Towards a field of PIM inquiry

William Jones, Manuel Perez-Quinones, Marcia Bates, Mike Franklin, David Karger. David Levy, Mel Knox (student volunteer).

The discussion group was organized to consider key questions relating to PIM as a field of inquiry including:

1. What does it mean for PIM to be a field of inquiry? What does it take? Is this necessarily what we want?
2. What is PIM (at its core) and isn't? What are its components?
3. Is there a conceptual framework which might help (as a way to approach PIM and its components)?
4. How do we measure progress in PIM as a field? With what benchmarks?

Since another discussion group led by Diane Kelly was also consider question #4, our group focused on the first 3 questions. We considered the possibility that what's needed for progress in PIM is a community – from this may emerge a field over time. There is no field. A community is emerging. There is certainly interest and a need. The group agreed that many important PIM concerns were currently "falling through the cracks" . On the other hand, PIM as an area of study, provides a good meeting ground and area of application for the work of several different disciplines including information retrieval, database management, artificial intelligence, human-computer interaction and cognitive science.

The discussion then moved to a consideration of what PIM should encompass as an area of study. PIM is a large area with uncertain boundaries. It includes all efforts to work with, deal with, and react to information at a personal, individual level. PIM includes various activities to search for, find, encounter, interpret, decide to keep (or not), file and organize for re-use, re-access and ultimately use information. Good, timely information is critical to a wide range of tasks, professional and personal.

A deeper understanding of what PIM is, at its core, and at its broad periphery of overlap with other fields of inquiry, begins with consideration of definitions for PIM and associated concepts.

Some working definitions

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

Information and the information item

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

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

An ***information item*** is a packaging of information. Examples of information items include: 1. paper documents. 2. electronic documents and other files. 3. email messages. 4. web pages or 5. references

(e.g., shortcuts, alias) to any of the above. Some might prefer to use the term “information object” to emphasize the point that an information item can be acted upon. Items encapsulate information in a persistent form that can be created, stored, moved, given a name and other properties, copied, distributed, deleted., moved, transformed, etc.

The support that we depend upon for our interaction with paper-based information items includes our desktop, paper clips, staplers, filing cabinets, etc. In our interactions with digital information items we depend upon the support of various computer-based tools and applications such as an email client, the file manager, a web browser, etc. The “size” of current information items is partly determined by these applications. There are certainly situations in which some of us might like an information item to come in smaller units. A writer, for example, might like to treat paragraphs or even individual sentences as information items (to facilitate their re-use). A salesperson might view the individual entry in a contact management database as an information item. Applications exist in each case to help (e.g., contact management software, writer’s software such as DevonThink)..

An information item has an associated ***information form*** 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 report are paper documents, e-documents and other files, email messages and web bookmarks.

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

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

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

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

Personal information.

The discussion group considered several senses of p*ersonal information*:

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

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

The third case -- information we experience but do not keep in our PSI – can sometimes pose a special kind of PIM problem: We remember the information, but maybe not enough about the information to be able to find it again later. For example, we might see information on a web site about a concert in another city by our favorite musical group. But since we can't attend, we take no special steps to keep this information. Later, we find out that we must attend a business meeting in that city on the same week as the conference. We want to get back to the web page but can't recall how we got there to begin with and can seem to formulate a query to return the web page as one of the results (we're not sure since we don't know what web site's name is or how it would appear in the listing of results).

A Personal Space of Information

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

A few other things to note about a PSI:

- Although we have some sense of control over the items in a PSI, this is partly illusory. For example, an email message can be deleted so that it no longer appears in an inbox. However, the message is very likely still around somewhere (as some have learned to their chagrin).
- A PSI does not include the web pages we have visited but may include copies (in a cache) and does include the bookmarks we create to reference these pages.
- Does PSI include our own internal memories? On the one hand, the answer must surely be yes. What could be more personal? No one else owns our memories but us. But, paradoxically, an argument can be made for "no". How much control do we have over what goes into our memories? Or what comes back out? Some things lodge in our minds even though we wish they would not. We cannot forget, i.e. we cannot simply press a "delete" key.
- In general, there are large unavoidable grey areas. For example, the files we place on a network share should probably be considered a part of our PSI even though they may not be under our exclusive control. Similarly, a PSI should probably include the many icons that applications like to leave on our computer desktops and the bookmarks and folders that are automatically created.
- A PSI is, by definition, "everything". We each have only one PSI.
- A PSI is distinguished from a Personal Information Environment (PIE) which, as used in the literature, commonly refers to subset of a PSI in combination with supporting tools. The physical space of an office including papers piled and filed, the stapler, filing cabinets, etc. is a PIE. A notebook computer is a PIE. A person can have several PIEs.
- The size of our PSI continues to grow, especially with respect to digitally encoded information. The PSI is a potential source of information for use a number of different ways. The PSI might be used, for example, to customize our experience of the Web (see the section below on finding/re-finding). The information of a PSI might be "mined" to extract important patterns in our information (and our interactions with this information). Effective re-use of the information in the PSI promises to improve our productivity. At the same time, the growing size of our PSI also raises serious questions of privacy and security.

