

The Value of Persistence: A Study of the Creation, Ordering and Use of Conversation Archives by a Knowledge Worker

Christine A. Halverson
 IBM T.J. Watson Research
 krysa@us.ibm.com

Abstract

This paper argues that designers of Computer Mediated Communication Systems (CMCs) need to pay attention to the storage, organization and retrieval of conversations. It presents an ethnographic study of 'Bob,' an expert consultant to consultants in a large organization, and examines the ways in which he fashioned what he calls his "external memory pack" from the thousands of conversations he has had via instant messaging, email, and other forms of CMC. Particular attention is paid to the way in which he organizes, searches, and weaves together conversations to achieve his ends. We conclude by outlining the design implications of his use.

1. Introduction

A large organization is a distributed network of people who, through a set of procedures and processes, gain access to a private pool of resources which they recruit and deploy to achieve organizationally sanctioned ends [1]. Resources include capital, infrastructure, supplies, services, knowledge and skills. The procedures that provide access to resources range from formalized routines (e.g. creating and executing purchase orders, requesting marketing studies, and booking travel), to informal, ad hoc processes (such as borrowing equipment from another department, telephoning an associate for advice, or calling an impromptu meeting).

Conversations, whether occurring face to face, or via media which permit the conversations to span distance or time, play an essential role in supporting organizational procedures and processes. Even the most formalized procedures—for example, filing an expense report for a trip—may require conversations to clarify which organizationally defined categories certain expenses belong in, to work around bugs or errors in the execution of the process, or to support the occasional audits that are required by the organization's accountants. And, of course, the less formal processes are even more dependent on conversation.

We foreground this perspective because we are interested in technologies that support conversation among people in organizations. In pursuing this interest, we find ourselves in the company of many other investigators. Yates [2], in her seminal book *Control Through Communication*, traces the impact of various

machine-age communication and information technologies on the evolution of organizations from the nineteenth century on. Similarly, Sproull and Kiesler [3], examine some of the ways in which digital communications technologies such as email, effect communication in the organizational sphere. Most recently, a number of studies have investigated the rapid adoption of instant messaging applications in the organization, attempting to characterize its modes of use and the practices which are evolving around it [4-6]. The investigation we describe here is driven by many of the same concerns and interests. However, we shall focus more narrowly, in contrast to most of the other work in this area which, as we shall see, aims to produce general understandings of the role of communications technologies in large organizations. In fact, this paper reports on an ethnographic study of a single individual.

This individual, whom we will call Bob¹, is a domain expert in the consulting service arm of a large organization. Because of his depth of expertise, Bob functions as a consultant to other consultants within his company; that is, ordinary consultants on customer engagements talk with Bob when they encounter problems they can not solve. And Bob, in turn, calls on a wide network of contacts, as well as an extensive set of archives developed from previous interactions with his client-consultants, to solve their problems. As we shall see, this is a gross simplification of what is going on — Bob is *not* known to most who end up talking with him (there are over a hundred thousand people in his company's consulting service); Bob is not a solo worker, (although he works out of his home, he is a member of a distributed group that has a history and a set of negotiated practices); Bob's conversations are not just face to face (they occur over a variety of media and at a variety of paces); Bob conducts his activities so that they result in the generation of persistent archives which he regularly accesses; Bob's 'solutions' to problems are rarely cut-and-dried answers (the problems that reach him are typically too complex for that). And Bob does all of this without using any paper.

Clearly there can be no hope of generalizing from a study of a single individual; however, we argue that our focus on a single individual offers advantages that compensate for the loss of generality. First, it makes it

¹ Bob, other names of individuals and companies are pseudonyms.

possible to gain a deep understanding of the individual and his work contexts, within which his communicative practices can be situated. Second, it makes it possible to examine Bob's use of IM and other CMC technologies against this background, to see how the supported communicative practices play out in that context. Third, we can examine Bob's use of IM and other media over time, looking both at the question of whether and how conversations are saved, and how they are retrieved and used. Fourth, as von Hippel [7] points out in his work on "lead users", examining the practices of pioneers or early adopters can provide insight on how technologies may be adopted by the larger population in the future. Finally, an examination of the practices of lead users can provide interesting insights on designing technologies.

We begin by reviewing work on communication in the workplace; we focus, in particular, on instant messaging (IM), as that is one of Bob's primary means of communication. Next, we describe our approach, and then move on to a close examination of Bob and his practices. We begin with a description of the settings within which Bob operates; this will help in making sense of what Bob does and why. Next we do a walk through of Bob's communicative practices and tools, and then look at the way in which Bob uses his tools and resources to deal with a single incident. We examine the kinds of conversations upon which the resolution of such an incident depends. We end with a look at the ways in which his reuse can be seen as a series of persistent conversations, rather than as information retrieval.

2. Related work

Initially, CMC in the workplace meant email [8, 9]. Going beyond email, early studies of CMC were primarily focused outside the workplace, looking at chat rooms, discussion lists (DLs) and MUDs [10].

