Personal Information Management Strategies and Tactics used by Senior Engineers

Christina K. Pikas
University of Maryland College of Information Studies and The Johns Hopkins University Applied Physics Laboratory; 11100 Johns Hopkins Road; Laurel, MD 20723-6099; Tel: 240-228-4812; cpikas {at} umd.edu

This paper reports the results of an exploratory qualitative study of how senior engineers in a research laboratory environment do personal information management (PIM). Responsive, semi-structured interviews were conducted with four senior engineers. Thirteen themes in four groupings emerged. The four groupings are: organization and retrieval, un-organized aspects, information keeping and preservation, and use of specialized tools. Themes not seen in other studies are: writing to remember and reporting to retrieve, and personal handbooks. The themes are described in detail. Implications for the design of information systems and future work are discussed.

Introduction

The purpose of this study is to explore how senior engineers manage their personal, work-related information. Much research has been done on the information behavior of engineers; however, this research has typically centered on the information seeking and scholarly communication processes related to external information. Many studies mention "personal information stores" and technical files, but frequently do not describe how these files are created, maintained, organized, and used. In contrast, this study investigates how people manage their personal information, a topic of increasing interest. Specifically, the study investigated how engineers with more than ten years post-graduate experience and who work in an applied research and development laboratory setting find, keep, use, organize, re-find, and share their research-related data, literature, working materials, reference materials, and electronic files.

Engineers are knowledge workers who receive, generate, manipulate, and otherwise deal with data and information as a function of their work. The data and information are at three levels: external, organizational, and internal or self-generated. External information can come from the scholarly community in the form of journals, newsletters, conferences, textbooks, handbooks, and so forth, as well as from trade publications, vendors, and their customers. Organizational information can come from their workplace's administration, their co-workers, and their project teams. Internal or self-generated information includes working papers, project reports, data collections, and peer reviewed literature. Successful engineers surely must have developed coping strategies to deal with being inundated with information. Further, one would expect engineers who have been on the job for a long time to have developed some organization schemes and methodology for successful information retrieval from their personal data and information collections.

The goal of this exploratory research, then, is to achieve a rich understanding of some of the methods successful senior engineers have developed. This understanding will assist in developing or deploying systems to support these methods and in educating junior engineers on ways they can improve their personal information management (PIM).

Literature Review

External information seeking and use of engineers has been well studied. Tenopir and King (2004) published an extensive review of more than 30 years' work on how engineers seek, use, and communicate information and many others continue to contribute research on this subject. Conversely, PIM has not been well studied until recently. There is a scattering of research over the past 40 years, but it has only been in the past five years that there have been special issues of magazines[1], workshops[2], and sessions[3] at professional conferences on the topic of PIM. I will highlight some of the findings on the information behavior of engineers

[1] *Communications of the ACM 49* (January 2006).
[2] See, for example, PIM05: NSF-sponsored workshop on Personal Information Management (PIM) held in Seattle, WA January 27-29, 2005 and Personal Information Management, SIGIR 2006 Workshop, held in Seattle, WA, August 10-11, 2006.
[3] See, for example, sessions at ASIS&T Annual Meeting 2004, 2005, 2006.

that are salient to PIM, then review some of the findings from the field of PIM that are most salient to the study of engineers in a research environment.

The Information Behavior of Engineers

"Scientists produce knowledge (facts) and engineers produce designs, products, and processes (artifacts)" (Kennedy, Pinelli, Barclay & Bishop, 1997, p.182), so they use information differently. The information use environment of engineers has the following characteristics:

- information is encoded in the systems they build
- the scholarly literature is less likely to be cumulative
- gains can be made without referencing the literature
- personal contacts are extremely important information sources
- information is selected for convenience as much as relevance (Taylor, 1991).

Ellis and Haugan (1997) discuss three different types of projects industrial engineers undertake:

- incremental
- radical
- fundamental,

as well as three different project phases:

- evaluation of alternative solutions
- development and testing
- summary of experiences.

Each of the types and phases of projects is associated with different information behavior.

From these general statements about the information behavior of engineers, three major themes relevant to PIM emerge.

First, engineers use different types of resources and communicate their results through different channels than researchers in other disciplines. Engineers need specific information on products and material properties as found in

- product catalogs
- specification sheets
- handbooks
- textbooks
- technical reports.

They also use the scholarly journal articles and monographs that scientists use most heavily (Cool & Xie, 2000; Hirsh & Kraft, 2000; Kwasitsu, 2003; Taylor, 1991; Tenopir & King, 2004). Engineers typically rely more heavily on internal organizational information sources than do scientists because of the proprietary nature of their work and specificity of the information needs (Cool & Xie, 2000; Hertzum & Pejtersen, 2000; Leckie, Pettigrew, & Sylvain, 1996; Shuchman, 1981). Engineers communicate their results through patents, sponsor presentations, technical reports, conference papers, and, to a certain extent, journal articles (Tenopir & King, 2004).

Second, engineers select information resources based on convenience, experience, and accessibility. Anderson, Glassman, McAfee, and Pinelli (2001) review literature from 1949 through 1997 showing that engineers, in particular, select information sources based less on technical quality and more on convenience, familiarity, and ease of use. The authors operationalized ease of use as reliance on

- local sources over external sources,
- oral communication channels over written,
- direct information channels over indirect
- use of personal information stores over other information sources.

