

PROGRAM CONTACT:
Dr Kristine Willis
301-594-0943
kristine.willis@nih.gov

SUMMARY STATEMENT
(Privileged Communication)

Release Date: 04/20/2015

Application Number: 1 R25 GM116149-01

Principal Investigator

SCHLOSS, PATRICK DAVID PHD

Applicant Organization: UNIVERSITY OF MICHIGAN

Review Group: ZRG1 CB-P (50)

Center for Scientific Review Special Emphasis Panel

RFA-GM-15-006: Training Modules to Enhance Data Reproducibility (R25)

Meeting Date: 04/08/2015

RFA/PA: GM15-006

Council: MAY 2015

PCC: T230KW

Requested Start: 07/01/2015

Dual IC(s): AA, AG, AT, DE, EY, NS, RI

Project Title: Development of reproducible informatics skills among microbiome researchers

SRG Action: Impact Score: 17

Next Steps: Visit http://grants.nih.gov/grants/next_steps.htm

Human Subjects: E2-Human subjects involved - Exemption #2 designated

Animal Subjects: 10-No live vertebrate animals involved for competing appl.

Project Year	Direct Costs Requested	Estimated Total Cost
1	68,500	73,980
2	70,255	75,875
TOTAL	138,755	149,855

ADMINISTRATIVE BUDGET NOTE: The budget shown is the requested budget and has not been adjusted to reflect any recommendations made by reviewers. If an award is planned, the costs will be calculated by Institute grants management staff based on the recommendations outlined below in the **COMMITTEE BUDGET RECOMMENDATIONS** section.

1R25GM116149-01 Schloss, Patrick

RESUME AND SUMMARY OF DISCUSSION: This application proposes to develop a set of autotutorials to teach microbiome researchers habits for engaging in reproducible research. The reviewers noted that the project's focus is unique in addressing specific needs of the microbiome researchers; however, it was unanimously considered to be significant. The reviewers agreed that introducing important skills on documenting data analysis workflow and implementing computational lab notebooks that can be shared and collaborative are needed in this research area, and will be relevant for other bioinformatics research disciplines beyond microbiome. The investigator is experienced and has already piloted this approach. The reviewers noted the strong module design and the effective workshop format; however, it was neither clear nor explained how this workshop format will be extended to autotutorials, i.e. the delivered product. In addition, it was unclear who will be in charge of the development; concerns were raised about the proposed allocation of this role to a postdoctoral fellow. The evaluation plan was viewed as strong. Following the discussion, the concerns remained, but were viewed as minor. Overall, the panel members concluded that the project is appealing and expressed very high enthusiasm for this application, which has the potential to strongly advance research education and enhance data reproducibility.

DESCRIPTION (provided by applicant): Enthusiasm for understanding how changes in the human microbiome affect health has led to an influx of researchers who have little experience in documenting and performing bioinformatic analyses. Compounding this problem is that over the past 10 years microbiome datasets have grown from hundreds to millions of sequences. With the increased size of the datasets the complexity of the analyses has also grown. Researchers that once used spreadsheets to analyze their data now struggle to use command line tools. Traditional training programs have not been able to meet the needs of these researchers and so it is essential that instructional materials be developed to train these researchers on the best practices for documenting and disseminating their analyses so that they can be reproduced by others. The objective of this proposal is to develop a training module that microbiome researchers can use to improve the reproducibility and overall quality of their research. Aside from the general importance of insuring that all research is reproducible, the significant growth of the community makes it urgent that such instructional materials are developed now. The specific aim of the proposed effort will develop a set of autotutorials to teach microbiome researchers habits for engaging in reproducible research. This aim will be achieved through an iterative process of development, evaluation, and refinement using a wide network of microbiome researchers to assess the materials prior to broader dissemination. The expected outcomes of the proposed modules are the improvement of the reproducibility of research within the microbiome research community, increased accessibility to raw original data, and a greater use of literate programming tools for constructing manuscripts and oral presentations. Furthermore, if it is possible to improve the reproducibility of the original research, then it will be more likely that other researchers will use those data, results, and methods to perform additional analyses resulting in a greater understanding of the microbiome. Such a process is rare within the microbiome literature given the vast size of many of these datasets. The long-term goal of this project is to establish a broader community- supported resource devoted to disseminating best practices in performing reproducible microbiome research. Given the significant role of the microbiome in human health and the significant growth in our understanding of how it shapes health and disease, improving the reliability of the results from these studies will have a meaningful positive impact. This project will yield a significant vertical step in the field because it will put tools into the hands of researchers performing microbiome-focused studies empowering them to perform sophisticated and reproducible analyses. The approach taken in the proposed research is innovative because it represents the first concentrated effort to develop formal, public, and open training modules directed at the microbiome research community. Finally, we anticipate that the materials we develop for the microbiome research community will be easily disseminated across other bioinformatics research disciplines.

