SUMMARY of Testimony, Dr. Victoria Stodden, Columbia University

Before the House Committee on Science, Space and Technology
Subcommittee on Research

Hearing On Scientific Integrity & Transparency, March 5, 2013

Federally Funded Digital Archives are Necessary for Scientific Integrity and Accelerate Scientific Discovery

Massive computation has begun a transformation of the scientific enterprise that will finish with computation absolutely central to the scientific method. Convenient access to the scientific data and software associated with published scientific findings is now a necessary step in enabling reproducibility in computational science. Not only will this allow the results to be validated, but others can use the data, the methods embodied in the software, and the scientific findings to accelerate scientific discovery, applications, commercialization, and STEM education. This is vital for future economic growth and competitiveness. For computational reproducibility, access to the scientific data and software associated with validating the published result is essential.

Because of their broad impact, the federal agencies that fund scientific research play a key role in facilitating the dissemination and archiving of the data and software associated with scientific findings. Data archives that are discipline specific and directly federally funded are necessary and permit others to use these resources to accelerate economic and scientific competitiveness. This is a solution to the public good problem of access to scientific data and code. I believe dedicated federal funding is required to establish and maintain such archives, since using research grant funds is unpredictable and unreliable. Funding agencies need to treat this as a mandate and plan to protect data and code availability for 25 years. Archived data and code should be linked with all publications that use either of them, in order for reproducibility to be effective. Researchers should receive citation credit for data and software contributions to the scholarly record.

The creation of directly federally funded data and code archives is supported by the OSTP Executive Memorandum released February 22, 2013: "Increasing Access to the Results of Federally Funded Scientific Research."

Data and software availability do not, by themselves, ensure reproducibility of published computational findings, but they are an essential step toward the solution.