To Have and to Hold: Exploring the Personal Archive

Joseph 'Jofish' Kaye[1], Janet Vertesi[2],

Shari Avery[3], Allan Dafoe[2], Shay David[2], Lisa Onaga[2], Ivan Rosero[4], and Trevor Pinch[2]

[1, 2, 3] Cornell University

[1] Information Science, 301 College Ave, Ithaca NY 14850

[2] Science & Technology Studies, Rockefeller Hall, Ithaca NY 14853

[3] Veterinary Medicine Information Technology, Ithaca NY 14850

[4] Amazon.com, Seattle WA 98144

jofish | jav38 | slk9 | sd256 | lao9 | ad285 | idr2 | tjp2 at cornell dot edu

ABSTRACT

The personal archive is not only about efficient storage and retrieval of information. This paper describes a study of forty-eight academics and the techniques and tools they use to manage their digital and material archiving of papers, emails, documents, internet bookmarks, correspondence, and other artifacts. We present two sets of results: we first discuss rationales behind subjects' archiving, which go beyond information retrieval to include creating a legacy, sharing resources, confronting fears and anxieties, and identity construction. We then show how these rationales were mapped into our subjects' physical, social and electronic spaces, and discuss implications for development of digital tools that allow for personal archiving.

Author Keywords

Archiving, ethnography, identity, filing, email, bookmarks.

ACM Classification Keywords

H.3.m [**Information Storage and Retrieval**]: Miscellaneous.

INTRODUCTION

The problem of personal archiving, in a nutshell, is that we collect more documents and objects than we can immediately access. We would like to have everything ready at hand, but given our limited physical capacity to access materials, we end up having to store most of our collections. We therefore organize this storage in such a way that we hope that we can retrieve the material on demand. Thus archiving can be seen as a functional problem involving storage, retrieval and a system that organizes this storage and retrieval. But in order to understand what is stored and retrieved, and why and how it is organized, it is necessary to address what archiving

means to the people who choose and use this material. If we want to design better technologies to support archiving activities in a digital arena, we must seek to understand the personal archive not only as a problem of building an efficient information-retrieval system, but also as a human practice. We must look for the needs that drive archiving behavior, and ask, "Why archive?"

Archiving in HCI

Archiving in its myriad forms has received a considerable amount of attention in HCI. For example, Henderson has looked at how people organize their desktops [17] and Jones et al. have reported on the extensive use of desktop folders [18]. Other researchers have studied in detail the use, archiving and/or storage of emails [4,5,10,23], of documents [8], of time management tools [7], of To Do lists [3] and personal information management software [3,6,19,20].

In addition to these medium-specific approaches, there is work in HCI on archiving itself as a practice. In their 2001 paper Whittaker & Hirschberg [22] examined the role of paper archive. They categorized office workers into two categories, pilers and filers, and noted that filers accumulated more information and accessed it less frequently than pilers. They also pointed out that only 49% of the archives they studied consisted of unique documents, over a third of the documents were publicly available, and that 'informal librarians' accumulated a given set of information and share it with their colleagues.

We also found Voida et. al.'s observations of practices surrounding iTunes sharing to have strong parallels to archiving practices. [21] It is not immediately apparent that the situation they describe is one of archiving, but when the use of iTunes is considered as one of accumulating, organizing and, importantly, sharing a collection of media, the similarities become striking.

We build on this significant body of work by examining the variety of diverse functions and structures in the personal archive. We posit archiving as an activity that occurs across media, locations, careers and time, an ongoing practice of selection, organization, collation, display, storage, retrieval

and disposal. As such, in order to design comprehensive archiving tools, we believe it is important to look at the archive as a whole, rather than focus on one medium, such as the email folder or filing cabinet, or a set of needs, such as sharing papers with a group. In this way, we hoped to uncover overarching concerns or needs that drive archiving as an activity: needs that may be otherwise overshadowed by affordances or limitations of a single medium.

METHODS

We started our enquiries with a set of questions we hoped to answer. For example, how do academics solve problems of storage and access wherever their personal space is located, be it a basement, a barn, a lab, a home study or a university office? How do such archiving practices change in response to evolving careers, new archiving and retrieval technologies, fears or experiences of archiving disasters, the growing amount of information available and the struggle to cope with ever-increasing accumulation of stuff? Can we identify systematic practices, coping strategies or what we might call different styles of organizing personal archives? And are there best practices we can learn from to maximize the utility of our own archiving systems?

We investigated these questions through an empirical study of forty-eight scholars working in various academic fields, most of whom were at a single Ivy League University. Our subjects ranged from graduate students to professors emeritus, included scholars of a variety of sciences, social sciences, and humanities subjects and a wide selection of national and ethnic backgrounds. While we recognize the limitations in studying only academics, we believe that that the rich field results and focused nature of this study will undoubtedly have lessons applicable to other domains.

The study consisted of forty-eight subject visits, involving semi-structured interviews and tours of their office space. After communally drafting a shared set of questions, the study was piloted by each member of the team interviewing a fellow teammate. The interview topics were further refined, a list of questions standardized, and each member of the team selected 5-7 people to interview. Collectively, we reviewed the subject pool to make sure that we had a relatively varied group by subject, gender, age, ethnicity and seniority. We individually contacted the subjects, requesting to interview the person in the location where they did most of their research. In some cases this was their work office, in some cases it was their home office, or some combination thereof.

