Personal Information Management

William Jones
University of Washington

Introduction

Personal Information Management (PIM) refers to both the practice and the study of the activities a person performs in order to acquire or create, store, organize, maintain, retrieve, use, and distribute the information needed to complete tasks (work-related or not) and fulfill various roles and responsibilities (for example, as parent, employee, friend, or community member). PIM places special emphasis on the organization and maintenance of personal information collections (PICs) in which information items, such as paper documents, electronic documents, e-mail messages, Web references, and handwritten notes, are stored for later use and repeated reuse.

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 needs. Tools and technologies help us spend less time with labor-intensive and error-prone information management activities (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 is now available for the management of personal information. But this diversity has become part of the problem, leading to *information fragmentation* (Jones, 2004). A person may maintain several separate, roughly comparable but inevitably inconsistent, organizational schemes for electronic documents, paper documents, e-mail messages, and Web references. The number of organizational schemes may increase if a person has several e-mail accounts, uses separate computers for home and work, makes use of a personal digital assistant (PDA) or a smart phone, or uses any of a bewildering array of special-purpose PIM tools.

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

schemata for files, e-mail messages, or Web references. Users reasonably complain that this is one organization too many (Boardman & Sasse, 2004; 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. Many excellent studies have focused on uses of, and possible improvements to, e-mail (for example, Bälter, 2000; Bellotti, Ducheneaut, Howard, Neuwirth, & Smith, 2002; Bellotti, Ducheneaut, Howard, & Smith, 2003; Bellotti, Ducheneaut, Howard, Smith, & Grinter, 2005; Bellotti & Smith, 2000; Ducheneaut & Bellotti, 2001; Gwizdka, 2000, 2002a, 2002b; Mackay, 1988; Whittaker, 2005; Whittaker & Sidner, 1996; Wilson, 2002). Other studies have examined the use of the Web or specific features such as bookmarks or history information (for example, Abrams, Baecker, & Chignell, 1998; Byrne, John, Wehrle, & Crow, 1999; Catledge & Pitkow, 1995; Tauscher & Greenberg, 1997a, 1997b). Yet other studies have considered the organization and retrieval of documents in paper and electronic form (for example, Carroll, 1982; Case, 1986; Malone, 1983; Whittaker & Hirschberg, 2001).

Research that focuses on people and what they want or need to be able to do with their information also comes under the PIM umbrella. The completion of a task depends critically on certain information: For example, returning a telephone call may require knowing a person's first name and telephone number. Thus, the study of how people manage various tasks in their lives is relevant to PIM (Bellotti, Dalal, Good, Flynn, Bobrow, & Ducheneaut, 2004; Bellotti et al., 2003; Bellotti et al., 2005; Czerwinski, Horvitz, & Wilhite, 2004; Gwizdka, 2002a; Matthews, Czerwinski, Robertson, & Tan, 2006; Whittaker, 2005; Williamson & Bronte-Stewart, 1996). Research into digital memories (Gemmell, Bell, Lueder, Drucker, & Wong, 2002) and the "record everything" and "compute anywhere" possibilities enabled by advances in hardware are also germane (Dempski, 1999; Lucas, 2000).

The past few years have seen a revival of interest in PIM as an area of serious inquiry that draws upon the best work from a range of disciplines including cognitive psychology, human–computer interaction, database management, artificial intelligence, information and knowledge management, information retrieval, and information science.[1]

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? (See Czerwinski, Gage, Gemmell, Marshall, Pérez-Quiñones, Skeels, et al., 2006 for a recent review.) 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 of computing

devices can make it possible to take our information with us wherever we go and 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.

However, renewed interest in PIM is also spurred by the awareness that developments in technology and tools, for all their promise, invariably create new problems and sometimes exacerbate old ones, too. Information that once existed only in paper form is now scattered in multiple versions in both 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 and applications—for all the help they provide—can further complicate the challenge of information management.

The Problems of PIM

In the real world, we do not always find the right information in time to meet our current needs. The necessary information may never be found or it may arrive 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 are 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 a different format so that the hassles of 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 at hand. We may fail to make effective use of information even when it is directly in view.

These are failures of PIM. Some of these may be memorable. 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 e-mail message—that we know is "there somewhere." Over the course of an already busy day, we may spend precious minutes, sometimes hours, looking for lost information. Other failures of PIM may go unnoticed as part of background *information friction* associated with getting things done. In his highly influential article, "Man–Computer Symbiosis," Licklider (1960, p. 4) made the following observations about his own work day:

> About 85 per cent of my "thinking" time was spent getting into a position to think, to make a decision, to learn something I needed to know. ... My choices of what to attempt and what not to attempt were determined to an embarrassingly

great extent by considerations of clerical feasibility, not intellectual capability.

Many of us might reach similar conclusions concerning our own interactions with information. A seemingly simple e-mail 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, e-mail messages, Web references, and so on. Can you give a presentation at a meeting next month? That depends. ... What did you say in previous e-mail messages? When is your child's soccer match? Better check the paper flyer with scheduled games. Does the meeting conflict with an upcoming conference? Better check the conference Web site to get dates and program information. What have you already scheduled in your calendar? And so on. In their observations of people processing e-mail, Belloti et al. (2005) have noted instances in which a single e-mail message initiates a task involving several different software applications and lasting an hour or more.

The Potential of PIM

Information is a means to an end. Not always, not for everyone, but often. We manage information in order to have it when we need it—to complete a task, for example. Information is not an inherently precious resource. In truth, 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 (1971, p. 40) elegantly expressed this point with respect to resource optimization:

> What 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.

This quotation still rings true even if we replace "attention" with "time," "energy," or "well-being." Certainly the nagging presence of papers representing unpaid bills, unanswered letters, or unfiled documents can distract, enervate, and demoralize. We cannot "see" our well-being, our attention, 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 e-mail messages, and other forms of information. It is through these personal information items that we seek to manage the precious resources of our lives.

The payoffs for advances in PIM are large and varied:

- For each of us as individuals, better PIM allows us to make better use of our precious resources (time, money, energy, attention) and thus, ultimately, improves the quality of our lives.

- Within organizations, better PIM means better employee productivity and teamwork in the near term. Over time, PIM is key to the management and leveraging of employee expertise.

Advances in PIM may also translate into:

- Improvements in information literacy programs (Eisenberg, Lowe, & Spitzer, 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 significant in domains such as intelligence analysis or medical informatics. Better PIM may help doctors and nurses to balance a large and varied caseload. Potentially of even greater impact may be PIM support for individuals undergoing long-term or sustained treatments for chronic or acute health conditions (Pratt, Unruh, Civan, & Skeels, 2006).

Objectives, Scope, and Structure for this Chapter

The remainder of this chapter covers the following:

- *Influences on PIM* reviews key historical influences on the study of PIM and also considers the considerable synergistic overlap with existing disciplines including cognitive science, information science, and human–computer interaction.

- *Analysis of PIM* introduces key concepts of PIM and its conceptual framework, which, in turn, provides the organizational structure of the subsequent review of PIM-related research.

- *Research to Understand How People Do PIM* reviews research squarely focused on PIM and also considers a sampling from a much larger collection of PIM-related research.

- *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.

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

The chapter concludes with a return to a key problem of PIM: information fragmentation. Research relating to PIM is similarly fragmented. The progress in PIM depends upon an integrated approach involving several fields of inquiry. This progress, in turn, may promote important integrations in the practice of PIM.

Influences on PIM

Broadly defined, PIM includes the management of information going into our own memories as well as 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, Yates, 1966).

Although definitions of PIM vary (see the section "Analysis of PIM"), they generally include, as a central component, the management of external forms of information. The difficulties of managing paper-based information have long been recognized and tools have been developed over time to address these challenges. Yates (1989) notes, for example, that the vertical filing cabinet that is now such a standard (if increasingly old-fashioned) feature of office, home, and workplace was first commercially available in 1893.

The modern dialogue on PIM probably began with Vannevar Bush's inspirational article "As We May Think" (Bush, 1945), in which he presented his vision of a memex device that would greatly increase a person's ability to record, retrieve, and interrelate information (see the chapter by Houston and Harmon in the present volume). Licklider (1960, 1965), Engelbart (1963), and Nelson (1982) each advanced the notion that the computer could be used to extend the human ability to process information and, even, to enhance the human intellect. The phrase "Personal Information Management" was apparently first used in the 1980s (Lansdale, 1988) in the midst of general excitement over the potential of the personal computer to augment the human ability to process information (Goldstein, 1980; Johnson, Roberts, Verplank, Smith, Irby, Beard, et al., 1989; Jones, 1986). The 1980s also saw the advent of so-called "PIM tools" that provided limited support for the management of appointments and scheduling, to-do lists, telephone numbers, and addresses.

A community dedicated to the study and improvement of human–computer interaction (HCI) also emerged in the 1980s (Card, Moran, & Newell, 1983; Norman, 1988) and much of the applied research reviewed in this chapter was initiated by practitioners in this field. However, much HCI research has remained focused on specific forms of information (e.g., e-mail messages, Web pages, digital photographs), specific devices to aid interaction, and, increasingly, group and organizational issues. The study of PIM focuses primarily on the individual but also broadens to include key interactions with information over time and across tools. PIM considers our personal use of

information in all of its various forms—including paper. Although today it is difficult to imagine a practice of PIM that does not involve computers, *information* in all its forms is the primary focus.

In recent years, there has been discussion of human–information interaction (HII) by way of contrast to HCI (Fidel & Pejtersen, 2004; Gershon, 1995; Lucas, 2000; Pirolli, in press). Interest in HII is due in part to a realization that our interactions with information are more central to our lives than are our interactions with computers. This realization is reinforced by developments in ubiquitous computing. Success in computing and, perhaps paradoxically, in HCI, may mean that the computer will come to "disappear" (Streitz & Nixon, 2005) into the background of our daily lives, much as electricity currently does. With "transparent interfaces," we are left with information. Much in HII remains to be defined, but when this happens, PIM will likely be an important element.

The study of human cognition also informs, and is informed by, PIM. The common ground shared by PIM and cognitive science is considerable and largely unexplored. Of relevance are not only the classic findings of cognitive psychology (e.g., Neisser, 1967) but also more recent work on situated cognition, distributed cognition, and social cognition (e.g., Fiske & Taylor, 1991; Hutchins, 1994; Suchman, 1987). Also very relevant is the study of affordances provided by the environment and by the "everyday" (and often overlooked) objects of a person's environment (Gibson, 1977, 1979; Norman, 1988, 1990, 1993).

Synergies with the field of information science and the study of human information behavior are largely unrealized. For example, the work by Erdelez and Rioux (2000) on information encountering has clear relevance to an essential decision of PIM—whether and how to keep new information. To take another example, Dervin's (1992, 1999) work on sense-making certainly relates to a person's efforts to maintain and organize personal information collections (PICs) over time. The large subfield of information seeking, although focused on the retrieval of public information from external sources (e.g., a conventional library or the Web), certainly relates to the PIM activities of finding and refinding (see Pettigrew, Fidel, & Bruce, 2001).

The study of information and knowledge management in organizations also has relevance to the study of PIM (e.g., Garvin, 2000; Selamat & Choudrie, 2004; Taylor, 2004; Thompson, Levine, & Messick, 1999). Issues seen first at an organizational level often migrate to the PIM domain. The merits of various schemes of classification or the use of controlled vocabularies, for example, have long been topics of discussion at the organizational level (Fonseca & Martin, 2004; Rowley, 1994). But these topics may find their way into the realm of PIM as the amounts of personally held digital information continue to increase. This migration has already happened with regard to privacy, protection, and security (e.g., Karat, Brodie, & Karat, 2006).

