

# 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*. We explore how and why email is currently used to support these PIM functions, and the problems arising from using email for PIM. We evaluate new technological approaches that address these problems and identify their limitations. We conclude with a discussion of outstanding issues email generates for PIM relating to *inter*-personal information management, and discuss future trends in email and PIM.

## Email as the critical site for PIM

For many people in organisations, work is interpersonal rather than solitary, and email is the main conduit through which their work and information are distributed. They 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 (Bellotti, et al., 2005). Email’s role as conduit naturally leads it to be used for three key PIM functions: *task management*, *personal archiving* and *contact management*.

*Task management* involves reminding oneself about current tasks, tracking task status and maintaining information relevant to those tasks. Email’s conduit function leads many people to exploit the inbox for task management. They leave information about current tasks in the inbox, knowing that when they access it to pick up new messages, they will be reminded about those tasks (Jones et al., 2001; Whittaker, 2005). They even send themselves email to create a message in the inbox as a reminder and perhaps a link to useful information. Some users also organise emails that relate to current tasks into active folders, returning to these folders when they need to deal with those tasks (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 for many people, email is the primary work conduit, it is also natural that they should use it to store *contact information* (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, such as a contact manager, calendar, todo list or the user’s file system. 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. The most salient retrieval cue for an email attachment is often the sender, and this cue is unavailable if the attachment is accessed through the user's file system (Bellotti et al., 2005). In their attempts to retain such contextual information, users may end up duplicating folder hierarchies, leading related documents to be stored partially in email and partially in the file system – making it hard for them to collate information (Boardman and Sasse, 2004).

- Email was not originally designed as a PIM application, leading to *a lack of direct support for PIM functions*. For example, users may schedule meetings and appointments using email, but email itself does not provide dedicated support for calendaring functions, such as meeting reminders or scheduling repeated meetings.

Two different techniques have been proposed to address these problems: *centralization* and *information extraction*.

- *Centralization* 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.
- *Information extraction* 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

### ***Centralization: Email as the unifying application for PIM***

One weakness of the *centralization* approach is that current email clients do not handle 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**

Many users leave task-related information in the inbox to manage tasks. But this approach does not scale well when users receive large numbers of messages. As messages accumulate in the inbox, this decreases the salience and accessibility of individual messages, which often get pushed out of sight by incoming items.

The alternative strategy of placing messages into active folders has the advantage of grouping messages allowing them to be more efficiently worked on together. But this strategy only works if users develop the habit of routinely returning to inspect those folders, as most users do with the inbox.

Yet another way to support *task management* is to classify messages by task *in the inbox itself*. Classifying inbox messages makes it easier to process tasks, because related items can be collapsed into a single list item, that is seen every time the user accesses the inbox. This reduces overall inbox clutter, increasing task salience and improving reminding. Various visualisations have been developed to represent inbox tasks, including tree representations based around threads, or flat representations of 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. Threads are known to be a weak indicator of tasks due to topic drift and email responding practices. For this reason Bellotti et al., (2005) 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 PIM (Dumais et al., 2003, <this volume>), it is only a partial solution to *task management*. Search can be effective for accessing information already identified as being relevant to a given task. But it 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 (Horvitz et al., 2003).

## **Personal Archiving**

Email is an important information repository for personal information, but archiving can be problematic. Users currently apply three main strategies 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 manual classification is a cognitively difficult task, because it requires users to predict future usage contexts. As a result, users are often inconsistent in their classifications and they may also forget the existence of long-term folders. A given folder may therefore end up containing very different messages, or there may be duplicate folders containing very similar materials (Whittaker 2005). This situation is exacerbated when users have to update their folder definitions or add new folders in response to changes in their job.

One solution is assisted filing. Machine learning techniques can be used to analyze message headers and content, derive folder definitions and make recommendations to users about how they might categorize incoming inbox documents (Segal & Kephart, 1999). Although this technique has been shown to be effective in user tests, it will only work if users have already created folders, and not all users do this. And for those who *do* create folders, assisted filing cannot help them to identify and create *new* folders.

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. But sorting and search have limitations too. Sorting by

sender or date exploits users' ability to remember partial information about a message, but access using sorting is nevertheless an indirect way to find information. For those relying on search, new tools such as Stuff I've Seen (Dumais et al., 2003), Gmail™, and Google Desktop™ are vast improvements on previous email and desktop search that greatly facilitate access to archives. They also partially address *fragmentation* by accessing information from both email and the file system. Nevertheless, defining a search query can be just as hard as classifying information in folders. And by not creating folders, sorting and search promote inbox clutter - reducing the effectiveness of task management.

