Improving My Lab, My Science With the Open Science Framework

February 28, 2014

Brian A. Nosek will be featured in two symposia at the [2014 APS Annual Convention](#), May 22–25 in San Francisco, California. He will also be conducting a workshop on "Open Science Framework: Tools for Your Workflow" at the convention.

My lab has a problem. We do research, time goes by, and some research materials and data get lost. I forget why we did the study; we can't find the final version of the materials that we used. Data just disappears.

Gremlins are not stealing it. Machines break; people leave; organizational strategies break down. We presume that we will just remember what, where, and why. Then, we don't. This loss of data wastes resources and makes our work less reproducible.

We should know better. We *do* know better. But the problems persist. Basic principles from psychological science offer at least three reasons why we struggle to preserve our own research products. First, knowing what one should do is not sufficient to ensure that it gets done. Second, behavior is often dominated by immediate needs and not by possible future needs (e.g., "I know what var0001 and var0002 mean, so why waste time writing the meanings down?"). And third, the necessary changes require extra work; we are too busy for things that make our lives harder.

These factors are nontrivial. And I don't think it is just us. Everyone has an anecdote about the loss of research products because of disorganization, overconfidence in memory, or the complexity of managing information in collaborations. How can this problem be solved? A consultation of the psychological literature suggests that behavior is more likely to change if the solutions:

- provide immediate rewards;
- integrate easily with existing behavior; and
- are easy to do.

Open Science Framework

Jeff Spies, then a University of Virginia graduate student and now codirector of the Center for Open Science (COS; www.cos.io), took on this behavior-change challenge, starting with our lab at the University of Virginia. His dissertation project involved creating the [Open Science Framework](#) (OSF) to stop the hemorrhaging of research material and create incentives for preservation and transparency. Now, the OSF is a free, open-source web application backed by the COS, a nonprofit technology start-up founded by Jeff and me. COS is supported by four foundations and staffed by a team of more than 20 scientists and software developers.

The Center for Open Science hosted an open house to introduce its mission of improving transparency and research preservation to the surrounding Charlottesville community.

The OSF helps individuals and research teams organize, archive, document, and share their research materials and data. Users have accounts and create projects. The projects can have components such as study materials, analysis scripts, and data. Each project and component can have a wiki, attached files, or add-on services from other vendors, like a repository at web-based hosting service GitHub. Users can give other users access for collaboration. The OSF logs actions and retains version histories of the wikis and files so that the history of the research process is recoverable.

Our laboratory has integrated the OSF into our daily workflow. Project leaders create an OSF project and add collaborators. Following a checklist, the team posts study materials, submissions to institutional review boards, notes about research goals, and analysis plans during the study-development process. After data collection, the data, codebooks, and analysis scripts are added to the projects. Later, posters, lab presentations, or manuscripts are added as well.

Members of each project team now have a shared space for accessing everything connected to the project. When a graduate student leaves the lab, the materials are not lost. When a computer breaks, the materials are not lost. When I forget why we did the study, I can reconstruct the purpose and see exactly what we did. We are losing less of our work, and we can more easily reproduce what we did and what we found. By integrating the OSF into our daily workflow, our lab is being a better steward of its resources and work products.

OSF Enables Greater Transparency and Reproducibility

The mission of COS is to increase openness, integrity, and reproducibility of scientific research. The OSF is central to that mission as an infrastructure that provides value to researchers for their needs and simultaneously enables greater transparency.

OSF can be used for private collaboration — like my lab does for most active projects — and making those private projects public takes just a couple of clicks. Users can share with certain users or with the public

at large as well as control which parts of the project are public. For example, my laboratory collects some sensitive data. We can make the study materials, codebooks, and some anonymized data from those projects public but leave the identifying information private. Further, we may have projects for which we will be willing to make some parts available immediately but other parts available later. For example, we might share the research protocol but hold the data until we meet our primary research objectives.

Using the OSF also enables us to improve the reproducibility of our research. With our scripts, code, and data made public, other researchers can reproduce our analyses and findings, or [reanalyze our data for their own purposes](). Also, with our materials made public, other researchers can access and use them with greater confidence that the original design is represented accurately. Have you ever tried to recreate materials just by the description in a published methods section? It can be a harrowing effort, fraught with error. Sharing the actual materials we used publicly extends the efficiency gains from my lab to others by reducing the error in attempts to recreate those materials.

