Dissertations

School of Computing

2008-01-01

Knowledge sharing: using searchable email databases

Frank Wedgeworth
Dublin Institute of Technology, brendan.tierney@dit.ie

Knowledge Sharing

Using Searchable Email Databases

Frank Wedgeworth

A dissertation submitted in partial fulfilment of the requirements of

Dublin Institute of Technology for the degree of

M.Sc. in Computing (Knowledge Management)

August 2008

I certify that this dissertation which I now submit for examination for the award of MSc in Computing (Knowledge Management), is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the test of my work.

This dissertation was prepared according to the regulations for postgraduate study of the Dublin Institute of Technology and has not been submitted in whole or part for an award in any other Institute or University.

The work reported on in this dissertation conforms to the principles and requirements of the Institute's guidelines for ethics in research.

Signed: _____

Date:

ABSTRACT

In today's knowledge driven economy, a company's intellectual capital is increasingly becoming its most important asset. The knowledge of how to create value defines a company's success. Through knowledge management the value of knowledge within a company can be increased. One way of increasing the value of knowledge is by making it more accessible. The accessibility of knowledge can be facilitated by integrating the search for knowledge into the user's workflow. Another way to increase the value of knowledge is through the capture of undocumented, tacit knowledge and converting it into explicit, documented knowledge.

Email has been identified as the "killer app" and as a "habitat" for users. It is thus integrated into the user's workflow. Due to its combination of conversation and deliberative nature, email is suitable for capturing tacit knowledge. Repositories grow from interactive applications as a by-product of interaction and collaboration.

An investigation into knowledge sharing through the use of a shared email archive in a manufacturing plant is described. Users have been found to habitually store and search their email, confirming that email archives are being used as personal knowledge repositories. Making the information in these repositories available to be shared amongst a wider group was found to be technically feasible and clear benefits were identified. Advantages identified included reuse of information and the capture of resolution to problems. Disadvantages were identified concerning issues of privacy and confidentiality.

Key words: *email, knowledge, privacy, sharing, attachments*

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my supervisor Richard Lawlor for all his help and invaluable guidance during the course of this dissertation. I would also like to express my gratitude for the input provided by Deirdre Lawless at critical points. I would like to thank my managers Noel McShera and Ciaran Murray for their assistance and support during my studies.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

1   INTRODUCTION

1.1 Introduction

This research proposed to investigate the usefulness of a knowledge resource based on archived email that has been communicated between members of groups, involved with the development and manufacturing of analysers e.g. Manufacturing Engineers, R&D and Quality Assurance.

Evidence was collected to ascertain the extent that people make use of previously sent emails as a source of information and the usefulness of that information.

The research aimed to investigate the following aspects:

- What types of information do people search for?
- How far back in time do they search for information?
- How do they organise the email they receive and send?
- What type of search criteria do they use – do they search by keyword? Do they search for messages from particular people?
- Are there particular features that they would like to have available when searching?

Issues concerned with the use of a shared email archive were examined by investigating such aspects as:

- Do users consider that a shared email archive would contain information that they could make use of?
- What are users' opinions about allowing emails they send being made available to users who were not included in the original distribution?
- What are their concerns about privacy and access control?

Based on the outcome of the investigations above, the creation of a knowledge resource consisting of archived emails was investigated, and a prototype developed and evaluated.

1.2 Background

Knowledge creation, capture, sharing and dissemination are the areas of knowledge management of primary interest for this project.

Gartner Inc (2004), in their glossary of frequently used IT terms, defines KM (knowledge management) as "A business process that formalizes the management and use of an enterprise's intellectual assets. KM promotes a collaborative and integrative approach to the creation, capture, organization, access and use of information assets, including the tacit, uncaptured knowledge of people".

Knowledge management has also been defined as "the process of capturing and sharing a community's collective expertise to fulfil its mission." (Burk, 1999).

The organisation described in this investigation manufactures and sells analysers. Problems encountered during the development and manufacturing of analysers are recorded in databases, along with the details of the resolutions of these problems. During investigations of problems, information is transmitted via email. The various departments involved in design and manufacture are located across multiple sites in Europe and the U.S.A. As face-to-face interaction with a colleague or expert is not always possible during an investigation, this increases the likelihood that information will be transmitted by email, thus increasing the probability of transferring tacit knowledge into explicit knowledge. The need to respond to an email prompts the creation of explicit knowledge from tacit knowledge and is likely to occur due to the less onerous requirements of creating emails in terms of time, content and formatting, compared to more formal documentation.

Knowledge creation can occur through email, as information is transmitted back and forth in a question and answer format, gradually been refined as conclusions are reached

and answers provided. The essential information contained within these emails will make its way into the problem reporting databases and official documentation. However, it is proposed that being able to search through the emails communicated during the investigations may provide useful additional information on the problems that could augment more formal documentation.

The information contained within an email message is only immediately available to those to whom it is addressed. At a particular point in time, this group is the only one interested in that piece of information. Over time, the members of the group may change due to retirement, promotions or the members leaving the organisation, or the group may be extended to include members of other departments.

If email already sent is available in a shared archive, the information contained in those emails would then be available to the new members of the group. This would be a means of knowledge transfer and knowledge dissemination.

The benefits of making this information available could be increased by storing and presenting it in a structured, searchable format. The archive could be made available to other functions within the organisation, so that knowledge gained by one group could be of use to other areas.

1.3 Research problem

Project Objectives:

1. Identify the knowledge management requirements for the members of groups involved in the development and manufacturing of analysers.
2. Analyse the knowledge in email and its usage by members of groups involved in analyser development and design.
3. Based on analysis, develop a set of requirements for a knowledge resource to support Knowledge Management in this area.

4. Implement a prototype knowledge resource and evaluate its usefulness in practice for particular KM purposes.

5. Develop a plan for extending the knowledge repository based on investigation and experimentation.

1.4 Intellectual challenge

Emails are communications between a specific set of individuals, those that are specified by the sender when addressing the email.

Two main intellectual challenges were identified

1. Can the information communicated in emails be aggregated into a knowledge resource?
2. Can this knowledge then be made available to a wider group than those included on the address list of the original email?

1.5 Research objectives

The following objectives have been achieved throughout the dissertation and contributed to the overall outcome:

1. Literature review was conducted to identify
   - where use of email as a knowledge management tool would fit into existing models
   - whether email repositories contain knowledge
   - the retrieval of information from email
   - the use of email to share knowledge and identify expertise

2. Analysis of email usage among engineers and more specifically their use of email archives to search for knowledge

3. The analysis of the willingness of a group of engineers to share knowledge using a common email repository

4. Investigated the technical feasibility of creating a common email archive

5. Outlined a prototype that could be used to share knowledge through a common repository

1.6 Research methodology

During the investigations conducted in this dissertation both primary and secondary research methods were used:

- Primary research consisted of interviews and surveys. Both unstructured and semi-structured interviews were conducted. Unstructured interviews were held with two key stakeholders to explore the feasibility of the use of an email archive to share knowledge. The focus of the unstructured interview was to identify benefits of the proposed system, technical feasibility and probable costing. Semi-structured interviews were conducted with a cross section of potential users. The objectives of the semi-structured interviews were to identify the usage of email, exploring whether users archived email, how they archived email, how they searched and the type of knowledge they considered to be available in email archives. As part of the semi-structured interviews issues of privacy and sharing were discussed and examined. Based on the feedback from the semi-structured interviews a survey was distributed to a wider group of potential users. The survey contained both quantitative and qualitative questions. The nature of the questions elicited subjective responses.

- Secondary research consisted of literature review. A variety of sources on existing research into the use of email as a knowledge management tool was conducted including the following:
  - Journals

- Conference papers
- Books
- EU website on data protection
- Irish Data Protection Commissioner's website
- Company, Microsoft and IBM Lotus Notes websites

2   LITERATURE REVIEW

2.1 Introduction

The following chapter contains a review of literature associated with email, focusing on Knowledge Management issues.

Where email fits into an existing model of knowledge management was examined, along with issues such as locating expertise, knowledge sharing and the elicitation of tacit knowledge. An examination of the way people archive and retrieve email was also covered.

Several authors identify the importance of email. Bob Kahn, one of the pioneers of the Internet, which was developed for resource sharing, is quoted as saying in 1972 "..everyone really uses this thing for electronic mail" (Hafner K, Lyon M,1996, p. 186). A report on the completion of the Arpanet research program in the late 1970s stated: "The largest single surprise of the ARPANET program has been the incredible popularity and success of network mail" (Hafner K, Lyon M, 1996, p. 214).

One issue identified with email in the literature is that it is being used for tasks for which it was not intended – e.g. task management, project management, information exchange, scheduling and social communication. Whittaker & Sidner (1996) referred to this as *Email Overloading*. For this study, the focus was on the information contained in email and not on task management or scheduling.

2.2   Where Email repositories fit in to other models

Email repositories appear to fit what Zack (1999) terms as Interactive Applications. He describes these as focused primarily on supporting interaction among people holding tacit knowledge. Repositories grow from interactive applications as a by-product of interaction and collaboration among users. Email would fit into his description of "forums", interaction among those performing common practices or tasks, which tends to

be more ad hoc or emergent. Zack (1999) states: "The producers and consumers comprise the same group of people, continually responding to and building on each individual's additions to the discussion. The flow continually loops back from presentation to acquisition". This process matches the to and fro of email conversations and backs up the contention that knowledge can be created within email. Zack (1999) suggests that with the appropriate structuring and indexing of the content, a knowledge repository can emerge. From the point of view of this study, an archive of email may already be in this form – a structure exists based on time, sender and subject.

2.3 Do Email Repositories Contain Knowledge?

Lichtenstein (2004) confirms that knowledge can be created by a process of refinement during the course of an email exchange. She describes a knowledge creation life cycle consisting of five underlying processes – initiation, crystallisation, qualification and combination. Conversations carried out in emails were observed to follow a loop of qualification and combination, resulting in crystallisation and sharing and the creation of new knowledge. Knowledge sharing is described as taking place during and as a by-product of the development of the knowledge itself, during the transmission of emails. This knowledge will be shared among the micro-community of people who were involved in the email trail. However, no mention is made in this paper on the benefit of capturing this knowledge for future use, perhaps by members outside the micro-community.

Whitaker & Sidner (1996) also described this pattern when they provide reasons for archiving email. Issues may take several email exchanges to resolve, or require the responses of several individuals to reach a consensus.

Schirmer (2003), in research conducted for the development of the Lotus Discovery server, found that for some users, their email, rather than the enterprise's public data, represented most of their enterprise's knowledge.

Ducheneaut & Bellotti (2001) found that almost all the subjects in their study used email to exchange documents. This would imply that searching email would return the information within these documents. Perkiö et al. (2005) agree that email archives can contain knowledge where they state: "Considering the sheer amount of information in these archives…the current interfaces can be considered sub optimal for harnessing all the buried knowledge". They identify the information in email as multifaceted - it has a textual, temporal and a social dimension.

On the other hand, Sorensen & Gibson (2004) in their survey of 16 professionals found mixed views on email – one considered that it contained only data, while another considered that it contained important information. Another considered it a means of CYA – "Cover your Ass". This implies a poor opinion of the content of email – but at the same time it indicates the importance of being able to refer back to that information.

2.4 Knowledge Sharing and Identifying Expertise

Newman (2002) described extracting social information from email as probably most useful in identifying the linkages between individuals, and the kinds of issues they discuss. This could help new group members to identify the people involved, and to identify groups working on particular topics. In terms of knowledge management, this ties in with the identification of experts who hold knowledge and identifying communities of practice. Newman (2002) also points out that isolating these groups is a major technical challenge, and awareness among users that it is being performed may discourage participation. This is probably more true of the public email lists discussed by the author rather than internal company email, where the participants would be contactable and the range of topics more focused.

Leuski (2004) describes how pattern analysis of speech acts in email allows the identification of participants' roles, thus identifying the relationship between people. An example of a relationship is given as: *who started a project* and *who brought it to completion*. From the point of view of knowledge management this seems applicable to the identification of expertise. Similarly, Viégas et al. (2006) hypotheses that the patterns

of communication between people are significant and that these patterns can be identified from email using data mining. The patterns identify the interaction among individuals. Although the primary use of their research seemed to be outside the work environment, it could be useful for determining the communication flows, and therefore the knowledge flows within an organisation.

Email has been also been identified as a resource for Social Network Analysis (Zhou et al. 2005), (Zhang et al. 2006). Identification of leadership roles within an organisation, from analysis of the conversation history between people that is logged in email is described by Zhou et al. (2005). Zhang et al. (2006) in describing the availability of the Enron Corpus of emails made available by the US government, suggest that it "provides a promising resource for research on human interactions, and for discovery of the hidden patterns of collaboration and relationships in communities". They describe a method of graphically representing these interactions. In the context of Knowledge Management, this would be useful in identifying Communities of Practice, identification of experts, knowledge sources and knowledge flow.

Such a use is described in Gloor et al. (2003). They describe collaborative groups, COINS, defined as "Collaborative Innovation Networks (COINs) are groups of self motivated individuals from various parts of an organization or from multiple organizations, empowered by the Internet, who work together on a new idea, driven by a common vision." Gloor et al. (2003) describe a project to identify such groups through analysis of email, using visualization techniques. Their conclusion lists the following advantages of locating COINS:

1. *By locating COINs, organizations can learn about innovations which are underway. This enables them to spot hidden business opportunities and also cut the time to market for new inventions.*

Support for Innovation is a key goal of Knowledge Management as identified by Davenport & Prusak (1998) and Nonaka & Takeuchi (1995).