They found that aerospace engineers and scientists rely most heavily on their personal stores of information, consult co-workers within the organization, and then consult external colleagues. On the other hand, Fidel and Green (2004) provided a richer description of how engineers understand accessibility and how they select information resources. They found that engineers paid a great deal of attention to quality as well as issues of accessibility. Both studies and Kwasitsu (2003) found that engineers seem to maintain and frequently consult personal stores of information.

Third, engineers gain information from people as well as from published sources. As mentioned above, the preference for information seeking from people or through “oral channels” was a factor of accessibility in the study by Anderson et al. (2001) and was a finding in several other studies (e.g., Chakrabarti, Feineman, & Fuentevilla, 1983; Cool & Xie, 2000; Ellis & Haugan, 1997; Kwasitsu, 2003; Von Seggern & Jourdain, 1996). Hertzum and Pejtersen (2000) studied this phenomenon in more detail through two case studies of corporate engineers. The engineers provided multiple reasons for relying on oral channels: shared context and understanding of research problem, information not available in written form, and being unsure of what information will solve the problem (Hertzum & Pejtersen, 2000).

Personal Information Management

PIM has been studied for many years, but primarily in light of automating the activities or designing information systems to better support PIM. Jones (2007) recently published a review of the work from psychology, information science, and computer science. This section provides a definition and examples of personal information; reviews activities required to manage personal information, and describes organization and retrieval schemes for personal information.

First, personal information is defined broadly in PIM. Jones (2007) defines personal information as “information people keep for their own personal use.” He provides examples such as e-mail, appointments, web pages, books, articles, and documents. To these examples, Kaye et al. (2006) add letters, scientific specimens, and blackboards full of equations.

Second, PIM activities are categorized as acquisition, organization, maintenance, and retrieval (Jones, 2007). Acquisition of information may be through

- purposeful information seeking and subsequent decisions to keep information (Bruce, Jones, & Dumais, 2004; Bruce, 2005; Whittaker & Hirschberg, 2001)
- serendipitous information encountering (Erdelez, 1999), or
- unsolicited information in mail, e-mail, or in person.

Maintenance is “modifying, moving, or deleting documents from the PIM system” (Barreau, 1995, p.334). Malone (1983) grouped people into pilers and filers. Other studies have found elaborate organization schemes for print and electronic materials (Kaye et al., 2006). Improper maintenance can lead to binge organizing (Jones, 2007) or “premature filing” (Whittaker & Hirschberg, 2001). Retrieval from this collection can be by browsing, shuffling through papers, or even by desktop search (Jones, 2007).

Third, organization and retrieval schemes for personal information vary from other information collections. Malone (1983) found that some participants arranged piles as reminders. Classification of documents and subsequent filing is a problem for many as documents are classified under temporal, subject, importance, required action, potential use, and document attribute dimensions (Barreau, 1995; Jones, 2007; Kwasnik, 1991; Lansdale, 1988; Malone, 1983). Retrieval of this information is complex even though the searcher has seen and kept the documents. Memory and association of the document within its creation context are important and chronological clues can help (Malone, 1983).

Methods

Qualitative research was selected to provide rich, process-related information. In-depth, semi-structured, responsive interviews are well-suited to explore each participant’s PIM in detail yet guide each of the interviews so that all of the major points of interest are discussed. Responsive interviews have the following characteristics (Rubin & Rubin, 2005)

- the goal is to obtain the participant’s interpretations of their experiences
- the personality and style of the interviewer matter
- the interview is an exchange
- the questions are modified to explore what is said
- the design is flexible and adaptive.

The interview guide consisted of 11 broad questions regarding the nature of the participant’s work and information requirements, how the participant deals with data and information, and how electronic forms of communication and information have changed the participant’s PIM. The final interview guide is reproduced in the appendix.

Participant Selection

Potential participants were selected through searches of published papers and patents. The four participants were purposefully selected to represent variety in academic background and current work projects. All are professional engineers at the Johns Hopkins University Applied Physics Laboratory (JHU/APL) in Laurel, Maryland. JHU/APL is a university-based applied research and development laboratory that focuses on national security and has major efforts in space science and technology. Of the technical professional staff, 47% have degrees in engineering and 19% have doctorate degrees. JHU/APL projects run from basic research to full-scale operational testing, through all phases of project life cycles (http://www.jhuapl.edu/newscenter/presskit/presskit.asp).

Three of the four participants hold doctorate degrees in engineering, two in electrical and one in aerospace. The fourth participant holds a Masters of Science in Engineering Physics. One participant has been at JHU/APL just over ten years while the remainder have each been at the lab about thirty years. The participant with only a little more than ten years at the lab previously worked in a similar position in a peer organization.

Data Collection

The interviews took place over the Fall of 2006. The participants were invited through e-mail. The interviews were held in a library office because audiotaping was not allowed in participants' offices for security reasons. They lasted one to two hours and were audio taped and transcribed. Field notes were made on a tour of one participant's office and during a patent presentation by another participant.

Data Analysis

The transcripts and field notes were read and re-read and the cross-case constant comparison method was used to develop codes, integrate categories, and define themes (Dye, Schatz, Rosenberg, & Coleman, 2000). The various quotes, variations, and counter examples providing evidence for each code were reviewed, weighed, compared, and contrasted to synthesize summary information for each theme. These summaries were checked against the transcripts to ensure that they conveyed the participants' meaning. The participants were provided with copies of the results to complete a member check to support internal validity.

