

Personal Information Management – Literature Review

Compiled by Meriel Patrick, [Sudamih Project](#), Oxford University Computing Services

Introduction – Aims and Scope.....	1
Organizational Strategies.....	1
Aims and Purposes of Organization.....	4
Finding Information.....	5
Types of Information.....	5
Storage Behaviour.....	6
Summary and Conclusions.....	7
Bibliography.....	8

Introduction – Aims and Scope

This literature review was compiled in the autumn of 2010 as part of the JISC-funded Sudamih Project¹. Its aim was to draw together ideas, concepts, and study findings which are of practical relevance for personal information management, with a view both to understanding how individuals typically organize their material, and to inform training resources being developed as part of the Sudamih Project.

Consequently, it deals with only a selection of the literature on this topic: that which seemed most relevant and useful to the task in hand. It does not include the substantial proportion of the works in this area which are aimed at software developers, or those which simply give details of observational studies (e.g. investigations into the mean number and size of folders among a particular group of users) which are of little practical value from an information management point of view.

Organizational Strategies

Malone’s seminal 1983 essay “How do people organize their desks?”² identifies two key concepts for the organization of paper materials: *files* and *piles*:

“Files are units where the elements (e.g., individual folders) are explicitly titled and arranged in some systematic order (e.g., alphabetical or chronological). ... In piles, on the other hand, the individual elements (papers, folders, etc.) are not necessarily titled, and they are not, in general, arranged in any particular order. ... Furthermore, the piles themselves are not titled and they are not necessarily

¹ ‘Supporting Data Management Infrastructure for the Humanities’, based at Oxford University Computing Services: <http://sudamih.oucs.ox.ac.uk/>.

² In *ACM Trans. Inf. Syst.* 1, no. 1 (1983): 99-112.

arranged in any particular order on the desk or table top. Since piles have no systematic order, however, their spatial location is often especially important in finding them.” (p. 106.)

Whittaker and Hirschberg, writing in 2001³, recognize the distinction, but suggest that the difference between filers and pilers is one of degree: in their study, all participants filed some material and kept other items in piles (p. 159). Contrary to their expectations, those who tended more towards piling generally had smaller archives and accumulated material less rapidly, perhaps because filers sometimes file material prematurely, and are then reluctant to throw it away (pp. 160-2). Pilers did often experience more difficulty in locating material, but this problem was not unique to them: filers with complex systems sometimes also found it hard to remember where things had been filed (p. 161).

Henderson’s 2009 essay⁴ suggests a third category: *structurers* (p. 70). Structurers (who in Malone’s terminology would probably count as a type of filer) tend to have filing systems which are narrower (fewer top level folders) and deeper (more levels of folder nesting), and may also make more use of additional organization methods such as colour coding and tagging with keywords. Henderson points out that people with different styles of organization need different types of user interface: rather than attempting to encourage people to fit into one standard organizational model, the emphasis should be on supporting a range of systems that are designed to accommodate individual preferences.

Indratmo and Vassileva⁵ provide a useful overview of types of organizational structure. They divide the terrain slightly differently, identifying five types of system: hierarchical, flat, linear, spatial, and network.

Hierarchical systems use folders and subfolders. They are familiar to most people (Malone and Henderson seem to be thinking principally of this sort of system when they talk of filing and structuring), and help users to see relationships between concepts. Folders sometimes represent tasks, and so can play a role in task management. However, a number of studies have shown that “Creating a hierarchy and classifying information accordingly is a heavyweight cognitive activity” (p. 3). The effort involved in classifying material often leads to hierarchical systems not being consistently maintained. Malone points out that this can often be what leads to information piling: when users are uncertain where to file something, or don’t have sufficient time, piles of loosely related but unclassified material often form as a result (Malone, p. 108).

Additionally, projects develop over time, so hierarchies may become out of date, and information may fit in multiple places in the hierarchy (Indratmo and Vassileva, p. 3). Information fragmentation can also be a problem, as documents may be stored in several different hierarchies (e.g. hard drive file system, email client, etc.).

³ “The Character, Value, and Management of Personal Paper Archives.” *ACM Trans. Comput.-Hum. Interact.* 8, no. 2 (2001): 150-170.

⁴ “Personal Document Management Strategies,” 2009, <http://sarahhenderson.info/2009/personal-document-management-strategies/>.

⁵ “A Review of Organizational Structures of Personal Information Management.” *Journal of Digital Information* 9, no. 1 (May 19, 2008).

