This item was submitted to Loughborough's Research Repository by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

E-mail archiving for records management

PLEASE CITE THE PUBLISHED VERSION

PUBLISHER

EUNIS

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Booth, Garry, Graeme Fowler, and Carys Thomas. 2019. “E-mail Archiving for Records Management”. figshare. https://hdl.handle.net/2134/3091.

Loughborough University

This item was submitted to Loughborough's Institutional Repository by the author and is made available under the following Creative Commons Licence conditions.

creative commons

COMMONS DEED

Attribution-NonCommercial-NoDerivs 2.5

You are free:

- to copy, distribute, display, and perform the work

Under the following conditions:

Attribution. You must attribute the work in the manner specified by the author or licensor.

Noncommercial. You may not use this work for commercial purposes.

No Derivative Works. You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the Legal Code (the full license).

Disclaimer

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/

EMAIL ARCHIVING FOR RECORDS MANAGEMENT
EUNIS 2007

Garry Booth, Graeme Fowler, Carys Thomas
Computing Services Department, Loughborough University, Epinal Way, Loughborough, LE11 3TU, United Kingdom
c.m.thomas@lboro.ac.uk

1. ABSTRACT

The business of universities is increasingly transacted by email, and the management and preservation of stored email is an important part of an institution's records management process. Until recently at Loughborough University email was mostly sent from, and received into, email accounts assigned to individuals, and was stored and deleted by these individuals as they saw fit. Stored email was deleted when staff left. It was backed up for business continuity, but backups were kept for a short time, and not for archive purposes.

In 2002 the University successfully bid for funding from JISC (Joint Information Systems Committee) under the programme "Study of the Records Lifecycle, specialist electronic studies" for a project to examine institutional email. An overview of the findings of this project is given, along with results of surveys of UK universities in 2003, 2004 and 2006 into their practice in archiving email.

In 2006 the University embarked on the process of replacing the staff email system, the intention being to implement email archiving as part of the rollout, drawing on the work of the earlier project which had produced a draft email retention policy for the University.

The design of the archiving facility was determined by user consultation and by involving the University senior management in consideration of the risks to being covered. Technical issues to do with the choice of server and supported client platforms also heavily influenced the choice of solution that is now being provided.

The University is currently in the early stages of rollout of the email upgrade, and we describe the method we are using to implement the archiving facility.

2. PROJECT (2003) - INSTITUTIONAL RECORDS MANAGEMENT AND E-MAIL

The Loughborough University project[1], "Institutional Records Management and Email", was funded as part of the JISC[2] "Study of the Records Lifecycle" programme[2.1]. The aims of the project were to:

- Examine current working practices
- Develop policies for the retention and disposal of email
- Evaluate technical options for archiving email
- Inform Loughborough University and UK HE sector of the findings

Forty interviews with staff were undertaken, with a range of staff roles across the University selected. The report of interviews [1.1] concluded as follows:

> "All interviewees viewed e-mail as important to the way they carried out their work; it had become their main tool of communication and has largely superseded the use of the telephone and paper correspondence. There appeared not to be any area of business where e-mail was not used to transact the business of the university. The ease and speed with which e-mail can be used to disseminate information on a broad scale to staff and students alike was seen as a benefit. Similarly, staff felt the speed of communication made them more efficient in the way they worked. In many cases e-mail transactions were used to agree responsibilities and to deliver decisions that were viewed with having the same authority as if they had been signed by the sender.
>
> "Staff generally viewed e-mail as being of crucial importance and some felt they had some ownership of it rather than seeing it as corporate property in their custody. The way staff managed their e-mail inevitably varied and was dependent on their work style, they all however, made independent decisions on whether to retain or delete e-mails. Retention of e-mails was often based on having evidence of decisions or instructions or where sensitive information was thought important to keep; and were often printed and filed. Staff felt confident they could retrieve e-mails on request and had a reasonable understanding of the Data Protection Act. Knowledge of the Freedom of Information Act was much less certain and generally they would need advice to interpret requests related to it.
>
> "In any e-mail archive staff generally felt, although not exclusively, that they would like some ownership of it and be able to control access to it on the basis that some of the content was of a confidential nature. Most staff thought they would benefit from training which addressed e-mail management and the Data Protection and Freedom of Information Acts."