More recently, investigators have examined the use of asynchronous CMC in the workplace. For example, Churchill and Bly [11] report on the use of a MUD over the course of several years by personnel at Argonne National Laboratory. This MUD is used for both work and social talk, and supports opportunistic encounters, planned interactions, and coordination. Reporting on a more persistent type of system, Kovalainen, et al. [12] describe the use of an electronic diary in a Finnish paper mill. They argue that the diary entries constitute dialogues within and between work shifts, and that these dialogues share some characteristics with conversations used to coordinate work in face to face situations. Finally, systems such as Babble [13] and Loops [14], designed specifically for use in the workplace, combine persistent text with the lighter weight interaction style of instant messaging. A six month field study of Babble's use by six work groups [15] demonstrated usage patterns spanning the spectrum from focused work to coordination to social talk.

The recent explosion of IM, in and outside of the workplace, has sparked a host of new studies about how IM has been used to meet the needs of people immersed in large organizations. Using a range of techniques including interviews, observations and log analysis a number of researchers have found that IM use encompasses a wide range. This summary is necessarily simplified but Nardi et al. [6] find that IM is to check availability, and negotiate times, places and channels for work-focused communication. Isaacs, et al [5] found the primary use of their system was information exchange for work conversations, and that secondary uses were for simple, single purpose interactions (28%) and for scheduling or coordination (30%). Handel and Herbsleb [4], again found the content of a semi-persistent group chat system was focused on work talk and negotiating availability. Despite its use by six globally distributed work groups, they found that it was primarily used for bursts of synchronous chat.

These studies are all focused on broad patterns of use that occur across conversations. Deeper examinations tend to focus at a micro-level within conversations, such as Volda et al's [16] look at IM using sociolinguistic analyses, and Brennan and Ohaeri's [17] analyses of chat. Ideally we want to bridge these levels and understand what is going on not only within a particular conversation, but also as those conversations are embedded in one's work.

In summary, there is ample evidence that conversation-oriented systems are used, and are useful, in the workplace. They support planned work interactions, opportunistic interactions, scheduling and coordination work, and social talk. However, work covers a longer temporal extent than the immediacy of conversation. We argue that computer mediated applications must also support ongoing dialogs within work groups, self reflection, and providing an archive that acts as external memory for individual and group. Bob is one window into how this can be done.

3. Data Collection and Analyses

The primary method for data collection was ethnographic. A series of semi-structured interviews led to the author spending three weeks observing Bob and three weeks with his team lead (week long blocks spread over 9 months). Both of these involved trips to their work sites—the individuals' homes. In addition, five days (over two trips, separated by 9 months), were spent at the call center that distributes and directs calls to Bob and his colleagues. Data was recorded by note and videotape. Notes were transcribed within 48 hours, and reviewed against the tapes for accuracy. In addition, the videotapes were analyzed separately. We also have access to several of his corporate resources: those databases commonly available, as well as the more restricted, group-specific, TeamRoom. The content of the latter was sampled and

analyzed. Analysis was based on Hutchins' [1] framework of distributed cognition. Sampled TeamRoom content was coded based on ethnographically derived categories. We will introduce those categories and the coding method later in the paper.

4. The Site: Company, Group, Home

4.1. Global Corp

Bob is an employee of the services arm of Global Corp, a very large computer software and hardware company. Bob is a technical consultant in their Novel Business Services (NBS) area. More specifically he is the member of a *practice*² [18]—Web Business Application Development (WBAD)—that helps customers build web based enterprises. Creating projects supporting novel business approaches is an opportunity to sell Global Corp.'s software, but it often means that Global Corp.'s software usually needs to be customized in order to link to legacy systems and third party software. This customization is a development process, and so NBS is organized into a number of communities of practice, centered around particular areas of expertise.

Bob has 15 years of experience at Global Corp, starting as a programmer of software products. As a consultant he went out on customer engagements, and over time he worked his way into an elite team in his practice, the WBAD Center of Competency (CoC), which means that he is among the top 20 or so consultants in his practice. The CoC's mission ranges from examining meta-level issues that impact the practice as a whole, to providing crisis intervention for troubled projects. Bob was a member of such a team, "parachuted" in to solve difficult problems. Recently, a part of the CoC team was tasked to be part of a new enterprise that is providing help for other consultants in the field: the Technical Engagement help desk (TE).

As a consultant, Bob often worked at customer sites. It is no accident that his house is only a few minutes from the airport, as easy access was important. Now however, he works out of his home, with a team of people from the CoC that he has worked with for quite some time. They, of course, are not in his home, except virtually. They are spread across the United States. Chris, the team lead, is in Colorado, Bob is in North Carolina, Bill is in California, Anoop is in Michigan, and Harry no longer maintains a home, taking his free time in whichever city he happens to be currently. Bob's network of colleagues from previous consulting work is similarly spread across the country and the world.

4.2. Technical Engagement Help

It will help situate Bob's extensive creation of conversation to understand the basic process of the TE

² Practices are based on the notion of a community of practice

help desk. A consultant with a technical problem has several possible ways of contacting the TE help desk, but we will only focus on one here: calling up toll free number with a request. This call is routed to a call center. If possible, they answer the question, however, it is more likely that they determine which practice the question belongs to, and then refer the call to that practice's team lead. The team lead then assigns it to a Subject Matter Expert (SME) within her team.