2. *By supporting hidden COINs and making them transparent, organizations can become more efficient in working together. They can better identify their knowledge sources and streamline communication processes.*

This is obviously of benefit to Knowledge Management.

3. *Because key contributors can be identified through transparent COINs, organizations have a better chance to identify and reward leaders and important collaborators.*

This identifies one solution to a problem in knowledge management – encouraging collaboration and knowledge sharing.

4. *By making the communication flow transparent, a more open working environment can be created, generating additional trust among its members.*

Openness and trust are key requirement identified by Zack (1999) for a culture of knowledge management – he advocates an organizational climate and reward system that values and encourages cooperation, trust, learning, and innovation and provides incentives for engaging in those knowledge-based roles.

The concept of COINs is similar to "Community of Practice", a term used by Lave and Wenger (1991). Community of practice are discussed by Ardichvilli et al. (2003). They investigated the reasons why people contribute to knowledge sharing communities. Based on a study of an online knowledge network in Caterpillar Inc, they found that the majority of respondents were willing to share knowledge. One reason given by respondents was that it was in the nature of engineers to share knowledge. Another reason was the need to establish themselves as experts. A barrier to participation was identified by Ardichvilli et al. (2003) as the preference for people to rely on their contacts rather than the knowledge network. Other issues identified was "fear to lose face" if the information was not completely accurate or relevant. The motivations to contribute identified would be expected to apply to email based knowledge management system among engineers. An email repository would also be more likely to capture the information sent directly between individuals. The "fear to lose face" issue may also affect communication. The fact that an email may reach a wider audience than originally

intended by being placed in a repository could inhibit some engineers from responding to a request for information.

2.5 Tacit Knowledge Extraction From Email

Zack (1999) states that tacit knowledge is usually shared through highly interactive conversation, story-telling and shared experience. That email is similar to conversation is mentioned by a number of papers.

Newman (2002) describes email threads as "Persistent Conversations", although they are unlike conventional conversations where people take turns to speak – simultaneous replies may be received to the initial message. From the point of view of using email as a knowledge repository this aspect may present problems with trying to analyse responses.

Sands (2003) also agrees with the usefulness of email to impart information.

"The most successful form of persuasion and information transfer takes place in dialogue or duplex form rather than simplex transactions". He contrasts the dialogue form of email with the Web, which he states has up to recently been simplex. He gives this one way communication the term "Jug and Mug". He uses the term "considered spontaneity" for email communication. This term ties in with Lichtenstein (2004) characterization of email responses as being reflective, and therefore more valuable. Sands (2003) also states that both sender and receiver are likely to use less formality, informality often being a characteristic of tacit knowledge.

Whittaker & Sidner (1996) and Kerr (2003) suggest that the last email in an exchange often helps determine *conversational status,* by summing up the current state of the conversation or by containing questions or tasks that are still outstanding. In a knowledge management context, analysis of these open questions could be useful in identifying knowledge gaps.

The similarity of email to conversation is also supported by Goldstein & Evans Sabin (2006) in their investigation on classifying the communicative intent of email, where they conclude that "we believe our findings support the characterization of email as an

amalgam of unique communicative genres, where the common genre – email conversations is most similar to spoken communication". They identified genres they referred to as Email Conversations.

- **Responses -** Provide information in response to a question.
- **Response with forward function -** Provide information in response to a question and ask questions.
- **Information request -** Asks for information.

This provides evidence that an email archive would be a repository of email conversations, and would be likely to accumulate tacit knowledge in an explicit form.
This is backed up by Ruggeri Stevens & McElhill (2000), "It will often be true that purposive use of e-mail offers a medium by which individual tacit knowledge can be transferred to groups".

2.6 Problems With Email As A Collaborative Tool

Whittaker et al. (2004), states that email and voice mail, because they are message centric, do not support the social reminding (that you have undertook to do something for somebody) and social data mining processes (trying to find the name of a contact) that are a natural by-product of physical proximity. They describe *ContactMap*, a user interface showing photographs rather than a list of names, which is more effective for social reminding. From the point of view of the present study, such a concept could be a useful addition to a user interface.

Heer et al. (2007) describes the benefits of visualizations in analysing and creating new knowledge, referred to as *Sensemaking*. However they state that sensemaking has a social aspect as people have different interpretations of visual data and collaboration is required, allowing others to learn from their peers as consensus is achieved. They describe a scenario where visual data (graphs) were shared through email. This leads to

problems due to the scattered and disconnected discussion of the visualizations. They suggest a better way of asynchronously sharing visual data using a web page displaying the visualization along with a discussion box and allowing annotations. However, this limitation of email does not take from its value as a means of capturing knowledge. If the emails are archived, the visual data can be perused again.

Viégas et al. (2006) also point out this limitation of email in comparison to physical proximity. Email does not have the sensory detail available in face to face communication. They describe the archives of online interactions as "little more than white noise". This seems an extreme opinion, as simply sorting by user name or glancing at the subject lines will provide information on who a person has been talking to and the general areas of discussion. However, given their expectation that it seems more likely that users might use their system (*Themail*) to reminisce, rather than for work related issues, it does point to the personal aspects of visualizing relationships and provide a user friendly way of located expertise.

2.7 Why Do People Archive Email?

Evidence that users archive email is described by Dabbish et al. (2005), where they found from their study that an overwhelming majority of messages were retained. Venolia et al. (2001) found that archiving messages was a very common activity. Users wanted to use the information contained in the emails again. Whittaker & Sidner (1996) also confirmed that people archive email. One of their contributors involved in Service and Support provided a relevant quote: "*It's not only everything that's being said... It's every person that has been involved"*. This suggests a possible advantage to the use of email – in a lot of circumstances people are compelled to document their responses and share and store the details – identified in Whittaker & Sidner as "Cover Your Ass", and also referred to in Sorensen & Gibson (2004). An update on Whittaker & Sidner (1996) paper by Fisher et al. (2006) found that 50% of the participants in the study had email more than 4 ½ years old and 25% had email more than 6 ½ years old. Pagel et al. (2004) stated that 40% of users kept all their emails.

In a study conducted by Mariano & Casey (2007), 27 percent of those interviewed mentioned email folders as repositories to store and retrieve explicit knowledge. A quote: *"Sometimes people need to know something and they go get it through the email and eventually they'll have an answer through the email"* backs up the view that the accumulation of answers stored in email archives can serve as a source of knowledge.

2.8 How People Archive Email

Whittaker & Sidner (1996) identified several problems with filing emails – it's difficult to decide what folder a message could be filed in, difficulty in finding the messages again, some folders having too few items and others being so large that the purpose of filing was negated. They described three types of filer - *Frequent, No Filer and Spring Cleaner*. They noted that No Filers tended to be those with the greatest number of email messages, such as managers.

These categories of Frequent, No Filer and Spring Cleaner were backed up by Dabbish et al. (2005). The high proportion of messages people retain suggested to them that technology to aid in the location and viewing of messages is an important area of future research for electronic mail. They found that people had difficulty in finding the messages they needed.

Bälter (2000) points out that by leaving messages in the inbox, time spent moving messages to other folders is eliminated, but the time required for searching will increase if there are a large number of messages. If too many folders were used, the user spends longer trying to find the right folder - more than 30 folders were deemed to be inefficient. The most efficient strategy for users was to use no folders (everything in the inbox) but that this raises demands on search tools. Advantages are identified for the use of folders. They provide users with the context of other related messages, and may be used to group messages that are difficult to search for with a tool, but still must be read together. Also, all items in a folder can be deleted at the same time if necessary.

Ducheneaut & Bellotti (2001) found that experienced users used more folders. They suggested that this was because experienced users had built up a number of folders over time, some of which referred to old projects. They suggested that the limit of 30 folders found by Bälter (2000) did not take into account non-active folders. They found that people used sort more than search.

Venolia et al. (2001) found in their research that the average number of folders was 104, and that people did not seem to have any trouble selecting folders for particular messages. They noted that Outlook imposes minimum of 2 folders – Inbox and Sent, which may make following threads difficult as they may need to be followed across folders. They described Microsoft Outlook functionality Categories, but found that few users were aware of it.

The numbers of folders required has relevance to the present study, and whether the preferences for folders expressed by single users could scale up to multiple users and projects.

2.9   Retrieval Of Information From Email

Research into retrieval from email archives has been stimulated recently by the availability of 500,000 emails from the Enron corporation, provided for public access by the US government (Kalman & Rafaeli 2005), (Zhou et al 2005), (Zhang et al. 2006), Perkiö et al.(2005) profile a system for search and analysis of large scale email archives. They identify the information in email as multifaceted - it has a textual, temporal and a social dimension. With regard to textual/content searching, they identified two desirable criteria: (1) the user would not be required to have exact matches of query words, and (2) the ability to support queries containing only a few words, what they termed lazy queries. Topical trends and social networks were identified and visualization was used to present them to the user. They suggest that visualization facets help reduce the burden of exploring the large amount of emails returned by the search engine.

Aery & Chakravarthy (2005) in their review of related work, identified three broad categories of classification for email.

- Rule based classification to classify emails into folders.
- Information Retrieval based classification.
- Machine Learning based classification techniques – e.g. Naive Bayes for training.

In the system they present, *emailSift*, they suggest another method of classification - using the combination of header, subject and content to identify the structure of the email for classifying it into a particular folder.

Goldstein & Evans Sabin (2006) investigated ways of characterizing an email by its most important speech acts in terms of the intended action of the sender and expected action on the part of the receiver. One suggested use for this categorization is to track responses to the user's requests for information or action. They state that in emails, responses often contain a mixture of speech acts e.g. answers and comments to a sender's email as well as additional questions for the sender. They identified characteristics of email that indicate its type. Examining their characteristics, the following seem useful for detecting knowledge

- Presence of Re:
- Presence of Fwd:
- Attachment signified in header info or by an insertion in text body
- Fraction of interrogative sentences (sentences ending in '?'/total sent)
- Attachment indicators such as "attached, here is, enclosed"
- Opinion indicators: "think, feel, believe, opinion, think, comment"
- Information indicators such as "information", "info", "send"

They found that a combination of classifiers using a verb lexicon and email characteristic gave reasonable classification performance. This research would seem to indicate that the identifying knowledge within an email archive is feasible.

Newman (2002) describes the organising of email using subject header and the first few lines of their initial messages – this allows users to determine if the topic is of interest. Also, by listing the number of messages in the thread the user can determine which threads were "hot". An alternative is the use of a subject index, generated by extracting words from the subject lines whose lemmatised forms are not found in the list of 5000 or so of the most frequent words in English. The approach was found to be effective because subject lines usually capture the major issue being raised, and the terms used tend to be specialized ones.

Problems with the structure of information in email were identified by Whittaker & Sidner (1996). There is no convention on whether to include prior messages during an exchange of messages while an issue is being worked out. These prior messages can help provide important context. In addition, the thread of messages exchanged may be interleaved with unrelated information. Both these issues are of significance to using email as a knowledge repository. The repository is likely to be searched when the original context is no longer fresh in peoples minds or the search is being performed by a new member of the group - trying to extract relevant information will be made more difficult by the lack of all the messages in the thread and the interleaving of unrelated messages.

Newman (2002) points out other difficulties with searching email archives – unlike formal publications, they contain a very large number of components (conversations) and the components are not naturally organised into meaningful groupings. Another problem identified is that although email threads are often called "persistent conversations", they are unlike conventional conversations where people take turns to speak – simultaneous replies may be received to the initial message, in turn prompting multiple sub conversations.

2.10 Conclusion

This chapter examined the use of email as a knowledge resource in existing literature.

Email as a knowledge resource appears to fit into what Zack (1999) terms as Interactive Applications Email and would fit into his description of forums. That knowledge is contained in email archives was confirmed by several authors, where they describe a process of exchanging email to resolve an issue, or require the responses of several individuals to reach a consensus.

The identification of expertise through email has been explored in this literature review. Of particular interest is the identification of Communities of Practice by analysis of email communication. A barrier to participating in knowledge sharing was identified as peoples reliance on their contacts rather than consulting a knowledge resource. Engineers were identified as a group that were willing to share knowledge.

Email similarity to conversation – it has been described as "Persistent conversation"- is useful for the transfer of tacit knowledge. Problems with email as a collaborative tool were identified as lack of visualisation and the lack of sensory queues due to the absence of physical proximity.

The two main reasons why people stored their email were identified as the information contained and traceability (CYA).

Studies have found that people can be divided up into Frequent Filers, Spring Cleaners and No Filers. Although the optimum strategy for filing is to use a single folder, other studies have found that people prefer to use folders.

Information can be extracted from email through classification and visualization techniques.

3   PRIVACY ISSUES

3.1 Introduction

This study investigates the use of email as a source of knowledge. The area under investigation concerns email sent by engineers using the organisation's equipment and during the course of their work. Part of the investigation is to determine the level of access that an employer has to employee email. This chapter will examine the issue of Privacy with regards to using it as a knowledge resource.

3.2 Working Party Document 2002

The Working Party On The Protection Of Individuals With Regard To The Processing Of Personal Data adopted working document WP55 on 29 May 2002. The full title was "Working document on the surveillance of electronic communications in the workplace" 5401/01/EN/Final WP 55. The purpose of this document was to complement the work of the Data Protection Directive 95/46/EC. This working document offers guidance and concrete examples about what constitutes legitimate monitoring activities and the acceptable limits of workers' surveillance by the employer. It is focused on surveillance, but for the purpose of this research it is considered a useful guide to the extent that email can be examined by the employer. In particular, it will help clarify the question whether the fact that the employer owns the equipment that generated the mail, allows them to have complete authority to do whatever they want with it.