A second deliverable of the project was to draft an email retention and disposal policy. A generic version[1.2] was produced, for institutions to use as the basis of their own policy, along with a customised version for Loughborough University. In summary, the draft policy states that:

- Emails are owned by the institution;
- It is the responsibility of individual staff to decide what emails to archive, and to capture these as records and retain in accordance with the University's retention schedule;
- It is the responsibility of the University to provide the technology to do the archiving, plus guidance and training.

The project undertook research into technological solutions, by means of a market survey of products available at that time. See the final project report for further details[1.3]. It identified three broad strategies available for the archiving of e-mail:

- Automatically archive all e-mail in and out of the institution.
- Automatically archive e-mail selected on the basis of policy rules in terms of addresses date-stamps, folders and / or content.
- Provide users with a sophisticated archive tool and define institutional policies for its use.

The first strategy of automatically archiving all e-mail seemed at the time to be inapplicable to the HE sector, although it should now be noted that the regulatory environment has since moved in this direction, following the lead of the US. The second option of selectively archiving e-mail using a set of policies or rules was the approach adopted by leading products of the time, which predominantly used Microsoft Exchange server technology and Outlook

related email clients. No suitable products were found which matched the third option of providing a sophisticated archive tool for users to tailor their archiving of email records. The software cost of these systems was in the order of about £85 - 100K + VAT with 15 - 20% annual recurrent costs. The cost of additional server hardware and suitable long term storage would also need to be considered.

3. SURVEYS OF PRACTICE IN OTHER UNIVERSITIES

As part of the project, UK universities were surveyed[1,4] during 2003 to discover their current practice in terms of archiving email. The method used was an informal request for information to IT Directors, via an email list for such information sharing among this group, operated by UCISA, the Universities and Colleges Information Systems Association[3] . This survey has been repeated twice, in 2004 and 2006.

Survey of UK Universities, 2003

- 21 institutions responded
- Only general policies were evident
- Respondents were backing up their email rather than archiving it
- No institution was identified which had a well defined email archiving policy

Survey of UK Universities, October 2004

- 26 Universities and Colleges responded
- 22 not archiving email
- 4 in progress
- - Two taking a copy of all email
- - Other two – focus was management of mailbox sizes

Survey of UK Universities, March 2006

- 28 Universities and Colleges responded
- 1 has implemented
- 7 in progress
- 20 not archiving email
- Movement towards installing such systems
- Focus is technical management of storage, not records management

The response rate was around the 20-25% mark each year, with a different set of institutions responding each time, and so the results can not be considered in any way statistically conclusive. However the replies and the comments accompanying them did indicate that over the three year period there was an increased awareness of a requirement for archiving email, and that systems were increasingly being implemented.

4. WHAT IS "EMAIL ARCHIVING"?

It became noticeable from the survey of University IT Directors that there were different drivers for implementation of an email archiving system, and the term "email archiving" had several different meanings:

- Our project had focused on the records management aspect of email storage management, i.e. the selection of emails to be preserved as records, and kept in accordance with a defined retention schedule.

- Users often have a different idea about what "archiving" means – in some cases they define it as removal of email off the central servers onto their local hard disks. This latter meaning is almost directly the opposite of the "records management" meaning, which implies long term storage by the institution.

- In the IT industry, in a technical sense, "archiving" tends to mean the removal of "old" or unused material off expensive fast storage media onto cheaper, slower disks or tape.

- Another common meaning of "email archiving", in the US in particular, is keeping a copy of all email into and out of an organisation for reasons of legal compliance. This archive has to be demonstrably tamper proof so that an audit trail of verifiably accurate emails can be provided when requested by law enforcement agencies etc.

Most of the commercial systems available to support email archiving concentrate on the above two types of "archiving". Our survey of systems found few that provided adequately for the first (records management) type.

- One further potential driver for preserving emails, identified by Maureen Pennock in the Digital Curation Manual[4] is for future cultural, social or historical research, for instance by forming collections of correspondence of notable public figures. No standard technical approach is in evidence for this type of preservation.

5. DESIGN – MANAGEMENT AND USER CONSULTATION

The draft email retention and disposal policy, produced by the project in 2003, was broadly supported by Loughborough University Senior Management, through a series of presentations and discussions at various committees. However it was felt that it should be adopted as a model of good practice, rather than as a mandatory policy to be in any way enforced. Some of the issues raised were as follows:

- Ownership. Does email belong to the institution or the individual? There are drivers for institutional ownership of e-mail records, but users see e-mail in particular as an informal method of correspondence, and their own personal property. There are related issues of academic freedom and intellectual property rights.

