N d'ordre

SUPPORTING LIGHTWEIGHT REFLECTION ON FAMILIAR INFORMATION

AURÉLIEN TABARD

Thèse pour obtenir le grade de Docteur de l'université Paris Sud
Discipline : Informatique
Préparée au Laboratoire de Recherche en Informatique dans le cadre de l'Ecole Doctorale d'Informatique de l'Université Paris-Sud
Présentée et soutenue publiquement par Aurélien Tabard

October 7, 2009

JURY

Mme. Christine Froidevaux	Président
Mme. Abigail Sellen	Rapporteur
M. Pierre Cubaud	Rapporteur
M. James D. Hollan	Examinateur
Mme. Wendy Mackay	Directeur de thèse

Supporting lightweight reflection on familiar information
Aurélien Tabard

SUPERVISOR: Wendy E. Mackay

Paris, France, le 25 mai 2009

ABSTRACT

The introduction of computers in life-science laboratories has progressively changed how researchers handle information. This can be easily observed in the media used to manage information: paper in the form of articles, notebooks or scrap notes is now combined to a diverse set of digital resources: Digital libraries, electronic notebooks, wikis, etc. These changes are not simple technicalities, but are challenging how researchers capture, re-visit and reflect on information.

In this dissertation, I argue that information management is a reflective practice in which Biologists and, more broadly, knowledge workers make sense of the their ongoing activities. I present two field studies (Chapters 3 - 4) in two French research institutions (the Institut Pasteur and INRA) which emphasize that reflection is not only a post-hoc activity but happens as researchers manage information, when they decide what to save, what to discard and how to do it.

Based on these observations, I propose two systems, Prism and PageLinker, addressing different aspects of researchers' reflective practice. Prism (Chapter 5) is a hybrid notebook combining paper and digital notes with other sources of digital information such as bookmarks, emails or documents around streams of activity. Prism supports active and selective saving of information its longitudinal use showed how users dedicated a particular notebook (i.e. master) to synthesize information. PageLinker (Chapter 6) demonstrates a lightweight way to capture information by piggybacking on users' existing interaction with computers, namely with the Web browser. PageLinker field evaluation showed how it supports users' reflection on the navigation while making browsing faster.

RÉSUMÉ

Le développement de l'informatique dans les laboratoires de sciences de la vie a progressivement transformé la façon dont les chercheurs gère l'information. Ce constat peut être facilement observé au travers des médias utilisés pour gérer l'information. Le papier sous forme d'articles, de cahiers ou de brouillons est maintenant combiné à des ressources numériques: bibliothèques numériques, cahiers électroniques ou wikis. Ces changements ne sont pas seulement techniques mais recouvrent des évolutions de pratiques dans la façon dont les chercheurs enregistrent, retrouvent et réfléchissent à l'information qu'ils manipulent.

Cette thèse décrit la gestion d'information comme une pratique réflexive au travers de laquelle les biologistes, et plus généralement les travailleurs de la connaissance, appréhendent leur activité. Deux études de terrain à l'Institut Pasteur et a l'INRA soutiennent cette thèse (Chapitres 3 et 4). En se basant sur ces observations, je présente la réflexion non pas seulement comme une activité a-posteriori, mais comme une composante active de la gestion d'information, qui se déroule alors que les chercheurs décident d'enregistrer ou de mettre de côté de l'information.

Ces études ont mené à deux systèmes: Prism (Chapitre 5) et PageLinker (Chapitre 6), qui explorent deux aspects de la pratique réflexive des chercheurs. Prism est un cahier hybride combinant flux papiers et numériques. Son utilisation longitudinale à montré que les utilisateurs dédient un cahier particulier (maître) à la synthèse d'information. PageLinker illustre comment la capture d'information peut être active tout en ne distrayant pas de la tâche principale, en prenant avantage des interactions existantes entre les utilisateurs et l'ordinateur, dans ce cas précis avec le navigateur Web. L'évaluation de PageLinker montre que cette capture légère permet aux utilisateurs de réfléchir sur leur navigation tout en la rendant plus rapide.

PUBLICATIONS

Some ideas and figures appeared previously in the following publications:

Mackay, W. E., Van Kleek, M. G., and Tabard, A. (2009). Interacting with temporal data. In *CHI Extended Abstracts'09: Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, pages 4783–4786, New York, NY, USA. ACM.

Tabard, A., Mackay, W. E., and Eastmond, E. (2008). From individual to collaborative: the evolution of prism, a hybrid laboratory notebook. In *CSCW '08: Proceedings of the ACM 2008 conference on Computer supported cooperative work*, pages 569–578, New York, NY, USA. ACM.

Yuan, S., Tabard, A., and Mackay, W. E. (2008). Streamliner: A general-purpose interactive course-visualization tool. In *Proceedings of KAM'08, International Symposium on Knowledge Acquisition and Modeling*. IEEE Press.

Tabard, A., Mackay, W., Roussel, N., and Letondal, C. (2007). Pagelinker: Integrating contextual bookmarks into a browser. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '07*, pages 337–346.

Tabard, A. (2008). Interfaces réflexives, faciliter la co-adaptation avec des historiques d'interaction. In *RJC-IHM 2008 Actes des rencontres jeunes chercheurs en IHM*. 4 pages.

Roussel, N., Tabard, A., and Letondal, C. (2006). All you need is log. In *WWW 2006 Workshop on Logging Traces of Web Activity: The Mechanics of Data Collection*. 4 pages.

VIDEOS

The videos available on http://tabard.fr/videos/ illustrate Prism and PageLinker.

Ton acte toujours s'applique à du papier ; car méditer,
sans traces, devient évanescent, ni que s'exalte l'instinct
en quelque geste véhément et perdu que tu cherchas.
Écrire

— Mallarmé, *Divagations, Quant au livre, L'action restreinte*

REMERCIEMENTS

Ces remerciements viennent presque deux ans après avoir défendu ce travail, et avoir été éloigné de tous ceux qui y ont contribué:

Wendy qui m'a tant appris.
Catherine pour avoir toujours été là.
Nicolas R. pour les différents pieds à l'étrier.
Jean-Baptiste pour m'avoir guidé du début à la fin,
toujours le décalage.

Olivier & Evelyn for the fun.

Émilien, Sofiane, Nicolas M., Yann, Caroline et Renaud; Julie, Fanis, Matthieu, Guillaume, Emmanuel, Olivier, Stéphane, Michel continuez de faire et re-faire in-situ. Nathalie, jdf, Pierre en embuscade.

La Telecom team pour ces derniers mois passés ensemble. Éric merci. Sylvain, Thomas, Gilles, Anne et Mathias, au détour d'une coinche.

Je dois beaucoup à Abigail Sellen et Pierre Cubaud pour leurs commentaires sur les versions préliminaires de cette dissertation.

Plus loin.

Ev, MERCI.

Maud et Christophe pour ces soirées partagées. Frank, pour les autres ;) Florence, Vincent et Vincent, pour les dernières.

Pierre et Ina, Julien dire que vous êtes maintenant un de plus... la thèse est bien peu à côté. Jonas et Alexis, continuez de jouer!

Une note particulière pour toute la famille, la petite et la grande. Papa, Maman, Pauline, Grégoire, Mamie, Opa, Oma. Les semi-parisiens: Marie, Joan, François, Sandrine, Noémie et Ariane. Et la liste est longue, dans le désordre: J.B., Camille, Jhorland, Benoit, Olivia, Robin, Ambroise, Lucas, Arno, Diego, Thomas, Clothilde, et leurs parents.

Cette thèse à une bande son: Les Ogres, MAP, K. Arkana, Assassin, le grand Serge, les murmures du Limonaire et des cafés sans WiFi où elle a été écrite.

A.O.C. Paris XVIIIE.

À Marthe Habeillon

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Part I

THESIS

1 INTRODUCTION

This dissertation explores how knowledge workers engage in managing their information. It discusses the design of personal information management tools that not only support capture of information for subsequent retrieval, but also help users reflect on the information they manage.

Over the past two decades, the tools designed to handle information became central to everyday activities of knowledge workers (Barreau and Nardi, 1995). Yet, as they expose themselves to an increasing quantity of information, knowledge workers face the tools' shortcomings (Sellen and Harper, 2001): lack of integration, organization problems or dilution of attention. The quantity of information they manage may increase, but their abilities stay the same.

