

Habits and Open Science

March 31, 2016



Richard D. Morey



Candice C. Morey

In March 1665, Henry Oldenburg introduced the first publication of the *Philosophical Transactions of the Royal Society* to the world. In one remarkable paragraph, Oldenburg declared the purpose of the scientific journal — the second in the world, by only 2 months — to be the “clear and true” communication of scientific work to other curious people around the world. The printing press had revolutionized the spread of information and opinion, and the Royal Society was determined to take advantage of the technology to revolutionize science itself. Oldenburg declared openness as a central goal: The curious of all nationalities were invited to use the *Transactions* to “search, try, and find out new things, impart their knowledge to one another, and contribute what they can to the Grand design of improving Natural knowledge ... for the Universal Good of Mankind.”

The concept of publicly communicating diverse scientific findings in one publication using the printed page was a tremendous improvement over the typical practice of scholars writing letters to one another, or speaking about their findings at private or public meetings, or writing books. This concept is, of course, the model for dissemination of scientific findings today, though article formats have been standardized and lengthened.

We are in the midst of a second revolution in communication — one in which digital media have replaced the printed page and communication is electronic, essentially instantaneous, and of practically unlimited bandwidth. However, the similarity of modern scientific communication to 17th-century scientific communication is remarkable: Most scientific publications can still be rendered on the printed page. Material that is not printed in the scientific article is often lost forever. Scientists have taken advantage of technological advances to *do* science but have been slow to take advantage of technological advances to *communicate* science.

Relative to the possibilities afforded by new technology, modern scientific communication is not very open. When an article is published, we can easily share the supporting data, materials, code, and experimental programs with anyone at very little cost — but we often don't. Scientific values are progressive and open, but scientific communication is stuck in the 17th century. What accounts for the lack of progress in scientific openness?

One likely explanation is the inherent conservatism of scientific training: We inherit the practices and habits of those who trained us. Because we are evaluated on the basis of our scientific communication, we stick with what has worked in the past; thus, the scientific-openness deficit is a deficit of habit. If we want the next generation of scientists to be open, we must *train* openness.

Three Ideas for Creating Open Habits

We outline three ideas for training graduate students to be scientifically open that arise from [our own experiences](#) working with data both as students and with students. We selected them for several qualities: First, they should help students build habits that will enable them to make a *choice* about the level of openness they would like in their own research, rather than force them to be slaves to their habits; second, they should be easy to implement by students' advisers, requiring little effort and time; and third, they should benefit the advisers and the labs, even if the research in question is not ultimately opened.

Data Partners

In a scientific culture in which data are not typically shared, the same person or group of people are both producing the data and analyzing them. This makes it difficult for a student to take into account the needs of analysts who may use those same data in the future. Time can turn good data into useless data when the personnel most knowledgeable about the information inevitably move on. How were the data produced? What do the columns and values in the data represent? What sort of preprocessing was done? Students need to understand how to annotate data in such a way that it is useful to other researchers.

To this end, we suggest that advisers pair their students with peers from a related lab in a “data partner”

arrangement. Under this model, a student would, for example, collect data and perform her central analyses, then send her partner a draft of her methods and results sections, together with the data and a short metadata document describing the data. The data partner would be tasked with reproducing the central result using only the information given.

The data-partner scheme has several benefits. First, students must learn good data-curation habits for a concrete reason: Someone else — from a trusted partner lab — will be looking at the data. They also benefit from having someone check their results. Labs and advisers benefit from the scheme because future use of any data depends on good documentation.

The 5-Year Plan

A central part of making science transparent is ensuring the longevity of the data. An astounding amount of data is lost because researchers simply make inadequate plans to preserve them. Students naturally are thinking about the next hurdle in their careers rather than how the data they collect might be accessed in the future.

The “5-year plan” is one way that advisers can encourage their students to think about that future. About halfway through their projects, students should prepare a presentation that describes what they’ve done to ensure that their data will be accessible, intelligible, assessable, and useful in the future (The Royal Society Science Policy Centre, 2012). If the lab does not already have standardized practices in place that support these four goals, then such standards can be developed by discussing with the students what works and what doesn’t (for ideas, see van den Eynden et al., 2011). Other members of the lab can contribute to the discussion, allowing an adviser to emphasize what they see as important to the entire lab and focus the discussion. A clearly articulated 5-year plan helps ensure the longevity of the data for the members of the lab and anyone else who might request them.

The Submission Check

There are sometimes reasonable legal, ethical, or practical reasons for keeping data or materials from being shared. But in current scientific culture, data are closed by default. Generally, researchers are not expected to release their data and materials. In an open scientific culture, transparency is the default and closed research should be deliberate and justifiable.

We suggest that advisers meet with their students before the first submission of a manuscript to discuss whether the data, materials, and code from their projects will be released openly with the manuscript. This meeting — the “submission check” — should be used to ensure that the student is ready to release the materials openly, even if he ultimately decides not to do so. If the materials are ready to be released, then they are in good shape to be archived for later use by the lab. If the student chooses not to release the materials, he should defend that choice. If the student will release them, the discussion can revolve around how best to do so. Will he announce the release on social media, or on his blog? Perhaps he will contact other labs to announce the data and the submitted manuscript; maybe he will wait until the manuscript is accepted. With a data partner and the 5-year plan, the student already will have done most or all of the work and can make informed, deliberate, and free choices about the release of the materials.

Openness as a Choice

Researchers who are not trained to take advantage of modern technology for open science will have difficulty embracing transparency because their scientific communication methods are a matter of habit. To reduce the transparency deficit between scientific values and scientific practices, we must ensure that future scientists have the right habits. Data partners, the 5-year plan, and the submission check collectively offer low-maintenance ways for advisers to ensure their students can contribute to a more transparent future.

Richard D. Morey will be speaking at an APS–SMEP workshop on “BayesFactor and JASP: A Fresh Way to Do Statistics” at the 2016 APS Annual Convention, May 26–29 in Chicago, Illinois.

References

Oldenburg, H. (1665). The introduction. *Philosophical Transactions of the Royal Society*, 1, 1–2. Retrieved from <http://rstl.royalsocietypublishing.org/content/1/1-22/1.short>

The Royal Society Science Policy Centre. (2012). “Science as an open enterprise.” Retrieved from <https://royalsociety.org/~media/policy/projects/sape/2012-06-20-saoe.pdf>

van den Eynden, V., Corti, L., Woollard, M., Bishop, L., & Horton, L. (2011). Managing and sharing data: Best practice for researchers. UK Data Archive, University of Essex, United Kingdom. Retrieved from <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>