FINAL REPORT

ICPSR/Sloan Foundation Challenge Grant for

Replication of Empirical Research:
A Soup-to-Nuts Protocol for Documenting Data Management and Analysis

Richard Ball, Associate Professor of Economics
Norm Medeiros, Associate Librarian

Haverford College

June 27, 2014

This report is a narrative account of the activities we undertook with the support of the ICPSR/Sloan Challenge Grant. A separate financial report accounting for the expenses incurred under the grant is being prepared by the Haverford College Business Office and will be submitted by July 15.

In January of this year, we submitted a Midyear Progress Report that described the activities we had completed by then, as well as others that were to take place later this year. This final report is a revised and extended version of that midyear report. We have added information about events that have taken place since January, and we discuss our current thinking about how to build upon the experience of the past year.

Like the mid-year report, this final report presents the activities we have undertaken in four areas: (i) workshops about our documentation protocol for instructors of statistical methods courses, (ii) presentations at professional conferences and meetings (completed during the grant period, and upcoming), (iii) submitting articles to appropriate publications, (iv) additional outreach. It concludes with some reflections on the past year and discussion of our plans for the future development of Project TIER.

FACULTY WORKSHOPS

Test-run workshop for Philadelphia-area faculty, Haverford, College, June 4, 2013. (As described in the mid-year report.) The agenda and participant list for this workshop are attached to this report as Appendix 1. As described in our grant proposal, this workshop was intended mainly as an opportunity for us to get some experience running this kind of event, in preparation for the two workshops planned for fall 2013 and spring 2014. Participants included thirteen faculty members of the departments of economics, mathematics and statistics, political science and psychology from Bryn Mawr, Haverford and Swarthmore Colleges, as well as two current Haverford College economics majors.

We were fortunate that Scott Long agreed to participate in the workshop as an observer and advisor. In addition to making a number of insightful and constructive interventions during the workshop, Scott met with us (both the evening before and the morning following the workshop) to discuss the workshop and the related work we are doing in detail. His comments and suggestions led directly to several tangible results that we discuss later in this report.

First workshop for nationally recruited participants, Haverford College, October 18-19, 2013. (As described in the mid-year report, with revisions.) The agenda and participant list for this workshop are attached to this report as Appendix 2. As envisaged in our grant proposal, eight participants attended the workshop. Five of them were social science faculty members from liberal arts colleges, two were librarians specializing in information literacy and data management, and one was ICPSR's own Lynette Hoelter.

We were impressed and encouraged by the level of engagement demonstrated by every one of the eight participants. They all clearly understood the purpose of the protocol and saw the value of teaching it to students. Most importantly, several participants indicated they planned to incorporate what they had learned at the workshop into their teaching and research advising at their home institutions. We discuss the extent to which participants followed through with this stated intention in the last section of this report.

Second workshop for nationally recruited participants, Haverford College, March 7-8, 2014. (Briefly mentioned in the mid-year report; more fully described here.) We discovered that the costs of running the first workshop were less than we had anticipated, and therefore decided to increase the number of participants at the second workshop to twelve.[1] Nine of the participants were faculty from various social science departments at liberal arts colleges, one was a political science post-doc at Brown, one was a university librarian, and one was the director of a social science research institute at Mississippi State University. The participant list for this workshop is attached as Appendix 3. The agenda was very similar to the agenda of the first workshop.

As sincerely impressed as we were by the participants at the first workshop, we were even more amazed at how engaged, committed, and insightful the second cohort was. We learned a great deal from their comments and suggestions, and a number of them seemed eager to introduce the TIER protocol to their students. Again, the extent to which this has come to fruition is discussed at the end of this report.

PRESENTATIONS AT PROFESSIONAL CONFERENCES: COMPLETED DURING THE GRANT PERIOD

Presentation at the 2013 Stata Conference, New Orleans, July 18, 2013. (As described in the mid-year report.) Richard Ball gave a talk about Project TIER at this annual meeting for Stata users and developers. Questions and comments from the approximately 90 people in

---

[1] The workshops were less expensive than we anticipated because the participants' requests for reimbursement for travel were much less than we had budgeted for (most obtained funding for travel from their home institutions), and because Haverford College did not charge anything for the on-campus guest rooms the participants stayed in.

attendance indicated that they were well aware of the need for improved transparency and replicability of empirical social science research, and that they viewed our efforts to introduce good practices early in the undergraduate curriculum as an innovative and constructive intervention.

