

# How Do People Re-find Files, Emails and Web Pages?

William Jones<sup>1</sup>, Abe Wenning<sup>2</sup> and Harry Bruce<sup>1</sup>

<sup>1</sup> the Information School, University of Washington

<sup>2</sup> Avanade

## Abstract

Recent research suggests that a preference for navigation by folder to re-find files endures notwithstanding dramatic improvements in support for search as an alternate method of return. This paper describes a study that confirms this finding for files but that observes a distinctly different pattern of preference for re-finding email messages. After a delay of 2 to 4 weeks, search was the most common first choice for the return to email messages. A third, compound method was predominant for the return to Web information: The use of character-by-character “auto-complete” search was frequently followed by a hyperlink navigation to reach a targeted web page. Results point to the need for an integrated support of search and navigation methods during re-finding attempts. Results also suggest that support for re-finding begins with support for the initial “keeping” of information. Finally, results affirm two basic tenets of personal information management (PIM): 1. The need to consider multiple forms of information. 2. The need to consider a PIM activity such as re-finding within a larger context that includes other activities of PIM and considers the life cycle of personal information.

**Keywords:** personal information management, PIM, information behavior, re-finding, personal search

**Citation:** Jones, W., Wenning, A., & Bruce, H. (2014). How Do People Re-find Files, Emails and Web Pages? In *iConference 2014 Proceedings* (p. 552-564). doi:10.9776/14136

**Copyright:** Copyright is held by the authors.

**Acknowledgements:** This material is based upon work supported by the National Science Foundation (# 0534386).

**Contact:** williamj@uw.edu, agbram@gmail.com, harryb@uw.edu

## 1 Introduction

Re-finding – the return to information previously accessed – is both a common everyday activity and an important area of research in the field of Personal Information Management (PIM) (Capra & Pérez-Quñones, 2005; Elsweler, Baillie, & Ruthven, 2008).

People may most remember their protracted, time-consuming efforts to re-find or their outright failures. Less noteworthy but perhaps in aggregate even more costly are the routine steps taken with each incidence of re-finding. Re-find the email message you are meaning to answer. Re-find the several more you may need to reference in order to answer accurately. Re-find the Web page with information about an upcoming social event. Re-find the document you send people with parking instructions for a visit to your office.. Re-find the pictures you took last weekend. We may be reminded of Licklider’s observation in his highly influential article, “Man-Computer Symbiosis,” concerning his own work day:

About 85 per cent of my ‘thinking’ time was spent getting into a position to think, to make a decision, to learn something I needed to know . . . my choices of what to attempt and what not to attempt were determined to an embarrassingly great extent by considerations of clerical feasibility, not intellectual capability (Licklider, 1960, p. 4)

When we reckon with the many small acts of re-finding in a typical day, a few steps or seconds saved with each successful act of re-finding could really add up. It is, therefore, important to understand better the methods by which people choose to return to information currently as a step towards understanding how

we might improve upon these methods of re-finding. We aim for circumstances of PIM in which the “clerical tax” we pay is at a minimum.<sup>1</sup>

Consider the re-finding of files. A time-honored method of returning to documents on the “desktop” (i.e., on a person’s computer) is to proceed through a series of steps: navigate through a folder hierarchy to the folder where the file is thought to be located and then scan the folder contents to recognize the sought-for file. This method will be referred to throughout the current article as *folder-based navigation*. “Navigation”<sup>2</sup> is to emphasize the central importance of recognition as opposed to recall (Lansdale, 1988). “Folder-based” is to distinguish this kind of navigation from several others (navigation through the hyperlinks on a web page, for example). Folders and the folder hierarchy play an essential role in this particular variation of navigation.

Navigation, in its many variations, stands in contrast to a more direct method, commonly called “search”, wherein people type a few keywords in the hopes of recognizing the sought-for file in a list of matches that are returned. A search that bypasses layers of hierarchy would seem to be potentially much faster than folder-based navigation and so a preferred choice of re-finding method. To the contrary, Barreau and Nardi (1995) found that people overwhelmingly preferred navigation to search as a method of re-finding files.

The years since their studies (done in 1993 and 1994) have seen dramatic improvements in desktop search. And yet, a preference for navigation as a re-finding method apparently endures. Boardman & Sasse (2004) report that people expressed a preference for navigation to search although the strength and nature of this preference varied depending upon the form of information involved. Folder-based navigation was frequently mentioned as a preferred method of return to files – supplemented, on occasion, by a sort and scan of files within a targeted folder. Folder-based navigation was less commonly used for email and search was mentioned more frequently though still as a last resort after other methods of return – scanning and sorting of the inbox – had been tried first. Preferences for the return to Web pages were less clear with participants reporting the use of both bookmarks and direct search through a search service.