In order to cope with this flood of information, the study of Personal Information Management (PIM) focuses on supporting *"the activities people perform to acquire, organize, maintain, and retrieve information for everyday use"* (Boardman, 2004). The end-goal of PIM is then often portrayed as finding the *"right information, in the right form, at the right time"* (Jones and Teevan, 2007), or *"keeping found things found"* (Jones, 2007). While retrieving information through better search or browsing tools is necessary, the thesis focuses on a complementary approach: how can we support and let users benefit from engaging actively in PIM activities?

1.1 ILLUSTRATION

To better understand how knowledge workers manage information, I worked throughout the thesis with biology researchers. Biology researchers are at the forefront of the changes in the way information is created, collected, digested and managed, they face intense competition and need to work very efficiently (Dirks and Hey, 2007). They are lead users, *"users whose present strong needs will become general in a marketplace months or years in the future"* (von Hippel, 1988).

Consider Juliette, a 2nd year Ph.D student at Institut Pasteur who studies the genome of a fungus that causes disease. The following scenario illustrates how she engages in PIM activities and how the benefits she gets from managing information go beyond retrieving information.

> It's Thursday and Juliette is finally running an experiment that will teach her more about the effects of a new treatment on the expression of fungus genes. She already ran the same type of experiments in the past and is now quite familiar with the protocol. As she runs the experiment, she only notes in her laboratory notebook the modifications she did to the experimental protocol and the specific characteristics of the samples she used.

SELECTIVE SAVING: Juliette *selects and filters* the information she writes in her notebook. Since she knows the protocol, rather than copying it again, she only writes what she considers to be important. In doing so, she uses her past experiences and the hypotheses she had about the experimental results. Later on, when she will look back at her notebook, she will be able to identify more easily the salient properties of this particular experiment than if she had written all the information.

> Juliette leaves the samples she prepared in a scanning machine overnight. On Friday morning, she transfers the raw results into a spreadsheet on her computer. She starts to run preliminary statistical analyses but soon realizes she has to normalize her data in order to get significant results. She tries different normalization methods and compares their results. To do so, she draws a table in her notebook and fills it with a summary of the results.

ACTIVE SAVING: Juliette could have directly printed a spreadsheet and pasted it in her notebook. However she feels that by *writing the table manually* in her notebook *she better understands* the effects of the normalization methods.

> On Friday afternoon, she browses web sites about statistics for biology before analyzing the normalized data, since she is not very confortable with statistical methods. She prints a few pages describing different methods that seem appropriate to her analysis and skim through their descriptions. When she starts running her analysis the next monday, she is not sure what were the exact names of the methods that appeared interesting, but the printed documents in the corner of her desk remind her of the methods to use.

IMPLICIT SAVING: Juliette did not intend to save these pages but printed them to get informed. However, by seeing them on her desk, she benefits from contextual cues that help her remember.

The scenario above illustrates how researchers reflect on their activity in a lightweight manner as they manage their information. As researchers decide what they save, they transform the information they manipulate and frame it in the perspective of their ongoing activities. In contrast to hypotheses of total recall, researchers take advantage of their notebooks, personal information management systems and their context to filter what they want to remember.

1.2 THESIS

This dissertation argues that the personal information management tools designed for knowledge workers should not only support information retrieval but also the way knowledge workers reflect on their activity as they manage information. To do this, the dissertation:

1. Describes how researchers benefit from selectively saving information in their paper and digital tools, to make sense and reflect on the information they manipulate.

2. Shows how researchers adapted a paper+digital notebook we designed to support reflection by actively filtering, saving and synthesizing what they deemed important.

3. Demonstrates the efficiency of contextual bookmarks, which let users save information in context, in a lightweight manner that does not interrupt researchers' workflow, but also let them reflect on their past browsing experiences.

1.3 thesis overview

This dissertation is targeted to designers, developers and researchers. Designers and researchers will be interested in chapters 3 and 4 which provide insights on how knowledge workers manage information. Designers and developers can benefit from the design, architecture and evolution of a hybrid paper+digital notebook described in chapter 5, but also from the description of contextual bookmarks in chapter 6. Researchers should also find value in the chapters 5 and 6 which describe how knowledge workers adapted the tools we created.

Chapter 2 presents an overview of the related work. It discusses the life-logging and search-oriented approaches of information management and present alternatives based on studies of users' practices. It then presents how the scientific process is changing and the tools proposed to support researchers.

Chapter 3 characterizes the way Biologists reflect on activity as they manage information. Interviews of Biologists at the Institut Pasteur highlighted their use of various tools to collect and offload information in both notebooks and computer applications. These tools do not only serve to archive information, deciding what to save forces researchers to reflect on their activity: they filter, process and reframe the information they manage.

Chapter 4 presents the participatory design study conducted at INRA. We interviewed bioinformaticians and conducted participatory workshops which improved our understanding of personal information management practices in the laboratory. This field work also let us refine implications for the design of a laboratory notebook supporting reflection.

Chapter 5 presents the iterative design of Prism, a hybrid notebook which integrates interrelated streams of personal information: paper notebooks, digital notebooks and data from the computer. Bioinformaticians from INRA tested Prism and participated in its design for 9 months. This longitudinal use helped us understand reflection mechanisms.

Chapter 5 describes the benefits of integrating heterogenous streams of activity in order to gather information, and the difficulties that arise when users must manage this mass of information, aggregated from different locations and updated constantly. It finally shows how users kept master notebooks, where they organized thoughts and reflected on their activity rather than simply collecting information.

CHAPTER 6 presents the design and evaluation of PageLinker, a contextual bookmarking system. Based on the problems encountered by Biologists when they browse information on the Web, we designed PageLinker in a participatory way. PageLinker associates web pages together as users copy and paste data between them, creating local bookmarks to navigate from one page to the next.

When evaluating PageLinker, we tried to maximize internal and external validity with repeated time series experiments at users' workplaces. The evaluation of PageLinker shows that implicit capture improves navigation and retrieval of information.

CHAPTER 7 summarizes the contributions and outlines future work.

1.4 RESEARCH STRATEGY

Three approaches ground the thesis work: empirical findings, technological proofs and theoretical perspectives. This section presents triangulation as a method to navigate among these approaches, then presents grounded theory as an analytical method to justify empirical findings, and participatory design as a method to design and adapt tools with users.

1.4.1 *Framework: Triangulation*

To navigate among perspectives coming from social science, engineering and design, Mackay and Fayard (1997) applied the idea of triangulation to Human Computer Interaction (HCI). It is a simple framework that describes how the research and design models underlying HCI can be integrated. The framework promotes the use of different approaches, i.e., observation, design of artifacts and theory, to validate findings.

The research projects presented in the thesis instantiate specific paths among these approaches, e.g. from observation to design, or from a field evaluation to its theoretical generalization. The numbered arrows of figure 1 illustrates the relationships among the approaches developed in the thesis.

- Arrow 1 represents how we moved from a study of information management practices of Biologists and Bioinformaticians, to the characterization of reflection as a key activity of researchers.
- Arrow 2 represents how we drew implications for design from the conclusions of the field study.
- Arrow 3 represents the back and forth between the design of Prism, observations of users and participatory workshops.
- Arrow 4 represents how we applied observations of how Biologists' browse the Web to the design of PageLinker.
- Box 5 describes the participatory design of PageLinker.
- Arrow 6 describes the generalization we drew from PageLinker field evaluation.