Evening lecture at the ICPSR Summer Program in Quantitative Methods of Social Research, Ann Arbor, August 1, 2013. (As described in the mid-year report.) This lecture was presented jointly by Richard Ball and Norm Medeiros. We were very happy to have the opportunity to talk about our work to this audience and meet people at ICSPR. We are indebted to Bill Jacoby, who invited us to give this talk, and to Scott Long, who put us in touch with Bill.

Presentation at the annual meeting of the Southern Political Science Association, New Orleans, January 10, 2014. (As described in the mid-year report, with new information added.) Bill Jacoby was the Program Committee Chair for this conference. In August, when we met him in Ann Arbor, he suggested we submit a proposal to give a presentation about Project TIER. Richard Ball gave the talk at a session titled "Undergraduate Social Sciences and Research Training." (This session was organized by Lynette Hoelter, and Frederique Laubepin gave a presentation—small world!)

One of the reasons we were eager to give public talks like this one was to recruit people to apply for the workshops we held at Haverford. In this respect, this event produced a very tangible result: one of the audience members at the talk, David Blanding, applied to and ultimately attended the spring workshop.

Presentation at the midwinter meetings of the American Library Association, Philadelphia, January26, 2014. (As described in the mid-year report, with revisions.) This talk was presented jointly by Richard Ball and Norm Medeiros. The audience at this event was interested both in issues related to archiving and curating research data documentation and in the nature of the very productive faculty/librarian interaction upon which Project TIER is built.

Panel at NSF conference on robustness of social science research, Washington, DC, February 20-21, 2014. (As described in the mid-year report, with revisions.) This is another opportunity that arose thanks to Scott Long. Scott suggested we contact Myron Gutmann at the NSF; Myron referred us to Heather Dean, an AAAS Fellow in Science and Technology Policy in the NSF's Directorate for Social, Behavioral and Economic Sciences. At the NSF, Heather was working on issues related to the reliability and replicability of empirical research, and over the course of the summer and fall we had a productive correspondence about her work and ours. Early in the fall she invited us to participate in the February conference, which she was involved in organizing.

This was a two-day conference at which about thirty invited participants explored how professional standards for empirical social science research can be enhanced. Individual sessions were focused on the roles of various key constituencies, such as journal editors, university tenure and promotion committees, and research funding agencies. Richard Ball gave a presentation about Project TIER in a session on education and instruction.

PRESENTATIONS AT PROFESSIONAL CONFERENCES: COMPLETED AFTER THE GRANT PERIOD, AND UPCOMING

Summer Institute on Transparency in Social Science Research, hosted by the Berkeley Initiative on Transparency in Social Science, UC Berkeley, June 2-6, 2014. The Berkeley Initiative for Transparency in the Social Sciences (BITSS) was launched about a year ago by several faculty members from UC Berkeley, under the leadership of Ted Miguel, professor of economics at UC Berkeley. The 2014 BITSS Summer Institute, which in fact fell under the umbrella of ICPSR's summer quantitative methods program, was the first in an envisaged annual series of training programs for social scientists nearing the end of their doctoral studies or in the early stages of their academic careers. This year about 35 participants, from diverse social science disciplines and from several different continents, attended the institute. A team of instructors gave presentations and workshops on various aspects of research transparency and credibility. Norm Medeiros and Richard Ball were invited to give a presentation on Project TIER. (Thanks to George Alter for introducing us to the people at BITSS.)

Roundtable at the Joint Statistical Meetings of the American Statistical Association, Boston, August 5, 2014. Richard Ball will lead a 90-minute roundtable discussion on Project TIER.

Invited workshop at Mississippi State University, Starkville, MI, September 26, 2014. Richard Ball will give a workshop on documentation of data management and analysis aimed at several constituencies at MSU: faculty members and graduate students looking for a method of documenting their empirical research, faculty interested in teaching these methods to their undergraduate or graduate students, and managers of projects for collecting and archiving statistical data who would like to implement a stream-lined and reliable work-flow process. Richard was invited to give this workshop by Dr. Ronald Cossman, Research Director at the MSU Social Science Research Center, who was one of the participants in the spring 2014 TIER workshop at Haverford.