A *flat* structure uses tags to label items. Tagging is in some ways more flexible than a hierarchical system, as multiple tags can be assigned to one item (p. 4). Tags also lend themselves more to being shared – as is evident from many Web 2.0 applications (p. 5). However, there are drawbacks: tags are often assigned inconsistently, meaning that not all the relevant information will be retrieved (p. 5). Users are also often reluctant to annotate their collections extensively: when the collection is small, they feel it is unnecessary, but by the time it has grown, the job has become forbidding in scope. Tools to make manual tagging easier, or to automate the process, may help here (p. 5). A flat structure which makes use of tagging and keywords is the type of system favoured by Merrill⁶ (Merrill, p. 77).

Linear structures involve arranging items in an ordered list based on one attribute – alphabetically or chronologically, for example. Chronological ordering has something to recommend it: people can often recall roughly when they (for example) received a message; the ordering can be automated, reducing effort; and this sort of structure may help maintain contextual information (Indratmo and Vassileva, p. 6). However, relying heavily on the user's memory of one attribute does not always facilitate easy retrieval. Additionally, the linear structure means one can only search by one attribute at a time, and does not capture relationships between items in the way a hierarchical system does (pp. 5-6). There have been some attempts to develop PIM tools using a linear structure, but these have not yet achieved wide usage.

Spatial approaches use locations to organize items – different classes of items are kept in different places. In a digital context, this might involve arranging items on the computer desktop. This keeps items visible and easily accessible, but the approach only works for small collections – when the desktop becomes too cluttered, efficiency drops. Again, there have been attempts to develop a PIM tool that overcome this problem (by allowing users to see past states of the desktop), but this is not widely used (pp. 7-8). Merrill suggests a variant on this sort of system as a tool for solving a problem or understanding something challenging: he'll sometimes stick large Post-it Easel Pad notes all over his living room wall, then move them around into different arrangements to help him make connections between the information (Merrill, p. 113). But this is primarily a working tool rather than a storage system.

Network systems allow users to create links between items, without imposing any structural constraints on what can be linked. This method is flexible, and allows different types of information to be linked (e.g. documents and email messages), but it's harder to get an overview of one's whole collection, or to navigate through it successfully. Broken links can also be a problem, if information items have been moved. Semantic Web technology and RDF-based systems are examples of networks (as indeed is the Web as a whole). If meaningful links are to be created on a large scale, consistent terminology must be used (so that, for example, a single entity does not appear multiple times in the network under different names), which requires the use of fixed vocabularies and ontologies (Indratmo and Vassileva, p. 8-9). (The article does not discuss single-owner networks in detail, but it seems the terminology problem would be less acute here.)

Indratmo and Vassileva conclude that no single system emerges as a clear winner: some are better suited to certain types of material, working practice, or people than others, but none is obviously

⁶ *Getting Organized in the Google Era*. Broadway Books (2010).

better overall. They suggest that attention needs to be focused on augmenting or improving existing structures, rather than replacing one system with another (p. 10).

Civan *et al.* conducted a study⁷ examining the pros and cons of hierarchical file structures versus tag-based systems. Like Indratmo and Vassileva, they found that each had advantages and disadvantages, but there was no clear winner (Civan *et al.* pp. 10-1): preference for one over the other depended largely on the nature of the task in hand and the user's working style. For example, tags allow items to be put in multiple categories, but location-based searching (that is, hunting systematically through groups of files: see ***Finding Information***, below) can be easier with folders (with tagged material, this type of search often results in the same items appearing multiple times). There were some features of each system which some participants liked and others disliked – for example, folders hide material from the user's immediate view, which means that clutter is reduced, but also that items are more likely to be forgotten (pp. 9-10).

Perry's article "A Plea for the Horizontally Organized"⁸ is a more light-hearted look at the topic, but makes a serious point. He suggests people can be divided into the *vertically* and the *horizontally organized*: the former make extensive use of files and tend to put things away when they're not working on them, while the latter like to see everything spread out in front of them, and only file materials they aren't likely to consult again (though Perry doesn't use the terminology, his key point here seems to be that the reminding function of documents (see ***Aims and Purposes of Organization***, below) is extremely important to the horizontally organized). Office workspaces are generally designed to cater for the vertically organized, and as a result, horizontal organizers tend to have messy desks. He suggests that the solution to this is an alternative type of workspace designed for the horizontally organized (though his suggested model – a fifteen foot lazy susan – is unlikely to be a practical option for many people!).

