Personal Information Management

William Jones

University of Washington

Introduction

Personal Information Management (PIM) refers to both the practice and the study of the activities people perform in order to acquire, organize, maintain and retrieve information for everyday use. One ideal of PIM is that we always have the right information in the right place, in the right form, and of sufficient completeness and quality to meet our current need<jones 2004> (W. Jones & Maier, 2003). Tools and technologies help us spend less time with time-consuming and error-prone actions of information management (such as filing). We then have more time to make creative, intelligent use of the information at hand in order to get things done.

This ideal is far from the reality for most people. A wide range of tools and technologies are now available for the management of personal information[1]. But this diversity has become part of the problem leading to *information fragmentation <ref Jones 04>*.. A person may maintain several separate, roughly comparable but inevitably inconsistent, organizational schemes for electronic documents, paper documents, email messages and web references. The number of organizational schemes may increase if a person has several email accounts, uses separate computers for home and work, uses a PDA or a smart phone, or uses any of a bewildering number of special-purpose PIM tools.

Interest in the study of PIM has increased in recent years with the growing realization that new applications, new gadgets, for all the targeted help they provide, often do so at the expense of increasing the overall complexity of PIM. Microsoft's OneNote, for example, provides many useful features for note-taking but also forces the use of a separate tabbed system for the organization of these notes that does not integrate with existing organizations for files, email

[1] Though now slightly dated, the following reviews provide some idea of the diversity of commercially available PIM tools:

Blandford, A., & 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

Etzel, B., & Thomas, P. (1996). *Personal information management: tools and techniques for achieving professional effectiveness*. Washington Square, N.Y.: New York University Press.

Rosenberg, S. (1999, March 5). Personal information mismanagement : why hasn't the software industry given us more tools to get our lives in order? *Salon 21st,* 6.

messages or web references. Users can rightly complain that this is “one organization too many” (R. Boardman & Sasse, 2004; R. Boardman, Spence, & Sasse, 2003).

Interest in building a stronger community of PIM inquiry is further driven by an awareness that much of the research relating to the study of PIM is also fragmented by application and device in ways that parallel the fragmentation of information that many people experience. Excellent studies focus on uses of and possible improvements to email (for example, (Balter, 2000; V. Bellotti, Ducheneaut, Howard, Neuwirth, & Smith, 2002; V. Bellotti, Ducheneaut, Howard, & Smith, 2003; V. Bellotti & Smith, 2000; Ducheneaut & Bellotti, 2001; Gwizdka, 2000, 2002a, 2002b; Mackay, 1988; Whittaker & Sidner, 1996; Wilson, 2002). Studies similarly focus on the use of the Web or specific web facilities such as the use of bookmarks or history information (for example, (Abrams, Baecker, & Chignell, 1998; M.D. Byrne, John, Wehrle, & Crow, 1999; Catledge & Pitkow, 1995; Tauscher & Greenberg, 1997a, 1997b). And studies have looked at the organization and retrieval of documents in paper and electronic form (for example,(Carroll, 1982; Case, 1986; Malone, 1983; Whittaker & Hirschberg, 2001).

Additional research efforts fit well under a “PIM umbrella” that maintains focus on people and what they want to or need to be able to do with their information. The completion of a task depends critically on certain information. For example, returning a phone call depends upon knowing the person’s first name and phone number. As such, the study of personal task management clearly relates (V. Bellotti, Dalal, B, Good, N, Flynn, P, Bobrow, D. G. & and Ducheneaut, 2004; V. Bellotti et al., 2003; M. Czerwinski, Horvitz, E and Wilhite, S, 2004; Gwizdka, 2002a; Williamson & Bronte-Stewart, 1996) to PIM. Research into “digital memories” (J. Gemmell, Bell, Lueder, Drucker, & Wong, 2002) and the “record everything” and “compute anywhere” (Dempski, 1999; Lucas, 2000a) possibilities enabled by advances in hardware also relate.

Good research relating to PIM is scattered across a number of disciplines including information retrieval, database management, information science, human-computer interaction, cognitive psychology and artificial intelligence. The author led a 3-day workshop on January 27-29, 2005, sponsored by the National Science Foundation (NSF), at which thirty researchers from these disciplines and with a special interest in PIM met to discuss challenges of and promising approaches to PIM (http://pim.ischool.washington.edu/). A common sentiment expressed at this workshop was that research problems of PIM have often “fallen through the cracks” between existing research and development efforts.

The problems of PIM gone bad

In our real world, we do not always find the right information in time to meet our current needs. The necessary information is never found or it “arrives” too late to be useful. Information may also enter our lives too soon and then be misplaced or forgotten entirely before opportunities for its application arrive.

Information is not always in the right place: The information we need may be at home when we’re at work or vice versa. It may be on the wrong computer, PDA, smart phone or other device. Information may be “here” but locked away in an application or in the wrong format so that the hassles associated with its extraction outweigh the benefits of its use. We may forget to use information even when (or sometimes because) we have taken pains to keep it somewhere in our lives. We may fail to make effective use of information even when it is directly in view.

These are failures of PIM. Some failures of PIM are memorable. Other failures may recede into a background cost of "doing business" in our world. Many of us, for example, can remember the frustration of failing to find an item of information – for example, a paper document, a digital document, an email message – that we know is "there somewhere". We may spend precious minutes, sometimes hours, in an already busy day looking for this lost information.

But even a routine day when things proceed more or less as expected is often filled with many small failures of PIM. Smaller failures may occur so often that we stop noticing them in much the same way that we may no longer notice the scuff marks on the kitchen floor or the coffee stain on a favorite shirt. These failures form a part of an "information friction" associated with our practice of PIM. A simple email request, for example, can often cascade into a time-consuming, error-prone chore as we seek to bring together, in coherent, consistent form, information that lies scattered, often in multiple versions, in various collections of paper documents, electronic documents, email messages, web references, etc.

Can you give a presentation at a meeting next month? That depends… What did you say in previous email messages? When is your son's soccer match? Better check the paper flyer with scheduled games. Does the meeting conflict with a conference coming up? Better check the conference web site to get dates and program information. What have you already scheduled in your calendar? Can you get away with simple modifications to a previous presentation? Where is that presentation anyway? Here it is. No wait. This looks like an older version that still has some silly factual errors in it. Where is the current version?? Maybe you left it on the computer at home…

The benefits of better PIM

Information is a means to an end. Not always, not for everyone, but mostly. We manage information to be sure we have it when we need it – to complete a task, for example. Information is not even usually a very precious resource. We usually have far too much of it. Even a document we have spent days or weeks writing is typically available in multiple locations (and, sometimes confusingly, in multiple versions). We manage information because information is the most visible, "tangible" way to manage other resources that *are* precious.

Herbert Simon elegantly expressed this point with respect to the resource of attention:

***W**hat information consumes is rather obvious: it consumes the attention of its recipients. Hence, a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it. -- Herbert Simon, 1971*[2]

The quote still rings true if we replace "attention" with "time", or "energy" or …"wellbeing". Certainly the nagging presence of papers representing unpaid bills, unanswered letters or un-filed documents can distract, enervate and demoralize. We can't "see" our well-being or our attention or our energy or even our time (except through informational devices such as a calendar). But we can see -- and manage -- our paper documents, our e-documents, our emails

[2] Simon, H. (1971). Designing Organizations for an Information-rich World. In M. Greenberger (Ed.), *Computers, Communications and the Public Interest* (pp. 40-41): The Johns Hopkins Press.

messages and other forms of information. It is through these personal information items that we seek to manage the precious recourses of our lives.

The payoffs for advances in PIM are large and varied:

- For each of us as individuals, better PIM means a better use of our precious resources (time, money, energy, attention) and, ultimately, a better quality to our lives.
- Within organizations, better PIM means better employee productivity and better team work in the near-term. Longer-term, PIM is key to the management and leverage of employee expertise.

Advances in PIM may also translate into:

- Improvements in education programs of information literacy (Eisenberg, 2004). Progress in PIM is made not only with new tools and technologies but also with new teachable techniques of information management.
- Better support for our aging workforce and population in order to increase the chances that our mental lifespan matches our physical lifespan.

The payoffs for better PIM may be especially large in targeted domains such as intelligence analysis or medical informatics. Better PIM may help doctors and nurses to balance a large and varied caseload. Potentially of greater importance may be PIM support for individuals undergoing long-term or sustained treatments for chronic or acute health conditions.

For example, cancer patients commonly receive a primary intervention (e.g., surgery) which is followed by subsequent therapy lasting additional weeks, months, or years. Cancer patients are frequently in the situation of managing a regimen of longer-term, outpatient care – some combination of chemotherapy, radiation therapy, hormonal therapy, additional surgical procedures – while trying to maintain their normal lives at work and at home. They are thus saddled with all normal challenges of PIM and must also manage vast amounts of new and unfamiliar information, given by range of health care professionals from a range of different organizations and departments, often only aurally, often in inconsistent forms. Moreover, patients may experience heightened, if temporary, problems with memory loss – if not a product of the treatments and operations themselves, then the product of emotional reactions (anxiety, depression) to their situations.

PIM is not new

PIM broadly defined includes the management of information going into our own memories as well the management of external information. As such, an interest in PIM-related matters is evidenced in the study of mnemonic techniques going back to ancient times (see, for example, (F. A. Yates, 1966).

However, although definitions of PIM vary (see the section “An Analysis of PIM”), they generally include as a central component, the management of external forms of information[3]. For many centuries, paper (parchment, vellum) were the primary means of rendering information in external form. As information increasingly came to be rendered in paper

[3] See Levy (2001). *Scrolling forward : making sense of documents in the digital age*. New York, NY.: Arcade Publishing) for a discussion on the document in its many forms.

documents and these increased in number, so too did the challenges of managing these documents. In his autobiography, Benjamin Franklin describes his own difficulties with the attainment of the virtue of "order": "Order, too, with regard to places for things, papers, etc., I found extremely difficult to acquire".[4]

Tools in support of the management of paper-based information were developed over time. Yates (J. Yates, 1989) notes, for example, that the vertical filing cabinet that is now such a standard (if increasingly "old-fashioned") feature of offices, home and workplace, was first commercially available in 1893. New technologies embodied in new tools periodically spark an interest in ways of expanding the human capacity to manage and process information.

The modern dialog on PIM is generally thought to have begun with Vannevar Bush's highly inspirational article "As we may think" (Bush, 1945) published as World War II was finally nearing its end. Bush recognized a problem with the sheer quantity of information being produced and with its compartmentalization by an increasing specialization of scientific discipline: "The investigator is staggered by the findings and conclusions of thousands of other workers – conclusions which he cannot find time to grasp, much less to remember, as they appear". Bush expressed a hope that technology might be used to extend our collective ability to handle information and to break down barriers impeding the productive exchange of information. Bush described a *memex* as "a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility." The memex used small head-mounted cameras to record experiences and microfilm to store these experiences, but no computer.

In 1948, Claude Shannon published "A Mathematical Theory of Communication" (C. E. Shannon, 1948) describing what came to be called a theory of information. Key to this theory is the notion that the information content of a message can be measured for its capacity to reduce uncertainty. Such a measurement introduces the notion of a receiver or recipient of a message. The information value of a message depends upon the recipient and his/her state of knowledge The message that "Bob is coming to the meeting" has no information value, for example, if its intended recipient knows this already or if the message is given to the recipient in a language s/he does not understand. In neither case does the message do anything to reduce the recipient's "uncertainty" concerning who will be attending the meeting. Although the precise definition of information with respect to uncertainty will come to be seen as overly restrictive (e.g., (Capurro. R & Hjørland, 2003; Cornelius, 2002; Omar Aftab, 2001), a larger point in the work of Shannon (and Shannon and Weaver(C. E. Shannon, & Weaver, W. T., 1949/1963) remains: The value of information is not absolute but relative to a context that includes the intentions of the sender and the current state of a recipient's knowledge.

As the computer came into commercial use in the 1950s, Newell and Simon pioneered its use as a tool to model human thought (A. Newell, Shaw, J.C., & Simon, H.A, 1958; H. A. Simon, & Newell, A., 1958). They produced *The Logic Theorist*, generally thought to be the first running AI program. The Logic Theorist at proved many of the theorems of symbolic logic in Whitehead and Russell's *Principia Mathematica* illustrating, in the process, that many feats of

[4] Franklin, B. (1790). The Autobiography of Benjamin Franklin: Dover Thrift Editions.

intelligent problem-solving in people can be modeled on a computer using collections of elemental "if-then" production rules.

The computer of the 1950s was also an inspiration for Donald Broadbent's development of an *information processing approach* to human behavior and performance (Broadbent, 1958). By analogy to standard stages of information processing on a computer, people input information via their eyes, ears and other sensory organs, people store and process this information internally, and people output the results of this processing via their motor organs including hands and mouth. Information processing diagrams describing human behavior often resembled the flow charts used to describe a computer program.

If 1950's research showed that the computer, as a symbol processor, could "think" (to varying degrees of fidelity) like people do, the 1960's saw an increasing interest in the use of the computer to help people to think better and to process information more effectively. Working with Andries van Dam and others, Ted Nelson (who coined the word "hypertext", (Nelson, 1965) developed one of the first hypertext systems, The Hypertext Editing System, in 1968 (CARMODY, 1969). Douglas Engelbart also completed work in 1968 on a hypertext system called NLS (D. a. E. Engelbart, W., 1994 (video made in 1968)). Engelbart advanced the notion that the computer could be used to augment the human intellect (D. C. Engelbart, 1961, 1963). As heralded by the publication Ulric Neisser's book *Cognitive Psychology* (Neisser, 1967*)*, the 1960's also saw the emergence of cognitive psychology as a discipline in its own right focused primarily on a better understanding of the human ability to think, learn and remember.

The phrase "Personal Information Management" was itself apparently first used in the 1980's (Lansdale, 1988) in the midst of general excitement over the potential of the personal computer to greatly enhance the human ability to process and manage information (e.g. (Goldstein, 1980; Jeff Johnson 1989; W. Jones, 1986a, 1986b, 1988). The 1980's also saw the advent of so-called "PIM tools" that provided limited support for the management of such things as appointments and scheduling, to-do lists, phone numbers, and addresses. A community dedicated to the study and improvement of the human-computer interaction also emerged in the 1980's ((S. K. Card, Moran, T.P. and Newell, A., 1983; Norman, 1988).

With the 1990's came the Web and access, at ever increasing bandwidths, to increasing amounts of digital information – both traditional (text and graphics in books and articles) and novel (music, photographs, videos). The use of email also became commonplace in the 90's. The amount of storage space available to the average computer user increased dramatically, making it possible, for the first time, contemplate the use of the computer not only as a space for working information but also as a place to store information longer-term. Laptop computers approached desktop computers with respect to many features that users care about (speed, disk capacity, screen resolution) while decreasing in size and weight making it possible for many users to embrace the laptop as portable alternative to (and replacement of) their desktop computer.

A renewed interest in PIM

The past few years have seen a revival of interest in PIM[5] – not only as a "hot topic" but as a serious area of inquiry focusing the best work from a diverse set of disciplines including cognitive psychology, human-computer interaction, database management, information retrieval and library and information science.

Renewed interest in PIM is double-edged. On one side, the pace of improvements in various PIM-relevant technologies gives us reason to believe that earlier visions of PIM may actually be realized in the near future. Digital storage is cheap and plentiful. Why not keep a record of everything we have encountered? Digital storage can hold not only conventional kinds of information but also pictures, photographs, music – even films and full-motion video. Better search support can make it easy to pinpoint the information we need. The ubiquity of computing and the miniaturization computing devices can make it possible for us to take our information with us wherever we go and still stay connected to a still larger world of information. Improvements in technologies of information input and output (e.g., better voice recognition, voice synthesis, integrated displays of information) can free us from the mouse, keyboard and monitor of a conventional computer.

This is all very exciting. But, on the other side, renewed interest in PIM is spurred by the awareness that technology and tool development, for all their promise, invariably create new problems and sometimes exacerbate old problems too. Information that was once in paper form only is now scattered in multiple versions between paper and digital copies. Digital information is further scattered into "information islands" each supported by a separate application or device. This "other side" to renewed interest in PIM recognizes that new tools, new applications – for all the targeted help they provide – can still end up further complicating a person's overall information management challenge.

The plan for this chapter

The remainder of this chapter will cover the following topics, each in its own section.

A. **An Analysis of PIM** begins with a selective review of proposed definitions of PIM and associated concepts including the concept of *information* itself and the concepts of *personal information,* the *information item* (or information object) as an external encapsulation of information to be found, kept, used, re-used, etc. and *personal information collections (PICs)*.

B. **Research to Understand How People Do PIM, the Problems They Encounter and the Support They Need** reviews research squarely focused on PIM and also reviews a sampling from a much larger collection of PIM-related research. Relevant research is drawn, in particular, from the fields of human-computer interaction, information science and cognitive psychology. *Information fragmentation* appears throughout and in several different guises as a major, perhaps the major, problem of PIM.

---

[5] For example, at CHI'2004 there were 10 full papers (out of 93), 5 short-papers, and 4 posters focused on PIM-related topics. At CHI'2005 there were 9 full papers (out of 93), 5 short-papers or posters and 1 doctoral consortium presentation focused on PIM-related topics.

C. **Approaches towards a more integrative practice of PIM** includes a sampling of computer-based tool building efforts that show special promise in addressing PIM challenges. Some discussion is also given to techniques of PIM and to teachable strategies of PIM.

D. **Methodologies of PIM Inquiry** discusses some of the special challenges associated with the conduct of PIM fieldwork and with the evaluation of PIM tools and techniques. These circumstances favor more open-ended, qualitative methodologies of inquiry and evaluation. There is, nevertheless, a need for *reference tasks* (Whittaker, Terveen, & Nardi, 2000) to provide more quantitative, reproducible bases for the comparison of PIM tools.

E. **Theoretical Approaches and Conceptual Frameworks of Potential Relevance to PIM** looks at several theories that may help to deepen our understanding of PIM including 1. Cognitive Work Analysis. 2. Activity Theory. 3. Information foraging.

F. **The Future of PIM** concludes the chapter with a discussion of key challenges to be faced if PIM, as a field of inquiry, is to make significant progress. The section also explores some the synergies that wait to be realized between the study and support of PIM and the study and support of group information management and knowledge management (both at personal and organizational levels). This section also indulges in some fanciful speculation including PIM scenarios from 10 years hence.

Some disclaimers

- PIM is a broad. A review of relevant research, even as this pertains to core activities of PIM, is necessarily selective and cannot, given the space limitations of this chapter, be complete.
- In particular, this review provides only a cursory treatment of privacy and security. A more complete discussion necessarily involves considerations of law and policy (public, corporate and personal) that go well beyond the scope of the current chapter with its orientation towards empirical inquiry and technological remedies.

An Analysis of PIM

A deeper understanding of what PIM is, at its core, and at its broad periphery of overlap with other fields of inquiry, begins with consideration of definitions for PIM and associated concepts. This section then explores a conceptual framework that helps to connect several key concepts of PIM for purposes of the chapter's review. The section concludes with some discussion of what PIM is and is not in comparison to other fields of inquiry including cognitive science and human-computer interaction.

Some working definitions

Definitions offered here are working in their intended primary purpose to further the chapter's exposition. It is recognized that alternate, often better, definitions can be formulated for each concept and it is quite beyond the scope of this chapter to consider these alternatives.

Information and the information item

The statement above holds in particular for "information". A complete treatment of what information "is" has been a repeated topic for review in its own right. For example, Cornelius (Cornelius, 2002) and Capurro and Hijorland (Capurro. R & Hjørland, 2003) both provide excellent reviews in previous editions of ARIST.

In the context of this chapter's review of PIM, we focus on the capacity of information to affect change in our lives and in the lives of others. The information we receive influences the actions we take and the choices we make. We decide, for example, which of several hotels to book depending upon the information we are able to gather concerning price, location, availability, etc. Incoming information helps us to monitor the state of our world. Did the hotel send a confirmation? What about directions?

We also send information to affect change. We send information in the clothes we choose to wear, the car we choose to drive, and in the way we choose to act. We send information (often more than we intend) with every sentence we speak or write. It is with respect to the information we send, that it is most clearly necessary to go beyond Shannon's original notions of information as a collaborative exchange between sender and recipient. As Machiavelli might have said, we send information to serve our own purposes. Certainly one of these purposes is to be helpful and inform others. But we also send information to persuade, convince, impress and, sometimes, to deceive.

An ***information item*** is a packaging of information. Examples of information items include: 1. paper documents. 2. electronic documents and other files. 3. email messages. 4. web pages or 5. references (e.g., shortcuts, alias) to any of the above. Some might prefer to use the term "information object" to emphasize the point that an information item can be acted upon. Items encapsulate information in a persistent form that can be created, stored, moved, given a name and other properties, copied, distributed, deleted, moved, transformed, etc.

The support that we depend upon for our interaction with paper-based information items includes our desktop, paper clips, staplers, filing cabinets, etc. In our interactions with digital information items we depend upon the support of various computer-based tools and applications such as an email application, the file manager, a web browser, etc. The "size" of current information items is partly determined by these applications. There are certainly situations in which some of us might like an information item to come in smaller units. A writer, for example, might like to treat paragraphs or even individual sentences as information items to be re-accessed and combined in new ways (e.g., (Johnson, 2005) .

An information item has an associated ***information form*** which is determined by the tools and applications that are used to name, move, copy, delete or otherwise organize or assign properties to an item. The most common forms we consider in this chapter are paper documents, e-documents and other files, email messages and web bookmarks.

It is striking to consider how much of our interaction with the world around us is now mediated by information items. We consult the newspaper or, increasingly, a web page to read the headlines of the day and to find out what the weather will be like (perhaps before we even bother to look outside). We learn of meetings via email messages. We receive the documents we are supposed to read for this meeting via email as well.

