Information Scraps: How and Why Information Eludes our Personal Information Management Tools

MICHAEL BERNSTEIN, MAX VAN KLEEK, and DAVID KARGER
Massachusetts Institute of Technology
and
MC SCHRAEFEL
University of Southampton

In this paper we investigate *information scraps* – personal information where content has been scribbled on Post-it notes, scrawled on the corners of sheets of paper, stuck in our pockets, sent in e-mail messages to ourselves, and stashed in miscellaneous digital text files. Information scraps encode information ranging from ideas and sketches to notes, reminders, shipment tracking numbers, driving directions, and even poetry. Although information scraps are ubiquitous, we have much still to learn about these loose forms of information practice. Why do we keep information scraps outside of our traditional PIM applications? What role do information scraps play in our overall information practice? How might PIM applications be better designed to accommodate and support information scraps' creation, manipulation and retrieval?

We pursued these questions by studying the information scrap practices of 27 knowledge workers at five organizations. Our observations shed light on information scraps' content, form, media and location. From this data, we elaborate on the typical information scrap lifecycle, and identify common roles that information scraps play: temporary storage, archiving, work-in-progress, reminding, and management of unusual data. These roles suggest a set of unmet design needs in current PIM tools: lightweight entry, unconstrained content, flexible use and adaptability, visibility, and mobility.

Categories and Subject Descriptors: H5.2 [Information interfaces and presentation]: User Interfaces. – Graphical user interfaces.
General Terms: Design, Experimentation, Human Factors
Additional Key Words and Phrases: Personal information management, note taking, information scraps

This work has been partially funded by Nokia Research Center Cambridge VIRPI Project, a National Science Foundation Graduate Research Fellowship, the EPSRC (UK) Overseas Travel Grant EP/E035930, the Royal Academy of Engineering Global Research Award, and by the Web Science Research Initiative.
Authors' addresses: Michael Bernstein, Max Van Kleek, and David Karger, Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 32 Vassar Street, Cambridge MA 02139; email: {msbernst, emax, karger}@csail.mit.edu; mc schraefel, School of Electronics and Computer Science, University of Southampton, Southampton, UK, S017 1BJ; email: mc@ecs.soton.ac.uk.

1. INTRODUCTION

Despite the number of personal information management tools available today, a significant amount of our information remains out of their reach: the content is instead scribbled on Post-it notes, scrawled on the corners of sheets of paper, stuck in our pockets, sent in e-mail messages to ourselves, and stashed into miscellaneous digital text files. This scattered information ranges from ideas and sketches to notes, reminders, shipment tracking numbers, driving directions, and even poetry. The information may never make its way into our usual PIM applications – yet we carry it around with us, decorate our desks with it, and often even make sure to archive it. For a category of personal information with so little traditional support, it is all but ubiquitous in our lives.

We refer to these pieces of personal information as *information scraps*. The term suggests several images: notes that are written on a scrap of paper, that are incomplete, or that have been separated from our primary personal information tools. As a class of personal information, we have much still to learn about information scraps. What similarities exist among scrap features and management practices? Why are information scraps so often held outside of our traditional PIM locations and instead on scraps of paper or in text files? Why do we manage other scraps by co-opting our traditional PIM applications against their intended modes of use, such as by composing e-mails addressed to ourselves? If these unorganized bits truly indicate the limits of our PIM tools, how might we begin to build better tools? Our goal in this research is to open an investigation of information scraps, so that we might begin answering these questions.

In this paper, we investigate the nature and use of information scraps. We contribute a cross-tool methodology for studying existing information scraps, and apply this methodology to an investigation of information scrap practice. In our study, we investigate the *information types* [Jones 2007a] stored in information scraps, scraps' layout and language, the tools used in support of information scrap work, and the information scrap lifecycle. Our artifact investigation reveals a large diversity of information types captured in scraps; these uncommon types cumulatively account for a sizable percentage of all information scraps. Through analysis of our results, we derive a characterization of the typical roles that information scraps serve in personal information practice: temporary storage, cognitive support, reminding, information archiving, and recording of unusual information types. These roles suggest a set of unmet design needs in current PIM tools: lightweight entry, unconstrained content, flexible use and adaptability, visibility, and mobility. Finally, we describe approaches that we believe will be the most successful in the information scrap management tools of tomorrow.

1.1 Information Scrap Definition

Though our investigation began without a firm definition for the term *information scrap*, we have continued to refine our ideas. Firm boundaries around what items constitute (and don't constitute) information scraps allow us to relate our efforts to prior work, communicate our ideas to a general audience, and scope our research program.

An information scrap is an information item that falls outside all PIM tools designed to manage it. This definition suggests that information scraps include items such as address information not in the address book, electronic communication not in the e-mail client, and to-dos not in a to-do manager. It intentionally includes information items for which no PIM tools currently exist, as well as information items stored and managed in general-purpose (e.g., non-PIM) information tools. For the purposes of our work, we choose to include tools such as notebooks, spreadsheets, and text editors/word processors in the set of general purpose tools because they tend to be catch-alls for PIM data. Our definition also intentionally makes no distinction between paper and digital PIM tools. To illustrate, here are some examples of information scraps from our research:

- Note of how to make a call abroad saved as a text file in a "Miscellaneous" folder
- To-do on a Post-it note
- Photo of a whiteboard from a discussion kept on the computer desktop
- Meeting notes in a general-purpose paper notebook
- Serial number for an application saved in an e-mail to yourself
- A friend's phone number written on a piece of scratch paper
- Cooking recipes kept in a personal wiki
- Song lyrics and guitar tabs taped on the wall
- A copy of an academic transcript saved in a text file

By our definition, information scraps are the personal information items that have fallen between the cracks of our PIM tools. An information scrap is evidence that there is no appropriate tool at a time of need; the user deliberately chooses a tool with affordances designed for other forms of information. In our analysis, we treat the existence of such items as evidence of PIM design failures and thus suggestive of unfulfilled design opportunities.

We might also examine whether information scraps constitute a single, well-defined *information type*. Jones defines an information type (or *information form*) by the "constellation of tools and applications that make it possible to manipulate" a set of items [Jones 2007a p. 7]. Information scraps might be defined in terms of the characteristically large and varied set of tools that people use to typically create, manage and hold their scraps, including Post-its, e-mail clients, word processors and even the backs of hands. This approach fails to yield a useful definition, however, because these tools are sufficiently general that they could be used to hold, create, and manipulate nearly any other kind of information. Also, our informal notion of information scraps includes items that theoretically can be created and held in any application or tool. Thus, the set of tools supporting information scrap creation and manipulation is the set of all applications, which is not a useful definition for us either.

The definition given in this section was developed over time and engagement with the participants of our study via a grounded theory approach. As the study described in this paper was mainly exploratory, we developed an interim definition described in Section 4.1 as a theory-building placeholder.

1.2 Results Overview

Our study consisted of 27 semi-structured interviews and artifact examinations of knowledge workers across five organizations. Among the artifacts we collected, we found information scraps to encode a small set of popular information types, including to-dos, meeting notes and contact information, but also a wide variety of uncommon types, including fantasy football lineups, passwords and guitar tabs. Information scraps often exhibited abbreviated language and underspecified data, sometimes complemented by sketches and freehand drawings. Tools such as e-mail and paper notebooks were most popular, although more structured tools such as calendars were often adapted or co-opted as well.

Synthesizing these results, we consolidated a set of roles that information scraps commonly play in personal information practice:

- Temporary storage: created with a short life expectancy and retained as memory prostheses [Lamming et al. 1994] until such time as their usefulness has expired.
- Cognitive support: works-in-progress, brainstorms, and other instances of "thinking it through on paper."

- Archiving: intended to hold on to important information reliably for long periods of time – for example, web site passwords or descriptions of how to perform a complicated task.
- Reminding: placed in the way of our future movements and activities, thereby leaving reminders for ourselves.
- Unusual information types: captured because the information did not fit in the participant's existing PIM tools.

By examining the reasons why information scraps were used in the preceding roles, we derived a set of needs and design affordances for information scrap management:

- Lightweight capture: lowering the time and effort barriers associated with capturing personal information.
- Flexible contents and representation: allowing multiple capture modalities, and enabling the user to capture whatever information is important.
- Flexible use and organization: enabling users to devise their own organizational systems or to adapt tools and information in novel ways.
- Visibility and reminding: priming information so that it is likely to be encountered at a desired point in the future.
- Mobility and availability: migrating personal information to and from mobile scenarios.

Finally, we applied these insights to motivate PIM tool designs that provide these affordances and are thus better prepared to handle the exigencies of messy personal information practices.

2. RELATED WORK

In this section we situate our study of information scraps among the rich body of research already surrounding PIM-related activities. We begin by reviewing related research on the psychological underpinnings behind information scrap practice. Then, we canvas specific practices related to individual information scrap types such as to-dos and general personal information management practices as they bear on information scraps.

2.1 Psychological Foundations

There exists extensive psychological literature surrounding our motivations for creating and manipulating information scraps. Perhaps the simplest framing of the problem was formulated by Ross and Nisbett in *channel factors*, the “small but critical facilitators or barriers" to an action [Ross and Nisbett 1991]. Ross and Nisbett demonstrated the amplified effects that small difficulties or facilitators will have on human action, just as a pebble placed at the fork of a stream can dramatically divert the course of water. Seemingly small time and effort requirements such as booting up a laptop might thus be perceived as enough of a burden to cause us to use other means of capture such as writing on our hands.

There are many such channel factors encouraging us to create information scraps. Lansdale was the first to relate psychology to the study of personal information management [Lansdale 1988]; in his work, he noted classification, or filing, as a cognitively difficult activity of special note. This result suggests that information scraps may be created when the cost of filing a piece of information is perceived to be too high – whether choosing a point in a folder hierarchy or deciding which of several related applications to use. Csikszentmihalyi identified humans' desire to maintain a state of *flow* where uninterrupted concentration is highest [Csikszentmihalyi 1991]; Bederson translated this concept into interaction design principles in support of the flow state [Bederson 2004]. When the user is in a flow state, Bederson and Csikszentmihalyi argue that unrelated thoughts and interruptions may be undesirable, causing us to write them down as quickly as possible before our original thought was lost.

Information scraps often serve as a memory prosthesis [Lamming et al. 1994] or exosomatic memory, later used to remind us of the original thought. Scraps can help us index into our memory via a variety of cues. Location is a very powerful memory primer [Darken and Sibert 1993, Jones and Dumais 1986, Robertson et al. 1998a, Yates 1966]; a combination of knowing *what* and *when* can also effectively aid recall of the rest of a memory [Wagenaar 1986]. We are also able to recall a variety of contextual information about our documents to potentially aid in re-finding, such as textual content, visual elements, file type, or implicit narratives around file creation [Barreau 1995, Blanc-Brude and Scapin 2007, Gonçalves and Jorge 2004]. However, many information scraps do not include such metadata; it is unknown whether the highly abbreviated contents of many information scraps (e.g., "Joe the attorney" [Bellotti et al. 2004]) are more powerful memory cues than those above.

Often, we fail to create memory prostheses such as information scraps even when they might later be useful. We are habitually overconfident in our own knowledge and memory [Lichtenstein et al. 1982], leading to conscious choices not to remember an item that later becomes unexpectedly valuable. Further, even if we chose to write down or make an effort to remember, we do not always utilize this information when recall is needed. Our memory's faulty yet quick access is often preferred over accurate but slower external memory aids [Gray and BoehmDavis 2000, Gray and Fu 2001, Kalnikaité and Whittaker 2007]. Such a preference suggests that information scraps may only be deliberately re-accessed when our own memory has failed.

2.2 Information Scraps in Studies of Specific Data Types

Information scraps take many forms, and researchers have noted their existence across a number of type-specific studies. Here we review the relevant work by type.

Perhaps the most typical information scrap is the note-to-self. Through a series of semi-structured interviews, Lin et al. arrived at a model of such notes' lifecycle: trigger, record, transfer or maintain and refer, complete, discard or archive [Lin et al. 2004]. Campbell and Maglio identified salient characteristics of what they termed *notable information*, including transience, visibility, mobility, ability to post, transferability, short length, and ease of both creation and destruction [Campbell and Maglio 2003]. The authors further observed a strong preference for paper-based media over digital media. Dai investigated this preference by interviewing expert users of PDA memo applications to suggest future design directions; users were typically most hindered by a lack of organizational support for their digital notes [Dai et al. 2005]. Hayes et al. studied the phenomenon of *short important thoughts*, uncovering a strong need for ubiquity and mixed-initiative systems in the support of such information [Hayes et al. 2003]. Strikingly, 73% of Hayes et al.'s participants reported regularly transcribing such notes onto another medium, suggesting transfer as an especially important stage of the information scrap lifecycle.

Professional fields often encourage similar notetaking practices in the workplace. Paper engineering logbooks, long a common practice for professional engineers to use for recording notes and ideas, were found to commonly serve as reminders of work in progress and as a personal work record for future reference [McAlpine et al. 2006]. Meeting notes created by professional information workers contain a large number of facts (e.g., names, phone numbers, technical details, and procedures) and action items [Khan 1993]. The degree to which these practices translate across other logbook-intense professions such as the sciences [schraefel et al. 2004] is not yet clear.

Studies of e-mail use have revealed a wide range of behaviors from information keeping to collaboration and coordination, leading to characterizations of e-mail as a habitat in which we embed much of our personal information [Ducheneaut and Bellotti 2001]. As a result of this embedding, we see e-mail used for a variety of purposes in common with information scraps. E-mails are deliberately marked as unread or left unorganized in the inbox to serve as reminders or to-dos, and half-completed messages are saved along with notes for what to include [Bellotti et al. 2005]. Venolia et al. suggest that such coping strategies are due to the sheer volume of incoming messages [Venolia et al. 2001]. Whittaker and Sidner's early study found that 35% of folders contained only one or two e-mails [Whittaker and Sidner 1996], suggesting that many of these e-mails may have had no natural application and required small, artificial homes to be created. More recently, we have learned that nearly a third of all archived e-mail is actually sent by the owner to him- or herself [Fisher et al. 2006, Jones et al. 2005a] – another common information scrap pattern.

The ubiquity of to-dos, scattered in unorganized locations across the physical and virtual workspace, suggests that they might constitute a particularly common form of information scrap. Bellotti et al. undertook the most rigorous investigation of to-do practice to date [Bellotti et al. 2004]. Their findings suggest that to-dos are created by expending as little effort as necessary, and "only elaborated enough to provide a salient clue" to the original author (e.g., a to-do containing only the text "Joe the attorney"). To-dos are often integrated as resources into ongoing work, incorporating state or links to other artifacts. Bellotti et al.'s investigation uncovered a large number of separate tools (average 11.25 per person) being used to manage to-dos, noting that often to-dos are intentionally placed in the way of a typical routine to promote visibility, rather than in usual filing schemes. These fragments or notes are very much in keeping with our definition of information scraps: they are deliberately not kept in an application like a to-do list; they are in a specific "elsewhere." These locations might include the backs of hands, scraps of paper, unstructured text files, and post-its [Blandford and Green 2001].

