Summary

Overview:

Computers are now ubiquitous in scientific research, but many researchers in science, technology, engineering, and mathematics (STEM) are still not *computationally competent*: they do repetitive tasks manually instead of automating them, develop software haphazardly, and fail to track their work in any systematic, reproducible way. This lack of foundational skills impedes their ability to do research, and prevents them from engaging with new opportunities in open and web-enabled science.

This project will train STEM undergraduates in these skills, and assess the impact of that training on their productivity and career paths. We will do this by running software skills workshops for undergraduates likely to go on to graduate school, and by tracking the alumni of these workshops over several subsequent years both to improve the training itself and to encourage wider adoption of our model.

More specifically, we will adapt the two-day workshops run by Software Carpentry for researchers at the graduate level and above to teach undergraduate students drawn primarily from programs such as the NSF's Research Experience for Undergraduates (REU) program. These workshops will cover fundamental skills that are prerequisites for open and web-enabled science, including how to automate repetitive tasks, how to track and share work over the web, how to grow a program in a modular, testable, reusable way, and how to create, use, and share structured data. All materials will be made freely available to other educators and institutions under an open access license in order to promote the greatest possible uptake.

Together, these workshops will reach over 2000 students during the course of the project. Alumni will serve as the treatment population for a five-year study by a full-time professional researcher in educational assessment, who will explore the impact of this training on their careers in general, and their involvement with open and web-enabled science in particular.

Intellectual merit:

Our main contribution to knowledge will be an assessment of ways in which training of this kind can accelerate the careers of participants and the science they do. We will discover whether students who receive this training are more likely than their peers to continue to graduate school, to incorporate open and web-enabled science tools and practices into their work, to choose computationally oriented research topics and careers, and/or to develop new computational tools and practices. We will also determine the extent to which training of this kind changes students' outlook on the practice of science itself, and whether it can help level the playing field for women and minorities currently underrepresented in computing and science.

Broader impacts:

In the medium term, this project will have broader impact through the creation and dissemination of teaching materials and practices that other educators and institutions can adopt. All of the materials produced by and for this project will be made freely available under the Creative Commons - Attribution (CC-BY) license, while the results of our studies of the program's effect will be shared with other educators through science education journals and conferences.

In the long term, the project's greatest impact will be on scientific competitiveness. Computing is no longer optional in any part of science: even scientists who don't think of themselves as doing computational work rely on computers to prepare papers, store data, and collaborate with colleagues. The better their computing skills are, the better able they will be to conduct world-class research that aids national economic competitiveness.