Brian Summers BSc
Registrar & Secretary

University of East Anglia
Norwich
NR4 7TJ
England

Email: b.summers@uea.ac.uk
Tel: +44 (0) 1603 592771
Fax: +44 (0) 1603 507753
www.uea.ac.uk

Professor Peter Sommer
67 Mount View Road
London
N4 4SR

By Fax: 0208 3413472

"Dear Professor Sommer

Review of the back-up copies of the PCs of Professor Phil Jones, Professor Keith Briffa and Dr Tim Osborn

1. Thank you for agreeing to undertake work relating to an analysis of the back-ups of the personal computers of the above UEA staff which are being recovered from a UEA server currently in the possession of the Norfolk Constabulary and their consultants, Qinetiq.

2. We have asked Qinetiq to provide copies of all emails within those back-ups which have been authored, received or stored by these colleagues. We anticipate that we will receive this material on portable hard drives and we expect this will be available progressively from Monday 26 April. We will arrange for the hard drives to be delivered to you, we anticipate by courier.

3. You have received from us a weblink which shows around a thousand emails which were "hacked" from the server and published extensively on the web. Firstly we are asking you to identify the search terms or other methods by which it is most likely the emails which were published on the web were selected from the large volume of material represented by the back-ups of the various personal computers, and to ascertain whether there are other emails which would have been selected through such a process but which have not been published on the web and, therefore, could be retained by whoever was responsible for the "hack". As you know this is part of a review of the "hack" and the subsequent fall-out being conducted for us by Sir Muir Russell and his team. They will have some thoughts as to what the process may have been and will share these with you.

4. Secondly, we would like you to investigate whether there are emails which referred to any one of a number of third parties (I am currently finalising the list) which similarly were not published on the web.

5. Ideally we would receive from you, on appropriate storage media, copies of any emails which were not published but nonetheless are revealed by your searches under 3 and 4 above, separated as to the owners of the PCs from which they were backed up, as single emails not replicated in email strings, and ordered chronologically.

6. I understand that your approach would be, firstly to make contact with both the Norfolk Constabulary and Qinetiq to see what additional information you can gain from them and then to seek to scope out the exercise and report to us on the steps to be taken, the likely timescale and estimated future costs. At that point, and routinely during any future work, you will keep us informed of progress, discuss with us what further work can be done, and if it is sensible to continue with the work given the likely results and time and resources that will be required, or to modify the approach.

7. The University will pay you at the rate of £125 per hour plus VAT, I am happy to either receive interim invoices from you, or a single invoice at the conclusion of the work. The University will pay invoices within 30 days.

8. We have agreed that we must review progress routinely and that the University may call a halt to this work at any time, and reimburse you for work undertaken up to that point.

9. You will exercise the highest professional standard in undertaking this work and will always act with the utmost good faith towards us. All the materials submitted to you from the server will be kept securely by you and maintained in strictest confidence.

10. The external hard drives will be returned to us at the conclusion of the work, and no copies will be kept by you.

I would be delighted if you could indicate that you accept this engagement on the terms set out above by signing a copy of this letter and faxing it to us on 01603 507753.

Regards

Brian Summers

Registrar & Secretary

26 April 2010

....

Professor Peter Sommer Date"

Rymarz Elaine Mrs (REG)

From: Rymarz Elaine Mrs (REG) on behalf of Summers Brian Mr (REG)
Sent: 27 April 2010 09:17
To: 'peter@pmsommer.com'
Cc: Williams Lisa Ms (VCO)
Subject: COMMERCIAL IN CONFIDENCE: The Review of the Climatic Research Unit (CRU) emails

As from Brian Summers

Dear Peter

Thank you for returning the signed contract. I understand you are having some discussions with Julian Gregory, just to satisfy the police as to your "credentials". I understand you are also, through this process, intending to arrange for external hard drives to be delivered directly to you (or the data sent to you by some other mechanism) rather than coming via the University. I am entirely happy for this to be the case.