Similar to the Technology Tour [1], each interview began with a tour of the subject's office or workspace, with subjects identifying archived material, explaining the contents of their desk drawers, of their temporary and longer-term filing systems, of the various ephemera of the office. We asked about the organization of their books, and the division of materials between this office and their home or other workspace. Tours frequently included explanations of devices for workflow and personal organization,

including calendars, diaries, notebooks and to-do lists. The tour typically finished at the subject's computer, and we then asked for a tour of their computer: their email program and its folders (or absence thereof), their internet bookmarks, their techniques for organizing and sorting documents, and their systems for backing up.

The second half of the interview probed particular aspects of the system we had just seen, with emphasis on preferred methods and the relationship between digital and material filing strategies. For example, we asked how they took notes on articles and other material, such as margin notes, or highlighting, and then asked how they marked up and stored digital texts. We asked how and when they read journals, online or in print form, and what they did with them afterwards. We also asked about the history of their technique: did they learn it from someone else, did they have formal training, and what system did they aspire to? We asked about the impact of the physical environment on their techniques, such as the difficulties of filing A4 papers in a letter-sized environment, built-in bookshelves or cabinets, moving between multiple offices, and the negotiation between private and public space. Finally, to get a sense of their practice in action, we asked them to retrieve a reference or two taken at random from one of their published papers. The entire process took about two hours.

Analyzing qualitative data in group research is often problematic. In order to share data from individual interviews, we audio recorded each interview and photographed the office and the archiving system. Each author then presented a detailed summary of their interviews to the group, sharing photos, stories, and quotes which fueled lengthy discussion. As these summaries progressed, we noticed trends and themes in archive function and structure across examples, which were discussed in weekly meetings and on a Wiki, and which we collectively coalesced into our final written result.

RESULTS I: THE GOALS OF ARCHIVING

As expected, every single one of our subjects archived for multiple reasons and in multiple ways. As we started to put the results of our interviews together, we noticed similarities in the *functions* of the individual archives: goals and uses of the archive that were common across disciplines, media, and academic seniority.

Finding it later

Not surprisingly, one reason why people archive materials is so that they can easily find them later, a need which currently drives many of the design strategies for personal archiving tools. Indeed, most of the archives we visited were usually easily accessible: filing cabinets, bookshelves, and boxes were either within arms' reach or a short push away on a wheelie chair.

However, when we chose a reference at random from one of their published papers and asked our subjects to retrieve it, we found little variation in retrieval times regardless of

their system. For example, we noticed no significant difference in retrieval times between subjects who kept their bookshelves in alphabetical order and those who clustered their books by subject. Positional memory seemed to take over, as subjects referred to the last time they used it or worked through the structure of their individual archival systems to find where it 'should be'. This suggests there is no single 'best practice' among the variety of structures we witnessed that radically influences the efficiency of information access: in the words of one subject, "*I do what works for me.*"

Subjects were almost always disappointed in their archiving system, becoming most frustrated when they could not find things in it. They seemed to suggest that a well-functioning archive is effectively invisible, only noticeable when it breaks down. Thus one subject, who maintained an elaborate physical and electronic system, claimed it was actually "*not entirely functional, or non-existent*" because "*it's very hard to find things.*" They would describe fantasies of magic, psychic archives: "*I'd just like to snap my fingers and it would appear…*" They would also trumpet the successful adoption of tools that supported this magical retrieval, or deride those that were ineffective; for example, while one subject strongly urged his interviewer to "*fight the hegemony of alphabetization*", another claimed "*my achievement here early on was to realize that the only way organize [my books] was alphabetically.*"

Subjects were also disappointed as things piled up. Filing was invariably a task perpetually deferred, resulting in a secondary archive-in-waiting. This smaller, limited 'archive' was sometimes seen to have the advantage of speedy retrieval and being more ready-to-hand than formally filed documents in the archive itself, and thus became part of the archive. One subject was happy maintaining a 'to-file box', saying "*I don't actually put things right away into files.*" However, other subjects saw the secondary archive as a breakdown in the archive itself, and frequently as a personal failing.

While larger institutional archives, such as libraries or shared laboratory spaces, may rely on rigid categories and search engines to speed up retrieval time, *finding things* alone does not explain the other types of activities and needs that surround and drive the personal archives. We noted four other significant values we will discuss below which were more prevalent than efficiency in the rich variety of structures and practices we observed.

Building a Legacy
When asked how often he accessed materials from his personal archive, one professor shrugged. Gesturing to his wall-to-wall bookshelves and two rooms of filing cabinets, stuffed either with letters alphabetized by author or his own papers numbered in chronological order into the 500's, he said slowly, "*Well, I sometimes look at the books... I would say I use the correspondence file often [to put things away into] ... I think [I make] relatively light use of the files.*" In

Figure 1: A subject's life work as a legacy. He gives each paper he writes a new, sequential number: paper #330 can be seen in the foreground.

spite of the reams of material and the rigid filing system of which he was very proud, most of his stored material was never accessed or retrieved. While the organizational structure made retrieval possible, finding things was clearly not the main priority of this archive.