Strengths and Limitations

This work is one of few dealing with how non-academic researchers outside of computer or information science conduct PIM. Although the sample is small, the participants gave fully of their experience and the data allowed for some strong themes to be developed.

A major limitation of this study was the requirement that the audio recording take place away from the participants' offices. Interviewing in a naturalistic setting is very important for building rapport, establishing common ground, and achieving understanding in interviews (Rubin & Rubin, 2005). Future work should include additional visits to the engineers' primary work spaces, even if these visits cannot be mechanically recorded.

Results

Thirteen themes emerged which can be grouped as shown in table 1. The remainder of this section will provide detailed explanations and supportive evidence for each theme.

Theme	**Subtheme**	**Description**
1. Organization and retrieval	1.1 Memory	Memory as a key factor in successfully managing all types of personal information (Malone, 1983).
	1.2 Journaling over time	Narratives to document procedures, meetings, phone calls, and activities over time (Shankar, in press).
	1.3 Commenting software for retrieval and reuse	Annotations in analysis and modeling and simulation programs.
	1.4 Writing to remember and reporting to retrieve	Formal scholarly communication for memory and archival retention.

	1.5 Specific file organization: Personal Handbooks	Creation of handbooks from selection of other materials.
	1.6 Specific file organization: Technology and project files	Folders are arranged topically and by work project.
	1.7 Seeking information from people: Maintaining a collection of experts	Use and maintenance of networks for information seeking (Allen, 1977; Hertzum & Pejtersen, 2000).
2. Un-organized aspects	2.1 Organized messes	Mismatches between descriptions of organization schemes and messy desks.
	2.2 E-mail, the overlooked collection	E-mail collections are not fully exploited for PIM purposes.
	2.3 No bibliographic reference databases	
3. Information keeping and preservation	3.1 The nitty gritty is lost	Decisions not to retain working documents.
	3.2 The engineer provides the best preservation	Personal responsibility for data retention and preservation
4. Use of specialized tools	4.1 Software, whatever works	Use of office productivity software in place of specialized high-end engineering and analysis software.

Strategies not seen in other studies are bolded.

Table 1: Themes and Subthemes

1. Organization and Retrieval

The information collections consisted of data in three different stages: raw data exported from experimental equipment, processed data in some stage of analysis, and analyzed data published in reports. Information in the collections was in many different formats including:

- books (textbooks, handbooks, manuals)
- print documents in binders, notebooks, and loose (articles, work in progress, correspondence, photocopies)
- electronic files (e-mail, articles, work in progress, reports, address books).

Within the collections, there were three competing organization schemes. Some folders, binders, and piles were arranged by topic, whereas others were arranged by time or project.

The participants reported that they were generally successful in finding and retrieving information from their personal collections.

1.1 Memory

All of the participants emphasized the importance of their memory for information retrieval. One participant stated, “I firmly believe that to be successful at any job let alone a technical job, you have to have a very good memory” (C[4]; 39:1[5]).

Some of the participants save time by not organizing some parts of their collections and instead rely on memory: “I don’t have [project folders] as finely organized as I do the data folders that I maintain … memory helps a lot” (D; 17:8).

Sometimes the participants associate information with spatial and temporal features such as office location, project, conference, time of year, season, and so forth.

[4] Denotes participant. Letters are randomly assigned.
[5] Indicates page and starting line number from the interview transcript.

> I have a strange weird associative memory but I can tell when things occurred in time... The key element is when it was done. It seems to be that if I can locate it in my mind with the time, then we can find it. (A; 11:15)

This agrees with earlier findings in PIM from Malone (1983) and as summarized by Jones (2007).

1.2 Journaling Over Time

The engineers maintain detailed diaries or lab notebooks over time. These chronological collections contain data, observations, equipment printouts, information from the literature, as well as contact information and meeting notes. The notebooks may even list file names to associate print notes with the digital data and literature. One participant reported that "I do have notebooks which are sort of like a diary if you will… typically they'll last over a couple of years. So I use them not only for when I make measurements but also when going to meetings or something like that so everything is in the notebook" (D; 17:20)

The engineers use their memory to associate projects with time, and then use their journals to retrieve information related to the project. A participant described his method in detail:

> I started keeping a technical diary. I got a spiral bound book… I couldn't keep track of all of the phone calls I was making, who I was talking to, and who was calling me and I got this book out and I took the concept of the laboratory manual and now had a technical diary ... important things would be on the front cover …Those became collection points and then I had a calendar or a time sequence… I keep a record of everything I do. Print something out, cut it off, trim it, paste it in the book (C; 35:13).

In addition to the cross-project on-going journal, one participant maintains an open document in MathCad software throughout his literature searching, reading, data exploration, analysis, and reporting. He captures everything in one place while he is working and he can copy the plots and graphs into his final word processing document directly from that program. He also prints out a copy of his work and maintains it in a project binder.