More recently, Bergman, Beyth-Marom, Nachmias, Gradovitch & Whittaker (2008) report that people still prefer folder-based navigation as a means of return to files. Search was a last-resort after other methods of return had failed. A strong preference for folder-based navigation re-asserted itself even for study participants who were introduced to better search support (Google Desktop) during the course of the study. After reporting an initial spike in the use of search, participants reported that their preferences settled back on folder-based navigation as the primary method of return. Barreau (2008) also reports a persistent preference for navigation (or browsing) as a means of return. In a related finding, Bergman, Gradovitch, Bar-Ilan & Beyth-Marom (2013) report a strong preference for the use of folders over tags as the primary means of organizing files and emails.

On the other hand, with now standard availability of fast, index-supported search for email and increasing allowances for email storage, there appears to be some tendency away from the use of folders for organizing email and a tendency, instead, to leave emails, read and un-read, in the inbox (Whittaker, Matthews, Cerruti, Badenes, & Tang, 2011).

The study described in this article complements previous research in two respects:

---

<sup>1</sup> See Jones (2013, Chapter 9) for a comparison of *clerical tax rates* as these have changed over time and are likely to change in the future. For samples of “The Future of Personal Information Management”, Parts 1 and 2 (and soon Part 3) visit <http://keepingfoundthingsfound.com>.

<sup>2</sup> The term “browsing” is also commonly used in the literature. The term “navigation” is used in this article to emphasize the purposeful nature of the re-finding activity. The person is looking for a *specific* item of information.

1. Participants were observed for their actual re-finding behavior as they completed a delayed cued recall task. By contrast, data from the Barreau and Nardi, Boardman and Sasse, and Bergman et al. studies come from self-reports<sup>3</sup>.
2. In the spirit of cross-form research established by Boardman & Sasse, participants were observed as they returned not only to files but also to email messages and to web pages. How does the observed method of re-finding vary depending upon the kind of information to be re-found?

## 2 Method

*Note: The re-finding study described here was part of a much larger longitudinal investigation, conducted in 2009, into the ways that people manage information over time. Qualitative results relating to this larger investigation have been described elsewhere (Bruce, Wenning, Jones, E., Vinson, & Jones, W. 2010). Results presented here are from a more focused, quantitative analysis of this data.*

Seventeen participants (nine male, eight female, ranging in age from 18 to 49) were recruited through a Craigslist announcement and through flyers placed around the University of Washington's main campus. Five participants were students at the university (two graduate students); also participating were three teachers (high school), two system administrators, three software developers, a journalist, a network technician, a systems analyst and a librarian.

The procedure followed is similar to that used previously to study the re-finding of Web pages (Bruce, Jones, & Dumais, 2004). Participants completed a set-up session followed by a test session, two to four weeks later. Each session lasted about an hour. The procedure involved personal information – files, emails (received) and web pages actually viewed by the participant.

Given the potentially sensitive nature of this information, special steps were taken to insure that participants had ultimate control over the materials used in the sessions. Participants, under observer direction, generated three different lists – one for files, one for email and one for web pages – in an ordering that was counter-balanced across participants<sup>4</sup>. Each list was sorted by date (i.e., received date or last accessed date) so that more recent items were topmost.

The attempt was also made, for each list, to create a sampling of test items that was distributed, roughly, over the previous 7-day period. This meant sampling from a list beginning with items viewed “Today” until two acceptable test items were selected. Items were then sampled from “Yesterday” until two test items were selected and so on until 14 items test items had been selected for each form of information. (If the next day for sampling had already been reached because of “skips”, as described below, then participants continued from their current position in the list.)

To select items in a list within a given time period, participants worked through items one by one. The observer did not see the items during this stage of the set-up session. Participants were instructed to bypass items that, for whatever reason, they preferred not to include in the study. For items not skipped, participants rated the likelihood that they would want to re-find this item again over the next twelve months. For items where the likelihood was rated as 75% or higher, participants were then asked to briefly describe a reason for re-finding the item. Participants were encouraged to be as specific as possible but without referencing the item by “name” (e.g., file name, domain name, sender name or subject tagline).<sup>5</sup> Participants were not told that they would later be tested on these items.

---

<sup>3</sup> Though Bergman et al.(2008) also provide corroborating evidence from logged data for Linux users.

<sup>4</sup> For email messages, participants were instructed to simply work through the inbox of their primary email account. For web pages, participants viewed a listing generated by the history facility of their primary browser. For files, participants were instructed in the steps to generate a search of files viewed in the past seven days, ordered from most to least recently viewed.

<sup>5</sup> Examples of reasons included “to use as reference for what wife wants for birthday next year”, “want to change prescription to mail order; need to use form attached to email”, “use receipt when filing income taxes next year”.