Several other fields, including information retrieval, database management, and artificial intelligence, have potential relevance to the development of supporting tools for PIM. A proper review of PIM and its overlap with any one of the fields just mentioned would require a chapter in its own right. This review focuses on core activities of PIM, the challenges people face in the completion of these activities, and, in a much more limited way, approaches to the support of these activities.

Analysis of PIM

A deeper understanding of what PIM is begins with definitions and core concepts. This section sets out a conceptual framework that helps to connect several key concepts of PIM and compares PIM to related fields of inquiry.

Information and the Information Item

The question of what information "is" has been a topic of repeated discussion, excellent overviews of which have been provided by Cornelius (2002) and Capurro and Hjørland (2003) in recent volumes of the *Annual Review of Information Science and Technology (ARIST)*.

This chapter focuses on the capacity of information to effect change in our lives and in the lives of others. The information we receive influences our actions and our choices. For example, we decide which of several hotels to book based on the information we are able to gather concerning price, location, availability, and so on. 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 effect change. We send information in the clothes we choose to wear, in 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 (1948) original notion 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. E-mail messages

4. Web pages

5. References (e.g., shortcuts, aliases) to any of the above

Items encapsulate information in a persistent form that can be created, stored, moved, given a name and other properties, copied, distributed, deleted, transformed, and so forth.

Our interactions with paper-based information items are supported by, among other things, the desktop, paper clips, staplers, and filing cabinets. Our interactions with digital information items depend upon the support of various computer-based tools and applications such as an e-mail application, a file manager, and a Web browser. The "size" of current information items is determined in part by these applications. There are certainly situations in which some of us might like information items to be packaged in smaller units. A writer, for example, might like to treat paragraphs or even individual sentences as information items to be reaccessed 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, e-mail messages, and Web bookmarks.

Consider how much of our interaction with the world 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 the weather (perhaps before we even bother to look outside). We learn of meetings via e-mail messages and receive the documents for these meetings via e-mail as well.

As regards sending information items, we fill out Web-based forms; we send e-mail 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 proxies. We project ourselves and our desires across time and space in ways that would never have occurred to our forebears.

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

Personal Information

The term *personal information* has several senses:

1. The information people keep for their own personal use.

2. Information about a person kept by and under the control of others. Doctors and health maintenance organizations, for example, maintain health information about their patients.

3. Information experienced by a person but outside his or her control. The book a person browses (but puts back) in a traditional

library or the pages a person views on the Web are examples of
this kind of personal (or personally experienced) information.

This chapter is concerned primarily with the first sense of personal
information.

A Personal Space of Information

A personal space of information (PSI) includes all the information
items that are, at least nominally if not exclusively, under an individ-
ual's control. A PSI contains a person's books and paper documents,
e-mail messages (on various accounts), e-documents, and other files (on
various computers). A PSI can contain references to Web pages as well
as include applications, tools (such as a desktop search facility), and con-
structs (e.g., associated properties, folders, "piles" in various forms) that
support the acquisition, storage, retrieval, and use of the information.

There are several other things to note about a PSI:

- Although people have some sense of control over the items in
  their PSIs, this is partly illusory. For example, once an e-mail
  message has been deleted, it will no longer appear in one's inbox;
  however, the message is very likely still in existence (as some fig-
  ures 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 computers) and
  the bookmarks we create to reference these pages.

- A person has only one PSI.

Personal Information Collections

Several researchers have discussed the importance of collections in
managing personal information. Karger and Quan (2004) define a col-
lection quite broadly, taking it to comprise a variety of objects ranging
from menus, to portals, to public taxonomies. Boardman (2004, p. 15)
understands 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."

The characteristic features of a personal information collection (PIC)
will be listed here but no attempt will be made to provide a formal defi-
nition. A PIC might best be characterized as a personally managed sub-
set of a PSI. PICs are "islands" in our PSIs where we have made some
conscious effort to control both the information that goes in and the
manner in which it is organized. PICs can vary greatly with respect to
the number, form, and content coherence of their items. Examples of
PICs include:

- The papers in a well-ordered office and their organization, including the layout of piles on a desktop and the folders inside filing cabinets.

- The papers in a specific filing cabinet and their organizing folders (when perhaps the office as a whole is a mess).

- Project-related information items that are initially "dumped" into a folder on a notebook computer and then organized over time.

- A carefully maintained collection of bookmarks to useful reference sites on the Web.

- An EndNote database of article references.[2]

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 take the same form—all will be e-mail messages, for example, or all files. But this is not a necessary feature of a PIC. Later on, we review research aimed at supporting an integrative organization of information items, regardless of form. Such efforts aim at building a "form-neutral" layer of support for the management of information items.

The concept of a PIC will prove useful as we review research on the ways people approach the organization of their information. Statements such as "I've got to get my ____ organized!" often refer to a PIC. The organization of "everything" in one's PSI is a daunting, perhaps impossible, task. But we can imagine organizing our Web bookmarks, our e-mail inbox, or our laptop filing system (but probably only selected areas thereof).

Definitions of PIM

PIM is easy to describe and discuss, for we all do it and we all have had first-hand experiences with its challenges. But it is much harder to define.

Lansdale (1988, p. 55) defined PIM as "the methods and procedures by which we handle, categorize and retrieve information on a day-to-day basis," whereas Bellotti et al. (2002, p. 182) understood it to be "the ordering of information through categorization, placement, or embellishment in a manner that makes it easier to retrieve when it is needed." Barreau (1995, p. 327) characterized 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" (p. 327). Recently, Boardman (2004, p. 13) noted 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

guided 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 the input-storage-output model of actions associated with a PSI, the framework used in this chapter to help organize its discussion of PIM-related research will provide the following grouping of essential PIM activities:[3]

- *Finding / refinding activities* move from need to information and affect the output of information from a PSI.

- *Keeping activities* move from information to need and affect the input of information into a PSI.

- *Meta-level activities* focus on the PSI itself and on the management and organization of PICs within it. Efforts to "get organized" in a physical office, for example, constitute one kind of meta-level activity.

The remainder of this review is guided by a framework that derives from a basic assumption—namely, that *PIM activities help to establish, use, and maintain a mapping between information and need.* This simple statement can be expanded and the relationship between the various PIM activities visualized by reference to the diagram in Figure 10.1. Needs, as depicted in the leftmost column, can be expressed in several ways. The need may, more or less, originate internally—that is, within a person as she recalls, for example, that she needs to make plane reservations for an upcoming trip. Or it may be derived from an external source—for example, a question from a colleague in the hallway or a manager's request. Needs are evoked by an information item such as an e-mail message or a Web-based form.

Information, as depicted in the rightmost column, is also expressed in various ways—for example, as aural comments from a friend, as a billboard seen on the way to work, or via any number of information items including documents, e-mail messages, Web pages, and hand-written notes.

To make a connection between need and information is to create a mapping. Only small portions of the mapping have observable external representations. Much of the mapping has only a hypothesized existence in the memories of an individual: Indeed, large portions thereof 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 one from a need to desired information.

But parts of the mapping can be observed and manipulated. The folders of a filing system (whether for paper documents, electronic documents, e-mail messages, or Web references); the layout of a desktop (physical or virtual); and the choice of names, keywords, and other
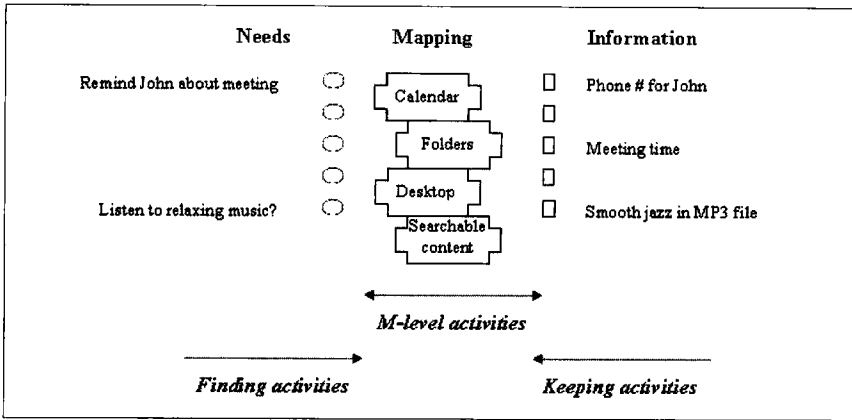
Figure 10.1 **PIM activities viewed as an effort to establish, use, and maintain a mapping between needs and information.**

properties for information items all form parts of an observable fabric helping to knit need to information.

Research to Understand How People Do PIM

Finding: From Need to Information

A person has a need and finds information in order to meet it. Needs can be large and amorphous—the need for information to complete a review of a research area, for example—or small and simple—the need for a telephone number. Many needs correspond to tasks (e.g., "get schedules and make airplane reservations"). But other needs may not fit tasks except by the broadest definition (e.g., "see that photograph of our vacation again").

Wilson's (2000, p. 49) definition for *information seeking* applies equally well to *information finding*, or simply *finding* as used in this chapter:

> the purposive seeking for information as a consequence of a need to satisfy some goal. In the course of seeking, the individual may interact with manual information systems (such as a newspaper or a library), or with computer-based systems (such as the World Wide Web).

In their efforts to meet a need, people seek. They search, sort, and browse; they 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. These activities are all examples of finding activities. Finding includes both acts of new finding, where there is no previous memory of the

needed information, and acts of refinding. The information found can be personal, residing in a PSI, or public, originating outside of the PSI.

There are several reasons for preferring the term "information finding," or "finding," to that of "information seeking" in relation to PIM:

- Although Wilson's definition of information seeking is inclusive, research on information research has tended to focus primarily on efforts to find information outside a PSI—from a "brick and mortar" library, for example, or from the Web (Pettigrew et al., 2001).

- "Finding" more directly expresses the goal of a finding activity: the location of items meeting a current need.

- People find, or try to find, not only information items but also physical items such as their car keys, cell phones, or television remote controls.

- The act of "finding" is complementary to that of "keeping." "Finders, keepers," as the saying goes: What we find, we can (try to) keep. With both physical items and information items, there is often a trade-off between investing more time now to keep or more time later to find. For example, time can be invested now to carefully pair the socks in a pile of freshly washed laundry—an act of keeping. Or, instead, more time can be spent later to find a matching pair of socks within the pile in order to meet the current need (e.g., nicer black socks for a business meeting).

This chapter focuses on *refinding private information*—that is, situations in which people are attempting to return to information they believe is in their PSI. But other variations of information finding are also PIM activities as discussed briefly here.

Finding and Refinding Public Information

There is an impressive body of work on information seeking and information retrieval that applies especially to finding public information (see, for example, Marchionini, 1995; Marchionini & Komlodi, 1998; Pettigrew et al., 2001; Rouse & Rouse, 1984); however, a comprehensive review of this literature is beyond the scope of this chapter.

There is a strong personal component in efforts to find new information 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 PSI. And information inside the PSI can be used to support a more targeted, personalized search of the Web (e.g., Teevan, Dumais, & Horvitz, 2005).

An online search to meet a need for information is often a sequence of interactions rather than a single transaction. Bates (1989) has presented a *berrypicking* model of online searching according to 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, Alvarado, Ackerman, and Karger (2004) note that users often favor a stepwise *orienteering* approach even in cases where the user knows where the information is and could presumably access it directly 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 lessen the cognitive burden associated with query articulation. The examples of berrypicking and orienteering suggest that it might be useful to preserve the search state within the PSI.