Regards
Brian

Elaine Rymarz
PA to the Registrar & Secretary
University of East Anglia
Norwich
Norfolk NR4 7TJ

e.rymarz@uea.ac.uk

Telephone 01603 592771

91348

2

Peter Sommer
DIGITAL FORENSICS SERVICES
67 Mount View Road London N4 4SR UK
☎+44 20 8340 4139 🖷+44 20 8341 3472 ☽ 07802 898135
peter@pmsommer.com P.M.Sommer@lse.ac.uk
http://www.pmsommer.com

Brian Summers
Registrar & Secretary
University of East Anglia
Norwich
Norfolk NR4 7TJ

26 May 2010

INVOICE
Climate Change Enquiry: Email Aspects

Work as per contract letter 26 April 2010 between 26 April and 24 May 2010. Advice, examinations, phone calls, correspondence with police and Qinetiq and report in connection with the emails review. 37 hours @ £125/hr	4625:00
Vat @ 17.5%	809.38
Total	**£5434.38**

UK VAT Registration No 554 2866 23

Please pay:

[Exempted pursuant to s.40(2), Freedom of Information Act]

INVOICE PAYMENT AUTHORISATION
ORDER No EIS00573/11
PAYMENT 5,434.38
VAT INC 1
NAME: BRIAN SUMMERS
SIGNATURE:
DATE: 26-5-10

UEA-CRU Review

Initial Report and Commentary on Email Examination

Peter Sommer

Terms of Reference

1. I am asked by the University of East Anglia to conduct a preliminary review of emails potentially relevant to to the UEA-CRU enquiry into what is colloquially referred to as the Climategate Hacked Emails scandal. A number of emails relating to the work of the CRU appeared on various websites where they were subjected to hostile interpretation. They were of course only a very small proportion of the total number of emails sent and received by the leading figures in the CRU over the period March 1996 to November 2009. I am asked to provide the means by which the totality of emails can be considered in relation to those that were published and to provide some initial filtering. The filtering is partly to assist the review team but also to ensure that emails located on UEA computers but which are not relevant are appropriately protected.

2. I am not part of the review team headed by Sir Muir Russell – the Independent Climate Change Email Review - nor have I any part in investigations by Norfolk Police.

3. Although there are also some longer term objectives, my initial terms of reference were to form a view of the feasibility of searching through emails forensically recovered by Norfolk Police's specialist consultants, Qinetiq from UEA computers and servers. Searching would be directed at:

    - Locating emails which appear to cover the issues in the Climategate scandal but which have not appeared on public websites; the aim here would be to make any potential candidates available to the team

    - Attempting to form a view of how the published emails were selected from the totality, perhaps by identifying search terms, etc. It should be noted that I have at this stage no information about how the emails were obtained by the website publishers and in particular any specific technical means deployed.

4. The enquiry under Sir Muir Russell wishes, I understand, to publish relatively quickly to meet the need to provide an account of events to the public. To that end I am asked to concentrate on solutions for them that might produce quick results rather than an ultra-reliable exhaustive examination.

5. I also understand that the University may have a separate requirement to be able to search through the emails in order to respond to requests from the Information Commissioner and under the Freedom of Information Act.

6. I have been supplied by the University with a thumbdrive said to contain copies of all the emails known to have been published on websites. I have also been supplied by Norfolk Police with three further thumbdrives containing emails extracted from UEA servers and representing messages current and passed associated with Keith Biffra, Phil Jones and Tim Osborn. The extractions were carried out by Norfolk Police's consultants, Quinrtiq.

7. I have also been supplied with a list of keywords which the team say will assist my search,

Background: Email Programs and Archives

8. All modern email programs have facilities to store messages received and sent. In the vast majority of such programs the messages are not stored directly as simple text but are held in databases which can facilitate searching through the collection of messages. Different programs use different database internal structures even if the underlying principle is the same. .

9. Each separate database is sometimes referred to as a "mailbox" or "mbx". All email clients tend to have "in" and "out" or "sent" mailboxes. The precise terminology varies between different programs. Given the large numbers of messages that an active email user will acquire over a period of time, some email programs allow them to store messages in separate, thematic databases or "mailboxes". The theme could relate to a particular activity or interest. There may also be a requirement to maintain a formal archive for business or regulatory purposes.

10. Where some-one wishes to try a new email program but also read older messages within it, they must seek out a conversion program. If as in the present case one wishes to review a large number of emails one cannot, therefore, perform a simple string or keyword search across a file. The search facility has to "know" the specific database structure. Where one may have to examine emails from several different sources, almost certainly it will be necessary to convert them into a single format.