Definitions of Personal Information Management

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

Lansdale<ref> refers to PIM as "the methods and procedures by which we handle, categorize, and retrieve information on a day-to-day basis". Bellotti<ref> describes PIM as "the ordering of information

through categorization, placement, or embellishment in a manner that makes it easier to retrieve when it is needed".

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

Boardman <ref> notes that "Many definitions of PIM draw from a traditional information management perspective – that information is stored so that it can be retrieved at a later date".

In keeping with this observation, as exemplified by Barreau's definition, we might analyze PIM with respect to our interactions with a large and amorphous PSI. From the perspective of such a store, the essential operations are input, storage (including organization) and output.

In rough equivalence to input-storage-output breakdown of actions associated with a PSI, the group considered a conceptual framework with the following grouping of essential PIM activities:

- **Keeping** activities affect the input of information into a PSI.
- **Finding/re-finding** activities affect the output of information from a PSI.
- **"M-level activities"** (e.g., "m" for "mapping" or for "maintenance and organization") affect the storage of information within the PSI.

This framework was discussed in the group and then presented to the larger group of all workshop participants. There was consensus to elaborate upon this framework for the final report. The following section makes a first attempt to do this.

A conceptual framework and focus: PIM activities that map between information and need

Note: the following section is a substantial elaboration on the keeping/finding/m-level framework initially discussed in the discussion group and then presented to the entire

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

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

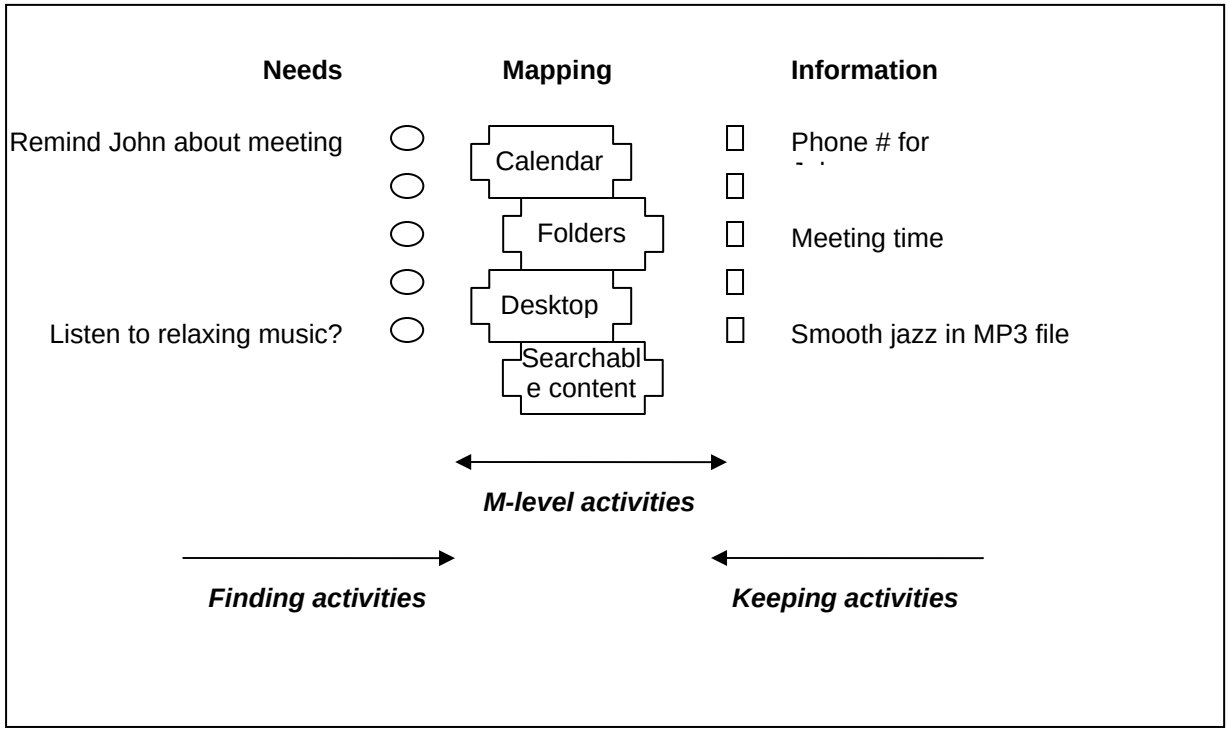

Figure 1. PIM activities viewed as effort to establish, use and maintain a mapping between needs and information

