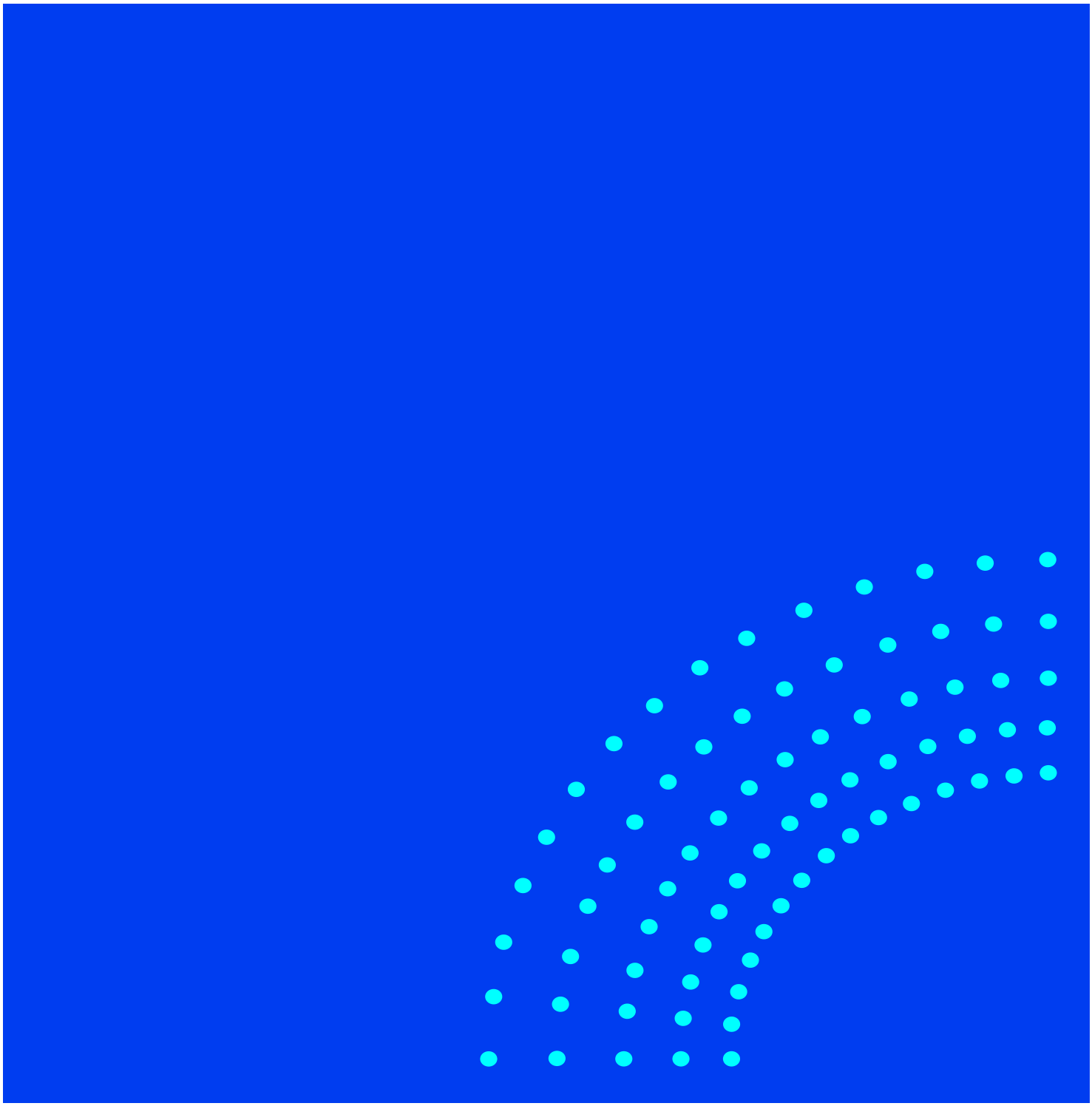
Master's Thesis in Informatics

Personal information management now and in the future

Tim Gahnström

Geneva, Switzerland 2004



REPORT NO. 2004/51

Personal information management now and in the future

Tim Gahnström

Department of Informatics
Göteborg University
IT UNIVERSITY OF GÖTEBORG
GÖTEBORG UNIVERSITY AND CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2004

Personal information management now and in the future

Tim Gahnström

© Tim Gahnström, 2004.

Report no 2004:51

ISSN: 1651-4769

Department of Business Technology

IT University of Göteborg

Göteborg University and Chalmers University of Technology

P O Box 8718

SE – 402 75 Göteborg

Sweden

Telephone + 46 (0)31-772 4895

Chalmers Repro

Göteborg, Sweden 2004

Personal information management now and in the future

Tim Gahnström

Department of Informatics, Göteborg University

IT University of Göteborg

Göteborg University and Chalmers University of Technology

SUMMARY

Decision makers today do often not have the problem of yesterday, lack of information. Instead they have an abundance of information flowing through their organizations and computers but the tools for managing this data are not good enough, much precious working time are spent organizing and locating data with inferior tools. The purpose of this study was to characterize the gap between the current operational tools for personal information management and the tools available in research laboratories. This was done as a case study at UNOSAT. The first conclusion of the study is that the researchers are far ahead in functionality. They have tools that adapt themselves to the user needs and they provide the user with a richer interface that saves time in all aspects of his personal information management tasks. Current tools don't take into account the fact that it is the same person both storing and retrieving the information but rely instead on principles for general information management, structured in a way so anyone should be able to use it. The second conclusion is that the tools in the laboratories will soon be seen in common operating system and applications.

This report is written in English.

Keywords: personal information management, information pollution

Personal information management now and in the future

What is the difference?

Tim Gahnström

Department of Informatics, Göteborg University

IT University of Göteborg

Göteborg University and Chalmers University of Technology

SUMMERING

Beslutsfattar av idag har ofta inte det tidigare vanliga problemet med brist på information. I stället har dom ett överflöd av information som strömmar genom deras organisationer och datorer men verktygen för at hantera informationen är för dåliga. Mycket dyrbar tid spenderas i onödan på att organisera och söka efter information. Målet med den här undersökningen var att jämföra de verktyg som används praktiskt för att hantera det personliga informationsflödet med de verktyg som finns i forskningslaboratorierna. Detta gjordes genom en fallstudie på UNOSAT. Den första slutsatsen av underökningen var att det som forskarna tittar på idag är mångt bättre än de verktyg som används praktiskt. Forskarna har verktyg som anpassar sig till användaren och som ger användaren fler möjligheter samtidigt som de hjälper till att spara tid. De verktyg som används praktiskt använder sig inte av det faktum att det är samma person som både arkiverar och söker efter informationen. I stället baseras de på samma principer som system för generell informations hantering där vem som helst skall kunna hitta informationen. Den andra slutsatsen är att verktygen som idag endast finns i laboratorier snart kommer att finnas även i vanliga operativsystem och program.

Rapporten är skriven på engelska.

Nyckelord: personal information management, information pollution

1 Introduction 8
1.1 About UNOSAT 8
1.2 Research question 8
2 Method 9
2.1 Empirical data collection 10
2.2 Information gathering and comparison 12
3 Theoretical framework 13
3.1 Information pollution 13
3.1.1 Information pollution as distractions 13
3.1.2 Information Pollution as misinformation 14
3.1.3 Information pollution as information overload 14
3.2 Information management systems 15
3.2.1 How people organize their desktops 16
3.2.2 The user subjective approach 17
3.2.3 Raton Laveur 19
3.2.4 Stuff I've seen 20
4 Empirical data 23
4.1 Current situation 23
4.1.1 Presentation session 23
4.1.2 Problem and feature session 24
4.1.3 Ethnographic study 26
4.1.4 Summary of findings 27
4.2 Client solution 28
5 Analysis 31
6 Discussion 34
7 Conclusions 35
8 References 36

Preface

This paper is a master thesis in Business technology, written for the IT-University of Gothenburg, a part of the University of Gothenburg. The thesis was written at the United Nations organization for Satellite imagery, UNOSAT.

I would like to thank my supervisors for their support; at UNOSAT Alain Retiere, Olivier Senegas and Einar Björgo and for the IT-University Hans Björnsson who supervised the project from his position at Stanford University.

Many thanks to Francois Grey at CERN OpenLab for initiating the contacts, helping us out with Office space, computers and all the details and infrastructure needed to make this possible. Without his support I would never have reached Switzerland in the first place.

1 Introduction

Technology is deeply changing human work; increasing reliance on computers, advanced communication networks and distributed working environments where people collaborate over vast distances. This is true also in the world of humanitarian relief work. Despite the low tech work on the actual sites the organizations and people are knee deep in the interconnected world. There is a big need for communication, collaboration, and large scale problem solving.