Aims and Purposes of Organization

Ease of retrieval is clearly a key aim: however, desktop organization is not solely about being able to find material again. Malone observed that piles in particular often have a second purpose: to remind people of tasks that need to be done (p. 106). Barreau and Nardi⁹ report similar behaviour with electronic files – users arrange files on the computer desktop or elsewhere to remind themselves to deal with them (p. 41) Barreau and Nardi also suggest that these two functions (finding and reminding) need to be taken together – attempts to separate them and focus on one or the other will simply result in a system that doesn't do what people want it to (pp. 42-3).

Jones *et al.*¹⁰ conducted a study on the use people make of folder hierarchies in organizing projects. They concluded that folders serve a purpose which goes beyond simply aiding information retrieval.

⁷ "Better to Organize Personal Information by Folders Or by Tags?: The Devil is in the Details." *68th Annual Meeting of the American Society for Information Science and Technology* (2008).

⁸ On his website *Structured Procrastination*,
<http://www.structuredprocrastination.com/light/organization.php>.

⁹ "Finding and Reminding: File Organization from the Desktop." *SIGCHI Bull.* 27, no. 3 (1995): 39-43.

¹⁰ "Don't Take My Folders Away! Organizing Personal Information to Get Things Done", 2004,
<https://digital.lib.washington.edu/dspace/handle/1773/2031>.

“Folders, if only crudely, summarize as well as organize – they represent an emerging understanding of the associated information items and their various relationships to one another” (p. 2). The folder structure may emerge gradually over time, as files relating to a particular aspect of the project are grouped together. Williams *et al.*¹¹ confirmed this finding.

Finding Information

Barreau and Nardi’s 1995 study found that when searching for documents on a computer, most people favour location-based finding (going to the most likely folder and scanning the list of files) over logical finding (using a search function). Searching (for the filename or text contained in it) tends to be used as a fallback – that is, if location-based searching fails (pp. 40-1). In 1996, Fertig *et al.*¹² argued that this behaviour resulted not from an active preference for this way of finding material, but from the fact that it was the least bad option available to users (p. 66). However, although they suggest a number of innovations that would provide alternative means for retrieving material, only one of these (improved search functions) has become mainstream in the fourteen years since the article was written, and location-based searching still appears to be the norm. This seems to provide some support for Barreau and Nardi’s position. A study reported in Jones’s 2007 work¹³ indicated that over 90% of respondents viewed search as a last resort, and that most still prefer location-based finding (p. 470). (The Sudamih Project interviews¹⁴ also suggest that while people do now use the search function more than they used to, they do so only when location-based finding fails them.)

Despite this, Merrill argues that changing this behaviour (that is, making searching the default option rather than a last resort) is the foundation of effective organization in the information age (p. 77): filing and then hunting for information is time consuming, and he suggests that we will be able to operate more efficiently if we instead invest our time in making information searchable (e.g. adding tags or keywords, storing as much as possible in one searchable database) and brushing up our search techniques.

Types of Information

Barreau and Nardi identified three main categories of information: ephemeral, working, and archived (pp. 41-2). The first of these has a short shelf life, and includes to do lists and memos. Managing large quantities of this can prove problematic, as it is not easy to file. Some ephemeral information hangs around for longer than expected (perhaps because colleagues fail to reply promptly), and this intensifies the problem.

¹¹ “The Personal Curation of Digital Objects: A Lifecycle Approach.” *Aslib Proceedings* 61, no. 4 (2009): 340-363.

¹² Fertig, Scott, Eric Freeman, and David Gelernter. “Finding and Reminding’ Reconsidered.” *SIGCHI Bull.* 28, no. 1 (1996): 66-69.

¹³ “Personal Information Management.” *Annual Review of Information Science and Technology* 41, no. 1 (2007): 453-504.

¹⁴ See the Sudamih Project Researcher Requirement Report, available from

<http://sudamih.oucs.ox.ac.uk/docs/Sudamih%20Researcher%20Requirements%20Report.pdf>.

Working information is used frequently and has a shelf life of weeks or months. Users generally have no problem locating this, as they use it regularly and can remember where things are. Difficulties can, however, arise, towards the end of a project when items are used less frequently: the categories into which the material is arranged then becomes more important as a means of locating things.