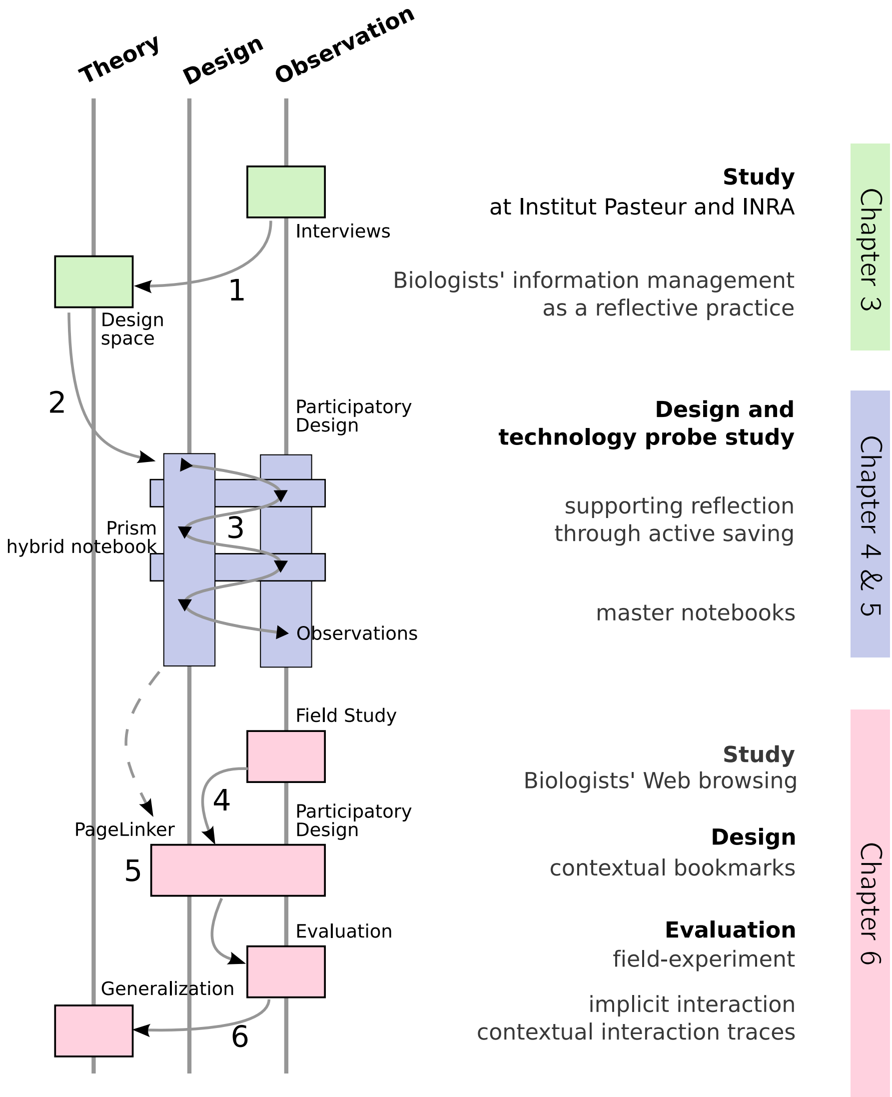


Figure 1: Thesis triangulation process, navigating between Theory, Design and Observation.

1.4.2 *Grounded Theory*

Grounded Theory (Strauss and Corbin, 1990) provides a set of methods and strategies for describing and analyzing field observations. Different techniques, including interviews, documents and observations, provide heterogeneous data which are iteratively coded. The codes are then sorted, critically analyzed, compared amongst themselves and grouped into categories. Finally, categories are analyzed by looking at relationships emerging from their descriptions. Grounded Theory does not aim to validate theories by illustrating phenomena, but rather highlights emerging phenomena which can be further explored by various studies and re-used in a design context.

1.4.3 *Participatory Design*

To better understand systems in use and the implications for design, I do not rely solely on observations but also on generation of technology. Throughout the thesis, I used participatory design methods to gather information about use, to prototype, to discuss design directions with users, and evaluate the systems produced. Muller (2003) defines Participatory Design (PD) as *"a set of theories, practices, and studies related to end-users as full participants in activities leading to software and hardware computer products and computer-based activities"*.

Participatory Design aims to produce an outcome closer to the needs of users by bringing designers and users together. It goes beyond user-centered design by letting users become actors at the different steps of the design process: problem definition, idea generation, prototyping and evaluation. Nonetheless, users do not act has designers per se, they share knowledge and help focus the design on *users'* problems rather than the *designers'*. The goal of participatory design is not only to gather ideas and feedback from users, but also reflect on what they do, to better understand the design context, and various aspects of the users' needs and desires that are unknown to the designers (Mackay and Fayard, 1999).

1.5 CONTRIBUTIONS

In this dissertation, I discuss how managing personal information is a reflective process. I detail how researchers manage their information, the systems developed to support their practice and how they adapted them. The dissertation shows that information management is not only about the support of acquisition, organization, maintenance and retrieval of information, but that it should also support the filtering, manipulation and transformation of information. The thesis offers the following empirical, technological and theoretical results:

1. **Empirical Findings.** The results of two field studies highlighted the reflective nature of Biologists' work practice. Based on Grounded Theory, I describe the dynamics of researchers' reflection. I discuss how the term hypomnemata (tools supporting reflection) is well suited for Biologists' notebooks and information management applications. I also describe how saving information happens at two levels: when Biologists capture information and when they articulate what was captured in an intelligible way.

   A subsequent study of Biologists' web browsing provides insights on the complexity of automating repetitive yet alternative tasks. I propose repeated time-series field-experiments to evaluate the benefits of tools supporting the management of personal information. This evaluation method balances internal and external validity. Experiments are replicable and are also ecologically valid since they are based on realistic scenarios designed with users and tested repeatedly in-situ.

2. **Innovative and adaptable technologies** helped us validate our field observations and revealed unexpected uses as researchers adapted them.

- **Prism**, a hybrid notebook, integrates streams of information coming from paper notebooks, digital notebooks, the computer and the Web. Prism evolved from a desktop application to an online application integrating heterogenous streams from users' activity, and which users adapted by creating "master notebooks" dedicated to reflection.
- **PageLinker**, a contextual bookmarking system improves webpage re-visitation by putting past navigation into context. The contextual traces of PageLinker support users by associating together related webpages they visit. These traces do not rely on chronology or hierarchies which minimizes users' organization efforts. Based on the implicit creation of contextual bookmarks, PageLinker lets Biologists reflect on their past navigation paths without disrupting their workflow.

3. **Theoretical perspectives.** Based on these different studies, I argue that the tools knowledge workers use to manage information affect how they perceive information. As knowledge workers transfer information from one medium to another, filter information, synthesize it or articulate their activity, they reflect on information. This reflective practice does not happen only after the fact as they try to make sense of their activity and imagine new directions. It happens in the course of action, as knowledge workers manage information, write in their notebooks, take classificatory decisions and sort information.

2 RELATED WORK

This chapter describes the literature relevant to the design of personal information management systems. It provides background on personal information management and discusses technological solutions. It then presents how Biologists must handle information both in the laboratory and on their computer, which leads to a discussion of technologies integrating paper and digital information.

The increasing use of computers and the Internet in research laboratories challenges scientists' established work practices. As they handle information on different media, physical or digital, and access information faster than ever before, both locally and remotely, they face personal information management challenges. In this chapter, I discuss studies of personal information management practices, the design of systems supporting these practices and how this applies to Biologists.

2.1 PERSONAL INFORMATION MANAGEMENT

The study of Personal Information Management (PIM) drew upon different disciplines, i.e., human-computer interaction (HCI), digital libraries, database management, information retrieval, and artificial intelligence. This led the initial research on PIM to focus on retrieving information, by applying principles of traditional information management to personal information (Boardman, 2004). As an illustration, Bellotti et al. (2002) described PIM as *"the ordering of information through categorization, placement, or embellishment in a manner that makes it easier to retrieve when it is needed"*.

Kidd (1994) departed from this position by insisting on the importance of *"getting informed"* over *"passively filing large quantities of information"*. In order to clarify the approach to PIM developed in the thesis, I discuss the terms composing personal information management, before discussing the design of personal information management systems.

Information

Information has a diversity of meanings coming from different disciplines such as Semiotics, Information Theory, or Computer Science, to name a few. Buckland (1991) provides a usable approach to information. He considers that information does not only denote the *"process of informing"* (the communication of knowledge) or *"knowledge imparted"* (knowledge about a particular fact). Information also *"refers to things that are informative"*. Thus information is not just an abstraction but situational and concrete; documents or objects, i.e. *"things"*, can be informative.

Personal or Familiar Information

Boardman (2004, chapter 2) defines personal information as *"information owned by an individual or under their direct control"* which suggests a binary division between what is personal and what is not. However, ownership of information can be difficult to define in the digital domain (Lessig, 2001; Stallman, 2002). A visited webpage is not really owned by the people visiting it, yet the information it conveys progressively becomes familiar.

In the thesis, I take into account the notion of familiar information (Roussel et al., 2005), which limits the importance of ownership of information. Familiar information is *information that users have manipulated or are aware of, with which they have a degree of familiarity*. Bergman et al. (2003) showed that as people manipulate information they consider it differently. The relationship to information is progressive and involves different steps where the same information can be considered differently over time.

