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

Michael Bernstein[1], Max Van Kleek[1], David R. Karger[1], MIT

mc schraefel[2], University of Southampton

Abstract

In this paper we investigate *information scraps* – personal information where content has been scribbled on Post-it notes, scrawled on corners of sheets of paper, buried in the bodies of e-mail messages sent to ourselves, and typed haphazardly into text files. Information scraps hold our ideas, sketches, notes, reminders, driving directions, and even our poetry. We here introduce a line of PIM research on information scraps, defining them to be information held outside of our usual personal information management tools. Although information scraps are ubiquitous, we have much still to learn about these loose forms of information practice. Why are they so often kept outside of our traditional PIM locations and instead in small, freeform tools? Why must we sometimes circumvent our traditional PIM applications to hold on to our scraps, such as by e-mailing ourselves? What are information scraps' role in the larger space of personal information management, and what do they uniquely offer that we find so appealing? If these disorganized bits truly indicate the limits of our PIM tools, how might we begin to build better tools? We have pursued these questions by undertaking a study of 27 knowledge workers at five organizations. In our findings we describe information scraps from several angles: their content, their location, and their lifecycle. We identify common roles that information scraps play: temporary storage, archiving, work-in-progress, reminding, and

[1] 32 Vassar Street, Cambridge, MA 02139
[2] School of Electronics and Computer Science, University of Southampton, Southampton, UK, S017 1BJ

handling of unusual data. We conclude our work with a set of design considerations derived from an analysis of our study results.

ACM Classification: H5.2 [Information interfaces and presentation]: User Interfaces. – Graphical user interfaces.

General terms: Design, Experimentation, Human Factors

Keywords: Personal information management, notetaking, information scraps

1. Introduction

Despite the number of personal information management tools available today, a striking amount of our information remains out of their reach: the content is instead scribbled on Post-it notes, scrawled on corners of sheets of paper, buried inside the bodies of e-mail messages sent to ourselves, and typed haphazardly into text files. This scattered information contains our great ideas, sketches, notes, reminders, driving directions, and even our 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. Information scraps can be to-dos [Bellotti et al. 2004], notes to yourself [Lin et al. 2004], passwords, song lyrics, or a variety of other information. We use these information scraps to manage the unexpected or uncommon elements of our lives, often when we do not have time to encode the information with explicit detail or when our usual tools are unavailable. Information scraps can fail us, however, if we cannot re-find them later when their contents are needed or if, when we do, they are unintelligible.

As a class of personal information, we have much still to learn about information scraps. Why are they so often held outside of our traditional PIM locations and instead on Post-its or in text files? Why must we sometimes subvert our traditional PIM applications to hold on to our scraps, such as by e-mailing ourselves? What are information scraps' role in the larger space of personal information management, and what do they uniquely offer that we find so appealing? 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 scraps, the constitution of scraps, tools used in support of information scrap work, and the information scrap lifecycle. Our artifact investigation reveals a large diversity of information types encoded in information scraps, contributing an appreciable percentage of the total scraps we investigated. 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. We then identify the needs that information scraps place serve in support of these roles: 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 of tomorrow.

1.1 What is an Information Scrap?

To understand what information scraps are and why they exist, we first need first to develop an understanding of what is – and isn't – an information scrap. To ground the discussion, 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 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 recipe kept in a personal wiki

- Song lyrics and guitar tabs taped on the wall
- Copy of academic transcript saved in a text file

In search of a formal definition for information scraps to match our intuitive notions above, we might 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]. Thus the information type of information scraps could 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 our 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.

We therefore propose an alternate definition, grounded in practice: *an information scrap is any information item that falls outside all PIM tools designed to manage it*. Address information not in the address book, electronic communication not in the e-mail client, and to-dos not in a to-do manager are all examples of typical information scraps admitted by this definition. Information scraps are the personal information items that have fallen between the cracks of our PIM tools; they are generated when the user chooses a tool other than the PIM tools designed to handle that type. This definition intentionally includes as scraps information items for which there currently exists no PIM tools designed to manage them, as well as information items stored and managed in general-purpose (e.g., non-PIM) information tools. For the purposes of this study, we chose to include in the set of "general purpose tools" notebooks, spreadsheets, and text editors/word processors because they tend to be catch-alls for PIM data. The definition also intentionally makes no distinction between paper and digital PIM tools.

Our objective in this study is thus to study failures in PIM tools, not in individuals' personal information practices. While an individual's preference for keeping to-dos on Post-it notes is a perfectly legitimate practice, it also implies the existence of design failures in to-do and task list management tools. Further, commonplace though it is to keep meeting notes in a paper notebook, it is important for us to ask ourselves how we could support their management better than with notebook pages.

1.2 Results Overview

Our study consisted of 27 semi-structured interviews and artifact examinations of information workers. We find that information scraps encode a small set of popular 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 are characterized by abbreviated language and underspecified data, often complemented by sketches and freehand drawings. Tools such as e-mail and paper notebooks have proven particularly popular, though tools are often adapted or co-opted in unusual ways to fit participants' particular information practice.

We have identified a set of common roles that information scraps play in our everyday lives:

- Temporary storage: created with a short life expectancy and retained as exosomatic short-term memory 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.

We suggest that these roles might be best supported by a particular set of design affordances:

- Lightweight capture: lowering the time and effort barriers associated with capturing personal information.
- Flexible contents and representation: allowing multiple capture modalities, and allowing the user to capture whatever information or metadata 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.

We use these insights to motivate the development of a new class of PIM tools better prepared to handle the exigencies of our messy information practice.

2. Related Work

Information scraps cross many boundaries in PIM – they are a cross-tool, cross-information type phenomenon that serve many purposes in our wide range of daily information-related activities. In this section we thus situate our study of information scraps among the rich body of research already surrounding PIM-related activities. We begin this section with related research on the psychological underpinnings of the needs behind information scrap practice. Then, we review work informing organizational practices as they bear on information scraps, and investigations into the practices surrounding individual information forms (such as to-dos).

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 done by Ross and Nisbett in what they termed *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 the water. In our terms, this means choosing to scribble on one's own hand because it is too cumbersome to wait for one's laptop to boot up.

There are many such channel factors that encourage 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 filling, 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. 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 ideas may be undesirable, explaining why we may attempt to write them down as quickly as possible before our original thought was lost. This practice has more recently been transferred into the popular Getting Things Done productivity practices [Allen 2001].

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]. It is likely that this contextual memory will carry over to information scraps as well.

Ironically, memory triggers like information scraps are often not recorded even when they should be. We are habitually overconfident in our own knowledge and memory [Lichtenstein et al. 1982], potentially leading to conscious choices not to commit something to memory that later becomes valuable. Our overconfidence can also lead us to *prospective memory failures*, when we forget to remember (see [Kvavilashvili 1992] for a review). 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 many information scraps may only be deliberately re-accessed when our own memory has failed.

2.3 Organizational Practice

The organization of our offices and paper archives has much influence on information scrap practice. Malone's seminal paper on office organization [Malone 1983] directly examined the existence of unorganized *piles* in office work. He notes that while most workers' desks were piled rather than filed, computers of the time required users to file rather than pile. We note this trend here, as the often uncategorizable nature of information scraps suggests that information scraps are particularly prone to piling. Whittaker and Hirschberg [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 were handwritten and irreplaceable, a population with likely overlap with information scraps (*e.g.,* meeting notes and brainstorms). In *The Myth of the Paperless Office* [Sellen and Harper 2003], Harper and Sellen detail numerous reasons for the continued prevalence of paper in the workplace, including its ease of annotation, flexible navigation, spatial reorientability and support for collaboration. Many of these same forces are at play when we choose to use paper to record our information scraps.

Similar investigations of our digital organizational practices have been performed. Barreau and Nardi [Barreau and Nardi 1995], in an investigation of digital file hierarchies, discovered that most participants' hierarchies were surprisingly flat due to low perceived future usefulness of complex archives. Rather, digital hierarchies were structured in the service of what they refer to as *location-based finding*: navigating to a directory of interest and proceeding to browse. This result is remarkable in that users preferred to depend on their own cognitive capacities for recognition of documents rather than ensure that everything is elaborately filed, suggesting that information scraps may be recalled in the same manner. In parallel to Whittaker and Hirschberg's description of working papers, Barreau and Nardi also detail what they term *ephemeral information* – that which has a short shelf life. Many information scraps exhibit the characteristics and difficulties of ephemeral information: they are "loosely" filed or not filed at all, and difficult to manage in large quantities [Barreau and Nardi 1995].

Jones extended this work through an investigation of folders in the service of ongoing projects [Jones et al. 2005c]. The semantics of the folders he investigated were continually adapted to reflect each participant's "evolving understanding of a project and its components." This result will likely hold for information scraps as well, whose boundaries are less clearly delineated than project folders. 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, we anticipate that the forces driving these artifacts to remain unfiled will also be placed on information scraps.

Generating accurate and personally relevant filing schemas for information scraps can be a cognitively difficult process. Malone's participants complained of the difficulty of accurately filing paper information [Malone 1983]; Bowker and Star [Bowker and Star 2000] 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] across devices, applications, and media is a consequence of information scraps' keeping activities occurring across a variety of media. In a cross-tool study, Boardman [Boardman et al. 2003] 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. 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]. Information scraps can be particularly susceptible to fragmentation, as information of the same type can be scattered across multiple tools. While fragmentation is mainly between applications, when we turn to mobile situations we see the fragmentation instead across devices such as cell phones, laptops, notebooks and other devices we carry while mobile [Oulasvirta and Sumari 2007]. Information scraps are susceptible to mobile fragmentation, as freeform, unsync'ed data inevitably becomes isolated from related data on other devices. Data unification is viewed as an eventual goal to remedy this situation [Bergman et al. 2003, Boardman et al. 2003, Jones et al. 2005b, Karger 2007, Karger and Quan 2004].

