UEA – CRU Review
Initial Report and commentary on email examination

Professor Peter Sommer

Terms of Reference

1. In November 2009 a number of emails relating to the work of the Climatic Research Unit (CRU) at the University of East Anglia (UEA) appeared on various websites and were subjected to hostile interpretation on various "blog" sites and in the mainstream media. They were believed to have been "hacked" from a CRU back-up server which itself was removed by the Norfolk Police as part of their investigation into the "hack". I am asked by the University of East Anglia to look at the back-ups of the computers of the key researchers in CRU as they are held on the back-up server to see if it is feasible to identify email traffic which was not publicised on the various websites, but nonetheless related to the same issues and might justify further investigation by the Independent Review into the publication of the emails and the allegations of inappropriate scientific and other practice which had subsequently been made.

2. I am not part of the Review Team headed by Sir Muir Russell, nor have I any part in the investigations by Norfolk Police. I have at this stage no knowledge of the technical means by which the emails were acquired from the CRU.

3. I have been supplied by the University with a "thumb drive" said to contain copies of all the emails known to have been published on the websites. I have also been supplied by Norfolk Police with three further "thumb drives" containing the emails extracted from the back-up server associated with the computers of the researchers in question. The extraction was carried out by Qinetiq, as contractors to Norfolk Police.

Conditions for Examination

4. As the back-up server and its contents remain a part of an ongoing police investigation, Norfolk Police has attached high importance to maintaining the security of the data supplied to me by their forensic consultants for the search. I was given access to the material only on 14 May 2010. The material has been given a very high level of security classification which requires that I work at secure facilities and follow particular protocols which, for example, preclude computers being left to run unattended or overnight and at weekends. These procedures, while providing a very high standard of protection to the data, are also very time consuming, particularly in the light of the need for the Review to conclude its work in a timely manner.

Initial Examination of the Thumb Drives

5. In most conventional email software, the emails are stored in a series of archives or mailboxes. In addition to "in" and "sent" mailboxes there may be one for "trash" which covers recently deleted material and any of a large number of thematic mailboxes in which the user has saved selected sent and received emails on a "project" basis. Different email packages use different formats for these archives. Most email packages allow for a very limited amount of searching of the archives so that users can trace and read old correspondence.

6. The emails relating to researcher "A" go back to at least 1996 and amount to 1.99Gb as stored on the thumb drives. There is a reasonable likelihood of emails being duplicated, for example in "in" and "out" boxes. The stored emails for researcher "B" amount to 3.93Gb. The email archives for researcher "C" amount to 2.03Gb. Researcher "A's" emails are contained in 46 "thematic" mail boxes and Researcher "C's" some 300 "thematic" mail boxes. Researcher "B's" emails are contained in some 50 mail box archives, but these are mainly "in" or "out". The emails as provided to me are in the format of an email program called "Thunderbird" and before they can be searched require indexing. There are large numbers of un-indexed emails and time constraints in preparing this initial report preclude indexing and any form of sophisticated analysis.

Problems of email analysis

7. I am told there are some 1,073 primary emails and their associated threads in the published material, which itself amounts to 18.7Mb. The further emails provided for analysis amounted to 7.95Gb, approximately 425 times the amount published on the web. An analysis, against selected terms or combinations of author and recipient will require the deployment of specialist software with more sophisticated searching facilities than are available in regular email software. Almost certainly there would need to be conversion of the "Thunderbird" archives into a format that the specialist analysis software requires; together with the "published" emails. The analysis package will have to export selected emails so that they can be printed out. I strongly suspect that any high level analysis I can conduct within a reasonable time would produce an unmanageable quantity of material. Any further analysis would have to be conducted by those familiar with the material and they would have to learn how to use the analysis programme. There is the further practical problem, familiar to me from various legal instructions, that email traffic is often highly informal and allusive, with the consequence that any investigator has to relate large numbers of emails to other types of evidence of particular events.

8 The processes of analysis to identify (and then review) additional email traffic which might be associated with the issues which are the subject of the allegations which have been levelled against CRU, is likely to take at least several weeks. It would be for the Review Team and the University to determine whether the cost, inevitable time delays and (at this time) uncertain outcomes could be justified. Until the material is subjected to a much downgraded security level, the likely position will be that the University and its appointed team will not be able to carry out any meaningful analysis.

Professor Peter Sommer

http://www.pmsommer.com

17 May 2010