We witnessed many personal archives such as this one, whose primary purpose was to allow visitors to make a visual sweep of the room in order to take in important aspects of the subject's personality and life's work. We called this kind of archive a "Legacy" archive. Essentially, these were testament to the subject's "life work", a self-constructed permanent record of the achievements and movements of the archiver. These subjects kept everything, but rarely went back to those files to retrieve a rough draft of a paper, a contributing author's responses, or a colleague's updates: indeed, another such archivist surprised the interviewer when, asked to retrieve a significant paper, he ignored the dozens of filing drawers in his office and turned to the internet. This example drove home the fact that such archives were meant for storage, but not necessarily for retrieval: *putting things away* and into the right place was much more important in this type of archive than ever retrieving items again.

The Legacy archives we visited often exhibited rigid structures, each unique and developed by the author alone, to maintain the large number and variety of materials that required keeping. As keeping everything often requires an extremely rigid filtering system (so as to know where to put everything!), subjects who displayed this form often exhibited great pride in the structure of their system, without necessarily referring to any particular aspect or item. Instead, they were most proud of the unified body of materials that reflected their career trajectory. Their archives' structures reinforced this need. While we recognize this archive value may be more typical for academics than others, we suggest that archivers in many

areas alphabetize volumes of correspondence, sort their libraries by theme and author over a number of shelves, or number their output chronologically. All these techniques result in elaborate organizational structures that both support the storage requirements and contribute to a sense of coherence and grander narrative of the archiver's work.

Sharing Resources

Flies and files, cards and computers dominated the shared space of a hallway we visited in an Entomology department, where the head of a large research group had personally compiled a sizable collection of materials for his lab. The collection was comprised of papers, slides, digital photographs, and specimens, all of which had different requirements for preservation, and was kept up by the professor's wife and his secretary, both of whom were responsible for the cataloguing and organization of the collection. While the laboratory archive spoke of the volume of specimens that the professor and his students had managed to acquire and preserve, this time the extreme volume and organization supported an opposite archival goal to just legacy building: instead of keeping things that spoke of one individual alone, this structure supported the necessary retrieval of documents by a large number of people who shared a particular resource.

We identified *Sharing* as an archival goal common to many of the personal archives we visited which were accessed often and by many individuals. Although the archive may be maintained by one or two individuals who were not, strictly speaking, librarians, a strong home-grown organizational structure supported the key activity of sharing materials with others, not only in terms of finding thing, but also in terms of knowing where to put things.

While the entomology archive was formerly a personal collection made into a larger shared resource, it was still private in terms of access and certainly not institutionalized in spite of its setting. Indeed, many of these archives occupied an interesting position between private and public. For example, one of our subjects keeps cut-outs from older (purged) scientific journals and catalogues them in EndNote software, which she acquired with a new computer at her new job. Newer articles are also stored digitally on the

Figure 3: Beyond *My Documents*: a shared archive of entomology research materials.

Figure 2: The distilled essence of a dissertation: in the event of a fire, the subject could '*grab this box and run'*.

computer. Now, although her collection is privately maintained, the contents of her catalogue are part of her lab's networked EndNote file. As such, the archive is now searchable by others in her workgroup. While the sense of accessibility defined this type of archive as a "Sharing" resource, it did not become fully institutionalized: its sense of ownership, home-grown organization, and belonging to a small group enabled it to remain, fundamentally, a personal archive.

Fears of Loss

"*If there ever were a fire, I would grab this folder right here*," declared one subject early in her interview. This sentiment was shared by a number of our subjects, exhibiting a fear of loss that was slightly different than that which underlay the Legacy or Sharing archives. Rather, a sense of anxiety or fear of a specific catastrophe involving the potential loss of irreplaceable information fueled the drive for preservation. [12] For example, we encountered fear of theft or fire, worry about an upcoming move, distrust of computers and the fear of a hard-drive crash, appearing unprofessional, or losing something important. It is critical to note that this anxiety was not always driven by rational risk or experience; however, in many cases, this fear influenced the archive's physical structure, from regular personal backups to specialized folders.

For example, one subject's entire Ph.D. research archive exhibited a funneled structure, in which larger boxes contained general material organized by chapter, with successively smaller boxes and eventually binders and files containing ever more important or core information from each chapter. Gesturing to this last stage of files, the subject claimed that if there were ever a fire, she could "*grab this box and run*". In her case, the question the archive was meant to solve (and which the subject repeated during the interview) was: "what would I take with me if there were a fire?" Many of our subjects exhibited practices that spoke

to this fear, even when it was not their archive's central concern. For example, one professor demonstrated her fireproof safe but could not find the key anywhere.

We made a particular effort to speak to academics who had suffered a loss in their personal archives due to fire or theft, and the effects of this loss were varied. We witnessed some subjects take up the archiving value again with vigor, focusing on recovery of their archive through purchasing the same books again or making up a catalogue of all the book titles one could remember so as to be able to find them again somewhere. However, one subject admitted that after a house fire when she was thirteen, she never kept anything of an archival nature, whether academic or personal, preferring now to display her judgment or discrimination at not keeping *anything* at all.

