Email and PIM: Problems and Possibilities

Steve Whittaker,
University of Sheffield,
UK

Victoria Bellotti,
PARC,
USA

Jacek Gwizdka
Rutgers University,
USA

PIM includes three areas of special importance *task management*, *personal archiving*; and *contact management*. Email plays a critical role in PIM. In this article we explain why almost every important PIM function currently involves email. Our analysis of the interrelations between email and PIM leads us to evaluate technological solutions that address problems with both and identify their limitations. We conclude with a discussion of outstanding issues email generates for PIM relating to *inter*-personal information management.

Email as the critical site for PIM

Most modern work is interpersonal rather than solitary, and email is the main conduit by which work and information are distributed. Because email serves as a distribution channel, people tend to 'live' in email, as evidenced by the sheer amount of time they spend using it, and their evaluation of its importance for everyday work (Ducheneaut & Bellotti, 2001). Its role as a conduit naturally leads to email being used for the three key PIM functions mentioned above: *task management*, *personal archiving* and *contact management* (Gwizdka 2004, Bellotti et al., 2003, 2005, Whittaker & Sidner, 1996, Whittaker, 2005).

Task management involves reminding oneself about current tasks, tracking task status and maintaining relevant information to those tasks. Because email is the conduit for work, people exploit the inbox for task management, both for reminding and to keep current task information readily accessible. They leave working information relevant to current tasks in the inbox, and even send themselves email to create a message in the inbox as a reminder and perhaps a link to useful information. They know they will frequently access the inbox to pick up new messages - leading them to be reminded about current tasks (Ducheneaut & Bellotti, 2001; Jones et al., 2001; Gwizdka, 2004, Whittaker 2005). Some users also organise working information by gathering emails that relate to current tasks into "active" folders. They return to these folders when they need to deal with information relating to that task (Bellotti et al., 2005).

Email is also used for *personal archiving*. Reference information delivered through email or information about completed tasks is often filed in email folders for future use (Bellotti et al., 2005, Whittaker, 2005). And because email is the primary work conduit it is also natural that people should use it to store *contact information* (Ducheneaut & Bellotti, 2001, Whittaker et al., 2004).

Problems arising from Email's critical role in PIM

Two problems arise from PIM functions being performed in email: *fragmentation* and *lack of direct support for PIM functions.*

- *Fragmentation* occurs because information delivered through email may be left there rather than relocated to dedicated PIM applications. Information may be left in email either because of the effort involved in relocating it to a separate application, or because users feel the information is more meaningful and accessible left in email. For example, the most salient retrieval cue for an email attachment may be the sender, and this cue is lost if the attachment is only stored in the user's file system. (Bellotti et al., 2005). In their attempt to retain such contextual information, users may be end up duplicating information, e.g. documents may be stored both in email and the file system – making it hard for users to collate information.

- *Lack of direct PIM support* arises because email was not originally designed to provide PIM functions. For example, users may look up meetings and appointments arranged using email, but email itself does not provide dedicated support for calendaring functions.

Two different techniques have been proposed to address these problems: *imperialism* and *integration.*

- *Imperialism* addresses *fragmentation* by locating all PIM in email, and provides *direct PIM support* by explicitly building PIM functions into email. This is the approach taken by Microsoft Outlook, which aims to provide task management, contact management and calendaring within a single application.

- *Integration* takes the opposite viewpoint, aiming to migrate PIM functions and information from email into dedicated applications to provide direct PIM support. It addresses *fragmentation* by making email data straightforwardly accessible to those applications. Various examples of this approach are discussed in article <cite Karger/Jones this issue>.

Addressing Email (and PIM) problems

Email Imperialism: Email as the unifying application for PIM

An obvious weakness to the approach of situating all PIM in email is that no current email client currently handles all PIM functions well (Bellotti et al., 2005, Whittaker, 2005). In the following section we explain how we might modify email to explicitly support core PIM functions of task management, personal archiving and contact management.

