14. Personal Information Management

INFO 202 - 15 October 2007

Bob Glushko
Plan for Today's Class

Defining "Personal Information Management"

PIM Technology and Tools

Personal vs Common Information Spaces

PIM Activities and Strategies
Vannevar Bush's Personal Information Manager
Personal Information Management (PIM) - A Fanciful Definition

Personal Information Management is a game of catch
... in which a person tosses their personal information into the future
... in the hope of being able to catch it later

Maybe "later" is "forever"
PIM - A More Serious Definition

"The practice and the study of the activities that people perform to acquire, organize, maintain, and retrieve information for everyday use"

So we limit PIM to cover actions (or inactions) that are the result of individual choices

It is also personal -- rather than social -- is that we decide on the activities and organizational schemes and carry them out by ourselves

Having this discretion about information organization and "making sense of it" to make decisions is a defining characteristic of professional, as opposed to clerical, information work
Why PIM Matters

PIM matters to us as individuals and professionals because better PIM results in better use of our time and attention and ultimately in better quality of life.

PIM matters within enterprises because better PIM means increased productivity (in the short term) and better knowledge management (in the long term).

Advances in PIM may also translate to improvements in education and in helping old people "match their mental lifespan to their physical lifespan."
PIM and Prevailing Technology

We can view many inventions as responses to the need for PIM.

Because PIM is embedded in user tasks and work context it reflects the prevailing technology support for information work.

The personal computer has had the greatest impact (so far) on PIM technology.

But as processors and connectivity is increasingly embedded in objects of all kinds information "from and about stuff" will have to be managed and the PC may lose its central role.
19th Century PIM Technology Breakthroughs
The Contemporary PIM Challenge

Vastly more personal information of all kinds

Much of this information is in incompatible digital formats produced and consumed by a variety of devices and applications

These devices and applications may separately provide support for PIM that is collectively incompatible

But our work often requires that information be created, managed, and retrieved in ways that cut across these format and device boundaries
# Data Unification in PIM

<table>
<thead>
<tr>
<th>APPROACH</th>
<th>EXAMPLE</th>
<th>OPERATIONS</th>
<th>ENABLES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard datatype</td>
<td>Text</td>
<td>Cut/copy/paste</td>
<td>Unified search</td>
</tr>
<tr>
<td>Unified presentation</td>
<td>Window manager</td>
<td>Layout, tile, show, hide</td>
<td>Simultaneous view</td>
</tr>
<tr>
<td>Unified namespace</td>
<td></td>
<td>Reference, de-reference</td>
<td></td>
</tr>
<tr>
<td>Grouping</td>
<td>Directories</td>
<td>Group, ungroup</td>
<td>Organize, browse</td>
</tr>
<tr>
<td>Metadata</td>
<td>ID3 tags, XML</td>
<td>Annotate, query</td>
<td>Search, organize, browse</td>
</tr>
<tr>
<td>Cross-ref</td>
<td>Web links, OLE</td>
<td>Link, follow, embed</td>
<td>Simultaneous view, orient</td>
</tr>
<tr>
<td>Relations</td>
<td>RDF, Haystack</td>
<td>Record named relations</td>
<td>Unified search, browse, orient</td>
</tr>
</tbody>
</table>
"Data Unification" Via Standard Datatype Should Remind You Of...

- Content/structure-based text objects: XML, SGML, databases
- Formatted electronic text: HTML, EDI, word processing files
- Unstructured electronic text: ASCII
- Printed text
E-Mail and PIM

E-mail was designed as a communications application, but it almost always takes on task and information management functions.

A "touch once" strategy assumes that messages can be classified as either "information" or "correspondence".

But this can't be done reliably.

Strategies for organizing messages have varying effectiveness and costs.
Favorite / Least Favorite PIM Tools and Features

What's your favorite / most used / most usable tool or tool feature for your PIM?

Are your organizational schemes and document models the same, similar, or consistent across the PIM tools you use?

How do your choices affect others in your social or work networks?
"Personal" in Context

We want to focus on "personal" IM by excluding activities that are mandated by the social, institutional, or work context a person is in.

But that's impossible if we think that a person's conceptual systems are "embodied" in their perceptions, physical and social experiences, etc...

So PIM is always shaped by the context of work.

There may also be conflicts between what people would do as individuals and what they are required to do as employees, project members, etc.
The Concept of a Personal Space of Information

All of the information that is nominally under a person's control defines one and only one PSI for that person.

So the PSI is an abstract collection of all of the various physical and digital information items.

Some of it is located "in the network" and is thus only indirectly controllable.

A fundamental challenge is maintaining an integrated view of the PSI.

A corollary challenge is reconciling the tension between one's PSI and the need to contribute and use information from common information spaces.
Pushed vs Pulled Information

*Pushed* information comes to our PSI (not as the result of any *immediate* action on our part)

*Pulled or Retrievable* information is what we intentionally seek
Categorizing PIM Activities

*Keeping*
activities are carried out when you encounter information - they determine the input of information into the PSI

*Finding or re-finding*
activities within the PSI determine the output of info from the PSI or its use in work

*Maintenance and organization*
activities are those taken to organize, re-organize, or integrate the PSI
Decisions to Make on Information Exposure / Receipt

Consume immediately

Ignore

Keep
Does the "to be kept" information item fit into our existing classification scheme within our PSI?