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

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

Connecting between need and information is a mapping. Only small portions of this mapping have an observable external representation. Large portions of the mapping are internal to our own memories – memories for specific experiences with information, experiences with information sources and kinds of information and, more broadly, our memories for the fabric of the world around us, its conventions, its "language" – it all goes into the mapping. Large portions of the mapping are potential and not realized in any form – external or internal. A sort function or a search facility, for example, has the potential to guide us from a need to desired information.

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

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

From need to information

find(need) -> information

We have a need. We try to find information to meet that need. Needs can be large and multi-faceted – the need for information for a review, for example – or small and simple – a locating someone's phone number. Needs can originate, more or less, in our own heads or they can come from outside – a hallway request from a colleague, for example. Frequently, a need, itself, comes packaged in an information item – an email request, for example, or via a web-based form requesting certain information for its completion. A need frequently equates with or is a part of a task (e.g., "prepare for the meeting", "answer my boss's email", "return the client's call"). But other needs may not fit tasks except by the broadest definition (for example, "see that funny web site again" or "hear 'Five to one' for old time's sake").

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

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

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

Information items may occupy a virtual space, but such a space cannot, yet, compete with the richness of our physical spaces. On the other hand, we can search for digital information using computer-based tools in ways that we cannot (yet) use for the search of physical objects. But there are many similarities as well. We can fail to find an ingredient – walnuts, for example – that might be perfect for a salad we're making for any of several reasons – each with their digital analog. The walnuts may not be on the shelves we look through in the kitchen. Perhaps they aren't in the kitchen at all. Or the walnuts may be right there in front of us on the shelf but in a container that we do not recognize. Or, in the midst of everything else we are doing to prepare dinner, perhaps we forget to look for the walnuts – this, too, is a failure of finding. Similarly, we may fail to find a web site we have bookmarked for our current project for any of several reasons. We may look in the wrong folders, or perhaps the bookmark is on another computer entirely, or we may fail to recognize the bookmark though is there in front of us. Or, especially in our rush to complete the project, we may forget about entirely about the bookmark.

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

Several questions arise concerning the actions of finding vs. re-finding how these might change when the target is information inside a PSI vs. information on the Web or elsewhere "out there" (see also the breakout report on finding, http://pim.ischool.washington.edu/breakouts.htm). Several studies indicate an enduring preference for browsing (e.g., by going through a nesting of folders) as a means of return to information within a PSI. Certainly, search is widely used on to locate information on the Web but,

search is often used in common with a hyperlink-enabled kind of browsing. For example, a we might use a search service to find a web site of interest and then browse within the site to locate web pages containing specific information of interest. Some members felt that there wasn't much difference between finding activities targeting new information on the Web vs. re-finding activities directed towards information in the PSI that has already been experienced. However, other breakout participants noted that there is often a strong emotional component associated to attempts to re-find information in our PSI. We may get frustrated, feel back or feel that we're "losing control" when we can find information that we "know is in there somewhere". For items repeatedly re-accessed and re-used there is also a question concerning when we "keep" this information in a way that makes a subsequent effort to re-find (re-access) easier. For example, if we go the same web site via Google search two or three times a week and then follow hyperlinks to a specific page at that site, when do we decide to "keep" the page in some other way – by making a link or a bookmark, for example – so that re-access is faster?

From information to need

keep(information) -> need

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

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

There is then a middle area of encountered information. We may have a need for this information, but not now. We must then decide whether to keep this information and, if so, how. Even if we judge the information to be useful, we may still decide that no special action is required – perhaps because we already "have" this information somewhere in our PSI or because we can easily return to the information, for example, by repeating the same search or the same path of hyperlinks that brought us to the information in the first place.

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

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

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

If our information is fragmented between devices and applications we must also anticipate the form in which we will need the information. On which device? (Laptop or mobile phone?) In which application?

As the example of the phone number illustrates, the number of ways to keep information has grown considerably in recent years as part of an overall increase in the number of devices and applications that we depend upon to manage our information. Paper is still very much a part of people's lives. In addition, we now manage electronic documents and other computer-based files, web references (as bookmarks, for example) and, of course, large numbers of email messages often in multiple accounts. We have desktop computers, laptop computers, smart phones, PDAs and ordinary notebooks.