Managing Familiar Information

The progressive relationship of users to information leads to reconsider what personal information management systems should be. Jones (2007) notes that the *"input-store-output* [or capture, organize, retrieve] *characterization is seriously limited"*. He argues that information management and information use are interwoven. The designers of PIM systems should not only consider the information captured but also how it is used and in which context. The studies presented in the next section describe how knowledge workers manage and use information, and the problems they face with existing systems.

2.2 STUDIES OF INFORMATION MANAGEMENT IN-SITU

The immediate context plays a role in how people manage information. Departing from early theories underlying HCI, Winograd and Flores (1985) and Suchman (1987) criticized the cognitivist view that both human mind and computers are information processors manipulating representations of the world.

With a phenomenological perspective, Winograd and Flores insisted on the role of artifacts in how people interact with the world. In their view, computers are tools used in an environment that cannot be formerly described. It is only by interacting with artifacts that people can build an interpretation of the world.

Coming from an anthropological tradition, Suchman's Situated Action emphasized the interrelationship between people's actions and their context. Situated Action *"underscores the view that every course of action depends in essential ways upon its material and social circumstances. Rather than attempting to abstract action away from its circumstances and represent it as a rational plan, the approach is to study how people use their circumstances to achieve intelligent action."*

The situated approach motivated the study of personal information management practices in context, particularly in the workplace. The following sections describe four characteristics of PIM in the workplace: How knowledge workers leverage their environment to process and remember information; how they keep information for other purposes than future reuse; how search is limited when managing information users are familiar with; and how information relevant to knowledge workers spreads over different media and applications.

2.2.1 *Use of context in managing information*

The work of Malone (1983) illustrates the willingness to learn how people manage information in their everyday work context. He studied how people organize their desk and identified two key strategies: filing and piling. He notes that categorizing information is cognitively difficult and that informal piles on the desk allow people to avoid the cognitive effort required for long-term filing; *"desks are cluttered and seemingly function as a spatial holding pattern for current inputs and ideas"* (Kidd, 1994).

Barreau and Nardi (1995) confirm that users do not manage information simply to retrieve it later but also store items as tasks reminders. Norman (1993) argues that many of the tools we use to process information also help us offload our minds from thoughts, and let us look at information from another perspective. An argument further developed by Hutchins (1996), who described how cognitive processes are distributed over time, space and multiple actors, in his studies of control rooms of naval ships or aircraft carrier.

Rather than the result of some disorganization or bad practices we should consider papers spread on desks, post-it notes or other information scraps (Bernstein et al., 2008) as optimization of information management processes (Kidd, 1994). Whittaker and Hirschberg (2001) explain that filers may engage in premature filing: to clear their workspace, they archive information that later turns out to be of low value. Filers amass more information, but access it less frequently than pilers. The authors finally note the importance of older/discarded archived information for many people, this discarded information should be of no use but is still kept.

2.2.2 *Beyond utility*

Information management does not necessarily adhere to a well-defined rationale. Kwasnik's (1991) study of classification practices in a physical office identified a number of contextual factors that influence classification decisions. She concludes that people do not make classification decisions based purely on the document's attributes such as title and author, but that the organization depends on situational factors, such as how the documents will be used.

In a study of academics, Kaye et al. (2006) showed that personal archiving goes beyond efficient storage and retrieval of information. They describe their subjects' motivations for archiving as *"not only information retrieval but also creating a legacy, sharing resources, confronting fears and anxieties, and identity construction"*. They argue that these rationales were mapped into the subjects' physical, social and electronic spaces: facilitating mobility or improving serendipity for example.

Marshall and Jones (2006) note that *"utility, serendipity, and the pleasure of encountering what we save relies on more than search alone"*. Information can be kept for sentimental reasons, act as a reminder, or allow us to *"forget the manifold things [we do] not need to have immediately at hand"* (Bush, 1945).

2.2.3 *Limitations of search*

Not all information has the same status, which may explain the different strategies of individuals. Barreau and Nardi (1995) identified three types of information: ephemeral, working and archived. They observed that ephemeral and working items were mainly retrieved by browsing whereas archived items were searched for.

Paraphrasing Teevan et al. (2004), the perfect search engine may not be enough. They observed that users prefer orienteering, i.e. navigating one step at a time, over teleporting, i.e. jumping to a search result. Orienteering let users specify less information at once (they do not have to specify a query). This preference on browsing over search is not due to the quality of search engines (Bergman et al., 2008). In a study comparing different search engines, they found that users preferred navigating rather than searching within their file system. Accessing information on its computer is different from searching for new information on the web. Users are already familiar with the information on their hard drive or the one they browsed recently on the Web. Browsing lets them access information easily but also learn about available information on the way. The authors conclude that search is a last resort when users cannot remember information location.

2.2.4 *Intricate nature of information*

Another problem of search is that personal information management is embedded within other activities and applications. Boardman and Sasse (2004) show that from one application to the next, hierarchical classification are usually redundant. Working on a particular project, users jump from one application to another. Bondarenko and Janssen (2005) argue that electronic documents should be embedded within meaningful context of information and easily accessible for regrouping as the task goes on.

Bondarenko and Janssen base their conclusion on a cross study of documents' use across the physical and digital worlds. They observed that users could easily reorganize their paper documents on their desks but could not do it as conveniently on their computers. Corroborating earlier studies, they also note that filing does not embody all information management practices, but that it is strongly related to task management. For example, email systems do not act only as information and communication systems (Mackay, 1988), but also as ad-hoc time and task management or personal archiving systems (Whittaker and Sidner, 1996; Ducheneaut and Bellotti, 2001).

2.3 personal information management systems

Studies on personal information management practices influenced the design of PIM systems. In parallel, possibilities offered by the development of technology led researchers to investigate extensive capture of users' activity.

The research on information management has a long standing history in the field of Computer Science, which can be traced back to the Memex. In 1945, Bush imagined a mechanical desk that one could use to read a self-contained microfilm library, the Memex (figure 2). His vision of *"a device for individual use, which is a sort of mechanized file and library"* influenced different threads of research on personal information management.

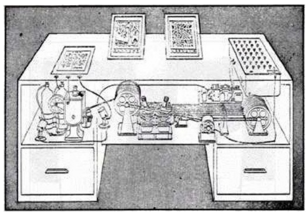


Figure 2: the Memex.

2.3.1 *A technology-driven approach to personal information management*

The Memex influenced engineers from Engelbart (1962) to Bell and Gemmell (2007), who explored how to manage large quantities of familiar information. With the development of ubiquitous computing (Weiser, 1991) and the increase in storage capacity, funding institutions recognized the challenge of storing and accessing large stores of personal information. Whether DARPA[1]'s LifeLog or EPSRC[2]'s Memories for Life initiatives focused on technology-driven developments.

LifeLog aims to centralize extensive records of information from individuals. Mann's (2004) *"life logging"* or *myLifeBits* (Gemmell et al., 2006) illustrate this approach of augmenting memory by capturing 'everything': *"every event we experience, conversation we participate in, and any piece of digital data we ever touch. According to this vision, these accurate digital records can then be accessed to re-live past events."* (Sellen et al., 2007).

Most lifelogging work focuses on technology, without trying to understand how users manage their digital and physical information (Sellen et al., 2007). Like many ubiquitous computing projects, the technological possibilities overshadowed the potential benefits for users (Rogers, 2006). Users are not necessarily interested in capturing extensive information (Petrelli et al., 2009). Furthermore, Chalmers (2004) and Dourish (2004) showed the complexity of capturing users' *context*. Cameras, audio recordings or sensors only capture a limited account of users' social relationships and interactions with their environment. Nevertheless, these recordings can provide cues helping people to remember past events.

2.3.2 *Memory prostheses*

Based on insights coming from psychological studies of memory, researchers from the Xerox EuroPARC investigated the design of "memory prostheses" (Lamming et al., 1994) whose goal was to extend human memory. The underlying idea was to capture information about

1 The Defense Advanced Research Projects Agency (DARPA) is the central research and development office for the U.S. Department of Defense

2 Engineering and Physical Sciences Research Council (EPSRC) is the UK Government's leading funding agency for research and training in engineering and the physical sciences

users' activities in order to provide cues for helping people to remember information and past events.

The memory prostheses projects shared the idea that the *"physical context can be a powerful cue for recall"*. This assumption was based on the observation that people organize memories of past events into episodes (Tulving, 1983).