This is often not an easy or binary decision

If we use an existing category or location, is the degree of "fit" acceptable or does it distort the existing organizational scheme?

Do we abstract / summarize / data mine / secure or otherwise process the "to be kept" information?
Where We Keep Information

*Action information*
- immediately available, organized by task, often in physical formats and organized spatially; often serves as reminders of what to do as much as information to help do it

*Personal work files*
- accessible, organized spatially or according to conceptual categories, but not intended for use by others

*Archive storage*
- not as immediately accessible, often files of completed work and possibly organized for use by others
"Pilers" and "Filers" in Malone's Classic Paper


- A classic paper that contrasted the strategies of "filing" and "piling" in managing information in offices
- Both files and piles are ways of collecting groups of elements into larger units; files are units where the instances are explicitly titled and arranged in some systematic order, and these structures may themselves to explicitly titled and systematically organized
- Piles tend not to have internal structure, other than access frequency; their spatial location is often the key to finding them
Diagram of a Piler's Office
Photo of a Piler's Desk
Diagram of a Filer's Office
Photo of a Filer's Desk
PIM and Search

Search can be considered as the unifying or integrating function that cuts applications or information sources to help make sense of your PSI. This is especially true in PC or intranet environments.

We make decisions about the likelihood of having to re-find something when we organize it and the expected cost of re-finding it versus finding it "from scratch."

How does "regular" search of the kind you'd use on Google not take advantage of the special properties of Personal information?
Finding vs Re-finding

Finding
implies when no previous experience (or memory of) having the needed information
there is often a stronger emotional dimension in re-finding previously experienced information

Re-finding
is facilitated because people generally know more contextual metadata about previously-found information
Kirsh on "Cognitive Overload"

Kirsh points out that "where we work" is a superposition of many specific environments and applications that we move in and out of.

Each of these environments and applications has its own cost structure for handling information based on the tools and resources it makes available.

These diverse cost structures result in "computational complexity" for making PIM decisions about keeping and finding information and encourage suboptimal reactive methods rather than careful planning.
Kirsh on "Information Utility"

People think about and value different kinds of information in different ways - their utility functions are non-monotonic and non-linear.

When you are looking for information, can you tell how hard or long it will be before you find it?

How long should you look before you give up?

How valuable will information you don't yet have turn out to be?
Kirsh's Information Utility Functions

- A little goes a long way
- A little is a dangerous thing
- Past a given threshold knowledge is suddenly valuable

Utility of knowledge vs. Quantity of Knowledge
PIM Strategies [1]

The lack of a coherent utility or demand function for information means that different people (or the same person at different times) will follow radically different PIM strategies

*Pack Rat or Blind Accumulation*
-- just save everything, usually spends excessive time filing

*Insurer*
-- Doesn't keep everything, but creates multiple copies (paper and digital) of information items to maximize re-finding (KFTF)

*Surface Clutterer*
-- Doesn't keep everything, but strives to keep information accessible, often in spatially organized piles
PIM Strategies [2]

Just-in-Case Learner
-- Spends excessive time consuming information when it arrives so they can always be prepared for some future information need

Just-in-Time Gatherer
-- Ignore all information needs except those needed immediately for current tasks. Maximizes the average value of information items, but some high-value information can't be found this way
Evaluating PIM Strategy

Is it better to pay a small and occasional incremental cost associated with a locally sub-optimal PIM strategy than to pay the certain large costs of an explicitly good strategy?

How can you evaluate your management of information in physical formats?

How can you evaluate your management of information in digital formats?

How does the representation of information change its "cost structure" in using it?

Is this change the same for all "computational agents" ... including people?
A Georgia Tech research project on “Personal Information Environments” has created a smart alarm clock that communicates wirelessly with other devices and information sources.

A change in a person’s morning schedule, whether in a PDA or another PIM product on a desktop will reset the alarm clock to accommodate the morning appointment.
A Lifetime of PIM

[Diagram showing relationships between self and others through information used and recorded]

- Self
  - Tutor
  - Mentor
  - Advisor
  - Diary/Journal
  - Meeting Prep
- Others
  - Babysitter
  - Companion Caretaker
  - Parole Officer
  - Pars Flight Recorder
  - Assistant for Elderly
  - Photo Album
  - Autobiography
  - Captain's Log
  - Personal Proxy
- Information used by
  - Conservatory
  - Biography
  - Baby Book
  - Executor
  - Obituary
- Information recorded by
  - Financial Manager
  - Medical Manager
Readings for INFO Lecture #15

Maureen Breitenberg, The ABC's of Standards-related Activities in the United States, NBSIR 87-3576, May 1987

Lois Chan & Marcia Zeng, "Metadata Interoperability and Standardization – A Study of Methodology Part I. Achieving Interoperability at the Schema Level (skip or skim "Sources and References")"