Finally, with the participant’s permission, the item was opened in its own window and a screenshot was taken of the item.

The test session involved selected testing of five items randomly chosen from each list (file, email, web page). Items from a given list were tested in a block and the ordering of blocks (for file, email, web page) was counter-balanced across participants.

For a given item, participants were given the “reason for return” which they had provided during the set-up session. With reference to a reason, participants were instructed to imagine a situation “now” where the item was needed<sup>6</sup>. Participants were then asked to re-find this item as quickly as possible by whatever method or combination of methods they chose. Participants were instructed to think-aloud as they did so.<sup>7</sup>

Participants were timed and their method(s) of re-finding were noted. Trials exceeding five minutes were stopped and marked as a failure to re-find. All methods used on a trial were recorded and independently cross-checked later with reference to an audio recording of the test session.

Participants worked from their own computer in each session. Operating systems represented on participant computers included Windows XP, Windows Vista, and the Macintosh OS X (seven, five and five participants, respectively). For participants running under XP, five used MS Outlook as their primary interface to email and two used Thunderbird. For participants running under Vista, two used Outlook as their primary interface to email; two used Gmail and one used the AOL web client. For participants running under the Mac OS, two used Apple Mail as their primary interface to email, two used Gmail and one participant used Yahoo Mail.

Four XP participants used the Microsoft Internet Explorer as their primary web browser; the remaining three XP participants used Mozilla FireFox. For Vista users, two used IE as their primary web browser, one used FireFox and two used both IE and FireFox. For Mac OS X users, four used FireFox and one used the Apple Safari browser.

### 3 Results

As shown in Table 1, the rate of successful return was high across conditions (93%, 88%, 84%, respectively, for files, email and web pages). An item was commonly found in three minutes or less. Moreover, success was likely to come with the first attempt. Results are consistent with high rates of successful re-finding reported by Boardman and Sasse.

	Files	Email	Web Pages
Eventual success	93%	88%	84%
Re-found in 3 minutes or less	90%	83%	81%
Success with first method of re-finding	78%	57%	83%

Table 1: Success rates and speed of return were high across information forms.

Across trials and forms of information, participants made an average of 1.48 attempts before successfully re-finding a targeted item.

<sup>6</sup> On those occasions where a participant had no recollection of an item given the reason for return, they were instructed to “give it a shot” i.e. by imagining some item they had previously encountered that would fit the “reason for return” and then to return to this item.

<sup>7</sup> Near the outset of the session, participants were given practice in thinking aloud by doing so as they created a file (e.g., MS Word document), typed “Hello world!” and saved the file in a folder of their choice. Participants were told only to describe only the “what” of their current thoughts and actions, not the “why”. The observer provided gentle reminders if the participant deviated from the instructions or stopped talking for more than 15 to 20 seconds.

In some cases, however, participants made several attempts and tried several different methods before either locating the targeted item or “timing out”. A look at some of these instances of multiple attempts is instructive:

- Participant SX-215, looking for an email, tries search, then search again, then navigation, then sorting (success):  

“I'm in Outlook, I'm gonna go into "Look for" and I'm gonna put in, uhh, "currency" and click "Find now." I: So the "Look for" is a search bar? R: Yeah. And that didn't find it, so I click "Clear" and then I'm gonna put in "Money" and see if that finds it. And that didn't find it. So then I'm gonnaing (?) to wonder if I put it in my Money folder. I didn't think I had, so I'm gonna look there. I: So you didn't remember that you had the Money folder? R: No, I have it. I just didn't think that I had, uhh, filed it away yet. And I definitely haven't filed it away yet, so I'm going back to the inbox, and then what the heck was that guy's name? I think this is an email that I sent to myself, so I'm gonna sort it by sender and scroll down to my name, where there's probably twenty five emails. And there it is. Foreign Exchange.”
- Participant JS-197, looking for a file, tries navigation four separate times (success).
- Participant AP-123, looking for an email, tries sort (by “From” field), then navigation through email folders, then search (on person’s name), then search again on the word “schedule” (unsuccessful).
- Participant BK-129, looking for a web page, looks first in a “Favorites” drop-down list, clicks one favorite link for a jump to “YouTube”, searches within this site twice, then searches from Google (successful).
- Participant ET-156, looking for an email, searches the Inbox for “house sitting”, then searches for a person’s name followed by a sort of results by “Received” date, followed by another search of the Inbox for “dinner” (successful).

The data summarized in Table 2 show a very large discrepancy between the three conditions with respect to the use of search. A planned paired t-test comparison between search as a first method for files vs. email was significant ( $t=4.41$ ,  $p < 0.001$ ,  $df=16$ ).<sup>8</sup> A second planned comparison between the use of search for files vs. web pages was not significant ( $t=0.18$ ).