Calendaring tools also display many of the characteristics of information scrap work. Users keep a plethora of non-appointment (but still time-based) information scraps in their calendaring tools: notes of which week of the semester it is, pointers from a diary entry to supporting materials, reminders, reports of how time was actually spent, and notes of prospective but not finalized events, among others [Blandford and Green 2001]. In their studies of e-mail and task management, Bellotti et al. also noted that participants would create calendar events as reminders [Bellotti et al. 2005].

The consideration of web and Internet material raises further issues. The Keeping Found Things Found project has investigated the means by which users keep web information [Bruce et al. 2004, Jones et al. 2001, Jones et al. 2005a]. Participants in these studies often captured information on the web using tools outside of browsers' native bookmarking and history facilities, instead e-mailing themselves URLs with comments, saving web pages to disk, or printing out information and filing away the hard copy.

In their ubiquity, camera phones have recently become a popular mechanism for information scrap collection. Ito describes the camera phone's ability to add meaning to the mundane objects in our lives, for example with photographs of the seashell we found on the beach, the street sign that will allow us to re-find a restaurant, and other objects that are simply interesting in some way [Ito and Okabe 2003]. Such one-off photos may fall into the domain of information scraps as they exist outside the domain of typical personal information categories, may serve unclear purposes, and may be difficult to categorize. These scrap pictures may be numerous, as well: images captured for personal reflection or reminiscence were the most numerous of those indexed by Kindberg et al. [2004].

2.3 Organizational Practice

Studies of physical document organization and information workers' offices have revealed several behaviors common to information scraps, including an aversion to filing and an affinity for paper media. Malone's seminal paper on office organization examined the existence of unorganized *piles* in office work [Malone 1983]. He found that piles served as reminders of unfinished tasks and lessened the cognitive effort associated with filing documents. Whittaker and Hirschberg [2001] discovered that *working notes* for current projects constituted 17% of the paper archives maintained across an office move – many of these working notes were handwritten and irreplaceable, likely containing many information scraps (e.g., meeting notes and brainstorms). In *The Myth of the Paperless Office*, Sellen and Harper detail numerous reasons for the continued prevalence of paper in the workplace, including ease of annotation, flexible navigation, spatial reorientability and support for collaboration [Sellen and Harper 2003]. These affordances align with many common information scrap needs.

Other studies have revealed similar organizational tendencies for information stored in digital tools. In parallel to Whittaker and Hirschberg's description of working papers, Barreau and Nardi detail what they term *ephemeral information* – that which has a short shelf life. Many information scraps exhibit the characteristics of ephemeral information: they are loosely filed or not filed at all, and difficult to manage in large quantities [Barreau and Nardi 1995]. Boardman and Sasse noted that their participants tended to combine filing and piling strategies based on item priority, regularly filing items of high perceived value but otherwise leaving their collections to spring cleaning or no organization at all [Boardman and Sasse 2004]. Boardman and Sasse further reported that 3% of files, 41.6% of e-mail, and 38.8% of bookmarks remained unfiled over their longitudinal study – again, the forces driving these artifacts to remain unfiled will likely also exist for information scraps.

Digital folder structures often remain ad-hoc and relatively flat. Jones investigated folders in the service of ongoing projects [Jones et al. 2005c], and found that folders' semantics were continually adapted to reflect each participant's "evolving understanding of a project and its components." This result may hold for information scraps as well, whose boundaries are even less clearly delineated than project folders. Similarly, Barreau and Nardi [Barreau and Nardi 1995] investigated digital file hierarchies and discovered that most participants' hierarchies were unexpectedly shallow due to low perceived future usefulness of complex archives. Rather, digital hierarchies were structured in the service of what they referred to as *location-based finding*: navigating to a directory of interest and proceeding to browse. This result suggests that users preferred to depend on their own cognitive capacities for recognition of documents rather than ensure that everything was elaborately filed. We suspect that this preference may also hold when filing information scraps.

Difficulty categorizing information at the time it is captured has been repeatedly examined as a modulator of how information is written down. Malone's participants complained of the difficulty of accurately filing paper information [Malone 1983]; Bowker and Star expand this point to the digital realm: "A quick scan of one of the author's desktops reveals eight residual categories represented in the various folders of email and papers: 'fun,' 'take back to office,' 'remember to look up,' 'misc.,' 'misc. correspondence,' 'general web information,' 'teaching stuff to do,' and 'to do.' We doubt if this is an unusual degree of disarray or an overly prolific use of the 'none of the above' category so common to standardized tests and surveys." [Bowker and Star 2000 p. 2]

Information fragmentation [Jones 2004] of information scraps occurs across devices, applications, and media. In a cross-tool study, Boardman et al. reported on three consequences of fragmentation: 1) file compartmentalization across tools, 2) lack of ability to coordinate work activity between tools, and 3) inconsistent design vocabularies [Boardman et al. 2003]. Fragmentation leads to undesirable effects such as an inability to gather all information about a single person or topic or to effectively link such data [Karger 2007]. While fragmentation mainly occurs between applications, mobile situations lead to fragmentation instead across devices such as cell phones, laptops, notebooks and other mobile devices [Oulasvirta and Sumari 2007]. We hypothesize that the low-effort, spontaneous capture needs of information scraps lead them to be particularly susceptible to both application and device fragmentation. Data unification approaches have been proposed as potential solutions to these problems [Bergman et al. 2003, Boardman et al. 2003, Jones et al. 2005b, Karger 2007, Karger and Quan 2004].

3. GOALS

In the research presented here, we have targeted a type of personal information with many unanswered questions – one that highlights the ad-hoc, unorganized underbelly of personal information. Our focus is on understanding why information scraps exist, what kinds of information they hold, why they end up in the tool or medium they do, and how they evolve through their lifetime. We entered our study hoping to understand to following:

- **Characterization of the phenomenon.** What is, and is not, an information scrap? Can we improve our intuitive understanding into a more precise characterization of the phenomenon?
- **Type variety.** What kind of data will be encoded in information scraps? How much variety will there be, and which information types will be the most popular?
- **Structure and expression.** How does a calendar item as an information scrap compare to a similar item in a digital calendar such as Outlook? Will the information scrap carry less information, or express it in a different way?
- **Tools.** Information scraps are by definition held in inappropriate or general-purpose tools. What tools are these, and why do we use them? How do we adapt the tools to hold information they may not have been designed to carry? To what extent does fragmentation take place across tools, and does this fragmentation inhibit later re-finding or re-use?
- **User needs.** What needs do information scraps serve? Why are they used in preference to other PIM techniques?

4. METHOD

We conducted a study consisting of 27 semi-structured interviews and artifact examinations of participants' physical and digital information scraps. From our previous work [Bernstein et al. 2007, Van Kleek et al. 2007] it had become clear that information scraps are a fundamentally cross-tool phenomenon, so we chose a cross-tool study inspired by Boardman and Sasse [2004]. This study design allowed us to examine the many locations and tools in which we believed scraps might appear. As our questions primarily surrounded information scrap content, organization, location, and lifecycle, we chose to focus on examining the scraps themselves rather than the capture or retrieval process. Diary studies and experience sampling studies would have also allowed us to record information scraps as they were generated, but we were concerned that the additional time and energy burden on participants would have conflicted with the overriding importance participants place on ease and speed when capturing scraps – participants may have simply chosen not to record the scrap to avoid the effort associated with writing in their diaries.

We carried out the interviews with participants from five different organizations, following a five-person pilot study in our lab. Three of the organizations we visited were information technology firms, focusing on mobile communication, interactive information retrieval, and wireless communication. One was small (start-up), another medium sized, and the third a large international corporation. The fourth organization was an Internet consortium, working internationally in a highly distributed fashion. The fifth was an academic research lab. We interviewed 7 managers (MAN), 7 engineers (ENG), 6 administrative or executive assistants (ADMN), 2 finance workers (FIN), 2 usability professionals (UI), 1 technical writer (WR), 1 campus recruiting officer (REC) and 1 industrial researcher (RES). There were 13 males and 14 females; the median and mode age range was 30-35. Educational level ranged between some college (4), college degree (11) and graduate degree (12). This population was a diverse group of professional knowledge workers with a skew toward those working in information technology.

Interviews were performed at each participant's main computer at his or her typical place of work. For privacy reasons, participants were free to refrain from sharing any particular piece of personal information. During the interviews, we did not use the term *information scrap*; rather, we asked participants to tell us about information that they had that was not formally recorded in a proper place, like a calendar or project folder. Participants then provided us with a stream of examples which we noted as they discussed these exemplar artifacts. Our questions focused on revealing the purpose of the item, the reason it was recorded the way it was, as well as where it fit in any context or process of use.

4.1 Information Scrap Operationalization

As discussed in the introduction, the term *information scrap* can be difficult to define. However, for purposes of internal validity, and lacking a rigorous definition at the time of our study, we required an operationalization of the term that would allow us to identify which artifacts to record. We used this operationalization to decide which artifacts were in scope; we did not generally attempt to convey this definition to the subjects of the study. Throughout the discussion of the study, we refer to an information scrap as a piece of personal information that:

• Is in a tool with no explicit support for that information's schema, or	e.g., a phone number on a Post-it note
• Has no tool specifically designed to handle that kind of information, or	e.g., an application serial number
• Is in a tool that does not seem particularly well suited to the information type.	e.g., a to-do with "where to remind me" information shoehorned into the details field

The following are examples of artifacts that we *excluded* from our capture: e-mail serving a communicative purpose, word processor documents with papers or full essays, and contact information in the computer address book. This definition carries a connotation that what is and is not an information scrap depends on each participant's tools, needs, and practices. Based on the results of this study, we refined our definition to that presented in Section 1.1.

4.2 Triangulation Method

Information scraps are distributed among tools and locations, and strategies vary from person to person. We faced a challenge in our artifact examination – specifically, that we might not uncover some classes of each participant's information scraps. We could canvas the space by asking about all known tools – but what if the participant used a tool we didn't know about, or used a common application in a way we did not think to investigate? We could instead query by location, such as Desktop or Miscellaneous folders – but what if the data lived in an application rather than a folder? We could ask how participants dealt with common scrap types such as how-to guides and URLs – but certainly there would be types we might leave out.

Our solution was to fashion a methodology by which we would look for information scraps along all three axes: tool, location, and type. By exploring along each axis with participants, we would be able to zero in on, or triangulate, the location of an appropriate artifact. We first examined their personal information by tool (which yielded the greatest number of information scraps), followed by location, and finally by type. To our knowledge, this methodology is novel. We began by running a pilot study (5 participants) to generate a broad set of tools, locations, and types that we used as a seed list for our final study. Participants were free to generalize to other tools as they desired. We recorded information scraps that were digital, physical, and mobile. Table I lists a sample of the script we followed.

Table I. The three categories of Tools, Locations and Types characterized the main starting points for our artifact study.

Triangulation Perspective	**Examples**	**What was targeted**
Tools	E-mail	Messages that do not serve communication purposes: e-mails sent to oneself, in the Drafts folder, or archived in the inbox.
	Calendar	Calendar entries that did not correspond to actual events; use of the details field.
	Bookmarks	Bookmarks carrying information beyond just a pointer to a web page – for example, "todo" or "toread" bookmark folders.
	Physical Notebooks	All available data (this location commonly holds information scraps).
	Physical Post-it Notes	All available data (this location commonly holds information scraps).
	Notetaking Applications	All available data (this location commonly holds information scraps).
	Freeform text files	"todo.txt" or "todo.doc" files containing personal notes, to-dos, and other data.
Locations	Computer Desktop	Documents of short-term interest and notes to self.
	Physical Desktop	Freeform notes and documents of short-term interest.
	"Miscellaneous" Folder	Data that was difficult to categorize.
	Office wall and whiteboard	Participant-authored decorations or annotations.
Types	Reminders and To-dos	To-dos not in the to-do manager or that did not fit the to-do manager's schema.
	How-to guides	All examples (no known application to organize this information).
	URLs of interest or quotes from web sites	Examples not held in a bookmarking utility.
	Contact information	Examples not held in a contact utility.
	Notes	All examples (common information scrap).
	Short pieces of data (*e.g.,* phone numbers, passwords, serial numbers, thank-you note lists)	All examples (common information scrap).

As the participant or interviewers pointed out information scraps in each tool, location or information type, the interviewers performed an artifact analysis and a semi-structured interview focused on techniques surrounding the items of interest. One of the interviewers, performing the artifact analysis, probed for as many specific instances of the class of information scrap as feasible given the time constraints. For each artifact, the following were recorded:

- Information type: the information type as described by the participant (e.g., to-do, URL, shopping list).[1]
- Tool: the tool used to author and edit the information scrap.
- Whether the scrap was completely authored by the participant or contained copy/pasted material.
- Whether the scrap contained each of the following types of content:
  - Text: written or typed text or words
  - Photographs
  - Pictorial Drawings: representational illustrations of real-world objects
  - Abstract Drawings: non-representational drawings, doodles, symbols, arrows, annotations, graphs representing relationships or quantitative data

Participants were responsible for identifying items and distinguishing multiple items from each other. The interviewers guided the interview so as to try and get a representative sample of information scrap from a variety of tools, skewing for breadth rather than depth, though we did spend extra time investigating tools with large numbers of scraps. At the conclusion of the study, we consolidated similar information type categories. To consolidate, we began with the specific types recorded by one of the interviewers and verified by the other. The two researchers then acted as coder/aggregators, consolidating types as aggressively as possible without sacrificing the participant's original intent with the scrap.

At the conclusion of the artifact examination the interviewers continued the semi-structured interview, following up on topics of interest. Interviews typically lasted sixty minutes.

---

[1] The data type was recorded as per the participant's primary classification of the artifact, even if it contained multiple data types. Thus, even though many to-dos included names of people, places to be, times of events, and so on, they were nonetheless coded as “to-do” if the participant viewed the overall artifacts as to-dos.

5. RESULTS

5.1 What do Information Scraps Contain?

In our artifact analysis, we coded each of the 533 information scraps for its information type, and then consolidated similar categories. The results can be seen in Table II and Figure 1.

Table II. The number of occurrences of each information type observed in our study. Distinct types on the right are separated by commas; e.g., "8 – Progress Report, Receipts/Confirmations" represents two types each occurring eight times.