> MathCad allows me to document. Put the equations in… and I can also bring in experimental data and plot that against the models and I find that very convenient and interactive environment so I use that a lot… I bring in like someone else's data, and I'll put in the reference to the paper and all the details into the MathCad document so it's there. So when I go to write the paper, I have the references there and I just go and cut and paste. (D; 8:14)

These journals are quite different from the pre-formatted templates Shankar (in press) found when studying the laboratory notebook practices of biologists. This is probably explained by the disciplinary and work setting differences between academic life sciences researchers and industrial engineers. In a presentation on patenting, one participant mentioned keeping a separate notebook for inventions. This invention notebook and the written observations kept with the data are similar to the laboratory notebooks mentioned elsewhere in industrial settings. The association of files with time is a well known feature of PIM (Bovey, 1996; Landsdale, 1988).

1.3 Commenting software for retrieval and reuse

Although none of these engineers are computer software engineers, they all work with computers for analysis and modeling and simulation. Two of the participants who work on project teams discussed annotating software for retrieval and reuse by labeling modules or components, commenting the program, or by attaching a "read me" file.

> Even in Simulink itself on these diagrams that you see when you're coding it there are comment boxes you can put in and you can give the boxes little names … or [describe] whatever it does. You can, if you want, keep version numbers and dates of changes in there as well. (B; 7:12)

Unfortunately, both participants mentioned frustration that co-workers seldom comment as much as is necessary or in the same manner.

> So in our group the theory is that the documentation for our software is these Simulink diagrams. You know we're supposed to put good comments on there…If the next [project] down the line needs it then they go to take it from you and you have no explanation in there what it's doing then they spend a lot of time looking at it and trying to make sure it's right. (B; 7:16)

I've had to force my EE who's writing this code to document it and put it into his report the code he used to unpack this data. I've tried to get him to take this code, put the MATLAB name – the name.mat format – put that on a disk, and then you have a read me file that's how to use the unpacking [program]. (C; 23:12)

This theme mirrors the practices generally reported for full time software engineers (Lethbridge, Singer, & Forward, 2003).

1.4 Writing To Remember and Reporting To Retrieve

Engineers described preparing internal technical reports and conference papers as a method for storing, saving, and providing for future information retrieval. Technical reports are often considered only as deliverables for sponsor work, but these engineers find that documenting their work in progress as interesting findings occur allows for archival storage of their information. They appreciate that technical reports and conference papers are not always easy to locate, but the reports are at least stored in common collections so that the engineer can re-find the information.

Well typically, the best way to do that is you keep raw data…[y]ou keep it on disk or you print it out or whatever. But, you also write about it. You document it. You write a report, you write a paper or you write something and *that* becomes your archival reference… (A; 9:16)

Archivists, records managers, and intellectual property specialists routinely consider laboratory notebooks or raw data to be the archival record (Shankar, in press). This differs dramatically from the participant's consideration of only the processed information as archival. A comparison of this to the theme "the nitty gritty is lost" and the use of separate invention lab notebooks in "journaling over time" shows a very different understanding of archival records between records management practices and the senior engineers.

1.5 Specific File Organization: Personal Handbooks

A standard resource in engineering libraries is the handbook. Typically, handbooks provide very general overviews of a discipline, sometimes as broad as "chemistry and physics", relevant mathematical tools, material property data, and other reference data. Two of these engineers reported crafting their own "personal handbooks" within their collections. These personal handbooks contain similar things to the published handbooks, but they combine items that uniquely address the engineer's work from several different disciplines. One of the engineers populated his handbook with his own measurements over time which became reference data for him. Another engineer described the contents and use of his personal handbook:

On the spine it says "useful information." That's been collected over the years. I've got oceanographic tables, I've got figures from satellite images, I've got wire gauge tables, I've got conductivity of metals, I've got reactivity of alkalis, whatever I'm working on gets put in to this book. It's usually a Xerox of a page that goes into a plastic holder … instead of pulling out my CRC handbook and finding the section in a 300 page book, I'll Xerox the 4 pages and put them in a book. Then when I want to do a circuit analysis, I just pull that book out, and there I've got my LaPlace transforms. (C; 37:11)

1.6 Specific File Organization: Technology Files and Project Files

The engineers generally maintain two types of files to organize their personal collections of literature, papers, presentations, and other completed work. First, they have project files that contain communications from the sponsor, Power Point presentations, and other deliverables. These files have start and end dates and are comprehensive of the project. For scientific and technical information, however, the engineers maintain ongoing folders with everything from fundamental and seminal papers to new, cutting edge conference papers. These files can be electronic or print and can be subdivided into cutting edge and historical.

I have technical files, which I file by topic… You know the key papers that have evolved, and people recognize and have recognized for years and cite a lot that cover the fundamentals and then I have files that contain interesting new papers. They're interesting to me based on what I know about [my specialty]… Hard copies are in a file, but, you know, as you write and as you build on a topic, you'll use that reference again. (A; 20:15)

In a variation of this method, one engineer who has worked on a single large project for years marks the folders by project and then by the topic, "I just have folders and I just put the name of the [project] and I have to have more than one folder then I'll put this was the […] determination or this was the […] model – whatever it was, I try to put on the tab of the folder." (B; 23:3)

These file types are similar to those found by Barreau (1995) and Kwasnik (1991).

1.7 Seeking Information from People: Maintaining a collection of experts

The participants mentioned relying on networks of expert colleagues as information sources. They build and maintain these networks over their careers as a personal information collection just as they build technical files and personal handbooks. One participant explained, “you can go to the library and do a literature search and also talk to experts and see what they think are key papers or references that they’ve used and just try to pull all of this stuff together” (D; 3:18).