2.3 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 information type.

Perhaps the most canonical information scrap is the self-note. Through a series of semi-structured interviews, Lin *et al.* arrived at a model of such notes' lifecycles: trigger, record, transfer or maintain and refer, complete, discard or archive [Lin et al. 2004]. Campbell and Maglio salient identified characteristics for 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 trend for personal notes: a strong preference for paper-based media over digital media. Dai challenged this supposition by interviewing expert users of PDA memo applications to suggest future design directions; the PDA memo users were typically stymied 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 participants reported regularly transcribing such notes onto another medium, suggesting transfer as an especially potent pain point for personal notes and thus some information scraps.

Closely related to the personal note in professional fields are engineering logbooks and meeting notes. Paper engineering logbooks, long a common practice for professional engineers to use for recording notes and ideas, were found to most commonly serve as reminders of work in progress and to serve as a personal work record for future reference [McAlpine et al. 2006]. For its part, 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 in the sciences [schraefel et al. 2004], however, is not yet clear.

E-mail is now not just a tool but a entire habitat, an ecology 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 information scrap purposes. 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]; other research suggests a complementary hypothesis that they may also be the result of information scraps being archived in the inbox. For instance, 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 (saved) e-mail is actually sent by the owner to 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 numerous 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 (and thus information scraps) will be 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 with only, "Joe the attorney"). To-dos are often integrated as resources into ongoing work, incorporating state or links to other artifacts. Bellotti's investigation uncovered a large number of separate tools (average 11.25 per person) being used to manage to-dos, noting that often they are intentionally placed not in a typical organization but instead *in the way* of a typical routine to promote visibility. These fragments or notes are very much in keeping with our definition of attributes 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 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 (KFTF) project has investigated the means by which users keep web information [Bruce et al. 2004, Jones et al. 2001, Jones et al. 2005a]. Use of bookmarks and browser history are relatively anemic; participants instead e-mailed themselves URLs with comments, saved web pages to disk, or printed out information. None of these locations is tailored to the support of URLs and bookmarks, so these strategies often generated information scraps.

Cameraphones have opened up new opportunities for information scrap collection and recording. Ito [Ito and Okabe 2003] describes the cameraphone's ability to elevate the mundane – allowing us to take photos of the seashell we find on the beach, the street sign that will allow us to relocate a new restaurant, or other objects that are simply "interesting" in some way. Such one-off photos may fall into the domain of information scraps, especially as the mundane objects can be difficult to categorize or be of predictable utility. 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.* [Kindberg et al. 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 the information scrap exists, what kind of information it holds, why it ends up in the (imperfectly suited) tool or medium it does, and how it evolves through its life. We entered our study hoping to understand to following:

- ***Definition.*** What is a precise definition of what is, and is not, an information scrap?
- ***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?
- ***Memory for information scraps.*** How long will memory for the meaning of the information scrap persist? How complete will it be? [Kalnikaité and Whittaker 2007]

- ***Tools.*** How many different tools do we use to manage information scraps? How do we adapt these tools to hold information they may not have been designed to carry? How do we spread our scraps across these tools, and does this fragmentation inhibit later re-use?
- ***User needs.*** What needs do information scraps serve? Why are they used in preference to other PIM tools?

4. Method

To explore these hypotheses, 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 was clear that information scraps are a cross-tool phenomenon, so we chose a cross-tool study inspired by the cross-tool work of [Boardman and Sasse 2004]. This study design allowed us to examine information scraps in many locations rather than a single (possibly sparse) one. As our questions primarily surrounded information scrap content, organization and 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 multinational corporation. The fourth organization was an Internet consortium, working internationally and 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 vested in information technology.

Interviews were performed at each participant's main computer, located 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 particular "proper place," like a calendar or spreadsheet. 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, 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 not in a tool designed to handle that kind of information, 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 bug list kept in a text file, or a to-do with a "where to remind me" field shoehorned in to the details field

This definition carries a connotation that what is and is not an information scrap depends on each participant's tools, needs, and practices. 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. Based on the results of this study, we later refined our definition of information scraps to the one found in the introduction of this paper.

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 types of a 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 inquired after tools (typically the most prolific approach), continued with location and finally 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 1 lists a sample of the script we followed.

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 freeform “details” field in the calendar.
	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 these).
	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 (phone numbers, passwords, serial numbers, thank-you note lists)	All examples (common information scrap).

Table 1. The three categories of Tools, Locations and Types characterized the main starting points for our artifact study.

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 general type of the information scrap as described by the participant (e.g., to-do, URL, shopping list). [3]
- Tool: the tool used to author and edit the information scrap.
- Presence of material not directly created by the author: a binary variable, true if the information scrap contained any content that was copied into the scrap rather than authored directly.
- Content encoding: four binary variables, each true if the information scrap contained the particular kind of content:
    - Text: written or typed text or words
    - Photographs
    - Pictorial Drawings: representative drawings; drawings intended to look like a particular object, as in a design sketch
    - Abstract Drawings: non-representative drawings; graphs, arrows, abstract visual layouts, diagrams

Multiple items were often coded from a single tool, as distinguished from each other by the participant. 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.

5. Results

[3] 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 artifact as a to-do.

5.1 What do Information Scraps Contain?

In our artifact analysis, we coded each of the 533 information scraps for its information type, then consolidated similar categories. The results can be seen in Table 2 and Figure 1.

Occurrences	Type
92	To-Do
44	Meeting Notes
38	Name And Contact Information
25	How-Tos
16	Work-In-Progress
14	Directory, File Path Or URL
12	Desired Items
12	Login/Password
9	Brainstorm, Calendar Or Event Details, Event Notes
8	Progress Report, Receipts And 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 And Goals, Timeline, Airplane Flight Information, Archived Document
3	Agenda, Configuration Settings, Jobs And Classifieds, Math Scratchwork, Project Notes, Shipping Information, Template E-Mail Response, To Read, Whiteboard Capture
2	(Mixed type), Academic Record, Bug List, Changelog, 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 And 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 Spellcheck

Table 2. We noted large numbers of scraps containing as to-dos, meeting notes, contact information, and how-to guides. However, there is was also a group of scrap types that only occurred once or twice across our entire investigation, highlighting the wide variety of personal information encoded in information scraps.

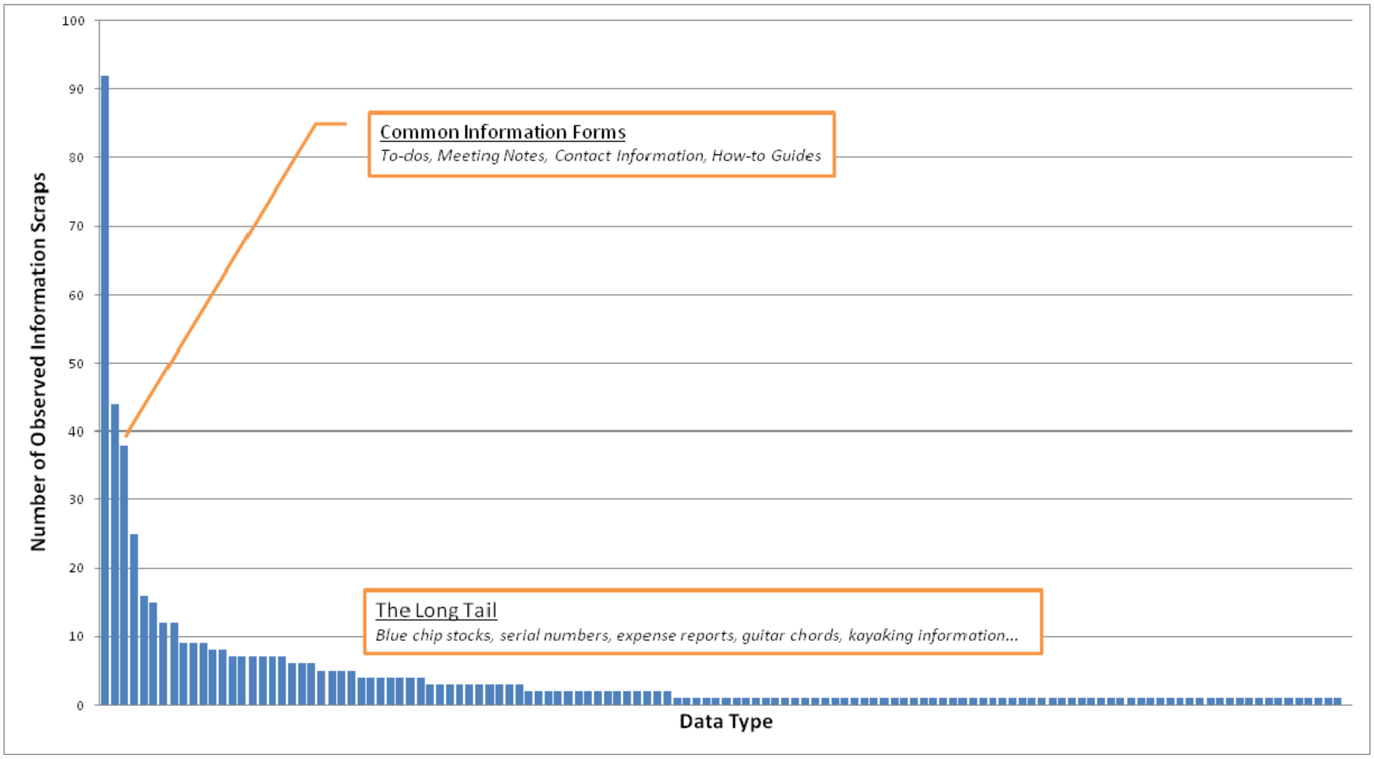


Figure 1. This visualization of Table 2 demonstrates the span of the long tail of personal information in information scraps. The ordering is identical to that in Table 2.

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: name, address, phone number or e-mail address.

