

Authors

Stephanie Wykstra
Research Manager

Workshop on Data-Sharing and Reproducibility



The importance of checking another person's work is easy to grasp, and is a norm in many domains. For example, good financial systems have audits and mathematicians check each others' proofs. Likewise, we should expect researchers to show their work by transparently sharing data and code underlying publications. This is true particularly for scientific research results that influence millions of dollars in spending on programs and policies. Yet studies show that transparency is the exception rather than the norm across many research disciplines ([Alsheikh-Ali et al. 2011](#), [Gherghinaa and Katsanidoua 2013](#), [Vines 2014](#)).

Why is this? In a survey of 1300+ researchers ([Tenopir et al. 2011](#)), a lack of time and funding were among the major reasons for not sharing. In our work on publicly sharing data/code from research studies conducted by [Innovations for Poverty Action \(IPA\)](#) and the [Center for Effective Global Action \(CEGA\)](#), we have found that it's much more time-consuming to publicly share data/code if making the materials usable to others was not considered early on during the project lifecycle. Research analysts and managers who worked with the project during data collection often move on after a couple of years, and if materials aren't

documented well, it can be difficult (or impossible!) to sort out messy files and datasets later on.

Organizing materials throughout a project was our motivation for our [research transparency workshop](#) co-organized by the Berkeley Initiative for Transparency in the Social Sciences (BITSS) and IPA. BITSS was established in 2012 to strengthen the quality of social science research and evidence used for policy-making. The initiative aims to enhance the practices of economists, psychologists, political scientists and other social scientists in ways that promote research transparency, reproducibility, and openness. IPA and CEGA have [research transparency initiatives](#) which support researchers in sharing their data but this is the first time that the two organizations came together to hold a workshop for researchers and research staff working in developing countries. All together we had about 35 participants at our 2-day workshop, including both researchers working at African universities and institutions as well as IPA's research managers from over ten countries within East Africa and beyond.

During the workshop (agenda [here](#)) we covered topics on how to make research reproducible. The sessions included an [overview](#) of why research is unreliable and how to make it more reliable, best practices for managing data and code, writing a pre-analysis plan, writing dynamic documents with [Markdoc](#) in Stata, and using Git/Github for keeping track of versions of code. We also had sessions on Open Science Framework ([OSF](#)), a tool created by the Center for Open Science for collaborating with other researchers. We hosted a guest speaker, Dr. Paulin Basinga, who works with the Gates Foundation as well as the Ministry of Health in Rwanda and discussed ([video here](#)) the importance of replications for providing a strong evidence base for policy. Finally, we dedicated several hours to hands-on time in which participants could work on improving their own data/code with instructors present to assist. All materials are available on our workshop [page on OSF](#).

In order to move transparency forward, we need more than requirements from journals and funders. We also need to develop and deliver training on transparency and reproducibility methods, and need to ensure adequate funding is available to make data/code usable from early on within a project. There are many topics to cover: while we covered important areas in our two-day workshop, we see a need for regular in-depth trainings on reproducible and sharing usable data (while maintaining participant confidentiality!) Coming up this summer, BITSS is offering a 3-day research transparency [Summer Institute](#) in Berkeley for grad students and researchers (deadline for applications March 31!) Groups such as [Software Carpentry](#), [Data Carpentry](#) and the [Center for Open Science](#) also offer resources and training in reproducible research, and researchers at Johns Hopkins also offer an [online course](#) through Coursera. Training in reproducible research is far from a complete recipe for changing the culture, but we believe that it is an important part of the change to come. We hope to look back in ten years and marvel at the fact that most researchers used to merely state their conclusions, without showing their work.

March 21, 2016