On the sending side, we fill out web-based forms. We send email messages. We create and send out reports in paper and digital form. We create personal and professional web sites. These and other information items serve, in a real sense, as a proxy for us. We project ourselves and our desires across time and space in ways that would never have occurred to our forbearers.

Another point concerning information items, in contrast, for example, to what we hear or see in our physical world, is that we can often defer processing until later. We can, and do, accumulate large numbers of information items for a "rainy day". This is quite unlike, for example, the scenarios of situation awareness where acceptable delays in processing information are measured in seconds (Durso, 1999).

Personal information

Personal information has several senses:

1. The information people keep for their own personal use.
2. Information about a person but possibly kept by and under the control of others. Doctors and health maintenance organizations, for example, maintain health information about us.
3. Information experienced by a person even if this information remains outside a person's control. The book a person browses (but puts back) in traditional library or the pages a person views on the Web are examples of this kind of personal(ly experienced) information.

This chapter is primarily concerned with the first sense of "personal information". However, we consider the 2nd sense of "personal information" in the context of an all-too-brief discussion of privacy and security. We consider the 3rd sense of personal information briefly as part of a later discussion of effort to personalize a person's experience of the web and web search.

A Personal Space of Information

A personal space of information (PSI) for a person includes all the information items that are, at least nominally, under that person's control (but not necessarily exclusively so). A PSI contains a person's books and paper documents, email messages (on various accounts), e-documents and other files (on various computers). A PSI can contain references to web pages. A PSI also includes applications, tools (such as a desktop search facility) and constructs (e.g., associated properties, folders, "piles" in various forms) that support the acquisition, storage, retrieval and use of the information in a PSI.

A few other things to note about a PSI:

- Although we have some sense of control over the items in a PSI, this is partly illusory. For example, an email message can be deleted so that it no longer appears in an inbox. However, the message is very likely still around somewhere (as some figures in the public eye have learned to their chagrin).
- A PSI does not include the web pages we have visited but does include copies we make (or that are cached on our computer) and the bookmarks we create to reference these pages.
- Does PSI include our own internal memories? On the one hand, the answer must surely be yes. What could be more personal? No one else owns our memories but us. But, paradoxically, an argument can be made for "no". How much control do we have over what goes into our memories? Or what comes back out? Some things lodge in our minds even

though we wish they would not. We can never completely forget what has become a memory, i.e. we cannot simply press a "delete" key.

- In general, there are large unavoidable grey areas. For example, the files we place on a network share can be considered a part of our PSI even though they may not be under our exclusive control. Similarly, a PSI includes the many icons that applications like to leave on our computer desktops and the bookmarks and folders that are automatically created as well.
- A PSI is, by definition, "everything". We each have only one PSI.
- A PSI is distinguished from a Personal Information Environment (PIE) which, as used in the literature (Goldstein, 1980; Malone, 1983), commonly refers to a subset of a PSI in combination with supporting tools. The physical space of an office including papers piled and filed, the stapler, filing cabinets, etc. is a PIE. A notebook computer is a PIE. A person can have several PIEs.
- The size of our PSI continues to grow, especially with respect to digitally encoded information. The PSI is a potential source of information for use a number of different ways. The PSI might be used, for example, to customize our experience of the Web (see the section below on finding/re-finding). The information of a PSI might be "mined" to extract important patterns in our information (and our interactions with this information). Effective re-use of the information in the PSI promises to improve our productivity. At the same time, the growing size of our PSI also raises serious questions of privacy and security (see section on managing privacy and security below).

Personal information collections

Several researchers have discussed the importance of collections in managing personal information. Karger and Quan (Karger & Quan, 2004) define a collection quite broadly to include a variety of objects ranging from menus to portals to public taxonomies. Boardman (Richard Boardman, 2004) defines a collection of personal information to be "a self-contained set of items. Typically the members of a collection share a particular technological format and are accessed through a particular application".

In this chapter, it will be useful to introduce the related concept of a Personal Information Collections (PIC). Characteristic features of a PIC will be listed here but no attempt will be made to provide a formal definition. A PIC might best be characterized as a personally managed subset of a PSI. PICs are "islands" in our PSI where we have made some conscious effort to control both the information that goes in and how it is organized. PICs can vary greatly with respect to the number, form and content coherence of their items Examples of a PIC include:

- The papers in a well-ordered office and their organization including the layout of piles on a desktop and the folders in filing cabinets.
- The papers in a specific filing cabinet and their organizing folder (where, perhaps, the office as a whole is a mess).
- Project-related information items that are initially "dumped" into a folder on our notebook computer and then organized over time.

- A carefully maintained collection of bookmarks to useful reference sites on the Web and their organizing structures.
- An EndNote database of article references including custom properties added by the user.[6]

A PIC includes not only a set of information items but also their organizing representations including spatial layout, properties and containing folders. The items in a PIC will often be all of the same form – all email messages, for example, or all files. But this is not a necessary feature of a PIC. Later we review research efforts aimed at supporting an integrative organization information items, regardless of form. Put another way, efforts aim at building a "form-neutral" layer of support for the management of information items.

The concept of a PIC will prove useful later as we review research on the ways people approach the organization of their information. When we complete a statement like "I've got to get my ____ organized!" it is often with respect to a PIC. The organization of "everything" in our PSI is a daunting, perhaps impossible, task. But we can imagine organizing… our Favorites, our email inbox, our laptop filing system (but probably only selected areas), etc.

Definitions of Personal Information Management

PIM is easy to describe and discuss. We all do it. We all have first-hand experiences with the challenges of PIM. But PIM is much harder to define. PIM is especially hard to define in ways that preserve focus on essential challenges of PIM.

Lansdale (Lansdale, 1988) refers to PIM as "the methods and procedures by which we handle, categorize, and retrieve information on a day-to-day basis". Bellotti (V. Bellotti et al., 2002) describes PIM as "the ordering of information through categorization, placement, or embellishment in a manner that makes it easier to retrieve when it is needed".

Barreau (D.K. Barreau, 1995) describes PIM as a "system developed by or created for an individual for personal use in a work environment". Such a system includes "a person's methods and rules for acquiring the information …, the mechanisms for organizing and storing the information, the rules and procedures for maintaining the system, the mechanisms for retrieval, and procedures for producing various outputs".

Boardman (2004) notes that "Many definitions of PIM draw from a traditional information management perspective – that information is stored so that it can be retrieved at a later date". In keeping with this observation, and as exemplified by Barreau's definition, we might analyze PIM with respect to our interactions with a large and amorphous PSI. From the perspective of such a store, the essential operations are input, storage (including organization) and output.

In rough equivalence to input-storage-output breakdown of actions associated with a PSI, the conceptual framework that is used in this chapter to help organize its discussion of PIM-related research will provide the following grouping of essential PIM activities:

- **Keeping** activities affect the input of information into a PSI.
- **Finding/re-finding** activities affect the output of information from a PSI.

[6] In personal communication, one researcher told me she uses 12 separate custom properties and "lives by" her EndNote database.

- **"M-level" activities** (e.g., "m" for "mapping" or for "maintenance and organization") affect the storage of information within the PSI.

These activities are described more fully in the next section.

PIM activities to map between information and need

The remainder of this chapter's content and organization are guided by a framework that derives from a basic assumption concerning PIM activities:

PIM activities are an effort to establish, use and maintain a mapping between information and need.

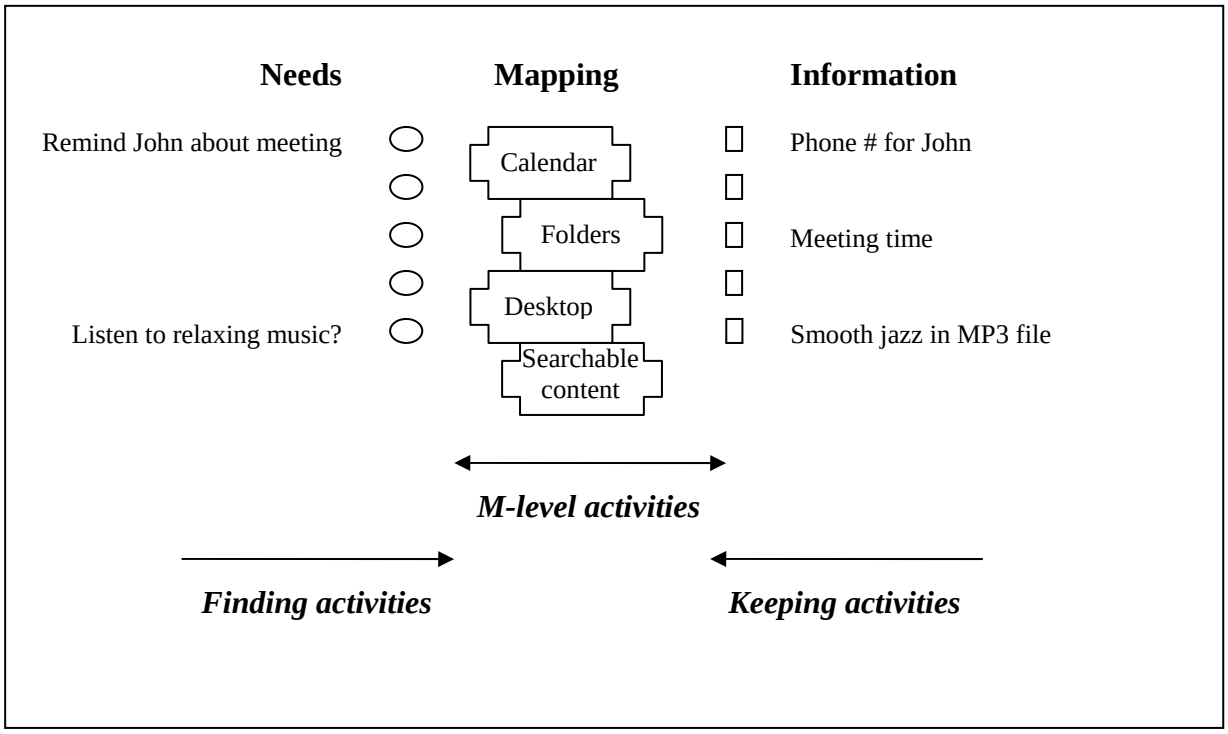

Figure 1. PIM activities viewed as effort to establish, use and maintain a mapping between needs and information

This simple statement can be expanded with reference to the diagram of Figure 1. Examples of information as listed in the rightmost column are expressed in various ways – as aural comments from a friend or colleague, as a billboard we see on the way to work or a message we hear over the radio and via any number of information items including documents, email messages, web pages and, even, hand-written notes.

Needs, too, as depicted in the leftmost column, can be expressed in several different ways: The need may, more or less, come from within us as we recall, for example, that we need to make plane reservations for an upcoming trip, or it may come via the question of a colleague in the hallway or a boss's request. Needs are often themselves packaged in information items such as email messages and web-based forms.

Connecting between need and information is a mapping. Only small portions of this mapping have an observable external representation. Large portions of the mapping are internal to our own memories – memories for specific experiences with information, experiences with

information sources and kinds of information and, more broadly, our memories for the fabric of the world around us, its conventions, its "languages" – it all goes into the mapping. Large portions of the mapping are potential and not realized in any form – external or internal. A sort function or a search facility, for example, has the potential to guide us from a need to desired information.

But parts of the mapping can be observed and manipulated. The folders of a filing system, digital or paper-based, the layout of a desktop, physical or virtual, the choice of file names and other properties for information items – each forms a part of an observable fabric helping to knit need to information.

PIM activities can be grouped, with reference to Figure 1, according to whether the initial focus is on a need, information or the mapping between need and information:

From need to information

find(need) -> information

We have a need. We try to find information to meet that need. Needs can be large and amorphous – the need for information for a review, for example – or small and simple – the need for a phone number of someone we wish to call. As noted above, a need itself, often comes packaged in an information item – an email request, for example, or via a web-based form requesting certain information for its completion. Many needs correspond to tasks (e.g., "prepare for the meeting", "answer my boss's email", "return the client's call"). But other needs may not fit tasks except by the broadest definition (for example, "see that funny web site again" or "hear 'Five to one' for old time's sake").

In our efforts to meet a need, we seek. We search. We browse. We scan through a results list or the listing of a folder's contents in an effort to recognize information items that relate to a need. Especially important, we remember to look in the first place. Sometimes, the information comes from our PSI and is information we've used many times before. Other times, the information comes from the Web and is new to us.

These activities are all referred to in this chapter as ***finding*** activities. "Finding" places the emphasis on the outcome – information meeting the need is "found" -- rather than the process. "Finding" includes "re-finding". We may repeat many of the same steps in an act of re-finding that we took to find the information in the first place. If a web search worked the first time, we may use much the same search with the same search terms a second time, for example. Finding is meant to include various information seeking activities.

Finding also applies, of course, to activities that target physical objects in our world and, as such, invites some interesting comparisons between our physical and digital worlds. We try to find a can opener that is, we believe, somewhere in our kitchen. Or we may look in our closet for a pair of shoes that we want to wear to a dinner party.

To be sure, there are differences between digital information items and physical objects. Information items may occupy a virtual space, but such a space cannot, yet, compete with the richness of our physical spaces. On the other hand, we can search for digital information using computer-based tools in ways that we cannot (yet) use for the search of physical objects.

But there are also many similarities between the finding of a physical object and the finding of an information item, digital or paper-based. We can fail to find an ingredient – walnuts, for

example – that might be perfect for a salad we're making for any of several reasons – each with their analog in the finding of information. The walnuts may not be on the shelves we look through in the kitchen. Perhaps they aren't in the kitchen at all. Or the walnuts may be right there in front of us on the shelf but in a container that we do not recognize. Or, in the midst of everything else we are doing to prepare dinner, perhaps we forget to look for the walnuts – this, too, is a failure of finding. Similarly, we may fail to find a web site we have bookmarked for our current project for any of several reasons. We may look in the wrong folders, or perhaps the bookmark is on another computer entirely, or we may fail to recognize the bookmark in though its listing is on the display. Or, especially in our rush to complete the project, we may forget about entirely about the bookmark.

Finding is broadly defined to include both acts of new finding where there is no previous memory of having the needed information and to include acts of re-finding. The information found can come from inside or outside a PSI. More broadly still, finding includes efforts to create information "from scratch" as in "finding the right words" or "finding the right ideas". When crafting an information item – a "simple" email response or a much longer, more structured document – we have many choices. We have choices concerning what information is referenced and from where. For example, is it faster to look for a bookmark in our PSI that points to needed information on the Web or is it faster to simply search again using a web search service? We have choices concerning how much of the item is "old" – composed with reference to, and perhaps a copy and paste from, other documents we have previously authored – and how much is "new" – coming directly from our own minds and through the keyboard without (conscious) reference to previous information. Our choices reflect an often complicated calculus of expected cost and expected benefit (e.g., (P. Pirolli & S. Card, 1999).

From information to need

keep(information) -> need

Many events of daily life are roughly the converse of finding events: Instead of having a need for which we seek information, we encounter information and try to determine what, if anything, we need to do with this information. We encounter information in many different ways and forms. We come across an interesting announcement for an upcoming event in the morning newspaper. A colleague at work may whisper news of an impending re-organization. An email may arrive with an announcement or a "for your information". While searching or "surfing" the Web for one need, we frequently encounter information that might be useful for some another need.

Decisions and actions relating to encountered information are referred to in this chapter collectively as ***keeping*** activities. Is the information at all relevant or potentially useful? Do we have an anticipated need for this information (Bruce, 2005)? We can safely ignore much of the information we encounter – the likelihood that we will need it is small and the cost of not having the information is small as well (W. Jones, 2004). Other information can be "consumed" immediately with no need to make special efforts to connect this information to need. Sport scores, weather reports, and stock market reports fall into this category.

There is then a middle area of encountered information. We may have a need for this information, but not now. We must then decide whether to keep this information and, if so, how. Even if we judge the information to be useful, we may still decide that no special action is required – perhaps because we already "have" this information somewhere in our PSI or

because we can easily return to the information, for example, by repeating the same search or the same path of hyperlinks that brought us to the information in the first place.

If we decide to keep the information we have encountered, then we must decide how. Information kept wrong may be useless when a need for it arises later on. In worst case, we may forget about the information entirely.

As an example, a salesperson gives us her business card that includes her phone number. Do we need to keep this information at all? The answer may be "no", either because we don't care to contact this person again or because we're certain we can easily access her phone number by another means – web lookup or via a friend or colleague, for example. On the other hand, we may decide this information is important enough to keep in several different ways. We may write the phone number down in a notebook or in a calendar to be sure of calling her again later. We may also enter this information into a contact database. But none of these methods of keeping may be any good to us if we're stuck in traffic and want to call her on our mobile phone to tell her we're running late to the meeting. (If only we had also entered the number into our phone…).

Keeping activities must address the multi-faceted nature of an anticipated need. When and where will we need the information? We must also assess our own habits and anticipate our own state of mind. Will we remember to look? Will we remember to look in this particular folder? Will we recognize the information? Will we even remember why we kept it?

There are additional variations in keeping that go beyond keeping documents and web references. We "keep appointments" by entering a reminder into a calendar. We keep good ideas that occur to us or "things to pick up at the grocery store" by writing down a few cryptic lines in a notebook or on a loose piece of paper. We frequently re-keep information inside our PSI. For example, as we encounter a forgotten web bookmark during a "spring cleaning", we may decide to move the bookmark to a new folder where we are more likely to notice it in the future. Or, as we comb through the documents associated with a completed project, we may decide that some of these documents still have value in connection with a new project and should either be moved to a corresponding folder or assigned a label for this new project.

A focus on the mapping between need and information

A third set of PIM activities is focused on the mapping that connects need to information. These are collectively referred to in this chapter as ***m-level*** activities. "M" as in "mapping" or "meta". "M" also as in "maintaining and organizing", "managing" (access to and distribution of the information in PSI), "measuring" (the effectiveness of a mapping and the structures, strategies and supporting tools associated with its creation, use and upkeep). And, also "M" as in "manipulating and making sense" of a PSI and its information. Each of these meanings of "M" is now described in more detail.

- ***Mapping.*** As noted earlier, only small portions of the mapping for a PSI have an observable external representation. These include the piles on a physical or virtual desktop and the folders, also either physical or virtual, of a filing system. Large portions of the mapping are internal to our own memories – memories for specific experiences with information, experiences with information sources and kinds of information and, more broadly, our memories for the fabric of the world around us, its conventions, its "language" – it all goes into the mapping. Much of the mapping is potential and not realized in any form – external

or internal. A sort function or a search facility, for example, has the potential to guide us from a need to desired information.

Note: It will generally be easier to discuss the mapping in the context of specific PICs. We'll also consider a kitchen and a clothes closet as physical counterparts to a PIC.

- ***Meta-level.*** One m-level activity is to "step back" and think about our PSI broadly or, more often, a PIC within. How should its information be structured? According to what schema? For common forms of information, this means deciding on a folder structure. But we can also expect increasing support for the organization of items by properties as well. It is at the meta level that we also consider potential utility of supporting tools that are proffered to help us. And we also consider strategies of PIM ("file everything right away", "don't file anything", "keep everything", "don't' keep any paper", etc.)

By analogy, we may think of or read about a "great new way" to organize our kitchen or our clothes closet. We may even consider a re-model that gives us more space or the purchase of a "tool" (e.g., a "drawer-design" refrigerator or stacking boxes for our clothes).

- ***Maintaining and organizing.*** We implement our "meta-level" scheme of organization through the actual creation of folders and a folder hierarchy (or through the creation of properties). Periodically, this structure needs to be updated. Some folders, for example, may no longer be needed. Some folders have grown too large and may need to be divided into subfolders. Folders may need to be moved or re-named. Information items themselves may similarly need to be deleted or moved.

By analogy, the food and utensils of a kitchen or the clothing of a closet may occasionally need to be re-distributed. We may also periodically attempt to weed out older items to be thrown out or given away.

- ***Managing*** *privacy, security and the distribution of items in a PIC.* A discussion of privacy and security brings us back again to a consideration not only of "our information" but also information "about us" and the large overlap between these two kinds of personal information. If our first reaction is to say "personal information is personal and no one else can see it" we are likely to have a later realization that some distribution of our personal information can be very useful. We want the travel agent to know about our seating preferences. We want colleagues and friends to know about our schedule. We may want close friends and family to know about our current condition if we are battling a serious illness. The increasing use of the personal web sites as a means to publish (and project) naturally brings a desire for technology that can support a "personal policy on privacy and security" that allows for finer distinctions that "everyone can access" or "no one can access". But, given this greater control, there is a need for user interfaces that can guide us in our choices and make clear their implications. This is a topic discussed briefly in a later section.

- ***Measuring*** *the effectiveness of a mapping and the structures, strategies and supporting tools associated with its creation, use and upkeep.* We must periodically ask ourselves "is it working?" Are the structures we've selected maintainable? Are the strategies we try to follow sustainable? Is this tool really helping or is it more trouble than it's worth? For paper documents, the signs that "things aren't working" are sometimes all too clear. For example, if paper documents continue to pile up in a "to be filed" stack and we never have time to

actually file these documents away, this may be a sign that our "great new organization", for all its promise, is simply not sustainable. The signs for digital information may be more subtle. As we look for efficient, accurate, objective ways to evaluate our own practice of PIM we run into many of the same problems, at an individual level, that are also in evidence for the field of PIM. We return to this topic in the next section and also in a later section on the methodologies of PIM.