Although emails archives are being examined for the purpose of knowledge capture and sharing, such an archive could possibly be used for monitoring, so these issues will have to be taken into consideration.

The working document states on p. 4 "Workers do not abandon their right to privacy and data protection every morning at the doors of the workplace". It argues that as workers develop relationships with others in the workplace, they have a legitimate expectation of privacy. However, this right must be balanced with other legitimate rights and interests of the employer – the right to run his business efficiently and above all to protect him from the harm that workers actions may create. The clearest example is where the employer is victim to a worker's criminal offence.

The following questions summarise the recommended assessment:
a) Is the monitoring activity transparent to the workers?
b) Is it necessary? Could not the employer obtain the same result with traditional methods of supervision?
c) Is the processing of personal data proposed fair to the workers?
d) Is it proportionate to the concerns that it tries to ally?

The Working Party opinion is that prevention should be more important than detection, that the interest of the employer is better served in preventing misuse rather than in detecting such misuse.
They suggest that employers may consider providing workers with two emails accounts:
a) one for only professional purposes, in which monitoring within the limits of this working document would be possible,
b) another account only for purely private purposes (or authorisation for the use of webmail), which would only be subject to security measures and would be checked for abuse in exceptional cases. This recommendation is also suggested by Pagell et al. (2004) when they recommend that employees: "Get a second, free web-based e-mail account for personal mail and use that address when you need to register for sites. This will not only keep your personal mail separate from work but will reduce the amount of junk coming to your business account."

The Working Party paper emphasised that the conditions of work have evolved so that it has become more difficult to clearly separate work hours from private life for example

people working from home. It identifies the relevant international legal instruments – Articles 8 and 10 of the European Convention for the Protection of Human Rights and Fundamental Freedoms. These articles can be summarised as:

Article 8: Everyone has the right to respect for his private and family life, his home and correspondence.

Article 10: Everyone has the right to freedom of expression

In the judgements given to date, the European Court has made it clear that the protection of "private life" enshrined in Article 8 does not exclude the professional life as a worker and is not limited to life within home. In the case of *Niemitz v. Germany* the European court rejected the argument that the professional and business life were not covered by Article 8. A quote from the judgement is relevant in showing how business and private life are not seen as mutually exclusive:

"There appears, furthermore, to be no reason of principle why this understanding of the notion of "private life" should be taken to exclude activities of a professional or business nature since it is, after all, in the course of their working lives that the majority of people have a significant, if not the greatest, opportunity of developing relationships with the outside world. This view is supported by the fact that, as was rightly pointed out by the Commission, it is not always possible to distinguish clearly which of an individual's activities form part of his professional or business life and which do not".

In the case of *Halford v. the United Kingdom* the Court decided that interception of workers' phone calls at work constituted a violation of Article 8 of the Convention. The court considered that Ms Halford would have a reasonable expectation for privacy for calls made from work, as there was no evidence that she had been warned that her calls were liable to interception. *Halford v. the United Kingdom* could be interpreted that advance warning to the employee about surveillance may make it allowable However the Working Party were not of that opinion. From examination of the case law the Working Party concluded: "The general principle of secrecy of correspondence covers communications at the workplace. This is likely to include electronic e-mail and related files attached thereto".

It is difficult to draw a firm conclusion from this document. It seems to indicate that employee's communication at work should be considered private – yet the employer must have some monitoring to protect themselves.

3.3 International Labour Office

In the general principles of the International Labour Office Code of Practice on Protection of Workers' Personal Data (1997), the following is considered relevant to the introduction of a common pool of email communication. Although the purpose of the proposed system is not to monitor workers, it does indicate that care may need to be taken in implementing such as system, and the importance of removing any personal data from such a system.

12.2. The workers' representatives, where they exist, and in conformity with national law and practice, should be informed and consulted:

a) concerning the introduction or modification of automated systems that process worker's personal data,

b) before the introduction of any electronic monitoring of workers' behaviour in the workplace

c) about the purpose, contents and the manner of administering and interpreting any questionnaires and tests concerning the personal data of the workers.

3.4 European Directive on Data Protection

The European Directive 95/46/EC concerns the processing of personal data.

Personal data was defined as data that could identify an individual?

Controllers must have a legitimate ground for processing personal data.

Data subject must be fully informed about personal data that is stored and their consent obtained to process this data. However, controllers should not rely on consent as a general means of legitimising such processing. Reliance on consent should be confined

to cases where the data subject has a genuine free choice and is subsequently able to withdraw the consent without detriment.

Does employees email contain personal information? Email will allow a user to be identified and will contain the users name email address and possibly telephone numbers. This information is similar to the information that could be obtained in a telephone book. Processing of any potential private data within email would appear to be covered in Article 7 (f) of the Directive, where it describes the scenario where processing is necessary for the purposes of the legitimate interest pursued by the controller or by the third party or parties to whom the data are disclosed.

A user who sends an email to a business is providing the information freely for a particular purpose. The organisation has a legitimate right to process this data for that purpose. What would contravene the principle of data protection would be for the organisation to use in for purposes other than the reason it was sent – for example, by mining emails to obtain addresses that would be used in a telemarketing campaign.

The purpose of the system being proposed for this dissertation is not to monitor employee's behaviour, but to extract and share knowledge. However, it has to be conceded that giving a supervisor the ability to see others email could enable monitoring.

3.5 Situation in Ireland

According to the website of the Data Protection Commissioners (www.dataprotection.ie), organisations have a legitimate interest to protect their business, reputation, resources and equipment, and monitoring of staff's use of email may be required to achieve this.

Basing their approach on WP55, it states: "Any limitation of the employee's right to privacy should be proportionate to the likely damage to the employer's legitimate interests. An acceptable usage policy should be adopted reflecting this balance and employees should be notified of the nature, extent and purposes of the monitoring specified in the policy".

It is advised:

- A balance is required between the legitimate rights of employers and the personal privacy rights of employees

- Any monitoring activity should be transparent to workers

- Employers should consider whether they would obtain the same results with traditional measures of supervision

- Monitoring should be fair and proportionate with prevention being more important than detection

3.6 Situation in the United Kingdom

Despite the concerns expressed in the Working Paper, monitoring of employee mail is allowed under the law of various countries. For instance it is allowed in the UK based on the Telecommunications (Lawful Business Practice) (Interception of Communications) Regulations 2000.

The interception has to be by or with the consent of a person carrying on a business for purposes relevant to that person's business and using that business's own telecommunication system.

The following is a summary of conditions where interceptions are authorised:

- To establish the existence of facts, to ascertain compliance with regulatory or self-regulatory practices or procedures or to ascertain or demonstrate standards which are or ought to be achieved (quality control and training).

- In the interests of national security (in which case only certain specified public officials may make the interception).

- To prevent or detect crime.

- To investigate or detect unauthorised use of telecommunication systems.

- To secure, or as an inherent part of, effective system operation.

- Monitoring received communications to determine whether they are business or personal communications.

- Monitoring communications made to anonymous telephone help lines.

- Interceptions are authorised only if the controller of the telecommunications system on which they are effected has made all reasonable efforts to inform potential users that interceptions may be made.

Examples where monitoring would be used is in financial institutions that must attach disclaimers to correspondence – monitoring of email would be performed to ensure that the employee was attaching disclaimers as instructed. The condition should be noted that the controller must inform users that interceptions may be made.

3.7 Situation in the United States

Under the US constitution, people have a right to a reasonable expectation of privacy (*Katz v. United States*, 1967). Several cases, e.g. *Smyth v. The Pillsbury Co,* 1996, *Bohach v. Reno,* 1996, *Bourke v. Nissan Motor Corp,* 1993, *Shoars v. Epson America Inc*. 1994 indicated that employees did not have a reasonable expectation of privacy for emails sent using their employers system. However, *Quon v. Arch Wireless,* 2008 indicated that a user did have a reasonable expectation of privacy for messages sent

electronically. In that case, the organisation did have a policy allowing monitoring. The expectation of privacy was based on the fact that the user's supervisor told the users he would not monitor their text communications. If an organisation has made users aware that monitoring may take place, and that this monitoring is being performed, then under these circumstances the user has no reasonable expectation of privacy.

Rasch (2006) notes that, although many companies have policies that state that the employee must agree to allow all their communication to be monitored, the reality is not so clear cut. There may be a disconnect between what company policy says, and what is actually done. If monitoring is part of the policy, but employees are aware that it is not actually being performed, this can be interpreted by the courts as a reasonable expectation of privacy by users.

Rasch (2006) also notes a disconnect between what employees say is their expectation of privacy, and how they act. Although the employees have agreed that the employer can monitor everything, in reality they will probably be offended if this monitoring affects them.

3.8 Conclusion

Creating a knowledge resource based on a shared email raises privacy issues. Allowing email archives to be searched by an employee's supervisor could be considered monitoring.

The document "Working document on the surveillance of electronic communications in the workplace" issued by the European Commission Working Party On The Protection Of Individuals With Regard To The Processing Of Personal Data declares: "Workers do not abandon their right to privacy and data protection every morning at the doors of the workplace". Judgements given by the European Court of Human rights would indicate that employees have an expectation of privacy in their communication at work. However, this is balanced by the needs of the employers to protect themselves from harm caused by abuse of the organisation communication system. In Ireland and the UK, the employer must inform users that such monitoring is taking place. Although the Working Party does not agree that informing users is all that is required to allow monitoring.

The US seems to be moving to the same conditions – users do not have a reasonable expectation of privacy if they are made aware that this is the organisation's policy to monitor email.

However, although employees may be aware that monitoring is taking place, they may be still unhappy with the situation. Both employees and employers need to be aware of the issues. This has implications for this investigation in that it would suggest that an approach based on volunteering information may be best. Monitoring of an employee's emails can take place, once an organisation has made it clear to the employee that it is the company policy, and that the employee is aware that the policy is enforced.

Regarding Data Protection – the content of personal data sent through work related email should be low. Details such as a persons name and email address would probably be available in a telephone book. Email sent from users outside the company would have been sent voluntarily, and any processing and storing of email would be for the legitimate reasons relating to the reason the email was sent. The email archive could not be used for other purpose, such as obtaining addresses for telemarketing.

4 EMAIL USE WITHIN THE ORGANISATION

4.1 Introduction

The focus of this chapter is on an investigation of the possibilities of using email as a knowledge repository in a manufacturing organisation.
It includes:

- Identification of key stakeholders
- Feasibility
- Requirements gathering
- Proposed system

4.2 Description of the organisation

The study was based in a manufacturing plant in Ireland, which is part of a multi-national company. The engineering department consists of approximately 27 engineers, with a mixture of software, mechanical and electronic specialists.

The engineers look after production of two product lines, each of which has two main models. In addition, older products need to be supported in order to continue shipping spare parts. The organisation contains a Quality Assurance Department. There is a local IT department. Recently the company have moved from using Lotus Notes to Microsoft Outlook as the company's email application.

4.3 Key stakeholders

The key stakeholders were identified as:

(1) The Engineering manager
(2) Email administrator
(3) Engineers

4.4 Feasibility

In order to determine if the project was feasible, interviews were conducted with the Engineering manager and the Email administrator.

Interview with the Engineering Manager

An unstructured interview was conducted with the Engineering manager over 2 sessions. The goal of the interview was to determine if the proposed system would provide benefits to the organisation.

The outcome of the interviews can be summarized as follows:

(1) With an average length of employment in the engineering group of 18 years, there were approximately 500 man years of knowledge in the department. A system to capture that knowledge would be beneficial.

(2) With regards to email, the reports that were attached to emails were of particular interest. The conclusion of investigations might eventually be included in official documentation including drawing and procedures, but a lot of routine reports would not make it that far.

(3) The different types of email attachments could represent different levels of information

- Slides/PowerPoint = Top level. summarized information – knowledge
- Word = Medium level - Information
- Excel = Low level – Data

(4) Following a chain of emails would reveal the sequence in which data was created. The latest email would be at the top of the hierarchy in terms of importance.

(5) The person to whom the email was addressed indicates the email's place in the information hierarchy. An email to the CEO would be more likely to contain summary analysed information rather than raw data.

(6)   Analysis of the people copied on emails could provide information on who the status holders were in the company.

(7) A problem identified with email was that there was too much of it and that it took too long to organise. Trying to file all the documents that arrived could take up to an hour a day. Simply storing all email in a single folder and using a search tool within the attachments would therefore be of benefit. This would be quicker than searching network drives where documents were stored. In effect, you would convert data into useful knowledge at the time when you need it.

(8)   Suggested scope for the work was to provide a tool or means of viewing, organising and extracting information from the previous 5 years work.

(9) Regarding privacy: Although the company owned the data, a key issue was that personal data should not be returned. This issue could be addressed with password protection and a series of permissions. Human resources could be used to identify sensitive emails, allowing emails with sensitive data to be filtered out or flagged.

(10) Another approach to the privacy issue would be to only allow users to search what they had sent. This would at least provide users with an archive of their own mail.

During a subsequent meeting, the following benefits of knowledge management were identified:

(1) A firm's intellectual capital is increasingly important – the knowledge of how to create value from physical assets, not the assets themselves, are the most valuable.

(2) Few people recognise the value of knowledge. Value is determined by its accessibility i.e. how easy is it to get at this knowledge.