A lot of the day to day work for upper management in UN agencies is carried out with computers. For several reasons the applications used for this work are not always chosen on solid grounds. Often the tools used are chosen from what happened to be readily provided by the organization. The client joined this project because he wanted a complete review of the tools he is currently using and a new computer setup with better tools. I joined the project because I wanted to se how well the current programs used in day to day work compares to what the modern researchers say the programs should look like. The kind of programs I have studied are collectively referred to as personal information management systems, this includes the operating system, communication programs and other office tools. As we will see in chapter five personal information management systems ate essentially there to fight the matter of information pollution.

The purpose of this study is to examine personal information management systems. The examination is done from both an operational and a research perspective. The operational perspective is provided by UNOSAT, here I study programs actually used in real organizations. The research perspective is gained from literature studies of the most up to date research on the subject.

1.1 About UNOSAT

UNOSAT is a UN organization working with satellite imagery for humanitarian organizations. Much time and money can be saved for the humanitarian community if they are provided with good supportive material in the form of satellite images and other earth observation data. The manager of this organization states that he is overwhelmed with information of different kinds and wants to have a better system for managing this. Chapter 5 and chapter 5.1 in particular explains the situation and the kinds of information in detail.

1.2 Research question

The purpose of this study is to examine personal information management systems (PIMS) from both an operational and a research perspective. Three issues in particular have been singled out as interesting.

Characterize the gap the gap between operational PIMS and the PIMS devised by current researchers?

What are the distinguishing characteristics of PIMS from the research community?

What are the distinguishing characteristics of the PIMS solutions at UNOSAT?

2 Method

This chapter explains the disposition of this thesis and the methods used to create it, it also explains the strengths and weaknesses of the chosen methods.

The main method I have chosen to use in my work is the case study. According to the definition by Merriam (1988) a case study is carried out within a small framework. According to Yin (1994) a specific phenomena should be studied within the framework. The framework I have chosen is UNOSAT and specifically the manager of UNOSAT (referred to as the client from now). The phenomena studied in this framework are the personal information management systems and the design of them. There are no defined methods that need to be used for collecting and study data in a case study, according to Merriam (1988). In this chapter I will instead discuss the methods I have chosen for this and their pros and cons.

There were some constraints on the project set by the client. He had a problem and wanted a useful and a practical solution to that problem, not an academic study of it. Unfortunately, solving this problem is not enough to qualify as a master thesis but together we found a good solution where we both got something interesting out of the cooperation. I was interested in comparing current tools with the tools proposed by the research community and the client wanted a review and an improvement of his tools. We decided that the way to proceed was that I should set up a good computer with modern tools for him and then compare this specific computer to what is suggested by the researchers. From this I try to draw some conclusions about how a computer like this will look in the future.

This chapter is divided into two parts, the first part explains how I defined the theoretical framework and how I compared the computer to this framework. The second part explains how I conducted the empirical pars of the thesis, how I found the needs of the client and what his demands were on the new computer.

This study is based on qualitative data collection and an inductive research approach. Whether to use an inductive or deductive approach as a way to gain knowledge has been discussed since ancient times. Induction means, according to Molander (1988), that you look at one or more specific cases and from there you try to formulate a general rule or assertion. Deduction is the opposite of induction, instead of looking at phenomena and make general rules, the deductive approach is to take a set of predefined truths, axioms, and combine them and build up a proof. In this thesis I have looked at the visions of four research groups and one computer set up with standard tools. Then I compared these five entities and drew conclusions about what improvements are likely to be seen in the computer in the near future. There are no doubt problems related to using this method and the results can not and should not be regarded as absolute truths. Instead this thesis is trying to create a better understanding of the problem and possible solutions to it.

The first and most obvious problem is that I have only used a small subset of the available research for my comparison, it was not possible to look into it all given the timeframe I had. Instead I have chosen four out of maybe a hundred different researchers. To get around this problem a little bit I carefully chose four studies that were carried out by somewhat larger groups and that were backed by institutions with an interest in the

subject and a track record of research in this area. I have also tried to verify the papers by carefully reading their backgrounds and study where they have gotten their ideas and foundations from.

The other immediate problem is that the computer I am comparing the research to might not be representative of computers generally used. It may be that I have not found the best solution and it might be that UNOSAT and the client doesn't have representative needs. I have taken measures to choose tools that are proven and that are known to be popular. I have myself been a heavy user of computers and office tools for the last decade and believe myself to have a good idea of what is available and the client, who is not a tech savvy person, easily found his way around the new computer even though it has a completely new operating system so I think this computer is similar enough to most other computers used in management environments to make the comparison meaningful.

The last of the interesting issues I am going to discuss here is that the research I have conducted have many ingredients of the classic action research as defined by Hanson (2003) without being a full fledged action research study. Action research means that the researcher not only study a phenomena but take an active part and does his best to improve the situation and thereafter measure the effects and improvements. The founding father of action research Kurt Levin once said, "No action without research and no research without action". The way I have conducted my study is most certainly through actions, I have actively improved the situation I have studied while I studied it but it is still not pure action research because I have not used my improvements as a basis for my conclusions. I have not compared the old solution to the new solution as devised by Hanson (2003), instead I have used the new and improved solution and compared it to other solutions, namely the once available in research laboratories. The old solution has not affected the research more than that I have used it as a knowledgebase for building the new solution.

2.1 Empirical data collection

The way we set this up was that I got rather free hands to create and improve the client's computing environment. I bought him a new computer and set it up with all the needed tools. Then I moved all the data from his old computer over to the new one and gave him a completely new and revised but more or less immediately productive computer.

I used a three step process to define how the new computer should look. All the sessions I had with the client took place in his office without any distractions or other people there to influence the answers. This was to make sure the client always felt secure and free to speak his mind which is very important since the questions some times touched on sensitive subjects.

First I got an overview of the organization together with the client. He made a standard presentation of the organization for me and then we had a long discussion about what he does and how the organization is positioned to its customers, investors and within the UN system. This was done to give me a starting point for my work. When I came to the organization the problem was still very vague and working on a solution at that point was not even thinkable. The chapter called "Presentation session" describes the result of this part. It is intended as a read for those who want to get a thorough understanding of the of

the client's problem. The client is assumed to be reasonably representative for a manager of a small organization within a large public organization; with representative I mean that he is using a similar set of tools as most others. After the initial introduction I read up on all written documentation related to the organization and wrote it up in a coherent way for this report and for future referencing.

The next step was to have a brainstorming session regarding the specific problem and possible solutions. This was a very open session where no ideas were discarded and where the talk flew fast in different directions. This kind of interview is called a free or open interview, it is more like an informal conversation than the common way where one person is clearly the interviewer and the other part is clearly the interviewee. This is the interview form where the least pressure is exercised on the interviewee. The form is also a form that encourages the interviewee to present his own thoughts and ideas which was the point of this session. (Lundahl and Skärvad, 1999)

It is just as important to come well prepared to an open interview as it is to a more formal and structured interview. To make sure that the brainstorming never came to a standstill I brought a number of ideas and questions to the meeting which I mentioned whenever I felt that it was necessary to keep the ideas flowing. Thanks to this and the fact that the client had a lot of ideas and thoughts of his own about the system, this session gave me a lot of useful material. I came to the session armed with a pen and a notebook and I took small notes during the whole event, these notes were rewritten immediately after to make sure I didn't miss or forget anything.

Once I had sorted out all the material it was clear that what the client was interested in was personal information management (PIM) systems. The final decision was then made to compare current PIM tools to the most up to date research about PIM systems.

After I had studied the subject of PIM systems thoroughly (read more about this in the next chapter "Information gathering and comparison"), I had a last session with the client. This session was a full day when I conducted an ethnographic study of him. I was with him the whole day in his office. I sat at a reasonable distance where I got a good overview of him and his monitor. I was not close enough to read the text on the monitor or to distract him in his work. This study was also carried out with a pen and papers as main tools for recording, they were chosen because of their ubiquity and discrete nature. I looked at the monitor most of the time but took notes all the time. As soon as I got a thought on my mind I took note of it so I shouldn't forget it, they related not only to what tools were used or needed but everything I came to think of, from the way the computer was used to how breaks were incorporated into the day.

Whether to use quantitative or qualitative method is always a choice that must be made on solid grounds. According to Lundahl and Skärvad (1999) the aim of a qualitative method is to analyze and understand a process or phenomena. All of the above depicted methods I used were qualitative. I tried to understand not only what the client wanted and thought he needed but also if he might actually have needed something that he wasn't aware of. In a qualitative study the basic tool is the interpretation of the data Lundahl, Skärvad (1999).

Quantitative research is generally done to statistically describe a phenomenon. Often the researcher gives a questionnaire to a large group of people or measures data from a

number of events. The data is then collected and put together and the researcher draws his conclusions from that.

Both methods have distinct advantages, a qualitative approach may give a deeper understanding of the phenomenon while the quantitative approach can be much more unbiased, often the data from a qualitative approach is unquestionable and the researcher is not so easily lead in the wrong direction by accident.

I could have used a very quantitative approach as a supplement to my qualitative studies but there was no time. The way I would have set it up would have been by installing a program on the client computer that recorded everything he did with it over a certain time. If I had done that, I would have had very solid information on how he used his computer. The reason I chose not to do it this way was because it would have taken to much time and the focus of this study was not to design the perfect computer but rather to compare a general computer to the research frontier.