- ***Manipulating, making sense of and using*** *information.* We try to understand the information and its implications for our lives. What does the information mean? How should it inform our decisions? As we look at the information for a conference, for example, we may be asking: "What kinds of clothes should I pack?". "How will get from the airport to the hotel?" "Will I arrive in time to make the welcome reception?" "If not, should I let Jim know?" "What else do I need to do to get read?" Making sense of our information often means that we literally try to arrange the information in various ways so that it is in view. Information in our field of vision is a powerful extension to our limited ability to keep "things in mind". For information in paper form, one time-honored method is to arrange paper documents on a desktop and other flat surfaces that surround the desk (perhaps even the floor). Although the virtual desktop of a computer may be inspired by the desktop metaphor, its own support for viewing and manipulation of information is much more limited. On the other hand, the computer supports news kinds of manipulation for digital information. For example, sorting through digital documents can take place in seconds vs. the minutes, hours or days required to do a comparable sort of paper documents. Actions of copy and paste or drag and drop are a much more effective means of doing a "paste up" of a digital document than old the fashioned scissors, glue and pasteboard are for doing comparable actions with paper documents.

  Nevertheless, as research reviewed in the next section makes evident, many useful m-level features of our paper world have not made the transition to our digital world. And many limitations of the paper world have. The folders, for example, that we still use to organize information items – digital and paper-based -- can also obscure. They can create barriers within a PSI not unlike the barriers Bush observed between an ever increasing number of scientific specializations: "*publication has been extended far beyond our present ability to make real use of the record. The summation of human experience is being expanded at a prodigious rate, and the means we use for threading through the consequent maze to the momentarily important item is the same as was used in the days of square-rigged ships."* (Bush, 1945). The wording in these sentences needs only slight modification to apply equally to the prodigious amounts of information we are able to store in a PSI. And we might indeed complain that the tools we have available for manipulating and making sense of our information have improved only modestly over the past two decades.

- ***Mañana***? *Or maybe tomorrow (but not today).* We might also say, jokingly but with considerable truth, that "m" stands for "maybe tomorrow but not today". The m-level activities described here are easy to postpone and to avoid altogether. None of them demand our attention in the way that an immediate need or even encountered information do. We perform activities of finding and keeping throughout a typical day. M-level activities can and are postponed for weeks on end. And then there is that messy closet…. Part of the problem is that we may prefer to pay the incremental, perhaps barely noticeable,

costs associated with the use of a poor mapping rather to suffer the certain and immediate costs of an m-level activity.

All three kinds of PIM activity are needed

One goal of PIM is to find an optimal combination of various PIM activities. As the clothing example helps to illustrate, different combinations are possible. We can follow a "no keeping" approach to the clean clothing we launder or get back from the drycleaners by simply leaving it all in a big pile and then sorting through the pile later, as needed – to dress for a dinner party for example. However, there are obvious problems with this approach: 1. It takes time to find clothing this way and this time doesn't get much shorter on successive searches. 2. Clothing kept in this way is likely to be wrinkled. By the same token if we keep information in poorly organized piles we may end up spending much more time to find items than we would have spent had we properly kept the items to begin with. We can also keep information in the wrong way so that it is not useable later on – for example, by keeping a web reference to an article that later disappears from a web site.

M-level activities are also important. Designing a good organizational scheme, implementing it and then maintaining the resulting organization over time are important activities of PIM and of the "thing management" too. An organization of clothing that's based on purchase date, for example, is not likely to be that useful. Short of this extreme lack of organizational utility are organizations that are too difficult to maintain over time or that don't make adequate provision for daily use (for example, that don't have an easily accessible place for frequently used clothing). The piles of "unkept" clothing surrounding a closet are often evidence of a failed organizational scheme. Again, analogs to paper and digital information are evident. Large, and aging, "to be filed" piles of information are frequently a sign of unsustainable organizational structure.

It should also be noted that there are grey areas of overlap between activities of finding, keeping and the m-level. For example, the act of keeping an item or, especially, the act of putting away or filing a group of items, may prompt an m-level creating of folders or even a re-thinking of organizational structure. Likewise with finding, we may be sufficiently frustrated with our efforts to find a document in a pile that we are moved to file away documents as we continue searching.

Other PIM-related activities

As noted earlier, one way to characterize activities of keeping and finding is with respect to the input and output of a PSI. We keep information in a PSI and we find (re-find) information later from the PSI. Similarly, activities of **reading** and **writing** are often characterized as activities that feed into and draw upon our own internal memory. Reading and writing are both active areas of research and represent productive meeting grounds between basic research in cognitive psychology and more applied research (e.g., in education and technical communication).

Clearly, skills of reading and writing are essential to our practice of PIM[7]. We must read, or at least skim, an encountered information item to determine what should be done with it. In our daily interactions with information items such as email or e-documents, we must also write.

But perhaps this discussion has it backwards. In a literate society, reading and writing are as essential to written communication just as listening and speaking are to aural communication. Rather than asking how reading and writing relate to PIM we might ask how our practice of PIM impacts these essential activities of communication.

There is evidence to suggest that ready access to vast amounts of digital information and the ability to view and work with this information on a variety of devices is changing our habits of reading in fundamental ways (Levy, 2001; Manguel, 1996; Marshall, 2006). For example, we may be reading in less depth and completeness, counting instead on ready and repeated encounter with the same information over time. To take the standard PIM example of an incoming email message, we may quickly scan to determine if an immediate action is required. If not, we may (or may not) perform a keeping action like setting a reminder or placing in a special folder in order to insure that the email is accessible to be read in more depth later on.

Access to large amounts of digital information is surely also changing our habits of writing. Students may be more tempted to plagiarize. On the other hand, search tools may make it easier to detect instances of plagiarism. Legitimate re-uses of information can represent a considerable savings in time. We may, for example, make small changes in a presentation previously given so that it can be used for a new audience. We then re-use the hours of work it took to put the original presentation together.

Beyond this, we may find that large portions of a document we write are sometimes more the product of "copy-and-paste" operations (from our previous writings) then they are of original writing. Certainly, management of "pasteable pieces" for re-use is a PIM activity (Johnson, 2005) and this raises several interesting questions. How do we go about deciding when to re-use and when to write from scratch? We may sometimes spend more time chasing down a paragraph we have previously written than it would have taken us to simply write a new paragraph expressing the same thoughts. Beyond this, we can wonder at what point a reliance on an increasing (and increasingly available) supply of previously written material begins to impact our creativity.

Other human activities that have long been a subject of basic research in psychology but that also have clear relevance to PIM include **problem-solving, decision-making**, and **categorization**. For example, work on project such as "plan my wedding" can be viewed as an act of problem-solving and folders created to hold supporting information may sometimes resemble a partial problem decomposition (W. Jones, Phuwanartnurak, A. J., Gill, R & Bruce, H., 2005). In turn, the decision to keep or not keep can be viewed as a signal detection task (Peterson, Birdsall, & Fox, 1954) and, as such, invites questions concerning the rationality of our keeping choices and concerning our ability to estimate costs and outcome (W. Jones, 2004). The folders people create in the course of completing a project may reflect internal *goal-derived categories* (Barsalou, 1991). Moreover, folders may reflect and support our mental

[7] There has also been some discussion concerning essential skills of PIM that might be taught (rather than learned "on the job"). See, for example, "Why Can't Johnny File? And What Can We Do to Help?", a panel discussion at ASIST04 (http://www.asis.org/Conferences/AM04/abstracts/71.html)

categorizations of information. If so, then folders and the act of filing may have significant impact on what we notice and remember about an information item (Brian H. Ross, 1999; B. H. Ross, 2000).

Results from basic research in psychology can certainly shape and inform our understanding of PIM. But PIM can also help to guide basic research. For example, an important goal of cognitive psychology is to understand how categories are learned and used. Several lines of research over past 20 years or so have attempted to move beyond the traditional study of taxonomic categories and the study of artificially-defined categories. And there has been a search for more ecologically valid alternatives to the traditional experimenter-supervised, forced-choice classification paradigm. An analysis of how people organize their information in support of a personal project and its goals could prove very useful to these efforts. It can be noted, for example, that the creation and use of folders takes place in the absence of immediate feedback (supervision) concerning their utility or the "correctness" of their use. Moreover, people have the option not to create folders or not to use them, once created, to "file away" a new information item.

Beyond this, we have reason to hope – to expect – that tools for the manipulation and organization of digital information will improve in expressive power and ease of use. If folder hierarchies will soon be obsolete as some have claimed (Dourish et al., 2000) we would hope new forms of representation support the functions of folders and then some. We should expect to get more out of a representation, once created, at less cost in our own time and aggravation. Do these representations then become a higher fidelity depiction of internal conceptualizations?

Although the discussion here and elsewhere in the chapter focuses mostly on cognitive aspects of PIM, it is clear that there are many dimensions to our interaction with information (see, for example, (B. I. J. D. Dervin, 1992; Fidel & Pejtersen, 2004; Pettigrew, Fidel, & Bruce, 2001). When people are asked to provide a "tour" of their information organizations, for example, or to describe an instance where they needed to re-locate a lost item of information, there is a clear emotional component and social component as well. People may feel angry or embarrassed. They may argue with their spouse or a co-worker.

Our information and our ability to manage this information come to represent and evoke a great deal more than is "there" in the information itself. This something to bear in mind even though most of the remainder of the chapter focuses on core activities of PIM: Finding activities, keeping activities and m-level activities.

PIM in relation to other fields of inquiry

As a final background task to complete before moving to PIM-related research, connections can be made and distinctions drawn between PIM and the following areas of inquiry:

Human-computer interaction. Much of the work reviewed here originates from practitioners in the field of human-computer interaction (HCI). However, a considerable amount of research in HCI remains focused on specific genres of application and associated forms of information, specific devices to aid the interaction and, increasingly, on group and organizational issues of HCI. In the study of PIM, the focus remains primarily on the individual but also broadens to include key interactions with information over time. PIM includes a consideration of our personal use of information in all of its various forms – including paper. Although it is difficult

to imagine a practice of PIM these days that doesn't involve computers, nevertheless, computers are not a primary focus; information is.

Human-information interaction. There has been discussion in recent years of human-information interaction (HII) in contrast to HCI (Fidel & Pejtersen, 2004; Gershon, 1995; Lucas, 2000b; Pirolli, 2006). Interest in HII is partly due to a realization our interactions with information are much more central to our lives than are our interactions with computers. This realization is reinforced by the trends towards ubiquitous computing. Success in computing and, perhaps paradoxically, in HCI may mean that the computer "disappears" (Norbert Streitz, 2005) into the backdrop of our lives much like electricity. If we get our "transparent interfaces" then we are left with our information. Much in HII remains to be defined but when this happens, it is likely that PIM will be a proper subset.

Personal knowledge management. Finally, there is sometimes discussion of personal knowledge management (PKM)[8]. Given the usual ordering of data < information < knowledge, we are tempted to think that PKM is more important than PIM. That may be so. One major challenge of PKM, just as with knowledge management more generally, is in the articulation of rules and "lessons of a lifetime" in a form that we (and possibly others) can understand. Knowledge expressed and written down becomes one or more items of information – to be managed like other information items.

PIM stands to benefit from advances in the fields of **information retrieval** and **database management.** For example, database techniques might be applied to "mine" and structure personal information (Xin Dong, 2005). Other techniques might realize essential efficiencies in support of deeper levels of unification in underlying data structures (David R. Karger, 2005; Karger & Quan, 2004). In the other direction, PIM offers these fields a very challenging area of focus.

Research to Understand How People Do PIM, the Problems They Encounter and the Support They Need

Levels of use

In their choice of external organizing structures for information, people appear to divide information items, regardless of form and content, into a small number of broad categories reflecting how soon and how often items are expected to be used. Different researchers use different terms to describe much the same categorization. For example, Cole (Cole, 1982) surmised that people interact with personal information at three different levels: 1.) "Action information" must be acted upon in the near future or has just recently been acted upon. 2.) "Personal work files" have ongoing relevance to a person's projects or areas of interest. 3.) "Archive storage" contains information of little or no relevance to ongoing projects. It is accessed infrequently and may be located somewhere distant from a person's office.

Sellen & Harper (A. J. Sellen & Harper, 2002) refer to "hot", "warm" and "cold" documents as, respectively, documents in current use, documents that will be used in the near future and other documents. Barreau and Nardi (D. K. Barreau & Nardi, 1995) refer to "ephemeral", "working",

[8] See, for example, the web site: http://www.global-insight.com/pkm/.

and "archived" information. Ephemeral information is valued in the short-term and includes to-do items and information for immediate use. Working information is valued over a longer period and may be kept for the period of a particular project over several months. Archived information is kept for the long-term but is not in day-to-day use.

People are also observed to impose a three-part division on their web bookmarks but with a slightly different meaning for each category (H. Bruce, Jones, W. and Dumais, S, 2004; W. Jones, Dumais, & Bruce, 2002): 1.) In the first category are bookmarks used repeatedly for quick access to reference material. 2.) Second come bookmarks that are used much less frequently – perhaps seasonally and only once a year (e.g., web sites pointing to tax information). 3.) Finally are bookmarks in what might be called a "check this out later" category. Associated sites were encountered but without sufficient time to examine in detail. Participants varied in their choice of location for each category. For example, two participants placed all three kinds of bookmarks in their Favorites area: Everyday bookmarks were placed at the top; less frequently used bookmarks were placed in the middle organized into folders; and "check this out later" bookmarks were placed at the bottom. For another participant, everyday bookmarks went into a "Links" area, less frequently used bookmarks went to a personal web site and "check this out later" bookmarks were placed loosely in Favorites.

The division of items into different regions (and associated spaces) with respect to access frequency can be seen as an attempt to make optimal use of the available "space" (Kirsh, 1995) and to make optimal use of attention – perhaps our most precious and limited resource.

Finding and re-finding: From need to information

Events of finding are roughly divided according to the location of the targeted information and a person's prior experience with this information as illustrated in Figure 2.

An information item is …	From a personal store	From a public store
Seen before	**A**	**B**
New	**C**	**D**

Figure 2. An event of finding can vary according to where targeted information is and whether the person is trying to find this information again.

Focus in this chapter is on events in the "A" quadrant – where a person is attempting to return to information s/he believes is in a personal space of information.

But events in all quadrants are acts of PIM. In attempting to find information to meet a need, people may range across several quadrants – looking for example, first for previously experienced information in the PSI (quadrant A), then, if this fails, looking for the same information on the Web (quadrant B) and, finally, to be complete, looking for new information on the web as well (quadrant D). The study of PIM and the application of personal information promises to decrease time and increase success rate for events in each quadrant.

B. Re-finding items from a public store

A return to items previously seen can be facilitated by the capture of context associated with a person's previous encounters with the item. When was the item encountered? In association with what other items? What task, activity, project? A person may often remember incidentals

in the context of a previous an item even if little is remembered of the item itself. Lansdale (Lansdale, 1988, 1991; Lansdale & Edmonds, 1992) argued for the capture, in usable, digital form, of contextual information to agree with the contextual associations that are made incidentally in a person's own memory of an encounter.

In the spirit of using a person's memory for previous encounters is the "History" facility of web browsers. However, in contrast to the popularity of bookmarks and the Back key, history lists are infrequently used. Several studies (M.D. Byrne et al., 1999; Catledge & Pitkow, 1995; Tauscher & Greenberg, 1997a) indicate that history lists in current browsers facilitate page access in less than 1% of page opens.

C. Finding items in the PSI

We now turn to a consideration of finding, not re-finding, as activity that may be increasingly directed towards the information in our PSI. Items may enter a personal store automatically so that a person has no memory of their existence. If these items are ever retrieved, it is through an act of finding, not re-finding. We may, for example, read only a fraction of the information in the magazines and newspapers that enter our homes in paper form via various subscriptions. Likewise, we may read only a fraction of the email that accumulates in our digital inboxes. Memories for a previous encounter with an information item may also fade so that its retrieval is more properly regarded as an act of finding rather than re-finding. Personal stores are becoming enormous. Some of their items may be more than 10 years old. As we use products such as Google Desktop Search[9], we may be surprised at the information we "have".

A step further is efforts to inform the user (unobtrusively) when items in her PSI match the text in view or being typed. One such system is called *Implicit Query* (*IQ*) (Dumais, 2004) (see also, (Budzik, 2001; M. Czerwinski, Dumais, S., Robertson, G., Dziadosz, S., Tiernan, S. and van Dantzich, M., 1999; Henzinger, 2003; Rhodes, 2000)).

D. Personalizing our experience of the Web and other public stores

There is an impressive body of work on information seeking and information retrieval that applies especially to events in the D quadrant and is beyond the scope of this chapter to review properly (see, for example,(G. Marchionini, 1995); (G. K. Marchionini, A., 1998; Rouse, 1984)).

It is important here to note that there is a strong personal component even in efforts to find new information, never before experienced, from a public store such as the Web. For example, our efforts to find information may be directed by an outline or a to-do list that we maintain in our personal space of information. Access to new information items may be via a query that we maintain in our personal space as a bookmark or even as a list of words we keep in written form (or "in mind").

An online search to meet a need for information is often a sequence of interactions rather than a single transaction. Bates (Bates, 1989) describes a "berry picking" model of online searching in which needed information is gathered in "bits and pieces" in the course of a series of steps where the user's expression of need, as reflected in the current query, evolves. Teevan et al. (J. Teevan, Alvarado, Ackerman, & Karger, 2004) notes that users often favor a stepwise

[9] *About Google Desktop Search*. (2005). from http://desktop.google.com/about.html.

"orienteering" approach even in cases where the user "knows" where the information is and could presumably access it directly ("teleport") using a well-formed query. The stepwise orienteering approach may preserve a greater sense of control and context over the search process and may also lesson the cognitive burden associated with query articulation. Examples of "berry picking" and "orienteering" suggest that a preservation of search state, in the PSI, might be useful.

There is much more that might be done to use existing personal information in efforts to find new information from a public space (J. Teevan, Dumais, S. T., Horvitz, E, 2005). For example, results might be filtered to highlight first those results that are "new" and unknown to the user. Also, analysis of the terms used in a personal information space might yield a customized thesaurus for application to improve precision and recall rates.

A. Re-finding items in a PSI

Events in the A quadrant – where a person is attempting to re-find information items in a PSI – can be broken into several steps:

1. A person must remember to look.
2. A person must find a way back to the desired information.
3. A person must recognize that the information, as displayed, is the information desired.
4. Items of information, frequently, must be assembled into a collection for purposes of completing a task.

Failure with any one of these steps can mean failure or delay in the retrieval and effective use of the information.

1. Remembering to look. Reminding is an important function of paper piles in the physical space of an office (Malone, 1983). Email messages in an inbox provide a similar function, at least until the messages scroll out of view (Whittaker & Sidner, 1996). Barreau and Nardi (D. K. Barreau & Nardi, 1995) observed that Macintosh users in their studies often placed a file on the computer desktop in order to be reminded of its existence (and pending actions). Barreau and Nardi also noted a reminding function in the dominant "location-based" search strategy observed in their sample of users. While looking through a folder in search of one file, the user may notice other relevant files and be reminded of actions to be preformed (see also (Herrmann, Brubaker, Yoder, Sheets, & A., 1999)).

People frequently do forget to look. In one study of web use, for example, participants frequently complained that they encountered web bookmarks, in the course of a "spring cleaning" for example, that would have been very useful for a project whose time had now passed (W. Jones et al., 2002). Participants had forgotten about the bookmarks and the web pages to which these bookmarks pointed. In another study, when participants were cued to return to a web page for which they "had" a web bookmark, this bookmark was used on less than 50% of the trials (H. Bruce, Jones, W., and Dumais, S., 2004; W. Jones, Bruce, & Dumais, 2003). Marshall & Bly (Marshall & Bly, 2005) and Boardman and Sasse (R. Boardman & Sasse, 2004) report similar problems of reminding. People send information to themselves (web references, for example) in email messages partly to realize a reminding function.(Jones et al. [year]).

2. Paths to the information. An effect that persists even as desktop search utilities improve is a preference for location-based finding, orienteering or, simply, browsing as a primary means to return to their personal information (D. K. Barreau & Nardi, 1995; G. Marchionini, 1995; O'Day & Jeffries, 1993; J. Teevan, 2003). In the author's own informal survey of people who have installed and use Google Desktop Search, people still express a preference for browsing – of their desktops, "My Documents" or through their folders. Over 90% of the respondents indicated that they used Google Desktop Search only as a "last resort" after other methods of return had failed.

3. Recognizing what's in front of you. When people actually name an information item such as a file, the research suggests that recognition accuracy is quite high (Carroll, 1982). High rates of recognition relate to a *generation effect* identified in research in human cognition (W. P. Jones & Landauer, 1985). Thinking of a name for an item causes people also to elaborate on connections between the name and the item. These connections persist in memory and aid in later recognition (and, to a lesser extent, recall).

However, we do not always name the information items in our PSI. Abrams, Baecker, & Chignell (Abrams et al., 1998) report, for example that when creating a web bookmark, users rarely change the default name provided by the browser At the same time, 86% of users in their survey reported that descriptiveness of bookmarks was a problem.

One powerful aid to the recognition of items in a results list returned by a search is to include excerpts from items in which matching search terms are highlighted (Golovchinsky, 1997 -a, 1997 -b). Highlighting of search terms is now a standard feature of many search facilities.

4. "Re-collecting": Finding a reasonably complete collection of information. The limited available evidence indicates that people are quite good at re-finding an information item (H. Bruce, Jones, W., and Dumais, S., 2004). However, in many instances, the need is not for a single information item but rather for several items that may be scattered in different forms in different organizations. To illustrate, we return to the example of deciding whether or not say "yes" to a request to give a presentation. Relevant information must be gathered from paper documents, previous email, an electronic calendar and previous versions of the presentation (where it is unclear which version is correct and current). The retrieval of each individual item may be a problem, but as the size of the set of relevant items increases, so too do the chances that one item may be overlooked. Of potential relevance are studies of information foraging and the notion of an "information scent" (P. Pirolli & S. Card, 1999).

