

Project Abstract:

An essential aspect of both translational medicine and human health-based team science is the integration of biomedical research results that are generated by multiple laboratories using widely varying experimental systems and data analysis methods. Beyond the uncertainties of experimental design and measurement, there are several critical points in the subsequent research workflow where lack of rigor and transparency may compromise the reproducibility of these laboratory results. In this proposal, we focus on data recording and data pre-processing as key steps of research workflows, and we propose to develop training modules that provide general principles, software tools, and exercises to broadly enhance data reproducibility of these steps in biomedical research. We aim to ensure these training modules are clear, relevant, and useful to laboratory-based researchers, whose attention is rather to their experimental technique and collection of accurate data, and who may have little or no background in the use of general purpose software tools. To ensure this, we will feature in these training modules examples from recent and ongoing NIH-funded microbiology and immunology research programs devoted to drug and vaccine development for infectious diseases at Colorado State University. There will be two instructional sequences of modules, “Improving the Reproducibility of Experimental Data Recording”, with eleven training modules, and “Improving the Reproducibility of Experimental Data Pre-Processing”, with nine training modules. The R programming language, and an ecosystem of related reproducibility tools, will form the technical basis for implementation of the modules in these sequences, while modules on principles and examples will be accessible to biomedical researchers regardless of programming knowledge. These training modules will be collectively published as an open on-line book using the *bookdown* technology, leveraging literate programming. Each module will form a chapter of this book, and will feature an embedded YouTube video of 10–25 minutes, with accompanying text in the book to provide trainees with a more detailed written reference they can refer to after completing the video module. Each module’s chapter will conclude with practical exercises or open discussion questions to complement the material taught in the video. To ensure this material is completely free and open to researchers in the United States, we will publish this online book, and its embedded videos and additional content, under a Creative Commons license.