- ***How-to Guides*.** Recipes noting how to perform certain tasks, kept for future reference. Examples included UNIX shell commands, login procedures for remote servers, and international calling/shipping/faxing instructions.

It is notable 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 3, 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. Information scraps encode a wide variety of personal information.

This result suggests that the distribution of information types contained with scraps exhibits the *long-tail* property of statistical distributions, where the scraps containing rarer forms cumulatively rival the number of occurrences of commonly occurring forms.[4] Our result is consistent with a variety of other natural occurrences distributed as power laws, patterns known variously as The Pareto Principle and Bradford's Law. These principles suggest that a large percentage of our information will fall into a small number of categories, and that a large diversity of unpopular categories will make up the remainder.

[4] This *long tail* effect has also recently gained fame in realm of Internet economics. [Anderson 2006]

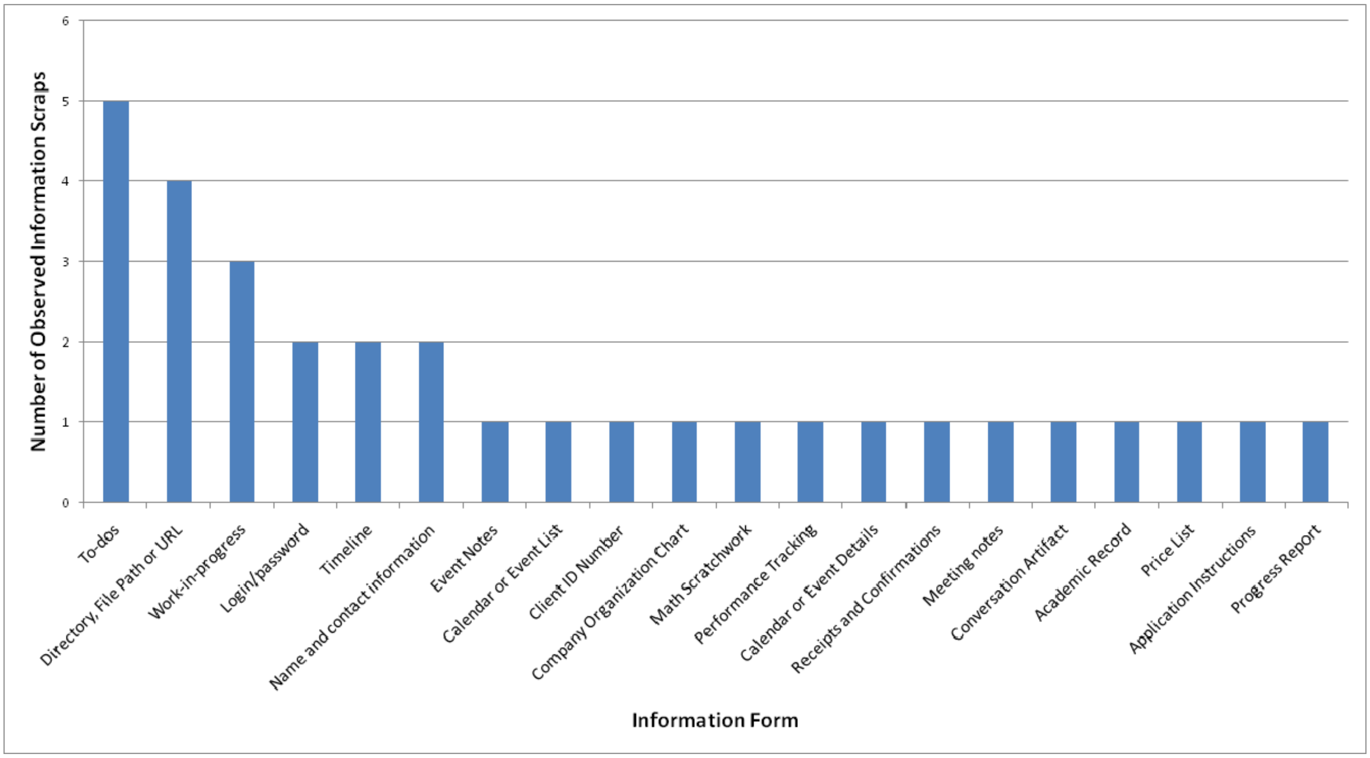
Upper Bound on Number of Occurrences of an Information Type	% of all Information Scraps
1	12.8%
2	18.4%
3	24.0%
4	29.3%
5	33.0%
6	36.4%
7	44.3%
8	47.3%
9	52.4%

Table 3. Information forms 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.

As can be seen in Figure 1, we uncovered a large number of rarely-occurring information types, including book wishlists, application serial numbers, expense reports, resumes, guitar chords, and information about kayaking. All of these types could benefit from an application tailored to their semantics (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 nonexistent 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' scraps, which implies that individuals use scraps to keep a large number of infrequently occurring information types themselves; see Figure 2a and Figure 2b. This result is again consistent with The Pareto Principle and Bradford's law, which state that subsets of the general population also follow similar power laws.

Figure 2b provides an interesting link between the global (i.e. inter-participant) and personal distributions: ENG3's most popular item are 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 personal distribution, the form still exists in what is globally the tail of the distribution. Several other participants exhibited such behavior, as well: ADMN1 maintained a set of artifacts she needed to share with her superior, ENG2 keeps an extensive set of personal progress reports, MAN6 has a sizable folder of documents to read in his free time, and so on. Yet each of these patterns was minor when averaged over the entire sample.



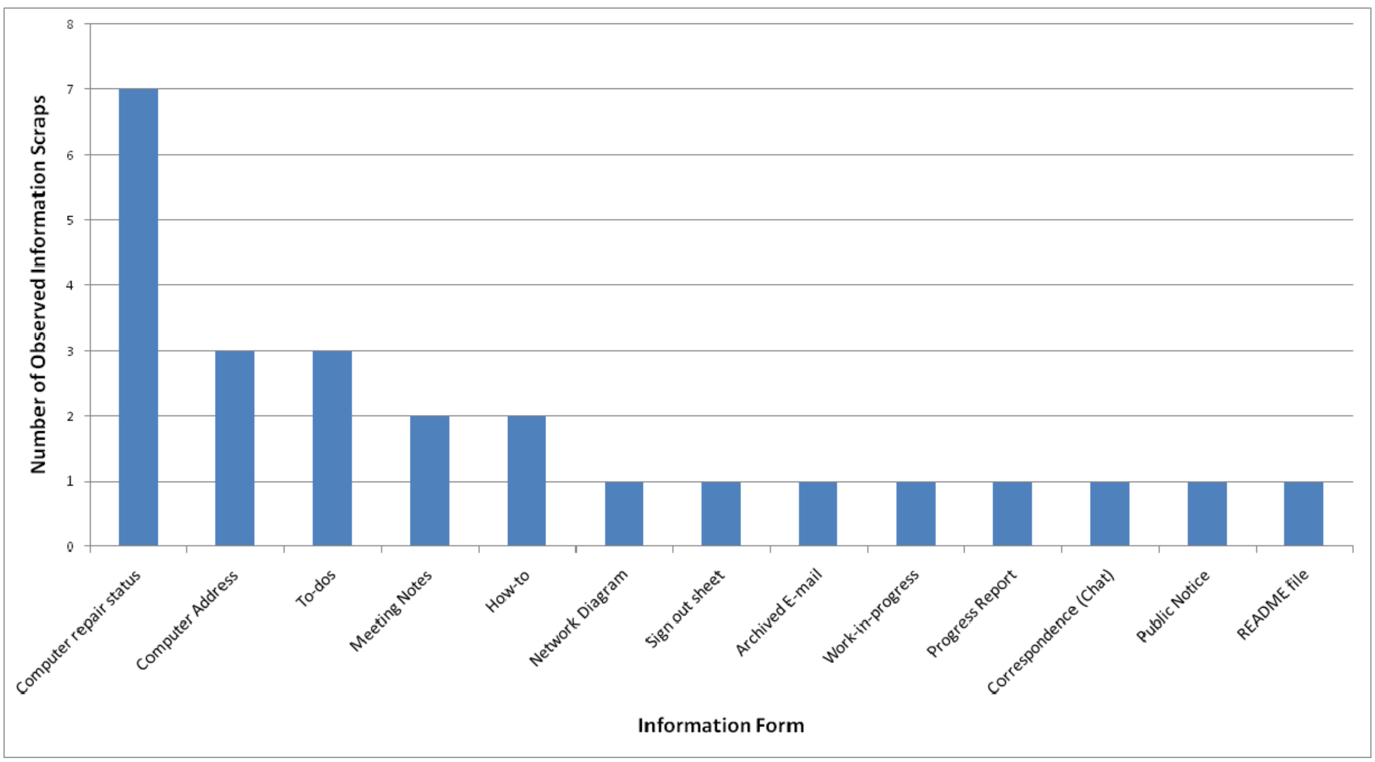


Figure 2. a) FIN1's information form distribution, which also follows a long tail. Here the histogram closely mirrors the accumulated distribution across all participants. b) ENG3's information form distribution, evidencing a large number (7) of computer status note scraps. Though these notes are in 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. However, it was not uncommon for information scraps to also contain abstract drawings (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.