The above should not come as a surprise, given that preservation is often identified colloquially as one of the central goals or values of most archives. However, the physical systems that arose from this underlying fear bear some attention, as the funneling practice and even the decision not to archive demonstrate how underlying fears may drive structural decisions in building the personal archive. In biology labs, these concerns of preservation from decay took on a different meaning, with different structural requirements: as one subject asked, "*How do you back up a protein? How do you keep a moth alive? How do you keep a cancer cell from… growing?*" Indeed, the goal of preservation of specific material deemed absolutely irreplaceable, which structures not only the physical layout of the archive but also the practice of archiving and sorting incoming material to fit the priority-ranged categories, provides an excellent example of an archival value.

While the above fears may seem unfounded, the experience of losing an archive is a devastating personal event. Reflecting on the experience of losing her entire professional archive in an office fire, one subject wrote: "*I never expected to experience such a strong emotional*
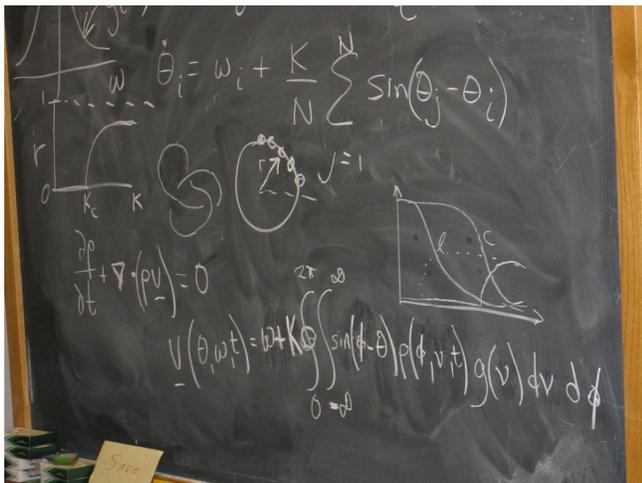

Figure 4: Although the equations have been untouched for years, this is clearly the chalkboard of a mathematician

reaction to the loss of my archive… After we had survived the first chaotic weeks I had a feeling of emptiness and deprivation and I felt very fragile. Working in our new, temporary office, I felt a big empty space behind my back (the place where my books used to be), and I felt cut off from the past and uncertain about the future. It is the first time in my life I experienced such a strong attachment to things. Talking with friends and family, I realized that books are an important part of my identity as an academic scholar."

Identity Construction: Impression Management in the Archive

One subject wrote her list of papers accepted to conferences on her white board; another kept mathematical algorithms on an otherwise unused chalkboard; parents displayed their children's drawings and new faculty hung their diplomas and awards on their office walls. As our study progressed, it became increasingly clear that personal archiving went well beyond the personal systems people used for filing, retrieving and referencing academic papers. In particular, we started to notice how archiving was used as an expression and crafting of identity, projected outwards towards the world as well as back at the individual to reinforce his or her sense of self. That is, archives serve a function in the subject's construction and maintenance of identity. Personal rationales, the built environment, mobility, and the social environment are all evident forces in the creation of and reflection on identity in archiving.

Claiming identity construction as a value for the personal archive requires some clarification. Goffman describes the change that occurs when a patient is institutionalized: they are stripped of their 'identity kit' – the clothes, wallet, books, briefcase, purse, handbag that are an intrinsic part of who they are. [14, 15] The personal archive is also a kind of 'identity kit': materials reflect and describe the owner, and the dispossession of these materials through purging, office moves, or fire brings discomfort and unhappiness to the archiver. Items in the archive serve as 'tokens', indicating who the archiver is and what they have achieved.

We found a strong parallel to Voida et. al.'s description of shared iTunes libraries.[21] Their subjects consciously constructed their publicly visible iTunes library to display the parts of their musical taste they felt they would most like to see emphasized, carefully building a public representation of their identity through their musical choices, much as we see happening in other areas of the personal archive. We differentiate identity construction from Goffman's notion of impression management [14] to make the point that the focus of this construction is not just outward, but is also part of the archiver's reinforcement of their perception of their own identity.

Another important goal of archiving, then, is to show to oneself or others who the archiver is, what role they play in their organization, what their achievements are, and what their interests are: "I am an organized person", "I am a

creative person", "I am the person that other people go to for copies of papers and reports." This archiving value was present in *every* personal archive we visited. Indeed, the 'visible personal' or a sense of identity in structure and content may be precisely what makes the archive personal, distinct from an institutional archive. While serving as a demarcation between public and private archives, identity construction also serves as a mark of ownership: personalization, the display of tokens, or peculiarities of structure make someone's archive clearly their own.

