The Endurance of University Data Records - Be Discouraged

Posted by Heading Out on March 5, 2010 - 10:21am
Topic: Policy/Politics
Tags: data destruction, data records [list all tags]

Much has been made of the destruction or loss of data from the files of the Climate Research Unit (CRU) of the University of East Anglia. Dr. Phil Jones, the Director until this all became public, did not do himself much benefit with his remarks before the Commons Select Committee in the UK that is looking into the Climategate matter. Destroying the raw data, or not making it available, so that all that one can use is the modified and gridded data, means that there are no checks that the adjusted data has been properly derived. But the destruction of research data is not only encouraged in some institutions, it is mandated by regulation. It is, however, a point that a lot of folk may have missed. So I thought I would mention this since I suspect that it affects much more than just the data at my own university.

I retired from the University last Friday, and have spent today throwing away about 80% of the material in one of the three offices that, transiently, it has been stored in. (It happened that in recent months the three folk who had worked with me on many of my research programs also retired, and so their records were boxed and collectively stored with mine until they could be sorted.) We worked through file after file, with data going back to the first experiments that I had run some 40-years ago when I came to this place as a very junior Assistant Professor. That data was still on graph paper, with hand-plotted curves. It went into the trash barrels. As did many of the journals that I had paid large chunks of money for over the years, and almost all of the correspondence dealing with the millions of dollars of contracts that I have managed during my term here.

I am working with the University Archivists. They and a couple of students helped me work through many of the files, and did most of the actual disposal. We have done some pretty interesting things over the years (I was incredibly fortunate to be involved in many of the activities that changed my discipline from an academic curiosity into something that impacts, in one way or another, many people's lives every day.) But not much of that is being kept.

A treatment for skin cancer that discriminates between healthy and diseased tissue – save the patent – the rest into the trash. Cleaning the Statue of Freedom atop the Capitol building in Washington, save the proposal, the final report and one paper. The rest – into the trash.

You might think that I am being deliberately destructive, but this is what the regulations require. For those as incredulous as I was, until last week, here is the information:

Grant Records			
‡ Pending/Rejected Grant Files	06/07 & Prior	Destroy	!
‡ Awarded Grant Files: Fiscal records including effort verification reports paid by grant funds		Send all fiscal material related to the grant to Office of Sponsored Programs Admin.	!
‡ Grant Project Director's Files and/or Administering Department Files: Awarded, completed grant files	06/07 & Prior	Contact campus grant's office for approval to destroy	!
‡ Grant History File: Grant proposal, report, sample of data	06/07 & Prior	Place copy of grant proposal, final report and sample of data in Campus Archives	!
Research Records			
Approved Animal Protocols - Expired	2006 & Prior	Destroy	!
Care of vertebrate animals as subjects in research and teaching records	2006 & Prior	Transfer to Campus Archives.	!
Research Data		Retain for seven (7) years after final report or project has been completed, then destroy.	!
Research Final Reports	2004 & Prior	Transfer to Campus Archives.	!

Notice that the applicable dates are for three years back. In other words three years after I get a contract or grant I am supposed to archive the proposal, report and the sample of data (the paper that I mentioned earlier), and then four years later, I am supposed to destroy all the research data. Hope the sanctions are not too onerous, since, until today I had kept everything. Now much of it lies in grey trash bags stacked down a hallway.

I actually got a bit annoyed about this last week, and it was then that this all came to my attention. As it happened when my pension was calculated (ours is based on length of service) the record did not show that I worked for part of 1997. Now I knew that I had, but if I (or actually the Center staff) had followed University rules, rather than what I had wanted, then there would have been no other record against which to compare the facts relative to the records that are in the Central Personnel Office. Given, however, that we had kept the records, the copy was found, sent up and the matter was straightened out within about an hour. (However, since a large number of boxes have recently left for the incinerator I don't believe that my replacement as Director has continued my cautionary practice).

In the past the Center has been audited, and had, on another occasion to defend a set of experiments that were investigated by a government agency. In the first case I found a record from a period that I suppose I should have had destroyed that showed that the audit inquiry was misinformed, and in the other I was able to supply all the documentation required (foregoing that it took several full weeks of several individuals time to copy – this being before much of our information was stored digitally). As a result, and based on that information, the inquiry was discontinued.

The amount of space that is needed to store digital records is trivial against the bookcases of material that have just gone into the trash. But storing the material only in digital form has some risks. I have just finished a comprehensive review of one of our programs, requiring data that was stored digitally back in about 1987. I cannot open be files for any of the information. I can't find readers that will read some of the disc storage that I recorded it on. (And where I made copies of the files and transferred them, the current versions of the software won't read files from that far

back.) It wasn't in this case too much of a problem since, in violation of policy, I had the paper copies and just scanned them in to get what I needed (and created a digital copy), but those paper copies will be in those grey bags next week.

There are problems with data storage. If I kept the written records, then when I vacate the room, the books will go onto a bookcase in the hall for the students who want them to help themselves (my colleagues already have), a small amount of material will go home, some will go to the Archives, but the majority will burn, or be landfilled. Because the person who follows me into that space has their own research and documentation, which they will put into the bookcases that I am vacating. There is not enough room to store the material. We used to use microfiche to do that – I haven't seen anyone use one of those readers in years, I stopped when ours broke.

The Federal Government and the National Labs are no different. I was on a National Panel which needed some information on a project from one of the National Labs and we wrote for it. It was about 15-years after the experiments. They no longer held the data, and there was no-one there who we could talk to about the work. (Which was one of those supposedly crazy ideas that folk go out and try, and bless my socks, this one worked, and might have been helpful if we could have found out more.)

So while I continue to think that it is madness to be spending the amount that we are on research into the possible problems of the greenhouse gases without a more robust set of raw data that everyone agrees has integrity and that has been compiled in a way that is logical and transparent, I have to point out that the protocols governing records at Universities are not supportive of my position. Not that this makes me feel any better; rather the reverse. And there are many, many research programs that do not have that level of visibility.

I used to joke in my class that disasters happen in about 20-year cycles, because nobody read anything that was older than that, and thus missed some less-than obvious design features, which became forgotten until their lack led to disaster. But I had not realized that the data was all gone. And in the digital age, if it isn't on the web who knows where to look for it.

Troubling thoughts!