

To the Independent Climate Change Email Review Committee,

I am making this submission with regard to the discussions of data sharing and archiving underlying your inquiry, specifically relating to the second element of your remit:

1.2. Review CRU's policies and practices for acquiring, assembling, subjecting to peer review and disseminating data and research findings, and their compliance or otherwise with best scientific practice.

and your sixth and seventh questions

6. The scrutiny and re-analysis of data by other scientists is a vital process if hypotheses are to rigorously tested and improved. It is alleged that there has been a failure to make important data available or the procedures used to adjust and analyse that data, thereby subverting a crucial scientific process.

7. The keeping of accurate records of datasets, algorithms and software used in the analysis of climate data. A key concern expressed by a number of correspondents and commentators has been as to whether datasets, and analyses based thereon, were deleted.

I am a Professor of Chemistry at Howard University, Washington, DC working in the area of chemical dynamics and kinetics. While my research focuses on the atmospheric chemistry of the outer planets, I have know and have worked with many chemists and physicists who study processes in the Earth's atmosphere and comment on climate issues at Rabett Run (<http://rabett.blogspot.com>)

Much of the commentary about data sharing and archiving issues at the CRU refers to periods one or even two decades in the past. Both official policies and informal practices in those areas has been rapidly changing and it would be a mistake to judge prior actions based on current regulations and practices.

For example, Nature which has one of the strictest data sharing policies, changed its instructions to authors in 1996 from

Nature requests authors to deposit sequence and crystallography data in the databases that exist for this purpose

those fields being the first to establish such data archives, to the current

Materials: As a condition of publication authors are required to make materials and methods used freely available to academic researchers for their own use. Supporting data sets must be made available on the publication date from the authors directly and by posting on Nature's web site, by deposition in the appropriate data base or on the internet.

Most other journals have much less stringent policies, yet those too have changed over time.

The issue of data archiving is another which has rapidly changed as the cost of storage has rapidly sunk and internet access has become ubiquitous. Even today, there is no requirement to retain data indefinitely. The US NIH Office of Research Integrity advises that

***Period of retention.** Data should be retained for a reasonable period of time to allow other researchers to check results or to use the data for other purposes. There is, however, no common definition of a reasonable period of time. NIH generally requires that data be retained for 3 years following the submission of the final financial report. Some government*

programs require retention for up to 7 years. A few universities have adopted data-retention policies that set specific time periods in the same range, that is, between 3 and 7 years. Aside from these specific guidelines, however, there is no comprehensive rule for data retention or, when called for, data destruction.

<http://ori.dhhs.gov/education/products/RCRintro/c06/3c6.html>

Guidelines offered by Kings College (London) are similar, recommending a seven year retention period for data from funded research and a four year period for data from unfunded research.

<http://www.kcl.ac.uk/content/1/c6/05/57/93/Howlongshouldikeepmyresearchdata.pdf>

The Climate Change Email Review has a two fold task First to evaluate past practices and second to make recommendations for future ones. I would urge you to carefully evaluate past practices in light of past expectations, which will require establishing what those expectations were. This can be time consuming, as I discovered today. If I may offer a small suggestion, the services of a research librarian would be of help to you in this. I would also urge you to avoid the trap of setting a requirement in the future that everything, every scrap of paper, every bit be archived. Besides requiring a huge amount of time for the researchers, such a massive effort would require significant additional resources and a large number of skilled personnel. I believe that in this, as much else, 90 to 95% or even a little less, is better than 100%, when the cost of the 100% is that everyone spend all their time archiving data.

I thank you for this opportunity. I apologize for not sending this letter before the 1 March deadline, but an exchange of email with your organization assured me that the Inquiry will continue to read submissions sent even at this late date.

I remain

Very truly yours

Joshua B. Halpern

Professor of Chemistry

Howard University

Washington, DC USA