

## CHAPTER 22 \*

# Teaching Integrity in Empirical Economics:

## The Pedagogy of Reproducible Science in Undergraduate Education

*Norm Medeiros and Richard J. Ball*

### Introduction

Professors and librarians choose careers at liberal arts colleges to forge meaningful relationships with students. A frequent and positive byproduct of this student-centered engagement is collaboration between professors and librarians on the provision of research services. Such collaboration is often a function of the library liaison model, a common organizational structure in college libraries that assigns a librarian to each academic department. Our collaboration began in this modest way, with associate librarian of the college providing guidance on acquiring data and appropriate literature for students in the associate professor of economics' introductory statistics course. Our collaboration has since grown into a curriculum development and outreach initiative that promotes the integration of transparency and reproducibility in the research training of students in the social sciences.

---

\* This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 license, CC BY-NC (<https://creativecommons.org/licenses/by-nc/4.0/>).

## Background

This collaboration, which we now call Project TIER (Teaching Integrity in Empirical Research), grew out of two initiatives that were underway at Haverford College in the early 2000s. Although these activities began independently of one another, combining them in a novel and productive way became the basis for Project TIER.

**The thesis repository.** One of these initiatives was the creation of an online archive of senior theses written by Haverford undergraduates. Haverford is distinctive in that every student writes an independent senior thesis, and in 2002 the college libraries established an online repository to preserve and make public all of these projects.<sup>1</sup> It has become a firmly established institution, and uploading PDFs of the theses of all graduating students is now a standard part of the yearly workflow of the libraries. The librarian has been the primary developer and administrator of the repository since its inception.

**Guidelines for documenting student research.** At the same time, another initiative was developing the economics introductory statistics course. Students in the course are required to write research papers involving original analysis of data obtained from public sources, and the librarian works closely with them to provide assistance finding previous research on their chosen topics as well as reliable sources of data. When this research paper requirement was introduced in 2000, the results were not entirely satisfactory. In most cases, the students' descriptions of the original data they had started the project with, the steps they had taken to process the data to prepare them for analysis, and the analytical procedures they had performed on them were opaque or completely garbled. Follow-up conversations indicated that the problem was not just a matter of poor exposition or writing style but that students' own understanding of the data they had worked with and what they had done with it was often shaky.

To address this problem, we developed a set of guidelines intended to help students better understand the statistical work they do for their projects, and thereby enable them to describe and interpret their research more clearly in their papers. The focus of these guidelines is on implementing the data processing and analysis for a project by writing and iteratively revising editable command files, rather than by using drop-down menus or interactively executing one command at a time. The guidelines also give specifications for a set of electronic documents—including the original data used in the study, the command files they write, and supplementary information that serve as comprehensive replication documentation that an independent investigator could use to reproduce all their reported results. When students turn in their final papers, they are also required to submit the electronic documentation. It took some trial-and-error to develop a workable set of guidelines, but after

several years we had arrived at a formulation that students were able to implement with little difficulty and a high rate of success. And as a result, the clarity of exposition and quality of argument in the papers students write in the introductory statistics class have improved as we had hoped. We have also introduced these research documentation guidelines to our senior thesis students in the economics department. Every year, several of our advisees follow the guidelines while they conduct research and then submit the prescribed replication documentation with their final theses.

**Student documentation meets the thesis repository.** The potential synergy between the online thesis repository and the new guidelines for the documentation of student research might have gone unnoticed had it not been for the fact that the librarian was closely involved in both initiatives. Inspired by Lynch's proposition that librarians need to think of digital objects as a distinct class of resources that present curators with novel opportunities and challenges rather than just as electronic manifestations of printed documents, he recognized that the practice of simply posting PDFs in the thesis repository could be enriched.<sup>2</sup> And the replication documentation we had begun teaching students to prepare for their theses (including data in multiple formats, computer code in several languages, etc.) presented an opportunity for doing so. The data and code for a research paper are certainly not just digitized versions of static print documents; they are of value only if users are able to interact with them creatively—to download, run, edit, and experiment with them. So, beginning in 2010, we put Lynch's idea into action by posting the replication documentation for selected economics theses along with the PDFs in the DSpace repository. In that first year, we posted replication documentation with three of the theses added to the repository, and the number grows each year.<sup>3</sup>