Conditions for Examination.

11. Norfolk Police have categorised the material supplied from Qinetiq as "Secret". A number of procedural consequences follow. I have signed an undertaking covering confidentiality and secure handling. The police also raised a number of issues relating to physical security. .

12. After some negotiation, I am working at law enforcement premises in London. They have secure facilities and are providing them partly as a favour to Norfolk Police and partly as a favour to me as I carry out separate consultancy work for them. They have no interest in the outcome of any climate change investigation.

13. I did not get proper access to the three thumbdrives until 14:25 on 14 May. Although I was not immediately aware of this, each thumbdrive had been encrypted – the drives initially appeared to be empty. Eventually I spoke to Qinetiq who told me about the precise encryption method used. I downloaded it and installed the relevant package while Qinetiq sought permission from ***[Exempted pursuant to s.31(1)(a), FOIA]*** It was not until 16:25 that the decryption had been fully completed.

14. "Secret" procedures require that all my working materials have to be securely stored over-night – they are placed in a sealed and tagged box and the box itself goes into a very large safe. There is an audit trail jointly operated by myself and the law enforcement agency . My own working hours are limited by the availability of law enforcement agency staff to facilitate this – between 0900 and 1700 week-days.

15. One consequence of this is that over-night indexing of material – an essential feature of all large-scale text and email analysis – is difficult to arrange.

Initial Examination of KB thumbdrive, coresponding to Keith Biffra

16. Qinetiq say: This drive contains emails located with 'Eudora' database files on the machine 'angara' belonging to Professor Keith Briffa. Four (4) backups were located within folders entitled '388', '389', '390' and '391'. The original database files were compressed on CRUBACK3 and bore a naming convention beginning with 'f' e.g. 'fIn.mbx'. This naming convention has been retained. Each of these database files has been imported to the 'Thunderbird' application on this drive. 'Thunderbird' can be launched by running the file '\ThunderbirdPortable\ThunderbirdPortable.exe'. On the left hand side of the screen are the imported 'Eudora' database files which have been named according to the machine they are from (angara) and the backup folder that the database was originally located in e.g. '388'. For example, the database 'angara-391-fIn.mbx' denotes the 'In.mbx' database held on 'angara' within the backup folder '391'.

17. This may need some translation: Keith Biffra (and indeed his colleagues) used an email client called Eudora (http://www.eudora.com/) Qinetiq have converted the back-up files generated by this application into a format used by another email program, Thunderbird (http://www.mozillamessaging.com/en-GB/thunderbird/) on the basis that it has easier searching facilities and there is a version which can be used without making a full installation on to a computer.

18. One significance of this is that I am not examining the material in the formats in which they were originally located on the various UEA machines. Qinetiq have provided the material in a "convenient" form. Although I have no reason to doubt their expertise I am not able to say of my own knowledge that the material is complete and reliable in its own terms. This "convenient" approach also limits the extent to which I can comment on how the published emails were selected.

19. There are 9 "390" back-up mbx files; 25 "388" back-up mbx files; 5 "389" back-up mbx files; 7 "391" back-up mbx files.

20. Emails go back at least to 1996. In the form supplied to me the mailboxes amount to 1.99 GB

21. It looks as though Keith Biffra stored his emails in a series of thematic mail-boxes which, if he has been rigorous and careful, might simplify searching. Each mbx mailbox is named and we can guess the contents from the name. One is called Ebay, for example, and relates to Ebay transactions. There are also in and out boxes plus spam boxes. This increases the likelihood of duplicates.

