Finding and Reminding
File Organization from the Desktop

Deborah Barreau and Bonnie A. Nardi

This paper summarizes and synthesizes two independent studies of the ways users organize and find files on their computers. The first study (Barreau 1995) investigated information organization practices among users of DOS, Windows and OS/2. The second study (Nardi, Anderson and Erickson 1995), examined the finding and filing practices of Macintosh users. There were more similarities in the two studies than differences. Users in both studies (1) preferred location-based finding because of its crucial reminding function; (2) avoided elaborate filing schemes; (3) archived relatively little information; and (4) worked with three types of information: ephemeral, working and archived. A main difference between the study populations was that the Macintosh users used subdirectories to organize information and the DOS users did not.

Introduction

Though personal computers have become increasingly important in the workplace for the last fifteen years, studies of file organization in offices have focused on the organization and use of *paper* documents (Cole 1982; Malone 1983; Suchman and Wynn 1984; Lansdale 1988; Blomberg *et al.* 1994). What about electronic files? The authors of this report became interested in electronic filing practices at almost the same time, and each conducted her own independent study, unbeknownst to the other. Nardi discovered Barreau's work (mentioned in the SIGCHI Bulletin, October 1994) after she had completed her study. She contacted Barreau and suggested a collaboration. The differences in the two study populations presented a tantalizing opportunity for a comparative analysis: Barreau studied a mixed group of primarily DOS/Windows users who were not computer experts (Barreau 1995) and Nardi studied Macintosh users with considerable computer experience (Nardi, Anderson and Erickson 1995). This reports presents our anal-

ysis of the similarities and differences in filing practices across the two user populations.

But why should we be interested in electronic filing and finding practices? The first reason is that filing and finding are such basic aspects of working with computers that while we scarcely notice their existence – hence the lack of research – every computer user spends time and effort in filing and finding every time the computer is used. As designers we should be concerned with optimizing finding and filing. The second reason for this work is that with the greater connectivity made possible by networks and servers, we are all poised on the verge on having to manage and find vastly more information than we have been accustomed to. Or are we? This paper considers the conventional wisdom about the supposed coming information overload in light of the activities of the users in our studies.

File Organization in Electronic Environments: Two Studies

In the spring of 1993, Barreau conducted a study of seven managers to observe how they organized and retrieved information from their electronic workspace (Barreau 1995). Interviews with the managers were audiotaped and transcribed. The managers were also asked to provide a tour of their electronic directories. The goal of the research was to identify the types of documents used and to determine the factors affecting individual decisions to acquire, organize, maintain and retrieve information. Managers were selected because they supervise multiple and varied projects, requiring the organization and retrieval of a variety of information types. Five of the managers were employed in the same company, a large information management company. Four of these managers used DOS and one

worked on a Macintosh. Another study participant was a project manager in a research department at a government agency and used OS/2. The final manager was a research scientist at a major corporation providing services to the government. He used Windows 3.0, which was not unlike DOS in the limitations of an 8-character file name (plus a three character extension). The OS/2 and Windows users were experienced computer users; the other study participants were relatively unsophisticated users. All users were networked to servers. Exclusive of servers, the amount of disk storage available to these users varied considerably. The OS/2 and Windows users had more than a gigabyte of storage, while three of the DOS users had diskless workstations and stored their files on the servers or on diskettes. One DOS user with a 120 MB hard drive stored most of his files on servers, and the Macintosh user with 300 MB of personal disk space made heavy use of diskettes. At the time of the study diskettes were used by some because they did not know how to transfer files to one another using the LAN and some felt more secure with diskettes. (They have since learned to transfer files electronically.)

In the winter of 1994, Nardi conducted an interview study of fifteen Macintosh users, most Apple employees (Nardi, Anderson and Erickson 1995). Study participants included managers, graphic artists, programmers, administrative assistants and librarians. Users were asked to provide a tour of their systems, and a structured set of questions was asked in a conversational style to elicit information about jobs and tasks as well as approaches to organizing and finding files. The interviews were videotaped in users' offices or cubicles to allow close observation of the desktop and to watch users as the interviewer requested them to find files observed during the tour. The number of files and amount of storage users had varied from as many as 31,000 files and as much as 1500 megabytes of personal storage (plus servers additional) to as little as 2,400 files and 80 megabytes available. All users were networked to servers. In this study some users, such as the professional programmer, were extremely sophisticated computer users, and all were experienced users. Two users were new to Apple and the Macintosh at the time of the study (one formerly UNIX, one DOS). This study group may represent a sophisticated population of users at the leading edge of filing and finding practices.

