software Carpentry  164 Ashdale Avenue  Toronto, Ontario M4L 2Y9  gvwilson@software-carpentry.org  +1 (416) 435 9779  December 29, 2012 |

Dear Dr. Stopak,

We are writing to ask PNAS to consider publishing our attached paper, “Best Practices for Scientific Computing”, as a Perspectives piece.

Science in all disciplines increasingly relies on computational approaches. From modeling and simulation to data manipulation and analysis, scientists use computers as a core part of their research process [1,2]. However, few scientists are aware of good practices for developing and using software in a rigorous and efficient manner. As a result, their work is often not reproducible and can include serious errors that can mislead the development of scientific ideas for years (e.g, [3,4,5]). Just as importantly, the lack of good “lab skills” for computational work means that many researchers spend far more time than necessary doing simple things, and correspondingly less time on the core scientific questions.

The growing recognition of these problems has triggered a flurry of editorials, commentary pieces, and reviews (e.g., [6,7,8]). However, none of these have explainedwhat scientists should actually do, or which software engineering “best practices” would improve the reproducibility, correctness, and efficiency of their work. Our paper – written with authors from a broad range of disciplines including physics, biology, mathematics, and engineering – aims to fill that gap. We outline a number of key software engineering practices that can improve computational science, discuss the reasons for these approaches, and provide citations to scientific research supporting their use. Rather than provide a “one size fits all” prescription, we lay out when and why specific practices should be used.

This manuscript has been specifically written to appeal to scientists and to demonstrate why these practices are important in the context of their research. While it may be somewhat atypical for a Perspectives piece, we believe that publication in a high profile, cross-disciplinary journal is crucial to improving this core aspect of scientific research, and we would be happy to adjust the format and language as needed.

Several Members of the National Academy of Sciences have agreed to support the paper, including:

James Brown (University of New Mexico)

Richard Lenski (Michigan State University)

Paul Sternberg (California Institute of Technology)

James M. Tiedje (Michigan State University)

Prof. Lenski is willing to serve as the Editorial Board member in charge of managing the review process.

As potential reviewers, we would suggest:

Carole Goble, University of Manchester (carole.goble@manchester.ac.uk)

Hilmar Lapp, NEScent ([hlapp@nescent.org](mailto:hlapp@nescent.org))

Stephanie Hampton, NCEAS and UCSB ([hampton@nceas.ucsb.edu](mailto:hampton@nceas.ucsb.edu))

Miriam Goodman, Stanford ([mbgoodman@stanford.edu](mailto:mbgoodman@stanford.edu))

Perry Greenfield, Space Telescope Science Institute ([perry@stsci.edu](mailto:perry@stsci.edu))

Andrew Lumsdaine, Indiana University (lums@cs.indiana.edu)

Victoria Stodden, Columbia University ([vcs@stodden.net](mailto:vcs@stodden.net))

Thank you for your consideration – we look forward to your response.

Yours truly,

Dr. Gregory V. Wilson

Director, Software Carpentry

Adjunct Professor, Dept. of Computer Science, University of Toronto

[1] J.E. Hannay et al.: “How Do Scientists Develop and Use Scientific Software?” *Proc. 4th International Workshop on Software Engineering for Computational Science and Engineering*, 2009; DOI:10.1109/SECSE.2009.5069155.

[2] P. Prabhu et al.: “A Survey of the Practice of Computational Science.” *Proc. 24th ACM/IEEE Conference on High Performance Computing, Networking, Storage and Analysis*, 2011; doi.acm.org/10.1145/2063348.2063374.

[3] G. Chang: “Retraction of 'Structure of MsbA from *Vibrio cholera*: A Multidrug Resistance ABC Transporter Homolog in a Closed Conformation' [J. Mol. Biol. (2003) 330 419–430]” *Journal of Molecular Biology*, 369(2), 2007.

[4] D.C. Lees and R.K. Colwell: “A strong Madagascan rainforest MDE and no equatorward increase in species richness: re-analysis of 'The missing Madagascan mid-domain effect', by J.T. Kerr, M. Perring & D.J. Currie [Ecology Letters 9:149–159, 2006]” *Ecology Letters*, 10(9), 2007.

[5] D.A. Kelt et al.: “Differential Responses of Two Species of Kangaroo Rat (*Dipodomys*) to Heavy Rains: A Humbling Reappraisal.” *Journal of Mammalogy*, 89(1), 2008.

[6] A. Morin et al.: “Shining Light into Black Boxes.” *Science*, 2012; DOI:10.1126/science.1218263, <http://www.sciencemag.org/content/336/6078/159.summary>

[7] Z. Merali: “Computational Science: …Error” *Nature*, 2010; doi:10.1038/467775a, <http://www.nature.com/news/2010/101013/full/467775a.html>

[8] G. Miller: “A Scientist’s Nightmare.” *Science*, 2006; DOI:10.1126/science.314.5807.1856, <http://www.sciencemag.org/content/314/5807/1856.summary>