2.2 Information gathering and comparison

Between the two last sessions I had with the client I made a thorough study of how prominent researchers devise that PIM tools should look in the future. The first thing I did was to sign up to a mailing list for PhD. students currently working with PIMS. There I read the archives and asked a few questions for starting points on the subject. The participants on the mailing list were most helpful and pointed me to a number of relevant studies conducted by various research groups in resent years about how to design PIM systems. I selected four of the larger studies conducted in the last few years and analyzed them more thoroughly. The aggregation of my studies of these papers can be read in the next chapter, the Theoretical framework. To get a good understanding of the points made in the research papers I also studied something called information pollution which is in essence the reason why we need tools for PIM.

In the analysis chapter I have mad an extensive comparison of the system I designed for the client and the solutions suggested by the researchers. This comparison is a straightforward item for item comparison later used to draw conclusions about the size of the gap between the two and what we can expect to see in the future.

3 Theoretical framework

The scientific framework to which I want to compare the standard tools (represented by the client solution in the following chapter) touch on two related subjects, the problem domain is in the area of information pollution and the solution domain revolves around information management systems in general and personal information management systems in particular. I will discus both parts in this chapter but the focus is on personal information management. More information on the specifics of the client problem and solution can be found in chapter five.

3.1 Information pollution

Information pollution works like a common denominator, holding the problem domain of this thesis together. As we will se in the ethnographic study, the client suffers from information pollution. He has learned to handle it with the cumbersome tools available today, but hopes for a better personal information management system in the future.

There is not yet a sole and widely accepted definition of information pollution, the topic was discussed already in the sixties, Kenneth Boulding (1966) wrote, "I am not going to try to define information pollution exactly, as it is one of these concepts that perhaps is most useful when it is rather vague". A number of renowned people have since written about information pollution in various ways. This chapter attempts to sort out these views and make the reader aware of what information pollution is and what kind of problems it creates.

Jakob Nielsen popularized information pollution to a broader public when he introduced it as a concept for too much information and various distractions at the Nielson Norman Group User Experience Conference 2003, there he said to BBC Online, "Information pollution is information overload taken to the extreme; it's where it stops being a burden and becomes an impediment to your ability to get your work done" (Twist, 2003).

3.1.1 Information pollution as distractions

In the fall of 2003 the user interface expert Dr. Jakob Nielsen proposed his view of information pollution. Besides giving a lengthy interview to BBC News Online (Twist, 2003), he has also used two Alert boxes to promote the ideas (Nielsen, 2003) and (Nielsen, 2004). The ideas put out on these occasions are not peer reviewed research but Jakob Nielsen is such a renowned person in this field that his thoughts cannot be neglected in a study of information pollution. His articles and alert boxes are well referenced over the internet and in journals and he is the author of numerous books and regarded articles.

Dr. Nielsen's views are written in contrast to the praise of instant messaging you often read in the popular press today. In Metz (2003) we can read "Instant messaging is quickly becoming an essential part of PCs, and if you haven't joined the IM party, it's high time you did" a few paragraphs further down in the same article you can read "If I were to poll ten CIO's and ask if instant messaging was being used strategically for

business communication within their organizations, nine out of ten would probably say no, and eight out of nine would probably be wrong, says Michael Gartenberg of research firm Jupiter media."

In (Twist, 2003) Nielsen emphasizes that it is not that all the tools and information is inherently bad, it is the accumulative effect of them that is a concern, "We don't really mind one polluting factory in the world, but we mind millions," he explains.

Nielsen defines information pollution as all the small distractions and all the unnecessary information you get every day. Some of the examples he uses are instant messaging and email. The problem with these programs, he says, is not that it takes so much time to answer the messages you get. What is more important is that you get distracted from your regular work. If you are a knowledge worker it can take a long time for your brain to get back to what you where doing, where you were and what line of thought you were in. "A one-minute interruption of your colleagues will cost them ten minutes of productivity as they reestablish their mental context and get back into "flow." Only the most important messages are worth 1,000 percent in overhead costs." (Nielsen, 2004).

An important countermeasure for this kind of information pollution, according to Nielsen is to take back control, do things when you decide you have time. Not when someone else or a machine decides. A first step is to turn of all alerts. You should choose when you want to read and respond to emails etc not the little blinking envelope in the bottom right corner of your screen. It is better if you work efficiently for a chunk of time and then do your communications instead of doing it in between your work every time someone comes around to ask you a question. (Nielsen, 2004; Twist, 2003)

3.1.2 Information Pollution as misinformation

Information pollution is a vague concept but it has at least two characteristics that can be defined. The first characteristic is when the information system produces images of the world which are unrealistic in the sense that they do not correspond to some external reality. This can be when people tell lies or when they make errors that get perpetuated. The other characteristic of information pollution is even more negative, it can be described as ignorance. When we have access to information that would help us, but we are ignorant of the fact that we have the information. (Boulding, 1966)

Cameron and Kuen-Hee (2000) describes a third kind of information pollution as misinformation. They describe Advertorials as information pollution. Advertorials are paid advertisements in magazines and newspapers that are made to resemble editorial material, this is very common and they are often referred to as advertorials. (Waltzer, Waltzer, 2001; Palser 2002).

This is deliberate misinformation from the advertiser, they try to deceive the reader to think that their, product, service or opinion is objectively written about and concluded to be good (Cameron, 1994).

3.1.3 Information pollution as information overload

The decision makers of all organizations rely on the information they receive when they make their decisions. Orman (1984) explains that most information is polluted to some extent. Information pollution is the contamination of information with incomplete,

inconsistent or irrelevant information. Although it affects the society as a whole, its most damaging effect is on the professional decision makers whose performance depends on the quality of information they receive. The same source explains that sometimes the information is more polluted and sometimes less but most often the information demands some kind of processing to be useful. The processing of information may involve discarding irrelevant information, analyzing to insure consistency or the triggering of further information gathering. As the costs of this processing approach the expected benefits, the users are forced to discard the information without adequate consideration. This situation is commonly referred to as information overload (Chervany, Dickson, 1974).

A number of researchers have observed that a major complaint of managers has drastically changed from lack of information to too much information (Orman, 1984). A popular book on this subject is written by David Schenk (1997), it is called Data Smog. While he was a writer for Wired Magazine in New York he got to experience information pollution first hand in his daily work. He shows that unless we find ways and tools to handle the enormous amounts of information that flows over us each day we might drown in it and not be able to use any of it (McKenzie, 1998).

3.2 Information management systems

A good place to start a study on information management should be at the roots, on the definition of information. Summers and Oppenheim (1999) references Yuexiao (1988) who estimates that there are several hundred definitions of the word information. They differ not only between and within, disciplines, but also as a function of the time when they were defined. Summers and Oppenheim (1999) continues with a short history and explains that Shannon and Weaver (1949) with a background in electrical engineering defined information as the actual bits in a data stream while they tried to figure out what was redundant and what was actually real information. In 1955 Farradane coined the term information science. This was to distinguish information workers of the time from the mere librarians. The information scientists were characterized by that they worked proactively while searching the literature and evaluated their sources of information and discarded the lesser ones. This goes on to today's broad definition of information where it can be text as well as images and numerous other forms of communication. The conclusion of this argument in (Summers, Oppenheim, 1999) is that there is no reason to put boarders or limits around the definition of information because they are bound to change and broaden as time passes.

Instead of trying to objectively define information from its content a more vague definition is proposed by Dumais et al. (2003) Personal information management (PIM) is the practice of managing information that helps us in our daily lives such as addresses, phone numbers, to-dos, appointments, notes, documents, folders, web pages and emails.