Due to their experience, each SME has a number of resources within Global Corp that are unavailable to most consultants. Bob is such a SME and he has access to a number of human resources that are spread across the country. Obviously communication is a major part of his day, as he connects to his team members, his resources, and the consultants with whom he is working. In this paper we focus on Bob's more private resources, looking at how they spread into more public spheres.

4.3. Bob's Work Site: Home Sweet Home

Bob gets up and sets up his physical tools on the dining/kitchen table. He has a laptop and two phones on the table next to him. One is a cell phone with free incoming minutes—this is his business phone, where calls for TE's help desk come in. This phone has a speaker on it—which he uses when he answers in order to keep his hands free for typing. The second phone is a mobile handset for his home phone, with a headset, again to keep his hands free. What is *not* there is much scratch paper. Bob tends not to write notes to himself on paper³. Instead he writes them on line.

5. Bob's World: Practices and Tools

What does it mean to be a consultant to consultants? Do you fly twice as much or dress twice as nice? Are you twice as smart? Or perhaps you know twice as many people? At least in Bob's case, it may mean that you have twice as many conversations, and you save them.

On the surface, Bob's job is to answer questions. Watching him work, his job sometimes appears to be *finding* an existing answer, by navigating to it in a database or by tapping his network of contacts. Or he may generate an answer by putting existing pieces together, either from data sources or by talking with his contacts. Occasionally he is a creator, when Bob writes a sample piece of code.

But Bob's job is not just finding or producing information. Bob talks about his job as often more "social"⁴ than otherwise. For example, a consultant may realize that they are having trouble with a production server. Bob helps them find out that they need to rebuild

³ In all of our observations we have only seen Bob write one note on paper, a phone message for his wife.

⁴ Social is Bob's term, although we might interpret it more as organizational.

the server with the latest fixes. However, now the customer is balking, and the consultant wants to know about how to get around the need for a rebuild. Instead, Bob helps the consultant negotiate through the social and political aspects around the issue, rather than just solving the technical problem. His answer covers how to handle the customer and convince them of the value of the rebuild, rather than trying to avoid it. Bob also sees part of his job as being a mentor, helping consultants learn how to work more effectively. His responses are full of pointers to information that go well beyond the root issue. All of these aspects of his work require conversations. Therefore, what facilitates his job is building and maintaining relationships across the company, even though it may not *be* his job.

In the following sections we step into Bob's world, starting with his daily setup routine. This will provide the context to introduce the tools he uses, and how he creates and uses the persistent text from the residua of CMC for a number of ongoing processes. In Bob's hands, many of his tools blur the line between communication media and repository. It is the computer mediated part of CMC that allows this, and provides the basis for his reuse.

5.1. Daily Practice

In the morning, often while eating his breakfast, he gets started—booting the laptop and setting up his environment. He manages the large number of applications and windows by using a shareware application that allows him to create a number of virtual desktops and page through them. As he starts up each application (or collection) this program lets him assign applications or documents to a particular virtual “page” or desktop. Switching from page to page automatically minimizes the current applications and brings up the other ones. This provides him the facility to manipulate a large number of programs without having them all cluttering a single desktop.

5.1.1 eMail. He begins by checking his email. Bob uses two email clients, one internal to Global Corp. and Netscape for external mail. While the external mail is a personal account, in fact he uses it to keep up with a range of technical information that he uses in his job—primarily subscriptions to various technical discussion lists. Only about 10% of the mail he gets daily is potentially useful. If he has time, he will scan through the messages. However, read or not, those related to technical issues are filed in a separate mail folder. The Netscape mail client he uses highlights unread messages in red. This allows him to come back to this folder when he has some free time and skim through the postings. In some sense these are an ongoing, if mostly one-sided, dialog that he has with the members of the various discussion lists. It keeps him up-to-date on the zeitgeist of various technical worlds. Having them saved also allows him to locally search for information that he

remembers having seen. Being segregated in a folder means that he can speed up a search by restricting the scope. In addition, he can use Netscape's full text search. Having read them he often remembers *when* he read it as much as *what* he read, or *where*. The date-wise organization of the email received helps him to skim for something that rings a bell, but for which he might not have remembered the right phrase.

After the external mail, he moves to his internal mail client, Lotus Notes. This is another application he brings up early, and assigns to its own virtual page. Global Corp. has a cap on how much mail he can save. Bob usually has trouble with this limit (175 Mb) because he also uses Notes to save a record of what projects he has worked on. Any time he receives email about work with an external customer, he files it in a folder bearing that company's name. In some cases, over his career with Global Corp. he has interacted with the same company in many different roles. In his mail however, each “conversation” has been saved in a company folder. He does this so that at the next question or engagement he can come back and check out details. Notes is also a platform for a number of databases he uses—including corporate-wide databases (e.g. such as product bug reports), databases specific to the services arm (e.g., the one he uses for recording his working hours), project or informal information databases, as well as project and task specific TeamRooms related to his team and job. We look at one of these TeamRooms more closely below (see section 5.1.3).

5.1.2 Instant Messaging. While Bob may spend more time first thing in the morning going over mail, overall he spends more time using his instant messaging (IM) clients. A separate virtual page is set up for his IM clients. Many of his “customers” – the Global Corp. consultants – are often at client sites without access behind the Global Corp. firewall. Thus, the corporate IM client, SameTime, is not always sufficient for contact, although it is the most populated “buddy list” he has. He also has clients for Yahoo, MSN, and AOL.

