To:
The Office of Science & Technology Policy

From:
Erich S. Huang, MD, PhD
Director, Cancer Research
Sage Bionetworks

Re:
OSTP RFI: Public Access to Digital Data Resulting from Federally Funded Scientific Research

Transparency in biomedical research isn't just good science, it's good medicine.

Why? Disease's complexity requires many eyes coming from many angles. This is clear on the front line where the practice of medicine can be a humbling process. I have had cases where a combination of knowledge, training, and luck have vanquished suffering, and also cases where the complexity of affliction has confounded me at every turn. This is equally true in our investigation of disease. No one has a monopoly on insight:

> *"Every intellectual has a very special responsibility. He has the privilege and opportunity of studying. In return, he owes it to his fellow men (or 'to society') to represent the results of his study as simply, clearly and modestly as he can."*
>
> –Karl Popper

Karl Popper's above quote is fitting. While you maintain confidence and decisiveness to effectively treat your patients, you are ever conscious of the repercussions of hubris when your patients' lives and well-being are at stake. A good physician-scientist is ultimately an intellectually "modest" doctor—you know that you don't always have all the answers.

With this in mind, the Office of Science and Technology Policy's Request for Information on Public Access to Digital Data Resulting From Federally Funded Scientific Research is particularly timely. The cloudburst of complex "big" biomedical data precipitated by genomic technologies is so promising, yet so difficult to contend with, that only an open exchange of data and ideas can unravel its complexity and maximize its benefit to our patients. This free market is threatened today by legislation such as HR 3699 that seeks to shut the door on free access to NIH-funded publications. Publicly-funded in the public interest? Throw the doors open.

I have seen the complexities of such data humble an institution. At Duke, where I received my medical and scientific training, a mixture of hubris and opacity in data and methods resulted in tenuous science. Institutionally, intentions were good—who wouldn't want to figure out how to use genomics to match the best therapy to a patient's disease? But the status quo (and this is *everywhere*) for managing Federally-funded data made open review by the scientific community difficult, if not impossible. While there are data-sharing plans mandated in the Federal grants process, they are unenforced.

Academic research has a cultural problem. There is no doubt that taxpayer funding of basic and translational research catalyzes progress in treating human disease and the basic discoveries that fuel advances, but there is also no doubt that funded individuals are conflicted.

As with any human endeavor, progress and discovery in the life sciences is driven by a complex mixture of desire to serve society and self-interest. And here we scientists must remind ourselves how we got to our positions: Taxpayer money. My doctoral training was Federally-funded. My stipend as a surgical resident was Federally-supported. My first faculty grant was Federal. It is easy to say "my grant", but really, it's the *public's* grant. My faculty development award's real purpose was to put me in the position to contribute meaningfully to the NIH's mission of mitigating disease. A *public* mission. Yet it is easy to see that many faculty see Federal grants as the mandatory offsets for covering their salaries, their research groups and, ultimately, their tenure. And surely university administrators often see things the same way.

So incentives are all in the wrong place: the currency of academic promotion is research grants and publications. If you're an Assistant Professor and are fortunate enough to obtain a 5-year Federal research grant that funds valuable and interesting cancer research, 4 years into "your" grant—with your eye cocked on competitive renewal (because you probably need two or three of these to obtain tenure), and publishing your findings in a prestigious scientific journal—why would you share these data? Why wouldn't you maximize your competitive advantage over all the other Assistant, Associate, and Full Professors out there?

We need to recalibrate how we do science. There's plenty of room for individuals to demonstrate their great ideas in a world where publicly-funded science is openly accessible. You can see a vibrant exchange of ideas in the open source software world on Github (www.github.com). This web service allows programmers to store their software projects in a frictionless open system that allows others to see their code, make copies of code, suggest improvements, or even write improvements themselves. For Facebook, Twitter and LinkedIn programmers, a job candidate's Github repository is considered far more valuable a record of their mettle than their resume. Even in this open, free-flowing milieu, programmers do pretty well for themselves; and their value can be measured by their contributions.

In science, the simple fact is that where the funding goes, scientists go too. Therefore an easy first step would be a stipulation that publicly-funded research be placed in an open-access repository and that standards for data quality, usability and accessibility be vetted by a neutral party. No open access? No renewal. Besides idealism, there are pragmatic reasons for doing this. See the Harvard Provost's response to the related RFI on open access publications for a cogent analysis of the economic benefits of open science. Another consideration is that public investments in research should provide maximal return on investment. We paid for it, right? Why not let the community (hackers, biotechs, colleagues, even Big Pharma) figure out how else to unlock a dataset's full value rather than leaving it to be skimmed by an individual marquee principal investigator? Really, we should see these data as *vital infrastructure*, like Eisenhower's Interstate Highway System and treat them as such—accessible to all.[1]

We owe this to our patients.

In the end, the scientific enterprise is about sharing and communicating "simply, clearly, and modestly". Why? Because none of us knows all the answers.

[1] And the economic benefits of our interstate system has far exceeded Eisenhower's investment.