Finding (Discovery of) Personal Information

Items may enter a PSI automatically (e.g., via the inbox, automated downloads, Web cookies, the installation of new software). People may have no memory or awareness of the existence of these items. If they are ever retrieved, it is through an act of finding, not refinding. Memories of 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 refinding. Personal stores tend to become enormous over time: Some items may be decades old. As the use of integrative desktop search facilities increases, people may be surprised by the information they already "have."

Refinding Personal Information

The remainder of this section focuses on the refinding of information in the PSI. Clearly, the ability to refind information in a PSI is essential if people are to make effective use of their personal information. If an information item is in the PSI and people remember that the information item is there, it is often because of some earlier, explicit act of keeping. Failure to find information is frustrating in general but would appear to be especially so for information that we know "is in there somewhere."

Lansdale (1988) has described a two-step process involving an interplay between recall and recognition. Recall may constitute typing in a search string or even an exact address for the desired information. In other cases, it is less precise. A person may recall in which pile a paper document lies but not its exact location within that pile. Or one may have a rough idea when an e-mail message was sent or an electronic document last modified. In a second step, then, information items or a representation of these, as delimited by the recall step, are scanned and, if one is successful, the desired item is recognized and retrieved. The steps of recall and recognition can iterate to narrow progressively the search for the desired information—as happens, for example, when we move through a folder hierarchy to a desired file or e-mail message or when we navigate through a Web site to a desired page. The two steps of recall

and recognition can be viewed as a dialogue between people and their information environments.

But a successful outcome in a finding effort depends upon completion of another step preceding recall: A person must remember to look. One may know exactly where an item is and still forget to look for it in the first place. It is also useful to consider a final "repeat?" step, although this is essentially a variation of remembering to look. Meeting an information need often means assembling or reassembling a collection of information items relating to the task at hand. The finding activity must then be repeated until the complete set of items is collected.

Failure to collect a complete set of information can sometimes mean failure for the entire finding episode. For example, a person may collect three of four items needed in order to decide whether to accept a dinner invitation next week. She consults a paper flyer, an events Web site, and her online calendar and, then, seeing no conflicts, accepts. Unfortunately, she did not think to look at a fourth item—a previously sent e-mail in which she agreed to host a meeting of her book club that same evening.

Finding events—especially when directed to previously experienced personal information—can, therefore, be viewed as a four-step process with a possibility of failure at each step:

1. Remembering to look.

2. Recalling information about the information that can help to narrow the subsequent scan.

3. Recognizing the desired item(s).

4. Repeating as needed in order to "re-collect" the set of items required to meet the current need.

Remembering (To Look)

Many opportunities to refind and reuse information are missed simply because people forget to look. This failure occurs across information forms. In a study by Whittaker and Sidner (1996), for example, participants reported that they forgot to look inside to-do folders containing actionable e-mail messages. Because of mistrust in their ability to remember to look, people elected to leave actionable e-mail messages within an already overloaded inbox. Inboxes were often further loaded with copies of outgoing e-mail messages that might otherwise have been forgotten in a "sent mail" folder.

Web information is also forgotten. In one study of Web use, for example, participants often complained that, while engaged in non-targeted activities such as "spring cleaning," they encountered bookmarks that would have been very useful for a project whose time had now passed (Jones, Dumais, & Bruce, 2002). Another study reported that, when participants were cued to return to a Web page for which they had a Web bookmark, this bookmark was used in less than 50 percent of the trials

(Bruce, Jones, & Dumais, 2004). Marshall and Bly (2005) have observed a similar failure to look for paper information (newspaper clippings). Many of us have had the experience of writing a document and then later discovering a similar document that we had previously authored.

If the old adage "out of sight, out of mind" is frequently true, then one way to aid memory is to keep items in view. Reminding is an important function, for example, of paper piles in an office (Malone, 1983). E-mail messages in an inbox provide a similar function, at least until the messages scroll out of view (Whittaker & Sidner, 1996). Barreau and Nardi (1995) have observed that users often placed a file on their computer desktop in order to be reminded of its existence and of associated tasks to be completed.

Visibility helps. But a person must still be prepared to look. Piles on a physical desktop can, over time, recede into a background that receives scant attention. Likewise, as online advertisers surely know, people can learn to ignore portions of a computer's display. Also, the ability to manage items and keep them in view—whether on a computer screen or on the surfaces of a physical office—degrades, sometimes precipitously, as the number of items increases (see, for example, Jones & Dumais, 1986).

Attempts to compensate for the limitations of visible reminders can introduce other problems. People who adopt a strategy of repeatedly checking their e-mail inboxes in order to respond to messages before these scroll out of view (and out of mind) may end up "living" in their e-mail application with little time or attention left to accomplish work requiring sustained levels of concentration. People who immediately click through to interesting Web pages, for fear of forgetting to look at these later (even if they bookmark them) may let their Web use degenerate into an incoherent sequence of page views scattered across a wide range of topics with little to show for the experience.

A computer-based device might remind people of potentially useful information in many ways (Herrmann, Brubaker, Yoder, Sheets, & Tio, 1999) including, for example, the spontaneous execution of searches that factor in words and other elements of the current context (Cutrell, Dumais, & Teevan, 2006). However, such reminding devices, like visible space, compete for a very precious and fixed resource—a person's attention—and so must walk a fine line to avoid the extremes of either being annoying or being ignored.

Why is reminding so important in the first place? Why do people forget? Part of the answer goes back to a key problem of PIM: information fragmentation. Information items are scattered in different forms across various organizational devices. Support for grouping and interrelating items is not well developed. The folder, for example, has changed little in its basic function since its introduction, as part of the desktop metaphor, over 20 years ago. Support for grouping, interrelating, and, more generally, creating external representations (e.g., of tasks or projects) that might complement our internal representations is a topic of further discussion in both the keeping and meta-level sections of this chapter.

Recall and Recognition

Recall and recognition constitute two parts of a dialogue between a person and his information world. For example, somebody types a search word (recall) and then scans through a list of results (recognition). He clicks on a folder (recall) and then scans through a listing representing the items (e.g., e-mail messages, files, Web references) within the folder. He sorts inbox e-mail messages by sender (recall) and then scans through messages from "Sally" (recognition).

Even as desktop search utilities improve, a preference persists for returning to information through what is known as location-based finding, orienteering, or browsing (Barreau & Nardi, 1995; Marchionini, 1995; O'Day & Jeffries, 1993; Teevan, 2003). Habits change slowly and desktop search support continues to improve. For example, in the author's informal survey of persons who have installed and use an integrative desktop search facility (i.e., one able to search quickly across files, e-mail messages, recently visited Web sites), people still expressed a preference for browsing their desktops, "My Documents," or through their folders. Over 90 percent of the respondents indicated that they used their search facility only as a "last resort" after other methods had failed.

And yet, desktop search is becoming increasingly integrative and ever closer to an ideal in which anything that can be remembered about an information item or the circumstances surrounding encounters with it (e.g., time of last use or nearby "landmark" events) can be used to help find this item (Cutrell et al., 2006; Lansdale, 1988, 1991; Lansdale & Edmonds, 1992). It is possible, then, that people may gradually shift to a greater reliance on search.

But the reasons underlying the preference for browsing may be more basic. In response to a cue (such as an expression of information need) people are usually, but not always, better at recognizing an item from a set of alternatives than at recalling it (Tulving & Thomson, 1973). Browsing reduces and distributes the amount that must be recalled and relies more on recognition (Lansdale, 1988). Teevan et al. (2004) discuss additional considerations favoring what they term "orienteering," such as cognitive ease (smaller steps, less burden on working memory), sense of location (and a greater sense of control), and a richer context in which to recognize and understand results. Basic research underlines the importance of context in recognition (Tulving, 1983; Tulving & Thomson, 1973).

If one assumes that people remember to look, how difficult is it to return to an information item such as an e-document, e-mail message, or Web page that has been previously seen? In a study on delayed cued recall by Bruce et al. (2004), participants were asked to return to Web pages they had last visited up to six months prior by whatever means they chose. Participants did so quickly (retrieval times were under a minute on average) and with success rates approaching 100 percent. The small number of failures and time-out delays (less than five minutes)

that did occur seemed primarily due to information fragmentation. For example, one participant looked for a Web reference first in her "Favorites," then in selected e-mail folders, then in folders under "My Documents" before finally locating the Web reference inside a presentation she had saved to a network drive.

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* that has been identified in research in human cognition (Slamecka & Graf, 1978; see also Jones & Landauer, 1985). Thinking of a name for an item causes people 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).

We do not always name the information items in our PSIs. Abrams et al. (1998) report, for example, that when creating a Web bookmark, users rarely change the default name provided by the browser. However, 86 percent of users in their survey reported that the 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, 1997a, 1997b). The highlighting of search terms is now a standard feature of many search facilities.

Repeating?

In many instances, one needs to find not a single, isolated information item but rather a set of items whose members may be scattered in different forms within different organizations. In the dinner scheduling example given earlier, four different items needed to be retrieved in order to decide whether to accept the invitation. If the likelihood of successful retrieval of each item is strictly independent of the others, then the chances of successfully retrieving all the relevant items decreases as their number increases. So even if the likelihood of success for each item is, say, 95 percent, retrieval of all four items drops to only 81 percent. In situations of *output interference,* items retrieved first may interfere with the retrieval of later items in a set—perhaps because the act of retrieval itself strengthens recollection of the items first recalled at the expense of unrecalled items (Rundus, 1971). Some of us may experience this effect when we try to think of everyone in a group of eight or nine friends. No matter whom we list first—and this can vary from time to time—the last one or two people are often the hardest to remember.

The chances of successfully retrieving all members of a set can also be much better than predicted by a strict independence of individual retrievals. Obviously, retrieval improves if all items are in the same larger unit—a folder or a pile, for example. It may also be better than predicted by strict independence if the items comprising a set have an internal organization or are interrelated so that the retrieval of one item actually facilitates the retrieval of others (e.g., Bower, Clark, Lesgold, &

Winzenz, 1969; Jones & Anderson, 1987). One quotidian instance of what we might call *output facilitation* seems to occur, for example, when remembering the characters of a well-told story or a good movie. Of potential relevance are studies of information foraging and the notion of an *information scent* (Pirolli & Card, 1999) that might guide people from one to another of the items in a fragmented set.

Summary: Finding Is a Multi-Step Process

Finding is a multi-step process with a possibility of stumbling at each step. First, people must *remember* to look. An item is retrieved through variations of searching or, more commonly for items in the PSI, browsing. Both browsing and searching involve an iterative interplay between basic actions of *recall* and *recognition*. Finally, in many situations of information need, people must *repeat* the finding activity several times in order to "re-collect" a complete set of information items.

Keeping: From Information to Need

Many events of daily life are the converse of finding events: People encounter information and try to determine what, if anything, they should do with it—that is, people must match the information to anticipated need(s). Decisions and actions relating to encountered information are collectively referred to in this chapter as *keeping* activities.

People may encounter information unexpectedly (more or less): For example, they may come across an announcement for an upcoming event in the morning newspaper or an "FYI" e-mail with a pointer to a Web site may arrive in their inbox. The ability to handle effectively information that is encountered by happenstance may be key to one's ability to discover new material and make new connections (Erdelez & Rioux, 2000).

People also keep information that they expect to receive and have actively sought but do not have time to process in real time. A search on the Web, for example, often produces much more information than can be consumed in the current session. Both the decision to keep this information for later use and the measures taken to do so constitute keeping activities.