Creating Incentives

Transparency and reproducibility are core values of science, but they are not presently part of daily practice. Before we started using the OSF, others had little access to the materials, data, and research processes in my laboratory. Transparency was limited to what we put into the published report. This can be a rather superficial summary of the actual research process. David Donoho, a computational scientist at Stanford University, has called published reports advertising for the science, not the science itself. At the same time, the present culture does not reward researchers for making their research more transparent or reproducible. The key incentive for my career success is publishing. Publishing does not require transparency or reproducibility, so why expend energy on it?

Center for Open Science Supporting the OSF

The OSF is an open-source project maintained by the nonprofit [Center for Open Science](). Jeff Spies and Brian A. Nosek founded COS in 2013 to increase openness, integrity, and reproducibility of scientific research. Now staffed by a team of more than 20 developers and scientists, COS is supported by $10M in grants from the Laura and John Arnold Foundation, John Templeton Foundation, Alfred P. Sloan Foundation, and an anonymous donor. COS established a preservation fund to ensure OSF user content survives even if COS does not.

Believing in transparency, and making it easy to accomplish, will certainly help to increase its frequency. But unless incentives for success are also aligned, it is not likely to change my daily practices. For effective change, we must nudge the incentives so that I feel some payoff for making these practices routine. Some simple nudges are integrated into the OSF itself. For example, a driver of scientific reputation is the influence of one's work on others. For public projects, OSF provides statistics documenting project views and file downloads. Metrics like these complement the slow-growing citation counts of published articles. OSF also has a novel citation type — forks. If other researchers want to use some of my public work, they can fork my public projects into their workspace. Their new project will always link back to my original work. They can revise a measure, reanalyze the data, or extend the work in some other way. The fork count is therefore a functional citation — others are using and extending my

research outputs.

Nudging incentives in the OSF alone are not sufficient to produce cultural shifts in behavior. COS also supports services to journals. Journals play a key role in managing research incentives and behavior. For example, Registered Reports is a publishing format in which the research design is peer reviewed prior to data collection or analysis. The APS journal *Perspectives on Psychological Science* offers a robust version of this format, specifically for replications. For the journal, the OSF provides back-end support for archiving of data and materials and preregistration of the research designs. Moreover, COS maintains a committee to refine, evaluate, and encourage adoption of Registered Reports by other journals. Another nudge is offering [badges for open practices](#). *Psychological Science* has adopted badges for open materials, open data, and preregistration. Authors who meet the specifications can earn the badges, which then appear with the published articles. OSF provides support for authors to meet the badge specifications, and COS maintains a committee of scientists for improving the specifications and encouraging adoption by other journals.

Improving Science From the Bottom Up

The purpose of science is to accumulate knowledge about nature. Openness and reproducibility are important because, in science, the believability of a claim should not depend on the authority or credibility of the person making the claim; rather, believability is contained in the evidence itself. The only way for this system to succeed is if the evidence supporting the claim is available for evaluation and reproducible by others. COS, with its OSF infrastructure, is trying to make it easier to behave according to these principles.

The OSF is providing value for our lab. Perhaps it will provide some for yours as well. Try it out.

Some Features of the OSF

Archiving	Preservation of materials and data
Collaboration	A shared private (or public) workspace for teams to maintain project materials, code, data, and reports
Logging	Project activity automatically tracked; good memory not required
Version control	Histories of files and wikis automatically preserved and recoverable
Integrates private and public workflows	Project owners can easily expose parts of the project (e.g., codebook but not data) to specific others or to the public
Registration	Current state of the project can be frozen to certify its content at a particular point in time (e.g., preregistration for confirmatory analysis)
Templates	Metadata can be attached to any OSF component to facilitate specific registration information, data collection, meta-analysis, and curation
Forking	Creates a copy of a public project in user's workspace with link back to original project (e.g., import another's scale for research use)
Add-ons	Other services can be integrated with OSF projects (e.g., data repositories, analytic and visualization tools)