Occurrences	**Types with the Given Number of Occurrences**
92	To-Do
44	Meeting Notes
38	Name/Contact Information
25	How-Tos
16	Work-In-Progress
14	File Path/Directory Path/URL
13	Desired Items
12	Login/Password
9	Brainstorm, Calendar/Event Details, Event Notes
8	Progress Report, Receipts/Confirmations
7	Computer Repair Status, Conversation Artifact, Correspondence (Chat), Financial Data, Products Of Interest, Reminder
6	Calendar Or Event List, Correspondence, Debugging Notes
5	Archived E-Mail, Computer Address, Ideas, Pre-Emptive Calendar Scheduling
4	Account Number, Airplane Flight Information, Annotations, Design Layout, People Of Interest, Plans/Goals, Timeline, Airplane Flight Information, Archived Document
3	Agenda, Configuration Settings, Jobs/Classifieds, Math Scratchwork, Project Notes, Shipping Information, Template E-Mail Response, To Read, Whiteboard Capture
2	(Mixed type), Academic Record, Bug List, Change log, Company Organization Chart, Debugging Program Output, Favorite Quote, File Backup, Frequent Flier Information, Hotel Information, Performance Tracking, Room Setup Diagram, Tax Information, Template Text, Time Log, To Share
1	(No Memory of Meaning), Announcement, Application Instructions, Architecture, Archived Document, ASCII Art, Baseball Schedule, Blue Chip Stocks, Book Margin Comments, Book Outline, Calculation Chart, Car Supply Shops, Citation, Class Assignments, Client ID Number, Concert Tickets, Correspondence (E-Mail), Credit Card Information, Deadlines, Decorative Drawing, Definition, Demographic Breakdown, Documentation, E-Mail Lists Of Interest, Employee Desires/Goals, Event Planning, Expense Report, Fantasy Football Lineup, File Transfer, Flow Diagram, Funding Options, Gift Certificate, Guitar Chords, Gym To Join, Insurance Claim, Kayaking Resources, Library Number, Moving Plans, Network Diagram, Newsletter Outline, Notes From Old Job, Parking Location, Part Number, Patent Information, Phone Payment Statistics, Picture Of Car, Picture Of Poster, Pictures Of Team Members, Planned Trip (Map), Presentation, Price List, Project Overview, Public Notice, Puzzle Answers, README File, Rebate UPCs, Recipe, Resume, Room Location, Salary Calculation, Serial Number, Sign Out Sheet, Song Lyrics, Talks Given, Travel Agent, Word To Spell-check

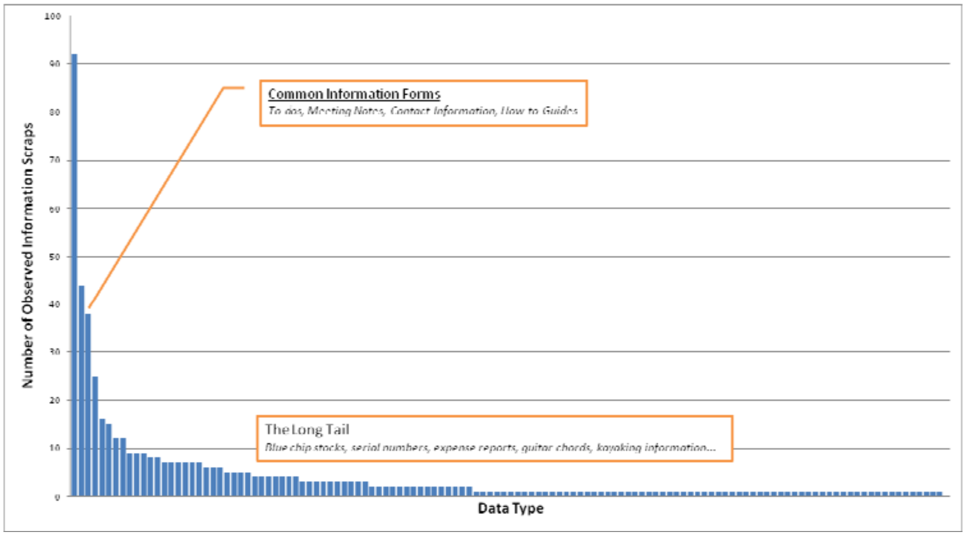

Figure 1. This visualization of Table II demonstrates the prevalence of uncommon types in information scraps. The ordering is identical to that in Table II.

5.1.1 Common information scrap types

The four most common information types we found in scraps were to-dos (92 instances), meeting notes (44 instances), name and contact information (38 instances), and how-to guides (25 instances):

- **To-dos:** Information scraps containing lists of items participants wanted to accomplish. Action items, traditional to-do lists, and other information interpreted as a to-do fell into this category.
- **Meeting Notes:** Notes taken down while the participant was in a meeting or discussion. These ranged from notes taken at formal meetings to hallway conversations.
- **Name and contact information:** Typical contact information such as name, address, phone number or e-mail address.
- **How-to Guides:** Notes containing instructions on how to perform certain tasks, kept for future reference. Examples included UNIX shell command sequences ("incantations"), login procedures for remote servers, instructions on ordering food for meetings/seminars, filing for reimbursements, and common office tasks such as shipping, calling, or faxing internationally.

We note that two of these top four categories (to-dos and contact information) are easily managed by many PIM applications. Section 5.5.1, discussing the capture stage, will give possible reasons why this information still may end up in scrap form.

5.1.2. Diversity of Data Types and the Long Tail

Another compelling view of information scrap forms appears when one focuses on the *least* frequently occurring items. Figure 1 illustrates the frequency of each information type we found within all participants' scraps, ordered from most to least frequent. Immediately noticeable is the fast drop in the histogram after the most common types described above, and the large mass of types with few occurrences. Furthermore, as can be seen in

Table III, the least frequently occurring types (the tail of the distribution) comprised a significant percentage of the information stored in all the scraps; in particular, forms that occurred only once comprised 13% of all scraps; twice or less, 18% of all scraps.

From our limited sample, the distribution of information types appears to follow a discrete power law probability distribution, containing a *long tail* of unpopular types. This correspondence between naturally occurring events and power law distributions is referred to as Zipf's law, and has been noted in several other domains such as the Long Tail of internet sales [Anderson 2006], the Pareto Principle (80-20 rule) in business economics and Bradford's Law in scientific citation patterns. In common with Anderson's conception of power laws in internet sales, we found that the scraps containing rarer forms in our study cumulatively rivaled the number of occurrences of

Table III. Information types that appeared only once make up approximately one eighth of all information scraps. The 50% threshold is crossed at nine occurrences out of 533 recorded information scraps.

Upper Bound on Number of Occurrences of an Information Type	% of all Information Scraps
1	12.8%
2	18.4%
3	24.0%
4	29.3%
5	33.0%
6	36.4%
7	44.3%
8	47.3%
9	52.4%

commonly occurring forms.

We uncovered a large number of rarely-occurring information types, including book wish lists, application serial numbers, expense reports, resumes, guitar chords, and information about kayaking. All of these types could benefit from an application tailored to their particular intended uses (e.g., comparing kayaking in Cambridge and Palo Alto, sorting resumes by years of experience, or finding the most similar expense reports); however, unlike information types like to-dos, for the majority of these rarely occurring types, such applications are either unavailable or unpopular.

In addition to looking at the distribution of types for all scraps as a single group, we also looked at each individual's scrap distributions to gain a sense for comparison. In doing so, we observe a similar distribution for individuals' information scraps, which implies that individuals also use information scraps to keep a large number of infrequently occurring information types; see Figure 2a and Figure 2b.

Figure 2b provides an interesting link between the global (i.e. inter-participant) and individual scrap type distributions: ENG3's most popular item is notes on the repair status of the computers he managed; however, he was the only participant to record such information. Thus, even though computer status notes are in the head of ENG3's individual distribution, they fall to the tail of the global distribution. We observed similar patterns in several other participants: ADMN1 maintained a set of artifacts she needed to share with her superior, ENG2 kept an extensive set of personal progress reports, and MAN6 had a sizable folder of documents to read in his free time. Therefore despite overall distributional similarities, individual differences are clearly visible among the frequencies of types kept by different individuals.

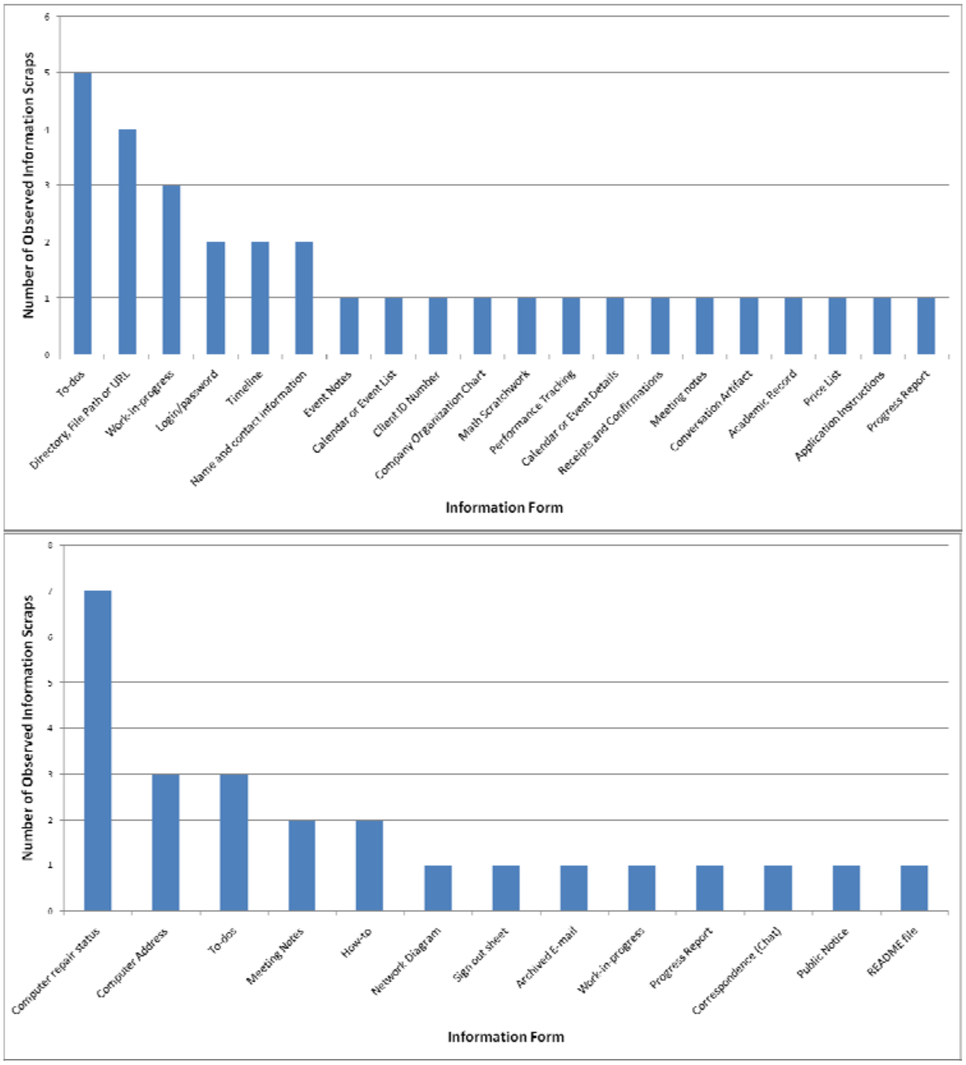

Figure 2. a) FIN1's information form distribution, which also follows a long tail. Here the histogram closely mirrors the accumulated distribution across all participants. b) ENG3's information form distribution, evidencing a large number (7) of computer status note scraps. Though these notes are in the head of his distribution, he was the only participant to collect such data, so when accumulated the computer status notes fall into the long tail.

5.2 Scrap Encoding, Composition and Layout

The predominant method of encoding information in scraps was text, typed or handwritten. Information scraps occasionally contained abstract drawings and annotations (Figure 3). Such drawings included arrows, graphs or timelines, stars, organizational lines and boxes, and markings indicating emphasis. We coded our data to record the number of information scraps that contained text, abstract drawings, pictorial drawings, or photographs. We found 96% of our information scraps to contain some sort of text (95% of the digital scraps, 96% of the physical scraps), 5% to contain abstract drawings (1% digital, 10% physical), 2% to contain pictorial drawings (no digital examples, 4% physical), and 2% to contain actual pictures (4% digital, < 1% physical). In two-proportion z-tests, the differences between digital and physical media for abstract drawings, pictorial drawings, and actual pictures are significant ($p < 0.01$); the difference for text was not significant. A small number of information scraps contained other kinds of media, digital or physical attachments such as laundry receipts.

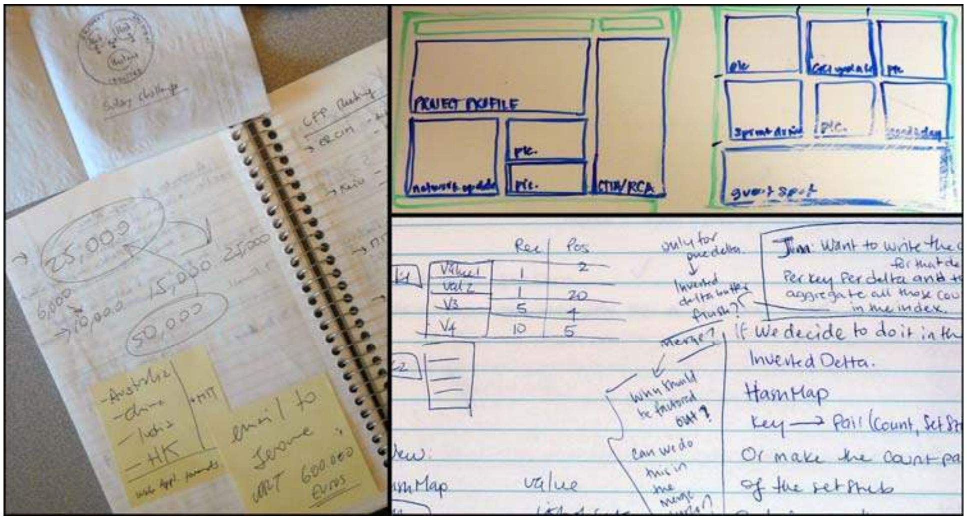

Figure 3. All three of these information scraps contain abstract drawings, including arrows, boxes, and so on.

Due to the varied nature of the types of information scraps, there was a corresponding variation in scraps' elements, such as phone numbers, e-mail addresses, or URLs. It was also common to see several elements intermixed in a single information scrap or several unlike information scraps together in one location – Figure 5b contains a URL, a PIN number, two UNIX commands, and a phone number sequence. The information types were rarely labeled, which occasionally made it difficult for the interviewers to understand the content of some notes (Figure 4).