Archived information is rarely accessed, but has a shelf life of months or years. Deciding what to archive and where is often difficult, but users generally found that there was little point in creating elaborate systems for archiving, as the benefits were not worth the time needed.

Whittaker and Hirschberg found that approximately half of people's personal archives were copies of publicly available documents. People kept these for a variety of reasons: ease of access, as a reminder to do something (e.g. read a particular paper), distrust of external sources, and sentiment (p. 157).

Storage Behaviour

The Whittaker and Hirschberg study looked at the nature of the paper archives of a group including a substantial proportion of researchers. They found that it was rare for people to take time out to sort through and rationalize their papers (p. 155), and even when motivated by an office move to do this, people kept a high proportion of the material (on average 78%), and found it hard to discard things (pp. 153-4). Material often accumulates in the first place because people are uncertain whether it will be useful, or because they don't have time to process it properly (p. 155). Less experienced workers accumulated material more rapidly than the more experienced, perhaps because those with more experience are better at identifying what they need to keep, or perhaps because they simply run out of room (p. 154).

Regarding digital archives, Williams *et al.* note that when storage space is plentiful, there is a tendency to accumulate material unintentionally, because this requires less effort than going through items and deciding what to keep and what to discard.

Marshall's book chapter¹⁵ highlights issues raised by storing information over time. We can accumulate a lot of material very quickly, which makes it hard to find what we want, even assuming we can remember what we have. It's often hard to predict what will be valuable in the future. Material is often highly distributed – stored in many different places and formats. Some of these formats will eventually become obsolete, and some will be highly application-specific – for example, it may not be easy to access emails outside the program through which they were originally received or sent. People are usually aware of the need to back up or take steps to preserve their material, but in practice, often don't actually do this: proper curation takes considerable time and effort. Application-specific formats pose a particular problem, as users may not know how to save the material in a more accessible form. Williams *et al.* also point out that people are sometimes hazy about what's stored on their own computer, and what's held on a server.

¹⁵ "How People Manage Information Over a Lifetime," in *Personal Information Management* (Jones and Teevan, eds.), University of Washington Press, Seattle, Washington, 2007.

Summary and Conclusions

- No organizational system is perfect for everyone: it's important to consider one's own needs and preferences, and pick the system (or combination of systems) which offers the best match
 - The most important feature of a system is that it does what the user needs and wants it to do, not that it fits some preconceived idea of what a good system should look like
 - A system which looks messy to an outsider may in fact be functioning perfectly well. (And a system which appears neat may not be serving the purpose it was designed for.)
 - Similarly, keeping an empty inbox or a clear desk may be a useful strategy for some, but not for all
 - Consequently, there's a need for honest self-evaluation
- Constructing and maintaining an organizational system takes a significant amount of time and effort – it is 'a heavyweight cognitive activity'¹⁶
 - This is a particular problem for hierarchical systems. Uncertainty about where best to file something often results in large piles of unfiled items
 - Many people find tagging easier than hierarchical filing
 - However, maintaining a comprehensive and consistent collection of tags may not ultimately be much less work
- The way in which information is organized often serves three distinct purposes
 - To enable easy retrieval of information
 - To remind the user that certain tasks need to be done
 - To represent the user's understanding of information items and how they relate to each other
- Some types of system are better at fulfilling some of these purposes than others. Users will naturally have different priorities, but ignoring the second and third and focusing solely on the first may result in a system that doesn't do everything that's needed
- There's generally a trade-off between ease of filing and ease of finding
 - The best system for an individual is the one that gives the best balance of these for his/her needs
 - Different types of system may work better for different projects
 - Using the right software or equipment can significantly reduce the amount of work involved
 - If a system is too complex or demanding to be easily maintained, it's likely to fail: the user simply won't be able to keep on top of the flow of incoming information

¹⁶ Indratmo and Vassileva, *op. cit.*, p. 3. The quotation refers specifically to hierarchical structures, but seems to be true to a greater or lesser degree of all organizational systems.