(3) Simply making information more accessible therefore increases its value.

(4) The way that we work determines how we search for information.

(5) When a decision has to be made, a search for information will only be undertaken if it's easy to find.

(6) To make the search for information easier it must fit into the work stream – as email is integral to how we do our work, an efficient search mechanism associated with email would be beneficial.

Interview with the Email Administrator

An unstructured interview was conducted with the email administrator. The goal of the interview was to determine if the proposed system was technically feasible and to try to identify the cost involved.

The outcome of the interviews can be summarized as follows:

(1) It was the policy of the company to store all email messages for a number of years. This storage is to tape and would be difficult and costly to restore these messages. This was important information – if it was desirable to get the last 5 years of data, it would be cheaper to use peoples local archives if they existed. Going forward, users could add email messages as they were generated.

(2) The recommendation for the repository was to create a Lotus Notes database. Lotus notes had been in use for over 10 years and had been recently replaced with Microsoft Outlook. Therefore, most of the archives were in Lotus notes format.

(3) Email is stored in Lotus notes as a database – adding email to a central database could be achieved by cutting and pasting mails from a user's mailbox into the central database. In Lotus Notes attachments are stored in a field of the email record in the database so any attachments would be added automatically when pasting.

(4) Email that has been sent using Microsoft Outlook data could be also be stored in this database. Alternatively, a separate central storage area for Microsoft Outlook could be used.

(5) An advantage that Lotus Notes had over Outlook was that the databases size could be allowed to grow much larger. The organisation had a policy limiting Outlook mailbox sizes to 500MB.

(6) Lotus Notes already has search and indexing capabilities.

(7) The knowledge retrieval functionality of the database could be increased by adding an extra field to the email record that could be used to indicate the category of the message. An example where this could be used would be to indicate that the message contains details of a resolution to an issue so it would have a higher level of knowledge than other messages.

(8) The security features on Lotus notes would be sufficient to configure access. There are 3 layers of security: server, client and database security. Encryption is also available.

(9) Regarding cost – the structure for creating such a database was already available. Some cost would be required for internal development, but it was not thought that this would be excessive. Storage cost should not be excessive because it would be included in the maintenance cost of the server which fits into the existing architecture. This central database would need an administrator but should not be costly as only a small amount of maintenance would be required.

Conclusion of Feasibility Study

- Searching email for knowledge was deemed beneficial.
- Creating a central archive was also technically feasible. Most of the existing email was in Lotus Notes databases, and creating a centralised database using this format was recommended.

- Obtaining email from tape archives was not practical. The possibility of users "donating" from any email stored locally would be investigated.
- The possibility of users adding emails to a central archive as they composed email would be investigated.

4.5 Investigation into use of Email among Engineers

Information was sought from the other key stakeholders – the engineers who would use the proposed system.

- Do users store the email they receive? (This would determine if it was possible to build a central archive using mail "donated" from individual archives)
- What types of knowledge do people search for?
- How far back in time do they search for information?
- How do they organise the email they receive and send?
- What type of search criteria do they consider most useful – do they search by keyword? Do they search for messages from particular people?
- Are there particular features that they would like to have available when searching?
- What type of knowledge do they consider could be obtained from searching emails archives?
- Do users consider that a shared email archive containing would contain information that they could make use of?
- What are users' opinions about allowing emails they send being made available to users who were not included in the original distribution?
- What are the concerns about privacy and access control?

This investigation was carried out in two phases:

(1) Semi-structured interviews with a subgroup of engineers. This was used to explore if the system would be of benefit to users and what would be the best way of implementing it.

(2) Structured survey distributed to a wider group of engineers. This was based on the feedback received from the semi-structured interviews. It was used to get a feel for the overall responses to the project. It was also less time consuming for the participants.

Semi-structured interview with engineers

A semi–structured interview was carried out with six engineers – three Software (two senior, one junior) and three Product engineers (two senior, one junior). See Appendix 1 for the format of the interview. The interviews took an average of 35 minutes to conduct. The questions were exploratory in nature in order to elicit information.

Summary of Feedback

(1) Position within the Company

Three software engineers and three product engineers were interviewed.

(2) When looking for information, how frequently would you perform the following search? (Yearly, Monthly, Weekly, Daily)

The structure of this question was suggested by Swaak et al. (2004)

In comparison to other methods the mailbox was searched daily by all respondents, indicating that it was used as a knowledge source.

One surprising finding was the extent to which paper archives was given as another source for searching. A reason given by a product engineer was that some of the older products he had worked on dated to before the time email and even personal computers were in common use, so some of the documentation was not readily available in electronic form. Although these products were obsolete, some of the information could be reused. There were two aspects to this older information: because it was so old it was less relevant and less frequently used, but at the same time, it would be much more difficult to recreate. The group of people who understood it has shrunk due to retirement, resignation or reassignment. It would be almost impossible to recreate. Information also could be surprisingly long lasting – the technology used in machines was not likely to be cutting edge. The cost of redesign and retesting would

mean that older technology is used for as long as possible, hence the usefulness of old information.

This indicates a life cycle for information. When an email is sent to group, it is very likely that everybody understands the context, and the e-mail may not be considered valuable, because the information is readily obtainable. As time passes, this will not be the case. It is possible that no one in the group that received the email will still be working on the product or even in the organisation. Making storage of emails as automatic as possible would overcome the perceived insignificance of a message or document at the time it was sent.

One engineer printed out important emails and filed them in a ring binder. This was unexpected, but it clearly allowed important information to be located quickly. It uncovered an assumption that paper based archives were obsolete – in some circumstances, a paper based archive could be easier to search than an electronic one.

(3) Do you save email messages you receive? Yes/No
All respondents stored the email. In general, if it wasn't deleted immediately, it was never deleted, except in cases where the administration required disk space and the large unimportant emails were deleted before archiving. One response was "even if it's remotely useful I will save it".

Reasons for saving: one interesting response was that you don't always get time to read information fully – later you can bring the email up and read it when the information is more relevant.
Another reason was that the email keeps a record of what you did – it's like a log.
The information contained was important especially technical information.
Traceability – "cover your ass" did not appear to be a strong reason, and was strongly denied by two engineers.
Habit was listed as a reason.
Another interesting reason was the ability to reuse information. For instance, how somebody else filled out a report could be used as an example for filling in new reports.
One respondent said searching email was faster than searching folders on a hard disk.

(4) If you save email messages:

How long do you save email? Days, Months, Years, Indefinite?

All of the respondents listed indefinitely, once they had not decided that it was relevant. Irrelevant emails were deleted once they were read. One exception was that situations where storage space was required urgently on the server – in order to save time, emails with the largest file attachments and low importance were deleted before moving the email to a local archive.

What determines for how long you store your emails

The consensus was that if email was not deleted when it was first read, it was never deleted.

What determines how long the information in messages retains its usefulness?

This depended on the lifetime of the product the email contained information about. This includes how long the system needs to be supported – several years after it's obsolete. This issue reiterates the usefulness of an email archive – the engineer who generated the information may well have left the company. In addition, some old product information can be used with new projects.

Do you file messages in folders– Always/Sometimes /Never

How many Folders?

Five out of the six engineers used folders. This was surprising as it was anticipated that saving into folders would be time consuming. However, these findings bear out what was found by Ducheneaut & Bellotti (2001).

Nobody who used folders found it difficult to file. The number of folders varied between 5 and 25. A mixture of organisation of storage by person and by project was used. These are relatively simple filing systems, which is probably why it is not difficult to file. Bälter's (2000) suggestion that the best method of storing is by using a single folder was not found to be in use.

Do you use features like Microsoft Outlook Categories to organise mail?

This was prompted by Venolia et al. (2001). None of the engineers was aware of this feature, possibly because Microsoft Outlook was relatively new to the organisation

What is the size of your mailbox file?

Although the size of mailbox file was asked, as mail was stored in multiple locations, a reliable answer was not always available. With everyone keeping this email indefinitely, a figure of >1GB per person seems reasonable.

(5) If you search email for information, what type of information, for example?

(a) To find out how to do something – procedures, workarounds?

This question was aimed at discovering if Procedural Knowledge could be found in an email archive.

Two engineers said no. One of the engineers stated that he would use the company intranet or the internet to find procedural information.

Four said yes. An example given of procedural knowledge was by looking at previous qualification plans and reports to understand how they should be filled out. It was noted that the official archive of these document is not easily accessed, so previous documents that were sent through email provide guidance.

(b) Information about something – part numbers, specifications?

This question was aimed at discovering if Declarative Knowledge could be found in an email archive.

All 6 engineers replied yes to this question.

An example of Declarative Knowledge given was *Conversational* – who said what and who made a particular decision.

Other examples given were finding drawings, test results or price quotations.

(c) To gain a deeper understanding something or the reasons why a device, system or organisation works as it does?

This question was aimed at discovering if Causal Knowledge could be found in an email archive.

Five engineers said yes, one said no.

- One interesting observation was made: "searching email can trigger association that leads to understanding". The combination of content, the sequence in

which it was received and the people involved provides context that aids in understanding.

(d) Would you search your email archive to find a contact or locate expertise?

This question was aimed at discovering if Social Knowledge could be found in an email archive.

Three engineers answered yes, three answered no. This can be interpreted that where only the users local email archive is concerned, it's likely that the user is already aware of any contacts or expertise contained in previous messages. A search for expertise or contacts would be sought elsewhere – perhaps by sending a new email. If users email archives were shared, it's more likely that a search would return expertise that the user was not aware of.

(6) If you search your email archive for information, what type of search would you consider most useful?

Search by keyword within email message (1-3) with 1 being most preferred

The answers were 1,1,3,1,2,1. This was the most popular type of search.

Search by Keyword within email attachments (1-3) with 1 being most preferred

The answers were 3,3,3,3,3,3. This was the least popular type of search, probably because no engineer was aware that Lotus Notes could index attachments. Microsoft Outlook does not have this capability, but searching is possible using Microsoft Desktop Search or Google Desktop. The engineers were not using this functionality either. At least one engineer stated that the combination of search within attachments and within the email message by keyword would make this type of search most preferred if available.

Search by Sender name (sort) (1-3) with 1 being most preferred

The answers were 2,2,1,2,1,2. This was the second most popular type of search

(7) Do you consider email to be better or worse than face to face communication for transferring information/knowledge?

The purpose of this question was to explore how email can be used to elicit tacit information.

During the interview, it was clarified that using the phone was included in the term face to face.

What do you consider the advantages of email over face to face and phone?

- It provides a record of what was said
- Gives you time to think
- Ideas are worked out in email by being written down – "Answers are considered – they are less off the cuff"
- Email's clearer
- If the person you are dealing with has poor English, it's easier to communicate in writing
- You don't have to locate the person
- Richer than phone – you can send pictures along with an email
- You can add complex information, such as paths to files

What do you consider the disadvantages of email over face to face and phone?

- Slower in reaching consensus
- Five minute call can finalise an issue that will take days in email
- Too much "junk" in email
- With email you lose interaction with people

Do you use Instant Messaging to ask a colleague for information?

This question was asked to see if detect the extent that instant messaging was used to resolve problems compared to email. However, only two engineers used it.

Do you consider that Instant Messaging has any advantages over email?

The advantage of instant messaging over email was identified as being that for an urgent question it generally provided a faster response.

(8) Sharing information within email archives:

Do you think that your email archive contains information that could be beneficial if shared with others in the organisation?

- Two engineers somewhat agreed with this, one was neutral and three strongly agreed with this statement.

Do you think you could obtain useful information if others in the organisation shared information in their email archives with you?

- Two engineers somewhat agreed with this, one was neutral and three strongly agreed with this statement.

Do you think you could obtain useful information in the email archives of employees that have left the company?

- One Engineer somewhat agreed, one was neutral and four strongly agreed.

Reasons for the answers:

An engineer who somewhat agreed with the idea made the point that the people who would find the information useful would have already been included on the address list of the original mail, so they would already have it in their own archives. The concept would be useful for accessing mail from people who were not available and had left the company, but the search mechanism would have to be good.

An Engineer who was neutral to the idea of sharing current employees email (due to privacy concerns) but strongly agreed with sharing the email of people that had left the company, explained: "that in a group, if you need to know something, you can ask them directly - it's better to ask them directly". This doesn't apply when someone is gone. By looking at their email, you see how they did it and it speeds up the process.

An engineer that was strongly in favour gave as reasons – (a) you are able to access information held by a few experts and (b) it gives new members of group's access to that information and captures the experience of experienced users if they leave.

Another engineer who strongly agreed gave as a reason that the conclusion to how problems were resolved is contained within email. Also, if someone is absent at the

moment, information in their mailbox is probably inaccessible - with a shared archive the information would be made available.

(9) If you consider that the organisation would benefit by sharing the information within mailboxes, can you suggest what you consider the best way to share that information?

Single central repository or the ability to search email archives on the users PC?

Answers provided included:

- Single (global) repository is better – files can be modified, and people who need the information need to know where it is placed. Also if individual PC is used it's more likely to lose data.
- Single archive would be better, as personal email archives are scattered (CD, network drives, hard drives) and so would be difficult to access.
- Single repository providing a single point of access.

  However, if user's data is organised by folder, it would be easy to exclude private data, so searching individual PCs would be possible.

Single folder or multiple folders?

Answers provided included:

- Prefer multiple folders. With the large amount of email, having a single folder makes it hard to find information.
- Prefer multiple folders organised by category. For a single folder a very efficient search mechanism would be required.
- Multiple folders, by project. When a topic comes to a conclusion, email it to an address. Also have rules based on the subject header to auto archive. However, if folder structure becomes complicated, they may become difficult to search.