A key example of identity construction is the archiving of objects of value: subjects kept artifacts that were important to them, artifacts that they never accessed or used, but the ownership and display thereof in the office environment brought them a sense of accomplishment or identity, much as similar practices have been observed in the home. [9] For example, one professor kept twenty years of letters written to him from professional colleagues, stating, "*I keep this correspondence and I value it and I would never throw this out*". Correspondence also took up significant archiving space for another senior professor, who rarely looks into the rows of metal file cabinets that his secretary maintains. These examples betray elements of the "Legacy" value of archiving, in the sense that these pieces are put away in the right places, never to be retrieved, with the storage and protection more important than the retrieval. However, these pieces of correspondence not only referenced personal and professional relationships that were dear to the professor, but were also kept in an impressive display in row upon row of filing cabinets with important names on the front of them labeled "Correspondence". This reinforced the professor's identity as a central part of his field, with connections to the right places, as well as reminding him of his friends, colleagues and collaborative work. Further, the sheer volume of this correspondence was itself visually imposing, filling an entire room of filing cabinets and demonstrating important aspects of the subject's personality, connections and life's work.

Subjects often archived projects on which they had

Figure 5: Years of correspondence: one drawer of many.

Figure 6: Old laptops kept for sentimental reasons.

expended large amounts of effort, such as PhD thesis research, even if they never planned to access it again. More than one senior professor still kept notebooks from their dissertation amidst books, correspondence, and journals pertaining to present-day work. Another subject kept the ancient laptop computer upon which she had typed her dissertation years ago: the computer would not turn on, and the files were inaccessible, but that was not the point. In these cases, old notebooks and hardware stand as tokens for a period of the archiver's life, their hard work and dedication, and their accomplishment of achieving a PhD. These tokens are not necessarily in direct support of their research or ongoing work, but rather form part of an attempt to maintain and present an identity based on what is important to the person, and what values they reflect.

We visited one senior administrator and researcher who was successful in keeping an almost totally paperless office: his concerns about reliability and reproducibility meant that all of his filing was done electronically, on his laptop, with frequent backups. However, he did have two half-height bookshelves, which appeared to contain his field's standard mix of reference manuals, instruction books and the occasional textbook. On closer examination, it turned out that none of the books were used in the conventional manner: instead, each book was kept because of a personal connection to the material, be it that he authored a chapter within the book, or it was signed by the author with thanks for the researcher's help. Paperweights, desk toys, art on the walls and even the desk lamp were all gifts of appreciation: an extreme case of the role of physical artifact as token. [2] Further, in spite of his otherwise paperless existence, a few standing files held every paper ever written by himself or his wife; somehow, it would not have been the same to just keep a digital copy of his or his wife's publications. Tokenism, in this case, evaded even the normal structures of the archive itself!

RESULTS II: THE STRUCTURES OF ARCHIVING

All of our subjects archived to some degree in both physical and digital space. These spaces ranged from custom-built offices with a built-in library to Rubbermaid containers transported in the back of a car, and from dedicated servers to inboxes with thousands of messages. While different kinds of tools are available in these distinct types of spaces, subjects appropriated or neglected these tools as necessary in building the structures they needed to support their archiving goals.

The Use of Physical Space

Many of our archivers chose a particular medium that solved a special problem for them, and extended it to cover almost the entirety of their archive, regardless of the range of subject matter contained. These included vertical and horizontal filing boxes, filing cabinets, bookmarks, and email or file folders, all pressed into use for a variety of situations perhaps far beyond those originally envisioned.

One post-doctoral fellow who maintained a number of teaching positions at different institutions without on-campus offices solved his archive problem by using his car as an archive location, and employing large containers for easy transport of his archived materials. Although the car archive was built to confront the problems posed by mobility, it was notable that the problem of mobility was not effaced by his solution, but rather given center-stage: the bags and boxes became symbols of his mobile situation, his capacity for mobility, and the stage of his career. Similarly, extreme use of a particular filing system, such as file boxes, vertical storage units, or filing cabinets gave a sense not only of the general organization of the office material but also of the unity of subject matter or approach.

Just as mobility structured the car archive, we also witnessed temporality in a similar role. One typical subject we interviewed kept his accumulated papers and books in an order that reflected not only similarity of themes but also his own intellectual history: the things he was thinking

Figure 7: One professor's files in a custom-modified barn: extreme use of one type of archiving system.

through at the time. This subject insisted that he was "*not particularly organized*", but was able to retrieve information without difficulty. This organizational structure also helped him to further his intellectual development in much the same way as it reflected it: as he put it, "*I don't mind actually having to rummage for stuff... I rather enjoy going through piles of old off-prints and reprints and stuff because I will find things that I've forgotten about and that I wouldn't have come across if I had everything organized and alphabetized by author.*" In this case, both the material's display and retrieval built a story about the professor's life's work; a story that continued to evolve in its process of use.

In quite an opposite vein, we met with one graduate student with an eclectic and disorganized collection of papers and notes and experiments and no evident organizing principle. But the lack of a system did not indicate a lack of identity construction. She made a point of saying to the interviewer, "*You should have interviewed [my brother]: he is super-organized.*" She then admitted to constant arguments with this elder brother over how to keep her material, which always culminated in her demand for him to respect her as a scholar and the fact that her ways of doing things were not the same as his. Clearly, she (and their colleagues, as they were in the same department) viewed the apparent disorganization of her personal archive juxtaposed against her brother's hyper-organized example, and this was a relevant indicator to her of her own approach to the material and her academic identity. In a similar style, another subject cultivated a 'mad professor' image with both his notoriously cluttered desk and his uncanny ability to find obscure documents buried on it nonetheless.