**A tale of two platforms.** DSpace has proved to be an excellent choice of platform for the Haverford College senior thesis repository. Several features, such as its robust metadata capabilities and the ability of search engines to discover materials posted there, make it an ideal platform for archiving completed projects. Through our experience advising students, however, we discovered the need for a more nimble and interactive tool for managing and sharing files throughout the process of conducting a research study. DSpace was designed specifically to support digital archives and lacks several features we deemed necessary for regular student engagement, such as an easy-to-use graphical interface for moving files, a means of authorizing individual and groups of students to have read/write access to their projects without having the same access level to other students' projects, version control, and the means of creating a directory structure to support students' visualization of the protocol. What's more, we sought an open repository that could be used by other instructors as a way of fostering a community around Project TIER.

After experimenting with several platforms, we found that the Open Science Framework (OSF), a tool developed by the Center for Open Science (COS), is ideally suited for our purposes. On the OSF website ([www.osf.io](http://www.osf.io)), users can create accounts instantly and for free, and then are able to create “projects” for particular research topics or studies they are working on. Each project is represented by a page where files can be uploaded and downloaded, and multi-level hierarchies of folders can be created and modified easily to create whatever directory structure is most convenient for the user. The OSF provides greater functionality and structure than common tools like Drop-Box and Google Drive, but students who don’t have a background in programming find OSF much easier to use than platforms like GitHub that are popular among computer and data scientists. We also found OSF attractive because it can be synchronized with Zotero, a bibliographic management application promoted and supported by the Haverford College Libraries.