Sometimes the networks consist of colleagues within the organization. Orientation programs for professional staff are the first opportunity the participants had to build a network within the organization. The seeds planted at orientation are then nurtured throughout the participant’s tenure. For other participants, depending on the size of their field or uniqueness of their work, the networks extend through professional societies and other organizations geographically proximate or topically similar. These networks are built and nurtured through professional conferences and meetings as well as the occasional cold call.

> Over the years I try to maintain and develop contacts other places… So you get to know people, find out what they’re expert in and certainly if it’s relevant to a certain class of problems that we’re interested in, then I try to develop a dialog so that I can ask them questions that they might know about or sometimes to share experimental data …It’s very helpful to get information from people who know all of the critical work that’s been done, to tap into that expertise that they have. (D; 12:7)

Engineers’ information seeking from other people is well-known in the literature (see The Information Behavior of Engineers above); however, a variation found in this study is the consideration of the network of contacts as a collection or resource to be built and cultivated. The development and maintenance of this collection is purposeful and calculated. The participants consider their networks to be valuable assets that they bring to new projects. Also, instead of just consulting physically proximate colleagues (Allen, 1977; Kraut, Egido, & Galegher, 1990), the engineers actively seek cross-lab consultations and external colleagues.

2. Un-organized Aspects

Along with all of the organization methods and tools, there are un-organized aspects of the participants’ PIM.

2.1 Organized Messes

Each of the engineers, at my prompting, described careful organization schemes for all of their personal information. Later in the interviews, however, each revealed “messes” at their desks. In fact, one engineer stated, “you go into a person’s office and you know if the desk is cluttered, the person is a worker. If the desk is pristine… well then… they’re not as productive” (C; 34:9). It is unclear from these conversations if these messes are failures of the organization system, items that have not been filed, or simply another organization scheme for a category of information not described here. After all, even though it may seem like a mess, the engineers are successful at retrieving the information: “you’ll never come in and find what you would consider a neat desk in my office but I always can tell you where something is” (B; 30:14).

Instead of Malone’s (1983) artificial separation of pilers and filers, these engineers appear to do both. Visits to the participants’ offices showed many books, binders, and files as well as open binders, notebooks, and papers in piles.

2.2 E-Mail, the Overlooked Collection

The engineers describe the importance of e-mail for project communication and how it has replaced memos in many cases. E-mail is also the channel for delivering data, analysis, and reports. Two of the engineers struggle with overloaded inboxes while the other two perform quick triage, save down what is important, and delete the ephemeral or unimportant. The engineers who reported full inboxes do not seem to view their e-mail collections as personal information resources to be searched and used. They know that there is valuable information there and are reluctant to delete project-related information, but they rarely seem to really tap this collection.

> A lot of the work that gets done now comes around through e-mail …I tried to get it organized because it’s a really good trail. There have been times that it’s been really good for me to have saved all that e-mail because something comes up and people say, how did we get here? And I have sent them four or five e-

> mails, I sent this, I sent this, I got this, I got this…And you can open them up and they have the dates right there. (B; 31:9)

2.3 No bibliographic reference databases

Bibliographic reference databases or citation mangers are databases used to store, find, and reuse bibliographic citation information. In many graduate schools these are ubiquitous to index and allow for searching of materials gathered in the course of the research. None of the participants reported using any sort of database to manage, search, and reuse citations.

3. Information Keeping and Preservation

3.1 The Nitty Gritty Is Lost

With all of the careful efforts to organize, store, and backup research data, it is surprising that the engineers often do not maintain the "nitty gritty" details of their work. They frequently do not save exploratory analysis, if-then analysis, or back-of-the-envelope working papers, "the day to day scribbley note stuff that you use just to keep the work going – most of that will get tossed" (B; 20:22). One engineer mentioned that the engineering work is rarely archived once the deliverable is complete.

Some of this information may be captured in lab notebooks and journals discussed above.

3.2 The Engineer Provides the Best Preservation

The engineers take a great deal of responsibility for guarding and maintaining their stores of data and information. Lab, department, and project tools exist to archive information and backup systems, but these engineers all backup their work by creating CDs, using multiple computers, and by saving things up to shared drives. For very important work, the engineer will often maintain print copies. The CDs are marked with years covered and stored in the office. The backups are organized by time and are done on regular intervals. One participant sums up his backup system:

> …if you're going to put everything on the computer you have to be careful to back everything up… I maintain three computers, so the probability that the hard drives would fail simultaneously on all three computers is unlikely. I do backups. I know the lab does backups on the hard drive, but I also do backups of the hard drive on CDs and of all the documents and data files and those types of things so and I probably do that every six months… Having binders though, those are more permanent unless there's a fire or something which is extremely rare. You have records of that data. (D; 19:15)

4. Use of Specialized Tools

4.1 Software, Whatever Works

The lab provides licenses and support for many very expensive and elaborate engineering and science computation, modeling, and visualization packages. In many cases, though, the engineers prefer using standard office productivity software for drawing and high level calculations. This is in part because of the ease of incorporating the media files into the Power Point and Word deliverables, "some of it was Matlab and some of it was Power Point. Most of it was Power Point which is not the world's greatest thing to try to draw" (B; 18:10). Also, it may be easier to share files in a Microsoft format than in proprietary high level engineering software.

Another aspect is the burden of maintaining proficiency in configuring and using the sometimes difficult interfaces of the engineering products. The office productivity software is powerful enough for most purposes.