PUBLIC HEALTH RELEVANCE: The proposed research is relevant to public health because it supports researchers within the domain of microbiome research, who have shown that it is impossible to separate human health from the structure and function of the human microbiome. Thus the research is relevant to the part of NIH's mission that pertains to the development, maintenance, and renewal of scientific resources that will assure our ability to perform robust and reproducible research in order to prevent disease.

CRITIQUE 1:

Significance: 1
Investigator(s): 2
Innovation: 4
Approach: 3
Environment: 1

Overall Impact: This is an excellent proposal. The planned modules are clearly laid out, and should provide training in several of the best tools available today. The list of supporting letters and arrangements for dissemination of the material are impressive. Providing case studies on GitHub in analyzable form can give users some of the hands-on experience required. The fact that the development of this proposal itself is reviewable in similar fashion is a nice touch. While these tools will definitely be of use to the microbiome community where the PI has a strong record, generalization for other audiences (e.g., molecular biologists, data analysts) should be straightforward. The engagement described should work quite well in a workshop setting.

It is less clear how the “autotutorials” will work in a context without direct involvement with a facilitator, and more clarity with respect to how much development will devolve onto the postdoc requested would help.

1. Significance:

Strengths

- Data and analysis tracking is becoming increasingly important for reproducibility. If anything, the importance is monotonically increasing with dataset size.
- The tools to be taught (R, markdown, GitHub) represent “best of breed”.

Weaknesses

- There is a somewhat unavoidable requirement for some programming experience beforehand.

2. Investigator(s):

Strengths

- The PI is an established researcher with a track record of pursuing and teaching methods for reproducibility.
- He has assembled a list of case study examples which could be useful for exploration.
- He has taught this material before.

Weaknesses

- None noted.

3. Innovation:

Strengths

- The key innovation is really the class of tools being taught. These deal with problems arising with increasing frequency in the biological sciences.

Weaknesses

- Some of this material is covered in courses already available on the web or through Coursera (Google for “reproducible research courses”).

4. Approach:

Strengths

- The module layout is very good, and the summaries are adequately detailed.
- Having the core material on GitHub is useful. Having this proposal there as well is a nice touch.
- The consideration of engagement in learning is nice to see, and should work well in a workshop setting.

Weaknesses

- It is less clear how well engagement and explanation can be triggered in an “autotutorial” session without access to a facilitator.
- The role of the postdoc in module development and dissemination should be clarified.
- I suspect the timings for introducing git are optimistic for a broad audience.
- It might be worth pointing out the limitations of reproducibility in module 1: just because it's reproducible doesn't mean it's right.

5. Environment:

Strengths

- The environment the PI has established for field testing these modules is exemplary, including buy-in from several investigators at other institutions and a professional society willing to run the proposed workshops for the next few years.

Weaknesses

- None noted.

Protections for Human Subjects:

No concerns. Human Subjects exemption 2 is considered and is justified.

Vertebrate Animals:

Not Applicable

Biohazards:

Not Applicable

Recruitment & Retention Plan to Enhance Diversity:

Not Applicable

Training in the Responsible Conduct of Research:

Not Applicable

Budget and Period of Support:

Recommend as Requested

CRITIQUE 2:

Significance: 1

Investigator(s): 1

Innovation: 2

Approach: 4

Environment: 2

Overall Impact: Documenting analyses using annotations of workflow is very important. This is one of the chief problems I encounter with undergraduates and graduate students. This proposal provides an exceptional set of illustrations about how autotutorials on documenting data analysis workflow annotations, computational lab notebooks that can be shared and collaborative. Design of modules is pedagogically sound and uses appropriate active learning techniques. Some concern about the nature of the GitHub platform and command line interfaces, as well as need for programming experience in some areas. The modules outlined however are well designed as is the assessment methods. Overall, this is a very good proposal with a few moderate concerns.