Newman's (1992) PEPYS system created logs of people's movements in order to build *"recognisable descriptions of past episodes"*, that could provide cues to help participants remember about past events. Based on this experience, Lamming and Flynn (1994) developed *"Forget-Me-Not"* as wearable system that would track users' activities such as users' location, email exchanges, phone calls, or file editions. Users could then visualize the personal data that had been collected and search for specific events.

Memory prostheses focused on improving users' memory by leveraging their context. They offered information about users' actions but provided limited access to users' information. The following section presents systems that try to incorporate the information people manage on the desktop with its context of use.

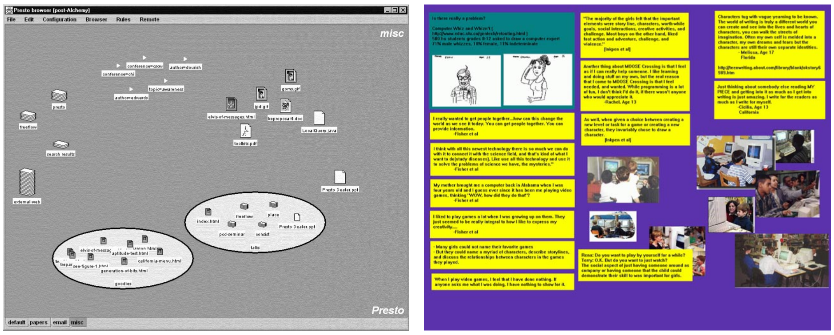
2.3.3 *Situated systems*

The studies presented in section 2.2 highlighted the importance of context when managing documents either in the office or on the desktop. However, existing operating systems isolate applications one from another which makes it difficult for users to re-organize information across applications but also to put digital documents in context. Users are forced to manage the cumbersome and redundant hierarchical organization of the file systems, bookmarks and email folders (Boardman and Sasse, 2004).

Alternative desktops

To reduce the problems of hierarchies and categorization, Dourish et al. (1999) proposed Presto, a document management system providing user-level document attributes (see the documents piles and groups on the Desktop in figure 3a,). These document properties let users informally group and stack documents as well as filter information. MessyDesk, MessyBoard (figure 3b) extend the desktop and let users organize information on a flat surface. Users can drop information snippets on their modified desktop as they would do onto a physical bulletin board (Fass et al., 2002).

Spatial organization allows users to informally organize their space in a personal way, by letting them connect documents together and control their relationships based on spatial placement. With Data Moutain, a 3D bookmark organizer, Robertson et al. (1998) showed that spatial organization of documents improves remembering. Furthermore, these spatial visualizations can improve the overall browsing experience by integrating navigation and reading activities into a shared space (Cubaud et al., 2002). Even though Spatial organization let users browse, organize and use information in a flexible, e.g. non-hierarchical, way, documents must still be organized.

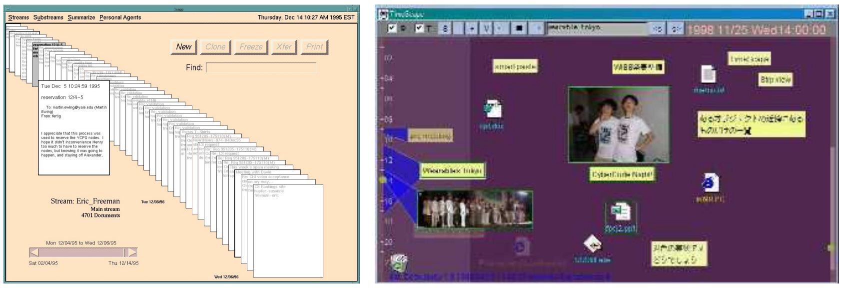


(a) Presto desktop (b) MessyDesk

Figure 3: Grouping and flexible organization of documents on the desktop.

Time based systems

Rather than directly confronting classification and organization, Lifestreams (Fertig et al., 1996) or Rekimoto's (1999) Time Machine Computing leverage the temporal dimension of data to organize information. They provide a *"time-centric approach to organizing information on computers"*, allowing users to navigate in the temporal states of the folders (figure 4).



(a) Lifestreams (b) Time Scape

Figure 4: Temporal representations of desktop documents.

To facilitate retrieval and recall of documents on the desktop, Stuff I've Seen (Dumais et al., 2003) associates to the opened documents the media manipulated at the same time (documents, photos, sounds, videos). When browsing or searching for documents, users can leverage their memory of the activity they were pursuing. Going further in integrating the temporal context, Ringel et al. (2003) added external landmark events such as the weather or news headlines to the documents context.

While the temporal context helps remembering information, organization problems may persist, particularly when users switch between tasks while working on their computers. If information is only organized according to its temporal dimension, it may lead to group unrelated information or activities together.

Activity based systems

UMEA (Kaptelinin, 2003) groups documents and applications based on their temporal proximity into projects. When a user works on a project and opens a document, the document he handles is associated with the project. However, unlike systems only based on time, users can define projects manually and associate or remove information associated with a project.

Rather than following a semi-monitoring approach like in UMEA, Bardram et al. (2006) proposed the activity-based computing (ABC) framework to let users manage activities on their desktop. The ABC framework let users create and manage activities completely manually. Centered on an activity bar, it allows users to group windows and resources into activities that can be resumed or suspended to switch between tasks.

In the same spirit, Giornata (Voida and Mynatt, 2009) allows users to group documents, share them or tag them. TAGtivity (Oleksik et al., 2009) takes a more lightweight approach to activity management. It focuses on rapid document tagging by enhancing windows with a tagging facility. Interestingly, supported by its quick input facility, TAGtivity was not only used to group documents for the long term, but also for transient and short term tasks or to increase the visibility of ressources.

2.4 THE CHANGING SCIENTIFIC PROCESS

Biologists, and scientists more broadly, are exploring new ways of managing their activities and their personal information. The availability of computational ressources and the Internet has radically changed how they work. They keep working in the laboratory but must also manage ever increasing amounts of information on their computers to do their jobs.

2.4.1 *Lead users*

Dirks and Hey (2007) qualify scientists as *"extreme knowledge workers"*. Scientists run experiments and must publish as quickly as possible, in a complex setting where they explore alternatives and make choices based on frequently changing resources. They rely on data gathered in the laboratory and their peers' work to produce knowledge, which they share in scientific publications or data repositories.

Biology researchers can be considered to be *lead users*: *"users whose present strong needs will become general in a marketplace months or years in the future"* (von Hippel, 1986). The study of how Biologists adapt to the changes in their work environment can help us to identify new solutions adapted to their complex practices. *"Since lead users often attempt to fill the need they experience, they can provide new product concept and design data as well"*. They can bring to light problems faced by everybody tomorrow.

2.4.2 *Improving researchers' workflow*

Termed e-Science or Cyber-infrastructure, Wouters (2004) describes the changes faced by scientists as *"the combination of three different developments: the sharing of computational resources, distributed access to massive datasets, and the use of digital platforms for collaboration and communication."* Many e-Science projects offer new tools to help scientist cope with the changes in their digital environment (Hey and Trefethen, 2003). For example, the Labscape project (Arnstein et al., 2002) is an interesting attempt to create a ubiquitous data capture, sharing and organization system across a large set of devices but also among people. More generally, e-Science projects involve digital tools that support scientists' workflow at different levels: the laboratory, the analysis process, the sharing of resources.

In industry, many Laboratory Information Management Systems (Paszko et al., 2002), focus on the integration of entire data collection and analysis tasks for whole laboratories. Other systems focus more specifically on the analysis part on the research work. For computer centered analysis work, the Taverna workflow management system helps Biologists automate analysis tasks (Oinn et al., 2006). These systems are useful, particularly associated with online sharing websites such as myExperiment (Roure et al., 2009).

Unfortunately, few e-Science workflow systems are user-centered (HPCwire, 2007). Beale (2009) explains that workflow systems integrate many different applications which may not have been originally designed with users in mind. The flaws of each application are passed to the system integrating them. Another limitations of workflow systems is their design orientation toward streamlining repetitive tasks, with little support for exploration and investigating alternatives. They are usually based on the idea that researchers follow work procedures that can be formally defined and to some point automated.

However, Latour (1988) argues that the ongoing scientific practice is very exploratory. In the laboratory, a typical experiment sometimes produces inconclusive data, which is attributed to failure of the apparatus or experimental method. A large part of scientific training involves learning how to make the subjective decision of what data to keep and what data to throw out, something which is hard to automate and control. This process requires learning and is part of what Schön (1984) terms *"a reflective pratice"*. This practice is largely hidden behind the idealised mechanisms of science, the tools and the methods. In the systems presented in the following chapters, my interests lie in supporting not only the well-defined processes but also the ways in which researchers rely on experience and intuition.