> To me it's a fantastic tool, I love Excel, I do all kinds of things in Excel. It's got great math functions behind it. It doesn't have a bad graphics package if everything is 2D…Only rare occasions have I used Matlab. One time I was doing a lot of 3-dimensional surfaces and something like Excel can't do it. (A; 24:10)

Discussion

Engineers maintain personal information collections over time as they accumulate information and build knowledge. A participant succinctly summarized this as: "You keep coming back to these same questions over and over again and you kind of build an understanding so I try to maintain a catalog of that understanding" (D;

7:16). They organize and access this information in different ways depending on whether it is related to project administration or their ongoing work.

The personal information collections maintained by the participants included a wide variety of data and information. Data were maintained in multiple formats and stages of processing. First, the participants maintained collections of raw data exported from laboratory equipment as well as their own and team members' hand recorded observations and photographs of experiments and field tests. These observations were sometimes transcribed or scanned into an electronic format, but were more often handwritten on forms designed for the purpose or written into special project laboratory notebooks (compare to Shankar, in press). Second, data collections included data that had been processed, reformatted, or converted for use in analysis or visualization software. Third, data collections included tables, graphs, reports of fully analyzed data, and the results of the equations and analysis that were ready for publication or pulled from reports of completed work.

The information in the participants' personal collections can be grouped by formality as well as by authorship and audience. Formal information included published textbooks, handbooks, journal articles, conference proceedings, and technical reports. Less formal information included e-mail, networks of experts, working papers, and calculations. Audience and authorship can be grouped into personal or self, organizational, and external. The collections included informal and formal materials authored by the participant for his or her own use, for organizational or project use, and for wider dissemination; informal and formal materials authored by the project team or organization's administration for the participant, his or her project, or external audiences; and materials from external authors for the participant, the project, or for external audiences.

Memory, and creation of resources to assist memory, plays an important role in the participants' PIM. The participants appear to work hard to achieve a balance between careful preservation and creation of finding aids and letting the nitty gritty, working papers, and e-mail get lost. While some information is backed up in multiple places, other information is discarded. The participants spend more time selecting and organizing some information that they expect to need to find and reuse while ignoring other information that they do not expect to need again.

The participants manage personal collections containing information from external, organization, and personal sources. They make their own handbooks from photocopies of published reference resources and from their own collections of measurements. They create and maintain project files that contain organizational information as well as published literature and their own work product. They create and maintain technical files that contain published articles including their own published articles.

Management of personal information includes:

- collecting, describing, and selecting information for retention;
- organizing, storing, and finding within the personal collection;
- reusing the information.

The engineers report very messy desks and elaborate organization schemes to find information. Even though there are careful organization schemes and backup systems for the personal information collections, these systems can fail, so the engineers rely on them frequently only as aides-memoire. Their memory of ideas presented at conferences, work they have done, and associations with time are the most important ways they have of finding information.

Implications for Information System Design

The participants in this study, all senior engineers in an applied research laboratory setting, have access to many PIM tools such as Microsoft Outlook, desktop search, electronic lab notebooks, citation managers, and so forth. Yet, as mentioned above, they either do not use these tools, or do not fully exploit the tools' features for PIM. Information technology staffers within the organization should work individually with senior staff members to set up email triage and management tools, to communicate the availability of desktop and intranet search, and to understand what is needed to allow engineers to trust central backups. Nevertheless, the results suggest some enhancements to systems. The following are ways PIM systems can be adapted to support the tactics of senior engineers found in this study.

Automating Journaling Over Time

Previous studies have suggested and evaluated software remedies for some of the issues mentioned by the participants. Cutrell, Dumais, and Teevan (2006) suggest that iterative search of a comprehensive personal archive may replace PIM tools. Such a system would have to track people, time, and other cues relevant in searching personal information. Some desktop search tools that are currently freely downloadable may provide the search engine. The comprehensive personal archive could be achieved by capturing everything a person sees and does as “MyLifeBits” (Gemmell, Bell, & Leuder, 2006) or “Stuff I’ve Seen” (Dumais et al., 2003) attempt to do.

While the creation of large archives may lead to serendipitous discovery in some cases, it may cause problems in others. Landsdale (1988) discusses how association can help memory and information retrieval. In fully digital systems, some of the visio-spatial and kinesthetic details are missing which may impede query formation and thus retrieval -- even if the items are in the archive. Second, print documents are important because of notes in the margin and other annotations which are still difficult in electronic documents (Whittaker & Hirschberg, 2001). Third, Kaye et al. (2006) found that engineers create personal archives to leave a legacy, share resources, and construct identity. With appropriate and granular security, this can be done with electronic archives, but setting of security restrictions might be overly burdensome.

Other studies suggest de-fragmenting information stores by labeling across different types of information such as e-mail, articles, and web pages (Jones, Bruce, Foxley, & Munat, 2005); by grouping heterogeneous items by project (Kaptelinin, 2003) or by creating a single hierarchy (Bergman, Beyth-Marom, & Nachmias, 2006). Again, this works for all electronic collections, or collections for which there are at least electronic pointers.

