

Project Title:

Training modules for improving the reproducibility of experimental data recording and pre-processing with examples from microbiology and immunology

Cover Letter:

Please accept the attached grant "Training modules for improving the reproducibility of experimental data recording and pre-processing with examples from microbiology and immunology" by Drs. Brooke Anderson and collaborators.

Best,

Brooke Anderson

Budget justification

First thing we need, more about why we need it.

Second thing we need, more about why we need it.

Project Abstract:

Summary of our proposed project.

Project Narrative:

Even shorter summary of our proposed project.

Specific Aims

A Significance

“Does the proposed program address a key audience and an important aspect or important need in training in rigor and reproducibility? Is there convincing evidence in the application that the proposed program will significantly advance the stated goal of the program?”

B Innovation

“Taking into consideration the nature of the proposed research education program, does the applicant make a strong case for this program effectively reaching an audience in need of the program’s offerings? Where appropriate, is the proposed program developing or utilizing innovative approaches and latest best practices to improve the knowledge and/or skills of the intended audience?”

C Approach

“Does the proposed program clearly state its goals and objectives, including the audience to be reached, the content to be conveyed, and the intended outcome? Is there evidence that the program is based on a sound rationale, as well as sound educational concepts and principles? Is the plan for evaluation sound and likely to provide information on the effectiveness of the program?”

C.1 Proposed Research Education Program Plan

“While the proposed research education program may complement ongoing research training and education occurring at the applicant institution, the proposed educational experiences must be distinct from those research training and research education programs currently receiving federal support. When research training programs are on-going in the same department, the applicant organization should clearly distinguish between the activities in the proposed research education program and the research training supported by the training program. The research education proposed must be **targeted to trainees and investigators at any level**. State the **goals for education** and **justify the area of training** selected for module development in terms of its **relevance and potential impact** on improving the development of skills and knowledge important for conducting rigorous and reproducible research. Describe the **subject material** to be covered. Describe the **format** for the training module proposed and **justify it in terms of the education goals**. The **length** of the proposed training module should be explained in terms of **scope and depth of coverage** of the subject matter. In addition, **how the research education will be utilized by trainees or investigators** should be described—for example, a module on how to avoid confirmation bias to be taken by all beginning laboratory workers, or a module on appropriate design of animal studies to be taken immediately prior to beginning such work. Describe the **plans for piloting and evaluating the effectiveness** of the training module. Describe **plans for making the proposed training module section 508 compliant of the Rehabilitation Act** (29 U.S.C. '794 d), as amended by the Workforce Investment Act of 1998 (P.L. 105 220; see <http://www.section508.gov/> for additional information). Provide a **time-line for module development, piloting and refinement, dissemination, evaluation, and maintenance**. This timeline must propose **making the training publicly available within two years** of the award date.”

C.1.1 Educational goals of the modules

C.1.2 Module subject material

We propose to develop two collections of modules, **Improving the Reproducibility of Experimental Data Recording** and **Improving the Reproducibility of Experimental Data Pre-Processing**.

The **Improving the Reproducibility of Experimental Data Recording** collection will include the following modules:

1. The principals of “tidy” data
2. Creating spreadsheet templates for experimental data collection
3. Example spreadsheet template: A template for collecting CPU data for a tuberculosis study [make more specific]
4. Choosing a spreadshet program for reproducible data collection: Excel, Google Sheets, and RStudio
5. Organizing data recording and meta-data recording through RStudio “Projects”
6. Creating “Project Templates” for consistency across projects
7. Example of creating a “Project Template”: A project template for a lab group studying tuberculosis [make more specific]
8. Harnessing version control (git and GitLab) to improve transparency in data recording
9. Using git from RStudio
10. Using GitLab for version controlled collaborations

The **Improving the Reproducibility of Experimental Data Pre-Processing** collection will include the following modules:

1. An introduction to R code scripts
2. The benefits of scripts for data pre-processing
3. Getting started with RMarkdown
4. The relationship between R code scripts and RMarkdown documents
5. Creating reproducible data pre-processing protocols using Rmarkdown
6. Example of a reproducible data pre-processing protocol: Automated gating for flow cytometry data
7. Example of a reproducible data pre-processing protocol: Measuring metabolite feature intensities for metabolomics LC/MS data
8. Complex data types in R and their use in Bioconductor packages
9. Converting from complex data types to “tidy” formats for data analysis and visualization with R’s “tidy” data tools