Presentation at the Annual Meeting of the American Economic Association, Boston, January 3-5, 2015 (proposal submitted). This is a very exciting opportunity that grew out of our interactions with Ted Miguel and his colleagues at BITSS. Ted has submitted a proposal to chair

a session at the AEA meetings that would consist of presentations of four papers on various aspects of transparency and credibility in social science research. One of the papers would be by Norm and myself, and would focus on computational reproducibility. The other papers would be by Brian Nosek of the Center for Open Science, Macartan Humphreys of the Columbia University political science department, and Eva Vivalt, a post-doc in economics at NYU. Ted submitted the proposal in April; we expect to find out next month whether this session will be given a place on the program.

SUBMITTING ARTICLES TO APPROPRIATE PUBLICATIONS

College & Research Libraries. (As described in the mid-year report, with revisions.) We plan to submit an article about Project TIER to this journal this summer. We have outlined an article based largely on the presentation we gave at the American Library Association meetings in January. We plan to flesh out that outline based on constructive comments we received there. *College & Research Libraries* is the official, peer-reviewed journal of the Association of College and Research Libraries, which was the sponsor of the session at which we gave our presentation at the ALA meetings.

Journal of Economic Perspectives. (As described in the mid-year report, with revisions.) The *Journal of Economic Perspectives* is a highly visible and respected publication of the American Economic Association. Articles for this journal are written upon invitation from the editors; unsolicited manuscripts are not considered. In our case, to be perfectly honest, we actually invited the editorial board to invite us to write an article about replicability of empirical social science research. After a couple of rounds of communication, and after considering a detailed outline of the paper that we produced for them, the editors have slated our article to appear in a spring or summer 2015 issue, and have requested a first draft by early September 2014.

ADDITIONAL OUTREACH

Development of the Project Tier website. (As described in the mid-year report, with revisions.) We created the Project TIER website (www.haverford.edu/TIER) in the spring of 2013. Since then, we have steadily enhanced and added new information to the site. It now contains a web-based version of our protocol—i.e., clickable step-by-step instructions for documenting an empirical research paper. The site also has links and citations for many articles specifically about Project TIER and more generally about issues of replicability and documentation of social science research. This site is also one of the main portals by which we publicized our two faculty workshops.

Webinar hosted by ICPSR, August 28, 2013. (As described in the mid-year report.) We are grateful to Linda Detterman for making this webinar possible. It served as a way to disseminate information about Project TIER to a wide audience, and in particular to publicize the two grant-supported workshops to be held at Haverford College.

Additional on-line presence of Project TIER. (As described in the mid-year report, with revisions.)

•Project TIER is now included in the teachingwithdata.org database of resources for teaching statistics. (Thanks are again due to ICPSR for making this possible.)

•Project TIER is now listed in the National Science Digital Library, nsdl.org.

•A post about Project TIER will be appearing soon on *The Signal*, a blog on digital curation produced at the US Library of Congress. Chelcie Rowell, a member of the Innovation Working Group of the National Digital Stewardship Alliance has interviewed us and will be writing the piece.

REFLECTIONS ON THE PAST YEAR AND PLANS FOR THE FUTURE

Our work on project TIER over the past year has been incredibly exciting and rewarding. The work we did with the support of the Challenge Grant was our first major effort to share what we have learned about teaching students good practices of research documentation and replicability with faculty and data curators at other institutions, and we frankly did not know how our ideas would be received. But the experience turned out to be very encouraging. The participants at our workshops were highly motivated and seemed to be convinced of the value of what we are promoting. The response was also very positive at the various talks and presentations we gave over the course of the year; we received many enthusiastic comments and constructive suggestions.

It has also been very encouraging to learn that Project TIER is one of a growing number of efforts that are bubbling up in several fields that share the broad goal of enhancing the transparency and credibility of research in the social sciences, notably those of ICPSR, BITSS, and the Center for Open Science. The efforts of all these groups are highly complementary, and many fruitful collaborations and interactions are underway among them. Each has a unique focus, but all are committed to the common goal of improving norms and practices with respect to transparency and credibility in social science research. Like Project TIER, many of these initiatives have been launched just within the past year or two. It is not clear to us why so many

related efforts have popped up independently at about the same time, but it appears that momentum is building that has a real chance of producing substantial changes in social science research training and practice in the coming years.

