

Scientists, Share Secrets or Lose Funding

By Victoria Stodden and Samuel Arbesman - Jan 9, 2012

The Journal of Irreproducible Results, a science-humor [magazine](#), is, sadly, no longer the only publication that can lay claim to its title. More and more published scientific studies are difficult or impossible to repeat.

It's not that the experiments themselves are so flawed they can't be redone to the same effect -- though this happens more than scientists would like. It's that the data upon which the work is based, as well as the methods employed, are too often not published, leaving the science hidden.

Many people assume that scientists the world over freely exchange not only the results of their experiments but also the detailed data, statistical tools and computer instructions they employed to arrive at those results. This is the kind of information that other scientists need in order to replicate the studies. The truth is, open exchange of such information is not common, making verification of published findings all but impossible and creating a credibility crisis in computational science.

Federal agencies that fund scientific research are in a position to help fix this problem. They should require that all scientists whose studies they finance share the files that generated their published findings, the raw data and the computer instructions that carried out their analysis.

The ability to reproduce experiments is important not only for the advancement of pure science but also to address many science-based issues in the public sphere, from climate change to biotechnology.

Consider, for example, a recent notorious incident in biomedical science. In 2006, researchers at Duke University seemed to have discovered relationships between lung cancer patients' personal genetic signatures and their responsiveness to certain drugs. The scientists published their results in respected journals (the New England Journal of Medicine and Nature Medicine), but only part of the genetic signature data used in the studies was publicly available, and the computer codes used to generate the findings were never revealed. This is unfortunately typical for scientific publications.

The Duke research was considered such a breakthrough that other scientists quickly became interested in replicating it, but because so much information was unavailable, it took three years for them to uncover and publicize a number of very serious errors in the published reports. Eventually, those reports were retracted, and clinical trials based on the flawed results were canceled.

In response to this incident, the Institute of Medicine convened a committee to review what data should appropriately be revealed from genomics research that leads to clinical trials. This committee is due to release its report early this year.

Unfortunately, the research community rarely addresses the problem of reproducibility so directly. Inadequate sharing is common to all scientific domains that use computers in their research today (most of science), and it hampers transparency.

By making the underlying data and computer code conveniently available, scientists could open a new era of innovation and growth. In October, the White House released a memorandum titled “Accelerating Technology Transfer and Commercialization of Federal Research in Support of High-Growth Businesses,” which outlines ways for federal funding agencies to improve the rate of technology transfer from government-financed laboratories to the private business sector.

In this memo, President Barack Obama called on federal agencies to measure the rate of technology transfer. To this end, agencies such as the National Institutes of Health and the National Science Foundation should require that scientists who receive federal funds publish full results, including the data they are based on and all the computer steps taken to reach them. This could include providing links to Internet sites containing the data and codes required to replicate the published results.

Exceptions could be made when necessary -- some information might need to be kept confidential for national-security reasons, for example. But standard practice for scientific publication should be full transparency.

Leaving this up to the scientific community isn't sufficient. Nor is relying on current federal rules. Grant guidelines from the NIH and the NSF instruct researchers to share with other investigators the data generated in the course of their work, but this isn't enforced. The NIH demands that articles resulting from research it finances be made freely available within a year of publication. But even if this policy were extended to all government-financed studies, the data and computer codes needed to verify the findings would still remain inaccessible.

As Jon Claerbout, a professor emeritus of geophysics at Stanford University, has noted, scientific publication isn't scholarship itself, but only the advertising of scholarship. The actual work -- the steps needed to reproduce the scientific finding -- must be shared.

Stricter requirements for transparency in publication would allow scientific findings to more quickly become fuel for innovation and help ensure that public policy is based on sound science.

(Victoria Stodden is an assistant professor of statistics at Columbia University. Samuel Arbesman is a senior scholar at the Ewing Marion Kauffman Foundation. The opinions expressed are their own.)