- Individual staff have different personal styles and preferences for organizing email.

- In the absence of any high profile (and high cost) legal cases in our sector, the cost of implementation, particularly in terms of staff time, could not easily be justified.

- Records management is not generally high on any manager's list of priorities, and it is not clear where the overall responsibility lies.

- It could be argued that a records management solution for email should be part of a future wider electronic document management system.

As a set of guidelines for email storage, a version of the policy[5.1] was adopted by the University's Email Advisory Group in October 2006.

As part of the wider implementation of a new email system, user consultation[5.2] was conducted via an online questionnaire. This included the following open ended question about archiving - our intention being to raise awareness of the issues of email retention and disposal, as well as seeking views.

> "E-mail is widely used throughout the University as a convenient way to exchange information to internal and external parties. Currently, saving e-mail messages in a safe retrievable format usually involves printing off and filing hardcopies of the relevant documents. However, this can be a time consuming and arbitrary process, and may not always capture all relevant information since it is dependent on the sender or recipient remembering to print and correctly file the e-mail message without deleting it.
>
> An e-mail archiving system would provide mechanisms to allow users to store e-mail messages safely for a period of time, or until the record is no longer useful. Furthermore, the introduction of an e-mail archiving system is of critical importance to the University as increasing regulatory compliance requires records to be kept about various areas of University business. The University has a duty to provide an audit trail of these records when required.
>
> A central e-mail archiving system would enable the University to respond to any such requests and enable the University to comply to requests for information made under the Freedom of Information Act. Any access to the email archive will also need to comply with the Regulation of Investigatory Powers Act 2000, the Data Protection Act 1998 and Human Rights Act 1998, which offer individuals protection against unwarranted access to any such archive.
>
> Please identify any e-mail archiving features you consider necessary or desirable."

Responses to this question indicated the following considerations which were fed into the design:

- Staff wanted control over their own e-mail and how it was archived. For example, one user responded:
  "I do not consider that it is right or proper for the university to centrally archive e-mail, apart from current e-mail in users' inboxes for backup purposes."

- They wanted a system that would be easy to use. It became clear that any system that involved a time-consuming operation of adding of metadata to individual emails would not be well received.
  "I'd like it to be the equivalent of moving a message to a folder, and just as easy to retrieve."

- There was a range of understanding of the term "archiving", with some clear indication of the need for a University repository of records.
  "The facility to file emails electronically on student files."
  "University File Structure so that user chooses to place emails (eg approving projects) into University space."

- However staff understanding of "archiving" did not always match with a corporate records management approach. Considerable education would therefore be needed if we were to achieve archiving for records management purposes, as distinct from personal archiving.

  "It's essential that I find a way of getting e-mails off the server and storing them easily."

  "Personal folders to store archived E-mails in."

  "I already save all my mail on my machine."

6. DESIGN – RISK ANALYSIS

The design of the archiving facility was greatly influenced by the analysis we undertook of the risks that the system was intended to address. This was a topic of considerable debate, as different groups of staff and management had radically different views on the relative likelihood and impact of the various risks identified. Several changes to the design were made during this process, and it is likely that this shifting basis for the implementation will continue over time. An additional requirement for our solution, therefore, was the need to allow for changes to the policies and implementation in future. The assessment of risks will inevitably change, and rapid response could be needed, for example in face of a high profile and high cost legal case in the sector.

The risks identified were as follows. We have not recorded here our assessment of likelihood and impact for each risk, as this remains a volatile area, but can supply the current analysis on request.

- The volume of email storage on the main email system compromises performance, ability to backup, length of time to restore, and ability to restore. This was identified as the most pressing risk by the IT staff responsible for managing the email system. The existing system allowed staff unlimited email storage, and management, including backup and restore, had become extremely problematic.

- Difficulty and great staff costs in providing email required for Data Protection subject access request, Freedom of Information requests, law enforcement investigations etc.

- Difficulty in finding emails due to huge amount of unnecessary email that is kept.

- Difficulty proving authenticity of emails produced in evidence, leading to legal cases being lost.