Email added automatically to a central repository or user chooses information to send?

- When documents are being reviewed there will be multiple modifications sent out for review – if there are six versions should all six be placed in the archive? Or only place the last one? If only the last one was placed in the archive there would not be confusion over number of documents. So voluntary would be better, as it would allow users to choose to submit the last version of the document to the archive. Also, subsequent deliveries should overwrite the previous one. Users need to where they can find the latest document. There should be a location that users can be directed to when files have changed. Versions of documents may change very quickly.
- One suggestion was to make it optional. Users could set up rules to send emails to the repository automatically. There could be two types of archive - a general purpose one that holds low status email and another one where emails with high knowledge content could be sent. However a problem was identified with this idea – everybody would send mail to the "important" archive and the less important one would be unused.
- It would be better to allow the user to choose what emails are placed in the archive.
- It would be better to allow the user to choose – best if a button or menu item was provided that sent the information into a location.
- It would be better to allow the user to choose – that way it would prevent personal information from being placed in the archive.

Would there be any particular useful way for presenting the information?
- Provide a brief description of contents of file so it's easy to browse search results. Allow attachments to be searched by keyword.
- It would be better to provide a Blog.

Any suggestions for the level of access to this information?
- Global access unless the information business sensitive.
- If the archive is organised in folders, access levels can be determined by folder. If email has been forwarded voluntarily, there should be no issue about access.

What type of search would be most beneficial?

- Search within attachments or identify attributes of emails.
- Search within attachments.
- Search by multiple criteria using filters. Search would have to be fast.
- Search by categories – Project/Product

(10) Would you be willing to allow your mailbox to be searched for information

Questions about privacy were asked to investigate the level of concern about privacy. At this point the feasibility of the most extreme form of email sharing was under investigation – that all email could be searched by an organisation.

Would you be willing to allow your mailbox to be searched for information by your supervisor?

Three engineers said yes, three said no, although one of the "Nos" was initially a yes as he believed that his supervisor could already monitor his email. The "Nos" were an indication of unhappiness with the idea – there was awareness that the company has the right to monitor mail (it's part of the email policy).

Would you be willing to allow your mailbox to be searched for information by a set of colleagues that may change without agreement (team changes?)

Three engineers said yes, three said no.

Would you be willing to allow your mailbox to be searched for information by anyone in the organisation?

All six engineers answered no to this question.

(11) Do you have any concerns about allowing others to search your mailbox?

Concerns about privacy? Yes/No

All six engineers answered yes

The response revealed that all engineers had concerns about privacy related to emails concerning work related matters. Examples provided were:

- Emails from Human Resources
- Emails from the Medical Centre

- Emails containing confidential information. A contact from a suppliers company sends their personal phone number in case they need to be contacted urgently
- Emails containing information about reviews with supervisors
- Email to former colleagues requesting a work reference

Most suggested that they would be willing to share with colleagues if this sort of information could be excluded.

One engineer did not think that sharing information by sharing email was a good idea. He stated that if information was required, it would be simpler to ask a colleague.

Another identified that it would be less of a problem with a central mailbox, where people volunteered information – they had already identified the information as not sensitive.

Another identified a problem with sharing with everybody in the organisation – although the expectation is that most colleagues would have little interest in items that did not concern them, you can never be sure that someone might take the email out of context and misinterpret it. This does not agree with the suggestion that a more open communication flow fosters trust, as identified by Gloor et al. (2003).

There did not seem to be as much concern about personal email that did not relate to work related issues, although two engineers pointed out that a user cannot prevent others from sending mail that is potentially sensitive.

Information out of date, incomplete, incorrect?

Five engineers had concerns about incorrect data. The engineer who didn't have a concern with this suggested that when a relevant document was identified, it would be expected that the user would be professional enough to confirm that it was the latest version available. A problem with the concept was identified – during review of documents, multiple revisions of a document could be placed in the archive. There would be difficulty ensuring that the latest version – or finalized version had been placed in the archive. A way would have to be identified informing people where the documents were, and when they were updated.

(12) Can you identify other means that would improve knowledge sharing within the organisation?

- Share information on a DVD or network drive.

- Have a method of informing user that information was wanted – others can add it to the database. That is, provide a pull system rather than a push system – this makes sure that the information being added is worth putting up – otherwise it may not be used.

- Forums – when conclusion is reached for an issue, email it to the forum.

- Web logs were identified as useful – like social networking sites such as Facebook and Bebo – more focussed, more relevant. Pointed to companies like IBM requesting their employees to use these networking sites as a means of keeping up with current trends in technology.

- One engineer identified asking people the best way of sharing knowledge.

Conclusions from Semi-structured Interviews

1   All engineers were using their stored email as a source of knowledge

2   Most were storing using multiple folders

3   All had concerns about privacy

4   Most had concerns about the information being up to date

5   Most thought they could get information from a shared email archive

Feedback from Survey

The semi-structured interviews were discontinued after six were conducted. A more focussed survey was devised – see Appendix 2. A survey was used in order to obtain information from more users - the interviews were taking 30-40 minutes, and it was difficult to secure the availability of engineers.

The survey repeated a lot of the questions in the semi-structured interview.

One item that was dropped was the issue of privacy. This was because the sample of six engineers confirmed that it would be an issue, and that the approach of making all email searchable would not suit everyone. A voluntary approach would be more suitable. Also, the issue might raise concerns if presented in a survey, without the feedback from an interviewer to explain the purposes of the questions.

Twenty surveys were distributed among the engineering group, Quality Control and the IT department. Nineteen responses were received.

(1) What is your job title?

The survey response came from IT, specialists, Product engineers, Software Engineers, and Quality Control. The Engineering manager, IT manager and Quality Control manager responded.

(2)   When looking for information, how frequently would you perform the following search? (Yearly, Monthly, Weekly, Daily)

The majority of the users searched their email daily. It was the most popular means of searching for information, followed by network drives, local disks, and contact a colleague. This confirmed that email is an important source of personal information management.

(3) Do you store email messages you receive?

18 out of 20 stored their email. The most important reason was information contained, followed by traceability.

The breakdown was interesting – members of the quality assurance department, the engineering manager and IT considered traceability most important – the product engineers considered the information contained in the email more important.
An IT engineer explained that traceability was important when dealing with suppliers – you needed a record of the email to confirm commitment.

(4) If you store email messages:

(a) How long do you retain them, before you delete?

Most (14 out of 19) stored email indefinitely.

(b) What do you think determines how long the information in messages retains its usefulness?

Reasons given:

- Work Projects - Open/Closed and Related Products – Current/Obsolete.

- Relevance to current or ongoing projects.

- My ability to recover things.

- When new information becomes available, it makes the old information redundant. The more you investigate an issue, the more relevant data is discovered.

- Legal requirements.

- Varies from project to project but pertinent information should be retained for as long as the project exists.

- Until the issue the problem relates to is resolved fully.

- Richness of technical or specialist content.

- History of event or communication - if deemed important it should be stored long term.

- Length of relevance to project/accountability/compliance.

- Evidence of complaints, faults with documentation or instruments.

- Lifetime of product.

- As long as product/subject is still in active use, discussion it is worthwhile. Also useful reference for historical purposes.

- Person who sent it or the product it relates to.

- The subject determines its usefulness, for example HR information would be held/retained for longer than IT information due to the fact that HR info does not change as often as IT does.

- Some information doesn't lose its usefulness i.e. cheat sheets.

- The importance of information contained.

- Importance, relevance to job, traceability and reference.

(c) Do you file messages in folders? Yes/No

17 out of 19 people stored their email in folders.

(d) How many folders?

The average number of folders was 17.2.

Note – for users who responded with ranges e.g. 6-7, then the upper limit was used. For users who responded with a value of greater than e.g. >7, then the value after the comparator symbol was used.

(e) Do you find it difficult to find the correct folder when filing? Yes/No

4 found it difficult to store in folders, compared to 13 who didn't.

(f) What is the approximate storage space required for your archived email

Size of mailbox – 7 users responded as >1 gigabyte. Most users did not know what their database size was. It is expected that most users would be using >1 gigabyte of storage.

(5) Would you consider an email archive a useful source of knowledge for the following?

(a) To find out how to do something – procedures, workarounds?
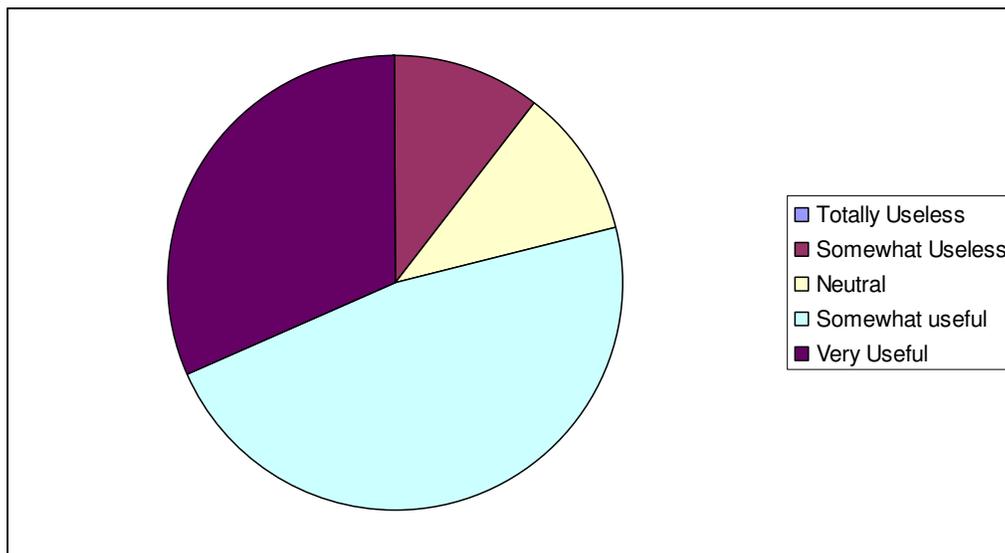
Figure 4.1: Usefulness of searching email to find Procedural Knowledge

(b)To find information about something; e.g. –part numbers, specifications?

Legend:
- Totally Useless
- Somewhat Useless
- Neutral
- Somewhat useful
- Very Useful
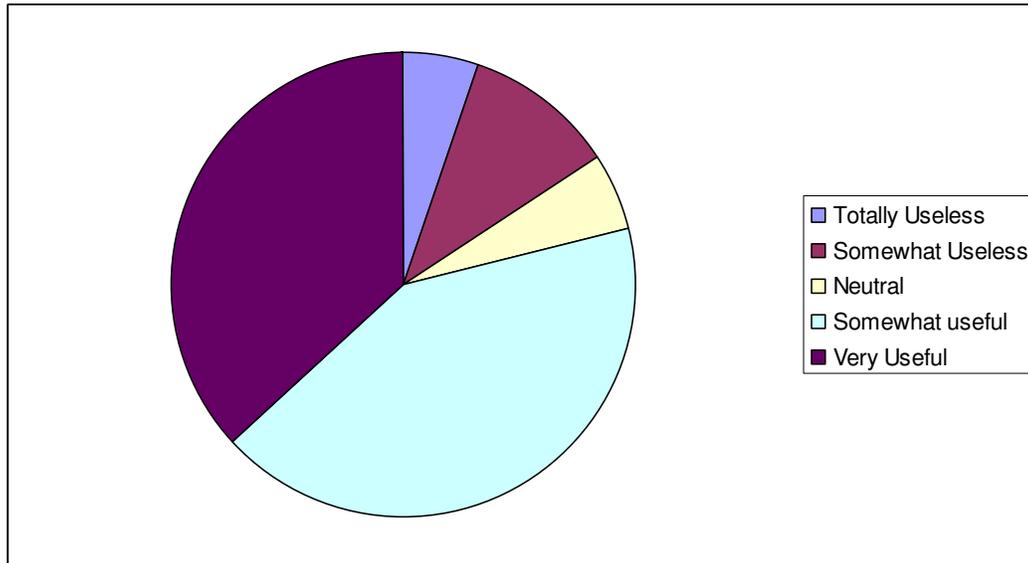
Figure 4.2: Usefulness of searching email to find Declarative Knowledge

(c) To gain a deeper understanding about something, or the reasons why a device, system or organisation works as it does?