Task Management

Task management is a central PIM function of email. We have described how users leave task-related information in the inbox. This approach does not scale well however, when users receive large numbers of messages. Accumulation of messages in the inbox decreases the salience and accessibility of individual messages, which often get pushed out of sight by incoming items.

The strategy of placing messages into active folders, mentioned above, for batch processing has the advantage of grouping messages that can more efficiently be worked on together. This strategy only works if users develop the habit of routinely returning to inspect those folders, as most users do with the inbox.

An alternative way to support task reminding and access is to classify messages by task in the inbox itself. Classifying messages makes it easier to process tasks, because task-related information is all in one place and related items can be collapsed into a single list item. This reduces inbox clutter, increasing task salience and improving reminding. Various visualisations have been developed to detect and represent inbox tasks, including tree representations based around threads, and flat representations of all information related to a task (Gwizdka, 2004, Venolia & Neustaedter, 2003, Wattenberg et al., 2005). One limitation of these approaches however, has been a reliance on using threads to determine whether messages relate to a common task. But threads are known to be a weak indicator of tasks due to topic drift and email responding practices. For this reason Bellotti et al., (2003) developed the idea of *thrasks*, which are user-customizable collections based on threads. Users can add unthreaded items to the collection or remove them, so that a thrask represents a task collection more than just a series of messages.

Although search has been touted as a solution to task management (Dumais et al., 1996), it is only a partial one. Search can be effective for accessing information identified as being relevant to a given task. But search cannot serve to remind the user about that task, as reminding is an extrinsic rather than a user-initiated process. Indeed it may turn out that to effectively support reminding, we need new automatic methods to detect and highlight critical tasks.

Personal Archiving

Email is an important information repository. But archiving is extremely problematic for email users who currently apply three main approaches for accessing archived information: folders (containing manually classified messages), search and sort.

Manual classification into folders has the primary function of organising information to make it more accessible later, but an important secondary function is to remove messages from the inbox, improving task management by reducing inbox clutter. Unfortunately manual classification is a cognitively difficult task (Whittaker and Sidner, 1996). Users may be inconsistent in their classifications or they may forget the existence of long-term folders. As a result, a given folder may contain very different messages, or there may be duplicate folders containing very similar materials. A further problem is that users have to constantly update their folder definitions or add new folders in response to changes in their job.

These difficulties lead some users to be reticent to create folders. They attempt to finesse the filing problem by relying on search or sorting using message headers to access long-term information. Both sorting and search have limitations, however. Sorting is an indirect way to find information and defining a search query can be just as hard as classifying information in folders. And by not creating folders, both approaches promote inbox clutter, reducing the effectiveness of task management.

One alternative is assisted filing. Machine learning techniques can be used to analyse folder contents, derive folder definitions and make recommendations to users about how they might categorise inbox documents (Segal & Kephart, 1999). Although this technique has been shown to be effective in user tests, one potential limitation is the requirement that users have pre-existing folders, whereas some users do not create folders (Ducheneaut & Bellotti, 2001, Whittaker and Sidner, 1996). Another critical problem is that the technique cannot currently help users who do file their email to create *new* folders.

Contact Management

A further important task in email is the management of names and addresses associated with key contacts. While most email systems can be customised to automatically extract email addresses into the address book, other information such as phone numbers and physical addresses has to be manually extracted from messages. This process is extremely tedious and error prone. However there is much information that can be automatically extracted from email. For example it is possible to identify important contacts automatically (Whittaker et al., 2004, Wattenberg et al., 2005). Having identified them it should also be possible to automatically extract additional information, for example from signature files or web pages, which could then be used to populate address fields.

Integration: making email data accessible to existing PIM applications

Integration takes the opposite approach from imperialism, aiming to extract information from email to make it accessible from dedicated PIM applications. Article <Karger/Jones this issue> describes various automatic and user-centric integration techniques, but how well do these techniques succeed in replacing email in resituating PIM in dedicated applications?