Folder navigation dominates as a first method of return to files. If the use of “recent” lists is also classed as a kind of navigation then the dominance of navigation increases to nearly 90%. Folder navigation as a first method of return to email (i.e., involving the use of email folders) was much less prevalent. When navigation of the inbox is also included, a measure of total navigation increases to just over 42%. Even so, in a paired t-Test, this measure is significantly lower than total navigation for files ( $t=3.84$ ,  $p < .002$ ,  $df=16$ ).

---

<sup>8</sup> Obviously, given the small sample size (17 participants) the power of these tests is low. Since tests involve pairwise comparisons of data points collected for the same participants (i.e., the design is within-subjects), statistical significance is not, however, unheard of even in such a small sample size and, some might say, all the more “practically significant” when it is observed.

	Files	Email	Web Pages
Search	6.28% (12.74%)	42.35% (34.39%)	8.00% (10.53%)
Folder Navigation	83.73% (22.57%)	19.61% (29.65%)	27.22% (26.28%)
Total Navigation	89.42%	42.16%	43.45%
Other	4.30%	15.49%	48.55%

Table 2: Mean percentages (and standard deviations)<sup>9</sup> for methods used first to re-find files, emails and web pages.

Included in the “other” category for email is sorting which occurred 10.87% of the attempts for email. Sorting was most commonly by sender. Of the 10 instances (7 participants) where sorting was a first method of return to email, 9 were sorted by sender (the other was sorted by presence or absence of an attachment). In the two instances where sorting was used as a second method of return, both were by sender. Sorting might be regarded as a variation on search where users must decide which attribute to focus on and must have an approximate idea as to the value of the attribute (e.g., Sender name). If sorting is added, then the dominance of search is even higher for email – at 53.53% of the trials.

For web pages, folder navigation equates to the use of a bookmarking facility (through which people can save and organize web references), usually as a feature of the browser. Other recognition-heavy methods of return that might reasonably be classed as variations in navigation include hyperlink navigation (9.56%), the use of a history facility (5.56%), and the occasional happenstance that the target web page is already open in a browser window or tab (1.11%). But a total navigation measure that includes these variations is still significantly lower than the total navigation measure for files ( $t=3.92$ ,  $p < .01$ ,  $df=16$ ). At an observed 8%, search as a first method of re-finding a Web page is in line with a previously observed occurrence of 13% (Bruce et al., 2004).

The web pages “other” entry includes one very popular method of return to web pages – auto-complete. A person types a few characters in the address well and selects from a list of matching web addresses that is generated and revised incrementally as the person types. Auto-complete might be regarded as a variation of search – albeit, where the search space is highly constrained to include only web addresses for pages that a person has visited. However, in over 50% of its uses in the current study, auto-complete was used as part of what might be called a compound method of return: Participants used auto-complete to get to the right web site. Participants then used hyperlink navigation to click to the targeted page within the site.

Participant comments attest to the popularity of auto-complete:

**ET-156:** Umm I’m going to guess that auto complete is gonna find the site for me. And it did not. Umm so I’m gonna continue typing the URL from memory and hit enter.

**EW-191** - Umm, there's only a few sites that I go to regularly, and they, umm, just because they're visited often, if I start typing them in, they'll finish for themselves, so.

What effect did a participant’s operating system and its level of search support have on a participant’s overall tendency to use search as a first method of re-finding? A simple non-paired comparison between participants with limited search support (Windows XP) and participants with advanced search support (Windows Vista and Macintosh OS X) reveals an overall intendency to use search more when search support is more advanced ( $t=1.83$ ,  $p < .05$ ,  $df=15$ ).

<sup>9</sup> Note that since measurements themselves are percentages, the means and standard deviations are also expressed as percentages.

% Search	1 <sup>st</sup> attempt	2 <sup>nd</sup> attempt	3 <sup>rd</sup> attempt	4 <sup>th</sup> attempt
Files	6%	53%	42%	50%
Email	42%	62%	60%	67%
Web	8%	33%	25%	100%

Table 3: The likelihood of search with each attempt.

Table 3 suggests an increasing tendency to use search on subsequent re-finding attempts. How did a person’s tendencies to “file” and organize impact their choice of re-finding method? We address this question by following the terminology of Boardman and Sasse but with adjustments according to the information form involved:

### Total Filers

- *Files/Documents*. Organize all or almost all of their files immediately upon creation. Make extensive use of folder hierarchy. No uncategorized files anywhere with rare exceptions.
- *Emails*. All incoming email is organized in some way; inbox contains zero messages at end of most days or all messages are tagged in some way, with rare exceptions.
- *Web pages*. All web references are organized or tagged according to some scheme. Extensive hierarchy may be present for large collections.