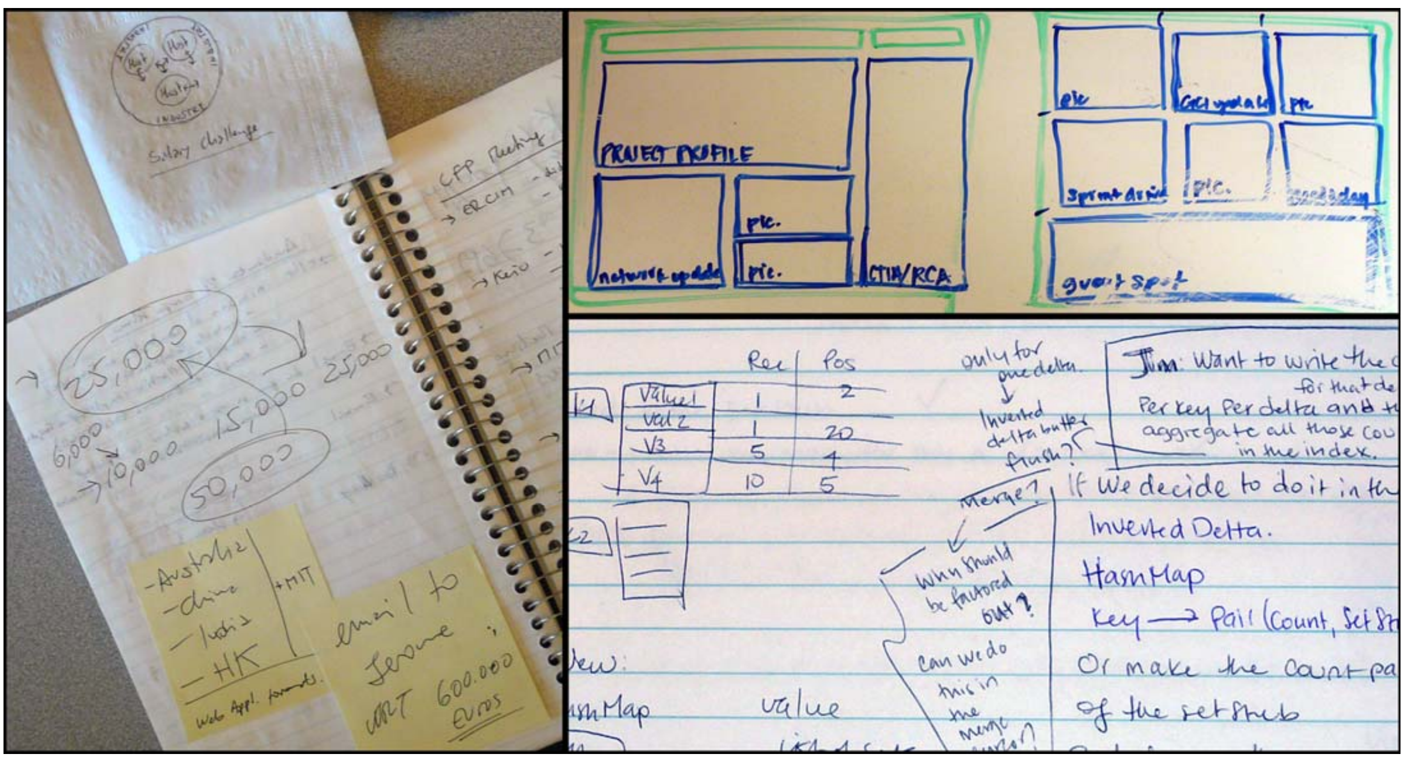


Figure 3. All three of these information scraps contain abstract drawings, including arrows, boxes, and so on.

Due to the varied nature of the types of information scraps, there was a corresponding variation in the type primitives encoded, such as phone numbers, e-mail addresses, or URLs. It was not uncommon to see several primitives 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, all mainly unrelated. The information types were rarely labeled, which occasionally made it difficult for the interviewers to understand the content of some notes (Figure 4).

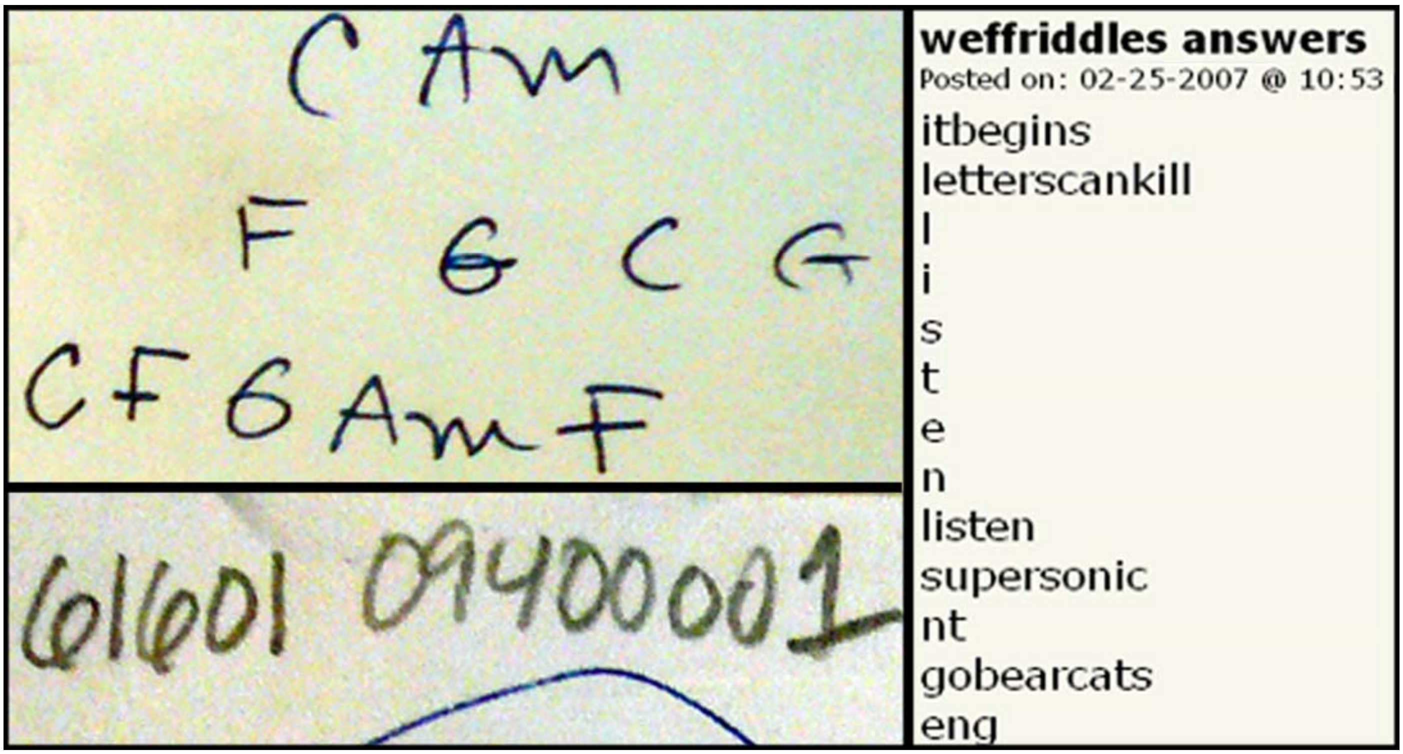


Figure 4. These three information scraps all contain unusual data. From left to right, top to bottom: guitar chords, an unknown string of numbers, and answers to an online riddle.

Information scraps were a common way to capture data that did not match existing schemas. Underspecifying data was an important affordance: "I don't know exactly when [the visitor] will come today," MAN3 explained about the calendar event in his to-do manager, "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.

Flexible content traded off the ability to manipulate the data in a structured manner later. Thus, many participants were reduced to keyword search or page flipping to re-find particular scraps. Groups of scraps that had grown large enough sometimes established their own lightweight structure to support manipulation, such as by using spreadsheet software with headings.

Several participants who kept free text files on their computer utilized this 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 (fully 10% of physical scraps involved some sort of drawing annotation).

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.

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. 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 scraps comprising notes, such as from 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."