Overall, finding becomes more difficult and more error prone when information is fragmented into several organizations. The reminding value of the inbox, for example, is likely to be diminished if we have several inboxes for several email accounts. Browsing to an information item becomes more difficult if the information item could be in any of several different organizations (H. Bruce, Jones, W. and Dumais, S, 2004). And certainly "re-collecting" is more difficult if the needed items are scattered across a PSI.

Keeping: From information to need

Research reviewed in this section illustrates several points concerning keeping activities of PIM:

- Keeping activities are common and important.

- An instance of keeping can be multi-faceted to reflect the multi-faceted nature of the needs for which information is being kept. Keeping is certainly about more than just saving information (in the same way that finding information is about more than just searching for information).
- The methods of keeping vary greatly.
- Keeping can be difficult, problematic and error-prone.
- Keeping is more difficult when information is more fragmented (e.g. by location, device and application).

This section concludes with a review of approaches in support that specifically target the keeping activity.

Keeping activities are common and important

We tie a string on a finger, place reminders on sticky notes, maintain to-dos in paper and electronic form, write down appointments in calendars – also both in paper and digital form. Keeping activities are as essential part of our efforts to avoid *prospective memory failures* (Ellis, 2000; O'Connail, 1995a; A. J. Sellen, Louie, G, Harris, J.E. & Wilkins, A.J., 1996; Terry, 1988).

Keeping activities are commonly triggered by at least two situations.

1. In the course of our current activity, completing a task, for example, we encounter information that may have relevance at some point in the future but not now. The ability to handle encountered information effectively is key to our ability to discover new material and make new connections (Erdelez & Rioux, 2000).

2. Also, we may be interrupted in the midst of our current task and we look for ways of "holding our place" so that we can return to this work later with minimal costs to re-start. We receive a phone call, we must attend a meeting, or we must re-boot our computer (thus losing the current arrangement of windows). In each case, we depend upon some trace of the current state as a means of reinstating the current task later on. Some useful information is often kept incidentally – a list of recently opened documents or last modified date, for example. We may keep additional information, mentally or in the notes of a "to do" list in order to aid our return. Task interruptions have been observed to occur as many as four times per hour for some professionals (O'Connail, 1995b) and this is quite possibly an underestimate.

Keeping activities can be multi-faceted to reflect the multi-faceted nature of the needs for which information is being kept

Kwasnik (B.H. Kwasnik, 1989) identified a large number of dimensions influencing the placement and organization of paper-based mail and documents in an office. Keeping behavior was influenced by document attributes (e.g., title, author), disposition (e.g., discard, keep, postpone), order/scheme (e.g., group, separate, arrange), time (e.g. duration, currency), value (e.g., importance, interest, and confidentiality) and cognitive state (e.g., "don't know," and "want to remember."). Overall, a document's classification was heavily influenced by the document's intended (anticipated) use or purpose – a finding subsequently re-produced by Barreau (D.K. Barreau, 1995).

Malone (Malone, 1983) observed that placing paper documents into piles was not necessarily (or only) a sign of disorder or laziness. Piles have several advantages over files (i.e., filing a document into a folder): 1.) Items can be placed in piles to defer the cognitively difficult job of classification required for filing ("what folder should this go in? … under what heading?") 2.) Documents in piles are more accessible. 3.) The visibility of piles provides an important reminding function that is usually not present in files (out of sight, out of mind).

Jones, Bruce and Dumais (W. Jones, Bruce, & Dumais, 2001; W. Jones et al., 2002) observed that the choice of method for keeping web information for later use was influenced by a set of different considerations or functions. Keep methods vary greatly in their ability to provide these functions. For example, a web address pasted into a self-addressed email provided important reminding function together with a context of relevance: The email arrives in an inbox which is checked at regular intervals and the email can include a few lines of text that explain the URL's relevance and the actions to be taken. Participants could also use self-addressed email messages when they wanted to be able to access the reference to a web page from both home and work. By contrast, for participants in the study, the bookmarking facility provided none of these functions: Participants could not access bookmarks at home that they made at work and vice versa. Participants could not easily, informally write notes explaining the potential relevance or use of the web page pointed to by a bookmark.[10] Most important, participants often failed to use and forgot entirely about a bookmark once created.

Marshall and Bly (Marshall & Bly, 2005) also note that the reasons for keeping information vary and are not necessarily task-related or even consciously purposeful. Some participants appeared to keep some information (e.g., newspaper clippings) for the please of expanding their collection of like items (e.g., recipes) and a few participants used the word "packrat" to describe their keeping behavior.

Great variation in the methods of keeping

In the Jones et al (2002) study participants were observed to use a diversity of methods to keep web information for later use. Participants used the bookmarking facility explicitly provided in commonly used web browsers. However, for several participants this facility had become a "mess" and fallen into disuse. (One participant opened her Favorites in a sidebar to show the observer and then quickly closed the sidebar again with much the same reaction one might have when opening a very messy closet.) Other keeping methods included printing out web pages, saving web pages to the hard drive, pasting the address for a web page into a document and pasting the address into a personal web site. One participant carefully wrote URLs into a notebook that she kept beside her computer. As already noted, several participants emailed web addresses (URLs) along with comments to themselves and to others. In a follow-on survey of 214 respondents, respondents reported on average having a weekly repertoire of just over five keeping methods (methods used at least once a week) (H. Bruce, Jones, W., and Dumais, S., 2004).

In the study, differences emerged between participants according to their workplace role and their relationship to the information they were gathering. Managers, for example, depended

[10] A common experience among participants later was then to encounter a bookmark (e.g. while "cleaning up" or "organizing") and have no idea why it was created (and little time to find out).

heavily on email to gather and disseminate information and did relatively little direct exploration of the Web. All four managers studied and two researchers as well described a keeping strategy of sending URLs to others as "for your information" (FYI). As one participant explained, he could then ask the recipient later "what do you think of the web site?" and, in doing so, strengthen a reciprocal relationship of information exchange while possibly also getting some good analysis of the web site at the same time. (And, if necessary email containing the web reference could always be located in the "sent mail" folder.) Friends and colleagues then became a part of the extended fabric to a person's own PSI. Managers in the Jones et al. study were especially adept at extending their own ability to store and process information through the exchange of email with colleagues and subordinates (e.g., "FYI" or "get back to me with your thoughts on …").

Pile-like and file-like analogues can be identified for various information forms. The folder hierarchy is the basic means of filing paper documents, e-documents, email messages and web bookmarks. Variations do, however, exist depending upon the supporting tools and applications of a form. For example, most paper document filing schemes do not exceed a depth of three or four. Folders can be re-ordered in the bookmarking facility of several web browsers but this ability is not supported the most commonly used file managers nor in the most commonly used email applications.

With respect to pile-like analogues, the email inbox provides is a path of least resistance and provides functions of accessibility and visibility. For web bookmarks, users can accept the top-level of a bookmarks folder as a least-effort, default location. However, if the bookmarks are not visible and not routinely checked, then this level lacks the reminding function of paper piles. Some users achieve higher visibility and accessibility, for a limited number of web references, by placing them in a "links" toolbar (W. Jones et al., 2002). The computer desktop is designed to provide some pile-like functions for e-documents. People may also use the top level of "My Documents" or a project-specific folder as a kind of pile-like dumping spot for information with an intent to sort through and organize this information later when there is time and when more is known about the context in which it will be used (W. Jones, Phuwanartnurak, A. J., Gill, R & Bruce, H., 2005).

The boundary between devices to file and pile frequently blur. For example, piles of paper documents can acquire a file-like meaning if these are placed in office locations (a table, a corner) reserved for information relating to a particular project. The paper documents of a pile may sometimes be organized by content or project into folders, stacked one on top of the other (Mander, Salomon, & Wong, 1992). Conversely, file folders can be placed in portable racks to have more "pile-like" visibility and accessibility. The folders of an electronic filing system may also acquire a kind of reminding value as a consequence of a person's method of access (D.K. Barreau, 1995), (D. K. Barreau & Nardi, 1995). Piles are certainly used to keep action information – information with an anticipated near-term purpose (B. H. Kwasnik, 1991). But piles may also include working and archive information. An office may include piles of documents that need to be kept but have no obvious relevance to any existing projects.

The variation in keeping methods is greater still we allow for variations in the size of the information item. Highlighting selected sentences in a document – digital or paper-based – can be regarded as a form of keeping. On a subsequent viewing of the document our attention is naturally drawn to the highlighted regions. We keep in ways that have nothing to do with a specific information item but that profoundly affect the information we have access to. We

keep appointments. We keep a daily schedule. We keep friendships and collegial relationships. We subscribe to magazines and newspapers. We have our favorite TV and radio channels. Even the routes we "routinely" take to work and to the grocery store affect the information we encounter.

The methods people choose for keeping information reflect not only their understanding of the information item itself and its anticipated situations of use but also the current situation. Also, people may decide to change the way as item is kept over time – by moving, for example, from a pile to a "file" (folder). If we're given a phone number at a party, for example, we may write it down on a napkin or even the palm of our hand, with the intent of "keeping it better" later on. Later, we may email the phone number to ourselves in order to have it accessible from both home and work computers. Or we may punch it into the directory on our mobile phone. Or we may add to a contact management database which we then print out to paper. Or … all of the above if the number is especially important and likely to be used in several different ways on several different occasions.

In summary, effective keeping means anticipating location, device and form of anticipated uses. Keeping must include support for remembering and for managing attention. Keeping is more complex than simply saving information somewhere. Just as finding is about more than just searching, keeping is about more than just saving.

Keeping can be difficult, problematic and error-prone

Several researchers have noted that filing information correctly into the right folders is a cognitively difficult and error-prone activity (Balter, 2000; Alison Kidd, 1994 ; Lansdale, 1988, 1991; Malone, 1983; Whittaker & Sidner, 1996). The same person, on alternating days, may follow different schemes of classification within the same folder hierarchy. A given folder may contain a jumble of different information items at the same time that very similar items are scattered across two or more folders. Filing depends upon an understanding both of the item being filed but also the anticipated contexts of is use (Bruce, 2005; B.H. Kwasnik, 1989). Sellen and Harper (A. J. Sellen & Harper, 2002) describe a study suggesting that 3% of the paper documents in a typical office are misfiled and that 8% are eventually lost. The act of filing also involves a potentially disruptive change in context as attention shifts from the current item itself, in the context of a current task, to a consideration of the organization into which the item is to be filed. Later, filed information is out of sight, out of mind and apt to be forgotten.

On top of all of this, filing is often not worth the trouble. Whittaker and Hirschberg (Whittaker & Hirschberg, 2001) use the phrase "premature filing" to describe a situation in which people go to the trouble to file information and then never use this information.

The decision to keep or not to keep, as well as the variations on how to keep, can be seen as signal detection tasks in which mistakes are unavoidable (W. Jones, 2004). Information not kept or not kept properly may turn out to be very important later on. Information kept may never be used. The cost of such a "false positive" (where information is falsely kept) is potentially twofold: 1.) the initial cost of keeping. 2.) an ongoing cost of clutter since the information may distract from or obscure other information that is needed. A similar twofold cost applies to information kept in the wrong way (e.g., the wrong desktop or folder): the information itself is inaccessible and it may also distract us from other, properly filed, information.

Placing (or leaving) information items in piles, as an alternative to filing, has its own problems. In Malone's study (Malone, 1983), participants indicated that they had increasing difficulty keeping track of the contents of different piles as their number increases. Experiments by Jones and Dumais (W. Jones & Dumais, 1986) suggest that the ability to track information by location alone is quite limited. Moreover, the extent to which piles are supported for different forms of information is variable, limited and poorly understood (Mander et al., 1992). The computer desktop may serve as a place to pile items for fast access or high visibility (D.K. Barreau, 1995; D. K. Barreau & Nardi, 1995). But if the desktop is often obscured by various open windows, the accessibility and visibility of its items is much reduced (Kaptelinin, 1996). The email inbox provides pile-like functions of accessibility and visibility but these functions are clearly reduced as the number of items in the inbox increases and especially for older messages that scroll out of view.

If filing is error-prone and costly and that if the ability to manage piles is limited, it is hardly surprising that people sometimes decide to do nothing at all – even for information they believe will be useful. This is especially true for web information. For example, Abrams et al (Abrams et al., 1998) report that users bookmark only a portion of the web pages they would like to re-access. A delayed cued recall study looked at how people re-find web information they find useful (H. Bruce, Jones, W. and Dumais, S, 2004; W. Jones et al., 2003). Participants used one of three "do nothing" methods (requiring no keeping activity) on over two thirds of the trials: 1.) search again (using a web-based search service). 2.) type in the first few characters of the URL for a web site and accept one of the suggested completions of the web browser. 3.) navigate to the web site from another web site. Overall participants were very good at getting back to "useful" web sites even when these sites were accessed only once or twice per year and hadn't been accessed for up to six months.[11]

Keeping is more difficult when information is more fragmented

If our information is fragmented between devices and applications we must also anticipate the form in which we will need the information. On which device? In which application? As the above example of the phone number illustrates, the number of ways to keep information has grown considerably in recent years with an increase in the number of devices and applications that we depend upon to manage our information. Paper is still very much a part of people's lives (A. J. Sellen & Harper, 2002; Whittaker & Hirschberg, 2001). In addition, we now manage electronic documents and other computer-based files, web references (as bookmarks, for example) and, of course, large numbers of email messages often in multiple accounts. We have desktop computers, laptop computers, smart phones, PDAs and ordinary notebooks.

The fragmentation of our personal information complicates our keeping decision and increases the chances of making a keeping mistake. We may keep the information on the wrong computer, the wrong device, in the wrong form or the wrong format. We may keep the information – an "e-ticket" plane confirmation number, for example, or directions to a hotel – at work when this information will actually be needed as we travel. We may keep information with a reminder to review in two weeks time when, in fact, the information is needed next week. Keeping information wrong may means it is effectively useless later. It does little good

[11] Given that the cue was effective in eliciting a memory for the web site, success rates were between 90% and 100% (across different conditions of access frequency).

to have kept a phone number in a contact management database on our laptop, for example, if we can't spare the time it takes to boot up the laptop, bring up the contact management application, and then locate the desired phone number.

An act of keeping might be likened to throwing a ball into the air towards ourselves where we expect to be at some future point in time and space. Worse still, given the many variations in how an item is kept and in how it may be used, we have to contend with more than merely time and location. But surely this overstates the case. Keeping and finding are complementary to each other (Lansdale, 1988) and many errors in keeping can be made up in later activities of finding. If we put the socks in the wrong drawer, we might still find them later, albeit at the cost of a longer search.

Given all costs of keeping and the may chances for error, given the preference for and success of "do nothing" methods of return to web sites; given the many improvements in search facilities with more on the way … it reasonable to question whether keeping is needed at all. This question is discussed further in a later section.

Support for keeping

Two different approaches in keeping support both aim to reduce the costs of keeping and to reduce the likelihood of error.

Keep the context, use one folder structure for many forms of information. A simple approach followed in the *Universal Labeler* (UL) prototype (W. Jones, Munat, C & Bruce, H., 2005) is to support an expanded use of the user's file folder hierarchy as a general classification scheme to organize email messages and web references as well as files. UL includes a "Label With" dialog that is a variation of the standard *Save As* dialog and referenced immediately below "Save As" on the *File* menu. Label With is used to select a folder as an additional "label" for the document in view without ever having to leave the context of the opened document. As one participant noted in a preliminary evaluation, *Label With* is a kind of "super shortcut maker". The Label With dialog can also be accessed for web pages and email messages that are in view. People can use Label With to turn their file folder hierarchy, which is typically the most elaborated of the folder hierarchies a user maintains (R. Boardman & Sasse, 2004), into an integrating organization of e-documents, email messages and web pages. Preliminary evaluation of Label With has been positive and a more complete evaluation is underway. Some users have been very positive – even asking to use Label With after the evaluation period. However, it is likely that for many people, Label With does not go far enough either in reducing costs (by preserving current context) or increasing benefit (by supporting the integrative use of the file folder hierarchy) to induce a change in current keeping habits.

Automated classification. Another approach is to automate or semi-automate the classification or filing of information items. Two prototype tools specifically target email, in recognition of our special need to sort, prioritize and file large quantities of incoming email. In *MailCat*, Segal and Kephart, (Segal & Kephart, 1999) train an adaptive classifier with the folder classifications of email messages from actual users. On test data, the correct folder is in the top three "guesses" given by MailCat on between 80% to nearly 100% of the trial, depending on the dataset. (Datasets were taken from six users). The researchers observed comparable levels of accuracy in their own everyday use of MailCat. If users are able to quickly select between one of three "guess folders" (or otherwise navigate to another folder as they would normally do)

then filing is likely to be faster and also potentially more accurate as well (e.g., by sometimes pointing out a "correct" folder that the user might otherwise overlook).

But some caveats are apparent: Only six datasets were used. Actual usage data was from the authors only. It's not clear, for example, how MailCat would perform with a deeply nested folder hierarchy or with folders like "stuff", "more stuff" … MailCat as a folder classifier is obviously less useful for people who prefer to leave large amounts of email in their inbox. However, it is possible to imagine a variation of MailCat that would assign sortable properties instead.

The "in-place" classification of email messages in the inbox is explored in the BiFrost prototype (Bälter & Sidner, 2002). BiFrost applies a set of rules in order to apply one of a small, fixed number of categories to each incoming email messages. Users are thus able to focus on, for example, on time-dependent email messages or messages sent by a "Very Important Person". In one evaluation nine of ten participants found BiFrost useful.

The M-level

M-level activities, e.g., "m" for "mapping" or for "maintenance and organization", apply to a collection of information (a PIC) as a whole and include choices made concerning strategy, structures and supporting tools. M-level activates are generally not forced by an event in the way that finding activities are triggered by a need or keeping activities are triggered by information. Keeping and finding activities are reactive; m-level activities are pro-active. In a busy world, this means m-level activities are often postponed, overlooked or ignored entirely – no only by people in their daily practice but also by researchers involved in the study of PIM.

Maintaining many organizations (maybe tomorrow)

The fragmentation of information by forms and their supporting applications poses challenges not only in the practice of PIM but also in the study of PIM. It is difficult and time-consuming to study and compare a participant's organizational schemes across several different forms of information. Several studies have now looked at how the same person manages across different forms of information (R. Boardman & Sasse, 2004; W. Jones, Phuwanartnurak, A. J., Gill, R & Bruce, H., 2005; Ravasio & Krueger, 2004) The following composite emerges:

- People do not generally take time out of a busy day to assess their practice of PIM as a whole or associated PIM strategies (e.g., always file, never file), organizational structures (folders and otherwise) and supporting tools. But, when asked to do so in the context of a fieldwork study, people seem quite willing and able to do this and may want to continue talking well after the time scheduled for an interview.
- People complain about the need to maintain so many separate organizations of information and about the fragmentation of information that results. Boardman et al., for example, observed a significant overlap between the folders a participant created to manage email messages and the folders created to manage files. When information is scattered in several organizations, the same activities must be repeated for each place and so associated costs are multiplied. Worse, since the same activities are likely separated in time, there is increased likelihood that different collections are maintained and managed in different, inconsistent ways. Also, the tools for managing differ.

- Even within the same folder organization, competing organizational schemes may suffer an uneasy co-existence with each other. People may apply one scheme on one day and another scheme the day after. There may, for example, be a tension between organizing files (images, articles, etc.) by project for current use and organizing these same files by content for repeated re-use (W. Jones, Phuwanartnurak, A. J., Gill, R & Bruce, H., 2005). Although research indicates that, given the right support, people are able to assign multiple categories to an information item (D. Quan, Bakshi, Huynh, & Karger, 2003), the support available to the average computer user (e.g., for creation of shortcuts in Microsoft Windows or for aliases on the Macintosh) remains quite primitive. As a result, people are more inclined to copy an item (e.g., a file) rather than to reference it.

Several participants in one study (W. Jones et al., 2002) reported making a special effort to consolidate organizations, for example, by saving everything as a file or sending everything as an email (or, in one case, by making paper printouts of important information to be organized into a paper-based filing system). One participant worked with his administrative assistant to design an organizational scheme which was then applied to work-related paper documents in both of their offices and also to work-related email messages and e-documents.

What triggers maintenance actions? People may decide to "file, act-on or toss" paper documents when these accumulate to beyond a certain level of acceptable clutter. People may also organize as they work through a pile in search of a specific, temporarily lost document. For email messages the trigger may be a limit on the size of the inbox as imposed by an email service or a system administrator. For files, the trigger was once the size of the hard drive. But when computer users routinely have 40 or more gigabytes of storage readily available, there may be few if any external forcing functions.

If users are freed from the necessity to delete or organize their digital information, this may be a good thing. The decision to delete information can be time-consuming and difficult to make. This has been referred to as the "old magazine" effect (W. Jones, 2004). The potential uses or benefits of the item in focus (e.g., an old magazine) may more salient than the ongoing cost of keeping (and never finding the time to read and use) the item. Similarly, Bergman (Bergman, Beyth-Marom, & Nachmias, 2003) refers to the "deletion paradox" to describe a situation where people may spend precious time on information items that are on average, of least value to them (e.g., old, never-used information items that are candidates for deletion). With the dramatic increases in digital storage capacity in the past few years, most users – unless they are keeping large numbers of music or video files – are no longer forced to delete anything, ever.

Even so, in fieldwork investigations of how people organize various forms of information, people frequently seem to express a unease with their current maintenance of information as expressed by apologetic comments like "this is a mess" or in references to themselves a "packrat" (Marshall & Bly, 2005). Or, as one participant in Boardman and Sasse's study (2004) said "Stuff goes in but doesn't come back out – it just builds up". In the Jones et al. (2005) study, all 14 participants made comments at the outset of the form "I really should clean this up". Four participants actually insisted on interrupting the interview while they move or deleted "old stuff that really shouldn't be there anymore". In the aftermath of an interview, participants both the Boardman & Sasse and Jones et al. studies were often inspired to re-organize their information so that the observer on a second or third session might hear the participant open with a comment like "since our last meeting I started thinking about how I organize my information and …".