Similarities and Differences Across the Two Studies

To our surprise, we found striking similarities across the two study populations. The similarities were:

• A preference for location-based search for finding files;
• The critical reminding function of file placement;
• The use of three types of information: ephemeral, working and archived;
• Lack of importance of archiving files.

The main difference across the two study groups was:

• Use of subdirectories

In the Macintosh study, all users employed subdirectories in their filing schemes. In what we shall call the "mixed" study

(DOS, Windows, OS/2, Macintosh), only the two experienced users and the Macintosh user used subdirectories. We will not dwell on this difference in the use of subdirectories as it may be largely attributable to level of experience and a statistical study would be needed to establish such a correlation. It is intriguing that the Macintosh user in the mixed study used subdirectories, even though she was not an experienced user, suggesting that the Macintosh interface to filing functionality may make the use of subdirectories more readily understandable. Statistical treatment of this question is needed.

We look at each of the similarities across the two studies in detail. More subtle differences arising from the different user interfaces are also noted.

Location-based Search

There are two basic strategies for finding files which we shall call location-based and logical. In location-based finding, the user takes a guess at the directory/folder or diskette where she thinks a file might be located, goes to that location, and then browses the list of files or array of icons in the location till she finds the file she's looking for. The process is iterated as needed. While users prefer to be able to go to the correct location on the first try, they often view the files in the target location by date, by name, or by some other characteristic in order to identify the file from the list they are scanning. In logical finding, a text-based search with keywords or file names is used to locate the file with such utilities as "Find" on the Macintosh or "whereis," a DOS utility.

In both studies, users overwhelmingly preferred location-based search. In the mixed study, managers often depended on *applications* to provide the tools for selecting the desired files; that is, files were stored in the default directory created by application software. For example, a user would put all his graphics files in the Harvard Graphics directory. Most users in the mixed study were not computer experts, so when separate directories or groups of files were needed, they were likely to subdivide the information by locating it on separate diskettes or in top level directories. Barreau (1995) reported that software applications "were accessible from the root directory of hard drives or in prominent places on the menus of networks." Thus users who did not use subdirectories still had convenient access to their files. These users rarely performed maintenance operations on their directories, preferring to use additional storage media rather than to archive, delete, or compress data no longer used with frequency.

Barreau (1995) reported that users in her study preferred browsing lists of files rather than trying to remember exact file names. They did not use text searching techniques or used them very infrequently. In one interview sequence, for example, a user was scanning a list of files to retrieve a file she had only created that morning, saying, "What did I call that file?" Scanning the list to recognize the file name was easier than trying to remember the exact name, despite the recency of creation of the file. While more research on this topic is needed, we hypothesize that users prefer location-based filing because it more actively engages the mind and body and imparts a greater sense of control (see Nardi, Anderson and Erickson, 1995). Users seem to prefer to actively search for an actual file that

they previously placed in a particular place rather than sitting there waiting for the computer to return a list of files that may or may not be relevant. It is, after all, the user's *personal workspace* in which the search is going on, and the user may feel that they command the space. Retrieving files on remote machines beyond one's personal control may feel quite different. Users may more willingly adopt logic-based search techniques since the space being searched has not been personally arranged by them.

In the Macintosh study, the predominant pattern for finding files was to look for a file in a particular location and then look in a different location if the file was not found. On the Macintosh a "location" could be either a specific folder, or a specific region of the Finder screen which holds icons. For example, some users kept most of their high level folders on the right hand side of the Finder screen, and then further subdivided that space according to other categories such as project. If a file was not found within a couple tries, then the Find feature was used to locate the file. If the file still was not found, then finally, as a last resort, a text search program (such as in OnLocation™) was used. Often users only needed to complete the first step as the file was frequently in the first place the user looked. Nardi, Anderson and Erickson (1995) suggest that reluctance to use the Find command may be attributed to the difficulty in remembering the name of the desired file, much as Barreau (1995) reported in the mixed study. Macintosh users found that text search programs are often slow and retrieve too many irrelevant files so they used them only when all else failed.

Macintosh users created long descriptive file names to enhance recognition when scanning lists of files or icons (both views are possible on the Macintosh and it is easy to switch between them). DOS/Windows users were more restricted in file naming possibilities but also carefully named files, and used meaningful file extensions to make file names more distinguishable. Overall, file naming was given careful attention by all users, but for the purpose of jogging the memory when scanning files at a location rather than for the purpose of searching for a particular name with a logical search (see Barreau 1995, Nardi, Anderson and Erickson 1995). Both studies highlight the fact that users consciously organize their files for easy retrieval. They have in mind the goal of being able to find a file when they name the file and when they place it in a specific location.