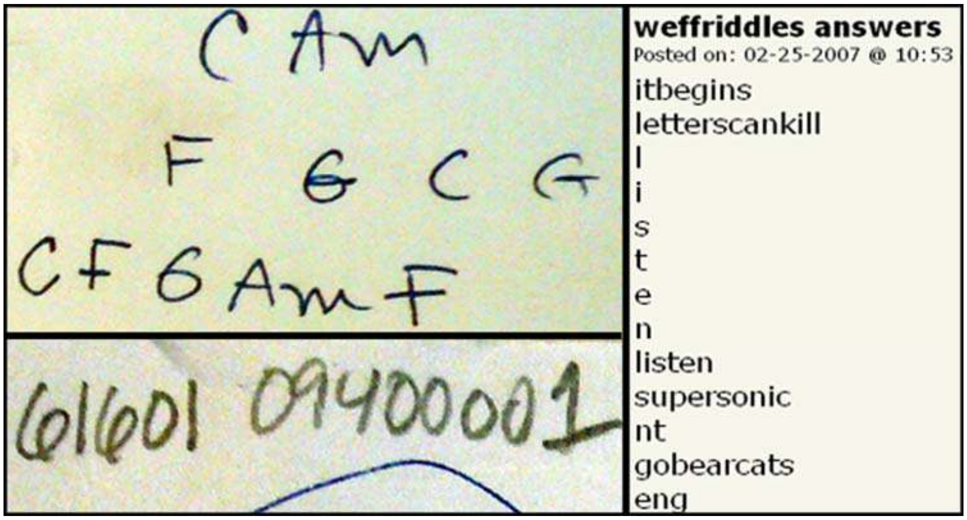

Figure 4. Information scraps containing unusual data. Counterclockwise from upper left: guitar chords, an unknown string of numbers, and answers to an online riddle.

Many examples of artifacts collected contained both incomplete and vague information. Furthermore, the need to capture information in such an incomplete or vague form occasionally impacted how and where something was written. For example, MAN3 explained the calendar event in his to-do manager: "I don't know exactly when [my visitor] will come today…If we'll agree on the details later, I prefer to use a to-do." Information scraps could also capture data with more fields than applications knew how to handle; for example, MAN maintained his own contact list where he could record previous deals and other personal notes on each person he worked with.

While most information scraps were very short (a few words or lines long), we observed several instances of scraps of approximately a handwritten page in length, particularly meeting notes. This result indicates more variation in length than we were originally expecting to see.

Several participants who kept free text files on their computer utilized the ability to mix types or lay out thoughts as they desired – even creating ASCII art in the case of ADMN6. Paper and physical tools were particularly preferred for their encoding flexibility, allowing participants freedom over visual structure and sketching (10% of physical scraps involved some sort of drawing annotation).

5.3 Use of Language in Scrap Text

We found that text written in information scraps used extremely terse language; many scraps consisted exclusively of key words, such as lists of names of people, places or objects, and raw bits of data, such as phone numbers, addresses, passwords, and other strings. Figure 5 gives examples of text used in scraps. Information scraps used as temporary storage locations in particular exhibited short language, listing single words or pairs of noun-object or noun-data value, often omitting the verb or relevant predicate, as well as articles and particles. For notetaking – in a meeting, class or brainstorming – phrase structure was more common. We also noticed a tendency to omit the subject title or description of what the data actually represented. Several participants, when sending e-mails to themselves, intentionally left the subject field blank or wrote something general such as "note to self."

Figure 5. Several of the information scraps we noted, focusing on typical examples of minimal use of language in scraps: a) an envelope with several scraps: guitar chords, stock ticker symbols, e-mail addresses, and an unknown number, b) a web site address, password and helpful SSH commands in a rolodex, c) use of the Outlook notes facility to maintain links of interest and the outline of a blog entry, d) “CFP Meeting, ERCIM, didn't use database #'s, up in [incomplete],” meeting notes, e) a brainstorm on a programming decision, f) “email to Leone w.r.t. 600.000 euros,” a reminder, g) several post-its on the laptop palm rest, reading “XIA HUA,” “ANDREW talk,” and “Gopal's cell #”, h) a reminder written on the participant's hand, a single word, i) text at the top of an e-mail the participant received, condensing it into a few memorable words, j) an annotated, copy/pasted chat transcript detailing a slightly arcane UNIX command.

5.4 Tools and Locations

We noted 51 different tools in use across our investigation, 33 digital and 18 physical. Figure 6 shows the distribution of the number of information scraps we located in each tool or location. Again there is a power law pattern, beginning with a set of extremely popular tools and trailing off to a large number of less popular ones. Participants maintained a small set of main tools for capturing information scraps, supplemented with a large number of less-used auxiliary tools.

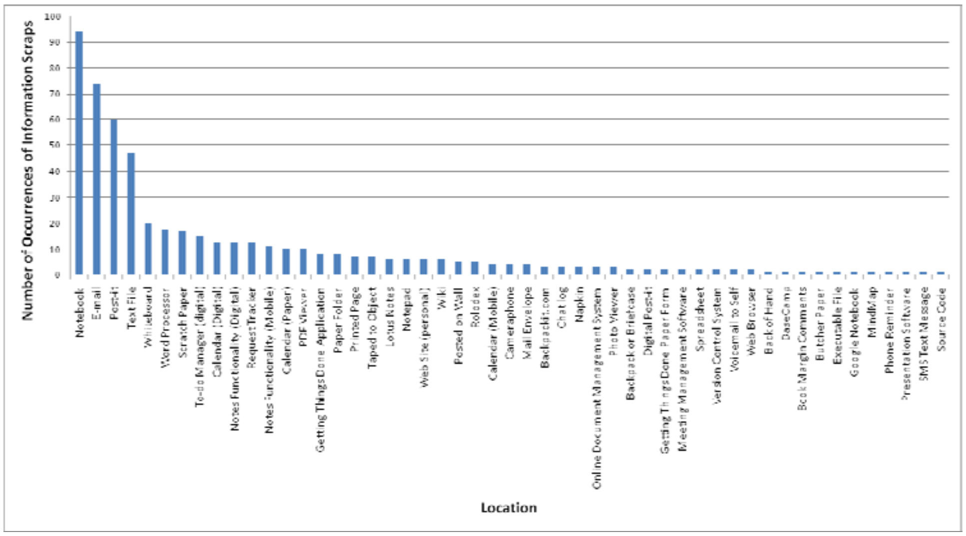

Figure 6. There are a small number of tools and locations used frequently, such as notebooks, e-mail, and post-its, and a large number of locations used a small number of times.

Among digital tools, e-mail was the most often used for recording information scraps (74 instances, 26.4% of digital scraps), followed by a text editor (47 instances, 16.8%; *e.g.*, TextEdit or Notepad) and word processor (27 instances, 6.4%; e.g., Microsoft Word). Text editors were preferred over word processors for being less complete or formal, particularly in early drafts of work. Several participants used text files to keep separate collections of contact information: for example, for business clients versus family and friends. In addition, some participants used text files to keep contacts because it was easier to add notes about particular contacts, such as the history of business negotiation with a particular contact, or the names of the contact member's spouses and family members.

In the physical world, paper notebooks (94 instances, 37.2% of physical scraps) and Post-its (60 instances, 23.7%) were the most popular choices. Participants reported that paper notebooks were often an appropriate choice because they were portable and more socially acceptable in face-to-face meeting settings. Thus, paper notebooks were the most popular tool for meeting note-keeping, and physical meeting notes were three times as common as digital meeting notes.

5.4.1 Physical/Digital Divide

Overall, there was an approximate parity in the number of physical (253, 47.47%) and digital (280, 52.53%) information scraps we gathered. However, this statistic is slightly misleading, as participants adopted widely different strategies. Examining the relationship between participant and percentage of scraps kept in digital vs. physical form, a chi-squared test rejects the null hypothesis (*p* < *0.01*). Figure 7 indicates how this dependence might have arisen: there is a bimodal distribution with most participants centered at the 50% mark and a smaller group being almost entirely digital. These digital participants tended to be technophiles or mobile workers, including two managers, an engineer, an administrative assistant and a research scientist. The existence of almost-digital practice is a somewhat surprising conclusion given previous research's claims that paper is still an overriding favorite for many information scraps [Campbell and Maglio 2003, Lin et al. 2004, Sellen and Harper 2003].

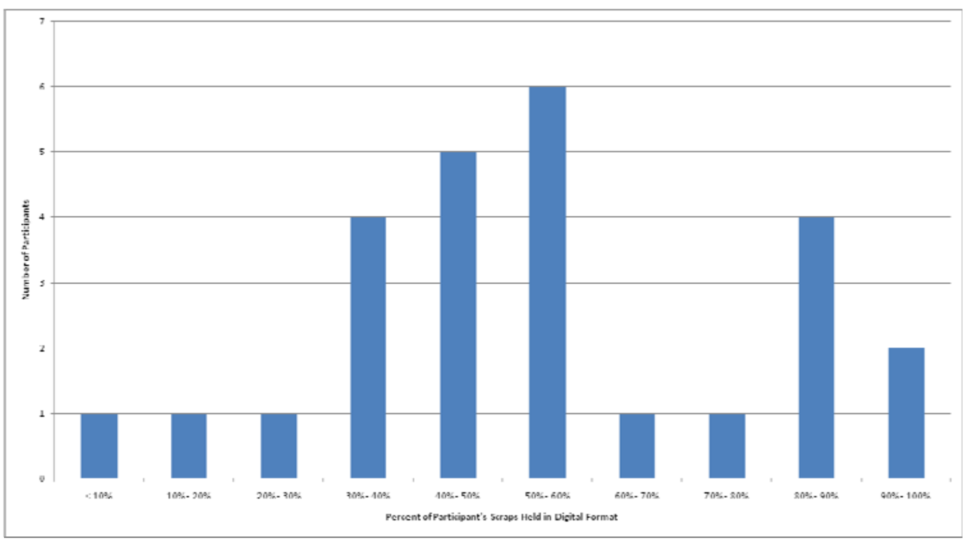

Figure 7. Examining the number of participants at each level of digital data, we see two groups: one centered around half digital, half physical, and the other almost completely digital.

5.4.2 Mobility

When the scenario called for information workers to go mobile, participants often generated information scraps to carry important data around or to capture information as events occurred. A small number of information scraps (22 instances) were in mobile digital form: primarily smartphones, but also PDAs, SMS messages and camera phone pictures. ENG7 and MAN7 in particular used smartphones heavily for capture, and relied on synchronizing functionality with their desktops. Though we were unable to note which physical information scraps were used in mobile scenarios and which were not, our interviews suggested that paper information scraps were particularly useful when mobile. UI2 is an illustrative example of a digital smartphone user: she described how she would reference a note file on her smartphone with relevant phone numbers, and was likely to send herself a voicemail or add a smartphone note file if the situation required. In addition to mobile scenarios, social constraints came into play: when laptops were not socially appropriate at meetings, paper notebooks were used instead.

5.4.3 Tool Adaptation

Many information scraps revealed ways in which participants adapted tools to better serve their purposes. Post-Its provided the best examples of this behavior. We observed Post-its adapted to deliver contextually-relevant information by being stuck in the places or to the physical objects to which they referred: in Figure 8, ENG3 affixed tape labels directly to the computers he was attempting to annotate. Post-its were also used as bookmarks in aid of re-finding and placed on the back of a cell phone to act as an extension of the device's note-taking facilities. Interviews revealed that Post-its were well-loved information tools; the wide variety of creative adaptation behaviors support this perception.

We also observed adaptation and reappropriation in digital tools. For example, ENG4 used a popular web-based software bug tracking tool as his personal to-do list because it afforded an organizational principle that he liked (individual issues as commitments with deadlines), he could access it from anywhere, and because he could easily update his list of commitments by emailing the system. E-mail was also often reappropriated, most commonly by participants seeking to archive information by sending themselves messages.

Annotation and revision were also quite common – documents were annotated with comments on what the recipient should do, calendar events contained explanatory notes, and last-minute amendments were appended to agendas. For example, ADMN5 printed out the day's schedule for her supervisor (produced using a calendaring tool) and marked it up with physical notes.

Figure 8. ENG3 wrote notes on masking tape, then affixed the information directly to the computers of interest.

5.5 The Information Scrap Lifecycle

In this section we build on Lin et al.'s micronote lifecycle [Lin et al. 2004], revisiting our results using the lens of the information scrap lifecycle. Here, we discuss Capture, Transfer, Organization, and Re-use:

- **Capture:** the process of translating a thought into a physical or digital information scrap.
- **Transfer:** optionally translating an information scrap from one form into another, either to put it in a more permanent form or enable mobility.
- **Organization:** the addition of structures and metadata to aid re-finding of scraps later.
- **Re-use – Reference, Retrieval and Recall:** the need to re-find scraps (reference), the process of re-finding those scraps (retrieval), and memory for scrap contents (recall).

5.5.1 Capture

We observed three major sources of information scraps: directly authored material, automatically archived material, and copy/pasted material. Directly authored material (the most common) was intentionally written in an effort to record information. Indirectly authored material consisted of scraps that were created as the result of some external action not initiated by the participant, such as receiving an e-mail or paper correspondence from someone else, and then explicitly kept by the participant. Thus, e-mails received and then saved in a "Miscellaneous" directory constituted indirectly authored material. In our interviews, copy/pasted information included examples such as photocopies of a credit card in a notebook, pieces of an online FAQ pasted into a text file, and internet chat transcripts manually saved. We coded our data to examine how often participants included any material they did not directly author in their information scraps. We found that 113 of the 533 scraps (21.20%) overall contained portions copied from other sources, breaking down as 28.93% of the digital scraps and 14.48% of the physical scraps. In a two-proportion z-test, this difference between physical and digital was significant ($p < 0.01$).

Among the directly authored material, the most commonly cited situation prompting information scrap creation was the need to write something down quickly before it was forgotten [Hayes et al. 2003]. To MAN6, writing information down quickly was essential to keeping: "Mind like water," he explained, is a critical component of the Getting Things Done approach to workflow organization [Allen 2001]. Others also reported how offloading information from the mind and into an information scrap freed them to focus on their primary task, such as holding a conversation, paying attention to a meeting, or even driving the car (as with MAN7).

Capture speed was one of the most important determiners of the tool participants chose to use. Even seemingly minor difficulties or annoyances with tools could deter use of a tool. "If it takes three clicks to get it down, it's easier to e-mail," FIN1 explained. MAN3 would write notes on Post-its and stick them to his cellular phone to transfer into Outlook later rather than enter the data directly into his smartphone, even though the phone supported note synchronization. When asked why not enter the note digitally in the first pass, he responded, "Starting in Outlook forces me to make a type assignment, assign a category, set a deadline, and more; that takes too much work!" Similarly, paper notebooks were often chosen instead of laptops because they required no time to boot up. The effect is similar to the one described with mobile applications by Oulasvirta and Sumari [2007] and with organic/digital memory tradeoffs by Kalnikaité and Whittaker [2007].