Limitations of storage no longer force us to delete and items remain. But our attention is still limited. Older items can still distract and obscure. Moving old information out of the way has its own problems. Moving information to another part of a folder hierarchy may mean losing some of the context, in containing folders, that gives the information part of its meaning. More serious, we never seem to find the time. There are possible solutions. For example, old information that is no longer used might be archived "in place", automatically, perhaps gradually (e.g. by fading through shades of grey). Information is still available and can be retrieved if needed but is no longer visible. The author is aware of no research efforts to improve support for the maintenance of personal information. Overall, maintenance seems to be an unglamorous and neglected area for tool support.

The value of external representations

We have too many folder organizations to maintain and we frequently postpone and ignore issues of maintenance they way we might avoid straightening up a messy closet. Keeping (filing) information into a folder structure is not easy and mistakes are common. Search continues to improve. Is it worth it to organize information anymore? Or can we leave our information "flat" and depend upon search (and possibly sorting) as our primary means of access?

The limited available research suggests that organization of information is still essential and that familiar structures such as the folder hierarchy provide important functions beyond support for the re-access to information. In one recent study (W. Jones, Phuwanartnurak, A. J., Gill, R & Bruce, H., 2005), 14 participants each completed a series of hour-long sessions which concluded with the question "Suppose that you could find your personal information using a simple search rather than your current folders…Can we take away your folders? Why or why not?" Participants were permitted to stipulate additional features of this hypothetical search utility and the "folder-free" situation. Issues of control and storage would be handled in some other way. The search utility itself would be fast, effortless to maintain, secure and private (no personal information would be communicated to the Web), etc. With these conditions in place, 13 of 14 participants gave a resounding "No!" to the question "Can we take your folders away?" (One participant gave a conditional "yes" if all his information could be ordered by date and time).

Participants would not give up their folders for a range of reasons including:

- "I want to be sure all the files I need are in one place. "
- "Folders help me see the relationship between things. "
- "Folders remind me what needs to be done. "
- "Folders help me to see what I have and don't have. "
- "I use empty folders for information I still need to get."
- "Putting things into folders helps me to understand the information better. "

The study gave special attention to the ways in which people organized information related to a project such as "planning my wedding" or "finishing course xyz". The file folder hierarchy for a project frequently represented a partial problem decomposition with subfolders representing important components of the project (for example, "plan our honeymoon" or "complete term

paper"). Many components were projects in their own right with sub-subfolders representing a further decomposition.

Project-related folder hierarchies seemed to do more than merely support a return to the information organized within. For example, hierarchies also seemed to function as a kind of project plan, albeit partial and imperfect. A display of subfolders can represent, in one view, many of the key components of a project. The hierarchy contains, intermingled, a record of tasks completed, decisions made and a reminder of tasks and decisions still to be addressed. This structure is, therefore, an important aid in the fundamentally difficult activity of anticipating information need (Bruce, 2005). Of course the folder hierarchy is no crystal ball into the future. But it does provide a reasonably concise representation of the likely information need.

The role of structure in the management of personal information has many facets. Finding the right structure for information helps people in *sensemaking* (B. I. J. D. Dervin, 1992), i.e., in efforts to make sense of the information. Russell et al. (Russell, Stefik, Pirolli, & Card, 1993) provide an excellent analysis in which structures (representations) are acquired and discarded depending upon a perception of costs – the costs to stay with a current structure vs. the costs to switch to a new structure. As a human activity, sensemaking and the search for structures involves both Internal Representations (IRs) and External Representations (ERs) which combine to form an integrated cognitive system (Hutchins, 1994; Kirsh, 2000). A recent investigation of information seeking by Qu and Furnas (Qu & Furnas, 2005) suggests that the searches for content and supporting structures are closely intertwined.

In the Jones et al. study, three participants made efforts to re-use a folder structure when a new project was seen to be similar to an older project. One participant reported making a complete copy of the folder for one project. He re-named the copy to represent the new project and then carefully deleted most of the files contained within. He said that the folder structure for the old project was a useful guide for the new project and he wanted to avoid re-creating this structure from scratch. Another participant created a file-free "xxx-xx-Course Name" template folder structure which he then copied and instantiated for each new course he took. Support for the retrieval and re-use of structures today's computing systems is minimal.

Is it working? Measuring the effectiveness of a mapping and its supporting structures, strategies and supporting tools

People are their own researchers with respect to their practice of PIM and the subjects of their own uncontrolled experiment. People try out new strategies, new schemes of organization, new tools and new features of existing tools. These change over time(Lansdale & Edmonds, 1992) and are uniquely configured for each individual. In one survey (H. Bruce, Jones, W. and Dumais, S, 2004) people were asked to list strategies they tried that had "worked" and also to list strategies that had not worked. Across respondents the same strategies were listed in both the "works for me" and "doesn't work for me" column including "file everything immediately", "file nothing", "keep the folder hierarchy broad but only one or two levels deep", and "develop a deeply nested hierarchy with only a limited number of subfolders at each step".

Survey respondents reported a similar experimentation with tools. For example, two respondents reported attempting to put all their information into InfoSelect[12], a special purpose PIM tool. One respondent used InfoSelect for over a year. But both respondents eventually abandoned the tool. InfoSelect was used in addition to rather than instead of the file system as a means of organizing information. Respondents were apparently unwilling to abandon the filing system and commit to InfoSelect entirely as an alternate means of organizing information. In the end then, InfoSelect represented an extra cost of PIM that was not sufficiently offset (in respondents' estimation) by the benefits of InfoSelect. Two respondents reported a similar pattern in their use of Microsoft OneNote. The application was initially embraced as a means of taking digital notes. Both participants eventually abandoned its use because OneNote required the use of a special tabbed system for the organization of notes which did not integrate with existing organizations for files, email or web references.

In evaluating the effectiveness of strategies, structures and supporting tools in their practice of PIM, people face many of the same challenges that researchers face in the study of PIM. The answer to the question "is it working?" is often 'that depends" or "it's too soon to tell". Proper evaluation must extend over a period of weeks or months and the new support must be exercised in many different situations of use. The appeal and benefits of a tool, for example, may be immediately apparent but its costs are only evident after a period of time. Also, people's subjective assessment of a tool may not align with objective measures of their use (Cockburn & McKenzie, 2000). The methodological challenges in the study of PIM are discussed in a later section.

Manipulating, making sense of and using personal information

We often retrieve and need to work with a collection of information items rather than a single information item. We try to understand how the pieces in a collection of information fit together. What does our schedule look like? How do these products compare? Do we have enough information? What actions should we take? What decisions should we make? A time-honored method for paper-based information is to arrange items on a desktop or other flat surfaces (and the floor if necessary). We can the "see" our information in a single glance. This visual memory becomes a powerful extension of our limited internal working memory. Further, we can readily re-arrange paper documents. We can stack them or otherwise group them as our understanding of their content and their relationships to each other slowly develops. We can highlight them or place "sticky notes" on them.

Bondarenko and Janssen (Bondarenko & Janssen, 2005) note that our comparable ability to manipulate digital information is still very limited. Moreover, it can be noted that assemblages of digital information needed to address a request or make a decision often vanish even though we may need much the same information again in a few days time. We dig through email messages, consult our calendar and look at one or more web sites to verify that a proposed meeting time next month will work for us. And a few days later, we find ourselves take many of the same actions again. We have a configuration of open windows that disappears between restarts of our computer. In our situation we may be reminded of the watchmaker in Simon's

[12] "InfoSelect 7.0 adds more information capture capabilities, numerous new features" Julianne Gilmore http://www.innovationtools.com/Tools/SoftwareDetails.asp?a=68.

parable who was unable to create stable subassemblies for a watch so that he had to start again from scratch after each interruption (H. A. Simon, 1969).

The ROOMs prototype(Henderson, 1986) is an example of an effort to support a persistence in state associated with an activity. Mander's work in support of digital piles (Mander et al., 1992) and Bondarenko and Janssen's work attempt to explore ways to provide rich paper-like manipulation of digital information. But clearly we have only begun to explore the possibilities. Our display of files, for example, is still limited to a small number of views – details, thumbnail, list, small icon, large icon and…?

Managing privacy, security and the distribution of items in a PIC

Issues of privacy and security apply to all three senses of *personal information* as previously discussed. We desire to control access to and distribution of the information we collect within our PSI – whether or not this information is directly about us. We desire some control over, or a least an awareness of, the information about us – medical conditions, financial statements, etc. – that are kept by others (medical insurers, credit reporting agencies, etc.) This extends to the record of web sites we have browsed or what books we have checked out of our local library. This, too, is personal information.

The frequently used watchwords "privacy" and "security" as applied to personal information connect to questions of technology, law and public policy that are well beyond the scope of this chapter to cover adequately. Moreover, although international standards of privacy and security are emerging there is still considerable variation, worldwide, by region and country (Cranor, 2002).

Even if standards for adequate protection are arrived at, mandated by law and enabled by technology, we face a significant challenge in deciding who should have access to what of our personal information and for which uses (Karat, 2005; Olson, Grudin, & Horvitz, 2005). We are understandably concerned about keeping our personal information from falling into the wrong hands. But at the same time, there is much information about us that we do want others to have. For example, we want our travel agent to have our seating our airplane seating preferences and our preference in rental car. Our credit card information, in the right hands, can make on-line transactions much easier. Our medical information, in the hands of a doctor, can save our lives.

Studies show that the willingness to share personal information can vary considerably depending on the type of information, who wants to see it, and for what reasons (M. S. Ackerman, 2000; M. S. Ackerman, Cranor, L. F., and Reagle, J., 1999; V. Bellotti, & Sellen, A., 1993). We may have vague, difficult to articulate, notions concerning different classes of personal information and different classes of people with access to this information, under different circumstances. For example, under normal circumstances, we may want only our doctor and close friends and family to know about a diabetic condition. But in cases of an emergency, it becomes essential that the people treating us have this information.

The sharing of information can be qualified further. For example, conditions and disclaimers may be attached to the information shared, or an audit trail maintained (Butler, 1991; Povey, 1999; Stevens, 2002; Stiemerling. O., 2000).

Giving people "security controls they can understand and privacy they can control for the dynamic, pervasive computing environments of the future" has been identified as a major

research challenge by the Computing Research Association (CRA) Conference on Grand Research Challenges in Information Security and Assurance (*Computer Research Association*, 2003). Whitten and Tygar (Whitten, 1999) point out that "security mechanisms are only effective when used correctly". Karat et al. (Clare-Marie Karat, 2006) note, for example, that privacy policies are often vaguely expressed (e.g.., "Customer service reps will only use your personal information for the efficient conduct of our business") and subject to considerable variation in interpretation when translated into rules can actually be enforced (and upheld in a court of law).

Brodie, Karat and Karat (Brodie, 2005) describe "Wizard of Oz" prototype explorations into ways of translating user statements of privacy and security (e.g., who can access what information under which circumstances to be used in what way) into precise enforceable rules. Olson, Grudin and Horvitz (Olson et al., 2005) describe an effort to determine meaningful groupings of information types and information recipients based on a small set of qualifying questions. Clustering is used to generate a hierarchy of classes for both information type and information recipients. Using categories within these hierarchies, users can choose to exercise a coarse-grained or fine-grained control over who has access to what information.

Given these developments, we can imagine a day when each of us has a coherent, enforceable policy of personal information sharing. Such a policy would be applied, for example, to a personal web site not only to control access but to determine who is notified concerning what updates in web site content. The policy would also apply to the information that others – credit card companies, medical providers, tax agencies, etc. – keep about us. And we might even hope that the policy could travel with the information as it is transferred (Cranor, 2002).

A policy of personal information sharing would help to automate the policy we already effectively, if informally, implement in our daily interactions. We whisper information to a friend or trusted colleague that we would not want others to hear. We give an employee a raise but do not want this known to other employees in the organization. We are back to the use of information and the dissemination of information items as a means of affecting change in our world. Policy information sharing, then, is not only about protection and privacy; it is also about projection and power.

Variations by individual, job role and project

People appear to vary greatly in their habits of maintenance and organization. Malone (Malone, 1983) distinguished between "neat" and "messy" organizations of paper documents. "Messy" people had more piles in their offices and appeared to invest less effort than "neat" people in filing information. Whittaker and Sidner (Whittaker & Sidner, 1996) grouped email users into three categories: 1.) "no filers" left most of their email in the inbox, which often grew to be quite large. 2.) "spring cleaners" also allowed their inboxes to grow quite large but then periodically transferred large numbers of email to an extensive folder organization. 3.) "frequent filers" made daily passes through their inboxes to file or delete messages. Comparable individual differences have been observed for the management of bookmarks (Abrams et al., 1998).

More recently, there is evidence that the same person may have very different habits of maintenance and organization for different forms of information and possibly for different PICs (R. Boardman & Sasse, 2004; H. Bruce, Jones, W. and Dumais, S, 2004). A person may have a

very organized collection of email messages but be relatively disorganized in her organization of paper documents. Or people may be very organized with respect to more specific collections of information – email relating to work, for example, or a specific project at work. In one survey people on average report be most organized for e-documents and other files and least organized for paper documents (H. Bruce, Jones, W. and Dumais, S, 2004). A person can be a "neat" with respect to one form or information, even with respect to a specific collection, and a "messy" with respect to other forms and collections of information.

Practices of organization are likely also to vary with job type and task. Kidd (A. Kidd, 1994) makes a distinction between knowledge workers, communication workers and clerical workers. Bondarenko, O., & Janssen (Bondarenko & Janssen, 2005) distinguish between "research" and "administrative" activities. Research activities are open-ended and unstructured (at least initially). Administrative activities are pre-structured, have little flexibility in their execution and do not change much from instance to another.

Variation by form

Brief mention is made here to research findings that are specific to the information based in paper, email messages and web pages.

Paper information forms endure and even thrive

Sellen and Harper (A. J. Sellen & Harper, 2002) note that for many people, paper is still the preferred medium for working interaction with information. The sentences of a paper document can be highlighted. Its margins can be scribbled in. Paper documents are portable and disposable. Many people also continue to prefer the paper medium as a way of working with a collection of documents. Paper documents can be easily piled, ordered, spread out and arranged for a better understanding of their contents.

Sellen and Harper also note that interactions between new technologies can actually result in an increase in paper use. For example, people have access to an increasing amount of electronic information via the Web, corporate networks, CDs, etc. But people still prefer to read information in paper form. And… people have ready access to high-quality printers. The result: An increase in consumption of paper – information delivered to a person in electronic form is printed out to paper.

In a study of 50 office workers, Whittaker and Hirschberg (Whittaker & Hirschberg, 2001) found little evidence to support the hypothesis that younger workers, having spent more of their lives in a "digital world", might use paper less to store information. (If anything, the opposite appeared to be true.) Paper is still very much with us and is likely to be so for some time to come.

Email is increasingly "everything"

Email applications increasingly provide an environment in which people "live" to accomplish a variety of PIM-related activities included time and task management, personal archiving and contact management (V. Bellotti et al., 2003; Ducheneaut & Bellotti, 2001; Mackay, 1988; Whittaker & Sidner, 1996). These studies also indicate that, for some people, email has become a primary means saving and organizing information items for later use – not only email messages but also documents, web references and even personal notes. People exchange

documents and web references via email. Managers, in particular, make heavy use of email (W. Jones et al., 2002).

Web browsing poses special challenges of information management

The number of possible paths through the Web a person might pursue in a routine session of web use is, for all practical purposes, infinite. How to consider new, promising paths while still keeping focus on the original task? People use various strategies to manage navigation within a session of web use. For example, people make heavy use of the "Back" key (Greenberg & Cockburn, 1999) often in combination with a "hub and spoke" style of web exploration (Catledge & Pitkow, 1995). Or people may create new browser windows in order to track new lines of exploration and they may maintain a list of "check this out later" web references.

Within a session of web access, the problem of keeping found pages found is closely related to the problem of sequencing web page visits. On any given page, the user may see hyperlinks to several pages of potential relevance. The user must decide which of these to follow and what to do with the others. In turn, the next page presents still more hyperlinks of potential utility and so on. The user may find it extremely difficult to sequence page access into a meaningful, productive web session. In the worst case, the session may degenerate into an incoherent sequence of page views scattered across a wide range of topics with little to show for the experience.

Newfield, Sethi and Ryall (Newfield, Sethi, & Ryall, 1998) refer to this as the problem of "web surfing". The standard web browser promotes a depth-first traverse of web pages (Lieberman, 1997). After exploring one reference, users press the Back button to return to the "hub" page and from there, they jump to another promising reference. Problems arise when the referenced "spoke" page itself contains many promising references and so becomes a hub in its own right. When users eventually return to a previous hub after a chain of hub-and-spoke traverses, they may no longer remember their original intentions (Cockburn & Jones, 1996).. Conklin (Conklin, 1987) identifies this as a case of *cognitive overload* – users lack the working memory capacity to maintain state for several levels of navigation at the same time.

Bringing the Pieces Together: Approaches to the Integration of Personal Information

As research in the previous section makes clear, information fragmentation creates problems for keeping, finding and the various m-level activities. The obvious antidote to fragmentation is integration (or unification). This section considers some approaches to integration. Each approach is illustrated by a brief discussion of an ongoing research project and possibly a prototype tool. This chapter is not a review of PIM tools. Before approaches are discussed, it is useful to consider of the kinds of integration that apply to personal information:

Variations and levels of integration

Integration across physical location. Perhaps the most basic kind of integration is the integration of information from many physically distinct sources. It is a significant burden to move physically from location to location to get the information we need especially when these are separated by some distance. Computing technology, in several ways, has already done a great deal to integrate information across physical location. Data transfer protocols such as the

File Transfer Protocol (FTP) long existed to bring information from where it is stored to where we, as users, need it. More recently, tools such as network file systems (NFS) and the Web free us from even having to think about the physical location of the desired information.

As the capacity and portability of storage devices continues to increase we can now bring with us a substantial proportion of the information we use regularly – on a laptop, for example, or even on much smaller device such as a "smart phone". We can access still more information via a wireless connection. Moreover, we now have access to many new kinds of information in digital form – text, of course, but also pictures, music and even full-motion video. Computing technologies already combine to enable a high degree of integration with respect to physical location.

For information in digital form, then, we have reason to hope that physical location will be come less and less a factor in information fragmentation. For digital information, several additional integrations remain to be accomplished:

Information in the means of access and organization. As noted in the previous section, desktop search utilities already provide one integrative means of information access. However, studies continue to show that people have a strong preference for browsing or "orienteering" styles of access to their information (D. K. Barreau & Nardi, 1995; O'Day & Jeffries, 1993; J. Teevan et al., 2004). People use search only after these preferred methods of access fail. There are more basis reasons to organize information – we understand the information better. People often have, and complain about having, several distant organizations of information – usually folder hierarchies Integration means providing at least an option to bring these organizations together. Some people may still find it useful to have distinct organizations for email, e-documents and web references. But this would be a choice freely made, not a separation imposed by supporting applications.

Integration by the grouping and inter-relating of items. It is often useful to group and inter-relate items – to each other and to the tasks for which they are needed. We might, for example, want to inter-relate all the information for a particular person in our lives. Frequently, we group together information relating to a particular task we wish to complete. For example, we might group together information concerning hotels in a city in order to select a hotel for our stay in that city.

Traditional folders provide one means for grouping information together. We also bring together information in a more ad hoc way through the windows we open on our computer display as we try to complete a task. A collection of a kind may be implicit in the trail of information items we have viewed. However, the trail does not persist and open windows add to our clutter and obfuscate as well as group our information.

Research has explored the more general and flexible notion of a *collection* (Dourish et al., 2000; Dourish, Edwards, LaMarca, & Salisbury, 1999a; Karger & Quan, 2004). Items can be manually assigned to a collection (e.g. files placed into a folder) but items (or at least suggested items) for a collection can also be generated based upon a match between items and a "definition" (e.g. a query) for the collection. Limited support for the automated creation of collections is available now via features such as Microsoft Outlook's "Search Folders". Variations on this are now expected in new releases of both the Macintosh and Microsoft Windows operation systems (Fried, 2004).

It may also be useful to assign properties to a collection as a whole. For example, if a collection of information relates to a task ("Find a hotel"), then it may be useful to assign task-like properties like "remind by" and "due by" (W. Jones, Munat, C & Bruce, H., 2005) which might then appear as appointments in an electronic calendar or trigger a reminder (via pop-up or email message) later on.

Integrative views of information. A fourth kind of integration of digital information takes us inside a collection or grouping of information (however defined or created). We seek to "view" the items within a collection. We look for recurring patterns among and important connections between information the items in view. For paper documents, the desktop and other flat surfaces of an office traditionally serve as a view space. We may move paper documents from filing cabinets to the desktop in order to "see" the information better. Computers provide several alternatives for comparable viewings of digital information including the computer desktop, a folder listing of files (or email messages or web references) and the window displays of opened documents, email messages, web sites. Our view of items can act as a powerful extension to our limited internal working memory for information (Larkin, 1987).

Unfortunately, as we attempt to arrange information on a computer display, we experience problems. For example, applications involved in rendering the items of a collection may each consume a large part of the display. Documents, email messages, web pages, etc. may each "live" in a large window with attendant menus, toolbars, jumping-off points and default presentations.

A computer display may be filled with windows, often obscuring each other and each competing for our attention. We can experience similar problems with the computer desktop and, of course, and with the top of a physical desk. The information we lay out in order to see and understand can turn into a jumble that actually impedes our ability to work effectively. Worse, a carefully arranged configuration of windows is typically not preserved between restarts of the computer.

Current computing support for the creation of more workable, integrative views of information is quite limited. There has been little progress in file managers, for example, beyond the standard icon, list (possibly with properties) and thumbnail views.