Reminding

The location of information on the desktop also serves a critical reminding function. Users in both studies across all environments were observed placing files in locations where they were likely to notice them. For example, users left electronic mail messages in the in-box as a reminder of a meeting or placed files in the upper level of a directory structure as a reminder to complete work contained in the files. Macintosh users reported behaviors such as moving icons near the trash can as a reminder to delete them. One Macintosh user, an administrative assistant, put all incoming mail and documents on her desktop and made sure they were attended to by the end of the day. If there were any documents left over at the end of the day, she put them on the right hand side of her screen as a

reminder to work on them first thing next day. The OS/2 user clustered most of the icons for executable files in one area of his screen while icons to open windows were scattered over the rest of the screen surface as a reminder of what they did and as a tool for opening their windows, placed so that they were not too crowded together. He also left icons of old versions of his files on the desktop (with the color removed from the icon) until he was sure he wouldn't need the old files anymore. The Macintosh, Windows and OS/2 users had more scope for creating ways to use file location as reminders because of the greater flexibility of icons in this regard.

The reminding function of file placement dovetails back into users' preference for location-based finding. If file placement is to serve two masters – finding and reminding – then a "physical" system in which a specific location is associated with the file is more useful than a purely logical system. No matter how good text-based search is, there is no reminding function that comes along with logical search as there is with location-based schemes where the user can physically place files, whether in directories, folders or on the desktop, in a specific place where they can be noticed. With a file in a specific location, the user knows he will see the file and be able to use it as a behavioral trigger to remind him to take some action.

Three Information Types: Ephemeral, Working, Archived

Nardi, Anderson and Erickson (1995) discovered a pattern of electronic information handling similar to that found in Cole's (1982) study of paper documents. Cole found three types of information: "action information," "personal work files," and "archived information." The types of information found in the Macintosh study were strikingly similar: ephemeral, working, and archived information. We describe each type of information and then discuss how these categories relate to the mixed study.

Ephemeral information has a short shelf life and includes items such as (some) electronic mail messages, "to do" lists, note pads, memos, calendars, and news articles downloaded from databases. The central problem of organizing ephemeral information concerns where and how to file information that is needed for only a short time. Macintosh users prefer to keep such information visible, "loosely" filed with other information at the top directory level, or not filed at all, i.e., just sitting as a file on the desktop. With limitations on the amount of information that can be viewed on a screen at one time, managing large quantities of ephemeral information can be problematic. Users did not have perfect solutions to this problem; it is an area for further research.

Working information is frequently-used information that is relevant to the user's current work needs and that has a shelf life of weeks or months. Working information is often created by the user or is the product of the user's work groups. It is usually important enough to be organized by location and category in its own folder or location on the

desktop. Users reported having no difficulty finding their working information as they use it repeatedly and thus can easily remember where it is. Users know where the files are spatially, knowing precisely where to look for what is needed. As

projects near completion and the information is accessed less frequently, the categorical structure of the information becomes more important than the spatial location for organizing and finding files.

Archived information has a shelf life of months or years, but is only indirectly relevant to the user's current work. It is infrequently accessed. Most archived information represents completed work, including final reports and project histories. Users in the Macintosh study indicated that selecting files to place in the archive once a project is over is often the most difficult part of the process. Every user in the study indicated that their attempts to establish elaborate filing schemes for archived information failed because they proved to require more time and effort than the information was worth.

These information types map to the data in the mixed study as well. Of particular concern to the users in Barreau's study was the volume of ephemeral information they had to cope with. Many files were created as quick-and-dirty reports or memos to other co-workers. These files were supposed to stay in the user's environment only long enough to determine if follow-up was needed. But due to lack of timely response from co-workers, such documents often cluttered up the user's filing system longer than intended. These documents lived in a kind of uncomfortable no-man's land between ephemeral and working information. The problems of filing this information were intensified because the information outlasted the normal shelf life of ephemeral information.

Discussion

One of the most striking findings across both studies is that users keep little archived information in their systems. Many researchers assume *a priori* that archiving information is a central problem for users (e.g., Rao, *et al.* 1994) but this assumption may be a reflection of researchers' own needs for archival support. Researchers *do* keep a good deal of information around for a long while because it is in their interest to do so. For most jobs, however, information that is not current has much less relevance and utility. This was true for the managers, administrative assistants, graphic artists, and so forth, in the two studies reported here. An insight of the studies is that in designing finding and filing systems, it is important to consider the mix of information types people use, i.e., the relative proportions of ephemeral, working and archived information. Ephemeral information, we argue, has received too little attention in terms of tool development, while archived information has been emphasized.