2.4.3 *Digitizing notebooks*

In order to support the exploratory nature of scientists work in the digital world, Butler (Nature, 2005) describes digital notebooks as a normal evolution of paper laboratory notebooks. A broad range of electronic notebooks exist (Lysakowski, 1997; Polonsky, 2006). Schraefel et al. (2004) provide a classification of e-notebooks around two axes: *"the degree to which paper is kept/replicated or entirely replaced, and the degree to which the system for the device is personal (like a lab book) or distributed (like the Web)"*.

In their design of an e-notebook, Schraefel et al. (2004) focused specifically on transferring the physical properties of laboratory notebooks on a tablet PC, while taking advantage of the PC possibilities (search, computation, sharing). In the context of knowledge workers, Notime (Lamming, 1991), Filochat (Whittaker et al., 1994) or Dynomite (Wilcox et al., 1997) are early portable devices for the capture and retrieval of handwritten (and audio) notes. The authors explored how to transfer the properties of paper on digital systems.

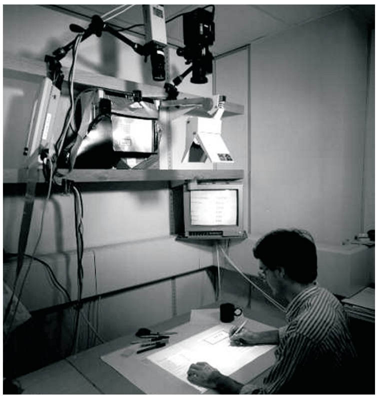
However, going from paper to digital is rarely neutral (Sellen and Harper, 2001); in the move, some of the particular properties of paper such as grasping, carrying, folding are lost (Sellen and Harper, 2001). Some aspects of paper notebooks such as durability or ease of reading (O'Hara, 1997) are also lost. The domain of tangible interaction shows that it is possible to keep interacting with physical objects or paper documents while manipulating digital data.

2.5 HYBRID WORLDS

Mackay (1998) presents two opposite directions to bridge the gap between digital tools and paper:

1. Redesign computers with paper affordances.
2. Augment paper with digital capabilities.

While starting in opposite directions, these two approaches share the common goal of blending the paper and digital worlds. Designers should explore the most appropriate forms of capture, storage, retrieval and, most importantly, interaction with information, with regard to its media form. Rather than opposing physical and digital realms, designers should try integrate them.



(a) The Digital Desk

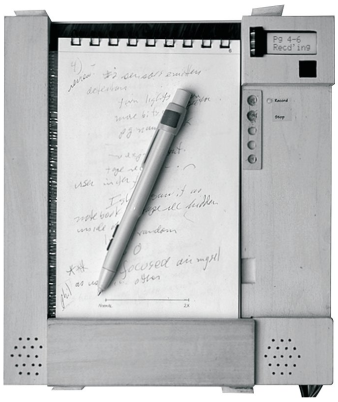
(b) Video Mosaic

Figure 5: Augmented desks

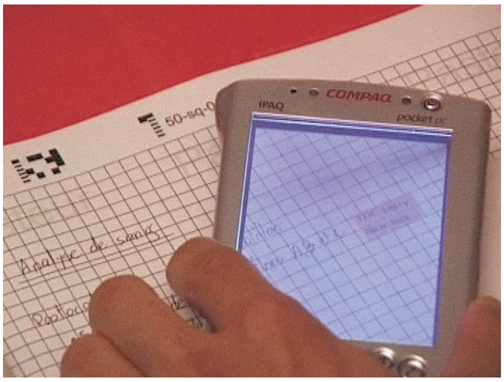
An early example of the research mixing paper and digital documents is the Digital Desk (Wellner, 1993). In this seminal work, Wellner explored how to augment a physical desk with digital capabilities (figure 5a). In Video Mosaic, Mackay and Pagani (1994) directly used paper storyboards as an input method for online video editing (figure 5b). Tangible representations of digital data do not only carry information in themselves, but also in how they are laid out or physically organized. Mackay et al. (1998) showed that the spatial organization of paper flight strips (used in air traffic control rooms) carried information about the ongoing activities of controllers, and offered a flexibility that could not be matched by the computer. It is not only the content of the medium that carries information but also the way people handle and organize artifacts within their environment (in this case the spatial layout of paper flight strips).

2.5.1 *Augmented notebooks*

Mackay et al. (2002) also built a series augmented laboratory notebooks. Biologists could write in a paper laboratory notebook, the handwriting or coded images were captured by a portable or desktop graphics tablet and linked to a searchable electronic version of the notebook. Based on similar technologies, a paper notebook and a tablet, Stifelman et al. (2001) developed the Audio Notebook which adds audio recording to a normal notebook (figure 6a). With the A-book (Mackay et al., 2002), a PDA became a *'physical information lens'* that provided an interactive window on any page of the paper notebook, with two-directional links between the paper and the computer (figure 6b).



(a) The Audio Notebook



(b) The A-book

Figure 6: Augmented notebooks

Yeh et al. (2006) developed ButterflyNet a mobile notebook for field biologists, based on the Anoto technology. With ButterflyNet, Biologists can associate handwritten notes with photographs, sensor readings, GPS track logs. ButterflyNet focused on heterogenous field-data capture, whereas Prism (described in chapter 5), focused on integrating information from the desktop and the Web with paper and digital notebooks.

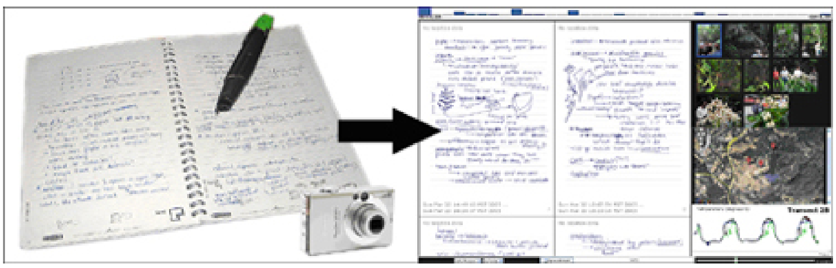


Figure 7: ButterflyNet

2.5.2 *Augmenting Paper with the Anoto technology*

The advent of Anoto[3] provided lightweight, inexpensive, hardware for capturing hand-writing on paper. The Anoto technology is split in two components: paper and pen. Anoto paper sheets are augmented with a tiny dot pattern (figure 8a, 8b), which encodes locations in a notebook (and even in a set of notebooks). According to Anoto, the coding mechanism can uniquely identify positions within a set of nearly 73 trillion sheets of A4 paper. The pen contains an internal camera tracking the dots on paper as users write. The Anoto technology provides a robust and mobile handwriting solution. As a result, most of the recent pen-and-paper literature is based on the Anoto system.

(a) Anoto dot pattern (b) Anoto notebook and pen (c) Pulse pen

Figure 8: The Anoto technology

Writing and Annotating paper-digital documents

The Anoto technology made it simpler for researchers to develop techniques for finding and modifying information on paper. Costa-Cunha and Mackay (2003) explored annotations and indexing mechanisms in notebooks. Using semantically significant gestures (such as an underline) users could index content (make it appear in the notebook table of contents) or link paper space to digital information.

With Paper Augmented Digital Documents (PADD, Guimbretière, 2003) paper printouts act as proxies for digital documents. Readers can annotate their Anoto printouts and the annotations appear on the digital documents. PaperProof (Weibel et al., 2007) goes beyond annotations and allows active readers to modify the digital documents from the paper printouts. In the same vein, PapierCraft (Liao et al., 2008) extends PADD by providing a gesture-based command system to let users to manipulate digital content directly on paper.

Tsandilas et al. (2009) designed Musink to deal with the lack of immediate feedback when writing or annotating music on paper with Anoto. Musink interprets the gesture once the pen strokes are uploaded to the computer but allows composers to correct or redefine the meaning of gestures on the computer. Until recently, the only feedback possible

3 http://www.anoto.com

with Anoto pens were blinking LEDs and pen vibrations. In 2008 Livescribe released the Pulse smartpen (figure 8c), an audio smartpen that can record audio and provides visual and audio feedback via a small screen and an embedded speaker.