Integrative facilities of data manipulation. In a fifth kind of integration, we move from "read" to "write" access. For example, we may want to give explicit, external representation to the patterns we notice, the connections we make and main points we note for information that we are viewing. Or we may want to transfer information from one application to another. A basic facility of data manipulation that we use repeatedly in a typical day is the copy/cut-and-paste facility (and the drag-and-drop facility). The cut-and-paste facility provides an intuitive way of moving data from one application to another – although in some cases, the transfer is still text-only.

Other facilities for manipulating data are still provided in a very fragmented, piecemeal fashion. For example, in ways that are analogous to those we use when marking up a paper document, it is possible to highlight and annotate selected text for a document in Microsoft Word. Similar, but not identical, operations can be made on "PDF" documents displayed in Adobe Acrobat. However, it is not possible to perform comparable operations on the selected text of an email message displayed in Microsoft Outlook nor is it possible to highlight the selected text of a Web page as displayed in most web browsers. Even the basic ability to

impose an ordering on information items is unsupported (e.g., for email messages) or accomplished only by a clever use of leading characters (W. Jones, Phuwanartnurak, A. J., Gill, R & Bruce, H., 2005).

Integration with the current context. Associations to various aspects of the current context are also a potential basis of unification. The time of our last interaction with a document (email message, web page) is recorded currently. But many other aspects of the interactive context are not (Lansdale, 1991). We may recall that, when we last viewed an information item at home, not at work, and the weather was warm and sunny outside. But these recollections provide little help in our interactions with the computer. As we create a new email message or e-document or as we browse to a web site, we may have a particular task in mind, but there is very little support for communicating this task to the computer. Worse, newly created documents are often placed, by default, in a place like "My Documents". In general, the context we "share" with the computer in our interactions with information items is very limited.

Other integrations are also relevant to PIM. Users may have considerable time and energy invested in existing folder hierarchies and other organizations. Moreover, these organizations and supporting applications are used in many ways that are not well understood (W. Jones, Phuwanartnurak, A. J., Gill, R & Bruce, H., 2005). Consequently, a new tool has a better chance of success if it is able to build upon these organizations and extend the functionality of existing applications rather than forcing a leap to an entirely new way of doing things.

This discussion is intended neither to be a definitive nor exhaustive treatment of the ways in which we might like to see a greater integration of our information. However, the discussion should provide a sense for the many facets of a general term like "integration" as applied to PIM. These facets (and others that may occur to the reader) can be used as a basis for comparing approaches to the integration of personal information that are discussed in the remainder of this section

Integration through email

The uses of email now extend well beyond its original use to send text messages between people separated from each other by time and distance. For example, email is now used for task management, personal archiving and contact management (V. Bellotti et al., 2003; Ducheneaut & Bellotti, 2001; Mackay, 1988; Whittaker & Sidner, 1996). Many of us might say that we practically "live" in email in a typical work day. (On the other hand, many of us may also go "offline" in order to do concentrated work without the constant interruption of email). One approach to current problems of PIM – in particular, the fragmentation of information by application – is to declare email the winner and to build additional PIM functionality into an expanded email application..

This approach is exemplified by *Taskmaster* (V. Bellotti et al., 2003)a prototype that deliberately builds task management features into an email client application. Taskmaster introduces support for "thrasks" as a way to automatically connect together task-related email messages based upon an analysis of message content. The thrask is intended to be an improvement on "threads". Email discussion within a thread can diverge widely from an original task while, at the same time, as other task-related email messages are sent outside the context of a thread. A thrask can also include links (e.g., web references) and documents that relate to the task. In this way several forms of information are brought together.

Following an "equality of content" principle, Taskmaster also displays attachments (links and documents) at the same level as the email messages associated with their delivery. Attachments are no longer buried within the email message. This makes it easier for the user to see and access all information related to a task, regardless of its form.

Email messages and associated content can be sorted and grouped by thrask but otherwise remain in the inbox until moved by the user. Users can also fine-tune by changing the thrask associated with an email message. The design intent is that Taskmaster adds new task-related functionality without taking away the functionality the user is already familiar with. Taskmaster provides several means of view thrask related email messages and also supports the assignment of task-relevant properties.

One potential limitation of the "integration through email" approach was alluded to above – people may want to spend *less*, not more, time in email. Also, adding additional facilities relating to task management and other PIM-related activities may increase the complexity of an email application that is already quite complex for many users. In addition, users are likely to have other reasons for continuing to use files in the file system – better backup, for example, or better, finer-grained control over access rights and security.

Integration through search

Desktop search facilities that can search across different forms of information – especially files, email and web pages that a person has visited – have a tremendous potential to support a more integrative access to information. Some of this potential has already been realized in available facilities such as Google Desktop.[13]

Fast, integrative cross-form searches are also supported in the Spotlight features of the Macintosh operating system X (Mac OS X)[14]. Spotlight also includes support for persistent searches and the related notion that "smart folders" can be populated and constantly updated to include the results returned for an associated query. Similar features are also planned for inclusion in the next major release of Microsoft windows (code-named "Longhorn")[15].

Microsoft's *Stuff I've Seen* (SIS) project is exploring additional integrations that build upon a basic ability to search quickly through the content and associated properties for the information items of a PSI. The UI for SIS supports sorting of returned results on several properties including a "useful date" (with a definition that varies slightly depending on the information form). Time intervals can be further bracketed in a *Memory Landmarks* add-on through the inclusion representations for memory events, both public and personal. An *Implicit Query* (*IQ*) add-on to SIS is a further step in integration. As a user is viewing an email message, content and properties associated with the message are used to form a query. Matching results are shown in a side panel. The panel may sometimes list useful information items that are in the user's PSI but have been completely forgotten.

---

[13] For a more complete review of desktop search engines available now for use see Answers.com (http://www.answers.com/ and then search, of course, for "desktop search").

[14] See http://www.apple.com/macosx/features/spotlight/.

[15] Spanbauer, S. (2005, August 2005). Longhorn Preview: The newest versions of the next Windows add graphics sizzle and more search features but lack visible productivity enhancements. *PC World*.

These and other search features make it clear that search is about more than typing a few words into a text box and waiting for a list of results. We return to a question posed earlier: Will the constellation of features enabled by fast, indexed search of content *and all associated properties* for information items in a PSI eventually eliminate the need for many PIM activities? In particular, does the need to actively keep information and to maintain and organize this information largely go away? Can people leave their information "flat" so that the need for conventional folders disappears?

There are two very different reasons for believing that the answer is "no":

1. A search can return many versions of the needed information. People create multiple versions of a document, for example, in order to represent important variations or to "freeze" a document at key points in its composition or, simply, because they need to use this document on different projects, in different contexts (and it's easier to copy than to reference). People may also save external items into their PSI several different times because they can't recall whether they have done so before or, again, because they want to access this item in different "places". Or people may receive several different variations of information in email. Airlines, for example, sometimes email several different variations of an on-line, e-ticket confirmation.

   When multiple versions are returned, considerable time may be spent deciding which version is correct or deciding which collection of items provides the necessary information. The problem of multiple versions becomes worse when people modify or correct a document or they save a new version of an item without tracking down and removing all the old versions. A CEO of a major financial services company told the author that he had recently spent over an hour trying to decide which of several versions of a PowerPoint presentation was the right one to use and modify for an upcoming meeting with a customer.

2. The second reason to believe keeping and organizing will remain essential PIM activities is more speculative but also more basic: The act of keeping an item, the act of organizing a collection of items, may be essential to our understanding of this information and our memory for it later on. If filing is cognitively difficult it is also cognitively engaging. Filing, as an act of classification, may cause people to consider aspects of an item they might otherwise not notice. People may forget to search for information later if they don't make some initial effort to understand the information. Folders, properties and other constructs can be seen as an aid in understanding information. Even if a tool like Implicit Query is wonderfully successful at retrieving relevant information anyway, people may fail to recognize this information or its relevance to a current need.

In a better world, we might hope realize the advantages associated with the current use of folders and other means of external representation without suffering their disadvantages. The penalty associated with misfiling currently, for example, is too severe. We may, for all practical purposes, lose this information until it is too late to use it. If folders become more "transparent" or more like tags, we might be more inclined to reference than to copy and more inclined to tag an item in several ways to represent different anticipated uses. We might still be able to search or sort through items as part of a larger set.

In this regard, improving desktop search facilities may have a paradoxical effect. With search, the cost of misfiling goes down. Even if an item is misfiled, it can still be found again later using search if necessary. Moreover, regardless of folder location, search can be used to

construct a useful set of results that can then be quickly sorted by time and other useful properties. Is it possible that people may sometimes be more inclined to file?

Integration through projects

It might be argued that information management and task or project management are two sides of the same coin. It certainly makes sense to try to organize information according to anticipated use and people are observed to do this (B.H. Kwasnik, 1989). ROOMS (Henderson, 1986) represents an early attempt to integrate information items and other resources (tools, applications) with respect to a user activity. For example, a user could set up a "room" for a programming project in which each window provided a view into a project- related resource. A task-based approach to integration, TaskMaster, has already been discussed in the context of extensions to an email application.

Another approach in tool support advances the notion of a "project" as a basis for the integration of personal information. When a distinction is drawn between tasks and projects it is typically with respect to length and complexity. In HCI studies of task management (V. Bellotti, Dalal, B, Good, N, Flynn, P, Bobrow, D. G. & and Ducheneaut, 2004; M. Czerwinski, Horvitz, E and Wilhite, S, 2004), for example a task is typically something we might put on a "to-do" list. "Check email", "send mom flowers for Mother's Day", "return Mary's phone call", or "make plane reservations". With respect to everyday planning, tasks are atomic. A task such as "make plane reservations" can certainly be decomposed into smaller actions – "get travel agent's phone number", "pick up phone", "check schedule", etc. – but there is little utility in doing so. In these studies, therefore, the focus is on a management *between* tasks including handling interruptions, switching tasks and resuming an interrupted task.

A project, by contrast, can last for several days to several years and is made up of any number of tasks and sub-projects. Again, the informal "to-do" measure is useful: While it makes sense to put tasks like "Call the real estate broker" or "Call our financial planner" on a to-do list, it makes little sense to place a containing project like "Buy a new house" or "Plan for our child's college education" into the same list (except perhaps as an exhortation to "get moving!").

In the UMEA prototype (Kaptelinin, 2003) uses the notion of a current project to bring together various forms of information – electronic documents, email-messages, web references – and associated resources (applications, tools). A design goal of UMEA is to minimize the user costs in setting up a project by automatically labeling items as these are accessed. Unfortunately, UMEA depends upon the user to signal a change in current project. Since users frequently forget to do this, projects are frequently associated with the wrong project. Users can go back later and edit project/item associations to correct for mislabeling but, understandably, users in an evaluation are not likely to do this. Kapetlinin sketches possible ways in which the system might detect a change in project but, to the author's knowledge, nothing along these lines has yet been implemented that can do this with any degree of accuracy. Another limitation of UMEA is that the project is essentially just a label and has no internal structure

Another approach in integration through projects is to have labeling of items occur as an incidental part of an activity that people might do in any case. People plan projects. Some of this plan finds external expression in, for example, to-do lists or outlines. The Universal Labeler (UL) prototype (W. Jones, Munat, C & Bruce, H., 2005), described earlier, encourages users to develop a project plan using a Project Planner module. The Planner provides a rich-text

overview for any selected folder hierarchy which looks much like the outline view of Microsoft Word. A hierarchy of folders appears as a hierarchy of headings and subheadings. The view enables users to work with a folder hierarchy just as they would work with an outline. As headings are added, moved, renamed or deleted, corresponding changes are made to the folder hierarchy. The goal is that the Planner is simply another view into the file folder hierarchy and is, in fact, integrated into the file manager. But, as part of UL's general support for shortcuts, folders for a project plan can be used to reference project-related email messages and web pages as well as files. The Planner also provides document-like features not available in a standard file manager:

- **Support for a "drag & link"** action of excerpting. It is possible to select text of interest, drag (or copy) and then drop (or paste) into a project plan. A link to the source of the drag (or copy) is automatically created. Often we are mainly interested in only a small part – a name, number or phrase – of the email message, web page, or e-document that we are reading. But we might like to return to the rest of the information item later on.

- **Support for "create & link"**. A link to the newly created information item (e-document, email message) is automatically created at the insertion point in a project plan.

- **Ordering of elements.** Users can order headings, subheadings and links of a project plan however they like. People depend on ordering as a way to establish priorities and to direct their attention to "first things first".

- **Notes** (annotations). Users can include notes just as they would in a document. Notes can, for example, be used to provide clarification for an associated heading.

Behind the scenes, the Planner is able to support its more document-like outline view by distributing XML fragments as hidden files, one per file folder, which contain information concerning notes, links and ordering for the folder. The Planner assembles fragments on demand to present a coherent project plan view including notes, excerpts, links and an ordering of subfolders (and sub-subfolders). The architecture allows for other views as well. Efforts are currently underway, for example, to support a "mind map" view (Tony Buzan, 2004).

A general goal of the UL is to make it easy for users to give external expression (e.g., in a to-do list, a conventional outline, a table, or a "mind map") to their understanding of a project they are working on. This representation is then used as a means of organizing the information needed to complete the project. Also, in support of project management, properties like importance, due date or "remind by" date can be associated to elements in the representation to further leverage that work that went into its creation.

Of course not all of our information directly relates to a project we are working on (or plan to work on). We keep information for a variety of reasons. Some items are more properly regarded as reference material that may have repeated use, whether or not in projects. For example, we may keep articles, music, recipes, quotes, and images with no specific project in mind. It may be more effective to organize reference items such as these by properties, perhaps using a faceted classification scheme.

Integration through properties

Dourish and Edwards argue that the folder hierarchy is limited, antiquated and should be abandoned outright in favor of a property-based system of filing and retrieval as in their

PRESTO/Placeless Documents prototype (Dourish et al., 2000; Dourish et al., 1999a; Dourish, Edwards, LaMarca, & Salisbury, 1999b). Such proposals are not new. Ranganathan's colon or faceted classification scheme (Ranganathan, 1965) is essentially an organization of information by a set of properties where items value assignment for one property can vary independently of its value assignment for another property. Recipes, for example, might be organized by properties such as "preparation time", "season", "region or style", etc.

However, an organization of information by properties depends upon an understanding of the information so organized. Meaningful, distinguishing, useful properties for special collections like recipes may be readily apparent but much less so for newly acquired information. Information for a project may more readily organize into the hierarchy of a problem decomposition.

Integration through time. One property of clear relevance across most items is time (as in "time of encounter" or "last accessed"). Several projects and prototypes are motivated by the integrative power of time as a means to organize information.

The MEMOIRS system (Lansdale & Edmonds, 1992) organizes information items in a sequence of events (which can also include meetings, deadlines, etc.) Perhaps best known of the time-based approaches to information integration is LifeStreams (Fertig, Freeman, & Gelernter, 1996; Freeman & Gelernter, 1996). In LifeStreams, documents and other information items and memorable events in a person's life are all placed in a single time-ordered "stream".

LifeStreams also permits users to place items into the future portion of the stream a point where a need for these items is anticipated. But it is with respect to the future that the LifeStreams timeline metaphor begins to wear. There are some future events are that are "fixed" (to the best of our ability to fix anything in the future) – meetings for example. It makes sense to place a presentation or report that is needed for a meeting at a point in the stream's future to coincide with the meeting. However, in many other cases we have no clear notion of when we will need an item in the future or when we will have an opportunity to use it. In these cases it may make more sense to organize items according to a need (goal, task, project) for which it may have use. Needs, in turn, are often organized into a hierarchy.

Integration through a common underlying representation

The digital information items discussed in this chapter – in particular the file – are high-level. The operations we can perform at the file level are useful, but limited. We can create, move, re-name, and delete files. The data with in a file is typically in a "native format" and readable only by a single application – the word processor, spreadsheet, presentation software, etc., used to create the file. In this circumstance, opportunities to share, consolidate and normalize data (e.g., to avoid problems with updating) are extremely limited. The user can initiate a transfer of data in a file to another file ("owned" by another software application) via mechanisms like "copy & paste" and "drag & drop". But this transfer is often little more than an interchange of formatted text. Information concerning the structure and semantics of the data stays behind in the source application. Moreover, the data is copied, not referenced, which can lead to many problems of updating later on.

As a result, data concerning a person we know – say Jill Johnson – may appear in many, many places in our PSI. This is another variation of fragmentation. Because of this fragmentation,

even simple operations, like correcting for a spelling mistake in Jill's name or updating for a change in her email address, become nearly impossible to complete. We may update some of copies but not all[16]. Also, we may experience the frustration of having some operations – name resolution, for example – available in one place (when sending email) but not in another (when working with photographs). Underlying these problems is a problem that there is no concept or "object" for a "person named 'Jill Johnson'" in the PSI and no means by which data associated with this person can be referenced, not copied, for multiple uses (as managed through various software applications).

The situation may improve with increasing support for standards associated with the Semantic Web (Berners-Lee., 1998) including XML (eXtensible Markup Language), RDF (Resource Description Framework) and the URI (Uniform Resource Identifier). RDF and XML, for example, can be used to include more semantics with a data interchange. URIs might be used to address data, in place, so that it doesn't need to be copied in the first place (thus avoiding problems with updating information about Jill Johnson, for example). Support for these standards may make it possible, in some future day of PIM, to work with information and data packaged around concepts such as "Jill Johnson" rather than with files. Data for Jill would be mostly referenced; not copied. We could readily add more information about Jill or to make a comment like "she's a true friend". And we could group information about Jill together, as needed, with other information. We could, for example, create a list of email addresses and phone numbers for "true friends" we would like to invite to our birthday celebration.

These and other possibilities are explored in the Haystack project (Adar, Karger, & Stein, 1999; David R. Karger, 2005; Huynh, Karger, & Quan, 2002; Dennis Quan, Huynh, & Karger, 2003). Haystack represents an effort to provide a unified data environment in which it is possible to group, annotate and reference or link information at smaller and more meaningful units than the file. In the Haystack data model, a typical file will be disassembled into many individual information objects represented in RDF. Objects can be stored in a database or in XML files. When an object is rendered for display in the user interface, a connection is kept to the object's underlying representation. Consequently, the user can click on "anything" in view and navigate to get more information about the associated object (e.g., to get Jill Johnson's birth date, for example) and also to make additions or corrections to this information.

Haystack creates a potential to explore, group and work with information many ways that are not possible when information is "hidden" behind files. However, many issues must be addressed before the Haystack vision is realized in commercial systems. For example, the use of RDF, whether via XML files or a database, is slow. Beyond performance improvements, a great deal must be done if application developers are eventually to abandon the control they currently have with data in native format in favor of a system where data comes, instead, from an external source as RDF.

Integration through a digital recording of "everything"

If a sequence of information events are recorded – the viewing of a web page, for example – there is the possibility later to facilitate not only retrieval of the web page again but also to

[16] But we might have good reasons to not update some copies. We may be keeping an older version of an address list. Her name and address may appear in an old paper that was already published and is part of our archive.

retrieve other items there were in close temporal proximity to this item. We might hope, for example, be able to access "the email message I was looking at right before I looked at this web page".

If enough events in our daily life are recorded, we might get significantly closer to a situation where virtually anything recall about a desired item – the contexts of our interaction with the item as well as its content – might provide an access route back to the item. For example, we might query our computer to "go back to the web site that Mary showed me last week".

In his article "As We May Think," Vannevar Bush described a vision of a personal storage system, a memex, which could include snapshots of the a person's world taken from a walnut-sized head-mounted camera and a voice recorder (Bush, 1945). Bush's vision has been realized and extended in wearable devices that can record continuous video and sound (Clarkson, 2002; Niedzviecki, 2001; Steve Mann, 2004)

A bigger question is what to with all this data once recorded? *MyLifeBits* (J. Gemmell et al., 2002; J. Gemmell, Lueder, Roger, and Bell, Gordon, 2003) is an exploratory project aimed at finding out. The project tries to "digitize the life" of computer pioneer Gordon Bell.

The study of "record everything" approaches, also called "digital memories" is becoming a very active area of research (Mary Czerwinski, 2006). For example, workshops on Continuous Archival and Retrieval of Personal Experience (CARPE) have been sponsored by the Association for Computing Machinery (ACM) in both 2004 and 2005.

A continuous recording of our life's experiences has many potential uses. For example, might use to refresh our internal memories concerning a meeting. It might be useful in some cases to support our version of events later on. Or we might like to review our digital recording in an effort to learn from our mistakes. Sometimes, we might review just for fun. But clearly, digital memories raise serious concerns of privacy and security which can only be partially addressed by technology alone.

Integration through organizing techniques and strategies

Approaches to integration are predominantly tool-based as inspired by developments in technology. But a degree of integration can also be accomplished through techniques and strategies of integration that make use of existing tool support. It has already been noted, for example, that people are sometimes observed to focus on a single form of information and the development of organizing structures for this form. Other forms of information are "squeezed" into this organization. Everything is printed out, for example. Or everything is sent as an email. Or everything becomes a file.

Some people are observed to create, instead, a single organizing schema which is then applied to different forms of information. The observation has prompted Jones (W. Jones, 2004) to speculate on the possible value of a *Personal Unifying Taxonomy (PUT).* A person's PUT would be developed after a review, as guided by a trained interviewer, of organizations for email, e-documents, paper documents, web references and other forms of information. Top-level elements in a PUT would represent areas with enduring significance in a person's life (high-level goals, important roles). A PUT would also represent recurring themes in the folders and other constructs of different organizations.

However, a great deal of work remains to be done to determine a process and principles of PUT development and to determine whether a PUT can be maintained over time to realize benefits that compensate for its costs of creation and maintenance. In the development of a process and principles of PUT development, we might hope to borrow from the field of library and information science. The larger point is that, in our fascination with the potential of new tools and technology, we should not overlook the potential to improve through changes in our techniques, strategies and habits of PIM.