- Vital University email records being lost when staff leave.

- Email records being lost because staff are unaware of the importance of keeping them and so delete them.

- Accidental deletion of email by staff.

- University email records deleted deliberately by staff for malice or to cover up a particular situation.

- Vital email records in individual mailboxes being difficult to discover due to staff absence and illness.

- The need to search across multiple staff mailboxes to get the complete picture or provide an audit trail.

- Vital University email records being unavailable because they are stored on personal archives on individual staff machines through use of POP, "archive" function in Outlook etc.

- Litigation or other problems due to emails being held longer than they should be. (e.g. staff disciplinary records.)

- Litigation or adverse publicity due to inappropriate email being kept and discovered. Assessment of this risk as high by University Senior Management caused us to abandon an initial proposal to archive all incoming and outgoing email, and keep it for a proposed period of 2 years. Implementing such an archive would have mitigated several of the other risks.

- Cost and difficulty in preserving emails across media when systems change etc.

- Potential legal action caused by keeping email – i.e. to do with confidentiality, personal freedom, Human Rights act, Regulation of Investigatory Powers Act.

- Research by historians/social scientists hampered by deletion of valuable historical and cultural records.

7. DESIGN – TECHNICAL CONSIDERATIONS

The design of the archiving solution was being undertaken alongside choice of email server platform and supported email client. The choice of main supported email client was made fairly readily. The University's supported email client at the time was Microsoft's Outlook Express. Users were familiar with this, but many were calling for the greater functionality provided by the full version of Microsoft Outlook. Furthermore, the cost of Microsoft Outlook licences was already covered by our Campus Agreement. The decision was thus taken early on that unless any insurmountable problems came to light, Outlook would be the supported client. However it was recognised that many staff would choose other clients, and the server platform chosen must support these. It was also a requirement that a Webmail interface must be made available - as fully functioned as possible, but also with a requirement for a basic version that would work well in low bandwidth situations.

The choice of email server platform was more problematic. A number of options were considered, and evaluated in terms of:

- Cost (initial and recurrent), end-user functionality;
- Staff skillset required for support;
- Standards compliance;
- Operating system and hardware requirements;
- Mail storage format;

- Management API;
- Resilience / High Availability.

After a formal tender process, the choice was made of Communigate Pro[6] for the server platform, using their MAPI connector to provide support for Microsoft Outlook and third party Webmail "skins" to provide the fully functioned and lightweight webmail interfaces.

When it came to looking at the choice of product for implementing the email archiving solution, a number of products were evaluated, and the following issues were considered:

- Cost. This included software licences. The cost of the archiving appliance was also an issue. This could be additional to the actual disk storage, which would be needed for most of the archiving solutions considered, although some provided hosted off-site services. Comparison of cost models was found to be a complex issue, as different solutions used a different basis for charging.

- Technical complexity. A hierarchical storage approach would be needed, with immediate online storage on fast disk medium, with slower disk technologies, or perhaps nearline or offline storage acceptable for parts of the archive. It was recognised that increased complexity and familiarity with proprietary systems meant that the cost of staff training and effort would be greater for some of the solutions considered.

- Integration with e-mail platform and institutional storage provision. This limits choice of solutions. Most archiving solutions were found to be heavily tied into specific e-mail systems and mass storage solutions. Most would support Microsoft Exchange and Lotus Notes. A few were agnostic and would work with RFC standards.

- Catering for diverse user base. Loughborough University has a range of computers using different operating systems, the most common being Windows (various versions), Apple MAC, and Linux. Some of the archiving solutions considered would have left MAC and Linux users having limited ability to interact with the archiving system. Users expect freedom to choose their email client, and do not appreciate any attempt to impose standards. A number of these systems assume Microsoft Outlook is the mail client being used and have written their software specifically to integrate with this. Non-Outlook users may have to use the archive appliance native interface to the archive rather than their preferred e-mail client.