People keep information not only to have it available at a later point in time but also to remember to look for it. A failure to remember to use information that has been kept is one kind of *prospective memory failure* (Ellis & Kvavilashvili, 2000; O'Connail & Frohlich, 1995; Sellen, Louie, Harris, & Wilkins, 1996; Terry, 1988). People may, for example, self-e-mail a Web reference in addition to, or instead of, making a bookmark so that an e-mail message with the reference appears in the inbox, where it is more likely to be noticed and used (Jones et al., 2002).

Keeping, more broadly considered, applies not only to information but also to channels of information. Subscribing to a magazine or setting the car radio to a particular station is a keeping decision. Even the

cultivation of friends and colleagues can be seen as an act of keeping (and certainly friends and colleagues often represent important channels of information).

Keeping activities are triggered when people are interrupted in the course of performing a task and look for ways of preserving the current state so that work can be resumed quickly later on (Czerwinski et al., 2004). For example, people keep appointments by entering reminders into a calendar or record good ideas or "things to pick up at the grocery store" by writing down a few cryptic lines on a loose piece of paper. For some professionals, task interruptions have been observed to occur as many as four times per hour (O'Connail & Frohlich, 1995) and this is quite possibly an underestimate.

Research relating to information keeping points to several conclusions: (1) keeping is difficult and error-prone; (2) "keeping right" has become more difficult as the diversity of information forms and supporting tools has increased; and (3) some costs of "keeping wrong" have gone away, but challenges remain.

Keeping Is Difficult and Error-Prone

Keeping actions, such as bookmarking a Web site or setting a reminder flag on an e-mail, are sometimes difficult both in the mechanics of execution and because these actions interrupt the current task (e.g., browsing the Web, reading e-mail). Even more difficult is the decision that guides these actions.

The keeping decision is multifaceted. Is the information useful? If so, do special steps need to be taken to keep it for later use? How should the information be kept? Where? On what device? In what form? Jones (2004) has characterized each keeping decision as a signal detection task[4] subject to a rational analysis of alternatives (Anderson, 1990).

There is a "gray area" where determination of costs, reciprocal benefits, and outcome likelihoods is not straightforward. In the logic of signal detection, this middle area presents us with a "damned if you do, damned if you don't" choice. If we keep the information, we may never use it. If we do not keep it, we may need it later. Moreover, if we keep information in the wrong way—in the wrong folder, for example—we may pay twice: We do not find the information when we need it and, worse yet, when we later need other information in the folder, the incorrectly filed information becomes an impediment to finding it.

Filing information items—whether paper documents, e-documents, or e-mail messages—into the right folders is a cognitively difficult and error-prone activity (Bälter, 2000; Kidd, 1994; Lansdale, 1988, 1991; Malone, 1983; Whittaker & Sidner, 1996). Difficulty arises in part because the definition or purpose of a folder is often unclear from the label (e.g., "stuff") and may change in significant ways over time (Kidd, 1994; Whittaker & Hirschberg, 2001; Whittaker & Sidner, 1996). Determining a folder's definition may be at least as problematic as determining a category's definition (e.g., Rosch, 1978; Rosch, Mervis,

Gray, Johnson, & Boyes-Braem, 1976; Wittgenstein, 1953; Zadeh, 1965). Worse, people may not even recall the folders they have created and so create new folders for the same, or similar, purposes (Whittaker & Sidner, 1996).

If a person's use of folders is sometimes inconsistent, such is also the case when it comes to the handling of incoming information. One's experience of the same information item can change considerably as a function of context (Martin, 1968; Tulving & Thomson, 1973). Kwasnik (1989) identified many dimensions that might influence the placement and organization of paper-based mail and documents in an office. In addition to attributes of the document itself (e.g., title, author), keeping behavior was influenced by 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, the classification of a document was heavily influenced by its intended use or purpose—a finding subsequently corroborated by Barreau (1995).

Jones, Bruce, and Dumais (2001, 2002) have observed that the choice of method for keeping Web information for later use was influenced by a range of considerations or functions. Marshall and Bly (2005) also noted that the reasons for keeping information vary and are not necessarily task-related or even consciously purposeful. Some participants in their study appeared to keep some information (e.g., newspaper clippings) for the pleasure of expanding their collection of like items (e.g., recipes) and a few used the term "packrat" to describe their keeping behavior (p. 117).

Sellen and Harper's (2002) work suggested that 3 percent of the paper documents in a typical office were misfiled and 8 percent were eventually lost. Perhaps the only surprise is that these percentages were not higher. Even when filing is done correctly, it is often not worth the trouble. Whittaker and Hirschberg (2001) have coined the phrase "premature filing" to describe a situation in which people go to the trouble to file information that turns out to have little or no value.

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

If filing is error-prone and costly and 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.'s (1998) study showed that users bookmarked only a portion of the Web pages they wanted to reaccess at a future date. A study of delayed, cued recall examined how people re-found Web information they considered useful (Bruce et al., 2004; Jones et al., 2003). Participants used one of three "do nothing" methods (i.e., ones requiring no keeping activity) in over two-thirds of the trials:

1. Searching again (using a Web-based search service).

2. Typing in the first few characters of the URL for a Web site and accepting one of the suggested completions of the Web browser.

3. Navigating to the Web site from another Web site. Overall, participants were very good at getting back to "useful" Web sites even when these were accessed only once or twice per year and had not been accessed for up to six months.[5]

Keeping "Right" Is Harder When Information Is More Fragmented

An act of keeping might be likened to throwing a ball into the air toward a point where one expects it to be at some future time. Keeping information in accordance with future need has never been easy (Bruce, 2005), but the current proliferation of information forms and supporting tools and gadgets makes keeping all the more difficult. The information people need may be at home when they are at work and 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 difficulty associated with its extraction outweighs the benefits of its use.

The information world that Malone (1983) described was largely paper-based. Today, paper documents and books are still an important part of the average person's PSI (Sellen & Harper, 2002; Whittaker & Hirschberg, 2001). However, people must also contend with the organization of e-documents, e-mail messages, Web pages (or references to these), as well as a number of additional forms of digital information (each with their own special-purpose tool support) including phone messages, digitized photographs, music, and videos. The number of keeping considerations increases further if a person has different e-mail accounts, uses different computers for home and work, or makes use of a PDA, smart phone, or some other special-purpose PIM tool(s).

People freely convert from one form of information to another (Jones et al., 2002). They make paper printouts of e-documents, Web pages, and e-mail messages and scan paper documents for inclusion in e-documents. They send e-documents and Web references via e-mail. They save e-mail

messages and Web pages into the same filing system that holds their e-documents.

People can keep information in several different ways in order to ensure that they have it later (Jones et al., 2002). Somebody may, for example, enter a client's telephone number into a calendar (as a reminder to call) and into a contact database. But doing so can increase the later challenges of updating and synchronization (e.g., when the telephone number changes). Moreover, such multiple registering of information may not cover all the contingencies when the information might be needed. Neither the calendar nor contact entry will help, for example, if the person needs to contact the client from his cell phone while stuck in traffic. We can hope that someday our information will be more integrated.

Some Costs of Keeping "Wrong" Have Gone Away, but Challenges Remain

Recent developments in technology have greatly reduced or even nullified some costs associated with mistakes made in the process of keeping. These reductions invite a consideration of two "decision-free" extremes in keeping strategy: that of keeping everything and that of keeping nothing at all (Jones, 2004). Unless one is engaged in video editing, the storage cost of a false positive—that is, of keeping digital information that is never used—is negligible. Why not keep it all? Facilities to sort, search, and filter may even help to clear away the clutter so that one can focus on the more useful information. Many people appear to be following a modified "keep everything" approach, for example, in the management of incoming e-mail by leaving it in the inbox, perhaps with occasional efforts to "spring clean" (Whittaker & Sidner, 1996).

Some costs associated with a "miss"—not keeping information that turns out to be useful—are also decreasing dramatically. With ever-increasing amounts of information available in readily searchable form on the Web (or intranet counterparts), people often rely on refinding methods that require no explicit keeping activity (Bruce et al., 2004). These "do nothing" methods include searching again or navigating from another Web site.

System support can also automate keeping in ways that combine local storage and reliance on the Web. The history and the "auto-complete" facilities in most Web browsers, for example, keep references locally to information that remains on the Web.

Approaches that automate keeping or that free individuals from the need to decide what is to be kept point to a dilemma identified by Lansdale (1988). People may not make the effort to keep information for later use either because doing so is too much trouble or because they are overly confident of their ability to retrieve the information at a later point in time (Koriat, 1993). Automated keeping can save people time and, more importantly, the distraction of leaving the current task in

order to decide whether and how an item in view should be kept for future uses. But if people do not take measures to keep the information that they have encountered, they may be less likely to remember to look for it at a later date when the need arises. The *generation effect* (Slamecka & Graf, 1978) has been observed in the assignment of names for text editing commands (Jones & Landauer, 1985) and in the assignment of tags to documents (Lansdale, 1991). Research in *prospective memory*—used to perform an action in the future—also supports a prediction that steps taken when information is encountered may reduce the likelihood of memory failure later on (Ellis & Kvavilashvili, 2000; O'Connail & Frohlich, 1995; Sellen et al., 1996; Terry, 1988).

An alternative to the "keep everything," "keep nothing," and "keep automatically" strategies is the "keep smarter" approach—making better decisions concerning future uses of current information (Jones, 2004). If a person has prepared a clear plan, for example, he is often more effective at keeping relevant information (including a recognition of its relevance) even when the plan and its goal are not the current focus of attention (Seifert & Patalano, 2001). One strategy is to apply technologies of information filtering to support the automation or partial automation of keeping decisions (e.g., Foltz & Dumais, 1992). E-mail applications, for example, commonly support the creation of special rule-based folders into which incoming messages can be copied or moved automatically. Establishing the rules, however, is not an easy task (Bälter, 2000). A step further in automation are tools that attempt to induce the rules for a folder based upon an analysis of its current members. Full automation of filing is problematic for two reasons: (1) rules, whether induced by the computer or created by people, are faulty; and (2) full automation reintroduces the dilemma already discussed—that without some involvement in the keeping activity, people may forget to look again later. One way to address both problems is for the computer to present a selection of likely folder destinations from which the person selects one or more (Segal & Kephart, 1999): He or she is then involved in the final decision and always has the option of selecting "none of the above."

A second approach is to tie acts of information access and creation (e.g., sending an e-mail message, making a new document, accessing a Web page) closely to the planning and completion of associated tasks and projects. The design of the Project Planner prototype (Jones, Munat, Bruce, & Foxley, 2005), for example, follows a guiding principle that information management and task/project management are two sides of the same coin. Moreover, with the right support, an integrative organization of information can emerge as a consequence of the efforts expended to plan a project and manage its tasks. Related to the advantage of a clear plan—at a higher level—is the potential keeping benefit of having an overall scheme of classification—a personal unifying taxonomy (Jones, 2004).

Summary: Keeping Is Multifaceted

Certainly keeping, like finding, can involve several steps. It may even trigger an act of finding—as in refinding the right folder or pile in which to place an information item. But the essential challenge of keeping stems from the multifaceted nature of the decisions about information needs. Is the information useful? Do special actions need to be taken to keep it for later use? Where? When? In what form? On what device? With no crystal ball to see into the future, answering these questions is a difficult and error-prone endeavor. But the attempt helps us to remember the information item subsequently. Some caution is advised against an overreliance on well-intended attempts to automate these decisions. Complementary tool support for planning may be one way to ensure that key connections are made between encountered information and expected need. And a well-formulated plan has other benefits as well.