Methodologies of PIM

The development of methodologies especially suited to PIM is still in its infancy. There is a need both for methodologies in *descriptive* studies aimed at better understanding how people currently practice personal information management and *prescriptive* evaluations to understand better the efficacy of proposed PIM solutions (usually involving a tool but sometimes focused, instead, on a technique or strategy).

The descriptive and the prescriptive can form a complementary and iterative relationship to one another:

1. Descriptive data from fieldwork observations, interviews and, possibly, broader-based surveys can suggest directions for exploratory prototyping of supporting tools (and supporting techniques as well).
2. Prototypes are built and evaluated to reach more definite, prescriptive conclusions concerning support that *should* be provided. The development and evaluation of prototypes can frequently suggest specific areas of focus for the next round of fieldwork.

This is a familiar, if somewhat idealized, process for the study of human-computer interaction (HCI) – although, all too often it seems, the descriptive component is overlooked or disconnected from the rush to build new tools (Whittaker et al., 2000).

PIM poses special challenges with respect to both its descriptive study and the prescriptive evaluation of proposed solutions:

1. **A person's practice of PIM is unique.** There is tremendous variation between people -- even between people who have a great deal in common with each other with respect to profession, education and computing platform. These truths are repeatedly reaffirmed in fieldwork studies (e.g., (W. Jones et al., 2001). People develop (and continue to experiment with) their own practice of PIM, including supporting strategies, structures, tools and habits with little or no formal guidance. A practice is uniquely suited to the individuals, his/her needs and information. This uniqueness makes it very difficult to abstract tasks or extract datasets that can be meaningfully used in a laboratory setting.
2. **PIM happens broadly across many tools, applications and information forms**. Moreover, people freely convert information from one to another form to suit their needs – emailing a document, for example, or printing out a web page. Studies and evaluations that focus on a specific form of information and supporting applications – email, for example – run the risk of optimizing for that form of information but at the expense of a person's ability to manage other forms of information.
3. **PIM happens over time.** Personal information has a life cycle – moving, for example, from a "hot" pile to a "warm" project folder and then, sometimes, into "cold" archival

> storage. Activities of keeping and finding directed to an information item may be separated from each other by days, weeks or months. Basic PIM events of interest – like filing, creation of a new folder or the protracted search for a lost item of information -- occur unpredictably and cannot be "scheduled". The effectiveness of an action to file information, for example, can't be assessed without also looking at later efforts to retrieve this information. People may initially embrace a solution but then, over time, tire of its ongoing use. Single-session studies and evaluations sample a point in time and can easily mislead. For example, a single-session evaluation of an automated categorization tool might show that users are quite happy with its categorization and the savings of their time that its use appears to offer. But later these users may find that they have more trouble finding information with the tool then without (e.g., perhaps because they attend less to the information initially when categorization is automated).

One approach then is to create ethnographies of PIM in which a person and his/her practice of PIM are the subject of an exploratory, longitudinal case study. Design methodologies that share in common an emphasis on context and situation have obvious relevance including *contextual inquiry* (Beyer H., 1998), *situated activity* (Suchman, 1983), and *situated design* (Greenbaum, 1991). These and other methodologies emerge from a participatory design movement that originated in Scandinavia (Namioka, 1993). Participants in PIM studies might also be encouraged to practice what might be called participatory observation or, more simply, self-observation. People are often quite interested in talking about their practice of PIM. Participants in longitudinal studies seem to attach a kind of therapeutic value to the opportunity to talk about their problems of information management with a sympathetic observer.

But longitudinal, case studies are time-consuming and it is not easy to find a representative sampling of participants able to commit to a multi-session study. The results of case studies may be very enlightening but do not, by themselves, form a proper basis for generalization.

However, a longitudinal case study can be followed by a much more targeted single session study or, even, with a survey. The case study can help to identify the effects to focus upon and the questions to ask in a single session study or a survey.

The effectiveness of PIM research can be improved through:

Development of reference tasks (Whittaker et al., 2000) For example, there is a need for validated keeping and finding tasks that can be administered to a participant as they work with their information.

Units of analysis.. One potential unit of analysis is the personal project (W. Jones, Munat, C & Bruce, H., 2005). The study of PIM puts an emphasis on helping people manage their information over time and in ways that cross the many boundaries set by current tools. This is a worthy if somewhat daunting ambition. How much personal information do we study? For how long? In what contexts? A personal project (e.g., planning a trip, taking a course, planning a re-model) is bounded in time and scope and still typically requires the use of a range of tools, computer-based and otherwise, and the use of many forms of information. A study of a person's management of information as they work to complete a project may, therefore, provide practical ways to approach PIM without "falling into" existing tool-based partitions (e.g. by studying only email use or only Web use)

It is important to note also that methodologies of PIM need to support the development and evaluation not only of tools but also techniques and strategies of PIM.

Theoretical Approaches and Conceptual Frameworks for PIM

There are, as yet, no theories of PIM. But several theoretical approaches and conceptual frameworks have potential application to PIM and might eventually lead to a theory of PIM. It is not possible to in this chapter even to mention, much less adequately review, the many approaches and frameworks that have application in PIM (but see (Pettigrew et al., 2001) and (Pirolli, 2006)) But a sampling provides some sense for the space potentially useful approaches and frameworks.

Cognitive Work Analysis

Cognitive Work Analysis (CWA) (Fidel & Pejtersen, 2004; J. Rasmussen, 1986; J. Rasmussen, Pejtersen, A.M. and Goodstein, L.P., 1994; Vicente, 1999)is a work-centered conceptual framework that emphasizes the importance of understanding the work people do, the behavior of people as they interact with their information, the context(s) in which people work, and the reasons behind the actions people take. In order to gain the understanding needed to design supporting systems, CWA calls for several dimensions of analysis including analyses of the work environment, work domain (including goals and constraints of the domain), tasks people must perform, the organization (including managerial style, organizational culture, social conventions, and allocated roles), decisions made (and supporting information), strategies, resources and values. Although much of the terminology and many of the applications of CWA are focused on organizational work environments, CWA has much broader application and would seem to have relevance to most situations of PIM.

Activity theory

Notwithstanding its name, **activity theory** (Nardi (ed.), 1996) is more properly regarded as a theoretical approach rather than an theory (i.e., with associated models and predictions that can be tested). Activity theory centers on the concept of an activity and upon the goal(s) that motivate an activity. The use, indeed the meaning, of objects in a person's world, including information items, paper-based and digital, should be understood primary with respect the roles they play in various activities the person is completing. One reasonable conclusion made evident by the logic of activity theory is that work environments should "help the user organize resources around their meaningful goals" (Kaptelinin, 2003). Other reasonable desiderata pointed to by activity theory are that a system should support a wide variety of work practices and should be as transparent as possible to that the user can focus on meaningful goals rather than on the system and its technology.

Sense-marking

Sense-making is a theoretical approach (also sometimes called a "meta-theory") under constant development and expansion by Brenda Dervin and her colleagues since 1972 (B. Dervin, 1999; B. I. J. D. Dervin, 1992). Sense-making characterizes people as theory makers on a journey through life in time and space. As people encounter a gap in their understanding, they seek information to improve their understanding, bridge the gap and "move on". Sense-making

regards a person's information behavior with respect to not only to thoughts (cognition), but also emotions, feelings and dreams. Sense-making has been applied to study information seeking and use and has been used by researchers in a wide variety of areas including media, education and telecommunications.

Cognitive architectures

Several theories and associated models of human information processing and human-computer interaction have potential application to PIM. These include the Model Human Processor (S. K. Card, Moran, & Newell, 1983), GOMS (Goals, Operators, Methods & Selector Rules; (S. K. Card et al., 1983; John & Kieras, 1996), EPIC (Executive Process-Interactive Control and Information, (Kieras & Meyer, 1997), Foraging Theory (P. Pirolli & S. K. Card, 1999) and the ACT-R (Anderson, 1990).

These theories are collectively referred to as *cognitive architectures* (see (Pirolli, 2006)). They can be seen as a response to a challenge issued by Newell and Simon (A. Newell & Simon, 1972) to researchers in cognitive psychology to develop unified, integrative theories of human cognition. ACT-R follows in a long line of ACT theories (stretching over a period of more than 30 years) that attempt to meet this challenge (Anderson, 1983). These theories also form the basis for the cognitive architecture in the Information Foraging Theory. Cognitive architectures listed here share in common a rules-based approach to the specification of human behavior. Selection or production rules are essentially "if-then" statements. A rule "fires" if its "if" condition matches (or is the best in a collection of matches).

Cognitive architectures have been usefully applied to predict behavior and performance in several PIM related activities including visual search through information displays (M. D. Byrne, Anderson, Douglass, & Matessa, 1999; Hornof & Kieras, 1997), Information Seeking on a PDA (Luo & John, 2005)and information seeking on the web (Pirolli & Fu, 2003).

Conclusions: Key challenges for PIM as activity and a field of inquiry.

PIM activities are usefully grouped according to their role in our ongoing effort to establish, use and maintain a mapping between information and need.

- *Finding* activities move us from a need to information that meets this need. Finding, especially in cases where we are trying to re-access items in our PSI, is multi-step and problems can arise with each step: We must remember to look; we must know where to look; we must recognize the information when we see it; and we must often do these steps repeatedly to "re-collect" a set of items
- *Keeping* activities move us from encountered information to anticipated needs for which this information might be useful (or a determination that the information will not be needed). Keeping activities are multi-faceted to reflect the multi-faceted nature of anticipated needs. We must make choices concerning location, organizing folder, form, and associated devices/applications.
- *M-level* activities focus on the mapping that connects information to need and on "meta-level" issues concerning organizing structure, strategies and supporting tools. We *maintain and organize* collections of personal information; we *manipulate, make sense of and "use"*

information in a collection; we also seek to *manage privacy and security* and we *measure the effectiveness* of the structures, strategies and tools we use.

One ideal of PIM is that we always have the right information in the right place, in the right form, and of sufficient completeness and quality to meet our current need. Although this ideal is far from reality for most of us, the research reviewed in this chapter should give us some reason to believe we're moving the right direction. There is clear interest in building a stronger community of PIM research to address a pervasive problem of *information fragmentation* – evident, even, in fragmented nature of PIM-related research.

PIM, as an emerging field of inquiry, provides a very productive point of integration for research that is currently scattered across a number of disciplines including information retrieval, database management, information science, human-computer interaction, cognitive psychology and artificial intelligence. On a personal level, improvements in our ability to manage personal information should bring improvements, not only in our personal productivity, but also in our overall quality of life.

Acknowledgements

I would like to thank Maria Staaf for her careful review and proofreading of previous drafts of this chapter.

References

Abrams, D., Baecker, R., & Chignell, M. (1998, April 18-23, 1998). *Information archiving with bookmarks : personal web space construction and organization.* Paper presented at the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 1998), Los Angeles, California.

Ackerman, M. S. (2000). *Developing for privacy: Civility frameworks and technical design.* Paper presented at the Proc.Computers, Freedom, & Privacy.

Ackerman, M. S., Cranor, L. F., and Reagle, J. (1999). *Privacy in e-commerce: Examining user scenarios and privacy preferences.* Paper presented at the Proc. ACM Conference on Electronic Commerce.

Adar, E., Karger, D., & Stein, L. A. (1999, November 2-6). *Haystack: per-user information environment.* Paper presented at the 8th Conference on Information and Knowledge Management (CIKM 1999), Kansas City, MO.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Balter, O. (2000, April 1-6). *Keystroke level analysis of email message organization.* Paper presented at the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2000), Hague, Netherlands.

Bälter, O., & Sidner, C. L. (2002). *Bifrost inbox organizer: giving users control over the inbox.* Paper presented at the Second Nordic Conference on Human-computer interaction.

Barreau, D. K. (1995). Context as a factor in personal information management systems. *Journal of the American Society for Information Science, 46*(5), 327-339.

Barreau, D. K., & Nardi, B. A. (1995). Finding and reminding: file organization from the desktop. *SIGCHI Bulletin, 27*(3), 7.

Barsalou, L. (1991). Deriving categories to achieve goals. *The Psychology of Learning and Motivation, 27,* 1-64.

Bates, M. J. (1989). The Design of Browsing and Berrypicking Techniques for the Online Search Interface. *Online Review, 13*(October), 407-424.

Bellotti, V., & Sellen, A. (1993). *Design for privacy in ubiquitous computing environments.* Paper presented at the ECSCW-93.

Bellotti, V., Dalal, B, Good, N, Flynn, P, Bobrow, D. G., & and Ducheneaut, N. (2004). *What a To-Do: Studies of Task Management Towards the Design of a Personal Task List Manager.* Paper presented at the Conference on Human Factors in Computing Systems, Vienna, Austria.

Bellotti, V., Ducheneaut, N., Howard, M., Neuwirth, C., & Smith, I. (2002, June 25-28). *Innovation in extremis : evolving an application for the critical work of email and information management.* Paper presented at the Conference on Designing Interactive Systems (DIS2002), London.

Bellotti, V., Ducheneaut, N., Howard, M., & Smith, I. (2003). *Taking email to task: The design and evaluation of a task management centered email tool.* Paper presented at the Conference on Human Factors in Computing Systems (CHI 2003), Ft. Lauderdale, Florida.

Bellotti, V., & Smith, I. (2000, Aug 2000). *Informing the design of an information management system with iterative fieldwork.* Paper presented at the Conference on Designing interactive systems (DIS 2000), New York City, New York.

Bergman, O., Beyth-Marom, R., & Nachmias, R. (2003). The user-subjective approach to personal information management systems. *Journal of the American Society for Information Science and Technology, 54*(9), 872-878.

Berners-Lee., T. (1998). Semantic Web roadmap: An attempt to give a high-level plan of the architecture of the Semantic Web.

Beyer H., H. K. (1998). *Contextual Design*: Academic press.

Boardman, R. (2004). *Improving Tool Support for Personal Information Management.* Imperial College, London.

Boardman, R., & Sasse, M. A. (2004). *"Stuff goes into the computer and doesn't come out" A Cross-tool Study of Personal Information Management.* Paper presented at the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2004).

Boardman, R., Spence, R., & Sasse, M. A. (2003, June 22-27). *Too many hierarchies? : the daily struggle for control of the workspace.* Paper presented at the HCI International 2003 : 10th International Conference on Human-Computer Interaction, Crete, Greece.

Bondarenko, O., & Janssen, R. (2005). Documents at Hand: Learning from Paper to Improve Digital Technologies In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 121-130). Portland, Oregon, USA ACM Press.

Broadbent, D. E. (1958). *Perception and Communication*. London, U.K.: Pergamon Press.

Brodie, C., Karat, C., Karat, J., and Feng, J.. (2005, July, 2005). *Usable Security and Privacy: A Case Study of Developing Privacy Management Tools.* Paper presented at the Proceedings of the Symposium on Usable Privacy and Security (SOUPS),.

Bruce, H. (2005). Personal, anticipated information need. *Information Research, 10*(3).

Bruce, H., Jones, W. and Dumais, S. (2004). Information behavior that keeps found things found. *Information Research, 10*(1).

Bruce, H., Jones, W., and Dumais, S. (2004). *Keeping and Re-finding information on the Web: What do people do and what do they need?* Paper presented at the ASIST 2004: Proceedings of the Annual Conference of the American Society for Information Science and Technology, Providence, RI.

Budzik, J., Hammond, K. and Birnbaum, L. (2001). Information access in context. *Knowledge based systems, 14*(1-2), 37-53.

Bush, V. (1945, July 1945). As We May Think. *The Atlantic Monthly*.

Butler, J. K. (1991). Toward understanding and measuring conditions of trust: Evolution of a conditions of trust inventory. *Journal of Management, 17*, 643-663.

Byrne, M. D., Anderson, J. R., Douglass, S., & Matessa, M. (1999). Eye tracking the visual search of click-down menus. *CHI 1999, ACM Conference on Human Factors in Computing Systems, CHI Letters, 1*(1), 402-409.

Byrne, M. D., John, B. E., Wehrle, N. S., & Crow, D. C. (1999, May 15-20). *The tangled web we wove : a taskonomy of WWW use.* Paper presented at the Conference on Human Factors in Computing Systems (CHI 1999), Pittsburgh, PA.

Capurro. R & Hjørland, B. (2003). The Concept of Information. In B. Cronin (Ed.), *Annual Review of Information Science and Technology (ARIST)* (Vol. 37, pp. 343-411).

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Card, S. K., Moran, T.P. and Newell, A. (1983). *The Psychology of Human-Computer Interaction,* . Hillsdale, New Jersey: Erlbaum Associates.

CARMODY, S., GROSS, W., NELSON, T., RICE, D., AND VAN DAM, A. (1969). A hypertext editing system for the /360. In M. F. a. J. Nievergelt (Ed.), *Pertinent Concepts in Computer Graphics* (pp. 291-330): University of Illinois Press.

Carroll, J. M. (1982). Creative names for personal files in an interactive computing environment. *International Journal of Man-Machine Studies, 16*(4), 405-438.

Case, D. O. (1986). Collection and organization of written information by social scientists and humanists: a review and exploratory study. *Journal of Information Science, 12*(3), 97-104.

Catledge, L. D., & Pitkow, J. E. (1995, April 10-14). *Characterizing browsing strategies in the World-Wide Web.* Paper presented at the Third International World Wide Web Conference (WWW 1995), Darmstadt, Germany.

Clare-Marie Karat, C. B., and John Karat. (2006). Addressing the Grand Challenge of Usable Privacy and Security Technology for Personal Information Management. *Communications of the Acm*.

Clarkson, B. P. (2002). *Life Patterns: structure from wearable sensors.* Massachusetts Institute of Technology,.

Cockburn, A., & Jones, S. (1996). Which way now? : analysing and easing inadequacies in WWW navigation. *International Journal of Human-Computer Studies, 45*(1), 105-129.

Cockburn, A., & McKenzie, B. (2000, September 5-8). *An evaluation of cone trees.* Paper presented at the People and computers XIV - usability or else! : proceedings of HCI 2000, Sunderland, UK.

Cole, I. (1982). *Human aspects of office filing: implications for the electronic office.* Paper presented at the Human Factors Society 26 Annual Meeting, Seattle, WA.

Computer Research Association. (2003, November 16-19). Paper presented at the CRA Conference on "Grand Research Challenges in Information Security and Assurance".

Conklin, J. (1987). Hypertext: an introduction and survey. *Computer, 20*(9), 17-41.

Cornelius, I. (2002). Theorizing information. *Annual Review of Information Science and Technology, 36*, 393-425.

Cranor, L. (2002). *Web Privacy with P3P*. Cambridge: O'Reilly.

Czerwinski, M., Dumais, S., Robertson, G., Dziadosz, S., Tiernan, S. and van Dantzich, M. (1999). *Visualizing implicit queries for information management and retrieval.* Paper presented at the Proceedings of CHI 1999.

Czerwinski, M., Horvitz, E and Wilhite, S. (2004, 2004). *A diary study of task switching and interruptions.* Paper presented at the Proceedings of the 2004 conference on Human factors in computing systems, Vienna, Austria.

David R. Karger, K. B., David Huynh, Dennis Quan, and Vineet Sinha. (2005). *Haystack: A General Purpose Information Management Tool for End Users of Semistructured Data.* Paper presented at the CIDR.

Dempski, K. L. (1999, September 27-29). *Augmented workspace : the world as your desktop.* Paper presented at the Handheld and Ubiquitous Computing : First International Symposium, HUC'99 : proceedings, Karlsruhe.

Dervin, B. (1999). On studying information seeking methodologically: The implications of connecting metatheory to method. *Information Processing and Management, 35*, 727-750.

Dervin, B. I. J. D. (1992). From the mind's eye of the user: The sense-making qualitative-quantitative methodology. In R. R. P. Glazier (Ed.), *Qualitative research in information management* (pp. 61-84). Englewood, CO: Libraries Unlimited.

Dourish, P., Edwards, W. K., LaMarca, A., Lamping, J., Petersen, K., Salisbury, M., et al. (2000). Extending document management systems with user-specific active properties. *ACM Transactions on Information Systems, 18*(2), 140-170.

Dourish, P., Edwards, W. K., LaMarca, A., & Salisbury, M. (1999a). Presto: an experimental architecture for fluid interactive document spaces. *ACM Transactions on Computer-Human Interaction, 6*(2), 133-161.

Dourish, P., Edwards, W. K., LaMarca, A., & Salisbury, M. (1999b). *Using Properties for Uniform Interaction in the Presto Document System.* Paper presented at the UIST'99.

Ducheneaut, N., & Bellotti, V. (2001). E-mail as habitat. *Interactions, 8*(5), 30-38.

Dumais, S. T., Cutrell, E., Sarin, R. and Horvitz, R. (2004). *Implicit queries (IQ) for contextualized search.* Paper presented at the SIGIR 2004.

Durso, F. T., & Gronlund, S. (1999). Situation awareness. In F. T. Durso, Nickerson, R., Schvaneveldt, R. W., Dumais, S. T., Lindsay, D. S., & Chi, M. T. H. (Ed.), *The Handbook of Applied Cognition* (pp. 284-314): Wiley.

Eisenberg, M., Lowe, C and Spitzer, K. (2004). *Information Literacy : Essential Skills for the Information Age*.

Ellis, J. K., L. (2000). Prospective memory in 2000: Past, present and future directions. *Applied Cognitive Psychology,, 14*, 1-9.

Engelbart, D. a. E., W. (1994 (video made in 1968)). ~ Research Center for Augmenting Human Intellect. *Reprinted in ACM SIGGRAPH Video Review, 106.*

Engelbart, D. C. (1961). Special considerations of the individual as a user, generator and retriever of information. *American Documentation, 12*(2), 121-125.