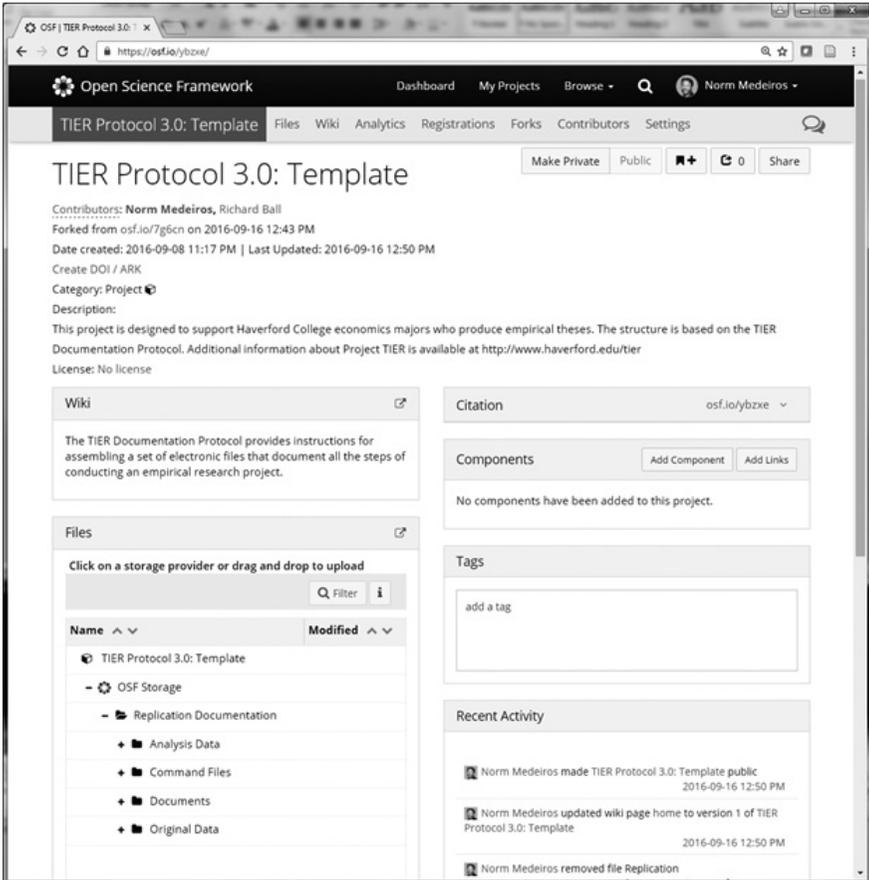
## Reflection

Given all these advantages, we now routinely direct our students to create projects on OSF for the research papers they write in our introductory statistics class as well as for their senior theses. They begin by cloning a template project we have built on OSF, consisting of a set of empty folders organized according to the workflow and documentation guidelines we teach them (see Figure 1). Students gradually populate these folders with the documents they assemble and revise as they conduct their research. The structure provided by this template enhances students’ conceptual understanding of what they are doing with their data, and the fact that everything is stored in the cloud facilitates collaboration among students working on group projects. Moreover, when students come to us for help with their research, being able to download their files to our own computers and explore them in detail dramatically enhances our ability to offer constructive guidance. By the time they finish their papers, some minor cleaning up of the documents they have accumulated on OSF is all students need to do to produce the accompanying replication documentation.

For Project TIER, OSF supports the workflow of students’ research while in progress, whereas DSpace is used to preserve and deliver the finished products of this research at an institutional level. Project TIER is also using Dataverse, a repository developed and supported by Harvard’s Institute for Quantitative Social Science, to provide an environment in which to showcase examples of reproducible student projects across the Project TIER network of institutions. Used together, these platforms support the many aspects of students’ scholarly communication needs throughout the entirety of their research projects.

## Figure 22.1

The TIER Protocol template in OSF



### *Project TIER*

These positive developments led us to begin some efforts at outreach to other faculty and librarians who might be interested in teaching research methods that emphasize transparency and reproducibility, data curation, and the intersection of those two areas. We began by writing a paper about our experiences and received a start-up grant from ICPSR and the Alfred P. Sloan Foundation that allowed us to host two Faculty Development Workshops in 2013–14.<sup>4</sup> It was at that point that (borrowing from the title of our 2012 paper) we began calling this initiative Project TIER.

Thanks to further grant support, notably from the Sloan Foundation, we have been able to continue and expand the work of Project TIER. We continue to offer two Faculty Development Workshops per year, and in 2016 extended our outreach by visiting five graduate programs in several fields of social science to conduct workshops for doctoral students.<sup>5</sup> Through a program of annual TIER Faculty Fellowships, we are collaborating with leaders in research transparency and statistical education. These fellows are creating new curriculum that will be integrated into TIER's outreach efforts, organizing paper sessions at professional conferences, and working with us to evaluate the effectiveness of the pedagogical methods we are developing. More information about these and other activities are available on the Project TIER website ([www.projecttier.org](http://www.projecttier.org)).

Project TIER is one of a surprisingly large number of initiatives for promoting research transparency that have emerged in just the last three to five years. Among the most prominent of these are the Berkeley Initiative for Transparency in the Social Sciences (BITSS) and the Center for Open Science (COS), but there are many others, all of whom interact and collaborate with each other extensively and productively. The niche we see for Project TIER is our focus on education and reaching students early in their professional development. We strongly subscribe to Thomas Carsey's view that "The best way to have an enduring impact on how research is conducted in the future is to affect how researchers are trained in the present."<sup>6</sup>

## Assessment

The leading assessment measure for Project TIER is adoption rate. Our goal is to foster a network of instructors committed to using the protocol in whole or part to promote reproducible and transparent research methods in their classes. Our faculty workshops and other outreach efforts are geared toward maximizing the curricular impact of Project TIER. Our Faculty Fellows program, now in its second year, is a means of more forcibly promoting the mission and principles of our initiative. Fellows take a leadership role in Project TIER's outreach and curriculum development efforts, while at the same time adding their experience and expertise to strengthen and extend the protocol's application.<sup>7</sup>

## Recommendations/Best Practices

Having instructed several dozen faculty and a small number of librarians on teaching reproducible research methods, we can attest that some of the most

successful instances of classroom application have been when an instructor and librarian partner on provision of services. As we've seen firsthand, support for the data management aspects of TIER enables faculty to focus on the pedagogical benefits of teaching transparent research practices. Librarians can bring expertise to the creation of repositories, projects, and metadata that are the building blocks of open science. Our experience suggests that these partnerships can make all the difference when it comes to incorporating robust reproducible research methods into quantitative social science courses.