22. The following is a list of the mbx boxes – where there are numbers in brackets this refers to the quantity of ***unread*** emails in the box (unread, that is, at the time of archiving) not the total number of emails, which could be many times more:

Inbox

Trash

Outbox

LFolders

Outbox

anag(3724)0-feBay.mbx

anaga(724)0-fECOCHANGE.mbx

anaga(1901)-fIMPRINT.mbx

anag(22548)-fIn.mbx

anaga(4490)-fitrdbfor.mbx

anag(10201)-fOut.mbx

anagara-390-fspam.mbx

anagara-390-fTrash.mbx

anaga(33417)fUEAinter.mbx

anga(408)8-f1996.mbx

angara-388-f1997.mbx

angara-388-f1998.mbx

anga(1421)-f1999.mbx

angara-388-f2000.mbx

angar(591)-fOut1999.mbx

angara-388-fOut2000.mbx

angara-388-fspam.mbx

angara-388-fTrash.mbx

angara-388-fUEA2000.mbx

angar(33925)UEAinter.mbx

anga(20768)fIn.mbx

angara-389-fitrdbfor.mbx

angara-389-fOut.mbx

angara-389-fTrash.mbx

angara-389-fUEAinter.mbx

anga(3770)-feBay.mbx

angar(736)-fECOCHANGE.mbx

angara-391-fIMPRINT.mbx

anga(22894)fIn.mbx

angara-391-fitrdbfor.mbx

anga(10207)fOut.mbx

angara-391-fspam.mbx

angara-391-fUEAinter.mbx

23. Thunderbird does not have immediate facilities for me to give a total number of emails associated with Keith Biffra. I can calculate a figure, but I would need time.

24. Thunderbird has some limited in-built searching facilities; in the first instance you need to build indexes for each mailbox. This is done automatically but takes time. In normal use this does not matter – emails are simply added as they arrive or generated. However in the current instance the program finds

large numbers of what are unindexed emails and once the program is opened,it sets about indexing them. Because of time constraints I stopped the indexing.

25. Once indexing is complete you can filter on sender, recipient, subject and entire content. There are few opportunities for combining the filters and no Boolean operators. Nor is there a function to export (ie make a copy of) an email other than on a one-by-one basis. Thunderbird's functions are aimed at helping the ordinary user locate the occasional email, not at bulk analysis. (As it happens I use Thunderbird myself).

26. Because I did not complete indexing I have been so far unable to use any of the keywords with which I had been provided, so that I cannot currently estimate the level of hits.

27. More sophisticated searching is possible using specialist external products such as Ad4Mail Forensic and Intella. (http://www.aid4mail.com/ and http://www.vound-software.com/Information/intella-overview) The search facilities permit identifying emails by "to" in combination with "from" and "subject" and "content" and "date", among others. However their use is not entirely straight-forward. Intella, the more sophisticated product, does not directly "read" the email data in which Qinetiq have provided the email messages. It would be necessary to convert. I have yet to carry out a test to see how easily and rapidly this can be achieved. Any analysis program will need to go through an indexing phase before it will function.

Initial Examination of PDJ thumbdrive, coresponding to Professor Phil Jones

28. Qinetiq say: Four machines were identified as being in use by Phil Jones (PDJ). Email from each machine is included on this flash memory drive. The machine names were:

    - crupdj2
    - m-crupdj
    - m-crupdj4
    - m-crupdj5

    On each machine the mbx files for each backup were found in Cdrive\EUDORA. Email from 'crupdj' can be accessed by clicking 'crupdj/ThunderbirdPortable.exe'. Several mailboxes were located within various backup folders relating to this machine e.g. '250', '265' and '276'. Each has been named according to the folder in which they were located and by the name of the 'mbx' mailbox, e.g. '250-In.mbx' refers to a mailbox entitled 'In.mbx' that was located within the backup folder '250'. Note that backup '250' had an extra set of mbx files stored in Cdrive\static\oldpc_thet\before_xp\oldpc\oldEudora. These mailboxes were named '250-old-<mailbox name>.mbx'. One (1) individual msg file was found

in Cdrive\RECYCLER\S-1-5-21-2947832195-1700217287-3046754539-1004. This file can be accessed by opening the file 'crupdj/Dc3714.msg'.

Email from 'm-crupdj' can be accessed by clicking 'm-crupdj/email.html'. This displays (10) emails each with a link (coloured blue) to the email content.

Email from 'm-crupdj4' can be accessed by clicking 'crupdj/ThunderbirdPortable.exe'. Five (5) mailboxes were located and are displayed on the left hand side of the 'Thunderbird' application window.

Email from 'm-crupdj5' was located within separate folders. The format of these emails differed from the usual 'Eudora' format, and were noted to have been saved as 'msg' files. These folders were copied to 'm-crupdj5/' and contain the original saved emails. Six (6) folders are located within the 'm-crupdj5/' folder. The 'msg' files can be opened using Microsoft Outlook.