Even when data was implicitly structured, such as with calendar events, participants chose the faster, structureless route of recording a scrap: for example, plain text such as "mtg @ 5pm in cafe." In interviews, participants explained that entering the data into a structured form or application could often double or triple the time it would take to simply type the information. Thus, there was a tension between the desire to capture the information quickly and the desire for rich representation and structure, often later achieved via the transfer process.

5.5.2 Transfer

Transfer, the process of moving an information scrap from one medium to another after it has been captured, only occurred for a small proportion of the information scraps we observed. Participants explained that the scraps that were transferred often held some particular importance. In particular, we discovered three major reasons for initiating transfer. The first was to transform and re-interpret the information to fill in incomplete details, making the notes appropriate for consumption by others or for permanent archiving. For example, MAN4 religiously transferred all of his handwritten meeting notes into e-mails to “fill in the gaps” and make the notes “sixty-day proof” (ensuring they would be understandable sixty days later). Second, transfer occurred when information was ready to migrate from a work-in-progress state to a more complete representation and needed a tool that offered additional functionality. Third was mobility: scraps were sometimes transferred onto other media to carry to another room, or sent in e-mail so that it would be retrievable from home.

5.5.3 Organization

Among information scraps that were archived, techniques varied; some consolidated information scraps of similar types/purposes, while others situated scraps with others that were created at the same time and thus created a chronological ordering. ENG1 in particular maintained three text files corresponding to three different types of how-tos accumulated over several years.

Several participants expressed difficulty filing information scraps accurately. This effect was especially powerful for single, unique thoughts: as ENG3 jokingly complained, “where *would* you put the last two octets of a MAC address?” REC concurred: “It's too much work to decide which section it should go in – because sometimes things don't fit in just one, or fit in multiple places. It's hard to decide what to do.” When such difficulties occurred, participants reported dedicating areas to unorganized information scrap collection, ranging from a special e-mail folder, “Miscellaneous” file folders, misc.txt text files, or a catch-all notebook.

We also noted that participants often copied information into multiple places to circumvent application limitations or fragmentation, resulting in replication across digital tools. For example, several participants copied contents from e-mails, wikis, bug tracking tools and groupware into their to-do list management tool or calendar. Participants reported this behavior was an effective coping mechanism for linking information from one tool into another which better fit their workflow. For example, ENG4 pasted emails into job tickets and summarized them in one line at the top (Figure 5i), because he wanted to keep all of the information relevant to an outstanding ticket in one location. Similarly, ADMN5 copied relevant email threads into calendar note fields when reminding her superior of a meeting so that the superior would be able to reconstruct the context and purpose of the meeting.

5.5.4 Re-use: Reference, Retrieval, and Recall

Participants reported that few of their scraps were actually referenced regularly. One group of information scraps, typified by the to-do list on a Post-it, was referenced actively until its usefulness was exhausted, and then was either archived or thrown away. A second group, including meeting notes, was archived immediately without a period of active reference. After archiving the scraps, participants reported not needing them except for special occasions.

Because our study methodology directly located the information scraps, we did not rigorously examine re-finding techniques; however, participants often spontaneously recalled the existence of a particular scrap and we could observe as they located it for us. When re-finding, participants used a technique similar to the orienteering behavior described by O'Day and Jeffries [1993] and Teevan et al. [2004]: direct navigation to a folder location thought to be relatively close to the desired information scrap, then small local steps to explore the results and their neighbors.

We found that participants exhibited good memory for the meaning of almost all of scraps we uncovered. Out of the 533 information scraps indexed, only one was ultimately left with the participant unable to identify its meaning. We did not investigate memory for the meaning of specific details in each information scrap, only focusing on the gist; specific details would likely have fared more poorly, due to human memory for meaning outperforming memory for details.

5.6 The Psychopathology of Information Scraps

During our interviews, we encountered a series of affective and psychological dimensions surrounding participants' perceptions of their own information scrap practice.

There was often perceived social pressure to be an organized individual; admitting to the existence of information scraps ran directly counter to this perception. Similarly to as recounted by Boardman and Sasse [2004], participants often began the interview proud to demonstrate their complex personal information solutions. However, when the interviewers began to inquire after those pieces of the workspace that were unorganized, the same participants would often become uncomfortable and embarrassed, much like Bellotti and Smith [2000] recount. Our participants reported:

- "I would like to have time to organize what I've captured, but this never happens." DOC adds that she wishes she were an adherent to the Getting Things Done methodology [Allen 2001].
- "By Friday, no stickies and no papers on the desk," UI2 preemptively excused the existence of information scraps on her desk.
- "Ideally, I wouldn't need this anymore," ADMN2 says of the disorganized notebook in which she keeps all of her important information. (Clearly, it's not going anywhere.)

Several participants apologized for the state of their office or computer desktop.

Exceptions to the embarrassment trend were found both in participants who put extra time into organizing their lives, colloquially known as *lifehackers* [Trapani 2007], and those who simply embraced the mess. MAN6, as a Getting Things Done devotee [Allen 2001], was quite proud to demonstrate his array of tools. Several participants expressed pride in keeping a tight reign over their information scraps. In contrast were those who had simply accepted that their lives would be messy; ENG1 repeated to the researchers what had become an affirmation of her love for a messy notebook: "It's <u>OK</u> to have a notebook!"

Often participants were forced into a cognitive dissonance between their perceptions of themselves as organized individuals and the messy reality of their lives at the time of the interview. UI2 described her regimen of "always" transferring all her Post-it notes into Outlook tasks, but when we noted that there were several such notes that had remained untransferred for some time, the response was defensive: "I've been too busy lately." Believing oneself to be an expert re-finder of information scraps also seemed *de rigueur*. When asked about problems participants might have re-finding information scraps later, responses were curt:

- “I just remember.” (ADMN3)
- “Generally, I remember where things are.” (RES)
- “I remember things.” (MAN5)

In contrast to these reports, participants usually spent considerable time while we observed them trying to re-find scraps they wanted to share with us. Many participants thus overestimated their memory for scrap locations.

6. ANALYSIS

6.1 Common Information Scrap Roles

We have consolidated a list of common information scrap roles (Figure 9):

Temporary Storage. Information scraps' small, discardable presence enabled their common use as temporary storage or exosomatic short-term memory. ADMN2 kept Post-it notes on her laptop palm rest for just this purpose, recording visitors' names and contact information later to be disposed of. Mobility was often coincident with temporary storage, for example directions written on a Post-it prior to a drive. Temporary storage information scraps were self-regulating in number, as they tended to be thrown away or to disappear in piles quickly after creation. Participants often expected this graceful degradation from temporary storage scraps.

Cognitive Support. Our participants shared with us many incomplete works-in-progress such as half-written emails, ideas for business plans, brainstorms, and interface designs – scraps used to aid the process of thought. “Before I put anything in the computer, I like to put it on the whiteboard first,” ADMN4 explained of her newsletter layout design process. Information scraps served cognitive support roles because scrap creation tools supported, and even encouraged, messy information work.

Two such support functions were epistemic action [Kirsh and Maglio 1994] and external cognition. Participants used information scraps in support of epistemic action to manipulate and reflect on external manifestations of their ideas, generating alternatives, refinements and elaborations. Information scraps also served as external cognition, enabling our participants to offload difficult thought processes onto the scrap. We observed a wide variety of information scraps being utilized in external cognition roles – for example, ENG2 and ENG5's in-progress notes taken down while debugging.

Archiving. In contrast to temporary storage scraps, which were intended to have short lifetimes, many information scraps were intended to hold on to important information reliably for long periods of time. For example, many participants used information scraps to archive notes from meetings – as well as information they could not rely on themselves to remember such as passwords. Several participants emphasized the importance of knowing that the information had been safely saved.

Reminding. Pre-emptive reminding was an important element of several types of information scraps collected, most commonly to-dos and calendaring (event) information. Participants preferred simple, reliable approaches to such reminding: they took advantage of information scraps' visibility and mobility by placing them in the way of future movements to create reminders. Colored Post-its, unread or unfiled e-mails and files left on the desktop reminded participants to take action or to return to a piece of information at a later date.

Unusual Information Types. This role was a catch-all for personal information that did not quite fit into existing tools. Taking advantage of information scraps' freeform nature, participants corralled unique information types that might have otherwise remained unmanaged. For example, ENG3 created an information scrap system to manage a library-style checkout for his privately owned construction tools, and MAN4 maintained a complex document of contact information annotated with private notes on clients. This role was particularly prominent in situations where the information did not match existing tools' schemas, such as calendar items with a date but no start time chosen.

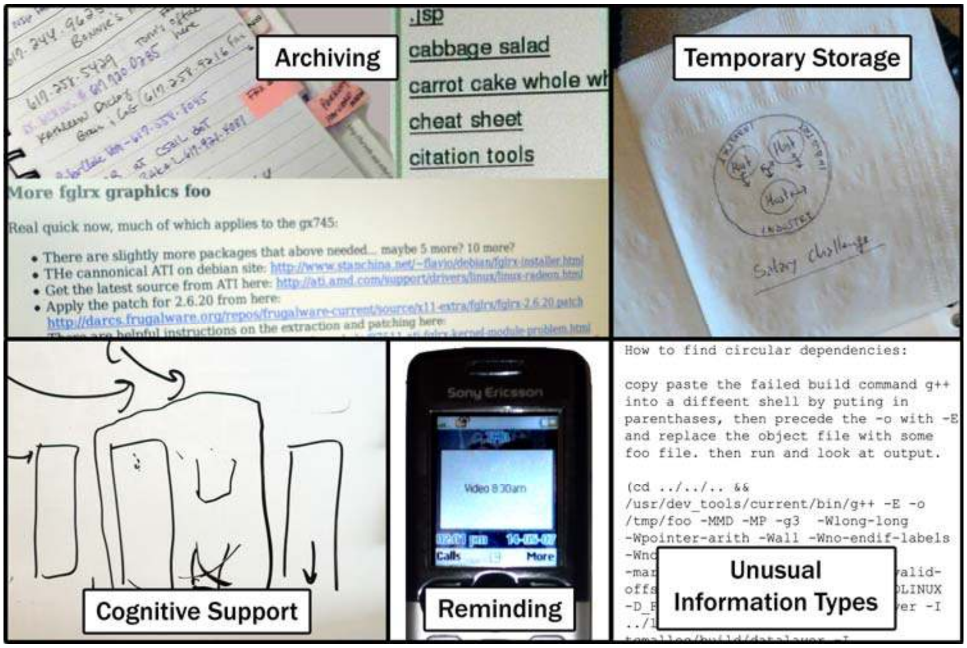

Figure 9. The five main roles information scraps played were archiving, temporary storage, work-in-progress, reminding, and a place to put information that wouldn't fit elsewhere.

6.2 Organization and Fragmentation

We may characterize information scraps of each type by the amount of effort that participants have invested in organizing them. We use web site passwords as illustrative examples:

- **Low Invested Effort:** scraps that are fragmented, unique, or separate from similar data. This includes information of a type that has not recurred often enough to warrant collection in a specialized repository, or temporary information such as might be encoded on a Post-it.

  Example: Web site passwords are archived using whatever is handy: e-mails, Post-its, or text files.

- **Medium Invested Effort:** information types with many instances archived together, but that remain unorganized.

  Example: The user has made sure that all of his or her passwords are archived somewhere in the e-mail inbox, though they are not tagged or filed in any consistent way.

- **High Invested Effort:** information types with many instances archived and organized.

Example: Passwords are all kept in the e-mail inbox in a special folder called "passwords."

Of the scraps we collected, a large proportion exhibited low effort – once captured, they were allowed to remain where placed. This placement was usually dictated by convenience of capture. In notebooks or text files, this pattern resulted in a chronological stream of scraps as new scraps were simply added to the beginning or end. For e-mail, participants seemed to leave scrap e-mails (*e.g.,* e-mails to self) in their inbox rather than filing them away in a sub-folder. Our results support earlier observations regarding engineers' lack of organization in their logbooks [McAlpine et al. 2006].

Fragmentation arose from participants' voluntary placement of scrap information in different places. The primary reason participants cited for writing information of the same type at different locations was convenience at time of capture. For instance, ADMN2 kept contact information (names, phone numbers, addresses) in her main notebook, a paper desk calendar, and a mini address book; and reported that where any piece of contact information ended up was determined by the location of the closest notepad. When asked about retrieval, she reported having to "rummage around" when she didn't remember where something was placed, that this often took time, and that she re-copied contact information between locations so that it would later be more easily found.

6.3 Constraints

Information scrap practice depends on particular physical, temporal, social or structural conditions of a situation that may necessitate tool use. We suggest that these conditions play a particular role in information scrap generation that may not be as evident in tools that have a stronger correlation between task and tool. One such strong correlation is a personal finance task: we have an application for balancing checkbooks, and are fairly indifferent to the physical environment, time constraints, and social conditions while we use it. There are also certain agreed social conventions around these tasks. Paying bills is usually a primary task in terms of one's attention – it would be unusual to pay bills while meeting with a colleague. As such, with such strong conventions and tool support around a particular practice, we do not see information scrap challenges in capture, storage and retrieval.

If we were to instead jot down thoughts while meeting with a colleague about a paper, the temporal, physical and social constraints would have more of an effect on our choice of tool. For instance, it may be socially taboo to be seen either using a computer in this context, or to take lengthy notes during the meeting. Either condition may predicate quick gestures on the back of a note card. The social conditions of an exchange may be such that it would break the flow of the conversation to reach for a more formal mechanism than a scrap of paper. Likewise, when mobile, more formal mechanisms for recording data may both be physically awkward and require too much engagement and time. So, currently, physical, temporal, social and structural factors have particular bearing on the devices selected and the kind of input generated with information scraps.

6.4 Caveats in Our Findings

The boundary between information scraps and the rest of our personal information remains fuzzy, especially in the case of high organizational effort. Once an amateur chef decides to collect all of her favorite recipes on a blog and tag them by cuisine type, are the individual recipes still information scraps? In one sense, no – she has devised her own organizational scheme for his or her recipes and appropriated a generic tool to support this scheme. On the other hand, a generic tool may lack some of the capabilities that recipe-specific tool might provide, thereby limiting its usefulness.

With regard to the variety of information types we catalogued, one difficulty with analyzing the frequency distribution of information types in scraps lies in drawing distinctions between similar information types. It is inevitable to question whether the categories we list could have been combined further, and thus the diversity lessened or erased. Our approach has been to group categories as aggressively as possible without losing the essence of the scrap's composition, attempting to find a rough lower bound on the strength of this long tail effect. A less aggressive grouping strategy produced results where ~25% of all scraps were unique types. Though individual pairs of categories might be further merged, we believe that the long tail effect is quite strong and worth noting.