Open science initiatives at the scholarly and curricular level are increasing in quantity and scope. Subject liaisons and repository managers need to stay abreast of these initiatives and communicate directly with faculty and institutional grant offices to offer meaningful assistance. There's real opportunity in helping faculty colleagues consider open data repositories that comply with journal and funding agency mandates. To the degree faculty believe teaching transparent research practices will advance their students' understanding and application of ethical research conduct, librarians can take a leading role in describing dimensions of replicable research and guiding student and faculty use of tools and repositories that promote open science.

## Conclusion

Our work in teaching computational reproducibility to instructors and students has helped us see the new opportunities in which data librarians should be involved. The traditional view of research data management as an activity that occurs after the completion of a research project neglects the opportunities we see in lifecycle data management; that is, teaching students that carefully managing data, computational code, and metadata should be inherent parts of the research process. Teaching undergraduates these skills provides benefits beyond just good organization; adherence to the protocol enhances students' understanding of their statistical analysis, as they are asked to describe the variables they are using and the exact steps they took to conduct their analysis. We believe taking a broader view of research data management, as we do with Project TIER, will provide numerous opportunities for data librarians to apply their expertise meaningfully.

Along the same lines, librarians need to make efforts to increase face-to-face interactions with the faculty and students they serve. For the past several years, the librarian has held weekly office hours in the economics suite as a means of gaining additional time with faculty and students. By being situated in the physical location of the economics department at Haverford, rather than waiting for faculty and students to come to the library, collaboration with faculty and consultation with students have both increased. This pro-

active approach to providing services is very much in keeping with the role data librarians need to assert in seeking out ways of supporting research data management activities.

## Acknowledgments

We gratefully acknowledge support provided to Project TIER by the Alfred P. Sloan Foundation.

## Notes

1. Subject to students' signing a release. This archive, which is built on the DSpace platform, can be accessed at <http://thesis.haverford.edu>.
2. Clifford Lynch, "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age," *portal: Libraries and the Academy* 3, no. 2 (2003): 327–36.
3. See <http://hdl.handle.net/10066/4899>, <http://hdl.handle.net/10066/6078>, and <http://hdl.handle.net/10066/4820> as early examples of senior theses built on the TIER protocol.
4. Richard Ball and Norm Medeiros, "Teaching Integrity in Empirical Research: A Protocol for Documenting Data Management and Analysis," *The Journal of Economic Education* 43, no. 2 (2012): 182–89.
5. These workshops were conducted for the economics departments of Duke University and Clark University, the University of Pennsylvania Department of Sociology, the University of Colorado School of Education, and the University of Colorado Institute for Behavioral Sciences.
6. Thomas M. Carsey, "Making DA-RT a Reality," *PS: Political Science & Politics* 47, no. 1 (2014): 72–77.
7. Selected comments from students about their experiences learning to conduct reproducible research in the classes of some past TIER Fellows can be found at <http://www.projecttier.org/tier-classroom/student-testimonials/>; a complete list of past TIER events, including conference presentations and paper sessions, is available at <http://www.projecttier.org/events/>.

## Bibliography

- Ball, Richard J., and Norm Medeiros. "Teaching Integrity in Empirical Research: A Protocol for Documenting Data Management and Analysis." *The Journal of Economic Education* 43, no. 2 (2012): 182–89.
- Carsey, Thomas M. "Making DA-RT a Reality." *PS: Political Science & Politics* 47, no. 1 (2014): 72–77.
- Lynch, Clifford. "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age." *portal: Libraries and the Academy* 3, no. 2 (2003): 327–36.