The management of personal information is an important part of an individual's learning. People continually collect information from a variety of sources and store them outside their cognitive system, the cognitive system is the brain and the memory. Outside the cognitive system might be in a file cabinet, a pile on the desk or, as in the case of digital information, it may also be a folder or a bookmark file. Usually an item is briefly reviewed by the person's cognitive system, and then cataloged, tagged, and put aside for

possible retrieval in the future. As the information to which a person is exposed expands, it becomes necessary to store more and more information, as does the need for effective mechanisms for organizing, retrieving, and using this information. (Bergman et al, 2003)

Common information retrieval tools, like popular web and intranet search engines (often called general information management (GIM) systems) are designed to facilitate information discovery. Given a short query they can find relevant materials using a variety of cues, such as content, anchor texts, and popularity. However, much knowledge work involves integrating and re-using information that has previously been created or accessed. For example, writing a presentation or paper may involve some web searching, but it also involves pulling together information from existing information sources like documents, spreadsheets, data analyses, email messages, etc. Dumais et al (2003) referenced a number of studies that have shown that about two thirds of web pages accessed were re-visits to pages previously seen. They also referenced articles that showed that similar re-access patterns have been observed in usage of library book borrowing, and human memory. (Dumais et al 2003)

The design of current computerized PIM systems such as PC operating systems and the surrounding programs unfortunately often rely on the same principles as those underlying the aforementioned general information management systems (Other good examples of GIM systems are Libraries and directories). PIM systems designed this way does not take into account the fact that a PIMS is organized and thereafter used by only one person: The same person who screens, classifies and stores the information, is the one who eventually retrieves it again (Bergman et al. 2003). Dumais et al (2003) even go so far as to say that today it is often easier to find information on the web than on your own computer. This is due to both the many different applications used to manage personal information, each with its own organizational hierarchy (e.g., email, files, web, calendar), and to the limited search capabilities in many of them. The approaches suggested by Bergman et al., (2003), Dumais, et al (2003), Barreau (1995) and Bellotti and Smith (2000) are manifold but an important part of them al is that context, value adding attributes and available metadata from the users and the environment, should relate to the data items These attributes are often temporal and personal in nature so they may make no sense to an outside observer but they still give good help to the user. Kwasnik (1991) found that only 30% of the attributes the documents were organized according to were document related. The rest were attributes related to the interaction between the user and the information (e.g., situation attributes, disposition, time, cognitive state). All of the aforementioned research projects, (Bergman et al., (2003), Dumais, et al (2003), Barreau (1995) and Bellotti and Smith (2000)) did research on their own but rely also heavily on what is done by Malone. No study of PIM would be complete without the foundation he lay in his study of how people organize their offices. A short summary of his results follow in the next chapter before the explicit results and guidelines from three other large studies are presented.

3.2.1 How people organize their desktops

Malone (1983) explored how people organize things in the context of their offices. He studied people in different jobs and analyzed the patterns in their organizing behavior. He was trying to find out the implications for how you should design something he called

office information system. He found that most people tended to organize their documents into “files” and “piles”. Files are well-organized, often labeled stacks or folders whereas piles contain miscellaneous documents that have no apparent organization or labeling. In the context of files and piles he made two interesting claims, they are equally important there to serve as reminders. The location and size help the user remember things related to the pile and the top documents also helps the user to remember its content etc. The other interesting claim was that one important reason for why people don’t file and classify their documents as well as one would expect is because it creates a heavy cognitive load on the brain. It is hard to decide what categories should be available and even when that is done it is hard to decide in which category a document should go, often, Malone noticed, a document should be in two different files at the same time. He concluded that automated systems could resolve many of the problems in these workspaces by supporting multidimensional classification, semi automated classification, piles as well as files and by letting these piles, files and even just documents work also as reminders, maybe by varying the size or color of icons based upon the importance of the document or having the item appear on the screen periodically.

I read four studies in depth that either have created experimental PIMS and or given solid design principles that should be followed in the design of such a system. The next chapter will go through these systems one at the time and discuss them or their design principles.

3.2.2 The user subjective approach

Bergman et al (2003) found three principles that should be used as guidelines in the design of a PIM system; the Subjective Classification Principle, all information items related to the same subjective topic should be classified together regardless of their technological format; the Subjective Importance Principle, the subjective importance of information should determine its degree of visual salience and accessibility; and the Subjective Context Principle, information should be retrieved and viewed by the user in the same context as in which it was previously used.

They claim that because these principles are only sporadically followed in the design of current system, be they computerized or physical, those systems often fail to decrease the load on a user’s cognitive system. This is why people behave the way Malone (1983) described. Bergman et al. (2003) referenced several sources that showed other examples of people not using the information procedures available in the PIM system, but instead rely on alternative strategies; they pile up papers instead of filing them and they keep hundreds of e-mails in their inbox instead of organizing them in folders. They even email links to themselves instead of using the bookmark feature or history feature of their web browser.

The three principles they suggest are described a little bit more in detail bellow.

The subjective classification principle suggests that all information items related to the same topic should be classified under the same category regardless of their technological format. A topic is a subjective value that is added to the information item by the user when storing the information. This means that in the design of a PIM system a user-driven model should be used. Not a common user driven model where the system is tailored to a specific group of user, instead it should be customizable by the individual

using it to enable easy classification and retrieval of the information items. In (Bergman, 1983) the use a common personal operating systems as an example of a PIM system that is more technology driven, instead of classifying information items according to topics they are first sorted according to technological format and only thereafter by topic. A given example are web links, they are not stored anywhere in relation to anything else, if you want to find a webpage you have to look in a specific folder or a specific file and thereafter you can find it according to subject (if you have organized your favorites). It would have been better if you only had to look at the folder where you store everything relating to a given subject to find the web links relating to that subject.

The subjective importance principle suggests that information items should be characterized by their importance and that this attribute should determine their visual salience and accessibility. This is because when a person is exposed to new information the first thing he determines is usually how important that information is. Important items should be easily accessible and noticeable, while irrelevant and unimportant information items should not distract the user. The importance of an information item is determined by the user relative to the importance of other information items. Subjective importance does not rest in the information itself, what is priceless to one person can be worthless to another.

Even within an information item, the user often needs to specify which sections are more important and which are less important because rapid and effortless accessibility to the important sections may be desired. This can be achieved if the user has the opportunity to highlight or otherwise mark the important parts.

The subjective context principle suggests that information should be retrieved and viewed by the user in the same context in which it was previously used. Research has shown that information is better recalled when it is stored in the context in which it was learned. Contextual characteristics are divided into three categories external, internal, and temporal.

The external context of an information item refers to the other items that the user dealt with while interacting with a specific information item. When viewing an information item in the same external context as the last time less effort is needed to reconstruct the mental processing involved in its creation, and the user will suffer from less memory load, because other relevant information items will be accessible from the item's working environment.

Internal context relates to the user's thoughts while interacting with the information item. In most encounters with an information item, there is some cognitive processing on the part of the user; the item triggers thoughts relevant to the item, responses concerning its relevance, significance, reliability, its association to other items, questions, etc. All these constitute the internal context of an information item and contribute to the construction of new or revised information. To be able to give the user its internal context back, the PIM system should allow the user to write annotations about what they read. These annotation should then automatically provide easy access to documents and be presented together with the original information

Temporal context relates to the state in which the user left the information item when he last interacted with it, and to his working plans regarding that information. In addition to

the internal and external context the user should be able to trace their previous steps and also to plan future steps of work without leaving what they are doing at the moment. There are already a number of aids for the temporal context aids in current PIM systems such as subjects of e-mail messages are in bold font if an e-mail has not been opened and in normal font if it was screened. Links are blue when they haven't been accessed yet, and turn purple when activated. An example of a temporal context aid that is often missing in current systems are something that allows the user to mark the information items that they plan to work with in the future, such as links within a text that will be accessed only after completing the reading of the text. Such marking will enable users to read a text that contains links without interruption.

3.2.3 Raton Laveur

Raton Laveur is a research project within the famed Xerox PARC. Numerous articles have been produced regarding Raton Laveur, in this study I rely on the work done by Belotti and Smith (2000). The initial goals of the project was to create a paper, scanning and printing based PIMS but it turned out that "The feedback we received from our paper prototypes was luke-warm at best" (Belotti and Smith, 2000). According to the article itself their initial goals were not at all on the same track as those of their interviewees. First of all, the prerequisites in terms of needed external equipment were just not available and to that it appeared to be to much overhead work with the first approach anyway. Instead, in the third iteration they came up with a completely computer based program for PIM. The program is called Raton Laveur and during their work with it, they came up with a number of interesting findings. They have summed up the most important of these findings in four design guidelines for a PIM system.

Embed PIM in an application that supports ongoing work: People dislike switching to a different application purely for the purpose of information management. "PIM functionality should be a part of the experience of the active online workspace" (Belotti and Smith 2003). In other words, the user should be able to work with and organize the information without leaving his normal working context. This requirement is based on the observation that people prefer to use an easy-to-access open application such as email to handle PIM.

Flexibility: Users must be able to customize the way the system looks and the way the system works. This requirement was based on the great variety and adaptation of PIM solutions that were present among the interviewees.

Lightweight: PIM-style information (e.g., project name, to-do, due-date) should be as easy to attach to any element as it is to place a sticky note on anything.

Simplicity: People dislike complex PIM tools. A successful solution must be easy to learn. This is based on the observation that most people just bothered to learn the most basic features of the tools they use. Even MS Outlook was repeatedly described as too heavyweight by the interviewees.

Raton Laveur is designed with these guidelines in mind. To keep it simple and lightweight the design team adopted a policy of adding no additional features unless experience-in-use dictated that the system was frustrating without them. It is designed around an email-client with both savable searches and groups as means of creating

collections in addition to normal folders. The groups are distinctly different from folders; they are a resource more like the piles observed by Malone (1983). They have a representative member, like the top document on a pile allocated by the user, which will always be displayed when any group member matches search constraints. The other members are then accessed through that document, the way you look through a pile on your desk. The other distinguishing feature of the program is the search interface to the documents. All documents are archived in the computer and indexed by available attributes. Emails for instance are searchable by, content, sender, recipient, subject and date. The searches are in no way constrained by documents being in groups or folders and the documents are not constrained by file type. The interface is centered around an e-mail client because this program is often an always on program with a good interface for displaying documents. To this they have added extensive support for other document types and extensive search and indexing capabilities.