Toolkits

Interacting with paper implies developing applications differently than on traditional computers (Signer, 2008). Yeh (2007) explains that *"programmers must work with the pen hardware and abstract raw input into high-level events. Program code must coordinate interactions across time and space, and send output to devices"* (since Anoto pens cannot display information). The lack of instant feedback also makes it difficult to debug and understand where the problems come from.

Based on the experience from a variety of systems (such as paper power-points (Signer and Norrie, 2007) or interactive exhibition leaflets (Luff et al., 2004)), researchers from ETH Zurich developed the iPaper framework supporting the rapid development and deployment of interactive paper applications (Norrie et al., 2006). During the same period, based on the experience of ButterflyNet, Yeh et al. (2008) developed the Paper Toolkit, which facilitates the design of interactive paper based systems. We used the Paper Toolkit in Prism (see chapter 5) to import and display the pen strokes coming from Biologists' paper notebooks.

2.5.3 *Alternatives to Anoto*

A few alternatives to the Anoto technology exist. We mentioned above early prototypes based on video capture and tablet technology. Researchers from the mi-lab in Austria combined vision algorithms and barcodes to augment sheets of Anoto paper. They can then project graphics on sheets or other surfaces, for creating interactive sketches (Block et al., 2008), or finding paper documents in a drawer (Seifried et al., 2008).

Another approach developed by EPOS[4] consists of using a base to triangulate the position of the pen using infrared sensors (figure 9). Mistry and Maes (2008) used this technology to create hybrid sticky notes that can send messages, be searched or located. However, this technology cannot differentiate on which page users are writing.

Figure 9: EPOS triangulation technology.

2.6 SUMMARY

I first discussed how managing information is a process in which users progressively get familiar with the information they manipulate. Then, I presented how the reconstructive and episodic properties of memory led to the design of systems whose goal was to augment human memory by providing cues about past events. The studies of information management in the workplace reveal the situated nature of PIM. These studies led to the design of digital systems introducing alternatives to the hierarchies of file systems, either using virtual layers, spatial organization, centering information around activities or time.

4 EPOS: `http://www.epos-ps.com/index.asp`
Mimio is technology similar to EPOS designed to augment whiteboards.

I then described how technology is changing how Biologists manage their information, and the technological answers presented in the literature. I finally discussed systems integrating paper and digital information, that could fit into laboratory work. The Anoto technology provides the most robust and usable solution to integrate paper and digital documents.

The changes in Biologists work practices described in this chapter motivated further studies of personal information management practices in the laboratory. Chapter 3 and 4 respectively present studies of Biologists' and Bioinformaticians' work practices. Both chapters highlight the role of information management in the way researchers reflect on their activity.

3 BEYOND INFORMATION MANAGEMENT: REFLECTION

This chapter presents a study of Biologists' personal information management practices. As information is distributed over computers, servers, paper and digital notebooks, researchers have a hard time making sense of the information they manage. Using Grounded Theory, we found that researchers do not save information only for future use, but also as a means to reflect on their ongoing activity.

3.1 INTRODUCTION

Biologists manipulate and transform information as they generate new hypotheses, design new experiments or make sense out of noisy data. Their increasing use of computers and of the Internet challenges how they "do science", by providing access to diverse experimental protocols, huge databases of information, and a wide variety of analysis techniques. This new wealth of information is difficult to manage and to account for. In this context, Biologists adapt the tools they use to their new questions and experiments, but also reflect on their activity in order to gather new ideas and justify their choices.

The study presented in this chapter investigates how the increasing use of computers influenced how Biologists manage their personal information and adapted their laboratory notebooks to support a reflective practice. Schön (1984) introduced the concept of *reflective practice* to describe how practitioners consider critical incidents in their life's experiences, to reflect and improve their practice. Schön distinguished two types of reflection: *Reflection-in-action* which occurs as a problem is being addressed, in the 'action-present', and *Reflection-on-action* which happens after the fact, when one looks back upon ones' activity.

> The reflective practitioner is the one who *"allows himself to experience surprise, puzzlement, or confusion in a situation which he finds uncertain or unique. He reflects on the phenomenon before him, and on the prior understandings which have been implicit in his behaviour. He carries out an experiment which serves to generate both a new understanding of the phenomenon and a change in the situation."* (Schön, 1984, p68)

Schön examined the reflective practices of five professions (engineering, architecture, management, psychotherapy and town planning), describing the discrepancy between the technical rationality practitioners learned and their everyday practice. Biologists face similar problems. Latour (1988) described how they navigate between a formal idea of science and its ad-hoc application in the laboratory. I argue in this chapter that Biologists also reflect upon their research. In their course of action, they reframe assumptions and generate new ideas as they manipulate information.

Biologists, and scientists more broadly, exhibit extreme needs and illustrate prominent characteristics that can help us understand future mainstream behaviours. Dirks and Hey (2007)[1] describe them as *"extreme knowledge workers"*. We observed how they explore new ways of managing information with regard to the emerging possibilities offered by technology. Their new uses can reveal how we should design future personal information management systems.

3.2 FIELD STUDY AT THE INSTITUT PASTEUR

We conducted a study at the Institut Pasteur between February and June 2007. Evelyn Eastmond and Wendy Mackay took part in the interviews and Catherine Letondal was instrumental in recruiting participants. The initial results were published in (Tabard et al., 2008), I present a second analysis, in this chapter, focusing on the use of notebooks in the management of Biologists' personal information.

3.2.1 *Setting*

The Institut Pasteur is a semi-private, non-profit foundation founded by Louis Pasteur in 1887 in Paris. It contributes to the prevention and treatment of disease. Approximately 2800 persons work on the Parisian campus in three fields of activity: research, education and public health.

At the Institut Pasteur, researchers must use laboratory notebooks. In an effort to manage scientific information, the French Minister of Research[2] even developed a *National Laboratory Notebook*[3] coming with a manual and a motivation booklet. Many French research institutions (INRA, INSERM, CNRS, the Curie network) echoed this effort and created their own laboratory notebook policy.

The management of the Institut Pasteur promotes the use of laboratory notebooks for three stated reasons:

- For research, the notebook is a memory tool that helps researchers document past protocols, choices, failed experiments or abandoned hypotheses. Laboratory notebooks help researchers retrieve information and avoid repeating experiments already done and losing information.

- For patenting, the notebook guarantees the traceability of research results and identifies the date and authorship of research results.

- For archiving and quality management, the notebook helps researchers benefit from the laboratory's expertise and facilitates in-house knowledge transfer. Louis Pasteur's notebooks are still available and readable (figure 10).

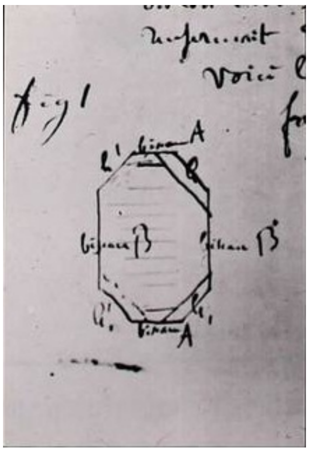


Figure 10: figure from one of Louis Pasteur's notebook.

Most researchers develop their own practices regarding notebooks use. However some rules apply to paper laboratory notebooks which make them different from less formal notebooks such as scratch pads. Biologists are not supposed to delete content or add new content to

1 T. Hey is the former leader of the UK eScience initiative and currently in charge of external research at Microsoft.

2 Ministère délégué à la recherche

3 http://www.cnrs.fr/infoslabos/cahier-laboratoire/docs/cahierlabo.pdf

past notes; they must also keep their notebooks in their offices as long as they work in the same laboratory.

When they leave for another position, the notebooks stay in the team for a few years until the ongoing projects are over. Then, notebooks are moved to the archives. When something deemed particularly important happens, Biologists may sign their notebooks and ask a witness to countersign, and archive it sooner than usual for better patent protection.

3.2.2 *Participants*

We interviewed 10 people at the Institut Pasteur: eight Biologists, a Bioinformatician and a computer scientist (CS). Two were Ph.D. students, seven were researchers and one was a senior researcher heading a team. Except the computer scientist, all had training in biology, and some followed subsequent courses in programming (three) or learned scripting by themselves.

The participants investigate very different topics: from yeast molecular genetics requiring laboratory experiments, to the study of viral infection mechanisms which involves breeding mice over many generation, or even more theoretical molecular models. They must be able to reason about their choices, decide among alternative approaches and produce academic results such as articles, theses, reports, project proposals or grants.