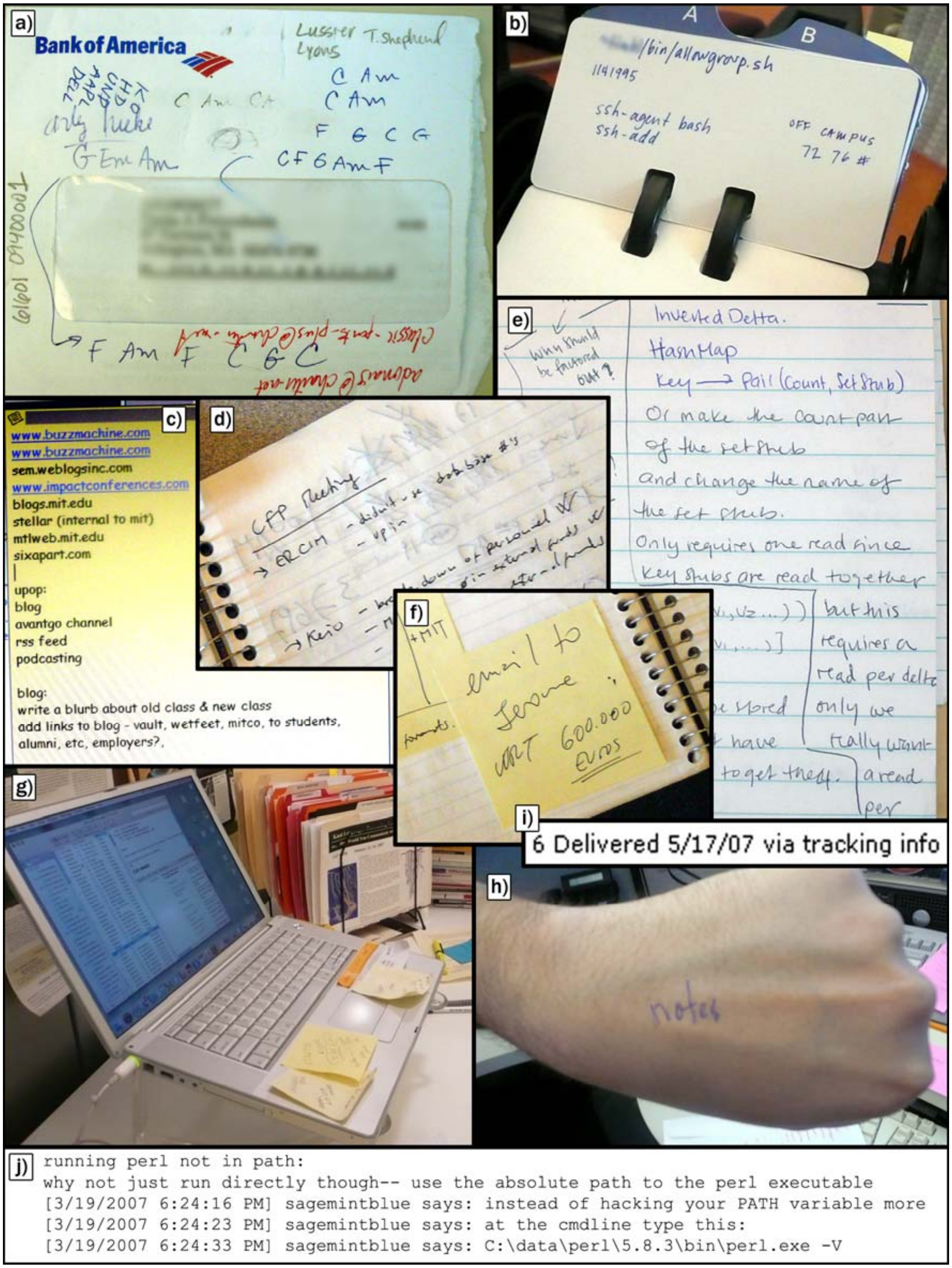


Figure 5. Several of the information scraps we noted, focusing on typical examples of minimal use of language in scraps. a) an envelope with several scraps: guitar chords, stock ticker symbols, e-mail addresses, and an unknown number, b) a web site address, password and helpful SSH commands in a rolodex, 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 (Pareto Principle) pattern, beginning with a set of extremely popular tools and trailing off to a large number of less popular ones. Participants maintained a small number of main locations for capturing information scraps, supplemented with less-used auxiliary tools.

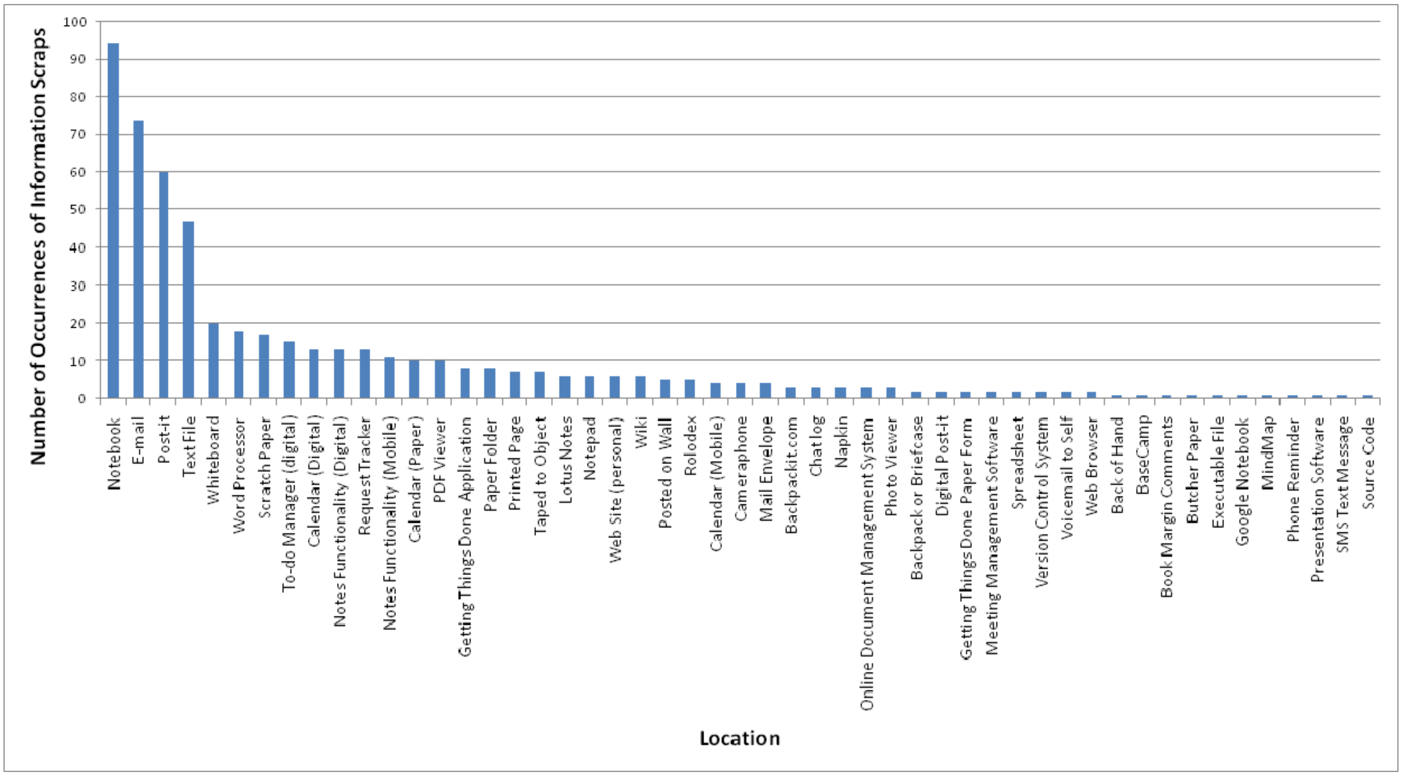


Figure 6. There are a small number of tools and locations used frequently, such as notebooks, e-mail, and post-its, and a large number of locations used a small number of times.

The digital and physical world each held popular tools. E-mail was by far the most dominant digital tool for recording information scraps (74 instances, 26.4% of digital scraps), followed by a text editor (47 instances, 16.8%; *e.g.*, TextEdit, Notepad, Wordpad) and word processor (27 instances, 6.4%; *e.g.,* Microsoft Word). Demonstrating the use of text files in particular, a few participants generated large, separate files of contact information because they wanted to keep the information conceptually separate from their contacts or needed data fields not available in the typical contact manager.

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. A chi-squared test comparing the number of digital and physical artifacts between each participant rejects the null hypothesis ($p < 0.01$), indicating that there is some dependency between the participant and the distribution of their information scraps between digital and physical tools. Figure 7 indicates what this connection might be: 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 who made special effort to have a paperless office. Their existence 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].

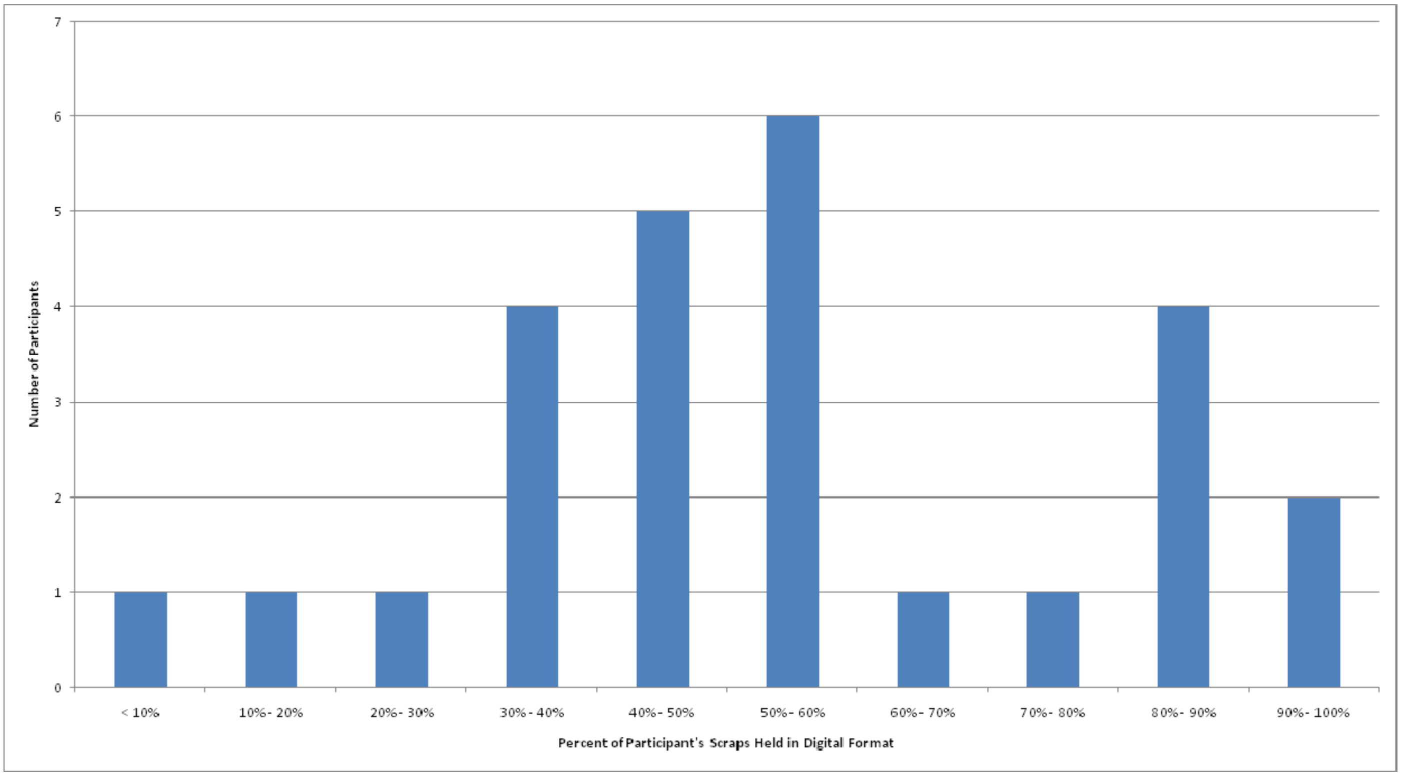


Figure 7. Examining the number of participants at each level of digital data, we see two groups: one centered around half digital, half physical, and the other almost completely digital.