- For rarely used material, it may not be worth spending a lot of time implementing a complicated system, as the time ultimately saved won't be more than the time invested
- Information fragmentation is often a significant problem: information relating to the same topic may be stored on multiple computers, in multiple formats (perhaps relating to different applications), saved to the hard drive or attached to emails, and in hard copy or digitally
 - Users can sometimes be hazy regarding what's stored on their own computer and what's saved elsewhere – stored remotely on a server, for example
- Information may be ephemeral (intended only for very short-term use), working (used regularly over a period of days, weeks, or months), or archived (stored and accessed only rarely)
 - Most users have a combination of these, and deal with them in different ways
 - Difficulties are most likely to arise when material makes a transition between categories
 - If ephemeral items are not dealt with as quickly as expected, it can be hard to know how and where best to store these
 - When a project ends or moves onto another phase, working information is often archived: if it's not filed appropriately at that time, it may be hard to locate later
- A majority of users still rely on location-based finding rather than searching as their primary method of locating material
 - For digital material, however, making good use of the search function is becoming increasingly important for effective information management

Bibliography

Barreau, Deborah, and Bonnie A. Nardi. "Finding and Reminding: File Organization from the Desktop." *SIGCHI Bull.* 27, no. 3 (1995): 39-43.

<http://portal.acm.org/citation.cfm?id=221307&dl=GUIDE&coll=GUIDE&CFID=97795080&CFTOKEN=56870801>.

Civan, Andrea, William Jones, Klasnja Predrag, and Harry Bruce. "Better to Organize Personal Information by Folders Or by Tags?: The Devil is in the Details." *68th Annual Meeting of the American Society for Information Science and Technology* (2008).

<http://kftf.ischool.washington.edu/rerecentpublicationaboutplanner/Civan%20et%20al,%202008,%20Better%20to%20organize%20bv%20folders%20or%20by%20tags.pdf>

Fertig, Scott, Eric Freeman, and David Gelernter. "'Finding and Reminding' Reconsidered." *SIGCHI Bull.* 28, no. 1 (1996): 66-69.

<http://portal.acm.org/citation.cfm?id=249187&dl=GUIDE&coll=GUIDE&CFID=97795080&CFTOKEN=56870801>.

Henderson, Sarah. "Personal Document Management Strategies," (2009).
<http://sarahhenderson.info/2009/personal-document-management-strategies/>.

Indratmo, J., and Julita Vassileva. "A Review of Organizational Structures of Personal Information Management." *Journal of Digital Information* 9, no. 1 (May 19, 2008).
<http://journals.tdl.org/jodi/article/viewArticle/251>.

Jones, William. "Personal Information Management." *Annual Review of Information Science and Technology* 41, no. 1 (2007): 453-504.
http://www3.interscience.wiley.com/cgi-bin/fulltext/121477218/main.html,ftx_abs.

Jones, William, Ammy Jiranida Phuwanartnurak, Rajdeep Gill, and Harry Bruce. "Don't Take My Folders Away! Organizing Personal Information to Get Things Done." University of Washington Research Works Archive (2004).
<https://digital.lib.washington.edu/dspace/handle/1773/2031>.

Malone, Thomas W. "How Do People Organize Their Desks?: Implications for the Design of Office Information Systems." *ACM Trans. Inf. Syst.* 1, no. 1 (1983): 99-112.
<http://portal.acm.org/citation.cfm?id=357430&dl=GUIDE&coll=GUIDE&CFID=97795080&CFTOKEN=56870801>.

Marshall, Catherine C. "How People Manage Information Over a Lifetime," in *Personal Information Management* (Jones and Teevan, eds.), University of Washington Press, Seattle, Washington, 2007. Also available via her website.
<http://www.csdl.tamu.edu/~marshall/PIM%20Chapter-Marshall.pdf>

Merrill, Douglas C. *Getting Organized in the Google Era*. Broadway Books (2010).

Perry, John. "A Plea for the Horizontally Organized," on his website *Structured Procrastination*,
<http://www.structuredprocrastination.com/light/organization.php>.

Whittaker, Steve, and Julia Hirschberg. "The Character, Value, and Management of Personal Paper Archives." *ACM Trans. Comput.-Hum. Interact.* 8, no. 2 (2001): 150-170.
<http://portal.acm.org/citation.cfm?doid=376929.376932>.

Williams, Peter, Katrina Dean, Ian Rowlands, and Jeremy Leighton John. "Digital Lives: Report of Interviews with the Creators of Personal Digital Collections." *Ariadne* 55 (April 2008).
<http://www.ariadne.ac.uk/issue55/williams-et-al/>.

Wilson, James A. J., and Meriel Patrick. "Sudamih Research Requirements Report" (2010).
<http://sudamih.oucs.ox.ac.uk/docs/Sudamih%20Researcher%20Requirements%20Report.pdf>