For all these reasons, we would say that the first year of Project TIER has been a success. But really it has been just a first step toward a larger goal. That larger goal is to make a real and substantial impact on the way students of social science are taught to take care of their data. The talks and workshops we have given are an important step in that direction, but the real indicator of our success will be the extent to which faculty throughout the country heed the call and start training their students in the TIER protocol or some similar method of research documentation. We are extremely happy at the tangible progress we have made in this direction during our first year. Two faculty members who attended our workshops have already adopted the TIER protocol in classes they taught this spring. (Both were classes on econometrics, one at Colgate University and one at Williams College.) A finance professor at Coe College is teaching the protocol to students participating in a summer research program there, and an economics professor at Bates College has indicated that she has firm plans to use it in an econometrics course she will teach this fall. Several of the other workshop participants have indicated a strong interest in using Project TIER methods in their classes in the upcoming year, and we are keeping in touch with them to see if we can help them do so.

In relation to our long-term goal of making an impact on how empirical social science research is taught at a national level, the handful of faculty who have already adopted or who are likely to adopt the TIER protocol for their classes is a small group. But in relation to what we anticipated for the first year of the project, it exceeds our highest expectations. The enthusiasm with which these faculty have embraced the goals of Project TIER makes us optimistic that future outreach will lead to further dissemination of the protocol in classes around the country. We also expect that the example set by the faculty who have already adopted the protocol will increase the credibility and visibility of the project, and that the rate at which the protocol spreads will therefore increase.

We have submitted a letter of inquiry to the Sloan Foundation about the possibility of continued support for the next 12-18 months, and are waiting for a reply. A copy of that letter is attached to this report as Appendix 4. Our goal is to use the next 12-18 months both to continue our outreach efforts and to develop a proposal for 3-5 additional years of support, to submit to the Sloan Foundation or some other grant-making agency. We believe that after this initial year of work, plus another year or so of experience, we will be in a good position to craft an effective multi-year strategy that a major funder would find credible.

Substantively, the activities we hope to undertake in the next round of Project TIER include the following:

• Holding two or three more faculty development workshops on the Haverford campus.

• Continuing to speak at meetings and conferences.

• Maintaining regular communication with all participants in past workshops, to share experiences and build a community of educators and scholars committed to promoting the training of students in data stewardship and documentation. Ultimately we would like to have an annual meeting of TIER educators, along with regular interaction via social media.

• Broadening our outreach to reach faculty in graduate social science programs, as well as undergraduate programs at institutions other than small liberal arts colleges.

• Inaugurating a "TIER Faculty Fellows" program that would recognize instructors who adopt our documentation protocol. Providing stipends to a cohort of fellows each year would provide an incentive for faculty to participate, enhance the credibility of the enterprise in the eyes of their colleagues, and allow us to require some feedback from the fellows—like giving presentations about their experiences teaching the protocol, and mentoring other interested faculty.

• Continuing to refine and develop the technical specifications of the TIER documentation protocol, and creating versions tailored to the needs of particular groups of users (e.g., for users of specific software, or for users from different disciplines). Work is underway on several aspects of this goal, and we expect to post updated versions of the protocol on our website by the end of this summer.

• Integrating our website—which provides the instructions for the protocol, but no interface on which to store the documentation files it describes—with a platform designed for managing data and code while a research project is underway and for archiving documentation after the study has been completed. While we were at the BITSS institute earlier this month, we met with developers from Dataverse and the Center for Open Science to discuss the possibility of integrating with their platforms. The prospects for cooperating with both of those groups appear very promising, and we are following up with them. We have a video-conference with Dataverse scheduled for July, and we have already set up a template on the Open Science Framework (the platform that the Center for Open Science supports) that we plan to use with our introductory statistics students this fall.