Reflecting on our methodology, a clear limitation of our study stemmed from our use of interview and artifact analysis instead of live observation via shadowing. We were thus unable to study how information scraps were used and created *in situ*, but instead only how people reported they used their scraps, along with evidence from their workspace and the physical and digital artifacts we collected. We were unable to capture the number of information scraps participants generated each day or significant contextual information surrounding capture and retrieval. We observed that participants often exhibited a kind of confirmation bias toward their own organizational skills by mainly acknowledging well-organized work, so it is further possible that this also affected our observations; for example, our participants may have ignored particularly embarrassing examples of disorganization. The methodology's strength was that it allowed us to observe a broad number of information scraps, perhaps more so than would have been possible *in situ*. We believe that an ethnographic shadowing methodology would complement ours well; it could objectively investigate many of the questions we could not.

Our triangulation method for locating information scraps was also not a perfect lens. We found that it was successful in unearthing a large number of information scraps in a variety of locations. On several occasions, only the last of the three dimensions we attempted (tool, location, and then type) successfully located a particular scrap. The triangulation method's strength lies in unearthing a wide variety of information; its weakness is that the number of information scraps found can be too numerous to examine thoroughly within the allotted time. In the future, we suggest the triangulation methodology might be modified to serve as a fast "tour" through a participant's information scrap landscape, allowing the investigators to then choose a small number of tools or locations to focus on.

7. IMPLICATIONS FOR DESIGN

In this section, we ask: what design affordances would enable personal information tools to better serve these important and underserved roles? How realistically can we expect tools, built using current technology, to support such affordances? We relate this discussion to our exploration of interface design ideas that we built into Jourknow [Van Kleek et al. 2007], a prototype note-taking tool in which we sought to better support information scrap activity (Figure 10).

Table IV outlines the set of affordances we have identified, including lightweight capture, flexible content, flexible organization, visibility and reminding, and mobility and availability. In Table IV we have contextualized each affordance in terms of the activities and constraints that influence scrap generation.

Table IV. The major design affordances necessary for management of information scraps derive from access, scraps' contents, and organization.

Category	**Observed Behaviors and Constraints**	**Derived Design Needs**
Access	Finding the easiest and fastest tool to use. Constrained by effort required, limited time and attentional resources. "If it takes three clicks to get it down, it's easier to e-mail." - FIN2	*Lightweight Capture:* Record information with minimal effort or distraction. Section 7.1
	Adapting tool use to physical locations and social situations. Constrained by available tools and social norms. "At off-site meetings, I don't have my infrastructure there and keeping notes on paper is less rude." - MAN1	*Mobility and Availability:* different capture methods may be necessary in different situations. Section 7.5
	Information scraps kept always in peripheral view, or in a location to be tripped over the next time the information would be relevant. Constrained by tools' ability to be placed in-the-way. "If it's not in my face, I'll forget it. Like if it's on the wiki – I have no idea it's there." - REC	*Visibility and Reminding:* information appears in the right place, at the right time. Section 7.4
Content	Scribbling, sketching, and annotation; in-progress, vague and underspecified information. Constrained by tools' expressiveness. "Drawing is the way you really see it!" - ADMN4	*Flexible Content:* record any kind of data, at any level of completeness. Section 7.2
	Coping strategies when information does not fit other applications' models, and collections of unusual information types. "There's this problem: I wanted to assign dates to notes, but Outlook would only allow dates on tasks." - MAN3	*Flexible Schema:* information may not fit existing molds. Section 7.2
Organization	Organizational strategies varying in degrees from completely disorganized to carefully filed, and avoidance of cognitively difficult filing decisions. "It's too much work to decide which section it should go in – because sometimes things don't fit in just one, or fit in multiple places. It's hard to decide what to do." - REC	*Flexible Categories:* Support for a variety of organizational strategies, as well as for transforming unfiled items into more structured, organized forms. Section 7.3
	Information scraps attached to or placed near related items. "I can never remember which [computer] is which. So I grabbed the gaffer's tape and marked them!" - ENG3	*Flexible Linkage:* enable information to be linked to and in view with arbitrary other information Section 7.3

7.1 Lightweight Capture

As described in Section 5.5.2, we found that participants often generated information scraps in response to a need to capture data quickly. This need occurred most commonly while the individual was performing some other attentionally, socially, or physically engaging primary task. For this reason, lowering both the actual and perceived cost of cognitive, social and physical effort may improve our tools. We see the following opportunities for reducing effort required during capture:

Avoiding upfront decisions and postponing disambiguation. Since information scraps are often captured at a moment when time, attention and cognitive resources are scarce, requiring individuals to make significant upfront decisions might incur sufficient cost to impede capture. Such decisions may include forcing the categorization of a new piece of information, choosing a reminder time, or setting parameters ultimately unimportant to the captured information. Thus, tools might aim to immediately handle information in whatever form provided to them by the user and postpone forcing user choices until a more appropriate time.

Avoiding task-switching, cognitive and navigational burdens. Navigational and cognitive costs associated with launching or switching applications contribute significantly to the time elapsed between the moment an individual forms an intention of writing something down and the moment they can actually start doing so. Since perceived time and effort during this critical interval have been found to dictate which tool will be chosen [Gray and BoehmDavis 2000, Gray and Fu 2001, Kalnikaité and Whittaker 2007], we believe that minimizing navigational effort will improve a tool's capture rate and therefore overall usefulness to the user.

Supporting abbreviated expression of information. As described in Section 5.3, we found idiosyncrasies in participants' language – notes often represented very little explicitly and instead served as memory primers. Our finding contrasts considerably with most PIM applications' requirements that users complete forms with formal expressions for properties such as who, when and where. Our belief is that tools can lessen the time and effort associated with entering information by supporting incomplete, informal capture methods.

Supporting diversity. If a tool is restrictive about the information forms it will accept, individuals will inevitably resort to a coping strategy – either imperfectly fitting the information into the tool, or fragmenting information by encoding it in another tool. Since coping strategies incur non-zero costs to devise and implement, and further lead to decreased effectiveness of future retrieval, we believe it worthwhile to accommodate whatever information the user wishes to express. We discuss further issues with supporting diverse information forms in Section 7.2.

Capture from other tools. Given that a significant portion of the artifacts we examined originated from other applications and devices, (*e.g.*, mobile phone, emails, web pages, IM conversations), we may reduce the need to create scraps by making it easy to select and pipe the relevant information from any application into an appropriate place. This desire also pertains to the need for ubiquitous availability of capture tools, discussed in Section 7.6.

Tablet-based notetaking tools such as Microsoft OneNote [Microsoft] have granted digital note-taking some of the expressive freedom of paper and pen. OneNote also allows users to categorize their notes post-hoc, such as by tagging to-do and contact items, thereby reducing the upfront time to capture. However, the OneNote user interface is large, consuming all available screen real estate, making it difficult to have "on the side" or to switch to while in the middle of another task.

Natural language expression interfaces for intuitive expressions of PIM information are another promising approach to reduce capture effort, because they incur little or no cognitive overhead for encoding information into an appropriate form for capture. Natural language interfaces allowing users to easily add information such as events, contact information and to-do reminders to their calendars have become increasingly available, including Google's Quick Add [Google], Presdo [Presdo] and I Want Sandy [Values of n, Inc.]. These tools have found some popularity and thus struck a niche; users prefer them to form-based GUI equivalents in some situations.

Another promising direction towards accelerating and reducing interaction effort with GUI applications is the use of keyboard accelerators that map repetitive GUI actions to reflexive key combinations. A number of application launchers that employ this technique have recently gained popularity: Quicksilver [Blacktree Software], Enso [Humanized], and GNOME Do [Siegel], for example, simplify simple tasks such as application launches via through the use of hotkey trigger, combined with small, unobtrusive pop-up windows that that provide feedback and support keystroke disambiguation. However, these tools have focused primarily on application launching rather than information capture and thus do not yet address the needs of information scrap capture.

Finally, with respect to tool integration, “snippet-keeper” application Yojimbo [Bare Bones Software] facilitates cross-application content grabbing by letting the user simply select the content they want grabbed and pressing a hotkey. Users can immediately tag their grabbed items or choose to defer organization; the items are then added to collections in the person's own tag-based Yojimbo repository.

While general-purpose drawing tools and word processors are potential candidates for information scrap management because they afford fast, unconstrained input of text or drawings, they are not ideal for several reasons. First, the design needs for creating published documents and free drawings differ significantly from those of scraps, especially in creation, use and semantic/structural characteristics. As discussed in Section 5.2, information scraps are often implicitly structured but sketchy and rough, whereas word processors and drawing tools are designed to create published or shared documents and illustrations. These tools thus foreground design affordances that are important to publishing but relatively useless for scrap creation. Also, these tools are not optimized for handling a large number of small data items: we observed a number of participants compiling large collections of scraps into a single text or word processor document in order to circumvent the overhead of creating and managing many documents (Section 5.5.1).

7.2 Flexible contents and representation

Our artifact analysis in Section 5.5.1 revealed that information scraps were considerably more diverse and irregular than the commonly considered set of PIM information types. Information often did not match expected schemas: some properties were missing, and others were introduced. Participants also commonly combined information types inside of a single scrap. These behaviors resulted in scraps such as a person’s first name and phone number, a time indicating when to contact the individual, and driving directions to that person's house – but omitting the contact’s last (family) name. Furthermore, as the distribution of types discussed in Section 5.5.1 suggests, there is a very large potential set of truly personal data types that collectively make up a significant portion of all information scraps but that do not fit at all in PIM applications today.

The widely heterogeneous fragments of information contrast significantly with the limited set of fixed schemas that constitute data types in PIM applications today. While the PIM tools such as Microsoft Outlook have started to blur distinctions in PIM types (e.g., to-do items with calendar entries), research tools such as Haystack [Karger and

Quan 2004] have taken a more radical approach in which general relational models such as RDF [W3C] are used as a basis of representation. In such a representation, rather than having disparate collections of data records of particular fixed schemas, instances are defined in terms of how they link their atomic data components (e.g., dates, times or names), and linking is possible among arbitrary data components. Thus, under such a model it becomes possible to create and represent the (often implicit) meanings implied by the freeform scraps we found in the study.

A remaining challenge surrounds developing a means by which the user can utilize this expressiveness. Current interfaces for directly specifying instances using similarly rich vocabularies (*e.g.,* [Musen]) carry high comprehension and execution overheads. Another option would be to automatically extract semantics; however, in practice this is a challenging problem because the language used in scraps (Section 5.5.3) is often incomplete, ungrammatical, and highly personalized. The difficult nature of the problem is reflected in the fact that personal notes are often ambiguous and unintelligible to people other than the author.

Controlled naturalistic languages such as those first proposed for databases [Popescu et al. 2003] provide a possible solution. In a simplified natural language, user are informed that the system can only interpret a restricted set of simple, common phrasings for information (or using some fixed syntactic convention) but that they are free to express anything they wish to using this language. This technique trades off expressiveness for perceived naturalness of expression.

7.3 Flexible usage and organization

The tool adaptations described in Section 5.4.3 are particularly interesting because they reveal individuals' needs: participants devised new custom organizational systems out of existing tools to better fit their needs, for example ENG4's re-appropriation of a bug-tracking tool as a personal to-do list and MAN3's use of Post-it notes on the back of his cell phone as a capture solution. Our study also revealed several instances where tools had to be adapted in order to accommodate information that didn't fit, such as when an application failed to provide free-text annotation capabilities.

Re-appropriation and adaptation by the end-user requires tools to be sufficiently flexible to accommodate novel use of their affordances or data, and for this flexibility to be simple enough for the end user to manage. Many high-functionality information tools (e.g., emacs [Free Software Foundation, Inc.] and Eclipse [The Eclipse Foundation]), have long allowed scripting and extension facilities for customization. The practical difficulty associated with developing application-specific plug-ins was such a high barrier to entry that few users attempted it. However, the past few years have seen new trends in open web-based APIs and data services that allow users to combine data sources and user interfaces to suit their information needs. Tools such as Intel's MashMaker [Ennals and Garofalakis 2007] and Dontcheva et al.'s card metaphor [Dontcheva et al. 2007] have attempted to make this process even more accessible to end users through drag-and-drop visual interfaces.

Participants also devised a rich set of organizational techniques and strategies, as summarized in Section 5.5.3. Some of these behaviors are already well supported in digital information tools, while others are poorly supported. For example, unlike notes written on physical media, digital PIM tools are good at automatically organizing collections of information records. However, these tools are generally less capable of adapting to novel organizational strategies, so our participants tended to use spreadsheet software if they needed to manage or sort novel fields. We propose three potential solutions: letting users manually specify organizational rules, specify organizational rules by example [Halbert 1993], or dynamically construct faceted views by combining of a simple set of operators (e.g., [Huynh et al. 2007, Yee et al. 2003]).

Supporting physical organizational strategies in the digital realm has been more challenging. Several systems have demonstrated methods of simulating aspects of the physical environment in the digital realm, including the creation of digital stacks and piles [Mander et al. 1992, Robertson et al. 1998b] and visual wear and tear on heavily-used items [Hill et al. 1992]. Others have demonstrated completely new ways of envisioning our organizational scheme, such as time-based metaphors [Fertig et al. 1996].

7.4 Visibility and Reminding

Nearly all participants employed a strategy of physically situating scraps in places where they could serve as references or reminders. As described in Section 5.4.3, the desire to be able to reference useful information frequently easily inspired participants to place items in prominent, always-visible locations, for example sticking Post-Its to their workstation monitor or writing information in the corner of whiteboards. For prospective reminding, several participants reported strategically placing notes in locations where they knew they would later serendipitously "trip over" them at the right time.

It is difficult to support these behaviors in digital tools because we cannot easily situate pieces of information in particular locations of easy access or strategic significance. The problem of physically situating digital information is still solved most straightforwardly by first converting the information to physical form (*i.e.*, printing it) and then sticking the physical version in the appropriate place. Display technologies could contribute to making it easier to physically situate digital information, including low-cost electronic displays such as e-ink [E Ink Corporation], multisensory ambient displays [Butz and Jung 2005, Dahley et al. 1998], and pervasive displays such as the EveryWhere Display [Pinhanez 2001]. An alternative approach is to build location sensors into portable displays and to display information grounded at the user's location; efforts in this vein include the Remembrance Agent [Rhodes and Crabtree 1998] and augmented reality research (*e.g.,* [Hirokazu Kato]). Still other work [Hsieh et al. 2006] has sought to reproduce the kind of passive reinforcement that occurs when we shuffle through our notes or flip through the pages of our physical notebooks while looking for something else.