Bringing up all of his IM clients is one of the first things that he does in the morning. Once he is connected he is drawn to IM over and over again. Obviously much of the draw is the ability to have quick connections with his broad range of personal resources. People IM him for information, or coordination, and he IMs others. His IM conversations interleave social with work questions as he maintains his relationships, even when asking simple questions.

However, Bob's buddy lists are another resource for him. For each problem, if he attempts to contact the consultant via IM he saves them in his buddy list, even though he may never communicate with them. In this way the buddy lists become a master list of all the consultants with whom he has dealt. A similar story could be told with respect to his other contacts,

colleagues and resources. The most recent count is 427 unique names on SameTime alone⁵. He has fewer contacts on each of the other IM clients—Bob estimates between 60 and 100 names on each. While some are personal (non work related) contacts, the majority are secondary handles for consultants to stay connected when they do not have access behind the firewall. In this sense, they tend to duplicate names on SameTime.

Bob has also saved every IM conversation from his internal client over the last three years, and where possible from external clients. Unlike his email, he saves these conversations by year. Generally, he can remember what year something happened, and then he can search more quickly within that year. He takes advantage of the operating system's ability to do full text search of indexed folders. He also takes advantage of the fact that SameTime saves each chat file with information about it (conversational partner, date, and time) in the file name. When a topic search comes up empty, he will search again using the name of the messaging partner (if he remembers). When that does not work, he may remember something more about the date, for example a particular month. In that case he may broaden the search out again to all the years, but search for the month.

This saving behavior enables him to locate information locally and respond quickly. It also provides grist from which he extracts conversation excerpts for the record, thus providing detail and context that a simple summary would not.

5.1.3 SPOC Team Room. TeamRooms are designed to be a shared resource for a work group. Their Notes heritage means that they have many of the mechanisms of a database, including key word search, and the notion of replication. Built into TeamRooms, via the database, is the ability to sort and search on the creator of an entry, the date of creation, the date it is assigned to be due, the kind of entry it is and a number of user created categories. However, one of the problems is that the notion of category is quite limited, and only one set of categories can be designated.

The structure of the SPOC TeamRoom duplicates Bob's method for handling his own mail system—that is being able to keep track of information by client company name, as well as by type of entry. This makes sense as the practice often deals with the same companies in different contexts as well as different engagements. However, this did not come for free—there were some technical problems that required Bob to customize the SPOC entries to support this.

While the TeamRoom might not seem like CMC, it provides that role between team members over time. The SPOC TeamRoom is *the* place to record information that may be useful to other members of the CoC. Over the last year, a major part of their work is involved with TE

help, so all of the problems handled for TE help are maintained here, as well as in the official application of record. Having the records here supports a number of communications between the members of the CoC. For example, entries in the SPOC TeamRoom provide Chris, the team lead, a way to monitor the solution of some problems for her weekly status report—without taking up her SMEs time. While this can also be true in a co-located work environment, there is a greater impact among remote workers. The TeamRoom record of a problem solution (or just process) provides a rich summary for the next SPOC team member who has to deal with it.

5.1.4 NoteTab Light. One tool that is critical for Bob, but does not look like a communication tool, is a shareware text editor called NoteTab Light [19]. NoteTab Light provides a number of advantages for him. It allows him to automatically access a number of open files all at once with a tabbed interface. And it provides a number of easy formatting transformations vital when switching from one application to another, particularly for moving HTML to “flat” text. Within that tabbed interface he also has the capability of full text search. More importantly, it provides a vehicle for Bob to have conversations with himself.

NoteTab Light is an application that he uses in a number of ways. Among the tabbed files are several which act as a combination of To-Do list and catchall for the electronic equivalent of notes on the back of an envelope, both personal and work related. While NoteTab Light is not the first application he opens, it is often where he starts his day, referring frequently back to the continuous To-Do lists he maintains. Equally, it is a personal archive of information which he searches through to direct his problem solving. Throughout his day as he finds resources—urls, databases, names—that he wants to not lose, he *harvests* them, adding them to the appropriate tab in NoteTab Light.

6. An Example: Constructing a Record

In this section, by following Bob through a particular example, we elaborate how his use of these tools creates the residua of a number of conversations that fit together during a problem solution. The persistence of many of these conversations, coupled with the range of timescales and individuals with whom the conversations occur, provide the means for Bob to reconstruct and officially document the work he does.

Monday mornings are generally slow, which is good because Fridays, when calls to the call center peak, are often frantic. This means that by Friday afternoon Bob is usually busy, just at the point where he wants to be winding down for the weekend. He often continues to work a problem right up until he is walking out the door for a weekend engagement, and thus he does not always get everything written down in the official record. So

⁵ With duplicates Bob counted 484, as of June 2003.

come Monday, he may have to reconstruct the record of the last few calls on Friday. This is one example of that.

