

Training for Open Science in Kenya

October 31, 2016



The benefits of transparency in scientific research – such as study pre-registration, development of pre-analysis plans, and publication of underlying data and code — are clear. For example, a benefit of sharing the data and code underlying published studies is that it enables others to check the results, and also can be very useful for carrying out further research. Yet in spite of the benefits, many fields have a long way to go when it comes to data-sharing.

Why is this? Researchers [cite](#) lack of time and funding as major barriers. In a culture where sharing isn't yet professionally rewarded by tenure review committees, making time to publicly share data in addition to the standard journal publication process can be a costly commitment. There are a few ways this incentive problem is being addressed. Through initiatives like the Transparency and Openness Promotion ([TOP](#)) Guidelines, journals themselves are moving toward requesting, or even requiring, submission of the underlying data. Another way to address these barriers is to provide researchers training and resources that can make it easier for them to share data, and also benefit from their contribution to the “public good.”

Research Transparency Workshop in Kenya

The [Center for Effective Global Action](#) (CEGA) and [Innovations for Poverty Action](#) (IPA) are organizations which work with academic researchers to carry out high-quality research studies on programs ranging from education to financial inclusion to health, primarily in developing countries. At both organizations, staff who have worked on sharing data from studies have found that it is much more difficult to prepare data and code after the analysis and publication are complete. Files easily can become messy and disorganized; unlabeled variables can be difficult to interpret later on (especially for those who didn't create them!); lack of documentation about which statistical code produces tables in a publication can make it difficult or impossible to replicate the study. The solution is to think early and often about how to prepare materials so that others can understand and use them (where “others” also

includes oneself in 6 months).

This past year, IPA and the Berkeley Initiative for Transparency in the Social Sciences ([BITSS](#)), CEGA's research transparency initiative, teamed up to hold a 2-day [research transparency workshop](#) outside of Nairobi, Kenya. BITSS was established in 2012 to strengthen the quality of social science research and evidence used for policy-making. The initiative offers resources and support to psychologists, economists, and political and other social scientists in promotion of research transparency, reproducibility, and openness. This was the first time the two organizations' [research transparency initiatives](#) came together to co-organize a workshop.

Workshop participants included researchers from African institutions and universities such as the University of Rwanda and the Ethiopian Economics Association, as well as research staff from IPA offices around the world. The workshop provided an [overview](#) of research transparency, hands-on sessions on best practices for managing code and data, advice on learning to use git (a version-control software) for version control of code, and a tutorial on using [Markdoc](#), a tool for writing dynamic documents. The workshop also included a demo of the Open Science Framework ([OSF](#)), a collaborative workflow platform created by the [Center for Open Science](#) (COS), and gave participants time to work on improving their own data and code. Finally, Paulin Basinga, who works with the Gates Foundation as well as the Ministry of Health in Rwanda, discussed the importance of replication — to verify results and check robustness of those results — for providing a strong evidence base for policy ([video here](#)). Full materials from the workshop are available in a public repository [page on OSF](#).

The Wider Research Transparency Movement

Psychologists have been leading the way on several initiatives within the research transparency movement: for example, Center for Open Science (COS) led the [Reproducibility Project](#), a collaborative project in which hundreds of researchers attempted to replicate studies from psychology journals in their own labs. COS is also providing leadership on initiatives aimed at tackling the incentive problem mentioned above. Their initiative for [Open Badges](#) reward researchers for providing open data, open materials, and preregistration.

Next Steps and Further Resources

BITSS offers regular workshops on research transparency and hosts an annual summer institute (see [2016 agenda and materials here](#)) and annual meeting ([see 2016 registration here](#)). [Software Carpentry](#), [Data Carpentry](#), and COS also offer workshops covering reproducible research. Johns Hopkins University offers an [online course](#) in reproducible research through Coursera.

For individual researchers, it's worth considering how transparency and reproducibility in your own research may affect your workflow and the tools you use as demand for more open social science grows. Here are some questions we discussed during the data sharing workshop:

- As a study progresses, are you keeping track of versions of your code used to clean and construct variables and to analyze data, ideally using software such as git?
- Are you leaving comments in your code and/or naming files to make it clear which parts of the

code produce tables in your paper?

- Are you labeling variables clearly so that you can understand them later and others can reuse them when the data is publicly shared?
- Have you considered what de-identification of the data may be required to share it publicly? How might these efforts affect replication of your analysis if you can only provide access to a de-identified public-use data file?
- Have you considered storing your materials in an established repository such as Dataverse or OSF, rather than on your own website, to make them more widely accessible? (If materials are archived in a repository rather than a researcher's website, they will be stored sustainably and will receive a unique digital object identifier so that others can cite the data and other materials if they use them.)

For the research transparency movement to succeed, there must be significant changes in norms and practices surrounding transparency in research. From funders requiring and offering support for data sharing to journals adopting new data-sharing policies and researchers changing their workflow to make reproducibility a priority, there are many aspects of the transparency movement. While much work remains, the good news is that the shift is well underway.

References and Further Reading

Alsheikh-Ali, A. A., Qureshi, W., Al-Mallah, M. H., & Ioannidis, J. P. A. (2011). Public availability of published research data in high-impact journals. *PLoS ONE* 6(9): e24357.

doi:10.1371/journal.pone.0024357

Gherghina, S., & Katsanidou, A. (2013). Data availability in political science journals. *European Political Science*, 12, 333–349. doi:10.1057/eps.2013.8

Kidwell, M. C., Lazarevi?, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., ... Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS Biology*, 14(5): e1002456. doi:10.1371/journal.pbio.1002456

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., ... Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS ONE*, 6(6): e21101.

doi:10.1371/journal.pone.0021101

Van den Eynden, V., Corti, L., Woollard, M., Bishop, L., & Horton, L. (2011). Managing and sharing data: Best practices for researchers. Retrieved from <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>

Vines, T. H., Albert, A. Y. K., Andrew, R. L., Débarre, F., Bock, D. G., Franklin, M. T., ... Rennison, D. J. (2014). The availability of research data declines rapidly with article age. *Current Biology*, 24, 94–97.

doi:10.1016/j.cub.2013.11.014