### Extensive Filers

- *Files/Documents*. Organize extensively, but leave some items unfiled (allow file to be saved according to system default of desktop, My Documents, etc.). Extensive hierarchy of files. (see Figure 1).
- *Emails*. Try to organize many messages on a daily basis. Inbox generally contains to-do items or does not extend beyond one page.
- *Web pages*. Organize many bookmarks as they are created or at the end of a browsing session. Most web references are organized in some way. Folders used but may have a quarter to half of bookmarks uncategorized.

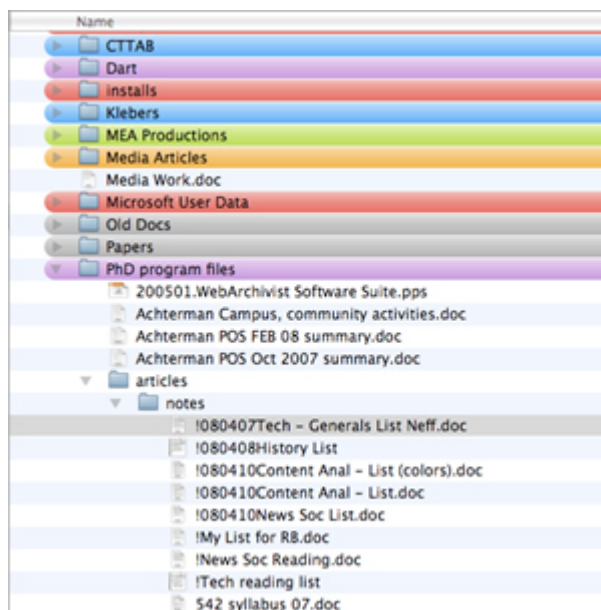


Figure 1. A portion of an “Extensive Filer’s” organization of files.

### Occasional Filers

- *Files/Documents*. Organize occasionally in spurts or during “spring cleaning”; leave most items unorganized and have relatively few folders. Rely heavily on system default location for saving electronic files.
- *Emails*. Organize only a few messages on a daily basis or do occasional “spring cleaning” to organize inbox/email messages. Inbox generally contains hundreds of uncategorized messages.
- *Web pages*. Organize bookmarks sporadically. Some organization of bookmarks exists but most are uncategorized. Bookmarks are only occasionally organized or deleted.

### No Filers

- *Files/Documents*. Do not organize any electronic files; all files are in one location such as the desktop or immediately under a default folder (e.g. “My Documents”). Rely on search or other method to find documents.
- *Emails*. Do not organize or tag any messages. No use of folders or labels. Inbox contains potentially hundreds or thousands of messages. Rely entirely on search for finding emails.
- *Web pages*. Don’t bookmark or otherwise keep references to web pages. Rely instead on return by Web search. If bookmarks are created, these are not organized.

	Files	Email	Web Pages
Total Filers	0	0	0
Extensive Filers	6	6	7
Occasional Filers	9	8	5
No Filers	0	3	2
Undetermined	2	0	3

Table 4: The number of participants classified in each of the "organizing tendency" categories. by information form.

	Files	Email	Web Pages
Higher tendency to organize	2.8% Search; 71.7% Navigation	13.9% Search; 71.7% Navigation	5.2% Search; 10.5% Navigation
Lower tendency to organize	8.2% Search; 90% Navigation	57.9% Search; 26.1% Navigation	4.8% Search; 7.6% Navigation

Table 5: The use of search and navigation as first methods appears to vary depending upon organizing tendency and is especially pronounced for email.<sup>10</sup>

Table 4 provides a breakdown of participants by “organizing tendency” and information form.

Table 5 provides a breakdown of search vs. navigation across different forms according to whether a person has a higher tendency to organize (Total Filers, Extensive Filers) or a lower tendency to organize (Occasional Filers, No Filers). The strongest apparent difference between higher and lower organizers is for email. Not surprisingly, perhaps, people who are more likely to organize are more likely to navigate.

<sup>10</sup> Note that in none of the cells of the table do percentages sum to 100% -- most notably not in the Web column. This is because so many of the methods of return observed resist classification as either “search” or “navigation”.

## 4 Discussion

Consistent with previous research, there is a strong preference for folder-based navigation as a first method to re-find a file. However, the picture changes when the target of a re-finding attempt is an email or a web page.

Better search support does matter. Participants are more likely to choose search under Vista and Mac OS X, where search support – notably the creation and seamless maintenance of an index that greatly speeds search – is built into the operating system. The greater inclination to search is seen across information forms but appears especially strong for email.

Even so, an overall preference for navigation persists that is especially strong for files. Why? We can consider two hypotheses.

A “*location is better*” hypothesis follows from a consideration of the affordances for location and a sense of place in both our physical and digital worlds. Bergman et al. (2008) note the power of the location metaphor. Teevan et al. (2004) point to a general importance for a sense of location.