• Developing systems of assessing the outcomes and effectiveness of the various activities undertaken for Project TIER. This assessment could take many forms for different aspects of the project: to assess the faculty workshops we hold on the Haverford Campus, we could ask participants to complete evaluations or give us comments in some other form. And as discussed above, another important measure of the success of the workshops, and in fact of Project TIER as a whole, is the extent to which instructors around the country introduce the TIER protocol in their classes. We would also like to go beyond just counting the institutions, the number of classes, and the number of students using the protocol, and try to assess the benefits students derive from learning the importance of documentation and replication of research: how does it affect their experience in the class? in other projects they undertake in their careers as students? in their later professional lives? We are also interested in assessing the degree to which, over the long run, Project TIER contributes to the development of stronger professional norms with respect to transparency of social science research. To this end, it might be possible to find a way to measure changes in norms (simply by using observable changes in typical practice as a proxy); it is less clear to us how we could measure the extent to which Project TIER had contributed to any observed changes in professional norms or practice. In any event, developing meaningful ways to keep track of the outcomes of our various activities will be an important piece of our strategy for identifying the activities and approaches that will be most productive as we continue our work.

The first year of Project TIER was more productive and engaging than we could have imagined. Building on this first year and accomplishing something that will have a significant impact on social science education in the coming years is going to require a lot more work. But we believe we have made a good start and identified some effective strategies to pursue in the future. And the fact that so many allied organizations committed to research transparency and credibility have emerged in the last year or two suggests that this is a propitious time for making real progress.

INAUGURAL TIER WORKSHOP
Haverford College   June 4, 2013

AGENDA

9:15-10:00am   Breakfast available in corridor outside the ITC

10:00-10:30am   Introductory remarks
        What is the TIER project?
        What are our goals for the workshop today?

10:30-11:30am   Demo of the protocol
        Examples of student work
        Principles and standards

11:30-11:45am   Break

11:45-12:30pm   Supporting documentation and metadata

12:30-1:45pm   Lunch

1:45-2:45pm   Curation of data and documentation
        Triceratops
        TIER web page: www.haverford.edu/tier
        Dataverse

2:45-3:00pm Break

3:00-3:45pm Discussion: we would appreciate hearing your comments and ideas

PARTICIPANTS

<u>Bryn Mawr College</u>
Steph Miller, Economics
David Ross, Economics
Richard Stahnke, Economics
Anjali Thapar, Psychology

<u>Haverford College</u>
Richard Ball, Economics
Saleha Jilani, Economics
Jeanna Kenney, Class of 2016
Norm Medeiros, Library
Weiwen Miao, Mathematics and Statistics
Shannon Mudd, Economics
Zach Oberfield, Political Science
Anne Preston, Economics
Josh Studnitzer, Class of 2014

<u>Swarthmore College</u>
Erin Bronchetti, Economics
Garret Christensen, Economics
Lynne Steuerle Schofield, Mathematics and Statistics

VISITING ADVISOR

J. Scott Long
Distinguished Professor and Chancellor's Professor of Sociology and Statistics
Indiana University

**Workshop on Teaching Good
Practices in Documenting Data
Management and Analysis**

18-19 October 2013

Haverford College
Whitehead Campus Center, 205A-B

AGENDA

Friday, October 18

2:00-2:30pm	Welcome/plan for weekend
2:30-4:00pm	Participant introductions
4:00-4:30pm	Break
4:30-6:00pm	Overview of the protocol
7:00pm	Dinner (Café San Pietro)

Saturday, October 19

8:30-9:00am	Breakfast
9:00-10:00am	Deep dive into nuts and bolts of the protocol
10:00-11:00am	Reactions; suggestions for improving and expanding the protocol
11:00-11:15am	Break
11:15-12:15pm	Dataverse
12:15-1:15pm	Lunch
1:15-1:45pm	Student perspective on the protocol
1:45-2:45pm	Discussion on feasibility of implementing the protocol
2:45-3:15pm	Break
3:15-4:00pm	Outreach; community building; staying in touch

List of Participants
October 2013 TIER Workshop at Haverford College

Melanie Brasher, Visiting Assistant Professor of Sociology, Dickinson College

Jay Chen, Assistant Professor of Finance, Coe College

Charlotte Flynn, Doctoral Student, Information Science and Technology, Syracuse University

Mark Foley, Associate Professor of Economics, Davidson College

Lynette Hoelter, Assistant Research Scientist, ICPSR, and Research Affiliate, Population Studies Center,
        University of Michigan Institute for Social Research