Reminding individuals of what their notes mean is another issue. As discussed in Section 5.5.2, the brief, incomplete nature of scraps meant that they were not necessarily future-proof, although memory for the general gist was strong. One participant, MAN4, strongly expressed a desire for helping him remember the significance of notes: “The thing I want the most in a note-taking tool would be to be able to ask it – Who? What? 'Why?' – 'Why' is absolutely crucial. If my sticky-note could answer this for me I'd be golden.” Research leveraging associative memory (the ability to recall information when it is contextualized with other relevant events) may be promising in this direction. In particular, the Stuff I've Seen project [Dumais et al. 2003] has focused on leveraging memorable events and past encounters with documents for re-finding. An extension to this approach has been lifelogging – the extended, automatic capture and retrieval of personal experiences [Hodges 2006, Gemmell 2002]. In particular, work using images taken at random from a wearable camera to prime recall has demonstrated substantial gains in duration and fidelity of recall of routine workplace situations and events [Sellen et al. 2007]. We believe that similar recall effects would occur for the meaning of notes, potentially also aiding in the re-finding of lost notes.

7.5 Mobility and Availability

Information scraps are often closely tied to mobile scenarios (Section 5.4.2). Social constraints may dictate the availability or appropriateness of tools in certain settings (Section 6.3). In response, many of our participants resorted to carrying legal pads, day planners or pocket sketchbooks whenever they were away from their desks.

To support capture in a mobile context, nearly every mobile smartphone and personal digital assistant (PDA) provides some basic PIM and freeform notetaking functionality. However, adoption varies widely among users [Dai et al. 2005]. Our study elicited two major impediments to the use of such devices for notetaking and personal information management: 1) a choice between fragmenting information across devices and synchronizing the mobile device, and 2) the difficulty, time and attention costs associated with mobile information entry.

Fragmentation can be ameliorated by synchronization, and the synchronization problem is mainly an engineering one. One option is for web- or desktop-based tools to offer mobile-accessible capture and access interfaces, as is the case with Google Calendar [Google] and Microsoft OneNote Mobile [Microsoft]. Increasingly high-bandwidth wireless networks with unlimited data pricing models have started to make feasible constant, transparent over-the-air synchronization of PIM data between mobile phones and desktop software [Nokia]. Another choice to let mobile devices serve as remote capture terminals directly to PIM software running on the desktop: for example, Jott [Jott Networks Inc.] translates phone voice commands into calendar appointments, to-dos, etc. via an API agreement with web services.

With respect to the barrier of data entry, significant progress towards improved text entry methods and new mobile capture modalities may improve their suitability for use in scrap capture. Both on-screen and physical "thumb-boards" are improving text entry speeds by incorporating better tactile feedback, predictive input and error correction facilities. Approaches that extend the reach of digital capture to handwriting on real paper are also gaining traction in new digital pen products [Logitech] systems that integrate handwritten notes with digital information are becoming visible in research [Stifelman et al. 2001, Yeh et al. 2006].

7.6 JOURKNOW

In parallel with our ethnographic study, we have designed an information scrap management tool called Jourknow (Figure 10) to explore the large design space unearthed by the previous sections [Van Kleek et al. 2007]. Instead of proposing to replace existing PIM applications as Yet Another Personal Information Manager, our intention for Jourknow is to serve as an exploratory platform for studying new input, retrieval and organizational affordances, for informing the design of the next-generation of PIM tools, to better support user needs.

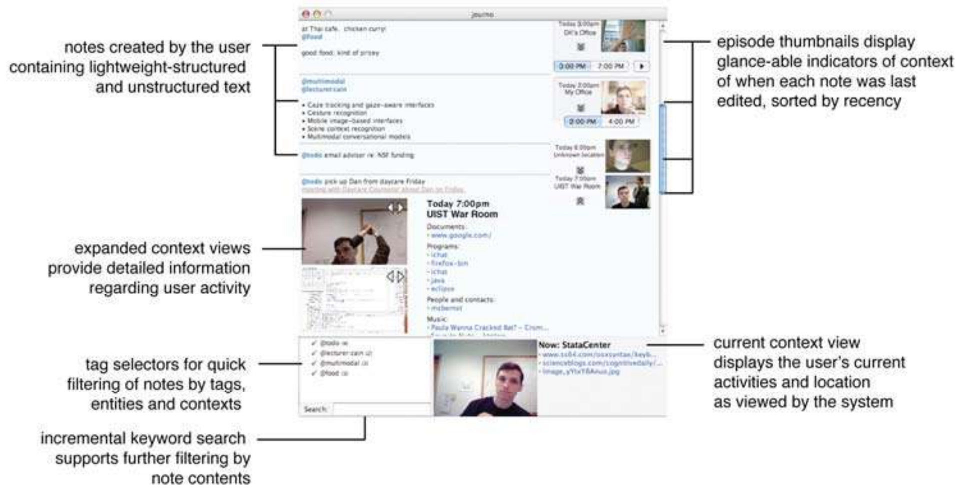

Figure 10. The user interface of Jourknow, our prototype information scrap capture and manipulation tool.

Jourknow has attempted to address the preceding design needs as follows:

- **Lightweight Capture.** To facilitate unstructured note capture into Jourknow, Jourknow provides cross-application hotkeys and a "heads-up display" for grabbing contents out of other applications or quickly writing down a piece of information while in the middle of something else.
- **Flexible Contents and Representation.** We intend Jourknow to be a single point of capture for all manner of personal information. As a simplifying maneuver, we have currently limited the system to text, cameraphone photos and voice memos as input techniques. To capture more structured data such as calendar items or bookmarks using text, the system incorporates a variety of different approaches to artificial natural languages which we call *pidgin languages* [Van Kleek et al. 2008] designed to explore variations in flexibility, extensibility and naturalness. We have also begun to examine interpretation techniques that try to leverage contextual information to aid disambiguation of short information scraps.
- **Flexible Usage and Organization.** Jourknow's data model is based on an open standard [W3] and thus can be accessed and re-interpreted by other applications. In support of flexible organization, Jourknow bases many of its

re-finding capabilities on flexible user-initiated or automatically-captured organizational metadata. The user may manually tag notes to create named sets. Leveraging the PLUM framework [Van Kleek and Shrobe 2007], Jourknow also automatically associates notes with a wide variety of contextual metadata such as location, music and chat activity, as well as open documents and programs. This metadata can be used to re-find notes later via a faceted browsing [Yee et al. 2003] interface.

- **Visibility and Reminding.** Jourknow contains two simple awareness mechanisms: a desktop widget allowing users to keep notes always visible, and an alarm mechanism that raises a later reminder of a particular note. We are also interested in exploring an interface to opportunistically display notes related to the ones currently being authored or viewed.
- **Mobility and Availability.** Our goal is to allow users to capture and access Jourknow data in a variety of different situations. The main Jourknow client is desktop-based; we have augmented the system with a mobile companion client called JourMini. JourMini can access Jourknow's notes as well as capture its own using the phone's keypad, built-in camera or voice memo features. Notes are remain synchronized between all of the clients.

Jourknow's most recent longitudinal evaluation [Bernstein et al. 2008] has demonstrated a variety of unsolved design and evaluation problems for information scrap management. We touch on these open problems in Future Work.

8. FUTURE WORK

An important next step for this work is to extend our artifact and interview study by observing scrap creation and re-finding *in situ*. We have focused thus far on understanding scraps' contents, tools and organization, examining artifacts after they have been created and before re-finding was needed. Our research does not paint a complete picture of creation and re-finding as they occur. What exactly triggers the need to record or reference an information scrap? What kind of information is recalled about each scrap at intervals after its capture? What are the most typical re-finding procedures, and how might we support them? A shadowing study would likely elicit many interesting (and likely unexpected) answers to these questions.

Just as we found it necessary to innovate methodology for the study information scraps, our experience suggests that it may be necessary to innovate again with the design process for scrap managers. Our ongoing investigation [Bernstein et al. 2008] has led us to the following set of open questions concerning the design and evaluation of systems like Jourknow:

- **Scope of the Design.** When a PIM tool such as Jourknow bases its usefulness on Gestalt integration of a wide variety of needs and uses, how can we identify subsets of these needs to prototype and test independently? The complete picture may be the only compelling one, but user-centered design suggests that monolithic development and evaluation is likely to fail. How do we isolate pieces of the system to design and iterate?
- **Prototyping.** Information scrap tools put themselves at the mercy of a wide variety of situational factors and constraints, as discussed in Section 6.3. How can we then craft effective experience prototypes [Buchenau and Suri 2000] to garner feedback on the rich context surrounding notes' capture and reuse? Experience sampling studies may be successful here, as they can force users to record information at unpredictable times and across a variety of scenarios.
- **Study type.** We have used longitudinal evaluations to give Jourknow a chance to ingratiate itself into our participants' practice, and to reflect on how that practice, once engaged, was or wasn't successful. However, such studies are high-risk, and it is difficult to force a change in practice in just a week. How can we combine evaluation methodologies to more effectively test our ideas?
- **Study population.** The quest for external validity dictates that researchers and practitioners randomly choose participants from the target population, rather than from a hand-picked subset. In the domain of information scraps, ironically, there are reasons why testing outside a friendly community might hurt a study. Due to mission critical aspects of PIM, there is little room for error – for example, while business students were excellent critics of our system, they were also unable or unwilling to overlook entry barriers to using the system such as outstanding bugs and performance issues.

Many of these issues are also raised by Kelley and Teevan in the more general context of personal information management [Kelley and Teevan 2007].

9. CONCLUSION

In this article we examined the phenomenon of *information scraps*: personal information that does not find its way into our current personal information management tools. Information scraps are pervasive – our participants shared with us an impressive number of scraps, both physical and digital, scattered over diverse parts of their information environments. We examined the wide variety of information held in scraps, the forms this information took, and the information scrap lifecycle – how scraps were captured, used and then stored or disposed. We identified a set of typical roles information scraps play in our lives: temporary storage, cognitive support, reminding, archiving, and capture of unusual information types. We then examined the needs that these roles demanded of our tools, and how participants' information scrap solutions addressed the needs. These needs included lightweight (fast, low-effort) capture, freeform expression of data, versatile representation of data, flexible organizational and usage capacities, visibility, proactive reminding, and mobility.

Our investigation uncovered evidence of unmet design needs in today's personal information management tools and thus suggests opportunities for improving those tools. By carefully considered redesigns, we can help users capture information into applications. We hypothesize that improved designs may thus reduce the number of unmanaged information scraps in our lives, though systematic study has yet to verify this hypothesis.

In particular, the wide variety of information contained in information scraps is galvanizing, as it suggests an opportunity for PIM to engage new types of information. The data indicates that a significant percentage of our personal information is beyond the reach of our current generation of tools, and furthermore will likely remain so without a significant recalibration of our goals. We identified over 125 information types from our sample of participants, and surely there are others. The diversity suggests that it may not be tractable for each of these information types to be managed by its own tool. Instead, we suggest that the future of personal information management may lie in finding a flexible approach that encompasses both traditional information and the unique data types that are currently underserved [Karger 2007]. First steps have been made (for example, [Van Kleek et al. 2007, Jones et al. 2005b, Karger and Quan 2004]), but it remains to investigate how these approaches influence practice.

Taken in sum, these conclusions specify a set of problems that will be challenging at best. Yet the challenge is necessary, even revolutionary. For PIM to move beyond its current limitations, it must venture beyond its established boundaries – and into the world of the information scrap.

ACKNOWLEDGEMENTS

The authors would like to thank the participants and pilot testers for their time, energy, and willingness to share their personal information. We are indebted to Nokia, Vanu Inc., Endeca Technologies Inc., the World Wide Web Consortium (W3C) and MIT Computer Science and Artificial Intelligence Laboratory for granting access to interview their employees, and Jamey Hicks, Kate Guarino, and Daniel Tunkelang for their help recruiting participants.

REFERENCES

ALLEN, D. 2001. *Getting Things Done: The Art of Stress-Free Productivity.* Penguin Books, New York, NY.

ANDERSON, C. 2006. *The Long Tail: Why the Future of Business Is Selling Less of More.* Hyperion, New York, NY.

BARE BONES SOFTWARE. *Yojimbo.* http://www.barebones.com/products/yojimbo/.

BARREAU, D.K. 1995. Context as a factor in personal information management systems. *Journal of the American Society for Information Science American Society for Information Science 46*, 327-339.

BARREAU, D. AND NARDI, B.A. 1995. Finding and reminding: file organization from the desktop. *ACM SIGCHI Bulletin 27*, 39-43.

BEDERSON, B.B. 2004. Interfaces for staying in the flow. *Ubiquity 5*, 1-1.

BELLOTTI, V., DUCHENEAUT, N., HOWARD, M., SMITH, I. AND GRINTER, R.E. 2005. Quality Versus Quantity: E-Mail-Centric Task Management and Its Relation With Overload. *Human-Computer Interaction 20*, 89-138.

BELLOTTI, V., DALAL, B., GOOD, N., FLYNN, P., BOBROW, D.G. AND DUCHENEAUT, N. 2004. What a to-do: studies of task management towards the design of a personal task list manager. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press, 735-742.

BELLOTTI, V. AND SMITH, I. 2000. Informing the design of an information management system with iterative fieldwork. In *DIS '00: Proceedings of the conference on Designing interactive systems.* ACM Press, 227-237.

BERGMAN, O., BEYTH-MAROM, R. AND NACHMIAS, R. 2003. The user-subjective approach to personal information management systems. *Journal of the American Society for Information Science and Technology 54*, 872-878.

BERNSTEIN, M., VAN KLEEK, M., KHUSHRAJ, D., NAYAK, R., LIU, C., SCHRAEFEL, M. AND KARGER, D.R. 2008. Wicked Problems and Gnarly Results: Reflecting on Design and Evaluation Methods for Idiosyncratic Personal Information Management Tasks. MIT-CSAIL-TR-2008-007.

BERNSTEIN, M., VAN KLEEK, M., SCHRAEFEL, M. AND KARGER, D.R. 2007. Management of personal information scraps. In *CHI '07: CHI '07 extended abstracts on Human factors in computing systems*. ACM Press, 2285-2290.

BLACKTREE SOFTWARE. *Quicksilver*. http://www.blacktree.com/.