Barreau and Nardi note a reminding value for location-based finding. Related to this is a value of serendipity: along the path to sought-for information, other information of relevance may be encountered that might otherwise be overlooked.

As Teevan et al. note, an “orienting” method of information access also enables a more stepwise progression towards desired information in which the expression of each step is relatively easy and, if a wrong step is taken, corrective backtracking is also easy.

There are related notions that recognition is generally easier than recall (Lansdale, 1988) and that the memory for how to access information is sometimes more “in our movements”, so to speak, as procedural knowledge rather than “in our words” as declarative knowledge (Bergman et al., 2008).

In some cases, we can’t know whether we have the right information item without the ability to inspect a larger context in which the item occurs (Teevan et al., 2004) – as we do routinely, for example, to locate the “correct” version of a document.

Finally, a sense of digital place with respect to a personal filing system is closely related to notions of familiarity, control and organization (Jones, Phuwanartnurak, Gill & Bruce, 2005). As people organize their digital information, this is most likely to be manifest in a personal file system (Boardman & Sasse, 2004). People express a sense of control, ownership and even pride (e.g., (Boardman & Sasse, 2004), p. 585) concerning their files and their organization.

At the same time, results demonstrate that people do search and, especially, as a method of return to email messages. How to account for this observation under the location-is-better hypothesis?

One rejoinder is to note that a constant stream of incoming email messages where the new automatically displace the old is profoundly disruptive for any sense of place. Other interfaces for email might induce a greater sense of place (S. Whittaker, Bellotti, & Gwizdka, 2007). And, in a slightly broader context, we can see the use of filtering techniques (that automatically sort email into different folders), blogs, the “Wall” of Facebook and even the use of multiple email accounts (one for business, one for friends and family, one for placing orders...) as variations on an attempt to impose a greater sense of location on email.

Notwithstanding the many useful features of location and the special challenges that “email overload” poses for any supporting tool – location-based or otherwise – we should not overlook a simpler explanation of results offered by a “*first-impressions*” hypothesis. The first-impressions hypothesis is essentially an alternate expression for the robust and ubiquitous *primacy effect* repeatedly observed in studies of cognitive psychology (Neisser, 1967).

Under this hypothesis, the method of re-finding follows from previously successful retrievals of the information and, ultimately, from an initial encounter when the information is created or otherwise experienced for the first time. People re-find files by navigating through folders because, in a traditional

filing system, people often specify the folder location of a file to begin with (even if this is to the desktop or a top-level such as “My Documents”). By the same reasoning, people pay special attention to properties of an email message such as sender or subject because these properties are used initially to screen and prioritize email messages. These same properties are later likely to be recalled during a retrieval attempt (Elsweiler et al., 2008).

People are observed to create email folders for email sent from specific people or pertaining to specific topics (Boardman & Sasse, 2004). But, when fast search (and other features such as grouping by conversational thread) support a comparable grouping by remembered attributes then why bother with folders? Why bother, in particular, when the flood of incoming email messages continues to increase? Both hypotheses have their validity. Both carry implications for system design.

Consider, for example, that the debate over the merits and eventual dominance of search has focused largely on the act of re-finding (Barreau & Nardi, 1995; Fertig, Freeman, & Gelernter, 1996; Nardi & Barreau, 1997).

But, in the spirit of the life-cycle emphasis of PIM (Jones, 2007), the first-impressions hypothesis prompts a shift backwards in time to the initial encounter with an information item and to the decisions made and actions taken during an initial keeping stage of PIM. Support for the tagging and annotation of information items remains basic and fragmented. Perhaps, with unified support for tagging and the application of searchable annotations, we can finally realize a “placeless” means of keeping (Dourish, Edwards, LaMarca, & Salisbury, 1999) that readily couples with a predominantly search-based means of re-finding.<sup>11</sup>

On the other hand, following a location-is-better hypothesis, we might continue to explore ways to situate email – especially in the context of a larger effort to manage tasks and projects (Jones, Hou, Sethanandha, Bi & Gemmell, 2010).

Suppose there is a best of both worlds? Suppose that these hypotheses aren’t really in opposition to each other? Consider, for example, the folder hierarchy that provides a basis for the placement of and subsequent navigation back to information items. This same hierarchy might be “flipped” so that folders become tags or labels to be applied to the information item (file, email, web page) currently in view. Our attention stays (more or less) on this item even as we label or “place” this item for anticipated future uses..