Personal Handbook and Enhanced Document Creation Tool

The personal handbooks mentioned above could be supported in an online environment with a combination of personalization, ebook, bookmarking, tagging, commenting and annotation, citation management, and clipping software. Electronic engineering handbooks are available primarily in PDF and html format. A personal handbook maker would allow the engineer to

- cut a chapter, table, equation, or data out of either a PDF document or an html document originating from an ebook, journal, conference paper, their output from analysis programs, or scanning of print;
- save the document to his or her personal workspace;
- annotate the information with handwriting, highlighting, or shapes such as boxes or circles;
- assign keywords or tags for retrieval;
- arrange and re-arrange these clippings as necessary;
- search or browse the clippings
- copy equations from this collection for use in analysis and word processing programs.

Citation information would be retained along with a link to the original source. The system would have a method for notifying the user if the original source is updated. This tool should be web-based, and default to private; however, the ability to share personal handbooks with colleagues would also be helpful. The ability to print sections or the whole book would make it easier for the engineer to use the material in closed areas or in the field where internet access is not practical.

The participants in this study expressed frustration in the preparation of documents for sponsor reports and conference and journal articles. Part of this could be helped through the adoption of a bibliographic management program. Existing bibliographic managers such as ProCite and RefWorks are weak in managing conference papers, standards, and technical reports – precisely the types of publications engineers use most frequently. These programs should be upgraded to better import information from government technical report stores and online resources.

The personal handbook software would also enable the engineer to create new documents through seamlessly remixing and transferring equations and graphics from the analysis, modeling, and CAD software as well as the personal handbook into a fully functional word processing or editing program. Citation information should automatically follow and land in the bibliography in the selected format. This bridging software must preserve the look of the graphic, and should also allow for manipulation and updating of the underlying information as is possible with office productivity products.

Blogs for Medium-Term Writing

Participants described preparing technical reports to archive and provide access to interesting results. There seems to be a gap between informal, internal documents and formal published reports that is not completely served by organizational written reports. Moreover, formal technical reports are time consuming to prepare and are not easily findable because departmental and project stores are often not indexed by the enterprise search engine. Engineers could write in smaller increments using blogging software. Blog posts are automatically reverse-chronologically arranged, time and date stamped, and fully searchable. Additionally, most blog software allows for the assignment of keywords and categorization for easier retrieval. Posting information in smaller chunks might seem less daunting than preparing a formal report while still allowing for information sharing, serendipitous finds, and archival storage. Seamless posting of equations, tables, graphs, and other media is a requirement for such a system.

Future Work

Additional interviews with additional senior engineers would show if the PIM methods reported here are more widely practiced. Also, similar interviews with senior physical scientists such as chemists and planetary scientists may indicate which of the methods are unique to engineers or scientists working in traditional engineering functions. Repeating this study with junior engineers may show what is common to the field, and what is learned over time working in a research laboratory.

Additional exploration of senior engineers' use of e-mail would enable training on existing tools to make triage and management easier, and may allow the development of systems to enable the engineers to retrieve more information from this largely untapped collection.

Acknowledgments

I would like to express my thanks to the participants for so freely giving of their time and expertise. Thanks also to Susan Fingerman for her assistance in participant selection and paper editing. I would like to thank Dr. Linda Aldoory and Dr. Dagobert Soergel at the University of Maryland for providing constructive criticism on drafts of this work and for their tremendous assistance in the preparation of this paper.

References

Allen, T. J. (1977). Managing the flow of technology: Technology transfer and the dissemination of technological information within the R&D organization. Cambridge, MA: MIT Press.

Anderson, C. J., Glassman, M., McAfee, R. B., & Pinelli, T. (2001). An investigation of factors affecting how engineers and scientists seek information. *Journal of Engineering and Technology Management, 18*(2), 131-155.

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

Bergman, O., Beyth-Marom, R., & Nachmias, R. (2006). The project fragmentation problem in personal information management. Paper presented at the *CHI '06: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* Montreal, Canada. 271-274.

Bovey, J. (1996). Event based personal retrieval. *Journal of Information Science, 22*(5), 357-366.

Bruce, H. (2005). Personal, anticipated information need. *Information Research-an International Electronic Journal, 10(3)*, paper 232. Retrieved January 9, 2006, from http://informationr.net/ir/10-3/paper232.html.

Bruce, H., Jones, W., & Dumais, S. (2004). Information behaviour that keeps found things found. *Information Research-an International Electronic Journal, 10*(1), paper 207. Retrieved January 9, 2006, from http://informationr.net/ir/10-1/paper207.html.

Chakrabarti, A. K., Feineman, S., & Fuentevilla, W. (1983). Characteristics of sources, channels, and contents for scientific and technical information systems in industrial R and D. *IEEE Transactions on Engineering Management, EM-30*, 83-88.

Cool, C., & Xie, H. (2000). Patterns of information use, avoidance and evaluation in a corporate engineering environment. . *Proceedings of the ASIS Annual Meeting, 37*, 462-472.

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

Dumais, S., Cutrell, E., Cadiz, J. J., Jancke, G., Sarin, R., & Robbins, D. C. (2003). Stuff I've seen: A system for personal information retrieval and re-use. Paper presented at the *SIGIR '03: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,* Toronto, Canada. 72-79.

Dye, J. F., Schatz, I. M., Rosenberg, B. A., & Coleman, S. T. (2000). Constant comparison method: A kaleidoscope of data. *The Qualitative Report, 4* (1/2). Retrieved May 4, 2007, from http://www.nova.edu/ssss/QR/QR3-4/dye.html.