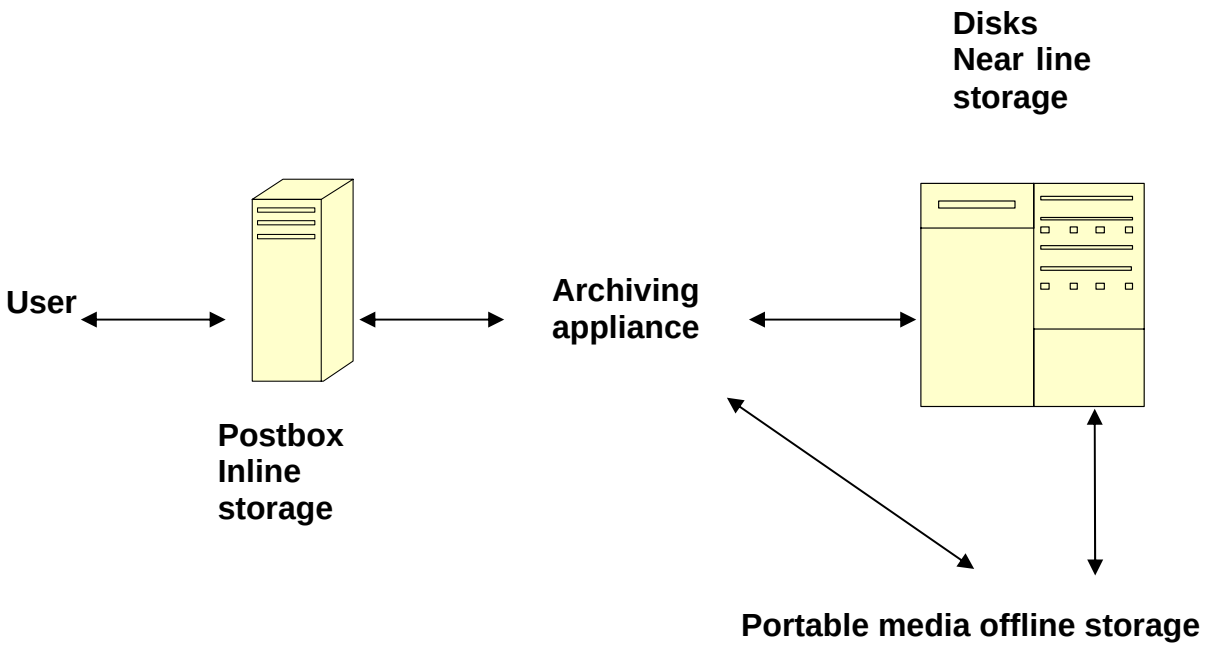

Figure 1 – Storage/archival infrastructure

8. ROLL OUT

A fundamental decision was made to archive email in the same format are regular email. It was felt that this would be beneficial on a number of grounds. Staff would be able to access the archive using the familiar interface of their normal email client, along with its search facilities. There would be no need to leave their email client and enter another interface to access the archive. Further, the archive would be accessible in this way to users whatever their choice of email client, and this had been identified as an issue of great importance to our user base. There would be no specialist application for our IT staff to manage and integrate into our email environment and storage infrastructure. There would be no additional proprietary format for storing archived email. We would avoid the cost of purchasing such a system, and the limitations that this would impose on our storage infrastructure now and in the future.

The major disadvantage of this approach was the additional storage requirement, as many specialist archiving systems use data compression techniques to reduce the storage, and also store each email once only, linking to it as required. We would be storing multiple copies of uncompressed email. However, by separating the archive from the live email storage we judged that we would still gain the management benefits of limiting the volume of our live mail email storage.

Staff would have no need to add metadata to their emails in order to archive them - categorization would be achieved simply by foldering the email correctly, providing the ease of use that had been identified as an essential requirement. Each department would define its own fileplan – a hierarchy of shared folders, each with defined retention period and access list, and a tool would be provided to allow nominated department administrators to edit the folder structure, to add new folders, delegate permissions of part of the folder structure to other users, etc. We would implement a quota for live storage, which would be held on fast disk

storage, but be more generous with the department archive storage, which would be housed on cheaper disk storage. Existing stored email would be moved onto archive storage, and kept for two years, so that users were only required to process their existing email if it needed to be kept for more than two years. In response to the requirement for users to have unrestricted email storage for themselves as individuals, as opposed to archived email that was stored in their department's archive, we also give staff the option to elect to have a "two-year store" – additional email storage that is kept for two years before automatically being deleted. Staff training about the new storage arrangements is being provided as part of new email rollout.