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 cameraphone 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 in mobile scenarios. UI2 was often mobile: she described how she would reference a note file on her smartphone with relevant conference call 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

Information scrap tools were characterized by an adaptivity to novel uses. Post-Its are perhaps most illustrative of this fact: we observed Post-its stuck on walls, on monitors, and inside notebooks; we saw them being folded up and carried into other rooms, crossed out and annotated, and stuck on to cell phones. The tools' generality enabled a wide variety of adaptation and innovation.

This adaptability enabled re-appropriation of the 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 (i.e., individual "issues" as commitments with deadlines), he could access it form anywhere, and because he could easily update his list of commitments by emailing the system. Similarly, MAN3, who stuck Post-It notes to the back of his cell phone, took advantage both the availability of an unused area on the back of his cell phone and Post-It notes of the right size to devise a solution that worked around the physical impossibility of taking down notes on the phone while using it.

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

We found that several participants attached commonly referenced artifacts to highly visible objects, such as the wall or computer monitor. Participants often depended on scraps' visibility and in-the-way qualities (as described by [Bellotti et al. 2004, Malone 1983]): for example, DOC left Post-Its with to-dos in an intrusive location right next to her monitor until she finished them. Sometimes the information was even attached to the item of interest; in Figure 8, ENG3 wrote on tape affixed directly to the computers he was attempting to annotate.

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], focusing our results using the lens of particular parts of the information scrap lifecycle. Here, we discuss Capture, Transfer, Organization, and Reference, Retrieval and Recall:

- 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
- 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

One of the most commonly cited situations prompting information scrap creation was the need to write something down quickly. Participants reported that quick capture was often prompted by situations such as uninvited distractions or being away from their desks. Thus they were a common means to archive important thoughts before the ideas were forgotten [Hayes et al. 2003]. "Mind like water," MAN6 explained, zenlike; getting the thought out of the head and written somewhere as quickly as possible was the first priority. This offloading of data to the scrap freed our participants to focus on their primary tasks, such as holding a conversation, paying attention to a meeting, or even driving the car (as with MAN7).

When time and effort were at such a premium, the fastest tool would often win out. Even seemingly minor difficulties or annoyances with tools could have striking effects on behavior. "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 syncing. 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 simply because they require no time to start up. The effect is similar to the one described with mobile applications by Oulasvirta *et al.* [Oulasvirta and Sumari 2007].

Even when data was implicitly structureable, such as with potential calendar events, participants chose the faster, structureless route of recording a scrap: for example, a string literal such as "mtg @ 5pm in cafe." In the semi-structured interviews, participants explained that structuring the data 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.

We observed three major sources of information for information scraps: directly authored material, automatically archived material, and copy/pasted material. Directly authored material, the most common, was intentionally written by the user in a direct 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 "Misc" 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 ($p < 0.01$).

5.5.2 Transfer

Transfer is the process of moving an information scrap from one medium to another after it has been captured. Information is moved from tool to tool, from physical to digital media, or from digital to physical media. Relatively few scraps were transferred; participants explained that those that were transferred often held additional importance.

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 transfered 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 in a work-in-progress state and needed to be available in a tool that offered additional affordances or the ability to accumulate several information scraps into one location. Third was mobility: scraps were sometimes written onto Post-its to carry to another room, or sent in e-mail so that it would be retrievable from home.

5.5.3 Organization

Among scraps that were archived, archiving techniques varied; some consolidated scraps of similar types/purposes, while others situated scraps with other scraps that were created at the same time, creating 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 used ad-hoc solutions to circumvent organizational barriers, replicating information scraps across multiple 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 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, such as 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 in [O'Day and Jeffries 1993, Teevan et al. 2004]. That is, they would teleport (directly navigate) to a folder location thought to be relatively close to the desired information scrap or sort the items in a useful manner, then take 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 study, we encountered a series of affective and psychological dimensions of information scrap management more powerful than we had expected. The most prominent factors will need to be accounted for in any information scrap management system, so we report them here.

There was often self-perceived strong social pressure to be viewed as an organized individual; admitting to the existence of information scraps ran directly counter to this perception. Similarly to as recounted by Boardman and Sasse [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, similarly to [Bellotti and Smith 2000]. 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 [Allen 2001] methodology.
- "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 interviewers what had become an affirmation of her love for a messy notebook: "It's OK to have a notebook!"

Often participants were forced into a cognitive dissonance between their perceptions of themselves as decently 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 refinder of information scraps also seemed *de rigeur*. 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:

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 came into play here, as information scraps could be used to bring information along, for example driving directions written on a Post-it. These 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 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 were commonly in cognitive support roles because scrap creation tools supported, and even encouraged, messy information work.

Two such support functions were epistemic action and external cognition. Epistemic action is a kind of thinking-by-doing, similar to the benefits expert Tetris players enjoy by rotating falling puzzle pieces to directly visualize and interpret the result of the rotation action [Kirsh and Maglio 1994]. Information scraps thus allowed authors to reflect on what they were doing, change text, cross out ideas, and reposition elements simply to see the result. 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, UI2's sketches of interface designs, and a large number of work-in-progress documents.

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. Our participants expressed comfort knowing the information had been safely saved.

Reminding. Many participants took advantage of information scraps' visibility and mobility by placing them in the way of their future movements to create reminders for themselves. Participants used techniques such as colored Post-its, unread or unfiled e-mails, and files left on the desktop, reminding them to take action later or that a piece of information might be interesting to return to at a later date. To-dos were an extremely common information scrap filling this role.

Unusual Information Types. This role was a catch-all for personal information that did not quite fit existing tools' expectations. 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 quite fit existing tools' expectations, such as calendar items with a date but no start time chosen.

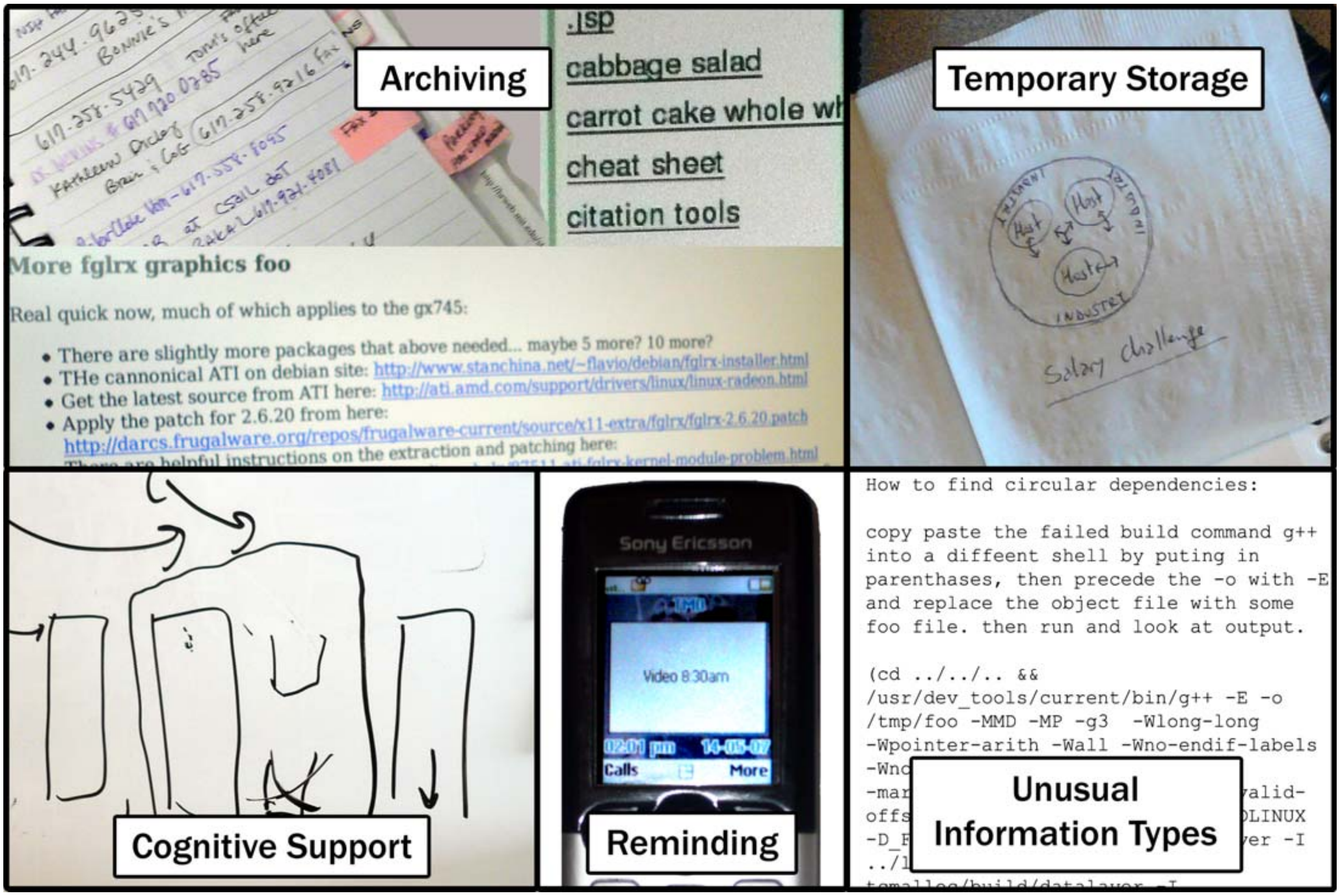


Figure 9. The five main roles information scraps played were archiving, temporary storage, work-in-progress, reminding, and a place to put information that wouldn't fit elsewhere.