The Meta-Level: Mapping between Need and Information

Meta-level activities, which constitute the third set of PIM activities, operate broadly upon collections of information within the PSI and on the mapping that connects need to information for these collections. At the level of keeping and finding, "managing" often equates with "getting by" (as in the sentence, "I finally managed to find the information"). The meta-level seeks to enhance personal control of one's PSI by stressing proactivity. How can people take charge of their PIM practice? How should the information be structured? According to what schema? Following which strategies? How can tools help either to structure or to obviate structuring? How is the effectiveness of current practice measured? Issues of privacy and security are also addressed at the meta-level (Karat et al., 2006). Who has access to what information under what circumstances? How can information (e.g., medical information, airplane seating preferences, a résumé) be distributed to best effect?

This section considers two meta-level activities that are (and should be even more) related to one another: (1) maintenance and organization, and (2) making sense of information and planning its use.

Maintaining (Too) Many Organizations

Differences between people are especially apparent in their approaches to the maintenance and organization of information. 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. Comparable differences have been observed in the ways people approach e-mail (Bälter, 1997; Gwizdka, 2002a; Mackay, 1988; Whittaker & Sidner, 1996), e-documents (Boardman & Sasse, 2004; Bruce et al., 2004), and Web bookmarks (Abrams et al., 1998; Boardman & Sasse, 2004).

Across information forms, differences in approaches to organization correlate with differences in keeping strategy. For example, people who

have a more elaborate folder organization—whether for paper documents, e-documents, e-mail messages, or bookmarks—tend to file sooner and more often. However, people are often selective in their maintenance of different organizations. Boardman and Sasse (2004), for example, classified 14 of 31 participants in their study as "pro-organizing" with respect to e-mail and e-documents but not with respect to bookmarks; only seven of the 31 participants took the trouble to organize their e-documents. (The study did not look at the organization of paper documents.)

The fragmentation of information by form poses special challenges for maintenance and organization. Folders with similar names and purposes may be created in different information organizations, especially for e-mail messages and e-documents (Boardman & Sasse, 2004). Maintaining consistency is difficult; for example, people may have a "trips" e-mail folder and a "travel" e-document folder. The fragmentation of information across forms also poses problems in the study of PIM (see the section on methods and methodologies). It is difficult and time-consuming to study and compare a participant's organizational schemes across several different forms of information and tempting to focus primarily on a single form of information such as e-mail messages or Web pages.

However, several studies have now examined how the same person manages across different forms of information (Boardman & Sasse, 2004; Jones, Phuwanartnurak, Gill, & Bruce, 2005; Ravasio, Schär, & Krueger, 2004). The following composite picture has emerged:

- People tend not to take time out of a busy day to assess their organizations or their PIM practice in general.

- People complain about the need to maintain many separate organizations of information and the fragmentation of information that results.

- 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 on the next.

- Several participants in one study (Jones et al., 2002) reported making special efforts to consolidate organizations, for example by saving Web references and e-mail messages into a file folder organization or by sending e-documents and Web references in e-mail messages.

The prefix "meta-" is commonly used to mean "beyond" or "about."[6] But the studies referenced here also invoke the original sense of "meta-" as "after."[7] For many people, meta-level activities such as maintenance and organization occur only after the more pressing activities of keeping and finding have been done. In many cases, this means not at all. Keeping and finding are triggered by many events in a typical day.

Information is encountered and keeping decisions are made (even if the decision is to do nothing). The information needed for a variety of routine activities (e.g., calling someone, planning the day's schedule, preparing for a meeting) triggers various finding activities.

Events triggering maintenance and organization of information are fewer and less frequent. For some people, these activities may be triggered by a corporate "clean desk" policy, a system administrator's message that an inbox is too full, or possibly a New Year's resolution to get organized. Studies of PIM themselves often serve as a trigger. For example, Boardman and Sasse (2004) reported that 12 participants in their study performed ad hoc tidying during the interview itself. In Jones, Phuwanartnurak, et al.'s (2005) study, all 14 participants made comments at the outset concerning a need to move or delete material that was outdated or no longer belonged in their files. Four participants actually insisted on interrupting the interview while they moved or deleted files or old folders.

As digital storage continues to increase in capacity and decrease in cost, maintenance and organization activities are seldom prompted by "disk full" events. People are freed from the need to delete or organize their digital information, and in many ways this is 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* (Jones, 2004). The potential uses or benefits of the item in focus (e.g., an old magazine) may be more salient than the ongoing cost of keeping (and never finding the time to read and use) the item. Similarly, Bergman, Beyth-Marom, and Nachmias (2003) refer to the *deletion paradox* to describe a situation where people may spend precious time on information items that are of little 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 people are no longer forced to delete anything, ever.

Even so, people often express unease about their current maintenance activities with apologetic comments or references to themselves as "a packrat" (Marshall & Bly, 2005, p. 117). Or, as one participant in Boardman and Sasse's study (2004, p. 585) said, "stuff goes in but doesn't come back out—it just builds up."

Making Sense of Information and the Value of External Representations

Much of the experimental work reviewed so far may make us question the value of organizing information in our PSI. We have too many folder organizations to maintain and we frequently postpone or ignore issues of maintenance just as we might avoid tidying a messy closet. Keeping (filing) information in a folder structure is difficult and mistakes are common. Storage is cheap. Search continues to improve. Is it worthwhile to organize information anymore? Or can we leave our information "flat"

and depend upon search (and possibly sorting) as a primary means of access?

We now review research demonstrating that people organize information not only to ensure its retrieval but for several other reasons as well. In a study conducted by Jones, Phuwanartnurak, et al. (2005, pp. 1506–1507), participants listed a number of reasons for using folders even if they had access to a perfect desktop search facility:

- "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."

In this study, a folder hierarchy developed for a project such as "wedding" often resembled a project plan or partial problem decomposition in which subfolders stood for project-related goals and also for the tasks and subprojects associated with the achievement of these goals. A "wedding dress" subfolder, for example, organized information and tasks associated with the goal of selecting and fitting a wedding dress (including, for example, a "wedding dress trials" sub-subfolder).

Barsalou (1983, 1985, 1991) has long argued that internal categories are used to accomplish goals. His research demonstrates people's ability to group together seemingly dissimilar items according to their applicability to a specific goal. For example, weight watchers might form a category "foods to eat on a diet." Rice cakes, carrot sticks, and sugar-free soda are all members of the category, even though they differ considerably in other ways. The best member is not necessarily like other category members. Instead, the best exemplar is the item that best accomplishes the goal or the ideal. Research by Markman and Ross (2003) suggests that an internal, goal-based organization for a set of items emerges as a by-product of the use of these items to accomplish goals. A person need not think explicitly about the goal-relatedness of items in order to internalize this organization.

This is not to suggest that a direct mapping exists between goal-directed folders as an external form of information organization and goal-directed categories as an internal organization of concepts. However, it is reasonable to suppose that folders (and piles, properties/value combinations, views, and so on) can form an important part of external representations (ERs), which, in turn, can complement and combine with internal representations (IRs) to form an integrated cognitive system (Hutchins, 1994; Kirsh, 2000).

Finding the right ER helps in *sense-making* (Dervin, 1992)—in efforts to make sense of information. For example, the right diagram can allow one to make inferences more quickly (Larkin & Simon, 1987). The way in which information is represented externally can produce huge differences in one's ability to use it in short-duration, problem-solving exercises (Kotovsky, Hayes, & Simon, 1985). Different kinds of representations, such as matrices and hierarchies, are useful in solving different types of problems (Cheng, 2002; Novick, 1990; Novick, Hurley, & Francis, 1999). Russell, Stefik, Pirolli, and Card (1993) have shown that ERs are acquired and discarded according to an assessment of relative costs and benefits.

What are the long-term costs and benefits associated with the use of ERs for PIM—the ER that results from use of a particular filing scheme, for example? And, can tools change the cost/benefit equation? What comes after the folder?

Efforts in tool support can benefit from basic research into how people plan. For example, support for progressive refinement (top-down or bottom-up) must also allow for the dynamic, flexible changes people make to accommodate new information or to exploit new opportunities. This opportunistic aspect of planning has been noted in experiments ranging from ill-structured domains, such as errand planning (Hayes-Roth & Hayes-Roth, 1979), to the highly structured Tower of Hanoi problem (Davies, 2003).

Summary: Meta-Level Activities Are Important but Easily Overlooked

Meta-level activities are critical to a successful PIM practice, but they are rarely urgent. Few events in a typical day direct our attention to meta-level activities such as maintenance and organization, making (overall) sense of an information collection, managing privacy and security, or measuring and assessing the effectiveness of strategies and supporting tools. As a result, meta-level activities can easily become afterthoughts. Research into meta-level activities and their support also appears to receive less attention than, for instance, research into finding (which can draw upon support from established communities in information seeking and information retrieval). But it is at the meta-level that we may realize some of the most productive synergies between applied research in PIM and basic research in cognitive science.

Methodologies of PIM Inquiry

The development of methodologies especially suited to PIM is still in its infancy. There is need for both methodologies in *descriptive* studies aimed at better understanding how people currently practice PIM and *prescriptive* evaluations to understand better the efficacy of proposed

PIM solutions (usually involving a tool but sometimes focused on a technique or strategy).

The descriptive and the prescriptive can form a complementary and iterative relationship with 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, Terveen, & Nardi, 2000).

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

1. *A person's practice of PIM is unique.* There is tremendous variation among people—even among those who have a great deal in common with each other with respect to profession, education, and computing platform—as demonstrated by many fieldwork studies (e.g., 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. PIM practice is uniquely tailored to the individual's needs and information. This uniqueness makes it very difficult to abstract tasks or extract datasets that can be used meaningfully in a laboratory setting.

2. *PIM happens broadly across many tools, applications and information forms.* People freely convert information from one form to another to suit their needs—e-mailing a document, for example, or printing a Web page. Studies and evaluations that focus on a specific form of information and supporting applications—e-mail, 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. The keeping and finding activities directed to a particular information item may be separated by days, weeks, or months. Basic PIM events of interest—such as filing, the 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, cannot be assessed without looking at later efforts to retrieve this information. People may initially embrace a solution but, over time, tire of its 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 time savings that it appears to offer. But these users may subsequently find that they have more trouble finding information with the tool than without it (perhaps because they attend less to the information initially when categorization is automated).

One approach 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 place an emphasis on context and situation have obvious relevance, including *contextual inquiry* (Beyer & Holtzblatt, 1998), *situated activity* (Suchman, 1983), and *situated design* (Greenbaum & Kyng, 1991). These and other methodologies have emerged from a participatory design movement that originated in Scandinavia (Schuler & Namioka, 1993). Participants in PIM studies might also be encouraged to practice participatory observation or, more simply, self-observation. People are often interested in talking about their PIM practices. Participants in longitudinal studies seem to derive therapeutic value from the opportunity to talk about their information management problems with a sympathetic observer.

But longitudinal case studies are time-consuming, and it is not easy to find a representative sample of participants able or willing to commit to a multi-session study. The results of case studies may be very enlightening but they 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 survey. The case study can help to identify the effects to focus on and the questions to ask in a single-session study or survey.

The effectiveness of PIM research can be improved through:

1. *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 participants as they work with their information.

2. *Tractable units of analysis.* One potential unit of analysis is the personal project (Jones, Munat, et al., 2005). The study of PIM emphasizes helping people manage their information over time in ways that cross the many boundaries set by current tools. This is a worthy, if somewhat daunting, ambition. How much personal information should we study? For how long? In what contexts? A personal project (e.g., planning a trip, taking a course, planning a

remodel) 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. Studying people's management of information as they work to complete a project may, therefore, provide practical ways to approach PIM other than tool-based analyses (e.g., the study of e-mail use alone or Web use alone).

