Open-ish Connectome

Gregory Kiar June 20th, 2019

As a Ph.D. student in Biomedical Engineering, I have focused my research on the evaluation of existing methods used in computational neuroscience. With the ominously phrased “reproducibility crisis” becoming a more popular phenomenon, my own involvement in a collection of initiatives dedicated towards open science, and software development experience under my belt, I believe I’m well situated to explore just how bad this problem is in our field.

Before I do that, it’s important that I take my own medicine and explore just how well I did. As a master’s student, I set out to develop an open, reliable one-click pipeline that would estimate structural brain maps from MRI scans and use it to create the largest database of these maps (connectomes) to date. Was my tool open? Was it reliable? Was the dataset I created useful? Two years into the project I had (with massive amounts of help and collaboration) created a pipeline that convinced me I succeeded. It was openly available on GitHub. It was robust on all the data we tested. I produced a database of over 100,000 connectomes that is still online. We packaged it, made a demo, wrote a paper, and claimed victory.

I moved on to a new project in a new lab and city for my Ph.D., and a new batch of students was tasked with taking over the project while I preached its adoption. I quickly came to discover that my tool wasn’t nearly as meaningfully open or widely robust as I thought, and the database was far from consumable. I think it’s important to mention the FAIR principles here. This set of guidelines—Findable, Accessible, Interoperable, and Reusable—provide goalposts that researchers can use to make sure they’re not just being open, but *usefully* open. I had critically failed to consider these ideas. The new developers struggled through limited documentation and failed to grab footholds on the beast I’d created. It was open, the code they needed was in front of them, but dissecting it was a formidable challenge. While the tool generated robust maps on the healthy adults I tested, I started to see systemic failures in other populations. Though the datasets I processed may have been online and accessible, the graphs were stored in a format that is now deprecated in two separate - and predictable - ways.

While I accomplished my short term scientific goals, I ultimately failed to make a resource that was effectively open, robust, and useful. This isn’t to say that these are impossible goals or that if I didn’t meet them nobody could, but that usefulness isn’t something we can get just by sharing our code and data. Accomplishing effective openness isn’t easy, but it’s also largely not taught. Feeling the need of my former self, I’ve thrown myself into teaching these skills over the last several years. I instruct with Software Carpentry, led the development of a “Training Track” for neuroimagers at hackathons, regularly hold “Brainhack 101” workshops, and have joined the leadership of training bodies within organizations such as the Canadian Open Neuroscience Platform, Healthy Brains for Health Lives, and the Organization for Human Brain Mapping, where I try to lower the barrier for trainees to gain skills in reproducible and open science.

While I consider my first attempt at creating a truly open resource that enabled others to stand upon my shoulders a failure, it is a big part of why I’m passionate about open science and education. In order for “open science” to become “science” we need to stop pretending that it’s easy or automatic, and start teaching researchers at all levels the skills needed to accomplish it.