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 that the user has 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 they were 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. Similarly 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].

A fragmentation-like problem 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

Table 4 frames the affordances in terms of three phases of the information scrap lifecycle: capture, organization, and use. We also highlight in the table the particular conditions that we see predicating the particular affordance requirements. These are the particular physical, temporal, social or structural attributes 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. For instance, if we are balancing a checkbook, we open a personal finance application and start a file; we do not create an information scrap. While paying bills, we are fairly indifferent to the physical environment, time constraints, or social conditions.The tasks of personal financial management predicate the use of the recognized tool for that genre of information. We may each use different tools for that task, but generally, we would agree there is a tool that exists to support that task, and that's the one we use when performing that task. There are also certain agreed social conventions around this task. Writing a paper usually means that paperwriting is a primary task in terms of one's attention. It would be unusual to have a tête–à–tête with a colleague while trying to pay bills. 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 on the paper while meeting with a colleague about that paper, the temporal, physical and social constraints would have more of an effect our choice of tool. For instance, it may be socially taboo to be seen either to use 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 to a remark. Likewise, one may be on the move, so that not only is it awkward to use a more formal mechanism for recording data, it may be that there is simply not the time to engage in a detailed recording of an observation. 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 private blog and tag them by cuisine type, are the individual recipes still information scraps? In some sense, no, because they have a place to reside and are reliably indexed. However, we cannot help but think that a blog might be a suboptimal method to collect and organize recipes, because it cannot take natively take advantage of recipe-specific needs such as ingredient filters or shopping list generation.

With regards to the variety of information types we catalogued, one difficulty with analyzing the frequency distribution of information types in scraps surrounds the question of where to draw 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 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 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, *in situ*, how information scraps were used and created, but instead only how people reported they used their scraps, along with evidence from their workspace and the physical and digital artifacts we collected. The kind of statistics that our procedure was unable to capture included the number of information scraps participants generated each day and significant contextual associations. 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, more so that 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 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 digital personal information tools to better serve these important and underserved roles of personal information practice? Do there exist any designs that support these affordances, and if not, how realistically can we expect such affordances to be supported by current technology? To ground this investigation, we discuss aspects of the design of Jourknow [Van Kleek et al. 2007], shown in Figure 10, our own note-taking research tool designed to better support information scrap activity.

We derive these design opportunities to identify the design deficiencies in our current tools that most explain users' preference for information scraps in fulfilling the roles in Section 6.1. While we cannot predict whether addressing these needs would cause the demise of unmanaged scraps, we have evidence that addressing these needs would positively impact daily practice.

Table 4 outlines the set of affordances we have identified, including lightweight capture, flexible content, flexible organization, visibility and reminding, and mobility and availability. In Table 4 we have contextualized each affordance in terms of the activities and constraints that influence scrap generation.

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. "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 [computer] which 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

Table 4. The major design affordances necessary for management of information scraps derive from access, scraps' contents, and organization.

7.1 Lightweight Capture

As described in Section 5.5.2, we found information scraps often to be generated in response to a need to capture data as quickly and easily as possible. 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 idiosyncracies in how individuals captured thoughts on unstructured media like paper and textual notes. Specifically, notes were often brief, incomplete fragments that explicitly represented very little and served more as memory primers than as information records. This finding contrasts considerably with most PIM applications' requirements that users complete forms with formal expressions for properties such as who, when and where. Our hypothesis is that tools can lessen the time and effort associated with entering information by supporting incomplete, informal representations.
- **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 messages, 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.

While general-purpose drawing tools and word processors are potential candidates for information scrap management because they afford 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. Finally, these tools are not optimized for handling a large number of small data items: we observed a number of participants compiled 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).

Tablet-based notetaking tools such as Microsoft OneNote [Microsoft] have granted digital note-taking some of the expressive freedom of paper and pen. OneNote in particular also allows a 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 monolithic and large, and therefore 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. These interface let users easily add information such as events, contact info and to-do reminders to their calendars and online PIM apps. Specifically, Google's Quick Add [Google] and I Want Sandy [Values of n, Inc.] have demonstrated potential for advantages in terms of navigational burden and efficiency over form-based GUI equivalents.

A number of cross-application launchers have recently gained popularity by turning common application launch and navigation activities into reflexive keyboard actions. The Quicksilver interface for MacOS X [Blacktree Software], for example, can be summoned from any application using a hotkey trigger, employs a small, unobtrusive interface to avoid obscuring the display of one's main task display, and adaptively re-ranks to ensure that the most frequently used commands are easily accessible. However, these tools are generally application-focused rather than information-focused – they are not designed to capture thoughts and personal information.

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.

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 contact items consisting of a name and a phone number, a time indicating when to contact the individual, and driving directions to the person's house – but no last 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.

These observations indicate a potential need for more flexible representations of PIM data. Some PIM suites such as Microsoft Office [Microsoft] 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 a general relational model called RDF [W3] is 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.,* [Mark 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.

Simplified naturalistic languages such as first proposed for Natural Language Interfaces 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 reappropriation 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 freetext annotation capabilities.

Re-appropriation requires tools to be sufficiently flexible to accommodate novel use of their affordances or data. 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 these application-specific plug-ins created such a high barrier to entry that this solution was largely beyond the realm of feasibility for most users. However, the past few years have seen a new trend in the popularization of open web-based APIs to data services. These APIs and reusable widgets have allowed web developers to combine functionality from multiple online applications into their own custom applications, spawning the phenomenon of mash-ups. 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.

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 tools like Excel [Microsoft] if they needed to manage or sort novel fields. We propose three potential solutions: letting users manually specify organizational rules, learn 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]) in a user interface familiar from online commerce.

Capturing physical organizational strategies in the digital realm have been more challenging. Visual layout systems involving stacks and piles [Mander et al. 1992, Robertson et al. 1998b] enable 2- and 3-dimensional layouts and implicit Gestalt grouping. Other possibilities include implicit "wear" on heavily-used items [Hill et al. 1992] and 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 somewhat 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.,* [ARToolworks 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.

A separate reminding problem surrounds reminding individuals of what their notes mean. As we discussed in Section 5.5.2, the brief, incomplete nature of scraps meant that they were not future-proof, although few admitted to having trouble understanding their notes after they had written them. 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 surrounding ways to support long-term recall of events includes Stuff I've Seen [Dumais et al. 2003], which attempts to contextualize events in terms of other more memorable events or features by long-term video transcripts [Czerwinski and Horvitz 2002] or images taken at random from a wearable camera [Sellen et al.

2007]. Both projects have demonstrated substantial gains in duration and fidelity of recall of routine workplace situations and events. We believe that similar recall effect would be occur for remembering the meaning of notes, as well as for helping in the re-finding of lost notes.

7.5 Mobility and Availability

Information scraps are often closely tied to mobile scenarios: scraps are often captured or referenced while away from one's desk (Section 5.4.2). Further, 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 have opened the door for automatic synchronization [Nokia], though automatic synchronization is not standard practice as yet. Another choice is for intermediary clients to pass information between the mobile client and 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, two options are innovation in text entry methods and innovation in mobile capture modalities. A variety of keypad form factors are exploring the former option, for example placing full QWERTY keyboards on phones or implementing text prediction as with the iPhone [Apple Inc.]. However, alternative approaches to notetaking incorporating paper-and-pen are also gaining traction in products [Logitech] and with additional functionality in research circles [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 [Van Kleek et al. 2007] to explore the large design space unearthed by the previous sections. 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.

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.

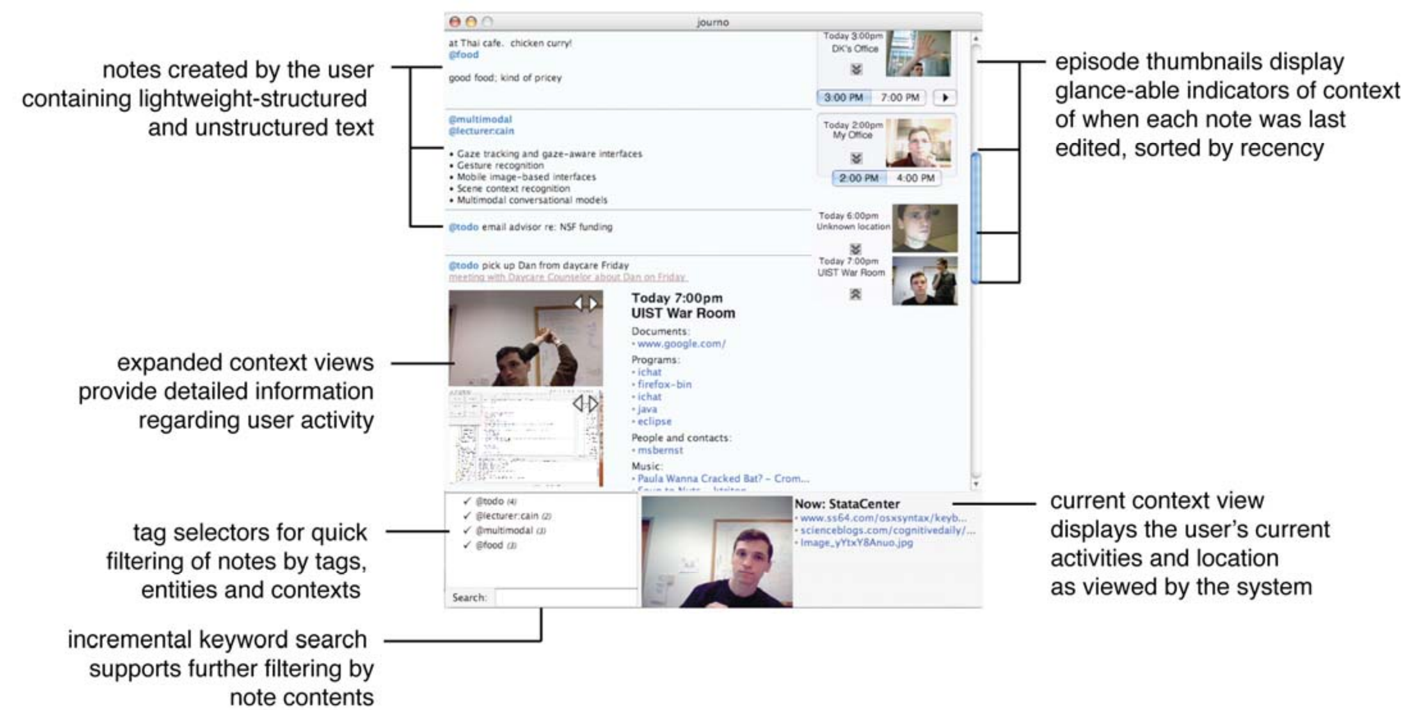


Figure 10. The user interface of Jourknow, our prototype information scrap capture and manipulation tool.