The other important day to day task that is well implemented in Raton Laveur is the reminder, also mentioned as fundamental in (Malone, 1983). One of the most important types of remainder is the notes placed on documents and the location of documents. Documents and piles are put at certain places where they will remind the users of things and give him easy access to them when he expects to need them the next time. It is also common to scribble notes on post-it notes and attach them to documents or to the monitor or even the door if it is reminder to bring something when you leave. All these features are, according to Belotti and Smith (2000) incorporated in an intuitive way in Raton Laveur. During their testing phase they noticed that people frequently used their inbox to keep remainders and to-do-lists. They simply marked an email as unread until they had dealt with the issue and sometimes they sent an email to themselves to remind themselves of something they should do in the future. In Raton Laveur any item can be made a to-do item with a simple clicking sequence, you can also enter reminders or deadlines on these items. When you filter the items according to remainders you will also have easy access to all documents relating to the particular remainder. To further resemble the post it notes you can ad comments to any document anywhere irregardless of the format. The enhancement compared to ordinary sticky notes is of course that you can easily search and filter the notes.

3.2.4 Stuff I've seen

Stuff I've seen is a research project from Microsoft. They have designed, deployed and evaluated a system that provides simple unified access to all the information a person has seen. What makes the report (Dumais et al, 2003) on Stuff I've Seen so interesting is the extensive evaluation of the system in a production environment. Stuff I've seen is as the name gives a way a PIM system that makes it easy for people to find information they have already seen at some point before. Two key aspects of the design support this, the system provides a unified index of information that a person has viewed on their computer, whether the information was an email, web page, document, media file, calendar appointment, etc. Today, people have to manage several different organizations of information – e.g., the file system hierarchy for files, the email folder hierarchy for email, favorites or history for web pages. With Stuff I've seen, all of these sources are integrated into one single index regardless of what form the information originated. Second, because a person has seen the information before, rich contextual cues such as

time, author, thumbnails and previews can be used to search for and present information. By providing a unified index across all these different information sources Stuff I've seen solves the problem of having to look in different places and to use different applications to find the information they are looking for. If a user wants to restrict search to a particular source they can, but this is not a prerequisite for finding information.

The user interface allows users to specify queries and to view and manipulate results in an intuitive way, because it works from a local index, query results can be returned very quickly, allowing an interactive and iterative query strategy. Many other common search interfaces only allow the user to specify his query properties and then hit a search button to launch the query, Stuff I've see instead launches its queries whenever any of the filtering widgets in the user interface are manipulated or when the user presses return. This allows a user to start broadly and quickly refine their query by interactively filtering and sorting the results. According to (Dumais et al., 2003) these search interface ideas relate to work done by both Belkin et al. 2001 in "Iterative exploration, design and evaluation support for query reformulation in interactive information retrieval" and to work done by Ahlberg, Williamsson and Schneiderman (1992) in "Dynamic queries for information exploration: An implementation and evaluation".

The evaluation research that was done after the deployment was large, it covered 234 persons and it used both qualitative and quantitative methods while analyzing the actual use of the system after deployment. A short summary of the most interesting findings follows bellow.

The users issued on average 4.4 queries per day but it varied a lot between users and days, on average the users used the system on 84% of the days. Users did not use advanced features much, similar to the statistics we know from web search engines, users preferred simple searches, only in 7.5% of the searches Boolean operators were used and the searches averaged 1.6 words at the time. This is apparently sufficient because there is a relatively small amount of data and the well know domain wherein the search takes place. It was very common (48% of the times) to apply one or more of the predefined filters to the search, for instance search only emails or web pages. 25% of the searches included a persons name suggesting that people are a powerful memory cue for personal content. It is easy t o remember who sent something or who it was written about etc.

Only about one third of the searches actually resulted in the opening of a file, the article does not claim to know why this is the case but they suggest that it may not be the naïve reason, that the search was a fail, which is the correct one. It may actually, they suggest be the case that the preview of the document were good enough or that the metadata was all that was needed.

Another interesting finding was the long time span that data was accessed in, the most commonly opened items were of course recent items but there is a very long tail in the distribution, items dating back to 8 years old were accessed through the system, it was not so frequent but items ranging from 2 to 8 years old were accessed with about the same frequency distribution.

The study also shed some light on user interface choices. Users could choose between having the input fields in the top row like the web search engines or have it in a frame to the left like in the windows operating system. It turned out that the top-design were most

popular. Another question they looked into were whether to sort the results according to rank or according to date, date were the winner here which is exactly the opposite to what have been seen to work with web search engines.

They also made questionnaires to the users where they asked general questions about tools that help them to keep found things found. On the basic question “Stuff I’ve seen-like search service should be an essential functionality in any computer.” They got an overwhelming 4.5 on the Likert scale were 5 means “Strongly agree”. The other questions also asserted the opinion that people in general liked and had use for the improved PIM system.

4 Empirical data

The client reason for supporting this research was to find a good solution to his problem with an ever increasing amount of information flowing over him. I used this as a way gather empirical data for the analysis. The chapter is divided into two main parts. The first part explains the efforts that went into defining the problem and the needs of the client. The second part describes the actual solution that was delivered to the client in the end.

4.1 Current situation

When I came in contact with this project it was still very unclear what the actual problems were, it was all vague and incoherent thoughts. To clear the fog and get something solid I had three different sessions with the client. The first session was just an informal presentation on what kind of organization Unosat is and how it operates. The second meeting was a brainstorming session to define the problems and the desired features of the final system. The last session was a full day's ethnographic study where I followed the client and took notes on his computer usage. The following three chapters will depict each of these sessions.

4.1.1 Presentation session

Unosat is a small organization within the UN working to promote the use of satellite imagery and other earth observation data in the international humanitarian community. They have access to and work together with all commercial data providers and the best value adding companies in the business.

This chapter is a short orientation about why the manager of a relative small organization like Unosat has such high amounts of information flowing over him. It is a deliberately short chapter since it is quite far away from the actual research issues.

4.1.1.1 Key competence

One of the key factors creating complexity for the client is that he is the person with most of the know-how about the UN politics and the needs of the humanitarian community. He is the only person in the organization that can do everything and often he is the only one with the needed set of skills for a given task.

The director has an assistant that helps out and two service managers that can do parts of the work but it is not enough. This problem is always in the process of being solved by constant schooling of the staff members. Within the next six month period another experienced person with a similar set of skills is planed to be hired to ease the workload.

4.1.1.2 Open standard organizations

UNOSAT is involved in, and sometimes a leading part of, many organizations that work to promote and / or standardize earth observation related issues. An Example of these organizations is the Open GIS Consortium where UNOSAT has taken an active part in its work to promote interoperability between different data providers and different tools that are commonly in use. Unosat is also working closely with the European Commission initiated Global Monitoring for Stability and Security (GMOSS) Network of Excellence

to develop methods and services using GIS and EO data that can be of benefit to the United Nations and its partners.

4.1.1.3 Funding organizations

UNOSAT is currently run as a project under the supervision of UNOPS, United Nations Organization for Project Service. This has a number of advantages but it also comes with a cost. The cost is mainly in terms of administrative work, the need too carry out the work in conformity with the rules and regulations set up by UN and UNOPS.

UNOSAT has two offices. One single room is located in the UNOPS building and the rest is located inside CERN, who is providing all the necessary computing and housing infrastructure.

The European Space Agency provides most of the in-cash funding, they see UNOSAT as their way into the humanitarian market.

UNHCR is the formal owner of the UNOSAT project at the moment. They are the once who formally requested the need of a service like UNOSAT from UNOPS.

Each one of these collaborations is individual and put their demands on the organization to do certain things or operate in certain ways.

4.1.1.4 Customers

The UNOSAT customers are a diverse group. Often they have no, or bad means of, communication. They might be in desolate or war torn locations. They always run on a tight budget and need the images now or preferably yesterday.

Even if they don't need the images in a hurry they don't have much time to spend on acquiring them, they want to have a one stop shop to which they can turn for everything. It is not possible for a person in these situations (some times the only one who speak the appropriate language) to spend a lot of time communicating their requirements and study complex things such as satellite imagery.

Another problem with the potential customers is that satellite imagery is reasonably new in the community. So the customers need to be informed about the existence, the possibilities and the use of satellite images. This must be done in a preemptive manner, once the catastrophe is at hand there is no time to try out new exciting technologies.

4.1.1.5 Employees

There are also some issues with the employees, several of them are not directly employed by UNOSAT but are instead provided to UNOSAT via one of the collaborations. This means that they don't only have responsibilities to UNOSAT but also to their respective organization.

4.1.2 Problem and feature session

To turn the first vague ideas into something more concrete we had a brainstorming session about the problems and ideas that the client had about features that might be useful in the system. The session was informal and at a high pace where nothing was discarded. I kept

some very small notes, only words here and there with pen and pencil and made a write up of the results immediately after when everything were still clear in my head.