The Use of Digital Space

While the dominant metaphor of contemporary user interfaces is that of an office environment, to a large extent this mimicking attempt is unsuccessful in capturing the richness of physical space, such as margin notes and physical location. The virtual office – complete with desktop, wallpaper, and trash can – is a veneer beneath which the dominant mechanism of information sorting and retrieval is that of virtual folders, rather than any other affordance of the physical office. Combined with common search tools, the ability to easily name and color folders, and the ability to easily build deep hierarchies, the folder system is designed primarily for easy categorization, storage, and retrieval. It would seem plausible that a system so designed would lend itself to better performance in terms of efficiency, but evidence from the field shows that most interviewees had a pretty good idea of where both electronic and paper documents should be and where they were, and most people retrieved paper and electronic copies without significant differences in latency. In some examples, subjects were able to retrieve paper documents much faster than they could find an email in their elaborate folder systems.
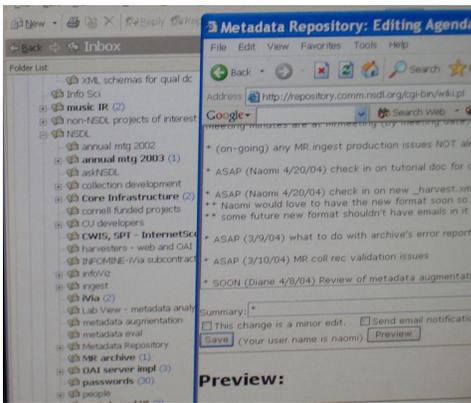
Figure 8: One subject's email filing system, with hundreds of email folders.

Ironically, instead of treating the spacelessness and fidelity of digital documents as features, subjects often perceived them as difficulties to overcome. For example, the spacelessness of electronic documents purportedly translates into two key advantages of electronic archiving: the ability to access the files from various geographical locations and the saving of physical space, which in most work environments is a scarce resource. For example, one of our subjects works primarily with images generated for research, be they micrographs, slides, or photos. The increasing digitization of existing slides or new images has facilitated her work, and made it easier to send copies to collaborators or others. However, this advantage of space was not without a drawback, in that the electronic files did not mesh with the analog indexing and categorization information. Similarly, another subject worked with large data files generated by computer controlled lab equipment. This equipment was isolated from the network for security reasons, making electronic archiving pose another type of problem. In order to share the data with her peers, or to use it on different workstations, the files needed to be burnt to DVD manually, an activity that took over an hour each day.

Several interviewees treated computer space as scarce: while they had plenty of office space, they behaved as if computer storage was expensive, and tried to save space on their computers. This may have more to do with the subject's mental model of their digital archive than the archive itself: keeping a clean and compact digital space meant a minimum of mental overhead to track items within it. For others, especially those that work in more than one office and do not have a portable computer, the digital nature of their content was actually a liability. Before digitalization they could simply take a folder of papers home, to a meeting, or to a long flight and continue to work on them. Now, unless they had a portable computer and a long-lasting battery, they needed to plan ahead, print copies and carry them around anyway. In those cases, just as in the case of the hard-to-backup microscope files, the electronic ideal is clearly disturbed by the necessity of physical objects. Even some of the interviewees that used laptops

reported that they printed papers, citing the convenience of reading print, the difficulties in conveniently making digital annotations, and the low resolution of computer screens: one subject said, "*I print [papers] out to read if I want to give them any real respect.*"

Another trait of electronic documents, the inherent ability to duplicate them easily, translates into the ability to back-up content rapidly. But few of our subjects bothered to make backups of their digital documents, and even fewer made backups of their physical documents and objects by translating them into digital files, such as PDFs, audio or video files. Most of them believed that if their documents were lost they could probably get duplicates relatively easily – a belief that on contemplation they found to be inaccurate when it came to original research and records from field research such as research notebooks and videotapes. On the other hand, those interviewees who reported that they performed some type of electronic backup, motivated by their own experience of computer crashes or stories by close peers, reported that this took time and effort.[1]

Much as in physical archiving, we saw our subjects rely on a single approach to organize their entire digital archive. Several of the figures demonstrate this: Figure 8 shows a subject who used email folders as her dominant filing system, while Figure 9 shows a carefully organized system of digital file folders. Another subject observed that "*the bookmarks are the most important research tool that I have.*" He kept over 120 folders of categorized bookmarks.