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

A focus on the mapping between need and information

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

- ***Mapping.*** As noted earlier, only small portions of the mapping for a PSI have an observable external representation. Large portions of the mapping are internal to our own memories – memories for specific experiences with information, experiences with information sources and kinds of information and, more broadly, our memories for the fabric of the world around us, its conventions, its "language" – it all goes into the mapping. Large portions of the mapping are potential and not realized in any form – external or internal. A sort function or a search facility, for example, has the potential to guide us from a need to desired information.

- ***Meta.*** One m-level activity is to "step back" and think about the mapping overall or for a subset of the PSI (e.g., the files on a laptop computer). How should information be structured? According to what schema? For common forms of information, this means deciding on a folder structure. But in the future we may also be able to organize items according to a rich set of properties. Certainly a challenge in such a property-based system will be to select properties that truly distinguish among the items, current and likely, without creating a lot of extra work. It is at the meta level that we also consider the potential utility of supporting tools that are proffered to help us. And we also consider strategies of PIM ("file everything right away", "don't file anything", "keep everything", "don't' keep any paper", etc.)

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

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

By analogy, the food and utensils of a kitchen or the clothing of a closet may occasionally need to be re-distributed. We may also periodically attempt to weed out older items for donation that are no longer in use.

Maintenance of a PSI also includes updating of information content as well as organization. When much the same information is scattered in different forms and many variations through a PSI, updating can be extremely difficult. If, for example, a friend's email address or phone number or, worse, name

changes, older information items with the incorrect information (e.g., birthday reminders, directions, holiday card lists, etc.) may linger for months or years after our initial efforts to update. More extreme, are situations for large parts of our personal information are rendered irrelevant or "wrong" by a new event. People recovering from a heart attach, for example, may want a radical update in their PSI to reflect the radical change they hope to accomplish in their style of life.

- ***Managing*** *privacy, security and the distribution of items in a PSI.* A discussion of privacy and security brings us back again to a consideration not only of "our information" but also information "about us" and the large overlap between these two kinds of personal information. If our first reaction is to say "personal information is personal and no one else can see it" we are likely to have a later realization that some distribution of our personal information can be very useful. We want the travel agent to know about our seating preferences. We want colleagues and friends to know about our schedule. We may want close friends and family to know about our current condition if we are battling a serious illness. The increasing use of the personal web sites as a means to publish (and project) naturally brings a desire for technology that can support a "personal policy on privacy and security" that allows for finer distinctions that "everyone can access" or "no one can access". But, given this greater control, there is a need for user interfaces that can guide us in our choices and make clear their implications.
- ***Measuring*** *the effectiveness of a mapping and the structures, strategies and supporting tools associated with its creation, use and upkeep.* We must periodically ask ourselves "is it working?" Are the structures we've selected maintainable? Are the strategies we try to follow sustainable? Is this tool really helping or is it more trouble than it's worth? For paper documents, the signs that "things aren't working" are sometimes all too clear. For example, if paper documents continue to pile up in a "to be filed" stack and we never have time to actually file these documents away, this may be a sign that our "great new organization", for all its promise, is simply not sustainable. The signs for digital information may be more subtle. As we look for efficient, accurate, objective ways to evaluate our own practice of PIM we run into many of the same problems, at an individual level, that are also in evidence for the field of PIM. We return to this topic in the next section and also in a later section on the methodologies of PIM.
- ***Manipulating and making sense*** *of our information.* As we consider a collection of informaiton, what are we seeing? What do we have? The folders that we still use as perhaps the most common way of organizing information items can also obscure. They can create barriers within a PSI not unlike the barriers Bush observed between an ever increasing number of scientific specializations: "*publication has been extended far beyond our present ability to make real use of the record. The summation of human experience is being expanded at a prodigious rate, and the means we use for threading through the consequent maze to the momentarily important item is the same as was used in the days of square-rigged ships." <ref>.* The wording in these sentences needs only slight modification to apply equally to the prodigious amounts of information we are able to store in a PSI. And we might indeed complain that the tools we have available for manipulating and making sense of, for example, a collection of computer-based files has changed little over the past two decades.
- ***Mañana****? Or maybe tomorrow (but not today).* We might also say, jokingly but with considerable truth, that "m" stands for "maybe tomorrow but not today". The m-level activities described here are easy to avoid and put off. None of them demand our attention in the way that an immediate need or even encountered information do. We perform activities of finding and keeping throughout a typical day. M-level activities can and are postponed for weeks on end. And then there is that messy closet…. Part of the problem is that we prefer to pay the incremental, perhaps barely noticeable, costs associated with the use of a poor mapping rather to suffer the certain and immediate costs of an m-level activity.