Or consider the popularity of direct entry (“auto-complete”) as a method of choice for the return to web information. On the one hand, this method can be regarded as a basic search (i.e., search focused on the text in addresses for Web pages previously visited). In some cases, the targeted page is directly returned through selection of a matching address. But in many other cases, auto-complete is only the first portion of a compound re-finding method: Search is a way to “parachute”, we might say, into the right region of a Web space. The person then navigates (or possibly searches locally) to reach the targeted item. In real situations of information access – web-based or not – the target may, in fact, not be a specific information item but rather a fuzzier region of information.

## 5 Conclusion

Consistent with other research, the current study observes a strong preference for folder-based navigation to re-find files. However, a different pattern is observed for other forms of information – most notably email, for which the inclination to use search as a first method is roughly on par with an aggregate measure that includes both inbox scans and folder-based navigation.

The data reported here were collected in 2009 at a time when local system-level (default), trouble-free (mostly) support for fast, index-supported search was present in the computing environments of some

---

<sup>11</sup> For example, Mac users have the option to add comments to files that can then be seen when browsing these files: <http://mactips.info/wp-content/uploads/2009/07/mt-comments-view.jpg> . These comments can also be searched using the search tools such as Spotlight. One participant in the study reported searching on comments as a means of return to files.

participants (e.g. those with Windows Vista and the Macintosh OS X) but not others (e.g. those with Windows XP). Participant computing environments, even in a relatively small sample of 17, differed in many other respects as well including, especially, with respect to email client (MS Outlook, Thunderbird, Gmail, AOL, Apple Mail, Yahoo mail) and web browser (IE, FireFox, Safari or, for some participants, a combination of these).

As of this paper's publication, our computing worlds are already considerably changed from the computing worlds of study participants. Had the data been collected more recently or, even, "now", we'd still find ourselves at a comparable remove from the data should we read this paper a few years from "now". If not a separation of time, then the diversity of computing environments might place us at comparable distance from the data, e.g.. "if I use or care about system/app/device X where things are done the `_right_` way, why should I bother with data gathered for users of system/app/device Y (where things are `_clearly_` done the `_wrong_` way)?"

What relevance does the data have to us? If indeed people still prefer to navigate to files even as they increasingly use search for email, so what? Maybe a preference for navigation is enduring. Or maybe we'll come gradually to use search as a first method of return to personal files. This might happen as personal (and personalized) search continues to improve. Or also as support for "search-facilitating" keeping methods (e.g. tagging) continues to improve. Or, indeed, as the sheer amount of what is "ours" continues to increase – as stored locally on our various devices and, increasingly, in the "cloud". But again, so what?

Consider that the value of the study described may be primarily in its choice of methods as these follow from two basic tenets of research in personal information management. First, it is important to consider a diversity of information forms. Results obtained for a single form such as files (or email or web pages) can mislead us in our attempts to generalize or to draw implications for design. Second, activities such as re-finding need to be considered in a larger lifecycle context that includes, for example, initial encounters with and organization of information that is later targeted in re-finding.

People exhibit a bewildering diversity of PIM behaviors<sup>12</sup>. Observed differences might be for any of several reasons: personality, experience, education, operation system, device, etc. Pity the poor academic researcher who lacks the resources to cover or control for the number of factors that `_might_` matter! At the same time, if we attempt to abstract away from this diversity we risk taking the "P" out of PIM.

But we begin to get a handle on diversity as observed variations in PIM behavior (an inclination to search, for example) occur *within-subjects*.. When the same participant exhibits different preferences in re-finding method depending upon information form, we can more confidently focus on a comparison of these forms, their supporting applications and the circumstances of creation and use for items of different forms. Why one way for files and another way for email? What implications can we draw for design?

The current study, like many other "field studies" in PIM, is exploratory in its efforts to identify and begin to understand the reasons for an observed diversity in the methods people use for PIM (and, specifically in this study, methods of re-finding). From these explorations we derive implications for the design of supporting tools and techniques of PIM. The study reaffirms a point made in other studies (as reviewed in the Introduction) that better support for re-finding should be "multi-flavored" in its support for techniques of both search and navigation. Moreover, support for re-finding may properly start much earlier in the lifecycle of information – with better support for the initial keeping of this information.

---

<sup>12</sup> Observed differences reported in this paper are for re-finding behavior. Comparable differences have been observed elsewhere in keeping behavior (e.g. Jones, Dumais & Bruce, 2002).