In spite of these similarities, due to the spaceless and replicable qualities of digital files and their systems' support of efficient retrieval, electronic documents on the whole were treated differently than analog files. However, we noted that people tried to maintain the same or similar values in their digital archives as they did in their analog archives. For example, one of our interviewees kept every single email he had ever sent or received in an elaborate folder structure, building a Legacy archive that perfectly achieved the function of elaborate storage, but not always, necessarily, retrieval. In another potent example, the researcher who kept a paperless office seemed to demonstrate an overwhelming preference for the digital

---

[1] While some institutions built infrastructure to support daily tasks like file synchronization and automatic backup, most interviewees reported that they had to improvise, each inventing his or her own electronic filing and backup system. Institutional support was limited to supplying the physical means with which to handle it: once filing cabinets, now hard disks. We found that serious hard disk crashes without adequate backup had negative effects on productivity that were as serious as office moves and fires. While the majority of institutional (rather than home) offices we observed were equipped with fire extinguishers, we suggest that the long term gains of providing extensive backup as a standard campus-wide service – like network infrastructure or sprinkler systems – may far outweigh the short term costs.

over the analog: except for his tokens, he did not keep anything that he could not file into his electronic folders. However, his elaborate back-up system belied a deep anxiety about losing his materials, and thus betrayed his fundamental *distrust* of the digital in being able to support this critical aspect of his archive. Thus, similarities between the digital and analog were not observed in terms of practices or structures, but in terms of the underlying *values*; values which, we believe, can better inform design.

DISCUSSION & IMPLICATIONS

In each of our interviews, asking "What is your personal archive *for*?" revealed a strong connection between each archive's function and its structure. Those who wished to build a story of their life's work or share their materials among a research group had to solve the problem of storing a wide variety of materials in a coherent way, and thus made use of a number of rigid and specific methods to maintain a strict filing system tailored either to storage or retrieval. Those who had no office to speak of or who had multiple offices on a campus had to solve the problem of information mobility or location, and therefore developed either highly specialized locals for different types of work or highly mobile workstations. Those who were afraid of fire or computer crashes either kept everything in such a way as to be able to save the most important information, or else kept nothing at all so as to avoid disappointment and loss. And every one of our subjects used their archives, their tokens or their visible personal information, to tell visitors and colleagues something about their past, their work, or their family.

We believe it is difficult and often inappropriate to emerge from a study of local users' practices with a checklist of global design standards or recommendations, no matter how coherent the results. However, we suggest that focusing on values, not simply parroting their resulting practices or
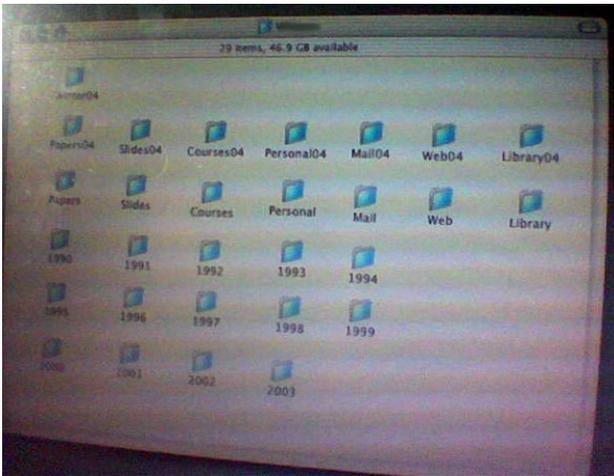

Figure 9: The digital filing system of one senior administrator in a technology field was based on a fundamental distrust of computers: he did not trust computers not to crash, so his entire filing system was designed for easy incremental backup.

organizational tools but investigating what the archive is for is paramount in designing successful systems for the personal archive. First, as these values of legacy, sharing, anxiety, and identity construction drive each archive's physical structure, *they also set the criteria for judging the archive's success*. That is, subjects judged the adequacy of their archives not only (or sometimes rarely) based on efficiency of retrieval, but also on how well or how poorly the archive helped them to identify and keep what they had decided they needed to keep and display what they needed to display. The questions that the archive is built to answer dictate the wide variety of archival practices in such a way that the user can always say, "it works for me".

Second, this finding reveals that there are no identifiable 'best practices' for archiving: rather, tools built to support the activity of archiving must accommodate many types of goals, methods, and styles. Flexibility and the user's own ability to tailor or fine-tune a system will be key to the success of any digital archive tool. Flexibility and fine-tuning are also essential as they facilitate both extreme use of a system as well as the hyper-customization that can be used to construct and project a sense of identity that accompanies ownership. Here is it important to note again that many of our subjects were *extreme users* of one or another system – filing boxes, cabinets, bookmarks, email folders – but not necessarily more than one. The challenge for the designer is to allow not only for customization in structure and extensibility for multiple functions, but also for the archive's use in identity construction, while maintaining appropriate levels of privacy.

Most personal digital archiving tools to date have concentrated on the value of 'finding it later'. In the manner of value-sensitive design[11], we suggest that the other archiving values discussed here provide opportunities for digital tools that support the full range of archiving practices. For example, Voida et. al have shown how iTunes represents an effective implementation of a digital archive that incorporates functions beyond storage and retrieval of digital information, such as impression management. [21] We suggest that studying or adopting practices from the customization of desktops, screensavers, blogs or personal websites may prove useful in informing digital archiving tools for identity construction. Similarly, perhaps backup systems ought to be *less* transparent, assuaging the user's fears of loss. File sharing tools need to be more nuanced, and easier for the novice to control from the desktop. More work needs to be done on electronic legacy building, confronting the concern that file formats will become obsolete and ultimately irretrievable. These gestures towards a larger project of value-laden archiving tools can provide a fruitful entry point for future design.