Legend:
- Totally Useless
- Somewhat Useless
- Neutral
- Somewhat useful
- Very Useful
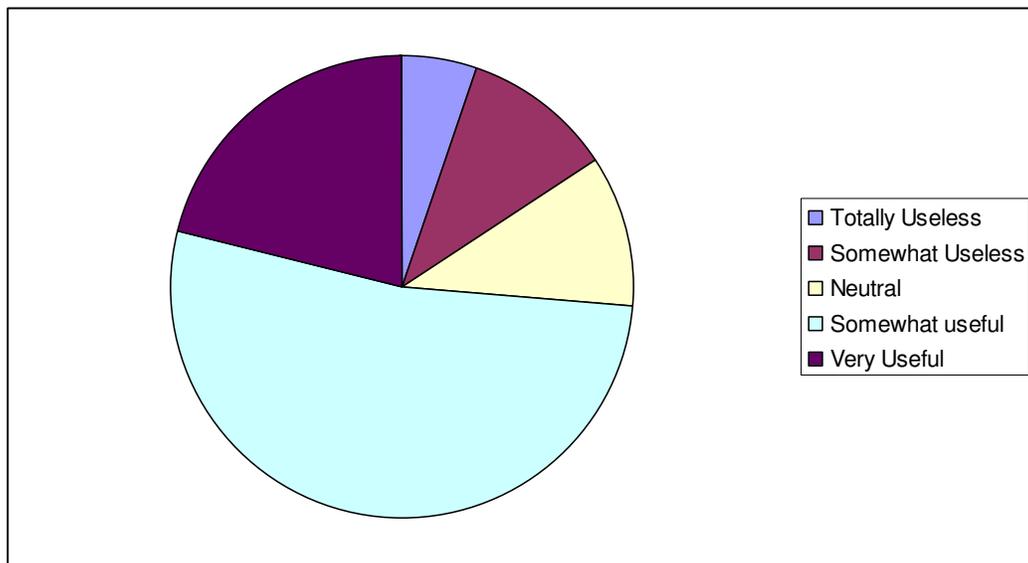
Figure 4.3: Usefulness of searching email to find Causal Knowledge

(d) To find a contact or locate expertise?
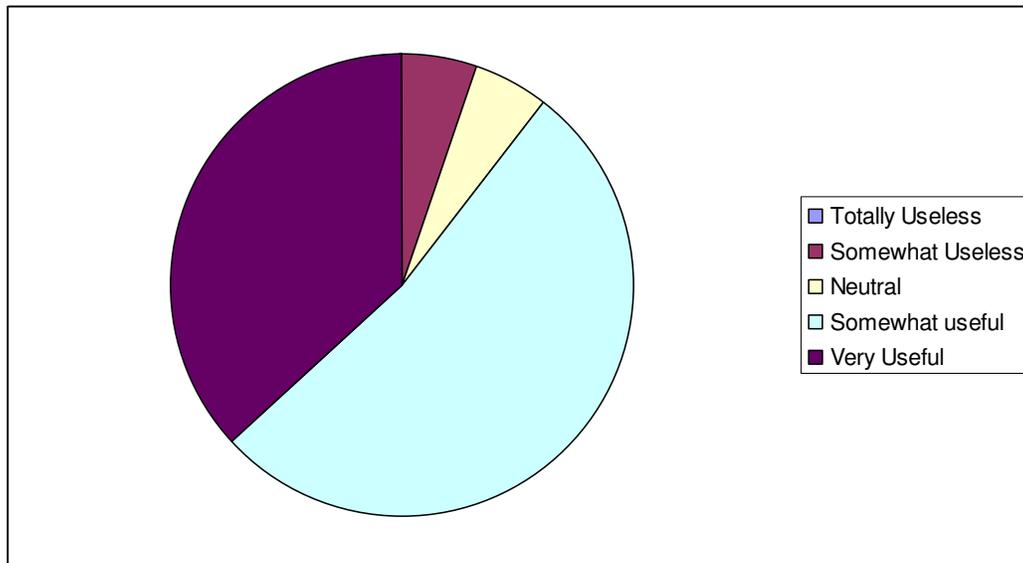
Figure 4.4: Usefulness of searching email to find Social Knowledge

Table 4.1: Usefulness of Email as a source of Knowledge

Would you consider an email archive a useful source of knowledge for the following?	Totally Useless	Somewhat Useless	Neutral	Somewhat useful	Very Useful
To find out how to do something – procedures, workarounds?	0	2	2	9	6
To find information about something; e.g. – part numbers, specifications?	1	2	1	8	7
To gain a deeper understanding about something, or the reasons why a device, system or organization works as it does?	1	2	2	10	4
To find a contact or locate expertise?	0	1	1	10	7

The results of the survey indicate that overall people thought that an email repository was a useful source of knowledge. The results were strongest for the category corresponding to social knowledge (to find a contact or locate expertise), indicating

that a map locating expertise against a particular person may be useful, e.g. Yellow Pages.

(e) Can you identify other type of knowledge that can be found in email archives?

- When starting on a project - a review of email can be a useful tool for gaining knowledge.
- To identify why a decision was made
- Justify some action at a later date
- History of an observed issue and the resolutions(s) associated with the same.
- Communications on a particular issue
- Reason for decisions
- To share a lost contact
- Contacts
- Contact details location of knowledge
- HR/legal info that does not change as often
- Project tracking
- Reminders of Work in Progress - i.e. task list.
- Drawings

(6) If you search your email archive for information, what type of search would you consider most useful to you?

Table 4.2: Comparison of usefulness of search types

If you search your email archive for information, what type of search would you consider most useful to you? Prioritise 1 -3, 1 being most useful	Score	1	2	3
Search by keyword within email message	34	7	9	3
Search by Keyword within email attachments	46	3	5	11
Search (or sort) by persons name	34	9	5	5

It was expected that the search by keyword within email attachments would have been the richest source of information and should have had the lowest total (highest preference). This finding confirms what was found during the Semi-Structured Interviews, and indicates a low awareness of the ability to search within attachments. The high preference for search by person name seems to indicate that information is associated with a particular person. This may have implications on how email is searched - people might be more inclined to go to a folder associated with an expert rather than a general search.

(7) In searching an email archive for knowledge, what would you consider the best method for presentation?

Table 4.3: Comparison of Knowledge presentation methods

In querying an email archive for knowledge, what would you consider the best method for presentation? Prioritise 1-3, where 1 is most preferred.	Score	1	2	3
Return a list of emails that match search criteria sorted by relevance, date or sender	29	11	6	2
Create a report presenting a summary and analysis of information	49	3	2	14
Provide a visual display showing where information is located(by sender or folder)	35	6	10	3

Returning a list of emails sorted by relevance is clearly the most useful.

The summary report option was clearly the least popular. It was considered that this option might be preferred by users identified as managers – however, all three users identified as managers chose the option to provide a visual display that would indicate where information is located.

(8) If a folder for sharing archived emails was made available:

(a) Would you be willing to add emails you consider useful to the folder as you compose and send them?

(b) Would you be willing to add any of your existing archived email to this central folder?

(c) Do you think you would search this folder for information?

Table 4.4: Comparison of methods of sharing knowledge

If a folder for sharing archived emails was made available:	Yes	No
Would you be willing to add emails you consider useful to the folder as you compose and send them?	11	7
Would you be willing to add any of your existing archived email to this central folder?	14	5
Do you think you would search this folder for information?	16	3

Overall, users seemed to support the concept. However, among the users who did not support the concept were the IT and the QA managers.

d) Can you suggest from the list below what you consider the best format for organising a central email folder?

Table 4.5: Comparison of methods for organising folders

Can you suggest from the list below what you consider the best format for organising a central email folder? Prioritise 1-3, where 1 is most preferred	Score	1	2	3
Single folder – users simply drop emails into the folder. Access is controlled by a central user.	45	1	4	12
Multiple Folders organised by Product. Access controlled by a central user.	24	10	7	0
Multiple folders organised by User. Access controlled by individual users.	33	6	6	5

The most preferred approach was multiple folders organised by product, with access controlled by a central user. If privacy was a primary concern, it would be expected that the options of multiple folders organised by user, with access controlled by the users themselves should have been most preferred. This indicates a preference for a logical or convenient structure over privacy concerns.

However, the Engineering Manager did not agree with any of these options and returned the following comments:

(1)   Adding new emails to the central repository would have to be automatic

(2)   Adding existing emails to the repository would depend on the degree of effort

(3)   All options presented require additional work and discipline on part of multiple users and comparative difficulty to monitor. The key is to avoid work during inputting but to save work for formatting outputs when found.

The QA manager was also strongly opposed, but for other reasons. A follow up meeting revealed these reasons:

(1)   Prescription of what should be shared. Everyone will have different views on what should be shared and they may share inappropriate details.
        Even within a community of practice there may be differences in what people think are useful emails.

(2)   Generally, emails are copied only to those who need the information. This will be known at the time. The use of a shared archive leaves the possibility open that at a future time, an email may make its way to an inappropriate recipient.

(3)   Subjectivity - other people's vocabulary may be different – making it difficult to search by keyword reliably. Everybody interprets emails differently. Each person has their own context, and will understand their own emails, but others might not. So extracting knowledge from the email conversations of others would be very difficult. Therefore, a shared email repository may be of limited use.

(4)   Regulatory – if this became part of the organisations system, it has regulatory implications. This is not necessarily bad, but could cause problems; for instance procedures for maintenance of access control.

Overall the feedback to Question 8 poses some difficulties. The initial interviews with the engineers suggested concerns with privacy, but feedback indicates that a

willingness to share with others through a shared project folder. Therefore, providing an automatic mechanism to add emails to a central project repository might be the best approach. However, feedback from the QA manager indicates that even with a group working on the same project there will be differences in deciding what to share.

(9)   Can you identify other means that would improve knowledge sharing within the organisation?

Table 4.6: Other means of sharing knowledge

Can you identify other means that would improve knowledge sharing within the organisation?	Yes	No	Don't Know
Wikis	10	0	9
Weblogs	9	0	10
Instant Messaging	12	2	5

The response for Wikis, Weblogs and Instant messaging indicate areas for future work. In particular, instant messaging – the Instant Messaging application was recently changed, so making users aware of the new application would be beneficial.
Other sources of knowledge

- Server folders for a number of items - product knowledge, test results etc.
- "When projects are completed - we never review - a review would be good for uncovering good practices and areas of improvement".
- Document management system accessed through intranet.
- Sharepoints: project information sharing.
- Lotus Notes database.

(10) Have you any other comments or suggestions?

- "Sometimes I think emails are abused by the fact too much data is put in them. An email should be a short message -anything with a lot of content should be an attachment to the email."
- "We use our current document repositories (DMS etc) for controlled document sharing."
- "Email archiving sounds like a very useful and helpful tool. Use of Instant Messaging would also greatly help."

- "Many emails contain sensitive details plus correspondence usually written as a result how the author felt at that time. They may not feel comfortable opening that to others, especially if it can be accessed by parties after written. Meaning of email can be misinterpreted. Allowing people access for non-business related activities can cause viruses!"
- "Any type of open communication facility is very useful for finding/sharing knowledge. It can be hard to adapt these tools but once in place they are very useful."
- "Usefulness should be determined by receiver, not sender. Sharing archived email would be very useful."

4.6 Existing Knowledge Management Applications

As well as investigating the use of email as a knowledge management tool, other knowledge management tools within the organisation were examined.

Lotus Notes provided the concept of team rooms, where documents could be shared among team members. However, due to recent changes in infrastructure, teams rooms were not always accessible.

An application that has been recently rolled out within the company provides file sharing and other collaborative tools such as chat, polling and discussion fora. An interesting observation about this is that the primary usage currently revolves around file sharing and that the collaborative aspects do not appear to be widely used. One possible reason is that the user has to be logged onto the application to avail of the collaboration tools such as chat groups and discussion fora. User activity has been analysed which observes that users logon, download files and then logoff the application. Email is used to communicate with users regarding latest versions and updates to documents that they may need to review. This backs up the view of email as user's primary habitat. The use of the file sharing facilities of the new application indicates the limitation of email as document sharing, reviewing and version management.

4.7 Conclusion

This chapter examined the feasibility of implementing an email archive as a knowledge repository. It was determined that such a system would be beneficial and technically feasible.

The use of email as a means of obtaining knowledge was examined. It was found that almost all users were storing email and consulted it on a daily basis. Most users stored their email indefinitely. The length of time email remained useful generally depended on the lifetime of the project they were working on. Older information was potentially less useful, but also more difficult to recreate, increasing its value. People were as likely to search by peoples name as by the content of email. Searching within the content of attachments was least popular, but this was most likely due to lack of awareness of this search capability.

Interviews with users suggested concerns about privacy of work related emails, rather than personal email. This suggests that the adding of email to a shared repository should be voluntary.

Overall users supported the idea of a shared archive. The preference for storing email was a central folder organised by product, with access controlled by a central user. It would be expected that concerns about privacy would favour the ability to control access by individual user. Feedback indicated that unless the adding of emails to a repository was made automatic, users would not use it. This suggests an approach of generating rules that add emails automatically to a project folder based.

Problems with the approach were identified – users have different ideas of what should be shared, even within a group. Differences in users' understanding and the context in which the email is read would make it difficult for others to interpret email.

5   EXPERIMENTATION & EVALUATION

5.1 Introduction

This chapter details the experimentation that was undertaken, to establish a shared email archive for a community of users, based on the existing features of email applications. Based on the survey results, these features were used to develop and implement a prototype system that could be used and implemented to meet the user requirements. The use of this prototype was user evaluated.

5.2 Structure of the system

Examining the response to the survey and interview, the following decisions were made.

- The system would be focused on Software and Product Engineers.
- Initially the system would be based on folders for each project.
- Within the last 6 months the organisation had changed the email system from Lotus Notes to Microsoft Outlook. As Lotus notes was in use for several years it was expected that most email archives would be in that format, except for new members of teams. As Lotus notes could still be accessed by all engineers, and the feasibility study had indicated that it was possible to create a shared archive using Lotus Notes, this would be used for the Lotus notes type email. It was proposed to investigate the creation of two separate storage areas, one in Notes and the other in Outlook, and then to investigate ways of providing a common search method across both systems.