C.1.3 Format for the training modules

- Online book created through the “bookdown” format, with each module as a book chapter. We can use Git Pages to host this (CSU options for web hosting?).
- Training videos embedded for each module, each 5–30 minutes. Videos will be similar to online course lectures and will be hosted using YouTube. Embedding in the book will allow users to watch videos without leaving the book’s webpage.
- Each chapter will end with exercise questions (around 10 questions, combination of discussion questions and applied exercises), as well as an embedded video with discussion of the discussion questions and a detailed walk-through of answers to applied exercises.
- Possibly host this through an online course platform like DataCamp?

C.1.4 Piloting and evaluating effectiveness of training modules

C.1.5 Insuring compliance with Rehabilitation Act

C.2 Team

C.2.1 Program Director/Principal Investigator

“Is the PD/PI capable of providing both administrative and scientific leadership to the development and implementation of the proposed program? Is there evidence that an appropriate level of effort will be devoted by the program leadership to ensure the program’s intended goal is accomplished? If the project is collaborative or multi-PD/PI, do the investigators have complementary and integrated expertise; are their leadership approach, governance and organizational structure appropriate for the project?”

“Describe **arrangements for administration** of the program. Provide evidence that the Program Director/Principal Investigator is actively engaged in research and/or teaching in an area related to the mission of NIH, and can **organize, administer, monitor, and evaluate the research education program**. For programs proposing multiple PDs/PIs, describe the complementary and integrated expertise of the PDs/PIs; their leadership approach, and governance appropriate for the planned project.”

C.2.2 Other members of the team

C.3 Institutional Environment and Commitment

“Describe the institutional environment, reiterating the **availability of facilities and educational resources** (described separately under Facilities & Other Resources), that can contribute to the planned Research Education Program. Evidence of institutional commitment to the research educational program is required. A **letter of institutional commitment** must be attached as part of Letters of Support (see below). Appropriate institutional commitment should include the provision of adequate staff, facilities, and educational resources that can contribute to the planned research education program.”

C.4 Evaluation Plan

“Applications must include a plan for evaluating the activities supported by the award in terms of their **frequency of use** and their **usefulness**. The use of **multiple evaluation approaches** is highly encouraged as is **testing several groups with different characteristics**. The application must specify **baseline metrics (e.g., numbers, educational levels, and demographic characteristics of test group)** in a structured format, as well as **measures to gauge the short and long-term success of the research education award in achieving its objectives**. Applicants are expected to **obtain feedback from test group** to help identify weaknesses and to provide suggestions for improvements, and **make the evaluation and feedback data** available to NIGMS staff.”

Pilot / test group evaluation:

- Work with GAUSSI to use some students as pilot testers?
- Recruit researchers / faculty as pilot testers?
- Work with CSU’s Research Ethics group to figure out ways to pilot?

Long-term evaluation:

- Google Analytics for online book. How often are people accessing the book? How long are they spending on the book website? Where are the people accessing the book?
- YouTube analytics for the embedded videos. How often are people accessing the book? How long are they spending on the book website? Where are the people accessing the book?
- Quiz for each chapter of the book? Use to evaluate how well they’ve mastered the material? (Possibly could use embedded Shiny apps for this? Other ways to do this?)
- Rating options for each chapter of the online book? Usefulness? What they learned?
- Survey within each chapter of the online book? Educational level, demographic characteristics.

C.5 Dissemination Plan

“A specific plan must be provided to disseminate the finished training modules **nationally** and make them **freely accessible**. In addition, links to these modules will be posted and maintained on the NIGMS web site.”

C.6 Timeline

“Provide a timeline for **module development, piloting and refinement, dissemination, evaluation, and maintenance**. This timeline must propose making the training publicly available within two years of the award date.”

D Works cited

E Environment

“Will the scientific and educational environment of the proposed program contribute to its intended goals? Is there a plan to take advantage of this environment to enhance the educational value of the program? Is there tangible evidence of institutional commitment? Where appropriate, is there evidence of collaboration and buy-in among participating programs, departments, and institutions?”