As Bob goes through his To-Do list for TE help in NoteTab Light, he sees a cryptic notation referring to ticket number 1234, with no other details. He then begins cycling through his other applications, looking for that number. He checks the SPOC TeamRoom and notes that it is not logged, so he will have to reconstruct what he did and thus, what needs to be done. He begins looking for the original text that came when the problem was assigned to him from the team lead. The usual model is that when his team lead Chris assigns it to a SME, she tries to connect the consultant and the SME in a number of ways. One of those is always a semi formal notification via email that the problem is being worked, and that it is assigned to a particular SME, who will be contacting the consultant. While the format of these notes is not formally prescribed, they do tend to be uniform in the information they contain: that is the ticket number, the consultant's name and contact information, and the client company. Searching on the number in Notes yields nothing, so he starts scanning his inbox for email from Friday. He finds it that way, and begins to create a record in the SPOC TeamRoom, copying and pasting the contact information he finds in the email.

Having found the consultant's name, his next step is to check for any IM conversations he had with the consultant on Friday. This kind of check is such a common occurrence for Bob that he has access to his chat folders set up on the start menu in Windows. Selecting the current year he uses Windows' search feature to search for the consultant's name. The search turns up nothing. From this Bob concludes that he must have only talked to him by phone. (This inference is only possible because Bob saves every chat—no record, no contact via IM.) Just to be sure (e.g., in case there was a misspelling), he scrolls through the listed chat files for that week and confirms no chats with the consultant.

He now goes back to the record in NoteTab Light and begins skimming notes from Friday again. He notices a remark that shows he talked to the consultant by phone. Now he remembers that he reached him on his cell phone. The consultant said he was recovering after an operation, so Bob told him he would catch up with him after the weekend. Now Bob has an idea of what happened Friday, and what needs to happen next. Bob then switches to the record he started in the SPOC TeamRoom. As he adds to the record he switches back and forth between the minimal record in NoteTab Light and the growing one in SPOC. Having started by pasting the contact information from the email, he now begins to add more information. As he types it sparks additional memories about the consultant's circumstances which he puts into the record. Once he reconstructs the record for the SPOC TeamRoom, he then goes on to construct the record in the more official record for the TE help desk.

6.1. Analytical traces

Much of his memory, and the reconstruction of the problem, is based on a range of conversations he had in the process of solving the problem. A detailed DCOG analysis was used to trace through numerous examples to construct a generic engagement with the TE help desk. (Detailed figure removed due to space constraints). Beginning with the consultant's first contact and continuing as the problem works its way through the process for each step we can detail the conversations and their representations and supporting media. What is interesting is how almost every step along the way involves a dialog. Some dialogs are with the other members of the SPOC group or the TE help desk staff, while some are with the consultant, or individuals who have resources necessary to answer the question. As we have seen, many such dialogs are recorded via CMCs. While the conversational records in this specific example are sparse, where Bob expects them is evident from where he looks.

However, there are interactions that are more ambiguous as conversations. Some searches of discussion lists and FAQs can be viewed as a longer term discussion with the writers. After all, their longevity is often one reason why they are written and posted. Some of the more official and public records that Bob writes can be viewed in the same way. They initiate a discourse on a very long time scale that may not be picked up until some time in the future. Participating in a conversation may be when a colleague adds something current to a record. Or, in the future, when they search for previous experience with a client and decide to contact the author through another mechanism, such as IM or email. Sometimes the discourse dies because the conversational partner is no longer available. However, we would argue that the residua of that conversation allows the conversation to live on in a sense—even if it is only in a self-reflective dialog between the reader and the otherwise unavailable author.

7. Components of an Event

A simple reconstruction like the one above provides a view into the process that creates a record. However, the simplicity that makes it a good expository example means that it is not the best example of the rich reuse of persistent text that we see in Bob's daily practice. To examine the extent of his (and others') use, we took a closer look at the content of the SPOC TeamRoom.

Overall, the SPOC TeamRoom contains 3570 entries, 2689 of which are related to the TE help desk. We sampled these by searching for instances of 'chat'⁶ or 'email' in the text. The resulting sample of 223 entries

⁶ Chat here is mostly referring to IM. Those cases where it was explicit or an apparent reference to phone contact were so coded.

(8.3% of the TE total), was coded for twelve categories (see Table 1).

Table 1 Categories for SPOC Team Room sample

A. Once per entry	
1. email excerpt	6. no chat possible
2. other excerpt	7. best contact is chat
3. chat excerpt	8. Phone contact
4. PMR excerpt	9. Assume phone
5. contact initiated through chat	
B. Coded multiple times per file	
1. chat file attached	2. lists chat file name

Each entry could be coded more than once for matching different categories, but could only be coded once for each category. The two exceptions were the reference to chat files in the body of the text or chat files attached. These were coded once for every file. Only one entry had more than one chat excerpt. The more usual case for multiple codes are to code for chat and another excerpt, or an email excerpt requesting further contact via chat. (Note that these records were spread across all the members of SPOC, so the use of persistent text originating elsewhere is not unique to Bob.)

Some larger patterns emerged from this analysis. The most common use was excerpting email (29.6%), most often email attempting to follow-up as a record for work done. More than two thirds of these emails contained the comment that the best way to contact the SME was via online chat. The two next most common contents that were explicitly excerpted were: PMRs⁷ (19.7%) and chat (19.3%). Chats having occurred were referred to in another 20.1% of the records and saved chat files were referred to by file name in 5.4%. Finally, 4% of the records had chat files attached.⁸ Overall, 81% of the entries analyzed had some sort of excerpted text or a referent to where the original text could be located.