The overall structure of the system is envisioned as shown below – it will rely on the existing IT infrastructure.
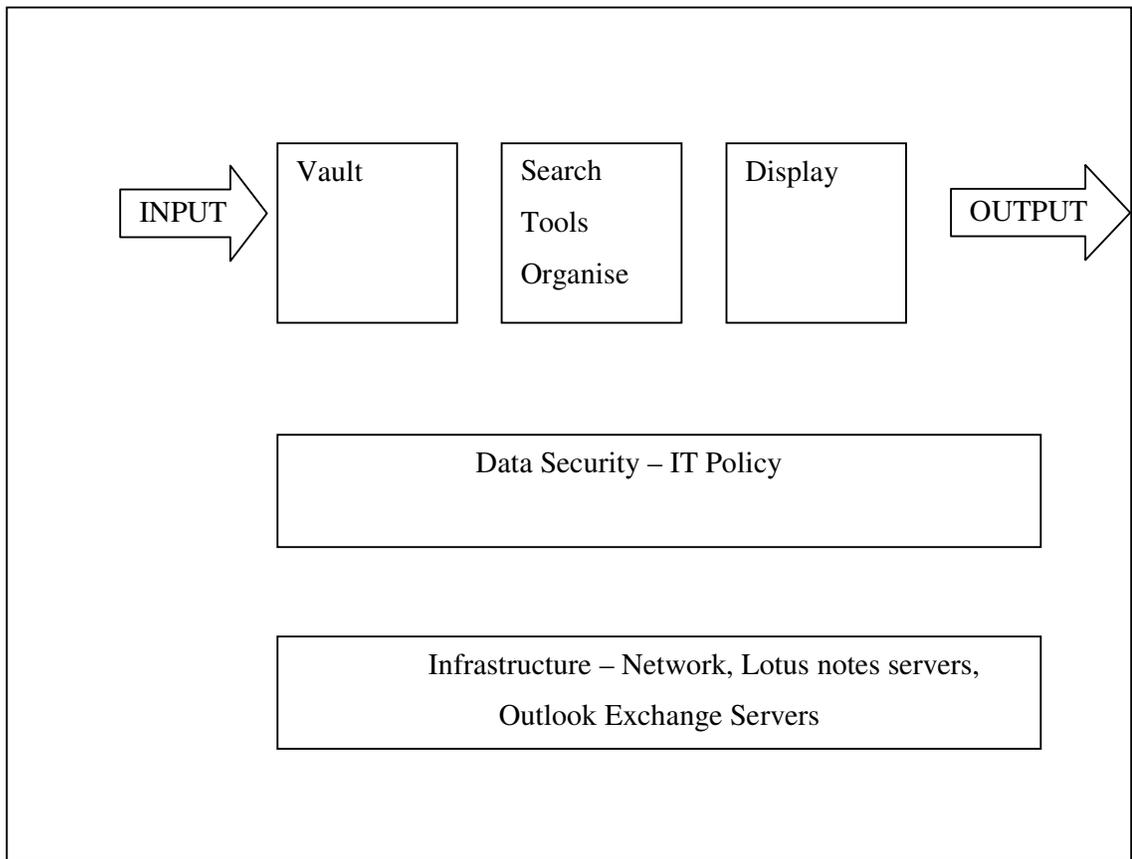
Figure 5.1: Environment

General description of LOTUS Notes

Lotus Notes is a collaborative application supplied by IBM (IBM Corporation, 2003). Part of the application is email. In Lotus each users email is stored in a database, formatted to store email. The databases can be located on a server or located on local computers. Users access the databases using the Lotus client.

The data inside a Notes database is stored as a set of records - each record is referred to as a document. The information in a document consists of one or more fields, which can be in a variety of formats. Text, numbers and dates are the basic data types, but formatted rich text and file attachments can also be stored in fields. A database is an

NSF (Notes Storage Facility) file, containing a basic unit of storage known as a "note". Every note has a UniqueID and a NoteID. The UniqueID uniquely identifies the note across all replicas within a cluster of servers, a domain of servers, or even across domains belonging to many organizations that are all hosting replicas of the same database. The NoteID, on the other hand, is unique to the note only within the context of one given replica. Each note also stores its creation and modification dates, and one or more items.

Attachments are transmitted as MIME but are extracted and stored in the $FILE field in a document.

Searching capability: The contents of a lotus notes database can be indexed – this includes the contents of attachments. This allows search by keywords, including operators such as AND, NOT and wildcard characters.

Creating the Lotus Notes Archive

This was relatively straightforward – a new mail database was created, where emails could be stored. Users given access to this database could create their own folders and set permissions – in effect, this database could be organised by both user folder and project.

Identifying emails to add to the Lotus Archive

An analysis of the mails suggested the following rules:

- The mail must not have originated from, or include in the Send field, anybody who worked in HR or the Medical department – this included past employees.
- The message must have been copied to at least one other person. This was used as an indication that the message was considered shareable at the time. It may exclude a lot of possibly useful email, but reduces the possibility of highly confidential data been shared.
- The From or To field must have at least one entry with the email address of the company. This reduces the possibility that a non-work related email is included. A stronger approach would be to have a rule where the email

addresses in the To field must all belong to the company, but this may be too exclusive. This requires further investigation.

Microsoft Outlook

Information on Outlook was obtained from DiGiacomo (2007). Microsoft Outlook is an application that includes Email and Scheduling. Using Exchange server, emails are stored on a central server, and replicated to individual mailboxes. Due to storage limits on the exchange server, users must regularly delete or archive messages in their mailbox.

Archived items in Outlook are stored in a Personal Folder Files (.pst). There is a limit of 5 GB on a person folder files.

Attachments are stored within the Mail Messages using the MIME (Multipurpose Internet Mail Extensions) format

Search: Outlook allows search by Keyword and operators. However, the contents of attachments are not searched.

Sharing data: Other users can be given access to items within a folder using the Delegate feature.

Security: Individual items can be marked as Private (DiGiacomo, 2007, P.585) – other people with access to the folder cannot view that item. The Delegate feature can be configured to allow access to Private items to delegated users.

Personal folder files can be protected using passwords. DiGiacomo (2007), P.586 warns that passwords can be cracked.

If the folder is shared with others using Server, access can be controlled to items via permissions.

Creating the Outlook Archive

The Public folder feature in Outlook could be used – however this feature has been deprecated in Microsoft Outlook 2007. Although it is still available, corporate policy would not allow new Public folders to be created.

However, a Functional Mailbox was created instead, which served a similar purpose.

The initial thoughts on a name were to call it EngArchive. At the Engineering Manager's suggestion, this folder was given the name Encyclo, to focus the users' attention on the reason for its creation – as a knowledge repository.

Once again, users could create their own folders and set permissions.

Using the Macros feature in Outlook, a button was devised that allowed users to store a mail in a folder with a single click.

5.3 Evaluation

Due to work conditions, implementation was restricted to the investigator and one other user in the software department.

The following issues were uncovered:

(1)   The filtering facilities within Lotus Notes were found to be very configurable and using this filter to identify Lotus Notes mail that can be copied to a central database was found to be relatively easy.

(2)   Data older than seven years was returned. Although it would contain useful information for anyone who had to work on similar projects, placing this data in a central archive could be used to identify a user as an expert. If the user had moved on from that project, they may no longer have that expertise. This may cause users to be reluctant to add older emails to the archive. This could be investigated further by expanding the evaluation to a wider group.

(3)   The use of a button to store email in the shared folder, despite being integrated into the email application, was still too time consuming to use under a heavy workload, even when the button was a novelty and the user was motivated to share. This provides strong evidence that an automatic rule for adding of emails would have to be developed if emails were to be added to the archive as they were generated.

(4)   The other user, a relative newcomer to the group, did find the ability to access older emails sent before he joined the group to be helpful.

.

6   CONCLUSION

6.1 Introduction

The aim of this project was to investigate the value of a shared email archive as a knowledge management resource in a manufacturing plant. The final chapter of this dissertation presents the conclusions and recommendations of this research project. It presents a summary of how the research aims and objectives were achieved, within the research definition and overview. It looks at what this research contributed to the body of knowledge. It evaluates the prototype, details any limitations and proposes future work.

6.2 Research Definition & Research Overview

This dissertation investigates the use of a shared email archives as a means of capturing and distributing knowledge, and how the use of an email archive can increase the value of knowledge within a company by increasing its accessibility and by converting tacit knowledge into explicit knowledge. Key stakeholders were interviewed regarding the benefits and technical feasibility of creating such an archive. Users were interviewed and surveyed on how they use, store, save and search email, the knowledge contained within email and their willingness to allow their email to be shared with a wider group.

The use of email in knowledge management was examined during the literature review. Issues of privacy of email within work were examined through analysing available information from the EU and Irish Data Commissioner and various other sources.

6.3 Contributions to the Body of Knowledge

The research confirmed findings by other researchers that the practice of archiving email was widely used and users regularly search these email archives as a source of knowledge. The preference for users to store emails in folders was also confirmed.

Privacy issues within emails were identified as being related to personal data that originated within the organisation, rather than non-work related emails that originated outside the organisation. Examples of messages that were of concern were identified as emails from HR, performance appraisals, medical appointments and sensitive information sent to or received from colleagues.

The preference that most users expressed for an archive based on project folders with centrally controlled access, indicates that privacy was less of a concern compared to convenience and logical organisation. It also indicates a willingness for users to share and trust others working on the same project, similar to the concept of a community of practice.

6.4 Experimentation, Evaluation and Limitation

Experimentation consisted of examining the features of different email applications to establish how the requirements identified during the investigation could be implemented. These features were then used to develop a prototype system. This prototype system was implemented but could only be shared among two users, so limited evaluation was conducted. Limitations were encountered due to reorganisation in the company which restricted a possible rollout to a wider group, and time restrictions due to conflicting projects.

6.5 Future Work & Research

(1) Extend the use of the email knowledge management prototype to more groups, and observe use patterns, perhaps using file access and keystroke monitoring to gather objective evidence of user activity.

(2) Investigate robust means of identifying email that are considered private. This will increase the likelihood of users allowing automatic rules to place emails in a stored archive.

(3) Investigate means of extracting information from emails, with context, while removing any personal or emotive words, in order to generate an FAQ.

(4) Investigate the presenting of email in a more graphical format.

(5) Investigate the generation of a Yellow Pages through the use of email in the organisation.

6.6 Conclusion

A company's knowledge is a vital and valuable resource. The email communication of employees within a company generates knowledge. Email is currently a major part of most users work environment – their "habitat". The generation of knowledge within email communication and its integration into a user's workflow makes it a primary target for positioning knowledge management applications within a company. This investigation found that an email archive would be useful as means of managing and sharing knowledge within a company.

Research conducted during this dissertation concluded that users were archiving email so that they could search for information contained within it in the future. Over time the knowledge contained in the email archive can increase in value as the experts who have this knowledge become unavailable thereby making it more difficult to regenerate this knowledge. Research identified users concerns regarding privacy and sharing email content that may be of a sensitive nature. Until a reliable method of removing emails with sensitive content can be identified, users seem to be more comfortable sharing email within a clearly identified group working on similar projects. Even though email was integrated into a user's workflow, knowledge sharing was not likely to take place unless the process was made as automatic as possible, making the reliable identification of sensitive data a critical requirement.

BIBLIOGRAPHY

Aery, M & Chakravarthy, S 2005, *eMailSift: Email Classification Based on Structure and Content*, Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM'05)

Ardichvilli, A, Page, V, Wentling, T 2003, *Motivation and barriers to participation in Virtual Knowledge Sharing Communities of Practice*, Journal of Knowledge Management, Volume 7 Number 1 2003

Article 29 - Data Protection Working Party 2002, *Working Document on the surveillance of electronic communications in the workplace,* retrieved on 1/June/2008 from http://eceuropa.eu/justice_home/fsj/privacy/docs/wpdocs/2002/wp55_en.pdf

Bälter, O 2000, *Keystroke Level Analysis of Email Message Organization*, CHI '2000 The Hague, Amsterdam

Burk, M 1999, *Knowledge Management: Everyone Benefits by Sharing Information,* Public Roads Magazine, Vol. 63,No. 3, November/December 1999

Davenport, TH & Prusak, L 1998, *Working Knowledge - How Organizations Manage What They Know*, Harvard Business School Press, Boston, MA

Dabbish, L, Kraut, R, Fussell, S, Kiesler, S 2005, *Understanding Email Use: Predicting Action on a Message*, CHI 2005, April 2–7, 2005, Portland, Oregon, USA

Data Protection Commissioner, *Guidance Notes – Monitoring of Staff*, accessed on 1/June/2008 from www.dataprotection.ie/docs/Guidance_Notes_Monitoring_of_staff /208.htm

DiGiacomo, P 2007, *Using Microsoft Office Outlook 2007*, Que Publishing

Ducheneaut, N & Bellotti, V 2001, *Email as habitat: An exploration*

of embedded personal information management. Interactions, Sep-Oct 2001, pp. 30-38

Fisher, D, Brush, A, Gleave, E, Smith, M, *Revisiting Whittaker & Sidner's "Email Overload" Ten Years Later*, *CSCW'06*, November 4–8, 2006, Banff, Alberta, Canada

Gartner, Inc. 2004, *The Gartner Glossary of Information Technology Acronyms and Terms*, retrieved 1/June/2008 from www.gartner.com/6_help/glossary

Gloor, P, Laubacher, R, Dynes, S, Zhao,Y 2003, *Visualization of Communication Patterns in Collaborative Innovation Networks – Analysis of Some W3C Working Groups*, *CIKM'03*, November 3–8, 2003, New Orleans, Louisiana, USA

Goldstein, J & Evans Sabin, R 2006, *Using Speech Acts to Categorize Email and Identify Email Genres*, Proceedings of the 39th Hawaii International Conference on System Sciences - 2006

Heer, J, Viégas, F, Wattenberg, M 2007, Voyagers and *Voyeurs: Supporting Asynchronous Collaborative Information Visualization*, CHI 2007, April 28–May 3, 2007, San Jose, California, USA

Hafner, K & Lyon, M 1996, *Where wizards stay up late: the origins of the Internet*, Touchstone Edition 1998

IBM Corporation, 2003, *Lotus Notes 6.5.1 Help* 2003, accessed from Lotus Notes application Help menu.

