Tue 12/20/2011 10:58 AM

Request for Information: Public Access to Digital Data Resulting From Federally Funded Scientific Research

To whom it may concern

My name is Andrew Vickers, PhD, and I am on the faculty of the Department of Epidemiology and Biostatistics at Memorial Sloan-Kettering Cancer Center. I have a long interest in data sharing in medical research. My scholarly papers include: the rationale for data sharing (Trials. 2006 May 16;7:15: http://www.trialsjournal.com/content/7/1/15); an empirical study of data sharing (PLoS One. 2009 Sep 18;4(9):e7078: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0007078); guidelines on how to share medical data (BMJ. 2010 Jan 28;340:c181: http://www.bmj.com/content/340/bmj.c181?view=long&pmid=20110312); an extensive set of raw data and code from a series of studies on radical prostatectomy (BMC Res Notes. 2010 Sep 2;3:234; http://www.biomedcentral.com/1756-0500/3/234) and an editorial on the policies of funding agencies (BMJ. 2011 May 4;342:d2323: http://www.bmj.com/content/342/bmj.d2323?view=long&pmid=21543405). I was also the author of a widely cited *New York Times* op-ed on data sharing (http://www.nytimes.com/2008/01/22/health/views/22essa.html).

I would like to address the first question about the specific Federal policies that would encourage public access to data arising from Federally-funded research. My own view is that this has to be a policy with teeth. We don't want to simply repeat the history of PubMed Central: this was originally voluntary, but compliance was dismal and so was then made mandatory. All researchers now know that, if they want to get more grants from the National Institutes of Health, they have to deposit papers to PubMed Central and prove that they have done so when they submit their next grant. Why not have something similar for raw data? I would propose that, if you want another grant, you have to prove that the raw data produced from your last grant is publicly available on a registry. Naturally, researchers could ask for waivers, just in the way that you can exclude women or children from clinical trials if there is a good reason to do so. For example, if a researcher can make a clear case that sharing data would pose a genuine threat to privacy (for example, genetic studies on unusual populations) , then that research could be exempt from the data sharing requirement. Moreover, researchers could request a reasonable period of time to exploit their data. For example, researchers could state that their raw data were deposited at a particular registry, but that the data will not be accessible for two years after the first publication describing the study results. Note that the proposal is not for a vague "data sharing plan" (which, could after all, be "we will evaluate your request and then refuse it"), but for mandatory depositing of raw data into a publicly accessible archive.

It is clear that there would be some technical obstacles to such a proposal. For example, how would registries to accept raw data be organized and who would run them? Just how "raw" should raw data be. However, it is clear and obvious that such obstacles are far from insurmountable and that they could be solved by methodical planning.

As regards question 9, attribution and credit, please note that this an issue that I have dealt with in the peer-reviewed literature (Trials. 2006 May 16;7:15). In that paper, I proposed some guidelines for conduct of investigators using raw data collected by another team and for journals publishing such data. In the following "independent investigator" is the individual wishing to publish a reanalysis of published raw data; a "trialist" is the individual who helped gather the data in the first place.

Code of conduct for independent investigators and journals

1. Independent investigators planning to *publish* a new analysis should contact the trialists before undertaking any analyses
2. One or more trialists should be offered a co-authorship on any resulting papers

3. If trialists disagree with the methods or conclusions of a new analysis:

a. They should not have veto power, unless this was agreed beforehand by the independent investigators

b. They should, however, be guaranteed the opportunity to write a commentary to be published alongside the new analysis

4. Journals should not publish new analyses of previously published data unless either a trialist is an author or a separate commentary from a trialist is attached

5. Published new analyses should cite the original trial

I would be delighted to respond to any questions or comments on these thoughts

Andrew Vickers

Memorial Sloan-Kettering Cancer Center