## Contact Management

A further important PIM task is the management of names and addresses associated with key contacts. While most email systems can be customized 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. But there is much information that can be automatically extracted from email. For example it is possible to identify important contacts automatically using message header information such as frequency, longevity of communication and likelihood of response (Whittaker et al., 2004, Wattenberg et al., 2005). Having identified contacts 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 contact address fields.

### ***Extraction: making email data accessible to existing PIM applications***

*Data extraction* takes the opposite approach from *centralization*, aiming to extract information from email, making it accessible to dedicated PIM applications, such as contact managers or todo lists. Article <Karger/Jones this issue> describes various automatic and user-centric data extraction techniques, but how well will these techniques succeed in replacing email in resituating PIM in dedicated applications?

## Task Management

There is currently little use of dedicated task management tools such as online todo lists or workflow (Blandford and Green, 2001). 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 outstanding tasks and to 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 therefore seems improbable that users will abandon using email for task management and rely exclusively on dedicated task management tools, because those tools fail to support these crucial aspects of task management.

## Personal Archiving and Contact Management

*Data extraction* 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 completely abandon it in favor of their contact managers or file systems. In particular, email provides important *contextual* information that may be lost when attachments or contacts are extracted from their original email context and integrated into dedicated PIM applications.

When users try to retrieve archival or contact information first delivered in email, they often use associative reminding based on indirect social and temporal cues such as the senders, recipients and date of the message (Whittaker et al., 2004). Users exploit these cues by accessing email folders and the inbox and then sorting to view by sender, date or a combination of both, to triangulate retrieval. The content of the email message is also 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 their file system or contact manager. Again this need for context potentially compromises the simplicity of the *data extraction* approach.

In summary, although both *centralization* and *data extraction* 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 *data extraction* from email into other PIM applications.

## Inter-Personal Information Management (IPIM): Email generates new challenges for PIM

In addition to being a critical site for PIM, email 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

We have argued why email is *the* critical PIM application, discussing two different technical approaches to address the problems it raises, and a distinct set of challenges email creates. However new developments in desktop search, machine learning, and text processing will generate new possibilities. The increasing capacity of systems to analyze text content 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:

- Provide organization at the task rather than the message level.
- Anticipate the importance of email and prioritize it accordingly.
- Detect obligations and message urgency.
- Provide new visualisations that allow users to view and organise information from multiple related messages.
- Propose actions based on email and make them easier to initiate.

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:

- List views, because they are convenient for viewing, archiving and sorting.
- Attachments, because messages often concern discussion of, and work around other content.
- Folders, search and sort, because, even if the system can help, people will still need multiple ways to find something.

- 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.

Other major new developments concern search-based email (Gmail™) and desktop search (Dumais et al., 2003, Google Desktop™). Unlike prior desktop or email search, these systems are fast and operate across entire archives and different applications. Efficient fast search will clearly promote better access to archives and contact information. However it will not tackle *task management*, as it does not support reminding, nor will it help to reduce inbox clutter in the way that filing and folders do for some current users.

A final issue is email's relation to other communication and collaboration technologies, which may partially replace certain email uses. Some organizations now routinely use instant messaging for quick conversations that previously took place in email. Others rely on blogs or wikis to distribute and comment on public information, rather than using email attachments. One important advantage of these alternatives is that they siphon information from email – reducing the overload problem. At the same time, however, they introduce problems of intrusiveness and notification that are not experienced with email.

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.E. & Grinter, R.E. (2005). "Quality Versus Quantity: Email-Centric Task-Management and its Relationship with Overload." *Human-Computer Interaction*, 20(1 & 2), 89-138.

Blandford, A., and Green, T. (2001). Group and Individual Time Management Tools: What You Get is Not What You Need, *Personal and Ubiquitous Computing*, 5, 4, 213-230.

Boardman, R & Sasse, M.A. (2004) "Stuff goes into the computer and doesn't come out": A Cross-tool Study of Personal Information Management. In *Proceedings of CHI 2004*.

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*.

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

Horvitz, E., Kadie, C. Paek, T., and Hovel, D. (2003). Models of attention in computing and communication, *CACM*, 46(3), 52-59.

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. (2005). Collaborative task management in email. Human Computer Interaction, Human-Computer Interaction, 20(1 & 2), 49-88.