It is a commonplace that we are overloaded with information, or just about to be. Is this really true? What the two studies reported here suggest is not that ordinary users are bombarded with potentially valuable information, if only they could find it, but rather that they are very busy with pressing short term needs! The ephemeral documents that represent these short term needs become problematic because there may be large quantities of them and, because they must serve a reminding function, they cannot just be buried away in a file somewhere. The problem is not that people are a hair's breadth away from finding that crucial piece of information that will make all the

difference, but rather, that Sally or Jorge has not responded in a timely fashion to a memo or report. In today's economy where people are expected to take on a large volume of work and deal with it efficiently, a pile up can occur with only short delays in response from co-workers. Tools to manage ephemeral information could help to relieve some of the pressure on today's workers.

What about better archiving for short-term information? Our studies did not tap the question of institutional memory and its potential importance. While our data that suggest that schemes for keeping around a lot of old information are not relevant to many classes of users, it is still possible that with better tools, archival information that is of relatively short term value, but is neither working nor ephemeral information, could become more valuable. Will a worker who is asked to write a brief summary of a company's resources for a proposal to be submitted within 24 hours be able to find a similar piece prepared for another purpose several months ago in time to avoid having to reconstruct it? And just where are those conference announcements and submission deadlines? What about those useful sounding URLs that one cannot fluently archive with current tools? Information of this type may hold great value for users and for their organizations over time as the files they are contained in may offer background information or lay foundations for similar or related future work.

Conclusion

Studies of user behavior can yield interesting results when the fuller context in which users work is considered. The mixed study was performed in a predominantly character-based operating environment with restrictions on the size of file names and limited help for computer novices. The Macintosh study was performed in the very different environment of a graphical user interface conducive to rapid learning. And yet the study results are still surprisingly consistent because both studies focused on how information *is used*, rather than narrowly scoping the focus of interest to be length of file names, or icon design, or some other user interface detail. Both studies suggest that the way information is used is a primary determinant of how it will be organized, stored, and retrieved in the personal workspace. The studies demonstrate consistent behaviors in organizing and finding information across different operating system environments. These behaviors are further consistent with those observed in organizing of physical offices. Thus we found that users prefer filing by location because it aids in helping them find what they need as well as serving a crucial reminding function. We found that users do not expend great energy on archiving because old information is generally not useful information. We found that users give up on elaborate filing systems because in the end they do not yield enough value. Users file information not according to systems of keywords or carefully architected logical schemes, but according to the dictates and vagaries of the kind of work they are doing and the type of information they are dealing with.

In designing studies of human-computer interaction, it is important to broaden our scope of analysis and not concentrate all efforts on narrowly focused atomic behaviors such as "retrieving a file." What both studies reported here show is that

retrieving and reminding cannot be torn asunder. An experimental treatment of retrieval methods, for example, in which all factors except retrieving are "controlled" for experimental rigor, would miss the whole point that finding and reminding are intimately linked in users' practice and should be considered together. The design of a new system of archiving should consider exactly which materials a user might find useful, where they originate, and how long they might be of value. Stepping back and taking a more panoramic view of users' practice will enable us to design more useful, congenial technology.

Acknowledgments

Both authors thank their respective study participants for generously donating their time and thoughts to the research.

Authors' Addresses

Deborah Barreau
College of Library and Information Services
University of Maryland
4105 Hornbake
College Park, Maryland, USA
barreau@wam.umd.edu

Bonnie A. Nardi
Advanced Technology Group
Apple Computer, Inc.
1 Infinite Loop
Cupertino, California 95014, USA
nardi@apple.com

References

Barreau, D.K. (1995). Context as a factor in personal information management systems. *Journal of the American Society for Information Science* 46, 5 (June) pp 327-339.

Blomberg, J., Suchman, L. and Trigg, R. (1994). Reflections on a work-oriented design project. *Proc. PDC'94* (Chapel Hill, NC, October 27-28).

Cole, I. (1982). Human aspects of office filing: Implications for the electronic office. *Proceedings Human Factors Society,* Seattle, Washington.

Lansdale, M. (1983). The psychology of personal information management. *Applied Ergonomics* 19, 55-66.

Malone, T. W. (1983). How do people organize their desks? Implications for the design of office information systems. *ACM Transactions on Office Information Systems* 1, 99-112.

Nardi, B., Anderson, K. and Erickson, T. (1994). *Filing and finding computer files.* Technical Report # 118. Cupertino: Apple Computer, Inc.

Rao, R., Card, S., Johnson, W., Klotz, L. and Trigg, R. (1994).Protofoil: Storing and finding the information worker's paper documents in an electronic file cabinet. *Proc. CHI '94* (Boston, 24-28 April). Pp. 180-185.

Suchman, L. and Wynn, E. Procedures and problems in the office. (1984). *Office Technology and People,* 2, 134-54.