The client started the session with pointing out that what he was mostly interested in was saving time. He always works long days and overtime is more the rule then the exception. A standard week is often around 60 hours, so the focus of all new efforts should be time savings.

The brainstorming then turned into issues that should be automated, the client brought up the slight annoyances of having to set up folder structures and project structures by hand for each new project. Project in this context is a loosely defined term. It is not necessarily a formal new project for the organization; it can also be the things like involving new people or setting up a new collaboration with another organization. In chapter 5.2.1 we can see that this feature is something that is related to the things Malone (1983) discussed and that was improved and even implemented by Bergman (2003).

Another annoyance the client brought up was that instead of using numerous programs for different tasks. He would like to see something like a control panel where he had access to all the documents and information from one unified interface. The issue here was not only documents but also contact information and relations between entities. If you for instance are looking at a document regarding project A you should have easy access to status data for project A as well as contact data to everyone related to the project. This is the kind of data that should be in this unified interface, but there is more to it. The unified part meant that you should not only have access to the contact data, the contact programs should be integrated in the interface. So if you want to send an email or contact one of the persons on MSN you should not have to look up the address in this program and then go to a second program to make the contact. You should have easy access to the actual communications channels from within the unified program.

On the subject of automated tasks I brought up the issue of backing up the computer. It turned out that at the moment only emails are backed up centrally and on a regular basis. The clients own documents and files are only backed up by him to a USB drive most irregularly whenever he has time. We agreed that an automated back up solution, possibly to the USB drive would be a small feature that might save a lot of problems in the future.

Since Linux is getting increasingly more popular and get more and more press the issue of whether to keep using Windows or switch to Linux eventually came up and I promised to look into it.

The client also mentioned the problems of working in a mobile environment. It would be preferable if the emails could be synchronized between the laptop and the server on any internet connection from any location. As it is now there are only two alternatives, one is to be at office 2 and the other is to use a slow web interface. The preferred way would be to be able to synchronize the server and the computer from any internet connection so that new emails (and other documents) are downloaded from the server and stored on the computer for offline reading and responding. On the next connection to the internet the computer should upload all written emails to the server for further distribution and download all new. It would also be a good thing if the e-mails and other documents were accessible from other internet connected computers. Like a web mail, but for all documents.

On the topic of synchronizing the use of the Palm also came up. The new computer needs to be able to synchronize flawlessly with the Palm pilot. This was mostly brought up as something that might be an issue if we decide to switch to Linux.

Less important but still nice features that also were mentioned were some kind of videoconferencing and document sharing within the organization. Online calendars and some scheduling program were also mentioned as desirable.

One last issue that I brought up was the question of how desirable the features really are, whether to always choose a proven and stable solution that might lack some features or to use the newest most feature packed program that might have some stability issues. The answer of course was more or less obvious. Always go with the stable solution, this will be a working machine that much is dependant on.

4.1.3 Ethnographic study

The ethnographic study was carried out a good two months after the brainstorming. This had two major effects on the outcome. The main reason for the delay was to give me enough time to study the subject on a theoretical and academic level, to learn more about the current research on personal information management (PIM). The side effect of this was that it gave the thoughts from the brainstorming session time to settle down and for new ideas to come forward.

During the day I sat at a reasonable distance from the person and the computer not to disturb or affect the workday more then necessary and took notes of everything that came to my mind. The better part of this particular day was spent working on an official proposal to one of the funding organizations.

Every day is very different; one day can be spent on site at a catastrophe, another at a summit in Rome and a third day can be spent only talking to colleagues and employees managing the day to day work and the organizational structure. So I am very aware that I have not gotten nearly all possible information out of my day. But there was not more time available for the ethnographic study. The study was carried out on a day that was spent mostly working with the computer, since that is the focus of the study. I tried to focus my attention on as general issues as possible. To get a better and broader picture of his work I have had a number of informal lunches with the client. During the lunches I have had minor question marks straightened out and he has tried to give me a general understanding of what he does.

When I came to the office at the set time 9.00 am, the client had just arrived. The first thing this day was to change the internet settings for the computer to the settings of office 2 because the day before had been spent at office 1. When this was done he started his email client, GroupWise, to check the emails that had arrived during the night.

Up until recently, he explained, he had had two e-mail accounts and had to check them at their respective office, but now all emails are forwarded to office 2. This is an improvement but not a perfect solution. Besides this, the e-mails can also be checked via a (5 times slower the client states) web interface from any internet connected computer.

Many e-mails (50% the client states) are only for keeping him updated with what is going on in the organization, they are mostly carbon copied or forwarded emails from or to the employees and these are only read and discarded. The rest either needs to be taken action on or contain important information that needs to be saved. The saved e-mails are moved into a well organized but large hierarchy of folders that he has created himself. The action e-mails

are either left in the inbox as reminders or acted upon immediately. When reading e-mails in GroupWise the e-mails are not previewed in a simple manner, instead each mail is opened and read. When sending e-mails the program automatically retrieves the e-mail addresses when the first letters of the name is written. The client states that about 30% of his time is spent traveling and those days he can only read e-mails through the slow web interface.

Not only the emails are well organized in folders but also other documents are well organized and accessed from windows explorer. Documents received by email are left in GroupWise and stored in the e-mail hierarchy.

The better part of this day was either handling e-mails or working with MS Office. The client is an advanced user of MS Office and uses complex documents with heavy formatting, linked files with Excel components inside word documents and word documents with forms that should be received or filed in. Sometimes they are returned as PDF documents and sometimes as Word documents. It was stated by the client that Word is pretty much the standard document format within the UN. The client does not only use advanced features in MS Office but also seems to have a good sense of how a computer works, what it can be expected to be good at and how it is supposed to be used. While working with a large Excel document and searching it for and fixing errors there were no sign of frustration, it was all done in a straight and analytical way. The familiarity with the computer does not go so far as to touch type or extensive use of keyboard shortcuts, everything except typing is handled with the mouse.

Much working time is spent, the client explains, either in coaching the employees or administrative work that needs to be done in a public organization like UN, this could be applying for funding or for making progress reports.

A quick break is taken at 11.30 with coffee; the coffee is drunk while doing more easy work such as email reading and a few small actions. Emails and MSN Messenger are only dealt with at certain times of the day and are not let to interrupt the work he states. In the late after noon most of the time was spent with MSN. MSN is preferred over the phone for small tasks and remarkably few phone calls were received during this day.

I will also mention a few other minor things that I took notes of during the day. They may or may not have implications on the set up of the computer.

Around 14.00 he did some administration of the Unosat mailing lists from a web interface.

He does not lock the monitor with an xlock kind of application but has a short timer on the screensaver that turns the monitor black.

The helping little animation (the cat) in MS Office is always present.

Caller ID is used on the phone to screen who is calling.

An exhaustive list of all applications used during the day is very short but contains most standard programs, Excel, Word, Acrobat reader, IE, messenger and GroupWise.

4.1.4 Summary of findings

In this chapter I make a summary of the findings of my user studies.

UNOSAT is a relatively small organization, but the director still has to manage lots of data. He has a stressful job and works long hours. The most important issue for him is to save time. Each little improvement in that area will ease the strain of his job.

The client has many ideas about how a new system should look like. Several of these are incorporated, as seen in chapter five, in the personal information management systems suggested by the research community, even thought solutions may not be available yet. Most of the ideas are about new ways of managing his huge data hierarchy. Today he keeps a well organized, but still cumbersome, hierarchy of folders and files. His wish is to be able to connect related files with each other in projects. He wants to have a unified workspace from which he can access all data in a project, from word documents to e-mails and instant messaging without having to change to a different program to view the files.

E-mails are an important part of the information flow to the client. He reads them systematically and organizes them into folders. E-mails that need to be acted on are kept in the inbox.

A lot of the client's time is spent traveling. Thus the mobility of the system is very important. Today e-mails may be read through a slow web interface. The best would be if even the documents on the server were accessible through a web interface, not only e-mails. Because of the mobile system, synchronization between the laptop and the server is an important issue, as is it with the Palm Pilot.

The client is an advanced user of MS Office so a new system must support these programs.

The current backup system is not satisfactory. An automated solution would be the best.

A good way of sharing information in the organization is desirable. An online calendar, video conferences and document sharing are all such features that a new system may support.

4.2 Client solution

This chapter describes the computer and the way it was set up for the client. The issue here was not to find the ultimate bleeding edge tools for each job but to set up a standards compliant working computer that "just works" and does the job it is expected to do.

When this project started the organization was a Windows only shop but the client had thoughts about a future shift to Linux. My initial reaction to this was that it is not as easy as it sounds and that a switch to Linux should not be made on a whim. I did not recommend a switch unless they had someone employed or planed to hire someone with Linux skills. I built this on the reasoning that even if both Windows and Linux have similar programs and in general are equally good, they are not the same and changing to a new operating system will from time to time cause problems or at least hiccups that might become real problems unless someone is there to solve them. Of course Windows also causes problems from time to time but there is already an established network of friends and colleagues that use Windows and might have had the problem before and have a solution to it. But as it turns out, the support organization at office 1 knows both Linux and Windows and UNOSAT is actually about to hire a person with Linux skills for other reasons. After some discussion we decided that this was an opportune moment to start a transition to Linux so instead of improving the old solution my job became to make a new solution in Linux.