1. Significance:

Strengths

- Identifies extremely important issues in microbiome bioinformatics, not the least of which is that senior investigators may also lack these skill sets.
- Provides interesting and informative outlines to challenges and solutions of providing annotative, collaborative pathways of analysis to increase reproducibility.
- Large community developed through his software package who will test and review materials.
- Diverse participants in previous workshops and community of researchers.

Weaknesses

- May need to think more about how the training in R, command line and other areas may limit participation or discourage new users.

2. Investigator(s):

Strengths

- Dr. Schloss is exceptionally well qualified to lead this project. His teaching and workshop experience in addition to his research specialty supports his knowledge of the field of microbial bioinformatics as well as the pedagogical knowledge to develop interactive materials that will be effective for the broader bioinformatics community. His knowledge of the issues of data reproducibility in the field is well documented. He is well funded and contributes actively to

disseminating techniques and analysis platforms. He teaches microbial informatics and has already developed useful tools. Schloss participated in the Evolution Synthesis Reproducible Research Hackathon.

- His group developed and disseminates a software package for analyzing microbiome data that reaches thousands through workshops and newsletters and wikisite.
- Plans to incorporate a teaching and Research postdoc.

Weaknesses

- None noted except that he may want to outline the postdoc professional development plan and how this work will advance the career of selected postdoc.

3. Innovation:

Strengths

- Autotutorials on documenting data analysis workflow annotations, computational lab notebooks that can be shared and collaborative are particularly needed and will be useful to an even broader informatics community.
- Design of modules incorporates pedagogically sound models using 5E's. Clearly understands the engage and explore aspects of learning and extend them to published works and reflections on practice.
- Acknowledges conceptually difficult content and potential challenges for participants with lesser programming expertise.
- Clearly identifies how misconceptions will be addressed.
- Clearly incorporates assessments for each learning outcome.

Weaknesses

- No real indication of support for videos and case study development.

4. Approach:

Strengths

- Sound modules design
- Really strong outlines of each modules learning outcomes, exploration activities
- Clear learning outcomes
- Interesting assignments
- Clear assessments and use of self-evaluation and evaluation of research group practices
- Acknowledges issues with long term tracking and behavioral change
- Acknowledges user unfamiliarity with command line tools, etc.
- Strong dissemination plan and reasonable evaluation plan that acknowledges long term tracking of behavioral change issues

Weaknesses

- No indication of how case studies and autotutorials will be developed? Who will do cases, script, etc. Will this primarily be the half time postdoc?
- Requires accounts on a number of different platforms GitHub, FigShare, DataDryad etc.

- No idea that will monitor people's reflections and assignments?
- Where will the less prepared get the programming knowledge and skill needed for module 3. He says he'll send them off, but where and to what level of skills?
- Address human subjects and perhaps develop a stronger plan for evaluation.

5. Environment:

Strengths

- Letters of support from many people who will test materials with research groups.
- Great letters of support from the software community.
- GitHub and riffmonas site already set up.
- Strong support from UM.

Weaknesses

- Evaluation support says fee based.

Protections for Human Subjects:

No concerns. Human Subjects exemption 2 is considered and is justified.

Vertebrate Animals:

Not Applicable

Biohazards:

Not Applicable

Recruitment & Retention Plan to Enhance Diversity:

Acceptable

- Workshops on murthur have strong female and UR participation which bodes well.

Training in the Responsible Conduct of Research:

Not Applicable

- Or this is all RCR...perhaps RFA should have asked for plan.

Resource Sharing Plans:

Acceptable

- Does need to acknowledge how this will be linked to NIH site, too.

Budget and Period of Support:

Recommended budget modifications or possible overlap identified:

- I would suggest getting help with evaluation and monitoring of participant reflections.

CRITIQUE 3:

Significance: 2
Investigator(s): 1
Innovation: 2
Approach: 1
Environment: 2

Overall Impact: This is a superb application written by one of the leaders in the field who really understand reproducibility. It addresses questions that are really important in this field and the material would be relevant beyond the microbiome field. This is not material covered in most courses. Integration into the BITTS and COS universe is great. Berkeley and BITTS are superb platforms. One noted weaknesses is that no detailed curriculum plan offered and it is unclear what would be format of delivery (slides, video, course?). The PI has experience with online courses, which somewhat balances this concern.