Task Management

There is currently little use of dedicated task management tools (Whittaker, 2005). Email's use as *de facto* task manager arises in large part from its role as information conduit. Users know they will frequently access email to process new messages. They exploit this frequent access to facilitate opportunistic reminding about other outstanding tasks and to quickly identify new, as yet undefined, tasks that may appear. But such opportunistic reminding is unlikely to occur with a dedicated task manager, because users have to actively remember to access that task manager and new tasks have to be identified somehow and entered into it. It seems improbable, therefore, that

users will abandon using email for task management and rely exclusively on dedicated task managers, because they fail to support crucial aspects of task management, namely reminding and new task identification.

Personal Archiving and Contact Management

Integration is more likely to be successful for personal archiving and contact management because they are not as closely tied to the conduit function of email. However there is still evidence that email offers significant benefits for both functions, making it unlikely that users will abandon it for dedicated PIM applications. In particular, email provides important contextual information that may be lost when attachments or contacts are removed from their original email context and integrated into dedicated PIM applications.

When users try to access archival or contact information first delivered in email, they often use associative reminding based on indirect social and temporal cues they remember about the original message context: 'I know I received the message containing the information and contact name around the time I was working on the X contract with A and B' (Whittaker et al., 2004). Users exploit these cues by accessing email folders and the inbox and sorting list views by sender, date or a combination of both to triangulate retrieval. So while integration will undoubtedly help with other facets of information access for archival and contact information, the importance of associative retrieval suggests that users are unlikely to stop using email and rely exclusively on other applications.

A related point is that the content of the email message may also be an important cue for users trying to relocate information. The salient cue for retrieval may be a keyword for a topic to which the contact or attachment is related. And the message itself may contain explanatory information that assists in making sense of the contact or attachment (Bellotti et al., 2005). The conduit function of email means that useful information is often first encountered in email – suggesting that users will frequently want to relocate it in the original context (perhaps an entire thread of email) rather than through a dedicated PIM application. Again this potentially compromises the simplicity of the integration approach.

In summary, although both imperialism and integration offer distinct benefits over the current situation, neither currently offers a compelling solution to email and PIM problems. A combination of both is needed - to provide greater dedicated support for PIM within email itself, as well as improved integration between email and other PIM applications.

Inter-Personal Information Management: Email generates new challenges for PIM

In addition to being a critical site for PIM, email also presents a more complex set of problems than other PIM applications. One key difference between email and other aspects of PIM is that email is *inter*-personal, serving as a conduit for tasks that

involve two or more people (Bellotti et al., 2005, Whittaker, 2005). As discussed in article <ref Erickson this issue>, email involves group information management; email information originates from and is also owed to others, who have expectations about how that information will be acted upon. In contrast, other PIM tasks such as information seeking, or archiving involve managing self-generated or self-discovered information. Such information does not usually require a response. Email information is therefore more complex to process because:

1. Email processing decisions have direct implications for others' work. Email is a work conduit, so that failure to respond appropriately to a message may directly jeopardise another person's work. Conversely, interdependent tasks are often subject to delays due to waiting for another person with different priorities to respond. Such delays can leave messages hanging around in the inbox (or, less often, actionable folders) for extended periods, often drifting out of sight and consciousness. Users therefore have to track both obligations and message status for email information.

2. Email requires constant processing. There are two reasons for this. Failure to respond quickly to colleagues' messages can compromise their work. And the sheer volume of messages users receive means that failure to deal with incoming messages can lead to a build up of messages in the inbox - compromising its task management function. This pressure is unrelenting, as new messages constantly demand to be processed. In contrast, filing personally generated digital files, contact addresses or discovered web resources, tend to be at one's own discretion with fewer externally imposed delays or deadlines.

3. Email information may lack adequate context, making it harder to process. Much personal information is self-generated or self-discovered information that arises in the context of specific user goals and interests. In contrast, email messages may not directly relate to user goals or interests, being generated by others who with different objectives. This lack of context makes it harder to act appropriately to a message, judge its value or to categorize it.