29. The stored emails occupy 3.93 GB

30. The **crupdj2** computer contains approximately 50 mbx archives. Professor Jones has used the facility to save emails into thematic mbx archives significantly less than Keith Biffra. Most of the emails are in "in" or "out" mbxs though in each instance there are several of these, presumably corresponding to points at which archiving took place. It is likely that there is some duplication of material. There are also some "trash" mbxs – these should contain material which have gone through a first stage of deletion but have not been completely deleted. (This is a function of the way in which many email client programs work). There are 22018 **unread** emails in the most recent "In" mbx. And 7184 in one "out" box. Again, because of limitations in the Thunderbitrd program I cannot immediately give a figure for the total number of messages . I notice there are a series of archives referring to a mbx called "deniers". These appear to relate to correspondence with, or about, climate change deniers – but this seems to be either the only or the main specialist themed mbx.

31. The computer **m-crupdj** had just 10 emails on it.

32. The computer **m-crupdj4** appears to have very little email and from a brief look, to have nothing to do with CRU's intellectual output – the messages are to do with UEA administration

33. Because of time constrainst I have not at this point looked at the computer m-**m-crupdj5.** The message data is stored in a different format from the others and it would take some time to set up an appropriate reader on the computer I am using.

34. For the reasons stated above, I did not allow Thunderbird to complete any indexing and as a result I currently have no idea about the level of hits on the keywords.

Initial Examination of TO thumbdrive, coresponding to Tim Osborn

35. Qinetiq say: Two machines were identified as being in use by Tim Osborn (TO). Email from each machine is included on this flash memory drive. The machine names were:

    - cruto3
    - cruto4

    Eudora was located on both systems. On 'cruto3' a compressed file known as a 'tar' file was located containing archived 'mbx' format files (email database files). All identified files were decompressed Two (2) user accounts were identified on the cruto3 system 'fe301' and 'ff055' - note that the initial 'f' is there due to the naming convention used by 'BackupPC' - the software in use within CRU. The original account names will have been 'e301' and 'f055'. Each account was found to contain emails and have been separated for clarity. Each were imported into portable installations of the application 'Thunderbird' to enable viewing. Eail from the 'fe301' account can be accessed by clicking 'cruto3/Thunderbird_fe301/ThunderbirdPortable.exe' Email from the 'ff055' account can be accessed by clicking cruto3/Thunderbird_ff055/ThunderbirdPortable.exe'. Note that due to the original location of some of the 'mbx' files within these user accounts it was necessary to rename them to enable viewing within 'Thunderbird'. To distinguish non-uniquely named 'mbx' files located within multiple folders, the filename was prepended by the folder name in which the 'mbx' file was located e.g. the 'mbx' file 'NCAS.mbx' was found in two folders, therefore the second instance of this file, which was located within the folder 'Projects.fol', was renamed to 'Projects.fol-NCAS.mbx'. For 'cruto4' as with 'cruto3', due to the original location of some of the 'mbx' files it was necessary to rename them to enable viewing within 'Thunderbird'. To distinguish non-uniquely named 'mbx' files located within multiple folders, the filename was prepended by both the backup folder name e.g. '123' and the folder name in which the 'mbx' file was located e.g. if the 'NCAS.mbx' was within the folder 'Projects.fol' in backup '123', it was renamed to '123_Projects.fol_NCAS.mbx'. Email from the 'cruto4' account can be accessed by clicking 'cruto4/ThunderbirdPortable.exe'.

36. The email archives, in the form supplied to me, amount to 2.03 GB

37. In the computer **cruto3** the user has created between 250 and 300 thematic mailboxes, each containing up to 400 separate messages. There are two separate archives which might need to be combined if one was to carry out a detailed search. It may be that those with more familiarity than I about the climategate issues, will be able to identify mbxes which can be safely excluded from a detailed search. The mbxes seem to cover regular teaching duties, and administration as well as specific research themes.

38. In the computer **cruto4** The large number of thematic mailboxes also appear, but there are also archives of the same mailboxes. Again, because of the limitations of Thunderbird it is not easy to produce an instant estimate of the total emails present, still less the extent of duplication.