THE FOLLOWING SECTIONS WERE PREPARED BY THE SCIENTIFIC REVIEW OFFICER TO SUMMARIZE THE OUTCOME OF DISCUSSIONS OF THE REVIEW COMMITTEE, OR REVIEWER'S WRITTEN CRITIQUES, ON THE FOLLOWING ISSUES:

PROTECTION OF HUMAN SUBJECTS (Resume): ACCEPTABLE Human Subjects exemption 2 is considered and is justified. No concerns about Human Subjects safety and protection.

COMMITTEE BUDGET RECOMMENDATIONS: The budget was recommended as requested.

NIH has modified its policy regarding the receipt of resubmissions (amended applications). See Guide Notice NOT-OD-14-074 at <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-074.html>. The impact/priority score is calculated after discussion of an application by averaging the overall scores (1-9) given by all voting reviewers on the committee and multiplying by 10. The criterion scores are submitted prior to the meeting by the individual reviewers assigned to an application, and are not discussed specifically at the review meeting or calculated into the overall impact score. Some applications also receive a percentile ranking. For details on the review process, see http://grants.nih.gov/grants/peer_review_process.htm#scoring.

MEETING ROSTER

Center for Scientific Review Special Emphasis Panel
CENTER FOR SCIENTIFIC REVIEW
RFA-GM-15-006: Training Modules to Enhance Data Reproducibility (R25)
ZRG1 CB-P (50) R
April 08, 2015

CHAIRPERSON

BOWERMAN, BRUCE A, PHD
PROFESSOR
INSTITUTE OF MOLECULAR BIOLOGY
UNIVERSITY OF OREGON
EUGENE, OR 97403

MEMBERS

BAGGERLY, KEITH A, PHD
PROFESSOR
DEPARTMENT OF BIOINFORMATICS AND
COMPUTATIONAL BIOLOGY
DIVISION OF QUANTITATIVE SCIENCES
UNIVERSITY OF TEXAS MD ANDERSON CANCER CENTER
HOUSTON, TX 77230

BLANCATO, JAN K, PHD
PROFESSOR
DEPARTMENT OF ONCOLOGY
THE LOMBARDI COMPREHENSIVE CANCER CENTER
GEORGETOWN UNIVERSITY MEDICAL CENTER
WASHINGTON, DC 20007

GOODMAN, STEVEN N, MD, PHD
ASSOCIATE DEAN FOR CLINICAL AND TRANSLATIONAL
RESEARCH
PROFESSOR OF MEDICINE AND EPIDEMIOLOGY
DEPARTMENT OF MEDICINE,
AND HEALTH RESEARCH AND POLICY
STANFORD UNIVERSITY SCHOOL OF MEDICINE
PALO ALTO, CA 94304

JOHNSON, ERIC J, PHD
NORMAN EIG PROFESSOR OF BUSINESS
COLUMBIA BUSINESS SCHOOL
COLUMBIA UNIVERSITY
NEW YORK, NY 10027

MARSTELLER, PATRICIA A, PHD
PROFESSOR OF PRACTICE IN BIOLOGY
DIRECTOR, HUGHES SCIENCE INITIATIVES
EMORY UNIVERSITY
ATLANTA, GA 30322

NAKAMOTO, ROBERT K, PHD
PROFESSOR
DEPARTMENT OF MOLECULAR PHYSIOLOGY
AND BIOLOGICAL PHYSICS
UNIVERSITY OF VIRGINIA
CHARLOTTESVILLE, VA 229080886

PRIBBENOW, CHRISTINE MAIDL, PHD
ASSOCIATE SCIENTIST
WISCONSIN CENTER FOR EDUCATION RESEARCH
UNIVERSITY OF WISCONSIN, MADISON
MADISON, WI 53706

SCIENTIFIC REVIEW OFFICER

SMIRNOVA, ELENA , PHD
SCIENTIFIC REVIEW OFFICER
CENTER FOR SCIENTIFIC REVIEW
NATIONAL INSTITUTES OF HEALTH
BETHESDA, MD 20892

EXTRAMURAL SUPPORT ASSISTANT

MASSAY, DIANNE
LEAD, EXTRAMURAL SUPPORT SPECIALIST AND
COORDINATOR
CELL BIOLOGY IRG
CENTER FOR SCIENTIFIC REVIEW
NATIONAL INSTITUTES OF HEALTH
BETHESDA, MD 20892

Consultants are required to absent themselves from the room during the review of any application if their presence would constitute or appear to constitute a conflict of interest.