Information Commissioners Office, *Lawful Business Practice regulations*, accessed on 22/June/2008 from:
http://www.ico.gov.uk/upload/documents/library/data_protection/practical_application/coi_html/english/supplementary_guidance/monitoring_at_work_3.html

International Labour Office, 1997, *Protection of workers personal data,* accessed on 23/June/2008 from www.ilo.org/public/english/protection/condtrav/pdf/wc-code-97.pdf

Kalman, Y & Rafaeli, S 2005, *Email Chronemics: Unobtrusive Profiling of Response Times,* Proceedings of the 38th Hawaii International Conference on System Sciences - 2005

Kerr, B 2003, *THREAD ARCS: An Email Thread Visualization*
IEEE Symposium on Information Visualization 2003,
October 19-21, 2003, Seattle, Washington, USA

Leuski, A 2004, *Email is a Stage: Discovering People Roles from Email Archives*, *SIGIR'04,* July 25–29, 2004, Sheffield, South Yorkshire, UK

Lichtenstein, S 2004, *Knowledge Development and Creation in Email,* Proceedings of the 37th Hawaii International Conference on System Sciences – 2004

Mariano, S & Casey, A 2007, *The process of knowledge retrieval*, VINE: The journal of information and knowledge management systems Vol. 37 No. 3, 2007, pp. 314-330

Newman, P 2002, *Exploring discussion lists: steps and directions, JCDL'02*, July 13-17, 2002, Portland, Oregon, USA

Nonaka, I & Takeuchi, H 1995, *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*, NY:Oxford University Press, 1995

Pagell, R, Meszaros, J, Valentine, D 2004, *Introduction to E-Mail Management and Knowledge Management*, World Library and Information Congress:
70th IFLA General Conference and Council 22-27 August 2004 Buenos Aires, Argentina

Perkiö, J, Tuulos, V, Buntine, W, Tirri, H 2005, *Multi-Faceted Information Retrieval System for Large Scale Email Archives*, Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI'05)

Rasch, M 2006 *Employee Privacy, Employer Policy*, Security Focus 2006, pp. 10-31

Ruggeri Stevens, G and McElhill, J 2000, *A qualitative study and model of the use of e-mail in organisations,* Internet Research: Electronic Networking Applications and Policy Volume 10, Number 4 2000, pp. 271-283

Sands, M 2003, *Integrating the Web and Email Into a Push Pull Strategy,* Qualitative Market Research: An International Journal, Volume 6, Number 1 2003, pp. 27-37

Schirmer, AL 2003, *Privacy and knowledge management: Challenges in the design of the Lotus Discovery Server*, IBM Systems Journal, Vol 42, No 3, 2003

Sørensen, C and Gibson, D 2004, *Ubiquitous visions and opaque realities: professionals talking about mobile technologies*, Volume 6 · Number 3 · 2004, pp. 188-196, Emerald Group Publishing Limited · ISSN 1463-6697

Viégas F, Golder S, Donath, J 2006, *Visualizing Email Content: Portraying Relationships from Conversational Histories*, CHI 2006, April 22-27, 2006, Montréal, Québec, Canada

Venolia, GD, Dabbish, L, Cadiz, JJ, Gupta, A 2001, *Supporting Email Workflow*, *MSR-TR-2001-88*, Microsoft Research, Collaboration & Multimedia Group

Whittaker, S. & Sidner, C 1996, *Email Overload: Exploring Personal Information Management of Email*, Proceedings of CHI'96, Vancouver, B.C., pp. 276-283

Whittaker, S, Jones, Q, Nardi, B, Terveen, L, Hainsworth, J 2004, *ContactMap: Organizing Communication in a Social Desktop,* ACM Transactions on Computer-Human Interaction, Vol. 11, No. 4, December 2004, pp. 445–471.

United States Court Of Appeals For The Ninth Circuit, 2008, *QUON V. ARCH WIRELESS*, retrieved on 23/June/2008 from:

www.ca9.uscourts.gov/coa/newopinions.nsf/D2CDDB4098D7AFB28825746C0048E
D24/$file/0755282.pdf

Zack, M 1999, *Managing Codified Knowledge*, Sloan Management Review, Volume
40, Number 4, Summer, 1999, pp. 45-58

Zhang, D, Gatica-Perez, D, Roy, D, Bengio, S 2006*, Modelling Interactions From
Email Communication,* ICME 2006

Zhou, Y, Hongyuan, S, Hongyuan, Z, Zhang, Y 2005, *Towards Discovering
Organizational Structure from Email Corpus,* Proceedings of the Fourth International
Conference on Machine Learning and Applications (ICMLA'05)

APPENDIX 1 SEMI-STRUCTURED INTERVIEW

(1) What is your position within the company?

(2) When looking for information, how frequently would you perform the following search? (Yearly, Monthly, Weekly Daily)

Search Company Intranet: Yearly ☐ Monthly ☐ Weekly ☐ Daily ☐

Search your local disk: Yearly ☐ Monthly ☐ Weekly ☐ Daily ☐

Search Network drives: Yearly ☐ Monthly ☐ Weekly ☐ Daily ☐

Search your Mailbox: Yearly ☐ Monthly ☐ Weekly ☐ Daily ☐

Contact a Colleague: Yearly ☐ Monthly ☐ Weekly ☐ Daily ☐

Other?

(3) Do you save email messages you receive? Yes ☐ No ☐

If yes ,why? For example,
Information contained?
Traceability?

(4) If you save email messages:

How long do you save email?

Days ☐   Months ☐   Years ☐   Indefinite ☐

What determines how long you store your emails for?

Storage Space Yes ☐   No ☐

Usefulness of Information   Yes ☐   No ☐

What determines how long the information in messages retains its usefulness?

Do you file messages in folders– Always ☐   Sometimes ☐   Never ☐

How many folders?

Do you find it difficult to find the correct folder when filing? Yes ☐   No ☐   N/A ☐

Do you use features like Microsoft Outlook Categories to organise mail? Yes ☐   No ☐

What is the size of your mailbox file?

(5) If you search email for information, what type of information, for example?

    (a) To find out how to do something – procedures, workarounds?

    (b) Information about something –part numbers, specifications?

    (c) To gain a deeper understanding something or the reasons why a device, system or organisation works as it does?

    (d) Social – would you search your email archive to find a contact or locate expertise?

(6) If you search your email archive for information, what type of search would you consider most useful?

    Search by keyword within email message (1-3)

    Search by Keyword within email attachments (1-3)

    Search by Sender name (sort) (1-3)

(7) Do you consider email to be better or worse than face to face communication for transferring information/knowledge?

- What would you consider emails   advantages over face to face communication?

- What would you consider its disadvantages?

- Do you use Instant Messaging to ask a colleague for information?

- Do you consider that Instant Messaging has any advantages over email?

(8) Sharing information within email archives:

- Do you think that your email archive contains information that could be beneficial if shared with others in the organisation?

  Strongly disagree 1 2 3 4 5 Strongly agree

- Do you think you obtain useful information if others in the organisation shared information in their email archives with you?

  Strongly disagree 1 2 3 4 5 Strongly agree

- Do you think you could obtain useful information in the email archives of employees that have left the company?

  Strongly disagree 1 2 3 4 5 Strongly agree

  Can you give a reason for your answers?

(9) If you consider that the organisation would benefit by sharing the information within mailboxes, can you suggest what you consider the best way to share that information?
Example

- Single central repository or the ability to search email archives on the users PC?

- Single folder or multiple folders?

- Email added automatically to a central repository or user chooses information to send?

- Would there be any particular useful way for presenting the information?

- Any suggestions for the level of access to this information?

- What type of search would be most beneficial?

(10) Would you be willing to allow your mailbox to be searched for information by?

- Your supervisor?
  Yes ☐   No ☐   N/A ☐

- Set of previously agreed colleagues
  Yes ☐   No ☐   N/A ☐

- Set of colleagues that may change without agreement (team changes)

  Yes ☐   No ☐   N/A ☐

- Anyone in the organisation?
  o Yes ☐   No ☐   N/A ☐

(11) Do you have any concerns about allowing others to search your mailbox ?

Privacy? Yes ☐   No ☐   N/A ☐

Information out of date, incomplete, incorrect? Yes ☒   No ☐   N/A ☐

Other concerns?

(12)   Can you identify other means that would improve knowledge sharing within the organisation?

Wikis?

Weblogs?

Other?

A survey to investigate knowledge capture using email archives.

Survey as part of an MSc in Computing

(Knowledge Management)

Dublin Institute of Technology

By Frank Wedgeworth

I invite you to complete this short questionnaire as part of my MSc in Knowledge Management. Your input is much appreciated.

This survey investigates the feasibility of capturing knowledge through the use of email archives.

Please answer all questions by ticking the box that best represents your viewpoint. Where further detail is required please write on the space provided.

(1) What is your job title? _____

(2) When looking for information, how frequently would you perform the following search? (Yearly, Monthly, Weekly, Daily)

Search Company network: Yearly ☐ Monthly ☐ Weekly ☐ Daily ☐
(eBrite, TeamCentre)

Search your local disk: Yearly ☐ Monthly ☐ Weekly ☐ Daily ☐

Search Network drives: Yearly ☐ Monthly ☐ Weekly ☐ Daily ☐

Search your email: Yearly ☐ Monthly ☐ Weekly ☐ Daily ☐

Contact a Colleague: Yearly ☐ Monthly ☐ Weekly ☐ Daily ☐

Search Paper Files: Yearly ☐ Monthly ☐ Weekly ☐ Daily ☐

Search the Internet: Yearly ☐ Monthly ☐ Weekly ☐ Daily ☐

Other: Yearly ☐ Monthly ☐ Weekly ☐ Daily ☐

(3) Do you store email messages you receive?   Yes ☐   No ☐

If yes, can you identify the reason in order of priority 1-4, 1 being most important?

Information contained	
Traceability	
Habit	
Other	

(4) If you store email messages:

(a) How long do you retain them, before you delete?

Days ☐   Months ☐   Years ☐   Indefinitely ☐

(b) What do you think determines how long the information in messages retains its usefulness?

_____

_____

_____

(c) Do you file messages in folders?   Always ☐   Sometimes ☐   Never ☐

How many folders? _____

(e) Do you find it difficult to find the correct folder when filing? Yes ☐   No ☐   N/A ☐

(f) What is the approximate storage space required for your archived email?

< 500 MB ☐   500MB - 1GB ☐   > 1GB ☐   Don't know ☐

(5)   Would you consider an email archive a useful source of knowledge for the following?

(a) To find out how to do something – procedures, workarounds?

Totally Useless	Somewhat Useless	Neutral	Somewhat useful	Very Useful
☐	☐	☐	☐	☐

(b)To find information about something; e.g. –part numbers, specifications?

Totally Useless	Somewhat Useless	Neutral	Somewhat useful	Very Useful
☐	☐	☐	☐	☐

(c)To gain a deeper understanding about something, or the reasons why a device, system or organisation works as it does?

Totally Useless	Somewhat Useless	Neutral	Somewhat useful	Very Useful
☐	☐	☐	☐	☐

(d) To find a contact or locate expertise?

Totally Useless	Somewhat Useless	Neutral	Somewhat useful	Very Useful
☐	☐	☐	☐	☐

(e) Can you identify other type of knowledge that can be found in email archives ?

_____

_____

_____

_____

(6) If you search your email archive for information, what type of search would you consider most useful to you? Prioritise 1 -3, 1 being most useful.

Search by keyword within email message	
Search by Keyword within email attachments	
Search (or sort)by persons name	

I don't search ☐

(7) In searching an email archive for knowledge, what would you consider the best method for presentation? Prioritise 1-3, where 1 is most preferred.

Return a list of emails that match search criteria sorted by relevance ,date, or sender	
Generate a report presenting a summary and analysis of information	
Provide a visual display showing where information is located (by sender or folder)	

I don't search ☐

(8) If a folder for sharing archived emails was made available:

(a) Would you be willing to add emails you consider useful to the folder as you compose and send them?

Yes ☐ No ☐

(b) Would you be willing to add any of your existing archived email to this central folder?

Yes ☐ No ☐

(c) Do you think you would search this folder for information?

Yes ☐ No ☐

(d) Can you suggest from the list below what you consider the best format for organising a central email folder? Prioritise 1-3, where 1 is most preferred.

Single folder –users simply drop emails into the folder. Access is controlled by a central user.	
Multiple Folders organised by Product. Access controlled by a central user.	
Multiple folders organised by User. Access controlled by individual users.	

(9)   Can you identify other means that would improve knowledge sharing within the organisation?

Wikis   Yes ☐   No ☐   Don't know ☐
Weblogs   Yes ☐   No ☐   Don't know ☐
Instant Messaging   Yes ☐   No ☐   Don't know ☐
Other   _____

(10) Have you any other comments or suggestions?

_____
_____
_____
_____

Thank you for your time.

If you have any questions please contact Frank Wedgeworth on ext 207

Please return completed questionnaire to Frank Wedgeworth in Software R/D