Peter Rogers, Information Literacy and Social Sciences Librarian, Colgate University

Jennifer Smith, Visiting Assistant Professor of Government, Wesleyan University

Tara Watson, Associate Professor of Economics, Williams College

Also: Peter von Allmen, Professor of Economics at Skidmore College, expressed very strong interest in the
        workshop. He would like to make documentation of empirical work a standard requirement for the
        senior theses of Skidmore econ majors, starting as early as spring 2014. He has a prior commitment that
        conflicts with the dates of the workshop, but has made plans to come to Haverford a week later to visit
        with us and learn about what we have been doing. It looks like he could be an active member of a
        network on teaching good data stewardship.

03/02/2014

Registered Participants for the
March 2014 TIER Workshop at Haverford College

Kenneth Abrams, Associate Professor of Psychology, Carleton College

Kathy Berenson, Assistant Professor of Psychology, Gettysburg College

David Blanding, Postdoctoral Research Associate, Brown University

Barbara Chapman, Laboratory Coordinator and Lecturer, Department of Psychology, Trinity College

Ronald Cossman, Professor and Research Fellow, Social Science Research Center, Mississippi State University

Monica Das, Associate Professor of Economics, Skidmore College

Sarah Dunphy-Lelii, Associate Professor of Psychology, Bard College

Margaret Maurer-Fazio, Professor of Economics, Bates College

Stephen Miller, Lecturer in Economics, Bryn Mawr College

Larry Milliken, Liaison Librarian for Humanities and Social Sciences, Drexel University

Michael O'Hara, Assistant Professor of Economics, Colgate University

Dan Ryan, Hannam Associate Professor teaching in Sociology and Public Policy, Mills College

HAVERFORD
COLLEGE

Daniel L. Goroff
Vice President and Program Officer
Alfred P. Sloan Foundation
630 Fifth Avenue, Suite 2550
New York, NY, 10111

Richard Ball
Associate Professor
Department of Economics
(610) 896-1437
rball@haverford.edu

June 5, 2014

Dear Dr. Goroff,

I am writing to ask whether the Sloan Foundation would consider supporting a project I am working on with my colleague Norm Medeiros, Associate Librarian at Haverford College (nmedeiro@haverford.edu).

Over the last six or seven years, Norm and I have developed a protocol for documenting empirical research papers in such a way that all the statistical results reported in the paper can be easily and exactly reproduced by an independent researcher. We teach the protocol to students in our introductory statistics classes as well as to economics majors writing empirical senior theses, and they have had high rates of success implementing the protocol to document their research. About a year ago, we began reaching out to social science faculty at other institutions to let them know about the protocol and to encourage them to incorporate similar instruction in responsible and transparent use of data into their curricula. Information about these efforts, which we have named "Project TIER" (Teaching Integrity in Empirical Research) can be found at www.haverford.edu/TIER. From June 1, 2013 through May 31, 2014, Project TIER was funded by a Sloan/ICPSR Challenge Grant. Among other things, the challenge grant supported two faculty development workshops that we hosted at Haverford College during the 2013-14 academic year.
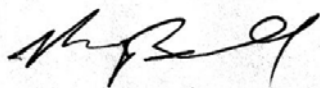
We are now seeking funds to continue Project TIER for another 18 months. During this time we hope to continue the activities we have been working on for the past year, notably the faculty development workshops at Haverford and travel to meetings and conferences. We also hope to expand our efforts in several directions: creating "TIER Faculty Fellowships" to encourage social science instructors to incorporate the TIER documentation protocol in their classes on statistical methods; reaching out to graduate (in addition to undergraduate) social science programs; establishing a presence on social media; and developing a web-based platform to facilitate the process of data management and archiving that the TIER protocol prescribes.

Finally, we hope to use the next 18 months to develop a comprehensive strategy for strengthening and sustaining Project TIER over the long term. Eighteen months from now, with two and a half years of experience under our belt, we expect that we will be well-positioned to prepare a strong proposal for three to five additional years of support, for submission either to Sloan or another major funder.

We expect our proposed budget for the next 18 months would be in the range of $50,000 to $75,000.

Thank you for your consideration.

Sincerely,

Richard Ball
Associate Professor