The Future of Email and PIM

In conclusion, we have argued why email is the critical PIM application, discussing two different approaches to address the problems it raises, and a distinct set of challenges email creates. However new developments in machine learning, text processing and the semantic web may lead to radical new possibilities. The increasing capacity of systems to analyze the semantic content of text and perform ever more powerful search functions will lead to some profound changes in what it is possible to do with email and PIM. We can expect to see new systems that can:

o Anticipate the importance of email and prioritize it accordingly.
o Quickly find related messages based on various criteria.
o Provide new visualisations that allow users to view and organise information from multiple related messages.
o Propose actions based on email and make them easier to initiate.
o Detect obligations and message urgency.

- o Support flexible organization and multiple views onto a message from appropriate PIM resources and project workspaces.

Of course these give rise to important issues that need to be addressed by careful interaction design, in particular in introducing automated processes into such a critical application – where the cost of algorithmic error without human oversight is high.

At the same time, we expect little change in other aspects of email:

- o List views, because they are convenient for viewing, archiving and sorting.
- o Attachments, because messages often concern discussion of and work around other content.
- o Folders, search and sort, because, even if the system can help, people will still need multiple ways to find something.
- o Information overload, because email continues to be an easy means to copy one message to many people and ever more collaborative work processes are being moved online.

Whichever approach to solving PIM and email is widely accepted, we need to keep in mind that email is constantly evolving and has been a site of continual functionality reinvention over the past decade. As email continues to evolve, new solutions will be demanded, thus, flexibility should be the key characteristic of any solution we create.

References

Bellotti, V., Ducheneaut, N., Howard, M., & Smith, I. (2003). Taking email to task: the design and evaluation of a task management centered email tool. In Proceedings of the Conference on Human Factors in Computing System CHI'2003, New York, NY: ACM Press. 345-352.

Bellotti, V., Ducheneaut, N., Howard, M., Smith, I.E. & Grinter, R.E. (2005). "Quality Versus Quantity: Email-Centric Task-Management and its Relationship with Overload." Human-Computer Interaction, 20(1 & 2).

Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R., and Robbins, D. (2003). Stuff I've Seen: A system for personal information retrieval and re-use. In *Proceedings of SIGIR 2003*.

Ducheneaut, N. & Bellotti, V. (2001). Email as Habitat: An Exploration of Embedded Personal Information Management. Interactions, 8(5), pp. 30-38.

Gwizdka, J. (2004). Cognitive Abilities and Email Interaction: Impacts of Interface and Task. Unpublished PhD Dissertation. University of Toronto.

Jones, W., Bruce, H., & Dumais, S. (2001). Keeping found things found on the web. In Proceedings of the tenth international conference on Information and knowledge management. 119-126. ACM Press, New York.

Segal R. & Kephart J. (1999). MailCat: An Intelligent Assistant for Organizing E-Mail. In Proceedings of the Third International Conference on Autonomous Agents. 276-282. ACM Press, New York.

Venolia, G. & Neustaedter, C. (2003). Understanding Sequence and Reply Relationships within Email Conversations: A Mixed-Model Visualization. Proceedings of ACM Human Factors in Computing Systems (CHI 2003), pp. 361-368.

Wattenberg, M., Rohall, S. Gruen, D., and Kerr, B. (2005). Email Research: Targeting the Enterprise. Human Computer Interaction, Human-Computer Interaction, 20(1 & 2).

Whittaker, S., Jones, Q., Nardi, B., Creech, M., Terveen, L., Isaacs, E., Hainsworth, J. (2004). ContactMap: Organizing Communication in a Social Desktop. *ACM Transactions on Computer Human Interaction.*

Whittaker, S. and Sidner, C. (1996). Email Overload: Exploring Personal Information Management of Email. In *Proceedings of ACM CHI'96 Conference on Human Factors in Computing Systems*, 276-283.

Whittaker, S. (2005). Collaborative task management in email. Human Computer Interaction, Human-Computer Interaction, 20(1 & 2).