BLANC-BRUDE, T. AND SCAPIN, D.L. 2007. What do people recall about their documents?: implications for desktop search tools. In *IUI '07: Proceedings of the 12th international conference on Intelligent user interfaces.* ACM Press, 102-111.
BLANDFORD, A.E. AND GREEN, T.R.G. 2001. Group and Individual Time Management Tools: What You Get is Not What You Need. *Personal Ubiquitous Comput. 5*, 213-230.
BOARDMAN, R. AND SASSE, M.A. 2004. Stuff goes into the computer and doesn't come out: a cross-tool study of personal information management. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press, 583-590.
BOARDMAN, R., SPENCE, R. AND SASSE, M.A. 2003. Too Many Hierarchies? The Daily Struggle for Control of the Workspace. In *Proceedings of HCI International 2003*, Vol. 1, 616-620.
BOWKER, G.C. AND STAR, S.L. 2000. *Sorting Things Out: Classification and its Consequence.* The MIT Press, Cambridge, MA.
BRUCE, H., JONES, W. AND DUMAIS, S. 2004. Keeping and Re-finding information on the Web: What do people do and what do they need. In *Proceedings of ASIST 2004*.
BUCHENAU, M. AND SURI, J.F. 2000. Experience prototyping. In *DIS '00: Proceedings of the 3rd conference on Designing interactive systems*. ACM Press, 424-433.
BUTZ, A. AND JUNG, R. 2005. Seamless user notification in ambient soundscapes. In *IUI '05: Proceedings of the 10th international conference on Intelligent user interfaces*. ACM Press, 320-322.
CAMPBELL, C. AND MAGLIO, P. 2003. Supporting notable information in office work. In *CHI '03: CHI '03 extended abstracts on Human factors in computing systems*. ACM Press, 902-903.
CSIKSZENTMIHALYI, M. 1991. *Flow: The Psychology of Optimal Experience*. Harper Perennial.
DAHLEY, A., WISNESKI, C. AND ISHII, H. 1998. Water lamp and pinwheels: ambient projection of digital information into architectural space. In *CHI '98: CHI 98 conference summary on Human factors in computing systems*. ACM Press, 269-270.
DAI, L., LUTTERS, W.G. AND BOWER, C. 2005. Why use memo for all?: restructuring mobile applications to support informal note taking. In *CHI '05: CHI '05 extended abstracts on Human factors in computing systems.* ACM Press, 1320-1323.
DARKEN, R.P. AND SIBERT, J.L. 1993. A toolset for navigation in virtual environments. In *UIST '93: Proceedings of the 6th annual ACM symposium on User interface software and technology.* ACM Press, 157-165.
DONTCHEVA, M., DRUCKER, S.M., SALESIN, D. AND COHEN, M.F. 2007. Relations, cards, and search templates: user-guided web data integration and layout. In *UIST '07: Proceedings of the 20th annual ACM symposium on User interface software and technology*. ACM Press, 61-70.
DUCHENEAUT, N. AND BELLOTTI, V. 2001. E-mail as habitat: an exploration of embedded personal information management. *interactions 8*, 30-38.
DUMAIS, S., CUTRELL, E., CADIZ, J., JANCKE, G., SARIN, R. AND ROBBINS, D.C. 2003. Stuff I've seen: a system for personal information retrieval and re-use. *In SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval.* ACM Press, 72-79.
E INK CORPORATION. *E Ink*. http://www.eink.com/.

ENNALS, R.J. AND GAROFALAKIS, M.N. 2007. MashMaker: mashups for the masses. In *SIGMOD '07: Proceedings of the 2007 ACM SIGMOD international conference on Management of data*. ACM Press, 1116-1118.

FERTIG, S., FREEMAN, E. AND GELERNTER, D. 1996. Lifestreams: an alternative to the desktop metaphor. In *CHI '96: Conference companion on Human factors in computing systems.* ACM Press, 410-411.

FISHER, D., BRUSH, A.J., GLEAVE, E. AND SMITH, M.A. 2006. Revisiting Whittaker & Sidner's email overload ten years later. In *CSCW '06: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work.* ACM Press, 309-312.

FREE SOFTWARE FOUNDATION, INC. *GNU emacs*. http://www.gnu.org/software/emacs/.

GEMMEL, J., BELL, G., LUEDER, R., DRUCKER, S., AND WONG, C. 2002. MyLifeBits: Fulfilling the Memex Vision. In *Multimedia '02: Proceedings of the 10th international conference on Multimedia.* ACM Press, 235-238.

GONÇALVES, D. AND JORGE, J.A. 2004. Describing documents: what can users tell us? In *IUI '04: Proceedings of the 9th international conference on Intelligent user interfaces.* ACM Press, 247-249.

GOOGLE. *Google Calendar*. http://calendar.google.com.

GRAY, W.D. AND BOEHMDAVIS, D.A. 2000. Milliseconds matter: An introduction to microstrategies and to their use in describing and predicting interactive behavior. *Journal of Experimental Psychology: Applied 6*, 322-335.

GRAY, W.D. AND FU, W. 2001. Ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press, 112-119.

HALBERT, D.C. 1993. SmallStar: programming by demonstration in the desktop metaphor. In *Watch What I Do: Programming by Demonstration*, CYPHER, A. Ed. MIT Press, Cambridge MA, 103-123.

HAYES, G., PIERCE, J.S. AND ABOWD, G.D. 2003. Practices for capturing short important thoughts. In *CHI '03: CHI '03 extended abstracts on Human factors in computing systems.* ACM Press, 904-905.

HODGES, S., WILLIAMS, L., BERRY, E., IZADI, S., SRINIVASAN, J., BUTLER, A., SMYTH, G., KAPUR, N., AND WOOD, K. 2006. SenseCam: A retrospective memory aid. In *UbiComp '06: Proceedings of the 8th international conference on ubiquitous computing.* Springer, 177-193.

HILL, W.C., HOLLAN, J.D., WROBLEWSKI, D. AND MCCANDLESS, T. 1992. Edit wear and read wear. In *CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems.* ACM Press, 3-9.

HIROKAZU KATO, A. *ARToolkit*. http://www.hitl.washington.edu/artoolkit/.

HSIEH, G., WOOD, K. AND SELLEN, A. 2006. Peripheral display of digital handwritten notes. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM Press, 285-288.

HUMANIZED. *Enso.* http://www.humanized.com/enso/.

HUYNH, D.F., KARGER, D.R. AND MILLER, R.C. 2007. Exhibit: lightweight structured data publishing. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*. ACM Press, 737-746.

ITO, M. AND OKABE, D. 2003. Camera phones changing the definition of picture-worthy. *Japan Media Review 29.*

JONES, W. 2004. Finders, keepers? The Present and Future Perfect in Support of Personal Information Management. *First Monday 9.*

JONES, W.P. AND DUMAIS, S.T. 1986. The spatial metaphor for user interfaces: experimental tests of reference by location versus name. *ACM Transactions on Information Systems (TOIS) 4*, 42-63.
JONES, W. 2007a. Introduction. In *Personal Information Management*, W. JONES AND J. TEEVAN, Eds. University of Washington Press, Seattle, WA, 3-21.
JONES, W., BRUCE, H. AND DUMAIS, S. 2001. Keeping found things found on the web. In *CIKM '01: Proceedings of the tenth international conference on Information and knowledge management*. ACM Press, 119-126.
JONES, W., DUMAIS, S. AND BRUCE,HARRY 2005a. Once found, what then? A study of keeping behaviors in the personal use of Web information. In *Proceedings of the American Society for Information Science and Technology 39*, 391-402.
JONES, W., MUNAT, C. AND BRUCE, H. 2005b. The Universal Labeler: Plan the Project and Let Your Information Follow. In *Proceedings of ASIST 2005.*
JONES, W., PHUWANARTNURAK, A.J., GILL, R. AND BRUCE, H. 2005c. Don't take my folders away!: organizing personal information to get things done. In *CHI '05: CHI '05 extended abstracts on Human factors in computing systems.* ACM Press, 1505-1508.
JOTT NETWORKS INC. *Jott*. http://www.jott.com.
KALNIKAITÉ, V. AND WHITTAKER, S. 2007. Software or wetware?: discovering when and why people use digital prosthetic memory. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press, 71-80.
KARGER, D.R. 2007. Unify Everything: It's All the Same To Me. In *Personal Information Management*, W. JONES AND J. TEEVAN, Eds. University of Washington Press, Seattle, WA, 127-152.
KARGER, D.R. AND QUAN, D. 2004. Haystack: a user interface for creating, browsing, and organizing arbitrary semistructured information. In *CHI '04 extended abstracts on Human factors in computing systems*. ACM Press, 777-778.
KELLEY, D. AND TEEVAN, J. 2007. Understanding What Works: Evaluating PIM Tools. In *Personal Information Management*, W. JONES AND J. TEEVAN, Eds. University of Washington Press, Seattle, WA, 190-205.
KHAN, F. 1993. A Survey of Note-Taking Practices. HPL-93-107.
KINDBERG, T., SPASOJEVIC, M., FLECK, R. AND SELLEN, A. 2004. How and Why People Use Camera Phones. HPL-2004-216.
KIRSH, D. AND MAGLIO, P.P. 1994. On Distinguishing Epistemic from Pragmatic Action. *Cognitive Science 18*, 513-549.
LAMMING, M., BROWN, P., CARTER, K., ELDRIDGE, M., FLYNN, M., LOUIE, G., ROBINSON, P. AND SELLEN, A. 1994. The design of a human memory prosthesis. *The Computer Journal 37*, 153-163.
LANSDALE, M.W. 1988. The Psychology of Personal Information Management. *Applied Ergonomics 19*, 55-66.
LICHTENSTEIN, S., FISCHHOFF, B. AND PHILLIPS, L.D. 1982. Calibration of probabilities: The state of the art to 1980. In *Judgment under uncertainty: Heuristics and biases*, KAHNEMAN, D., SLOVIC, P. AND TVERSKY, A., Eds. Cambridge University Press, Cambridge, 306-334.
LIN, M., LUTTERS, W.G. AND KIM, T.S. 2004. Understanding the micronote lifecycle: improving mobile support for informal note taking. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press, 687-694.
LOGITECH. *io2 Digital Pen*. http://www.logitech.com/index.cfm/mice_pointers/digital_pen/devices/408&cl=us,en.

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

MANDER, R., SALOMON, G. AND WONG, Y.Y. 1992. A "pile" metaphor for supporting casual organization of information. In *CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press, 627-634.

MCALPINE, H., HICKS, B.J., HUET, G. AND CULLEY, S.J. 2006. An investigation into the use and content of the engineer's logbook. *Design Studies 27*, 4, 481-504.

MICROSOFT. *OneNote*. http://office.microsoft.com/onenote/.

MUSEN, M. *Protege*. http://protege.stanford.edu/.

NOKIA. *Intellisync*. http://www.intellisync.com/.

O'DAY, V.L. AND JEFFRIES, R. 1993. Orienteering in an information landscape: how information seekers get from here to there. In *CHI '93: Proceedings of the SIGCHI conference on Human factors in computing systems.* ACM Press, 438-445.

OULASVIRTA, A. AND SUMARI, L. 2007. Mobile kits and laptop trays: managing multiple devices in mobile information work. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press, 1127-1136.

PINHANEZ, C.S. 2001. The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces. In *UbiComp '01: Proceedings of the 3rd international conference on Ubiquitous Computing*. Springer-Verlag, 315-331.

POPESCU, A., ETZIONI, O. AND KAUTZ, H. 2003. Towards a theory of natural language interfaces to databases. In *IUI '03: Proceedings of the 8th international conference on Intelligent user interfaces*. ACM Press, 149-157.

PRESDO. *Presdo.* http://www.presdo.com/.

RHODES, B. 1997. The wearable remembrance agent: A system for augmented memory. *Personal Ubiquitous Computing 1,* 4, 218-224.

ROBERTSON, G., CZERWINSKI, M., LARSON, K., ROBBINS, D.C., THIEL, D. AND DANTZICH, M.V. 1998a. Data mountain: using spatial memory for document management. In *UIST '98: Proceedings of the 11th annual ACM symposium on User interface software and technology*. ACM Press, 153-162.

ROBERTSON, G., CZERWINSKI, M., LARSON, K., ROBBINS, D.C., THIEL, D. AND DANTZICH, M.V. 1998b. Data mountain: using spatial memory for document management. In *UIST '98: Proceedings of the 11th annual ACM symposium on User interface software and technology*. ACM Press, 153-162.

ROSS, L. AND NISBETT, R. 1991. *The Person and the Situation: Perspectives of Social Psychology*. New York: McGraw-Hill.

SCHRAEFEL, M.C., HUGHES, G.V., MILLS, H.R., SMITH, G., PAYNE, T.R. AND FREY, J. 2004. Breaking the book: translating the chemistry lab book into a pervasive computing lab environment. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press, 25-32.

SELLEN, A.J., FOGG, A., AITKEN, M., HODGES, S., ROTHER, C. AND WOOD, K. 2007. Do life-logging technologies support memory for the past?: an experimental study using sensecam. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems.* ACM Press, 81-90.

SELLEN, A.J. AND HARPER, R.H.R. 2003. *The Myth of the Paperless Office*. MIT Press, Cambridge, MA, USA.

SIEGEL, D. GNOME Do. http://do.davebsd.com/.

STIFELMAN, L., ARONS, B. AND SCHMANDT, C. 2001. The audio notebook: paper and pen interaction with structured speech. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press, 182-189.

TEEVAN, J., ALVARADO, C., ACKERMAN, M.S. AND KARGER, D.R. 2004. The perfect search engine is not enough: a study of orienteering behavior in directed search. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press, 415-422.

THE ECLIPSE FOUNDATION. *Eclipse*. http://www.eclipse.org.

TRAPANI, G. 2007. *Lifehacker*. 2007. http://www.lifehacker.com.

VALUES OF N, INC. *I Want Sandy*. http://www.iwantsandy.com.

VAN KLEEK, M., BERNSTEIN, M., ANDRE, P., KARGER, D.R. AND SCHRAEFEL, M. 2008. Simplifying knowledge creation and access for end-users on the SW. In *CHI 2008 Workshop on Semantic Web User Interfaces*.

VAN KLEEK, M., BERNSTEIN, M., KARGER, D.R. AND SCHRAEFEL, M. 2007. GUI – Phooey!: The case for text input. In *UIST '07: Proceedings of the 20th annual ACM symposium on User interface software and technology*. ACM Press, 193-202.

VAN KLEEK, M. AND SHROBE, H. 2007. A Practical Activity Capture Framework for Personal, Lifetime User Modeling. In *Proceedings of User Modeling 2007*. Springer.

VENOLIA, G.D., DABBISH, L.A., CADIZ, J.J. AND GUPTA, A. 2001. Supporting Email Workflow. MSR-TR-2001-88.

W3C. *Resource Description Framework (RDF)*. http://www.w3.org/RDF/.

WAGENAAR, W.A. 1986. My memory: a study of autobiographical memory over six years. *Cognitive psychology 18*, 225-252. .

WHITTAKER, S. AND HIRSCHBERG, J. 2001. The character, value, and management of personal paper archives. *ACM Transactions on Computer-Human Interaction (TOCHI) 8*, 2, 150-170.

WHITTAKER, S. AND SIDNER, C. 1996. Email overload: exploring personal information management of email. In *CHI '96: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press, 276-283.

YATES, F.A. 1966. *The Art of Memory*. University of Chicago Press, Chicago.

YEE, K., SWEARINGEN, K., LI, K. AND HEARST, M. 2003. Faceted metadata for image search and browsing. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press, 401-408.

YEH, R., LIAO, C., KLEMMER, S., GUIMBRETIÈRE, F., LEE, B., KAKARADOV, B., STAMBERGER, J. AND PAEPCKE, A. 2006. ButterflyNet: a mobile capture and access system for field biology research. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM Press, 571-580.