According to the ethnographic study, the client is a heavy user of Microsoft Office and the Microsoft file formats are very common within the organizations and its partners. There are a

number of office suits for Linux which all have their strengths and weaknesses but there is one place where Open Office and Star Office wins hands down, that is compatibility with MS Office. Star Office and Open Office are based on the same code and offer, in most aspects, the same functionality. Star Office is sold by Sun Micro Systems and includes support and documentation at a reasonable price, Open Office is the free equivalent. Since Open Office and Star Office also are the most common office suit for Linux in the business world we choose to use Open Office as the standard office suit for the day to day work but since we know that only MS Office is 100% MS Office compatible we choose to also buy a $40 program called CrossOver Office from Codeweavers. This enabled us to run MS Office and a few other Windows programs in an emulated windows environment in Linux. The CrossOver Office is only a safety precaution, it is not supposed to be the main office suit of the computer, we did not want to take any chances and risk that an important document was not readable. In the coming months when the computer is used in the day to day work it will be evaluated wheatear or not the safety precaution was needed or will be needed in the migration of further computers.

Email is the second most important part of the day to day work for the client so a robust and feature rich tool is needed for this. Ximian Evolution was chosen as the tool for this job. It is a competent and stable program that is backed by Novell. Evolution has pretty much the same functionality as its competitors MS Outlook and Lotus Notes, the features include the ability to connect to a MS Exchange server and exchange calendars with others and use a centralized contact management with automatic address retrieval as well as a seamless synchronization with the palm organizer. Evolution was then configured to read emails from both the accounts at office one and office two via imap so the emails are always left at the server and are therefore accessible from any internet connected computer via a web based interface. The e-mails are also stored locally and synced whenever the program is open and the computer is online. One address book is stored locally on the computer and the computer is online it is also connected to the official address books of the organizations at the respective offices. The email software was configured to automatically retrieve and suggest completions from these address sources as soon as the beginning of a name or address was written in the address field. Evolution is not only an email package but can do al kinds of PIM related tasks and was set up to connect and sync with the palm pilot in the same way as it is done in Windows, with a simple press of a button.

A number of things that are taken for granted in Windows are often not available in standard Linux installations. Below I will just mention some of these, alone rather small things, that I had to change to make the computer acceptable and pleasant for a former Windows user, Many of them are actually added or changed as a result of the client trying the computer and even using them.

I installed common browser plugins such as java, shockwave and pdf. These were readily available and free but not present in the standard distribution. A modern look and feel was applied as a theme to the browser and the Gnome environment to make it look interesting and more familiar. Drivers for the wireless network were installed and the internet settings for the different offices where made easily available with ha point and click interface. This was a problem discovered in the ethnographic study. The external backup disk was converted to a format with Linux support and it was set to auto mount on connection, to show up on the desktop in a similar manner to the way it works in Windows. Drivers were also installed for other external peripherals connected to the docking stations at respective office, such as monitors and mice. To avoid confusion from information pollution, the equivalent of the start

menu was rearranged so that only the most common programs were immediately available and the more obscure programs were moved to a special folder called “other”. The last but most important of all the small details was to copy all the data and information structures to the new computer to make the user feel comfortable.

5 Analysis

This chapter is a step by step comparison between the findings I have from the literature studies of personal information management systems (PIMS) in the previous chapter and the common tools of today as applied in the client system (referred to as the client computer).

The broad definition of personal information proposed by Dumais et al. (2003) fits well with the information the client wants help to manage and what I observed during the ethnographic study. “Personal information management is the practice of managing information that helps us in our daily lives such as addresses, phone numbers, to-dos, appointments, notes, documents, folders, web pages and emails” Dumais et al. (2003). All except the web pages have been explicitly mentioned by the client during our discussions. One thing that is important to the client but is not mentioned by Dumais et al. (2003) is instant messaging. Much of the “small” communication is offloaded to instant messaging.

To use the concept of piles, mentioned by Malone (1983) is almost impossible on the client computer. The only way of organizing most documents is by filename and catalog structure. This system is, according to my findings, used extensively by the client. His information hierarchy is huge, tens of thousands of files in many gigabytes of data. The one good thing in this context is that everything is stored in simple hierarchy in the user’s home catalog, this makes it easy to backup the data and it keeps the personal data separated from the system- and program files. In the client computer, the files and folders manifest themselves to the user somewhat different in different programs, such as the email program, the web browser bookmarks and the ordinary files but they are all essentially the same, a hierarchy that starts at a root and diverge out to different branches. This folder system puts a lot of cognitive load on the user in terms of organizing. All folders have to be created and maintained in a good structure by hand or else they will become a cumbersome mess that is not possible to work with. There are a few collections of unstructured information that serve as actual piles; these are the top level of the structures. One example is the inbox. According to Belotti and Smith (2000) many users marked emails they needed to attend to later as unread in the inbox. These files are not organized in any way; they just lie there and wait to be taken care of. This is also the case with the client, though he instead removed all other mails except the ones in the “pile”. It is not exactly as the piles that Malone (1983) identified and not as good as the groups implemented in Raton Laveur, Belotti and Smith (2000) but it is a step in the right direction.

The folder system of today has almost all of the drawbacks of the files that were mentioned by Malone (1983). It is for instance only possible to store a document in one folder at the time so a document will never appear in both the folder related to a project and in a folder related to a person. It is possible to achieve this with certain files on the computer with something called symbolic links. They are similar to shortcuts in MS Windows but if you put a symbolic link to a file in a different folder it will appear to all programs and users as a real file even though they actually work on another file in another location. But none of the applications take advantage of this feature and it is too cumbersome to do by hand so it will not be used in the context of organizing information.

The unified index of the available personal information suggested by Dumais et al (2003) is in no way existing in the client computer. There is no centralization of all data files except that they are all stored somewhere in the user hierarchy of files. I have already mentioned that this has the advantage of making the user data and settings easy to backup but it does not help much in finding them and it does not relieve the user of the burden of organizing them by hand.

The suggestion from Belotti and Smith (2003) that the PIM system should be an integrated part of the experience is not supported at all by this setup. As I have already mentioned, many programs organize their own data in their own ways and the programs that do not, just store that data as files for the user to organize as he whishes. Most of the PIM programs are quite customizable and have a good flexibility which is important according to (Belotti and Smith 2003). The problem is that they all have to be customized separately. This is also the problem with the lightweight ness and simplicity, all programs have reasonably simple and lightweight interfaces for their PIM systems, but there are too many hierarchies of information spread in different places and there are too many different interfaces that deal with them so it becomes a burden to keep it all organized. There is also very limited computer support for this organization so most tasks have to be done manually.

The subjective classification principle suggested by Bergman et al. (2003) state that all information related to the same topic should be classified together regardless of their technological format. As seen in the problem and feature session, this was one of the things the client sought. A way of creating projects with all files that are related to the topic easily accessible. This is not followed in the today's system. Emails are stored in the email program's format, in the email folder and the internet bookmarks are handled by the web browser. Many other files are stored in the common information hierarchy and they can easily be accessed just by browsing through the hierarchy and double-click on the icons as they are seen, but other files than simple text files are not readily accessible within the hierarchy. The hierarchy starts an external program to view or manipulate the file so the user cannot go back and forth between documents or switch from one document to another of different type except by going back to the hierarchy and start a new program. Neither is it possible to access a document related to the current on any other basis then that it is located in the same folder. To a certain extent there are thumbnails available of some documents to make recognition and choosing easier while browsing through the hierarchy.

Another principle recommended by Bergman et al. (2003) is the Subjective Importance Principle; the importance of an information item should determine how visible or noticeable that item should be. There is a limited functionality for this principle already mentioned, such as the ability to store some data outside the standard structures or hierarchies such as in the inbox or on the desktop but there is no way to make the representation of the files more appealing or larger for more important issues. Neither is there a way to make them remind us of something or a way to make sure we deal with them on time. The closest we get is to the to-do list where we manually can add things like this, this works fairly well in many cases but you really have to put some effort into it because each item has to be added manually to the list and the parameters have to be set by hand.

The last subjective principle devised by Bergman et al. (2003) is the Subjective Context Principle; information should be retrieved and viewed by the user in the same context as in which it was previously used. There is limited support for some of the contexts identified by Bergman et al. (2003) but it is far from being good. The best of them is the support for the "temporal context". The most obvious example is the hibernate function of the computer. If the user is working on something but has to leave for a meeting or go home for the day he doesn't have to shut down his computer completely but can instead use a function that just stores all the computers volatile memories about running processes and open data files in one long stream on the hard drive so that the next time the computer is started it can all be restored in a few seconds; when the computer is turned on again all programs are still opened and running the same way they were when the computer was turned of. This helps restoring the temporal context. Another feature that somewhat supports this principle is the fact that the programs themselves often remember the location of the last files the user had open in a short, easily accessible, menu. This removes some cognitive load from the brain in the form of less searching for recently used data. Also the feature with virtual desktops that can hold their own set of open running programs gives support for the temporal context.