Why are these so interesting? For two reasons. One, we see his personal practices carried forward into other more public locations. In the reconstructed event we see how he mines his personal world to construct the record of what happened. By surveying the SPOC team room we see how such a record is finalized and eventually gets spread – at least to his colleagues. While we can make no claims about how Bob’s usage is transmitted, we do see evidence that the other members of the SPOC group similarly embed the residua of conversations in their records. This points to how while Bob may be leading the way, his colleagues are not far behind.

⁷ PMR stands for Product Management Record and is one way that GLOBAL manages problems and solutions of products in the field. Its easiest to think of these as a record of bug reports, although they contain much more in their text. PMRs are only maintained for 60 days after closure—one reason why SME’s include excerpts.

⁸ The data was recoded counting chat file references and attachments only once to calculate these percentages.

8. From Bricolage to Design

Bob is leading the edge, a *bricoleur* of CMCs, putting together tools in a way to accomplish his own needs, and thus makes new use possible [20]. As a bricoleur, Bob leads the way in showing how technology can be put together for new uses, satisfying the as yet unmet needs of the practice. However, our sampling of the SPOC database shows that he is not the only one who is saving text from various CMCs and repurposing it. Our advantage with Bob however, is that our fieldwork exposes a level of detail that helps us understand the needs that drive his adaptation and from which we can draw implications for design.

Figure 1 lays out his virtual world in the midst of his physical and conceptual world. The hexagons in the center of figure 1 are all the various applications that he uses. These are the tools at the foundation of his virtual world that are used in a number of contexts. We’ve mapped out his interactions using these tools along two primary dimensions. One dimension is whether the text is personal or public (vertical axis). The other is whether he is looking inside or outside of Global Corp (horizontal axis). Bob’s interactions using these tools serve to link all the various contexts of virtual, physical, and conceptual worlds, together.

For example, while a Google search operates on text that is outside of Global Corp., and outside of Bob’s personal and private world, excerpts in the form of urls end up in his NoteTab Light record, as well as in the more public (to others in his team) SPOC TeamRoom. From there they may migrate further within Global Corp. by being included in an FAQ, and probably outside of Global Corp., as the solution gets passed to an external customer.

In contrast, some items remain completely private, while others move only generally into a more public sphere. Bob’s chats, whether with one person or many, get saved locally. As he says, he uses them as an *external memory pack*, which can then be reused in a variety of ways. In some cases (e.g.the SPOC TeamRoom) they remain relatively private. They are referred to obliquely in the text, or more explicitly by file name. In either case, other members of the team with sufficient need can get to the information, but they have to go through Bob. In this way they remain private, and Bob retains control over the data. However, when he excerpts them, or attaches an entire file, he is opening them up, at least to the scrutiny of other team members. From there they can spread.

One of the key issues this raises is how aware or concerned the SPOC team members are with the potential sensitivity of such texts. An implication of this is providing help for how and when to easily make reused text anonymous. Clearly, locally saved and used texts do not *need* to be made anonymous. However, because information tends to propagate once it is in

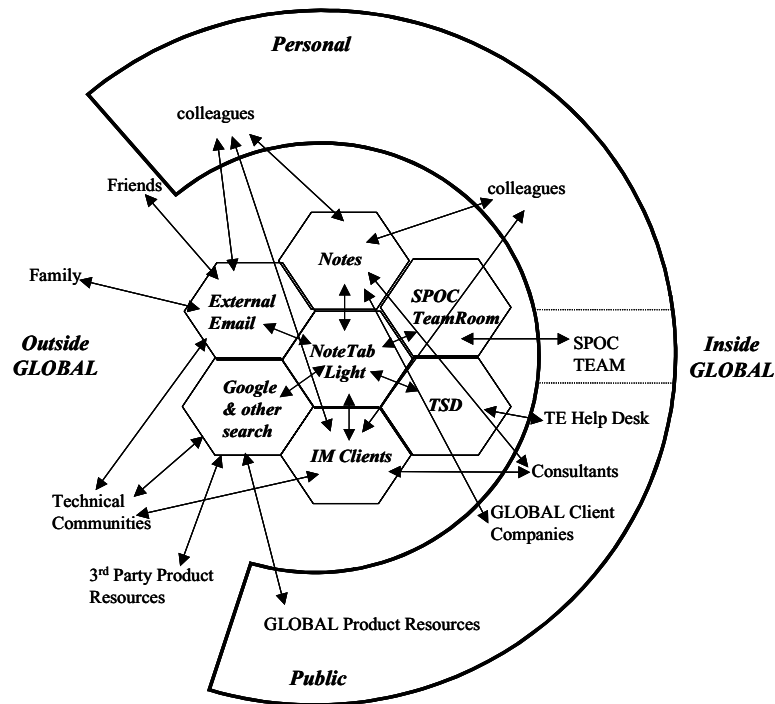


Figure 1. Bob's Virtual World, situated within his physical and conceptual world. The hexagons at the origin are the tools that make up his virtual world. The arrows between these tools and the categories of contacts illustrate the range of conversations he participates in and the range of tools used.

public, making a conversation anonymous trivially would be a boon.