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

Approaches to PIM Integration

As research has made clear, information fragmentation creates problems for keeping, finding, and meta-level activities such as maintenance and organization. The obvious antidote to fragmentation is integration (or unification). This section considers some approaches to integration.

Integration through E-Mail

The uses of e-mail now extend well beyond the sending of text messages between people separated from each other by time and distance. For example, e-mail is now used for task management, personal archiving, and contact management (Bellotti et al., 2003; Ducheneaut & Bellotti, 2001; Mackay, 1988; Whittaker & Sidner, 1996). Many of us practically "live" in e-mail 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 e-mail.) One approach to current problems of PIM—in particular, the fragmentation of information by application—is to accept the primacy of e-mail and build additional PIM functionality into an expanded e-mail application.

This approach is exemplified by *Taskmaster* (Bellotti et al., 2003), a prototype that deliberately builds task management features into an e-mail client application. Taskmaster introduces support for *thrasks* as a way to automatically connect task-related e-mail messages based upon an analysis of message content. The thrask is intended to be an improvement on threads. E-mail discussion within a thread can diverge widely from the original task even as other task-related e-mail messages are sent outside the context of a thread. On the other hand, 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 e-mail messages associated with their delivery. Attachments are no longer buried within the e-mail messages. This makes it easier for the user to see and access all information related to a task, regardless of its form.

E-mail 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 e-mail message. The design intent is that Taskmaster adds new task-related functionality without taking away the functionality already familiar to the user. Taskmaster provides several means of viewing thrask-related e-mail messages and also supports the assignment of task-relevant properties.

One potential limitation of the "integration through e-mail" approach has already been mentioned—people may want to spend *less*, not more, time in e-mail. Also, adding functionality for task management and other PIM-related activities may increase the complexity of an e-mail application that is already difficult to grasp for many users. Furthermore, 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, e-mail, and the 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 facilities such as Google Desktop.[8]

Fast, integrative, cross-form searches are supported in the Spotlight features of the Macintosh Operating System X (Mac OS X) (www.apple.com/macosx/features/spotlight). 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 (Spanbauer, 2005).

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 user interface for SIS supports the 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 the *Memory Landmarks* add-on through the inclusion of 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 views an e-mail 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 forgotten.

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 actively to 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."

First, 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, to "freeze" the document at key points in its composition, or, simply, because they need to use it on different projects and/or in different contexts. (Moreover, it's easier to copy than to reference.) People may also save external items into their PSI several times because they cannot 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 versions of information in e-mail. Airlines, for example, sometimes send several different e-ticket confirmations.

When multiple versions are returned, considerable time may be spent deciding which version is correct or which collection of items provides the necessary information. The problem of multiple versions intensifies when people modify or correct a document or save a new version of an item without tracking down and removing all the old versions. A chief executive officer 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 modify and use for an upcoming meeting with a customer.

The second reason to believe keeping and organizing will remain essential PIM activities is more speculative but also more fundamental: The acts of keeping an item and organizing a collection of items may be essential to our understanding of information and our memory of 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 fail to notice. If people do not make some initial effort to understand the information in their collection of items, they may forget to search for it subsequently. Folders, properties, and other constructs can be seen as an aid in understanding information. Even if a tool such as Implicit Query is successful at retrieving relevant information, people may fail to recognize this information or its relevance to a current need.

In a better world, we might hope to realize the advantages associated with the current use of folders and other means of ER without experiencing the disadvantages. The penalty currently associated with misfiling, for example, is too severe: We may, for all practical purposes, lose the misfiled information. 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 in order 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 decreases. Even if an item is misfiled, it can still be found again, using search if necessary. Moreover, regardless of folder location, search can be used to construct a useful set of results that can be quickly sorted by time and other useful properties.

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 expected future use and people are known to do this (Kwasnik, 1989). Rooms (Henderson & Card, 1986) represents an early attempt to integrate information items and other resources (e.g., tools, applications) with respect to a user's activity. For example, one 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 been discussed in the context of extensions to an e-mail application.

Another approach in tool support is the notion of a "project" as a basis for the integration of personal information. When a distinction is made between tasks and projects, it is typically with respect to length and complexity. In HCI studies of task management (Bellotti et al., 2004; Czerwinski et al., 2004), for example, a task is typically something we might put on a to-do list—e.g., "check e-mail," "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," and so on—but there is little utility in doing so. In these studies, therefore, the focus is on management *between* tasks, including handling interruptions, switching tasks, and resuming an interrupted task.

A project, by contrast, can last from several days to several years and is made up of any number of tasks and subprojects. Again, the informal to-do measure is useful, although 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.

In the UMEA (User-Monitoring Environment for Activities) prototype, Kaptelinin (2003) used the idea of a current project to bring together various forms of information—electronic documents, e-mail messages, Web references—and associated resources (applications, tools). One of UMEA's design goals was to minimize the user costs in setting up a project by automatically labeling items as they were accessed. Unfortunately, UMEA depended upon the user to signal a change in a current project. Because users frequently forgot to do this, items were frequently associated with the wrong project. Users could go back and

edit project/item associations to correct for mislabeling but they rarely took the time and trouble to do this. Kaptelinin sketched possible ways in which the system might detect a change in project but, to the author's knowledge, nothing along these lines has 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 label items as an incidental part of an activity that people might do in any case. When people plan projects, some of their planning finds external expression in, for example, to-do lists or outlines. The Project Planner prototype (Jones, Munat, et al., 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 that 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 with an outline. As headings are added, moved, renamed, or deleted, corresponding changes are made to the folder hierarchy. The Planner is simply another view into the file folder hierarchy and is, in fact, integrated into the file manager. But, as part of more general support for shortcuts, the folders of a project plan can be used to reference project-related e-mail messages and Web pages as well as files.

Behind the scenes, the Planner is able to support its more document-like outline view by distributing Extensible Markup Language (XML) fragments as hidden files, one per file folder, that 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 can handle other views as well. Efforts are currently underway, for example, to support a "mind map" view (Buzan & Buzan, 2004).

Integration through Properties

Dourish, Edwards, and their colleagues have argued that the folder hierarchy is limited, antiquated, and should be abandoned outright in favor of a property-based system of filing and retrieval such as that featured in their PRESTO/Placeless Documents prototype (Dourish, Edwards, LaMarca, Lamping, Petersen, Salisbury, et al., 2000; Dourish, Edwards, LaMarca, & Salisbury, 1999a, 1999b). Such proposals are not new. Ranganathan's (1965) colon, or faceted, classification scheme (Ranganathan, 1965) is essentially an organization of information by a set of properties in which an item's value assignment for one property can vary independently of its value assignment for another. Recipes, for example, might be organized by properties such as "preparation time," "season," and "region or style."

However, organization of information by properties depends upon an understanding of the information so organized. Meaningful, distinguishing, useful properties for special collections such as recipes may be readily apparent but this analysis is more difficult for newly acquired information. In particular, information relating to a project may be easier to organize into a hierarchy representing a plan or problem decomposition for the project.

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, and so on). Perhaps the 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 at points where a need for these items is anticipated. But it is with respect to the future that the LifeStreams timeline metaphor begins to falter. Some future events 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, we often have no clear notion of when we will need an item or have an opportunity to use it. In these cases, it may make more sense to organize items according to a need (goal, task, project). 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, rename, and delete files. The data within a file are typically in a "native format" and readable by only a single application—the word processor, spreadsheet, or presentation software 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 from one file to another ("owned" by another software application) via mechanisms such as "copy and paste" and "drag and 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 are copied, not referenced, and this can lead to many problems with updating.

As a result, data concerning a person we know—say, Jill Johnson—may appear in many, many places within our PSI. Because of this fragmentation, even simple operations, such as correcting for a spelling

mistake in Jill's name or updating her e-mail address, become nearly impossible to complete. We may update some of the copies but not all.[9] Also, we may experience the frustration of having some operations— name resolution, for example—available in one place (when sending e-mail) but not in another (when working with photographs). Underlying these issues is the 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, 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 they do not need to be copied at all (thus avoiding problems with updating information about Jill Johnson, for example). Support for these standards may make it possible to work with data and information packaged around concepts such as "Jill Johnson" rather than with files. Data for Jill would be primarily referenced, not copied. We could readily add more information about Jill or make a comment such as "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 e-mail addresses and telephone numbers for "true friends" we would like to invite to a birthday celebration.

These and other possibilities are explored in the Haystack project (Adar, Karger, & Stein, 1999; Huynh, Karger, & Quan, 2002; Karger, Bakshi, Huynh, Quan, & Sinha, 2005; 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 to information in units smaller and more meaningful 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 find Jill Johnson's birth date) and also to make additions or corrections to this information.

Haystack offers the potential to explore, group, and work with information in many ways that are not possible when it is "hidden" behind files. However, several 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, major changes in attitudes and practices will be required if application developers are eventually to abandon the control they currently have with data in native format in favor of a system where data come, instead, from an external source such as RDF.

Integration through a Digital Recording of "Everything"

If a sequence of information events is recorded—for example, those surrounding the viewing of a Web page—it should be possible to retrieve not only the Web page itself but also other items that were in close temporal proximity to it. We might hope, for example, to be able to access "the e-mail message I was looking at right before I looked at this Web page."

If enough events in our daily life were recorded, we might move significantly closer to a situation where virtually anything recalled about a desired item—the contexts of our interaction with the item as well as its content—could provide an access route back to the item. For example, we might direct the computer to "go back to the Web site that Mary showed me last week."

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

A bigger question is what to with all this data once they have been recorded. *MyLifeBits* (Gemmell et al., 2002; Gemmell, Lueder, & Bell, 2003) is an exploratory project aimed at addressing this by digitizing 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 (Czerwinski et al., 2006). For example, workshops on Continuous Archival and Retrieval of Personal Experience (CARPE) were 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, we might use it 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 that can only be partially addressed by technology alone.

Integration through Organizing Techniques and Strategies

Approaches to integration are predominantly tool-based and thus are generally inspired by developments in technology. But a degree of integration can also be accomplished through techniques and strategies that make use of existing tool support. As has been noted, people sometimes 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, for example; or everything is sent as an e-mail; or everything becomes a file.

Some people create a single organizing schema, which is then applied to different forms of information. This prompted Jones (2004) to speculate on the possible value of a *Personal Unifying Taxonomy (PUT)*. A person's PUT would be developed after a review, guided by a trained interviewer, of organizations for e-mail, 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 various information organizations.

However, a great deal of work will be required to establish a process and lay down principles of PUT development and to determine whether a PUT can be maintained over time to realize benefits that compensate for the 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. For example, many considerations that apply to library schemes of classification and their effective, consistent, sustainable use over time may have relevance to the development of a PUT. The larger point is that, in our fascination with the potential of new tools and technology, we should not overlook that of improving PIM through changes in our techniques, strategies, and habits.

Conclusions

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 that 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 have to remember to look, we have to know where to look, we have to recognize the information when we see it, and we often have to do these steps repeatedly to "re-collect" a set of items.

- *Keeping* activities move us from encountered information to expected future needs for which this information might be useful (or a determination that the information will not be needed). Reflecting the multifaceted nature of future needs, keeping activities are themselves multifaceted. We must make choices concerning location, organizing folder, form, and associated devices/applications.