The browser function tabbed browsing is related to what Bergman et al. (2003) suggested should exist to let a person mark something for later reading. With ha simple click it opens a webpage in a new tab in the background while you keep reading the old page. This only works in the browser and not for any other documents but it is a step in the right direction. There is almost no support helping in restoring the internal context of the user, annotation functionality is very limited in most programs and nonexistent in the rest. Same goes for the external context; the computer does not keep any track of which programs or data the user is using and it cannot give you easy access to other programs or datasets based on the current task.

Dumais et al. (2003) and Bergman et al. (2003) say that most of the information a person accesses is created by that person or has at least been accessed by the person before. This was seen clearly in my, albeit limited, ethnographic study. During the day I spent studying the client he never accessed any new information except in the form of emails. The theory is that a person searching for a file already has certain knowledge about it before he starts his search and this can be used to simplify the searching and improve accuracy. This is used to some extent by the client computer in that it transparently stores and maintains certain metadata about all files. This metadata is for example the last time the file was modified and the size of the file. This information is stored without inflicting any extra work on the user. However, the metadata stored in this fashion is only the simplest information, no complex information such as relating projects or working context is included as was suggested by Bergman et al. (2003). Another thing missing in this area is the complete lack of possibilities to ad your own metadata, such as notes about the information. The search interface is also too limited to make good use of the available data. Simple text searches exists for both emails and texts files and you can even make arbitrarily complicated searches with a command line tool called Grep where you can use so called regular expressions to create the perfect search filter. But none of these comes close to the tools proposed by to Dumais et al. (2003) neither in terms of speed nor in terms of flexibility or usefulness.

6 Discussion

Personal information management systems as a research discipline is fairly small an also quite new. I have not been able to study all available material but I have done quite extensive reading of the material from the largest studies in recent years. They do not contradict each other on significant issues, instead they complement and assert each other a number of times. So I dare to say that I have a reasonably solid basis for my conclusions and that I have fairly good insight into the world of PIMS. In this chapter I will discuss my thoughts on the future of PIMS in a non-scientific manner.

One of the most challenging things to implement in the terms of PIM functionality, but one that I still expect to see in a reasonable future, is a fully "pimified" operating system. The Holy Grail is the fully unified index of all information in the computer. The research systems I have looked into have partially working solutions for most information, but only the operating system can give us the perfect search on all data. As hard drives and memory become bigger and cheaper we will be able to store all information in a way optimized for searches instead of for space requirements.

If we don't get the unified index on an operating system level I think we will se a number of competing indexing programs.

Once we have the unified index in one way or another, there will be a plethora of different interfaces to this information. Some of the search interfaces will be focusing on familiarity and have text searches while some will focus on visualization of data in graphical environments, in two or even three dimensions.

One interesting thing here is whether the search interfaces will be so good that we will never have to organize our data in anyway at all. Maybe all we will have to do is to make sure that the data is somewhere on the disk and then the computer will find it for us. This would relive us from a lot of work, but the risk is that we would collect many times more data just because it is so easy. Maybe we would even accumulate so much information that we loose the familiarity we have with it today. Most people today have some sense of what data they have on their computer and they can retrieve it with relative ease, but this may change if it becomes a no issue to add more data.

Other things that I expect to have in future computers are links and relations between all kinds of data items. While looking at one item I expect it to always be effortless to find other data files that the computer deem relevant in my current context of work (correctly or not).

On the other hand what we will see may instead be that the computers take another other road. Maybe people don't want computers to be super easy and super adaptable since this comes at a cost in the form of more complexity, maybe not for the day to day user but very much so for the code. Many claim that the programs of today are already too complex. That is why we cannot protect them from viruses, hackers and crashes. If we let the computer be a little bit less integrated and a little bit less easy to use we could instead have a rock stable, blazing fast computer that never looses any data for us. What you cannot loose, you can always find.

7 Conclusions

The purpose of this study was to examine PIMS from an operational and a research perspective and to draw conclusions about the characteristics of the gap between the two. The study indicated the following:

The operational solutions as implemented at UNOSAT are characterized by being designed from a machine point of view; this means they are designed to work in ways that are easily implemented in various programming languages. It seems like after the first basic implementation is done, features are added to simplify PIM tasks, such as search functions and highlighting abilities. This gives the system certain characteristics, such as:

- Information is stored in a hierarchic way.

- Each program has its own file format.

- Only technical meta data is stored along with the file.

The solutions from the research community are designed with a basis in how ordinary, non computing PIM is handled, to this the scientists have added functionality to simplify common tasks that the computer happens to be good at. It is in no way perfected but has some desirable characteristics.

-The information management is simplified to be as ubiquitous as possible.

-Recognition is used instead of recollection to relieve the user of cognitive load.

-Extended use of meta data and subjective contexts adapts the system to the user's needs.

The gap, in functionality, between the operational tools and the tools proposed by the research community is very large. Many features are missing or are inadequately implemented in the client computer. This may not be so strange because the development of the software has roots in completely different mindsets. The positive thing here is that even if the gap in functionality is large, the gap in time seems to be small. Many of the revised systems are already implemented, albeit in special environments and / or on special hardware. But since they are working, they are proven to be implementable and the user response has generally been positive. This together with the trend of faster hardware and more capable software vouches for a positive future for our personal information management needs.

8 References

- Barreau, Deborah K. 1995, Context as a Factor in Personal information Management Systems Journal of the American Society for Information Science. 46(5):327-339
- Bellotti, Victoria, Smith, Ian 2000, Informing the Design of an Information Management System with Iterative Fieldwork. Proceedings of the conference on Designing interactive systems Pages: 227 - 237 ISBN:1-58113-219-0
- Bergman, Ofer, Beyth-Marom, Ruth, Nachmias, Ra, 2003, The User-Subjective Approach to Personal Information Management Systems. Journal of the American Society for Information Science And Technology, 54(9):872-878
- Cameron, Glen T., Kuen-Hee Ju-Pak, "Information Pollution? Labeling and Format of Advertorials in National Newspapers," Newspaper Research Journal 27(1),Winter,2000.
- Cameron, Glen T. 1994, Journal of Public Relations Research, 1994, Vol. 6 Issue 3
- Chervany, N. and Dickson, G. 1974, An Experimental Evaluation of Information Overload in a Production Environment. Management Science, 20. Vol. 10.(1974)
- Dumais, Susan, Cutrell, Edward, Cadiz, JJ, Jancke, Gavin, Sarin, Raman, Robbins Daniel C., 2003. Stuff I've seen: a system for personal information retrieval and re-use. Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval 2003. ISBN:1-58113-646-3
- Hansson, Anita 2003, Praktiskt taget aktionsforskning som teori och praktik - i spåren efter LOM, Intellecta Docusys, Göteborg
- Kenneth E. Boulding 1966 THE POLLUTION OF INFORMATION http://csf.colorado.edu/authors/Boulding.Kenneth/pollution-of-info.html
  15.00 the 23 January 2004.
- Lundahl U, and Skärvad PH, 1999, Utredningsmetodik för samhällsvetare och ekonomer, Third edition, Studentlitteratur, Lund
- Malone, Thomas W, 1983, How Do People Organize Their Desks? Implications for the Design of Office Information Systems, Xerox Palo Alto Research Center.
- McKenzie, Jamie 1998, Review Educational Leadership 1998, vol. 55 issue 5
- Merriam 1988: Case study research in education: a qualitative approach, Jossey-Bass, San Francisco Calif.
- Metz, Cade, Clyman, John, Todd, Mark 2003, IM Everywhere, PC Magazine, 11/11/2003, Vol. 22, Issue 20
- Molander, Bengt, 1988, Vetenskapsfilosofi. Bokförlaget Thales, Stockholm

- Nielsen, Jakob 2003, Information Pollution http://www.useit.com/alertbox/20030811.html 15.00 the 23 January 2004.
- Nielsen, Jakob 2004, Ten Steps for Cleaning Up Information Pollution http://www.useit.com/alertbox/20040105.html 15.00 the 23 January 2004.
- Orman, L 1984, Fighting information overload with decision support systems. Journal of Management Information systems fall 1984. Vol. 1 no.2
- Palser, Barb, 2002, ADVERTORIAL WITH A TWIST , American Journalism Review, May 2002, Vol. 24, Issue 4
- Schenk, David 1997, Data smog: Surviving the Information Glut; Harper San Francisco ASIN: 0060187018
- Twist, Jo 2003, Web guru fights info pollution BBC News Online 13 October, 2003. http://news.bbc.co.uk/2/hi/technology/3171376.stm 15.00 the 23 January 2004.
- Waltzer, Herbert, Waltzer, Miriam B. 2001, Political Communication, Jan 2001, Vol. 18, Issue 1
- Yin, R.K. 1994: Case Study Research, Design and Methods, Second Edition, Sage Publications, Inc., California.