One idea is if the underlying CMC system tagged and identified parts of the conversation so that an automated process of anonymization could be run on a subset of the tagged items. For example, meta information that might be tagged includes participants, start and end time of the discourse and the medium. Within the content it would also be good to tag and recognize personalized information such as names, phone numbers, email addresses and so on. Of course this raises issues when the intent of the saved information is to pass on specifics of another person – in which case you do not automatically want it to disappear when you prepare to send the conversation on to someone else.

Tagging the content in this way might also aid another issue of use—that is the cross-compatibility of file formats and their openness to various search processes. Text needs to be easily movable across applications, and therefore across file formats. To reuse text Bob is repeatedly converting from HTML to ASCII, or reformatting ASCII text for readability. Having this happen seamlessly, so that Bob could copy from a webpage and paste it into Lotus Notes and redact it

without first converting would save a lot of time. Time that could be spent adding to the content he's cutting and pasting.

With that kind of easy conversion, then perhaps a standard set of tags for tagging conversation could be used. This would aid the earlier issue of anonymization of the saved text. It would also enhance the possibility to search using text that normally doesn't have preset key fields. For example, Bob takes advantage of the fields in Lotus Notes for search and his browsing of IM filenames shows how having them is a good mental marker for retrieval. Tags in the conversation would provide a similar advantage for the content of saved IM chats.

However, Bob's repeated use of full text search indicates that we need to go farther. Text should always be searchable with a full text search engine in an easy way—that is easy to use rather than requiring people to learn Boolean operators to write a query. Bob's extensive use of regular expression matching points to the importance and power of fuzzy searches.

A conversational tagging structure could also provide the foundation for more flexible saving and the possibility for future visualization. Right now, Bob deals with the lack of structure appropriate to his needs within

an application by elaborately structuring what he can control, often outside of an application. Visualization could help Bob, both by visualizing what he has to enhance browsing and by helping him annotate for future use. Both could be done on the level of individual conversations as well as collections.

9. Conclusion

Bob's position as a member of the CoC, and my ability to gain access to him, are because he is seen by his management as a paramount multi-tasker, general fount of knowledge, and good communicator. One reason Bob is so successful is that he often harvests items of interest whenever he finds them, whether they are pertinent to his immediate task or not. In doing this he makes them both persistent and locally available. He does not structure them much to aid subsequent search, because he stores them where he can use full text search with regular expression matching.

Unfortunately he can not put everything in the Note Tab Light application and many of his other tools are limited when it comes to searching. His more elaborated structuring of other applications are to get around the limitations of their search mechanisms—both within and across the applications. There are many ways that we as designers could address these lacks – having seen their importance. Only some of these are elaborated above.

The fact that Bob needs all these different methods of structuring illustrates several other points. One, is that the applications themselves are insufficient to provide all the mechanisms that Bob finds necessary for future retrieval. Another less obvious one, is that while he has been successful at cobbling together a system that works for him, it does not necessarily work as well for others. He was able to mold the SPOC TeamRoom in a way that mimicked his personal email organization. However, a survey of his team members on TE help finds that they like it less, in part because they find its structure too complicated.

As various forms of CMC have moved into the workplace we begin to see repeating patterns. First is the concern about how it will change the workplace for the worse. Then there are studies that show the potential power, and the pitfalls, of the new medium. At the same time, but separate from these studies, is the spread of use as people like Bob discover how the new medium does or does not support them. Previous research has focused on one particular application: IM, or email, or MUDs.

And then there is Bob. Bob melds several CMCs through many different processes.. He uses them as an external memory pack – both private and public, in content and in structure. As such they become a teaching tool as he constructs saved bits and pieces into a coherent whole that not only answers the question, but expands it

into a mini-tutorial. Email is often excerpted as a record of what he has done—more as a CYA⁹ accountability kind of trace than anything substantive. IM messages seem to be reserved for problem definition, clarification and solving, as well as coordination and socializing. He uses NoteTab Light to keep track of TO-DOs, to maintain notes to himself, potentially useful URLs and to convert HTML to ASCII text.

We have come to accept that the modern knowledge worker lives in email [21]. Soon we may have to accept that they live in a number of other applications too. And what does that mean? If Bob's use is an indication, it means that, for example, IM is not just an ephemeral vehicle for conversation. IM's possibilities for persistence via computer mediation provide the opportunity and support for a variety of subsequent uses. Bob saves his IM conversations. He structures how he saves them in order to be able to retrieve them – taking advantage of, and getting around, aspects of the quirks of applications and operating systems.

He does this because the value in the conversations are not just from having them—but in being able to refer back to them and reuse them. For Bob the same is true of his other 'conversations', whether it be email or the self dialog he has with his personal notes or the writers of FAQs and DLs. IM is not unique for Bob, nor is it probably for the rest of us.

Bob is unusual in many ways: his technical skill and willingness to search for and cobble together many technologies, a job for which performing that overhead makes sense, and perhaps for not making notes on paper. However, he is not so unusual that we can not see the seeds of a possible future there for all of us. Bob's example opens our eyes, not just to having conversations, but designing in order to support doing something with them later: that is, the storage, browsing and retrieval of conversation. Right now, Bob is going to great lengths in order to do so. That need not be the case in the future.