- *Meta-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 and in the right form, and that it be 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 provide some reason to believe that we are moving in the right direction. There is clear interest in building a stronger community of PIM research to address the pervasive problem of information fragmentation in the practice of PIM.

Progress in PIM also depends upon overcoming related fragmentation in the conduct of PIM-related research. PIM, as an emerging field of inquiry, provides a productive point of integration for research that is currently scattered across a number of disciplines including information retrieval, database management, information and knowledge management, information science, human–computer interaction, cognitive psychology, and artificial intelligence. Ultimately, improvements in our ability to manage personal information should bring improvements not only to our personal productivity but also to our overall quality of life.

Acknowledgments

Endnotes

1. Thirty researchers from these disciplines and with a special interest in PIM met to discuss the challenges of, and promising approaches to, PIM at a special workshop (see the final workshop report at http://pim.ischool.washington.edu). Participants identified the potential of PIM to promote a synergistic dialogue between practitioners from various disciplines. Another sentiment expressed in several ways was that research problems relating to PIM often "fell through the cracks" between existing research and development efforts.

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

3. Certainly some events of finding and keeping involve no observable manipulation of information items and, therefore, fall outside the focus of PIM. A manager may see a recently hired employee, for example, and experience the need to retrieve his name. She may remember that the employee's name is "Ted" without reference to external information items. (But she might also find out the employee's name by referring to a paper printout that lists names of new employees.) Similarly, a salesperson with a facility for remembering telephone numbers might choose to commit the telephone number of a new client to memory. But if, instead, he writes the number on a piece of

paper, he has created an information item to be managed as part of his PSI.

4. The theory of signal detectability (TSD) (Peterson, Birdsall, & Fox, 1954; Van Meter & Middleton, 1954) has been applied elsewhere to a basic question of information retrieval: What does, and does not, get returned in response to a user query (see, for example, Swets, 1963, 1969)?

5. Given that the cue was effective in eliciting a memory for the Web site, success rates were between 90 and 100 percent (across different conditions of access frequency).

6. See, for example, the entry for "meta-" in the online encyclopedia Wikipedia (http://en.wikipedia.org/wiki/Meta-).

7. See, for example, the entry for "meta-" in the Merriam-Webster Online Dictionary (www.m-w.com/dictionary/Meta-).

8. For a more complete review of desktop search engines currently available for use, see Answers.com (www.answers.com) and then search, of course, for "desktop search."

9. But we might have good reasons not to 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 has already been published and is part of our archive.

References

Abrams, D., Baecker, R., & Chignell, M. (1998). Information archiving with bookmarks: Personal Web space construction and organization. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 41–48.

Adar, E., Karger, D., & Stein, L. A. (1999). Haystack: Per-user information environment. *Proceedings of the 8th Conference on Information and Knowledge Management*, 413–422.

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

Bälter, O. (1997). Strategies for organising email messages. *Proceedings of the Twelfth Conference of the British Computer Society Human–Computer Interaction Specialist Group*, 21–38.

Bälter, O. (2000). Keystroke level analysis of email message organization. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 105–112.

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

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

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition, 11*, 211–227.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 11*, 629–654.

Barsalou, L. W. (1991). Deriving categories to achieve goals. *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*, 407–424.

Bellotti, V., Dalal, B., Good, N., Flynn, P., Bobrow, D. G., & Ducheneaut, N. (2004). What a to-do: Studies of task management towards the design of a personal task list manager. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 735–742.

Bellotti, V., Ducheneaut, N., Howard, M., Neuwirth, C., & Smith, I. (2002). Innovation in extremis: Evolving an application for the critical work of email and information management. *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, 181–192.

Bellotti, V., Ducheneaut, N., Howard, M., & Smith, I. (2003). Taking email to task: The design and evaluation of a task management centered email tool. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 345–352.

Bellotti, V., Ducheneaut, N., Howard, M., Smith, I., & Grinter, R. (2005). Quality vs. quantity: Email-centric task management and its relationship with overload. *Human–Computer Interaction, 20*, 89–138.

Bellotti, V., & Smith, I. (2000). Informing the design of an information management system with iterative fieldwork. *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, 227–237.

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*, 872–878.

Berners-Lee, T. (1998). *Semantic Web roadmap: An attempt to give a high-level plan of the architecture of the Semantic Web.* Retrieved December 15, 2005, from www.w3.org/DesignIssues/Semantic.html

Beyer, H., & Holtzblatt, K. (1998). *Contextual design: Defining customer-centered systems.* San Francisco, CA: Morgan Kaufmann.

Boardman, R. (2004). *Improving tool support for personal information management.* Unpublished doctoral dissertation, 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. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 583–590.

Boardman, R., Spence, R., & Sasse, M. A. (2003, June). *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.

Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. (1969). Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior, 8*, 323–343.

Bruce, H. (2005). Personal, anticipated information need. *Information Research, 10*(3). Retrieved March 6, 2006, from http://informationr.net/ir/10-3/paper232.html

Bruce, H., Jones, W., & Dumais, S. (2004). Information behavior that keeps found things found. *Information Research, 10*(1). Retrieved March 6, 2006, from http://informationr.net/ir/10-1/paper207.html

Bush, V. (1945, July). As we may think. *Atlantic Monthly, 176*(1), 101–108.

Buzan, T., & Buzan, B. (2004). *The mind map book: How to use radiant thinking to maximize your brain's untapped potential.* London: BBC.

Byrne, M. D., John, B. E., Wehrle, N. S., & Crow, D. C. (1999). The tangled Web we wove: A taskonomy of WWW use. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 544–551.

Capurro, R., & Hjørland, B. (2003). The concept of information. *Annual Review of Information Science and Technology, 37*, 343–411.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human–computer interaction*. Hillsdale, NJ: Erlbaum.

Carroll, J. M. (1982). Creative names for personal files in an interactive computing environment. *International Journal of Man–Machine Studies, 16*, 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*, 97–104.

Catledge, L. D., & Pitkow, J. E. (1995). Characterizing browsing strategies in the World-Wide Web. *Third International World Wide Web Conference*, 1065–1073.

Cheng, P. C.-H. (2002). Electrifying diagrams for learning: Principles for complex representational systems. *Cognitive Science, 26*, 685–736.

Clarkson, B. P. (2002). *Life patterns: Structure from wearable sensors*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

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

Cutrell, E., Dumais, S., & Teevan, J. (2006). Searching to eliminate personal information management. *Communications of the ACM, 49*(1), 58–64.

Czerwinski, M., Gage, D., Gemmell, J., Marshall, C. C., Pérez-Quiñones, M., Skeels, M. M., et al. (2006). Digital memories in an era of ubiquitous computing and abundant storage. *Communications of the ACM, 49*(1), 44–50.

Czerwinski, M., Horvitz, E., & Wilhite, S. (2004). A diary study of task switching and interruptions. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 175–182.

Davies, S. P. (2003). Initial and concurrent planning in solutions to well-structured problems. *Quarterly Journal of Experimental Psychology, 56A*, 1147–1164.

Dempski, K. L. (1999). Augmented workspace: The world as your desktop. *Handheld and Ubiquitous Computing: First International Symposium*, 356–358.

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

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

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*, 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*, 133–161.

Dourish, P., Edwards, W. K., LaMarca, A., & Salisbury, M. (1999b). Using properties for uniform interaction in the Presto Document System. *Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology*, 55–64.

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

Durso, F. T., & Gronlund, S. (1999). Situation awareness. In F. T. Durso, R. Nickerson, R. W. Schvaneveldt, S. T. Dumais, D. S. Lindsay, & M. T. H. Chi (Eds.), *The Handbook of applied cognition* (pp. 284–314). Chichester, UK: Wiley.

Eisenberg, M., Lowe, C. A., & Spitzer, K. L. (2004). *Information literacy: Essential skills for the information age* (2nd ed.). Westport, CT: Libraries Unlimited.

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

Engelbart, D. C. (1963). A conceptual framework for the augmentation of man's intellect. In P. W. Howerton & D. C. Weeks (Eds.), *The augmentation of man's intellect by machine* (Vistas in Information Handling, vol. 1; pp. 1–29). Washington, DC: 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). Lifestreams: An alternative to the desktop metaphor. *Conference Companion on Human Factors in Computing Systems: Common Ground,* 410–411.

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). Retrieved March 6, 2006, from http://informationr.net/ir/10-1/paper210.html

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.

Foltz, P. W., & Dumais, S. T. (1992). Personalized information delivery: An analysis of information filtering methods. *Communications of the ACM, 35*(12), 51–60.

Fonseca, F. T., & Martin, J. E. (2004). Toward an alternative notion of information systems ontologies: Information engineering as a hermeneutic exercise. *Journal of the American Society for Information Science and Technology, 56,* 46–57.

Freeman, E., & Gelernter, D. (1996). Lifestreams: A storage model for personal data. *ACM SIGMOD Record, 25*(1), 80–86.

Garvin, D. (2000). *Learning in action: A guide to putting the learning organization to work.* Boston: Harvard Business School Press.

Gemmell, J., Bell, G., Lueder, R., Drucker, S., & Wong, C. (2002). Mylifebits: Fulfilling the memex vision. *Proceedings of the Tenth ACM International Conference on Multimedia,* 235–238.

Gemmell, J., Lueder, R., & Bell, G. (2003). The MyLifeBits lifetime store. *Proceedings of the 2003 ACM SIGMM Workshop on Experiential Telepresence,* 80–83.

Gershon, N. (1995, December). *Human–information interaction.* Paper presented at the Fourth International World Wide Web Conference, Boston, MA.

Gibson, J. J. (1977). The theory of affordances. In R. E. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 67–82). Hillsdale, NJ: Erlbaum.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

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

Golovchinsky, G. (1997a). Queries? Links? Is there a difference? *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 407–414.

Golovchinsky, G. (1997b). What the query told the link: The integration of hypertext and information retrieval. *Proceedings of the Eighth ACM Conference on Hypertext,* 67–74.

Greenbaum, J. M., & Kyng, M. (1991). *Design at work: Cooperative design of computer systems*. Hillsdale, NJ: Erlbaum.

Gwizdka, J. (2000). Timely reminders: A case study of temporal guidance in PIM and e-mail tools usage. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 163–164.

Gwizdka, J. (2002a). Reinventing the inbox: Supporting the management of pending tasks in email. *CHI '02 Extended Abstracts on Human Factors in Computing Systems*, 550–551.

Gwizdka, J. (2002b). TaskView: Design and evaluation of a task-based email interface. *Proceedings of the 2002 Conference of the Centre for Advanced Studies on Collaborative Research*, 4.

Hayes-Roth, B., & Hayes-Roth, F. (1979). A cognitive modeling of planning. *Cognitive Science, 3*, 275–310.

Henderson, A., & Card, S. (1986). Rooms: The use of multiple virtual workspaces to reduce space contention in a Windows-based graphical user interface. *ACM Transactions on Graphics, 5*, 211–243.

Herrmann, D., Brubaker, B., Yoder, C., Sheets, V., & Tio, A. (1999). Devices that remind. In F. T. Durso, R. Nickerson, R. W. Schvaneveldt, S. T. Dumais, D. S. Lindsay, & M. T. H. Chi (Eds.), *Handbook of applied cognition* (pp. 377–407). Chichester, UK: Wiley.

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. Retrieved February 24, 2006, from http://semanticweb2002.aifb.uni-karlsruhe.de/proceedings/Research/huynh.pdf

Johnson, J., Roberts, T. L., Verplank, W., Smith, D. C., Irby, C. H., Beard, M., et al. (1989). The Xerox Star: A retrospective. *Computer, 22*(9), 11–26, 28–29.