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? We suggest that experience sampling studies may be successful, as they can force users to record information at unpredictable times and across a variety of scenarios.
- **Study type.** We chose a longitudinal evaluation 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 few days. 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. We follow Jones [Jones 2007b] in suggesting that testing on more pliable participants such as coworkers or even oneself to start may lead to more useful feedback.

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 the *information scrap,* personal information that does not make 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 analyzed the wide variety of information types held in scrap form, underlining the criticality of supporting unusual or unique types in information scrap work. By examining the composition and lifecycle of information scraps, we have identified a set of user needs for information scrap management: lightweight capture mechanisms, freeform and potentially unstructured data representation, flexible organizational and usage capacities, visibility and the in-the-way property, and mobility. These strengths drive a set of typical information scrap roles, including temporary storage, cognitive support, reminding, archiving, and capture of unusual information types. Jourknow, our prototype information scrap management tool, has begun synthesizing these issues.

Our goal with this work is dualistic and somewhat contradictory: we want to understand and support information scraps, so that there might be fewer of them. Information scraps are an outcome – the result of a coping strategy – rather than an outright goal. The needs driving information scrap creation will remain constant, as we will always require lightweight capture, support for unusual data and flexibility. The scraps however, the information falling outside the purview of our personal information managers, can be reduced, eliminated or redirected by carefully considered redesigns of our personal information management tools.

The wide variety of information encoded in information scraps is particularly motivating, 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, unified platform to corral as much data as the user cares to support [Karger 2007]. First steps have been made in this direction (for example, [Jones et al. 2005b, Karger and Quan 2004, Van Kleek et al. 2007, Van Kleek et al. 2008]), but we have much distance yet to cover. Data flexibility will be in tension with a clear demand for rapid capture, driving a need for novel methods for capture and organization.

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 information scraps.

11. 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. 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.

12. 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.

APPLE INC. iPhone. http://www.apple.com/iphone/.

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. 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,* Vienna, Austria, Anonymous ACM Press, New York, NY, USA, 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,* New York City, New York, United States, Anonymous ACM Press, New York, NY, USA, 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,* San Jose, CA, USA, Anonymous ACM, New York, NY, USA, 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,* Honolulu, Hawaii, USA, Anonymous ACM Press, New York, NY, USA, 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,* Vienna, Austria, Anonymous ACM Press, New York, NY, USA, 583-590.

BOARDMAN, R., SPENCE, R. AND SASSE, M.A. 2003. Too Many Hierarchies? The Daily Struggle for Control of the Workspace. In *HCI International,* Anonymous , 616-620.

BOWKER, G.C. AND STAR, S.L. 2000. 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 *Proc. of ASIST 2004,* Anonymous .

BUCHENAU, M. AND SURI, J.F. 2000. Experience prototyping. In *DIS '00: Proceedings of the 3rd conference on Designing interactive systems,* New York City, New York, United States, Anonymous ACM, New York, NY, USA, 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,* San Diego, California, USA, Anonymous ACM, New York, NY, USA, 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,* Ft. Lauderdale, Florida, USA, Anonymous ACM Press, New York, NY, USA, 902-903.

CSIKSZENTMIHALYI, M. 1991. Flow: The Psychology of Optimal Experience. {Harper Perennial}, .

CZERWINSKI, M. AND HORVITZ, E. 2002. An Investigation of Memory for Daily Computing Events. .

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,* Los Angeles, California, United States, Anonymous ACM, New York, NY, USA, 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,* Portland, OR, USA, Anonymous ACM Press, New York, NY, USA, 1320-1323.

DARKEN, R.P. AND SIBERT, J.L. 1993. A toolset for navigation in virtual environments. *Proceedings of the 6th annual ACM symposium on User interface software and technology* 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,* Newport, Rhode Island, USA, Anonymous ACM, New York, NY, USA, 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 informaion retrieval,* Toronto, Canada, Anonymous ACM Press, New York, NY, USA, 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,* Beijing, China, Anonymous ACM, New York, NY, USA, 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,* Vancouver, British Columbia, Canada, Anonymous ACM, New York, NY, USA, 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,* Banff, Alberta, Canada, Anonymous ACM Press, New York, NY, USA, 309-312.

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

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,* Funchal, Madeira, Portugal, Anonymous ACM Press, New York, NY, USA, 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,* Seattle, Washington, United States, Anonymous ACM, New York, NY, USA, 112-119.

HALBERT, D.C. 1993. SmallStar: programming by demonstration in the desktop metaphor. 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,* Ft. Lauderdale, Florida, USA, Anonymous ACM Press, New York, NY, USA, 904-905.

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,* Monterey, California, United States, Anonymous ACM, New York, NY, USA, 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,* Montr\&\#233;al, Qu\&\#233;bec, Canada, Anonymous ACM Press, New York, NY, USA, 285-288.

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,* Banff, Alberta, Canada, Anonymous ACM Press, New York, NY, USA, 737-746.

ITO, M. AND OKABE, D. 2003. Camera phones changing the definition of picture-worthy. 2007, http://www.ojr.org/japan/wireless/1062208524.php.

JONES, W. 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. 2007b. Personal Information Management. In *Annual Review of Information Science and Technology (ARIST),* B. CRONIN, Ed. Information Today, Medford, NJ, 453-504.

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,* Atlanta, Georgia, USA, Anonymous ACM Press, New York, NY, USA, 119-126.

JONES, W., DUMAIS, S. AND BRUCE,HARRY CITEULIKE-ARTICLE-ID = {420503. 2005a. Once found, what then? A study of keeping behaviors in the personal use of Web information. *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 *In Proc. of ASIST 2005,* Charlotte, NC, Anonymous American Society for Information Science & Technology, .

JONES, W., PHUWANARTNURAK, A.J., GILL, R. AND BRUCE, H. 2005c. Don't take my folders away!: organizing personal information to get ghings done. In *CHI '05: CHI '05 extended abstracts on Human factors in computing systems,* Portland, OR, USA, Anonymous ACM Press, New York, NY, USA, 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,* San Jose, California, USA, Anonymous ACM Press, New York, NY, USA, 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,* Vienna, Austria, Anonymous ACM Press, .

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. .

KVAVILASHVILI, L. 1992. Remembering intentions: A critical review of existing experimental paradigms. *Applied Cognitive Psychology* 6, 507-524. .

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. *Judgment under uncertainty: Heuristics and biases* 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,* Vienna, Austria, Anonymous ACM Press, New York, NY, USA, 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. 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,* Monterey, California, United States, Anonymous ACM, New York, NY, USA, 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. 27, 481-504. .

MICROSOFT. Microsoft Excel. http://office.microsoft.com/excel.

MICROSOFT. Microsoft Office. http://office.microsoft.com.

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,* Amsterdam, The Netherlands, Anonymous ACM Press, New York, NY, USA, 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,* San Jose, California, USA, Anonymous ACM Press, New York, NY, USA, 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,* Atlanta, Georgia, USA, Anonymous Springer-Verlag, London, UK, 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,* Miami, Florida, USA, Anonymous ACM, New York, NY, USA, 149-157.

RHODES, B. AND CRABTREE, I.B. 1998. Wearable Computing and the Remembrance Agent. *BT Technology Journal* 16, 118-124. .

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,* San Francisco, California, United States, Anonymous ACM, New York, NY, USA, 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,* San Francisco, California, United States, Anonymous ACM, New York, NY, USA, 153-162.

ROSS, L. AND NISBETT, R. 1991. The Person and the Situation: Perspectives of Social Psychology. Temple University Press, .

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,* Vienna, Austria, Anonymous ACM Press, New York, NY, USA, 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,* San Jose, California, USA, Anonymous ACM Press, New York, NY, USA, 81-90.

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

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,* Seattle, Washington, United States, Anonymous ACM, New York, NY, USA, 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,* Vienna, Austria, Anonymous ACM Press, New York, NY, USA, 415-422.

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

TRAPANI, G. 2007. Lifehacker. 2007, .

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 *To appear in CHI 2008 Workshop on Semantic Web User Interfaces,* Florence, Italy, Anonymous ACM Press, .

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,* Newport, Rhode Island, USA, Anonymous ACM, New York, NY, USA, 193-202.

VAN KLEEK, M. AND SHROBE, H. 2007. A Practical Activity Capture Framework for Personal, Lifetime User Modeling. In *User Modeling,* Corfu, Greece, June, Anonymous .

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

W3. 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. 8, 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,* Vancouver, British Columbia, Canada, Anonymous ACM Press, New York, NY, USA, 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,* Ft. Lauderdale, Florida, USA, Anonymous ACM Press, New York, NY, USA, 401-408.

YEH, R., LIAO, C., KLEMMER, S., GUIMBRETI\`{E}RE, 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,* Montr\&\#233;al, Qu\&\#233;bec, Canada, Anonymous ACM, New York, NY, USA, 571-580.