## 6 References

- Barreau, D. (2008). The persistence of behavior and form in the organization of personal information. *J. Am. Soc. Inf. Sci. Technol.*, 59(2), 307–317.
- Barreau, D., & Nardi, B. A. (1995). Finding and reminding: file organization from the desktop. *SIGCHI Bull.*, 27(3), 39–43.
- Bergman, O., Beyth-Marom, R., Nachmias, R., Gradovitch, N., & Whittaker, S. (2008). Improved search engines and navigation preference in personal information management. *ACM Trans. Inf. Syst.*, 26(4), 1–24.
- Bergman, O., Gradovitch, N., Bar-Ilan, J., & Beyth-Marom, R. (2013). Folder versus tag preference in personal information management. *Journal of the American Society for Information Science and Technology*, n/a–n/a. doi:10.1002/asi.22906
- Boardman, R., & Sasse, M. A. (2004). “Stuff goes into the computer and doesn’t come out” A cross-tool study of personal information management. In *ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2004)*. Vienna, Austria.
- Bruce, H., Jones, W., & Dumais, S. (2004). Re-finding information on the Web: What do people do and what do they need? In *ASIST 2004: Annual Meeting of the American Society for Information Science and Technology*. Providence, RI.
- Bruce, H., Wenning, A., Jones, E., Vinson, J., & Jones, W. (2010). Seeking an ideal solution to the management of personal information collections. In *Information Seeking in Context Conference - (ISIC) 2010*. Murcia, Spain.
- Capra, R., & Pérez-Quiñones, M. A. (2005). Using Web Search Engines to Find and Refind Information. *IEEE Computer*, 38(10), 36–42.
- Dourish, P., Edwards, W. K., LaMarca, A., & Salisbury, M. (1999). Using properties for uniform interaction in the Presto Document System. In *The 12th Annual ACM Symposium on User Interface Software and Technology (UIST’99)*. Asheville, NC.
- Elsweiler, D., Baillie, M., & Ruthven, I. (2008). Exploring memory in email refinding. *ACM Trans. Inf. Syst.*, 26(4), 1–36.
- Fertig, S., Freeman, E., & Gelernter, D. (1996). Finding and reminding reconsidered. *SIGCHI Bulletin*, 28(1), 7 pp. Retrieved from <http://www.acm.org/sigchi/bulletin/1996.1/fertig.html>
- Jones, W., Dumais, S., & Bruce, H. (2002). Once found, what then?: A study of “keeping” behaviors in the personal use of web information. Presented at the 65th Annual Meeting of the American Society for Information Science and Technology (ASIST 2002), Philadelphia, PA.
- Jones, W., Phuwanartnurak, A. J., Gill, R., & Bruce, H. (2005). Don’t take my folders away! Organizing personal information to get things done. In *ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2005)* (Vol. 2005, pp. 1505–1508). Portland, OR: ACM Press.
- Jones, W. (2007). *Keeping Found Things Found: The Study and Practice of Personal Information Management*. San Francisco, CA: Morgan Kaufmann Publishers.
- Jones, W., Hou, D., Sethanandha, B. D., Bi, S., & Gemmell, J. (2010). Planz to put our digital information in its place. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems* (pp. 2803–2812). Atlanta, Georgia, USA: ACM.
- Jones, William. (2013). Transforming Technologies to Manage Our Information: The Future of Personal Information Management, Part 2. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 5(4), 1–179. doi:10.2200/S00532ED1V01Y201308ICR028
- Lansdale, M. (1988). The psychology of personal information management. *Appl. Ergon.*, 19(1), 55–66.
- Licklider, J. C. R. (1960). Man-computer symbiosis. *IRE Transactions on Human Factors in Electronics, HFE-1*, 4–11.

- Nardi, B., & Barreau, D. K. (1997). "Finding and reminding" revisited: appropriate metaphors for file organization at the desktop. *SIGCHI Bulletin*, 29(1). Retrieved from <http://www.acm.org/sigchi/bulletin/1997.1/nardi.html>
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century Crofts.
- Teevan, J., Alvarado, C., Ackerman, M. S., & Karger, D. R. (2004). The perfect search engine Is not enough: A study of orienteering behavior in directed search. In *ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2004)* (pp. 415–422). Vienna, Austria.
- Whittaker, S., Bellotti, V. M. E., & Gwizdka, J. (2007). Email as PIM. In *Personal Information Management*. Seattle, WA: University of Washington Press.
- Whittaker, Steve, Matthews, T., Cerruti, J., Badenes, H., & Tang, J. (2011). Am I wasting my time organizing email?: a study of email refinding. In *PART 5 ----- Proceedings of the 2011 annual conference on Human factors in computing systems* (pp. 3449–3458). Vancouver, BC, Canada: ACM.

## 7 Table of Figures

Figure 1. A portion of an "Extensive Filer's" organization of files. ....558

## 8 Table of Tables

Table 1: Success rates and speed of return were high across information forms.....555

Table 2: Mean percentages (and standard deviations) for methods used first to re-find files, emails and web pages.....557

Table 3: The likelihood of search with each attempt.....558

Table 4: The number of participants classified in each of the "organizing tendency" categories. by information form. ....559

Table 5: The use of search and navigation as first methods appears to vary depending upon organizing tendency and is especially pronounced for email.....559