It was further determined that all email into and out of staff mailboxes would be kept for 30 days, accessible to system administrators if individual emails needed to be restored or in response to requests. This mitigates to a small extent the risk of not being able to produce vital email records for whatever reason. Should there be a future requirement to increase the time period, e.g. due to changes in legal compliance expectations, this can easily be achieved.

Rollout of the new email system is currently under way, with about 10% of users having been migrated by mid May 2007. We are rolling out to a department at a time, so that we spread our support load and can provide customized training for the departments on the new email client including the new features such as the archive and calendar. A range of documentation has been produced, including a short introduction to email storage[5.3] and a FAQ[5.4] which explains the various types of storage in more detail. For guidance on the required retention period for different subject matter, we refer staff to three sources – Loughborough University's retention schedule for records containing personal data[7], JISC's Function Activity Model (FAM) & Record Retention Schedule (RRS)[2.2], and the JISCInfonet HE Business Classification Scheme and Records Retention Schedule[2.3].

Loughborough University's new storage implementation has to date been received reasonably well. The majority of users will not need to change their working practice as they will remain within the live mail quota. Departmental managers, administrators and IT staff are, on the whole, engaging well with setting up the fileplans. The "two-year store" is proving popular. Use of the archive is not yet great, but steadily growing. Our Information Science department was one of the first to be moved to the new system and their fileplan is being made available to other academic departments to help them in setting up their own.

REFERENCES (Websites visited May 2007)

[1] Project Website - Institutional Records Management and E-mail, (2003)
http://www.lboro.ac.uk/computing/irm/

[1.1] M. Norris, Loughborough University, Institutional Records Management and E-mail, Report of Interviews, (2003)
http://www.lboro.ac.uk/computing/irm/interviews-report

[1.2] M. Norris, Loughborough University, Institutional Records Management and E-mail, Generic Policy for Email Retention and Disposal, (2003)
http://www.lboro.ac.uk/computing/irm/generic-policy.html

[1.3] M. Norris, Loughborough University, Institutional Records Management and E-mail, Final project Report, (2003)
http://www.lboro.ac.uk/computing/irm/final-report.html

[1.4] M. Norris, Loughborough University, Institutional Records Management and E-mail, Summary of Responses for Information from External Sources, (2003) *http://www.lboro.ac.uk/computing/irm/external.html*

[2] Joint Information Systems Committee, *http://www.jisc.ac.uk*

[2.1] Joint Information Systems Committee, Supporting institutional records management programme (2003) *http://www.jisc.ac.uk/whatwedo/programmes/programme_supporting_irm.aspx*

[2.2] JISC Study of the Records Lifecycle Structure, Function Activity Model (FAM) & Record Retention Schedule (RRS) *http://www.jisc.ac.uk/whatwedo/themes/eadministration/recordsman_home/srl_structure.aspx*

[2.3] JISC Infonet, HE Business Classification Scheme and Records Retention Schedule *http://www.jiscinfonet.ac.uk/partnerships/records-retention-he/*

[3] Universities and Colleges Information Systems Association, *http://www.ucisa.ac.uk/*

[4] M. Pennock, UKOLN, Digital Curation Manual, Installment on Curating Emails: a lifecycle approach to the management and preservation of email messages, (2006) *http://www.dcc.ac.uk/resource/curation-manual/chapters/curating-e-mails/curating-e-mails.pdf*

[5] Loughborough University Email information and documentation (2006-7) *http://www.lboro.ac.uk/computing/mail/projects.html* (Some of the documents listed are only available to Loughborough staff and students – if you would like copies, please contact the authors of this paper.)

[5.1] Loughborough University, Staff E-mail Storage Policy (2006) *http://www.lboro.ac.uk/computing/email/email-storage-policy.html*

[5.2] Results of the E-mail Survey, Loughborough University (2006) *http://www.lboro.ac.uk/computing/mail/projects-survey-results.html*

[5.3] Introduction to New E-mail Storage (2007) *http://www.lboro.ac.uk/computing/email/intro-email-storage.html*

[5.4] Email Storage FAQ (2007) *http://www.lboro.ac.uk/computing/email/email-storage-faq.html*

[6] CommuniGate Systems Website, *http://www.stalker.com*

[7] Loughborough University's retention schedule for records containing personal data (2003), *http://www.lboro.ac.uk/admin/ar/policy/dpact/rec-ret.htm*