Johnson, S. B. (2005, January 30). Tool for thought. *New York Times*. Retrieved March 6, 2006, from www.nytimes.com/2005/01/30/books/review/30JOHNSON.html?ex=1264741200&en=c85978ec1eacfbe9&ei=5090&partner=rssuserland

Jones, W. (1986). 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. (2004). Finders, keepers? The present and future perfect in support of personal information management. *FirstMonday, 9*(3). Retrieved February 24, 2006, from www.firstmonday.dk/issues/issue9_3/jones/index.html

Jones, W., & Anderson, J. R. (1987). Short vs. long term memory retrieval: A comparison of the effects of information load and relatedness. *Journal of Experimental Psychology: General, 116*, 137–153.

Jones, W., Bruce, H., & Dumais, S. (2001). Keeping found things found on the Web. *Proceedings of the Tenth International Conference on Information and Knowledge Management*, 119–126.

Jones, W., Bruce, H., & Dumais, S. (2003). How do people get back to information on the Web? How can they do it better? *Proceedings of the 9th IFIP TC13 International Conference on Human–Computer Interaction (INTERACT 2003)*, 793–796.

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*, 42–63.

Jones, W., Dumais, S., & Bruce, H. (2002). Once found, what then? A study of "keeping" behaviors in the personal use of Web information. *Proceedings of the Annual Meeting of the American Society for Information Science and Technology*, 391–402.

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

Jones, W., Munat, C. F., Bruce, H., & Foxley, A. (2005). The Universal Labeler: Plan the project and let your information follow. *Proceedings of the Annual Meeting of the American Society for Information Science and Technology* [CD-ROM]. Silver Spring, MD: American Society for Information Science and Technology.

Jones, W., Phuwanartnurak, A. J., Gill, R., & Bruce, H. (2005). Don't take my folders away! Organizing personal information to get things done. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1505–1508.

Kaptelinin, V. (1996). Creating computer-based work environments: An empirical study of Macintosh users. *Proceedings of the 1996 ACM SIGCPR/SIGMIS Conference on Computer Personnel Research*, 360–366.

Kaptelinin, V. (2003). UMEA: Translating interaction histories into project contexts. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 353–360.

Karat, C. M., Brodie, C., & Karat, J. (2006). Usable privacy and security for personal information management. *Communications of the ACM, 49*(1), 56–57.

Karger, D. R., Bakshi, K., Huynh, D., Quan, D., & Sinha, V. (2005). Haystack: A customizable general-purpose information management tool for end users of semistructured data. *Proceedings of the Second Biennial Conference on Innovative Data Systems Research*. Retrieved March 6, 2006, from www-db.cs.wisc.edu/cidr/cidr2005/papers/P02.pdf

Karger, D. R., & Quan, D. (2004). Collections: Flexible, essential tools for information management. *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, 1159–1162.

Kidd, A. (1994). The marks are on the knowledge worker. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 186–191.

Kirsh, D. (2000). A few thoughts on cognitive overload. *Intellectica, 30*(1), 19–51. Retrieved March 6, 2006, from http://adrenaline.ucsd.edu/kirsch/articles/overload/cognitive_overload.pdf

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100*, 609–639.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology, 17*, 248–294.

Kwasnik, B. H. (1989). How a personal document's intended use or purpose affects its classification in an office. *Proceedings of the 12th Annual International SIGIR Conference on Research and Development in Information Retrieval*, 207–210.

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

Lansdale, M. (1991). Remembering about documents: Memory for appearance, format, and location. *Ergonomics, 34*, 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. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65–99.

Licklider, J. C. R. (1960). Man–computer symbiosis. *IRE Transactions on Human Factors in Electronics, HFE-1*, 4–11.

Licklider, J. C. R. (1965). *Libraries of the future*. Cambridge, MA: MIT Press.

Lucas, P. (2000). Pervasive information access and the rise of human–information interaction. *CHI '00 Extended Abstracts on Human Factors in Computing Systems*, 202.

Mackay, W. E. (1988). More than just a communication system: Diversity in the use of electronic mail. *Proceedings of the ACM Conference on Computer-Supported Cooperative Work*, 344–353.

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*, 99–112.

Mander, R., Salomon, G., & Wong, Y. Y. (1992). A "pile" metaphor for supporting casual organization of information. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 627–634.

Mann, S. (2004). Continuous lifelong capture of personal experience using Eyetap. *Proceedings of the 1st ACM Workshop on Continuous Archival and Retrieval of Personal Experiences*, 1–21.

Mann, S., & Niedzviecki, H. (2001). *Cyborg: Digital destiny and human possibility in the age of the wearable computer*. Toronto: Doubleday Canada.

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

Marchionini, G., & Komlodi, A. (1998). Design of interfaces for information seeking. *Annual Review of Information Science and Technology, 33*, 89–130.

Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin, 129*, 592–613.

Marshall, C. C., & Bly, S. (2005). Saving and using encountered information: Implications for electronic periodicals. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 111–120.

Martin, E. (1968). Stimulus meaningfulness and paired-associate transfer: An encoding variability hypothesis. *Psychological Review, 75*, 421–441.

Matthews, T., Czerwinski, M., Robertson, G., & Tan, D. (2006). Clipping lists and change borders: Improving multitasking efficiency with peripheral information design. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 989–998.

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

Nelson, T. H. (1982). *Literary machines*. Sausalito, CA: Mindful Press.

Norman, D. A. (1988). *The psychology of everyday things*. New York: Basic Books.

Norman, D. A. (1990). *The design of everyday things*. New York: Doubleday.

Norman, D. A. (1993). *Things that make us smart: Defending human attributes in the age of the machine*. Reading, MA: Addison-Wesley.

Novick, L. R. (1990). Representational transfer in problem solving. *Psychological Science, 1*, 128–132.

Novick, L. R., Hurley, S. M., & Francis, M. (1999). Evidence for abstract, schematic knowledge of three spatial diagram representations. *Memory & Cognition, 27*, 288–308.

O'Connail, B., & Frohlich, D. (1995). Timespace in the workplace: Dealing with interruptions. *SIGCHI Conference Companion on Human Factors in Computing Systems*, 262–263.

O'Day, V., & Jeffries, R. (1993). Orienteering in an information landscape: How information seekers get from here to there. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 438–445.

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. (in press). Cognitive models of human–information interaction. In F. T. Durso, R. S. Nickerson, R. W. Schvaneveldt, S. T. Dumais, D. S. Lindsay, & M. T. H. Chi (Eds.), *Handbook of applied cognition* (2nd ed.). West Sussex, UK: Wiley.

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

Pratt, W., Unruh, K., Civan, A., & Skeels, M. (2006). Personal health information management. *Communications of the ACM, 49*(1), 51–55.

Quan, D., Huynh, D., & Karger, D. R. (2003, October). *Haystack: A platform for authoring end user Semantic Web applications.* Paper presented at the 2nd International Semantic Web Conference (ISWC 2003), Sanibel Island, FL. Retrieved March 20, 2006, from http://haystack.lcs.mit.edu/papers/iswc-haystack.pdf

Ranganathan, S. R. (1965). *The colon classification* (Rutgers Series on Systems for the Intellectual Organization of Information, Vol. 4). New Brunswick, NJ: Graduate School of Library Service, Rutgers, the State University.

Ravasio, P., Schär, S. G., & Krueger, H. (2004). In pursuit of desktop evolution: User problems and practices with modern desktop systems. *ACM Transactions on Computer–Human Interaction, 11*, 156–180.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382–439.

Rouse, W. B., & Rouse, S. H. (1984). Human information seeking and design of information systems. *Information Processing & Management, 20*, 129–138.

Rowley, J. (1994). The controlled versus natural indexing languages debate revisited. *Journal of Information Science, 20*, 108–119.

Rundus, D. (1971). An analysis of rehearsal processes in free recall. *Journal of Experimental Psychology, 89*, 63–77.

Russell, D. M., Stefik, M. J., Pirolli, P., & Card, S. K. (1993). The cost structure of sensemaking. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 269–276.

Schuler, D., & Namioka, A. (Eds.). (1993). *Participatory design: Principles and practices.* Hillsdale, NJ: Erlbaum.

Segal, R. B., & Kephart, J. O. (1999). MailCat: An intelligent assistant for organizing e-mail. *Proceedings of the Third Annual Conference on Autonomous Agents*, 276–282.

Seifert, C. M., & Patalano, A. L. (2001). Opportunism in memory: Preparing for chance encounters. *Current Directions in Psychological Science, 10*, 198–201.

Selamat, M. H., & Choudrie, J. (2004). The diffusion of tacit knowledge and its implications on information systems: The role of meta-abilities. *Journal of Knowledge Management, 8*, 128–139.

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*, 483–507.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*, 379–423, 623–656.

Simon, H. A. (1971). Designing organizations for an information-rich world. In M. Greenberger (Ed.), *Computers, communications and the public interest* (pp. 40–41). Baltimore: Johns Hopkins University Press.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology, 4*, 592–604.

Spanbauer, S. (2005, August). Longhorn preview: The newest versions of the next Windows add graphics sizzle and more search features but lack visible productivity enhancements. *PC World, 23*(8), 20–22.

Streitz, N., & Nixon, P. (2005). The disappearing computer: Introduction. *Communications of the ACM, 48*(3), 32–35.

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

Suchman, L. (1987). *Plans and situated actions: The problem of human–machine communication*. Cambridge, UK: Cambridge University Press.

Swets, J. A. (1963). Information retrieval systems. *Science, 141*, 245–250.

Swets, J. A. (1969). Effectiveness of information retrieval methods. *American Documentation, 20*, 72–89.

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*, 97–137.

Tauscher, L. M., & Greenberg, S. (1997b). Revisitation patterns in World Wide Web navigation. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 399–406.

Taylor, A. G. (2004). *The organization of information* (2nd ed.). Westport, CT: Libraries Unlimited.

Teevan, J. (2003). "Where'd it go?": Re-finding information in the changing Web. *Proceedings of the MIT LCS/AI Student Oxygen Workshop*. Retrieved March 20, 2006, from http://sow.csail.mit.edu/2003/proceedings/Teevan.pdf

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. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 415–422.

Teevan, J., Dumais, S. T., & Horvitz, E. (2005). Personalizing search via automated analysis of interests and activities. *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 449–456.

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

Thompson, L. L., Levine, J. M., & Messick, D. M. (Eds.). (1999). *Shared cognition in organizations: The management of knowledge*. Mahwah, NJ: Erlbaum.

Tulving, E. (1983). *Elements of episodic memory*. Oxford, UK: Oxford University Press.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 359–380.

Van Meter, D., & Middleton, D. (1954). Modern statistical approaches to reception in communication theory. *Institute of Radio Engineers Transactions, PGIT-4*, 119–145.

Whittaker, S. (2005). Collaborative task management in email. *Human–Computer Interaction, 20*, 49–88.

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

Whittaker, S., & Sidner, C. (1996). Email overload: Exploring personal information management of email. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 276–283.

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.

Williamson, A. G., & Bronte-Stewart, M. (1996). Moneypenny: Things to do on the desk. *Adjunct Proceedings of the 11th British Computer Society Annual Conference on Human–Computer Interaction*, 197–200.

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

Wilson, T. (2000). Human information behavior. *Informing Science, 3*(2), 49–55.

Wittgenstein, L. (1953). *Philosophical investigations* (G. E. M. Anscombe, Trans.). New York: Macmillan.

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.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*, 338–353.