Ellis, D., & Haugan, M. (1997). Modelling the information seeking patterns of engineers and research scientists in an industrial environment. *Journal of Documentation, 53*, 384-403.

Erdelez, S. (1999). Information encountering: It's more than just bumping into information. *Bulletin of the American Society for Information Science, 25*(3), 25-29.

Fidel, R., & Green, M. (2004). The many faces of accessibility: Engineers' perception of information sources. *Information Processing & Management, 40*, 563-581.

Hertzum, M., & Pejtersen, A. M. (2000). The information-seeking practices of engineers: Searching for documents as well as for people. *Information Processing & Management, 36*, 761-778.

Hirsh, S. G., & Kraft, D. H. (2000). Information needs, information seeking, and communication in an industrial R&D environment. *Proceedings of the 63rd ASIS Annual Meeting,37.* 473-486.

Jones, W. (2007). Personal information management. In B. Cronin (Ed.), *Annual review of information science and technology* (pp. 453-504). Medford, N.J.: Information Today.

Jones, W., Bruce, H., Foxley, A., & Munat, C. F. (2005). The universal labeler: plan the project and let your information follow. *Proceedings of the ASIS&T Annual Meeting, 42.*

Kaptelinin, V. (2003). UMEA: Translating interaction histories into project contexts. Paper presented at the *CHI '03: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* Ft. Lauderdale, FL, 353-360.

Kaye, J., Vertesi, J., Avery, S., Dafoe, A., David, S., & Onaga, L., et al. (2006). To have and to hold: Exploring the personal archive. Paper presented at the *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* Montréal, Québec. 275-284

Kennedy, J. M., Pinelli, T. E., Barclay, R. O., & Bishop, A. P. (1997). Distinguishing engineers from scientists-the case for an engineering knowledge community. *Knowledge diffusion in the U.S. aerospace industry-managing knowledge for competitive advantage* (pp. 177-213). Greenwich, CT: Ablex Publishing Corp.

Kraut, R. E., Egido, C., & Galegher, J. (1990). Patterns of contact and communication in scientific research collaboration. In J. Galegher, R. E. Kraut & C. Egido (Eds.), *Intellectual teamwork: Social and technological foundations of cooperative work* (pp. 149-171). Hillsdale, NJ: L. Erlbaum Associates.

Kwasitsu, L. (2003). Information-seeking behavior of design, process, and manufacturing engineers. *Library & Information Science Research, 25*, 459-476.

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

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

Leckie, G. J., Pettigrew, K. E., & Sylvain, C. (1996). Modeling the information seeking of professionals: A general model derived from research on engineers, health care professionals, and lawyers. *Library Quarterly, 66*, 161-193.

Lethbridge, T. C., Singer, J., & Forward, A. (2003). How software engineers use documentation: The state of the practice. *IEEE Software, 20*, 35-39.

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.

Rubin, H. J., & Rubin, I. (2005). *Qualitative interviewing: The art of hearing data* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Shankar, K. (in press). Order from chaos: The poetics and pragmatics of scientific recordkeeping. *Journal of the American Society for Information Science and Technology*.

Shuchman, H. L. (1981). *Information transfer in engineering*. Glastonbury, CT: The Futures Group.

Taylor, R. S. (1991). Information use environments. In B. Dervin (Ed.), *Progress in communication sciences* (pp. 217-255). Norwood, NJ: Ablex.

Tenopir, C., & King, D. W. (2004). *Communication patterns of engineers*. Piscataway, NJ: IEEE Press.

Von Seggern, M., & Jourdain, J. M. (1996). Technical communications in engineering and science the practices within a government defense laboratory. *Special Libraries, 87*(2), 98-119.

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

Appendix: Final Interview Guide

1. Please tell me about your academic background.
>How do you feel that your academic background is meaningful to your job here at APL?

2. What projects are you working on right now?
>Please tell me about your role in these projects.
>What is an average day like working on these projects?
>Are these projects similar to the work you have been doing?
>What is the end product or deliverable of this work?

3. For your current or most recent projects, please explain to me how you collect and keep track of the data.
>How is this data stored?
>How is it shared?
>How is it archived for later use/re-use?
>How do you organize the stored data?

4. When you think of a recent project, please explain the process you used to find information to solve a particular problem you encountered.
>How did you find previous applicable work in the field?
>How did you locate experts to ask?

5. Tell me about your process when you go to write up your work and how you organize your information as you write up your work.
>How do you deal with incorporating images, tables, etc, into your papers?
>How do you incorporate mathematical equations into your papers?
>How do you attribute ideas you’ve found in the literature?

6. After you finish with a project, what do you do with the information and data that you've gathered?

7. Can you share any tips and tricks you use to keep track of journal articles, conference papers, or patents so that you can find them again?
>What does your organization scheme look like?

8. I'm interested in the differences in managing digital information and print information. How do you view the differences? Have you developed strategies to work in an environment where print and digital information sources are mixed?
>How do you read information you find in electronic format?
>How do you find information on the web?
>How do you keep track of information you find on the web?
>How do you deal with e-mail?

9. When you mentor less experienced staff members, do you advise them on ways to manage their work-related information? Please provide examples.

10. How do you share information among co-authors or among research team members?
>If you collaborated on projects, can you describe the process of how you co-author papers?

11. Are there any questions about your personal information management that I didn't ask and should have?