39. As before and for the reasons stated earlier, I did not allow Thunderbird to complete any indexing and as a result I currently have no idea about the level of hits on the keywords.

Conclusions and Observations

40. **Course of Investigation** I was originally instructed with effect from 26 April but did not receive the copies of the emails recovered by the police-contracted technicians until the afternoon of 13 May. I am asked to provide a preliminary view by close of business on 17 May but have not had access to the email material over the week-end of 15-16 May. As a result my findings thus far have been extremely limited.

41. I recognise that the team under Sir Muir Russell have to take a view of the latest date by which they must publish their report without losing public credibility. I recognise also that it is for the team to decide which material is relevant to their work.

42. **Problems of Email Analysis** There are, I understand, some 1073 primary emails and their associated threads which have so far been published on the "rogue" websites. I have been so far unable to calculate the number of emails associated with computers linked to Keith Biffra, Phil Jones and Tim Osborn, but whereas the published emails, printed out, amount to 18.7 MB, the files across the three Qinetiq thumbdrives amount to 7.95 GB. There are approximately 425 times as many emails on UEA facilities as have been published on the web.

43. Any analysis of the emails of the sort envisaged by the Russell team and the university will require the deployment of specialist analysis software. The emails as supplied in Thunderbird archives will almost certainly need to be converted into a format the selected specialist email analysis software requires. It will also be necessary to convert the "published" emails as supplied to me by the university into a common format and to find a way of marking them so that they can be readily distinguished from the totality. It will also be necessary to establish that the analysis package will be able to export selected emails so that they can be printed out. It will be necessary to have a series of trials to establish the best routes to achieve all of the above. The trials might take 1-2 days under ideal conditions.

44. After the trials the emails will need to be converted and then loaded into the analysis software. The software will then need to carry out some indexing before any actual analysis can take place. At the moment I can't forecast how long this would take, even under ideal conditions, but it could be a further 2 days. Further time may be required to carry out de-duplication. Duplication is likely to exist because of the methods used to back-up and also because there will be email traffic between the three CRU academics.

45. My own knowledge of the events that are the subject of the team's work is quite limited so that I could not do much of the detailed analysis. I could, as I

think has been suggested, produce all the emails corresponding to the keywords I have been given. However I strongly suspect that this product would prove unmanageable by virtue of quantity for the team. They will probably have to use the email analysis program themselves, having learnt how to use it.

46. **Methods of Email Selection** I have been asked to express a preliminary view about the likely processes used by those who published the selected emails on websites. At a very broad guess, the method would have been to use keyword searches to locate material of interest and that there would then be a manual process of refining. However much would depend on the method used to obtain the emails in the first place as this would have a significant impact on the level of access the perpetrators had to the totality of the archive. For example, if access was obtained via an insider with access to the computers of the three principals or the use of a Trojan (back-door) then one option would have been to use the search and filtering facilities within Eudora, the email program used by Biffra, Jones and Osborn. These facilities are similar to those available within Thunderbird – see paragraph 24 above. (In fact Eudora's facilities are rather more sophisticated than those in Thunderbird, and allows searching against regular expressions. This rather raises the question why Qinetiq chose to convert the mailboxes to Thunderbird).

47. On the other hand, if the material was obtained by means of access to back-up servers, then the perpetrators could have used any of a wide number of analysis tools.

48. I might be able to offer a further comment on this point if Norfolk Police are prepared to share with me, on a confidential basis, any views they have formed on the methodology of the hack. However I have to recognise that they are still in the midst of an ongoing investigation.

49. **Implications of "Secret" categorisation** There remains the problem of the decision by Norfolk Police that they regard the emails taken from UEA servers as "secret". I am not aware of the factors that led their risk analysis to this conclusion. At a practical level it has severely delayed and limited my own work thus far. I strongly suspect that the university and the Russell team will need to ask the police to consider whether their initial security concerns should be revisited and the material subjected to a much downgraded security level. In effect the likely position will be that the university and its appointed team will not be able to carry out any meaningful analysis of material which was originally created by / associated with, university staff.

50. I am at the disposal of the university and the Russell review team for further discussions and instructions.

Peter Sommer

17 May 2010.