Engelbart, D. C. (1963). A Conceptual Framework for the Augmentation of Man's Intellect. In *Vistas in Information Handling*. London: VI Spartan Books.

Erdelez, S., & Rioux, K. (2000). Sharing information encountered for others on the Web. *New Review of Information Behaviour Research, 1*, 219-233.

Fertig, S., Freeman, E., & Gelernter, D. (1996, April 13-18). *Lifestreams: an alternative to the desktop metaphor.* Paper presented at the Conference on Human Factors in Computing Systems (CHI 1996), Vancouver, B.C.

Fidel, R., & Pejtersen, A. M. (2004). From information behaviour research to the design of information systems: the Cognitive Work Analysis framework. *Information Research, 10*(1).

Freeman, E., & Gelernter, D. (1996). Lifestreams: A Storage Model for Personal Data. *ACM SIGMOD Record (ACM Special Interest Group on Management of Data), 25*(1), 80-86.

Fried, I. (2004). Tiger, Longhorn search for desktop answers. *C/net News.com*, June 30, 2004.

Gemmell, J., Bell, G., Lueder, R., Drucker, S., & Wong, C. (2002). *Mylifebits: fulfilling the memex vision.* Paper presented at the 2002 ACM workshops on Multimedia, Juan-les-Pins, France.

Gemmell, J., Lueder, Roger, and Bell, Gordon. (2003). *The MyLifeBits Lifetime Store.* Paper presented at the ACM SIGMM 2003 Workshop on Experiential Telepresence (ETP 2003), Berkeley, CA.

Gershon, N. (1995, December, 1995). *Human Information Interaction.* Paper presented at the Proceedings of the Fourth International World Wide Web Conference.

Goldstein, I. (1980, June 15-18). *Pie : a network-based personal information environment.* Paper presented at the Workshop on Research in Office Semantics, Chatham, Cape Cod, MA.

Golovchinsky, G. (1997 -a). Queries? Links? Is there a difference? . In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 407-414). Atlanta, Georgia, United States ACM Press.

Golovchinsky, G. (1997 -b). What the query told the link: the integration of hypertext and information retrieval In *Proceedings of the eighth ACM conference on Hypertext* (pp. 67-74). Southampton, United Kingdom ACM Press.

Greenbaum, J. a. K., M., Eds. (Ed.). (1991). *Design at Work: Cooperative Design of computer systems*. Hillsdale, AZ: Erlbaum.

Greenberg, S., & Cockburn, A. (1999, June 3). *Getting back to back : alternate behaviors for a web browser`s Back button.* Paper presented at the 5th Conference on Human Factors and the Web : the Future of Web Applications, Gaithersburg, MD.

Gwizdka, J. (2000, April 1-6). *Timely reminders : a case study of temporal guidance in PIM and email tools usage.* Paper presented at the Conference on Human Factors in Computing Systems (CHI 2000), Hague, Netherlands.

Gwizdka, J. (2002a, April 20-25). *Reinventing the inbox : supporting the management of pending tasks in email.* Paper presented at the CHI 2002 : ACM SIGCHI Conference on Human Factors in Computing Systems. Doctorial Consortium, Minneapolis, MN.

Gwizdka, J. (2002b, 2002). *TaskView: design and evaluation of a task-based email interface.* Paper presented at the 2002 Conference of the Centre for Advanced Studies on Collaborative research.

Henderson, A. a. C., S. (1986). Rooms: the use of multiple virtual workspaces to reduce space contention in a window-based graphical user interface. *ACM Transactions on Graphics (TOG), 5*(3), 211-243.

Henzinger, M., Chang B-W., Milch, B., and Brin S. (2003). *Query-free news search.* Paper presented at the Proceedings of WWW 2003.

Herrmann, D., Brubaker, B., Yoder, C., Sheets, V., & A., T. (1999). Devices that remind : chapter 13. In F. T. e. a. e. Durso (Ed.), *Handbook of applied cognition* (pp. 377-407). Chichester: John Wiley & Sons.

Hornof, A. J., & Kieras, D. E. (1997). *Cognitive modeling reveals menu search is both random and systematic.* Paper presented at the Conference on Human Factors in Computing Systems, CHI '97, New York, NY.

Hutchins, E. (1994). *Cognition in the Wild*. Cambridge, MA: MIT Press.

Huynh, D., Karger, D., & Quan, D. (2002). *Haystack: A Platform for Creating, Organizing and Visualizing Information Using RDF*. Paper presented at the Semantic Web Workshop.

Jeff Johnson , T. L. R., William Verplank , David C. Smith , Charles H. Irby , Marian Beard , Kevin Mackey. (1989). The Xerox Star: A Retrospective,. *Computer, 22*(9), 11-26, 28-29.

John, B. E., & Kieras, D. E. (1996). The goms family of user interface analysis techniques: Comparison and contrast. *ACM Transactions on Computer-Human Interaction, 3*, 320-351.

Johnson, S. B. (2005). Tool for Thought.

Jones, W. (1986a). The memory extender personal filing system. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 298-305). Boston, Massachusetts, United States: ACM Press.

Jones, W. (1986b). On the applied use of human memory models: The Memory Extender personal filing system. *International Journal of Man Machine Studies, 25*, 191 228.

Jones, W. (1988). "As we may think"?: Psychological considerations in the design of a personal filing system. In *Cognitive Science and its Application for Human/Computer Interaction R. Guindon (Ed.).* Lawrence Erlbaum, NJ.

Jones, W. (2004). Finders, keepers? The present and future perfect in support of personal information management. *First Monday,* *http://www.firstmonday.dk/issues/issue9_3/jones/index.html*.

Jones, W., Bruce, H., & Dumais, S. (2001, November 5-10). *Keeping founds things found on the web.* Paper presented at the Tenth International Conference on Information and Knowledge Management (CIKM 2001), Atlanta, Ga.

Jones, W., Bruce, H., & Dumais, S. (2003, September 1-5, 2003). *How do people get back to information on the web? How can they do it better?* Paper presented at the 9th IFIP TC13 International Conference on Human-Computer Interaction (INTERACT 2003), Zurich, Switzerland.

Jones, W., & Dumais, S. (1986). The spatial metaphor for user interfaces - experimental tests of reference by location versus name. *ACM Transactions on Office Information Systems, 4*(1), 42-63.

Jones, W., Dumais, S., & Bruce, H. (2002, November 18-21). *Once found, what then? : a study of "keeping" behaviors in the personal use of web information.* Paper presented at the 65th Annual Meeting of the American Society for Information Science and Technology (ASIST 2002), Philadelphia, PA.

Jones, W., & Maier, D. (2003). *Report from the Session on Personal Information Management. Workshop of the Information and Data Management Program.* Seattle, WA: National Science Foundation Information.

Jones, W., Munat, C & Bruce, H. (2005, October 28 - November 2, 2005). *The Universal Labeler: Plan the Project and Let Your Information Follow (accepted).* Paper presented at the 68th Annual Meeting of the American Society for Information Science and Technology (ASIST 2005), Charlotte, North Carolina.

Jones, W., Phuwanartnurak, A. J., Gill, R & Bruce, H. (2005). *Don't Take My Folders Away! Organizing Personal Information to Get Things Done.* Paper presented at the Conference on Human Factors in Computing Systems (CHI 2005), Portland, OR.

Jones, W. P., & Landauer, T. K. (1985). Context and self-selection effects in name learning. *Behaviour & Information Technology, 4*(1), 3-17.

Kaptelinin, V. (1996, April 11-13). *Creating computer-based work environments : an empirical study of Macintosh users.* Paper presented at the SIGCPR/SIGMIS'96 : Proceedings of the 1996 conference on ACM SIGCPR/SIGMIS, Denver, CO.

Kaptelinin, V. (2003). *Integrating tools and tasks: UMEA: translating interaction histories into project contexts.* Paper presented at the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2003), Ft. Lauderdale, Florida.

Karat, J., Karat, C. and Brodie, C. (2005). Privacy in Information Technology: Designing To Enable Privacy Policy Management in Organizations. *International Journal of Human Computer Studies*.

Karger, D. R., & Quan, D. (2004). Collections: flexible, essential tools for information management In *CHI '04 extended abstracts on Human factors in computing systems* (pp. 1159-1162). Vienna, Austria ACM Press.

Kidd, A. (1994, April 24-28). *The marks are on the knowledge worker.* Paper presented at the CHI 94 : ACM SIGCHI Conference on Human factors in Computing Systems, Boston, Mass.

Kidd, A. (1994). The marks are on the knowledge worker In *Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence* (pp. 186-191). Boston, Massachusetts, United States ACM Press.

Kieras, D. E., & Meyer, D. E. (1997). An overview of the epic architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction,*, 391-438.

Kirsh, D. (1995). The Intelligent use of space. *Artificial Intelligence, 73*(1-2), 31-68.

Kirsh, D. (2000). A few thoughts in cognitive overload. *Intellectia*(30).

Kwasnik, B. H. (1989, June 25-28, 1989). *How a personal document's intended use or purpose affects its classification in an office.* Paper presented at the 12th Annual ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 1989), Cambridge, MA.

Kwasnik, B. H. (1991). The importance of factors that are not document attributes in the organization of personal documents. *Journal of Documentation, 47*(4), 389-398.

Lansdale, M. (1988). The psychology of personal information management. *Applied Ergonomics, 19*(1), 55-66.

Lansdale, M. (1991). Remembering about documents : memory for appearance, format, and location. *Ergonomics, 34*(8), 1161-1178.

Lansdale, M., & Edmonds, E. (1992). Using memory for events in the design of personal filing systems. *International Journal of Man-Machine Studies, 36*, 97-126.

Larkin, J. S., H. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65-99.

Levy, D. (2001). *Scrolling forward : making sense of documents in the digital age*. New York, NY.: Arcade Publishing.

Lucas, P. (2000a). *Pervasive information access and the rise of human-information interaction.* Paper presented at the CHI 2000: Human Factors in Computing Systems, The Hague.

Lucas, P. (2000b). *Pervasive information access and the rise of human-information interaction.* Paper presented at the Human Factors in Computing Systems, CHI 2000, The Hague.

Luo, L., & John, B. E. (2005). *Predicting task execution time on handheld devices using the keystroke-level model.* Paper presented at the CHI 2003, ACM Conference on Human Factors in Computing Systems, Late Breaking Results,, Portland, OR.

Mackay, W. E. (1988). *More than Just a Communication System: Diversity in the Use of Electronic Mail.* Paper presented at the Conference on Computer-Supported Cooperative Work.

Malone, T. W. (1983). How do people organize their desks: implications for the design of office information-systems. *ACM Transactions on Office Information Systems, 1*(1), 99-112.

Mander, R., Salomon, G., & Wong, Y. Y. (1992, May 3-7). *A 'pile' metaphor for supporting casual organization of information.* Paper presented at the CHI 92 : ACM SIGCHI Conference on Human Factors in Computing, Monterey, CA.

Manguel, A. (1996). *A History of Reading*. New York:: Viking.

Marchionini, G. (1995). *Information seeking in electronic environments*: Cambridge University Press.

Marchionini, G. K., A. (1998). Design of interfaces for information seeking. *Annual Review of Information Science and Technology, 33*, 89.

Marshall, C. C. (2006). Reading and Interactivity in the Digital Library: Creating an Experience that Transcends Paper,. In D. B. M. a. G. George (Ed.), *Digital Library Development: The View from Kanazawa*. Westport, Conn: Libraries Unlimited, Greenwood Publishing Group.

Marshall, C. C., & Bly, S. (2005). Saving and using encountered information: implications for electronic periodicals In *Proceeding of the SIGCHI conference on Human factors in computing systems* (pp. 111-120). Portland, Oregon, USA ACM Press.

Mary Czerwinski , D. G., Jim Gemmell, Tiziana Catarci, Cathy Marshall, Manuel Perez-Quinones & Meredith M Skeels. (2006). Digital Memories & Ubiquitous Computing. *Communications of the Acm, Special Issue on Personal Information Management*.

Namioka, D. S. a. A. (Ed.). (1993). *Participatory Design: Principles and Practices*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Nardi (ed.), B. (1996). *Context and Consciousness: Activity Theory and Human-Computer Interaction*. Cambridge, MA: MIT Press.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century Crofts.

Nelson, T. H. (1965). *File Structure for the Complex, the Changing, and the Indeterminate.* Paper presented at the ACM National Conference.

Newell, A., Shaw, J.C., & Simon, H.A. (1958). Elements of a theory of human problem solving. *Psychological Review, 65*, 151-166.

Newell, A., & Simon, H. A. (1972). *Human Problem Solving, .* Englewood Cliffs, NJ: Prentice-Hall.

Newfield, D., Sethi, B. S., & Ryall, K. (1998, November 1-4). *Scratchpad : mechanisms for better navigation in directed web searching.* Paper presented at the UIST 98 : 11th Annual ACM Symposium on User Interface Software and Technology, San Francisco, CA.

Niedzviecki), S. M. w. H. (2001). *Cyborg: Digital Destiny and Human Possibility in the Age of the Wearable Computer*: Randomhouse (Doubleday).

Norbert Streitz, P. N. (2005). Special issue: The disappearing computer. *Communications of the ACM, 43*(3).

Norman, D. A. (1988). *The Psychology of Everyday Things*: Basic Books.

O'Connail, B. F., D. (1995a). *Timespace in the workplace: Dealing with interruptions.* Paper presented at the CHI '95 Conference on Human Factors in Computing Systems, Extended Abstracts.

O'Connail, B. F., D. (1995b). *Timespace in the workplace: Dealing with interruptions.* Paper presented at the CHI '95 Conference on Human Factors in Computing Systems, Extended Abstracts.

O'Day, V., & Jeffries, R. (1993). *Orienteering in an information landscape: How information seekers get from here to there.* Paper presented at the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 1993).

Olson, J. S., Grudin, J., & Horvitz, E. (2005). A study of preferences for sharing and privacy In *CHI '05 extended abstracts on Human factors in computing systems* (pp. 1985-1988). Portland, OR, USA ACM Press.

Omar Aftab, P. C., Austin Kim, Sneha Thakkar, Neelima Yeddanapudi. (2001). *Information Theory and the Digital. Age*. Cambridge, MA: Massachusetts Institute of Technology.

Peterson, W. W., Birdsall, T. G., & Fox, W. C. (1954). The theory of signal detectability. *Institute of Radio Engineers Transactions, PGIT-4*, 171-212.

Pettigrew, K. E., Fidel, R., & Bruce, H. (2001). Conceptual frameworks in information behavior. *Annual Review of Information Science and Technology, 35*, 43-78.

Pirolli, P. (2006). Cognitive Models of

Human-Information Interaction. In R. S. N. F. T. Durso, R. W. Schvaneveldt, S. T. Dumais, D. S. Lindsay & M. T. H. Chi (Ed.), *Handbook of applied cognition* (2nd ed.). West Sussex, England: John Wiley & Sons.

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review, 106*(4), 643-675.

Pirolli, P., & Card, S. K. (1999). .Information foraging. *Psychological Review, 106*, 643-675.

Pirolli, P., & Fu, W. (2003). *Snif-act: A model of information foraging on the world wide web.* Paper presented at the User modeling 2003, 9th international conference, UM 2003, Johnstown, PA.

Povey, D. (1999). *Optimistic security: A new access control paradigm.* Paper presented at the Proc. ACM New Security Paradigms Workshop.

Qu, Y., & Furnas, G. W. (2005). Sources of structure in sensemaking In *CHI '05 extended abstracts on Human factors in computing systems* (pp. 1989-1992). Portland, OR, USA ACM Press.

Quan, D., Bakshi, K., Huynh, D., & Karger, D. R. (2003). *User interfaces for supporting multiple categorization.* Paper presented at the Interact 2003.

Quan, D., Huynh, D., & Karger, D. R. (2003). *Haystack: A Platform for Authoring End User Semantic Web Applications.* Paper presented at the 2nd International Semantic Web Conference (ISWC 2003).

Ranganathan, S. R. (1965). *The Colon classification.* New Bruswick, N.J.,: Graduate School of Library Service, Rutgers, the State University.

Rasmussen, J. (1986). *Information processing and human-machine interaction: an approach to Cognitive Engineering.* New York, NY: North-Holland.

Rasmussen, J., Pejtersen, A.M. and Goodstein, L.P. (1994). *Cognitive Systems Engineering.* (Vol. Wiley). New York, NY.

Ravasio, P., Sch\&\#228;r, S. G., & Krueger, H. (2004). In pursuit of desktop evolution: User problems and practices with modern desktop systems *ACM Trans. Comput.-Hum. Interact. , 11* (2), 156-180

Rhodes, B. a. M., P. (2000). Just-in-time information retrieval. *IBM Systems Journal, 39*(3-4), 685-704.

Ross, B. H. (1999). Postclassification category use: The effects of learning to use categories after learning to classify. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 25*(3), May 1999, 1743-1757.

Ross, B. H. (2000). The effects of category use on learned categories. *Memory & Cognition, 28*, 51-63.

Rouse, W. B. R. a. S. H. (1984). Human Information Seeking and Design of Information Systems. *Information Processing & Management,, 20*, 129-138.

Russell, D. M., Stefik, M. J., Pirolli, P., & Card, S. K. (1993). The cost structure of sensemaking In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 269-276). Amsterdam, The Netherlands ACM Press.

Segal, R. B., & Kephart, J. O. (1999). *MailCat: An Intelligent Assistant for Organizing E-mail.* Paper presented at the Third Annual Conference on Autonomous Agents.

Sellen, A. J., & Harper, R. H. R. (2002). *The myth of the paperless office*. Cambridge, MA.: MIT Press.

Sellen, A. J., Louie, G, Harris, J.E. & Wilkins, A.J. (1996). What brings intentions to mind? An in situ study of prospective memory. *Memory & Cognition, 5*(4), 483-507.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal, 27*, 379–423, 623–656.

Shannon, C. E., & Weaver, W. T. (1949/1963). *he mathematical theory of communication*. Urbana-Champaign, IL: University of Illinois Press.

Simon, H. A. (1969). *The Sciences of the Artificial*. Cambridge, Massachusetts: MIT Press.

Simon, H. A., & Newell, A. (1958). Heuristic problem solving: The next advance in operations research. *Operations Research, 6*, 1-10.

Steve Mann, A. S., and James Fung. (2004, Oct. 15, 2004). *Continuous Lifelong Capture of Personal Experience Using Eyetap.* Paper presented at the First ACM Workshop on Continuous Archival and Retrieval of Personal Experiences (CARPE '04), New York, NY, USA.

Stevens, G. W., V. (2002). *A new dimension in access control: Studying maintenance engineering across organizational boundaries.* Paper presented at the Proc. CSCW 2002.

Stiemerling. O., W., V. (2000). Beyond 'yes or no'-Extending access control in groupware with awareness and negotiation. *Group Decision and Negotiation, 9*, 221-235.

Suchman, L. (1983). Office procedure as practical action : models of work and system design. *ACM Transactions on Office Information Systems, 1*(14), 320-328.

Tauscher, L. M., & Greenberg, S. (1997a). How people revisit web pages : empirical findings and implications for the design of history systems. *International Journal of Human-Computer Studies, 47*(1), 97-137.

Tauscher, L. M., & Greenberg, S. (1997b). *Revisitation patterns in World Wide Web navigation.* Paper presented at the Conference on Human Factors in Computing Systems (CHI 197), Atlanta, GA.

Teevan, J. (2003, September, 2003). *"Where'd It Go?": Re-Finding Information in the Changing Web.* Paper presented at the Student Oxygen Workshop, Gloucester, MA,.

Teevan, J., Alvarado, C., Ackerman, M. S., & Karger, D. R. (2004). *The Perfect Search Engine Is Not Enough: A Study of Orienteering Behavior in Directed Search.* Paper presented at the ACM Conference on Human Factors in Computing Systems (CHI '04), Vienna, Austria.

Teevan, J., Dumais, S. T., Horvitz, E. (2005). *Personalizing Search via Automated Analysis of Interests and Activities.* Paper presented at the SIGIR 2005, Salvador, Brazil.

Terry, W. S. (1988). Everyday forgetting: Data from a diary study. *Psychological Reports, 62*, 299-303.

Tony Buzan, B. B. (2004). *The Mind Map Book: How to Use Radiant Thinking to Maximize Your Brain's Untapped Potential*.

Vicente, K. J. (1999). *Cognitive work analysis*. Mahwah, NJ:: Lawrence Erlbaum.

Whittaker, S., & Hirschberg, J. (2001). The character, value and management of personal paper archives. *ACM Transactions on Computer-Human Interaction, 8*(2), 150-170.

Whittaker, S., & Sidner, C. (1996, April 13-18). *Email overload: exploring personal information management of email.* Paper presented at the Conference on Human Factors in Computing Systems (CHI 1996), Vancouver, B.C.

Whittaker, S., Terveen, L., & Nardi, B. A. (2000). Let's stop pushing the envelope and start addressing it: a reference task agenda for HCI. *Human Computer Interaction*(15), 75-106.

Whitten, A. a. T. J. D. (1999, August, 1999). *Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0.* Paper presented at the Proceedings of the 9th USENIX Security Symposium.

Williamson, A. G., & Bronte-Stewart, M. (1996, August). *Moneypenny : things to do on the desk.* Paper presented at the HCI 96 : adjunct proceedings of the 11th BCS Annual Conference on HCI, London.

Wilson, E. V. (2002). Email winners and losers. *Communications of the ACM, 45*(10), 121-126.

Xin Dong, A. Y. H. (2005). *A Platform for Personal Information Management and Integration.* Paper presented at the CIDR.

Yates, F. A. (1966). *The Art of Memory*. Chicago: University of Chicago Press.

Yates, J. (1989). *Control Through Communication: The Rise of System in American Management,* . Baltimore: Johns Hopkins University Press.