CONCLUSIONS

To the ongoing rhetorical question whether soon enough all archiving will be electronic, we will have to join the chorus [16] and answer with a resounding "no!" People do not take

full advantage of the backup potential of the digital, and often need to print documents out in order to use them as the functions they require in the analog domain are not well-supported digitally. Digital documents can be no easier to find naturally than analog files; as has been shown [18], subjects relied on the same techniques – a sense of where things ought to be or where they last put them – in order to both store and locate files. But most importantly, current electronic file systems do not offer solutions to the full range of answers to "Why archive?"

So what is the personal archive for? To a certain extent, it is for storing and retrieving information, but more often it is about other important values, such as building a legacy, sharing information, preserving important objects, and constructing identity. These values may affect structural decisions in designing the personal archive, and may be affected by institutional environments, but always reflect something about the archiver to whom they belong.

To return to one of our initial questions, we recognized that ultimately 'best practices' in this domain resist standardization: personal archiving is by nature a personal system, and effective archives are good at supporting values that form the underlying function and structure of the personal archive. Any digital system that attempts to replicate or complement analog archiving activity must take these factors into account in designing effective systems. Finding things fast is not always the answer: nor is adopting metaphors such as desktops and folders, or, arguably, abandoning them because they are not efficient or used to their imagined potential. Asking "Why archive?" reveals richer underlying rationales that can inform technological design, and if these underlying values are taken seriously, they will ultimately influence uptake.

REFERENCES
1. Baillie, L. (2002) *The Home Workshop: A method for investigating the home.* PhD Thesis, School of Computing, Napier University, Edinburgh.

2. Belk, R. The role of possessions in constructing and maintaining a sense of past, *Advances in Consumer Research* 17 1990, 669-676.

3. Bellotti, V., Dalal, B., Good, N., Flynn, P., Bobrow, D. What a To-Do: Studies of Time Management Towards the Design of a Personal Task List Manager. *Proc. CHI'04* 735-742.

4. Bellotti, V., Ducheneaut, N., Howard, M., Smith, I. Innovation in Extremis: Evolving an Application for The Critical Work of Email and Information Management. *Proc. DIS'02* 181-192.

5. Bellotti, V., Ducheneaut, N., Howard, M., Smith, I. Taking Email to Task: The Design and Evaluation of a Task Management Centered Email Tool. *Proc CHI'03* 345-352.

6. Boardman R, and Sasse, M.A ."Stuff goes into the computer and doesn't come out": a cross-tool study of personal information management. *Proc. CHI'04* 583-590.

7. Blandford A.E. and Green, T.R.G. Group and individual time management tools: what you get is not what you need. *Pers. & Ubiq. Comp.* 2001 5:213-230.

8. Bondarenko, O., and Janssen, R. Documents at Hand: Learning from Paper to Improve Digital Technologies. *Proc. CHI'05* 121-130.

9. Csikszentmihalyi, M., Rochberg-Halton. E. The Meaning of Things: Domestic Symbols and the Self. Cambridge University Press, 1981.

10. Ducheneaut, N., and Bellotti, V. E-mail as Habitat. *interactions* September + October 2001 30-38.

11. Friedman, B. (Ed.). Human values and the design of computer technology. CSLI, Stanford University, 1997.

12. Grayson, K. and Shulman, D. Indexicality and the verification function of irreplaceable possessions: a semiotic analysis. *J. Consumer Research*, 27 June 2000.

13. Featherstone, M. Archiving Cultures. *British Journal of Sociology*, 50(1) January/March 2000 161-184.

14. Goffman, E. *The Presentation of Self in Everyday Life.* Penguin, 1959.

15. Goffman, E. *Asylums.* Penguin, 1968.

16. Sellen, A. S. and Harper, R.H.R. *The Myth of the Paperless Office.* MIT Press, 2001.

17. Henderson, S. 2004. How do people organize their desktops?. In *Ext. Abs. of CHI'04* 1047-1048.

18. Jones, W., Phuwanartnurak, A., Gill, R. and Bruce, H. Don't take my folders away! Organizing personal information to get things done. *Proc. CHI '05* 1505-6.

19. Kim, P., Podlaseck, M., & Pingali, G. Personal Chronicling Tools for Enhancing Information Archival & Collaboration in Enterprises. *Proc. CARPE'04* 56-65.

20. Peters, R. Exploring the Design Space for Personal Information Management Tools. *Proc. CHI'01* 413-414.

21. Voida, A., Grinter, R. E., Ducheneaut, N., Edwards, W. K., & Newman, M. W. 2005. Listening in: practices surrounding iTunes music sharing. *Proc. CHI '05* 191-201.

22. Whittaker, S. and Hirschberg, J. 2001. The character, value, and management of personal paper archives. *ACM Trans. Comput.-Hum. Interact.* 8, 2 (Jun. 2001), 150-170

23. Whittaker S. and Sidner, C. Email overload: exploring personal information management of email. *Proc. CHI'96* 276-283.