# Report proposes standards for sharing data and code used in computational studies

December 8 2016



'Data, code and workflows need to be part of what gets disclosed as a scientific finding,' said University of Illinois information science professor Victoria Stodden, who with her colleagues is proposing a new framework for sharing data sets and software code in scientific publications. Credit: L. Brian Stauffer

Reporting new research results involves detailed descriptions of methods

and materials used in an experiment. But when a study uses computers to analyze data, create models or simulate things that can't be tested in a lab, how can other researchers see what steps were taken or potentially reproduce results?

A new report by prominent leaders in computational methods and reproducibility lays out recommendations for ways researchers, institutions, agencies and journal publishers can work together to standardize sharing of data sets and software code. The paper "Enhancing reproducibility for computational methods" appears in the journal *Science*.

"We have a real issue in disclosure and reporting standards for research that involves computation—which is basically all research today," said Victoria Stodden, a University of Illinois professor of information science and the lead author of the paper. "The standards for putting enough information out there with your findings so that other researchers in the area are able to understand and potentially replicate your work came from before we used computers."

"It is becoming increasingly accepted for researchers to value open data standards as an essential part of modern scholarship, but it is nearly impossible to reproduce results from original data without the authors' code," said Marcia McNutt, the president of the National Academy of Sciences and a co-corresponding author of the study. "This policy forum makes recommendations to enable practical and useful code sharing."

Sharing complete computational methods—data, code, parameters and the specific steps taken to arrive at the results—is difficult for researchers because there are no standards or guides to refer to, Stodden said. It's an extra step for busy researchers to incorporate into their reporting routine, and even if someone wants to share their data or code, there are questions of how to format and document it, where to store it

and how to make it accessible.

The report makes seven specific recommendations, such as documenting digital objects and making them retrievable, open licensing, placing links to datasets and workflows in scientific articles, and reproducibility checks before publication in a scholarly journal. The authors hope that disclosing computational methods will not only allow other researchers to verify and reproduce results, but also to build upon studies that have been done, such as performing different analyses with a dataset or using an established workflow with new data.

"Things like how you prepped your data—what you did with outliers or how you normalized variables, all the things that are standard in data analysis—can make a big impact on results," Stodden said. "Some researchers make code and data accessible on point of principle, so it's possible. But it takes time. We know it's hard, but in this report we're trying to say in a very productive and positive way that data, code and workflows need to be part of what gets disclosed as a scientific finding."

**More information:** "Enhancing reproducibility for computational methods," *Science*, science.sciencemag.org/cgi/doi … 1126/science.aah6168

Provided by University of Illinois at Urbana-Champaign

Citation: Report proposes standards for sharing data and code used in computational studies (2016, December 8) retrieved 10 April 2024 from https://phys.org/news/2016-12-standards-code.html