We have several sign posts for what direction to proceed. Bob himself, and his practices, are one sign post. The visualizations done by researchers to get an overview of large bodies of text are another [16, 22-24]. Bob's use is at a different level than these largely analytic visualizations. Bob's practice challenges us to alter our designs. Perhaps with help we will all someday be like Bob – carrying around our external memory pack on our laptops.

Acknowledgements

Our grateful thanks to Bob and his colleagues in the TE Help Desk. Without them this work would not have been possible. This paper benefited greatly from conversations with Tom Erickson, Ed Hutchins, Jim Hollan, Mark Ackerman and Wendy Kellogg. Thanks also to Juerg von Kanel and anonymous reviewers for their thoughtful comments. This project has been funded, in part, by NSF grant IRI-0124878..

⁹ CYA = cover your ass

10. References

1. Hutchins, E.L., *Cognition in the Wild*. 1995, Cambridge, MA.: MIT Press.
2. Yates, J., *Control through Communication: The rise of system in American management*. 1989, 1993, Baltimore, MD.: Johns Hopkins University Press.
3. Sproull, L. and S. Kiesler, *Connections: New ways of working in the networked organization*. 1991, Cambridge, MA.: The MIT Press.
4. Handel, M. and J.D. Herbsleb. *What is chat doing in the workplace?* in *Computer Supported Cooperative Work, CSCW02*. 2002. New Orleans, Louisiana: ACM.
5. Isaacs, E., A. Walendowski, S. Whittaker, D. Schiano, C. Camm. *The Character, Functions and Styles of Instant Messaging in the Workplace*. in *Computer Supported Cooperative Work, CSCW02*. 2002. New Orleans, LA: ACM.
6. Nardi, B.A., S. Whittaker, and E. Bradner. *Interaction and outeraction*. in *CSCW*. 2000. Philadelphia, PA.: ACM.
7. Von Hippel, E., *Lead users: A source of novel product concepts*. *Management Science*, 1986. **32**(7): p. 791-805.
8. Eveland, J.D. and T.K. Bikson, *Work group structures and computer support: a field experiment*. *ACM Transactions on Information Systems (TOIS)*, 1988. **6**(4): p. 354-379.
9. McKay, W.E., *Diversity in the Use of Electronic Mail: A Preliminary Inquiry*. *ACM Transactions on Information Systems (TOIS)*, 1989. **6**(4): p. 380-397.
10. Bruckman, A., *Identity Workshop: emergent social and psychological phenomena in text-based virtual reality*. 1992.
11. Churchill, E.F. and S. Bly. *Virtual Environments at Work: ongoing use of MUDs in the Workplace*. in *WACC*. 1999. San Francisco, CA: ACM press.
12. Kovalainen, M., M. Robinson, and E. Auramaki. *Diaries at Work*. in *Computer supported Cooperative Work, CSCW 98*. 1998. Seattle, WA: ACM.
13. Erickson, T., D. Smith, W. Kellogg, M. Laff, J. Richards, E. Bradner. *Socially Translucent Systems: Social Proxies, Persistent Conversation, and the Design of Babble*. in *Human Factors in Computing: The Proceedings of CHI 99*. 1999. Pittsburgh, PA: ACM Press.
14. Halverson, C.A., T. Erickson, and J. Sussman. *What Counts as Success? Punctuated Patterns of Use in a Persistent Chat Environment*. in *GROUP 03*. 2003. Sanibel Island, FL.: ACM Press.
15. Bradner, E., W.A. Kellogg, and T. Erickson. *The adoption and use of 'BABBLE': a field study of chat in the workplace*. in *Proceedings of the Sixth European Conference on Computer-Supported Cooperative Work*. 1999. Copenhagen, Denmark: Kluwer Academic Publishers.
16. Voids, A., W.C. Newstetter, and E.D. Mynatt. *When Conventions Collide: the tensions of instant messaging attributed*. in *CHI 2002*. 2002. Minneapolis, MN: ACM.
17. Brennan, S.E. and J.O. Ohaeri. *Why do electronic conversations seem less polite: the costs and benefits of hedging*. in *WACC '99*. 1999. San Francisco, CA: ACM.
18. Wenger, E., *Communities of practice: Learning, meaning, and identity*. 1998, Cambridge: Cambridge University Press.
19. *NoteTab Light*. <http://www.notetab.ch/ntl.htm>
20. Lévi-Strauss, C., *The Savage Mind*. 1962, Chicago: The University of Chicago Press.
21. Ducheneaut, N. and V. Bellotti, *E-mail as habitat: an exploration of embedded personal information management interactions*, 2001. **8**(5): p. 30-38.
22. Donath, J.S., *Identity and deception in the virtual community*, in *Communities in Cyberspace*, M.A. Smith and P. Kollock, Editors. 1999, Routledge: London. p. 29-59.
23. Donath, J.S., *A semantic approach to visualizing online conversations*. *Communications of the ACM*, 2002. **45**(4): p. 45-49.
24. Smith, M., *Tools for Navigating Large Social Cyberspaces*. *Communications of the ACM*, 2002. **45**(4): p. 51-55.