3.2.3 *Procedure*

We conducted semi-structured interviews of individual researchers in their laboratories or offices. We framed the interviews around a core set of questions (Appendix A) that evolved as new topics or problems were raised in the discussion. The interviews lasted approximately one hour and we videotaped them when possible. (Since we interviewed researchers in sensitive environments, some interviews were not videotaped at all. For others, we stopped recording when participants showed us sensitive data.)

We used a variation of the critical incident technique (Flanagan, 1954), to get specific use cases and to ground the interviews in problems regularly experienced by participants (Mackay, 2002). We asked participants to recall recent critical incidents, both positive and negative, and to describe them. For example, when was the last time they wrote something on their notebooks? When did they look for something on the computer, analyze data or perform other common tasks (such as designing a protocol or sharing data with colleagues)?

We looked at researchers' activity through a common entry point: their laboratory notebooks. The paper laboratory notebook is the central tool for managing information (Mackay et al., 2002). Like diaries, notebooks provide a record of users' activities which facilitates the recall and re-contextualisation of past events.

We focused on the personal information management practices in the laboratory and on the computer. We were particularly interested in the paper/computer file management, the way researchers came back to information they knew about, what and how they recorded their research.

3.2.4 *Analytical method: Grounded Theory*

Several qualitative methods are available for studying interactive phenomena among users, tools and their environment. We can distinguish several frameworks used in HCI: Phenomenology, Ethnography, Activity Theory, Grounded Theory. These frameworks come from different traditions, with different goals and require specific training to be used properly (Forsythe, 1999). Of these, Grounded Theory lays out a clear analytical method and seemed the most accessible.

I used Grounded theory (GT) (Strauss and Corbin, 1990) to understand the dynamic characteristics of personal information management in the laboratory, and the way in which researchers reflect on their activity. Starting from field data to let phenomena emerge, rather than (in)validating theory through observations, GT provides an open framework to the investigator for gathering and analyzing qualitative data. Figure 11 describes the three coding steps used to analyze the data coming from the interviews:

1. Open coding consists of a systematic examination of the qualitative data. The investigator defines questions in order to reduce complexity of the data and explore it in a systematic way. These questions allow the investigator to identify and categorize occurrences of the micro-phenomena of interest to the study. The results of this first analysis is the identification of the concepts and categories around their properties and dimensions.

2. Axial coding consists of relating categories to their sub-categories. The term "axial" refers to coding around the axis of a category. The investigator compares the concepts against each other in order to look for common themes along the categories, which reduces the number of concepts.

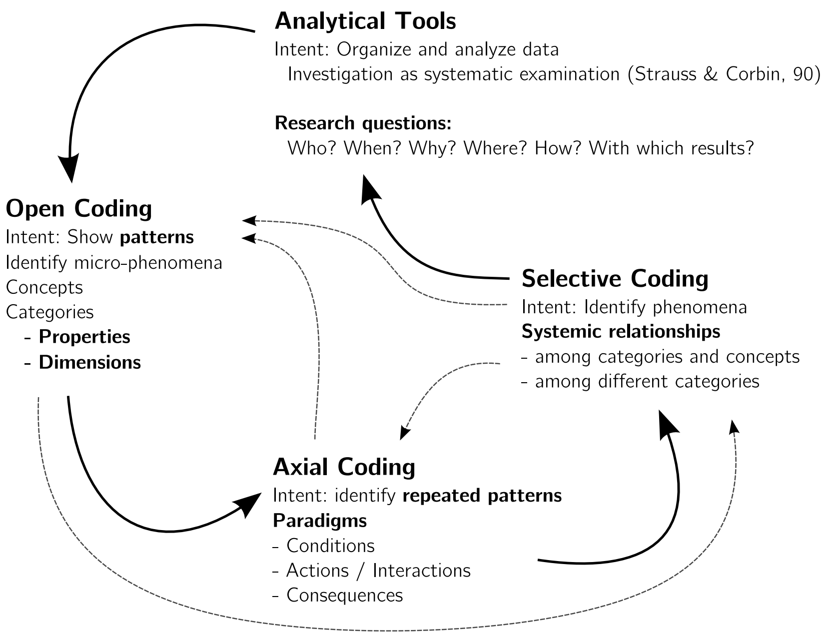


Figure 11: Grounded Theory Analysis process

3. Selective coding consists of integrating and refining the theory. The investigator looks for relationships among the concepts and categories. The interactions among the elements is what Strauss and Corbin call a paradigm: the set of causes, actions and consequences constituting a phenomenon.

The founders of Grounded Theory, Glaser and Strauss, disagree on the data coding methods. While both Glaser and Strauss use the same terms to present the analysis steps, they differ on the coding and categorizing methods. Glaser insists on induction, emergence, and the individual researcher's creativity within a clear set of stages. The analysis presented in this chapter follows the approach described in Basics Of Qualitative Research (Strauss and Corbin, 1990) which proposes a matrix to categorize and sort data. This explicit categorization helps non-experts analyze their data.

3.3 RESULTS

I present here repeated patterns observed among participants (axial coding, see Appendix B.2 for the overall organization of the categories). I detail the categories, properties and dimensions coming form the open coding phase of the grounded theory in Appendix B.1. First, I describe how paper and electronic notebooks let users manage their information in complementary ways. Then, I describe how participants filter and transform information as they process it and write or save it in notebooks. Finally, I describe the evolution of organizations schemes both in the laboratory and on the computer.

3.3.1 *Paper and electronic notebooks are complementary*

Table 1 summarizes the types of notebooks used by ten participants[4]. Most used a combination of notebooks, from scratchpad for calculating the concentration of a product, to paper laboratory notebooks for defining a protocol, or electronic notebooks for saving and annotating a DNA sequence. All but one (Victor) used paper or electronic notebooks. Seven out of ten participants used official paper laboratory notebooks, combined with less formal notebooks such as scratchpads or off the shelf paper notebooks. Six participants used what they considered to be electronic notebooks (text, Word or HTML files). None used commercial electronic notebooks.

Paper laboratory notebook use

Paper laboratory notebooks play a central role in the personal information management practices of Biologists. Biologists use their notebooks to remember information, reflect on their activity or prove that their work came first. They write in their notebooks regularly, from a few times a day to a few times a week. They consider it fundamental to their research, not some additional task.

Paper laboratory notebooks record the different aspects of Biologists' research. They hold action items, preliminary hypotheses and more definitive procedures such as protocols or analysis methods, so that the author (or someone else) can replicate experiments. Notebooks not

4 The names have been changed.

Participant		Paper Notebook	Digital Notebook
Marie	(CS)	-	HTML NB
Julie	(Bioinformatician)	1 official	-
Delphine	(PhD student)	1 official	HTML NB
Daniel	(PhD student)	1 scratchpad, 1 official	Word files
Victor	(Senior researcher)	- (only data folders)	-
Alain	(Researcher)	1 official	-
Sabine	(Researcher)	1 scratchpad, 1 official	-
Carole	(Researcher)	2 informal, 1 official	Word files
Christophe	(Researcher)	1 official	Text files and Blog
Pierre	(Researcher)	1 small paper notepad	HTML NB

Table 1: Laboratory Notebook use at Institut Pasteur

only organize information about experiments, they also help manage the wealth of information researchers handle daily, from laboratory material to computer data or notes about teaching, tutoring or collaborations with colleagues.

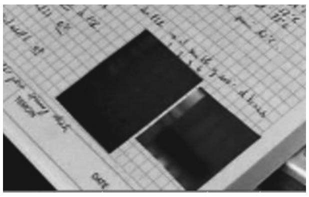


Figure 12: Gel in a notebook.

Figure 13: Pasted and annotated protein sequence.

Besides hand-written notes, lab notebooks contain information in diverse forms. Biologists insert temporary information as post-it notes or loose sheets of paper. They also paste in experiment results such as gels[5] (Figure 12), photo negatives, or computer printouts. The later may be raw data that has to be analyzed or the final version of a successful analysis script that a Biologist wants to save in case of computer failure. Paper notebooks contain many other references to digital information such as URLs written on the top of pages, printouts of data sequences (Figure 13), images, and references to articles.

While many Biologists also use e-notebooks, paper laboratory notebooks provide several benefits in terms of:

REACTIVENESS: They have no boot time. They are always accessible at the desk while exploring data, reading or writing articles but also at the bench during experiments. Delphine uses her paper laboratory notebooks when she conducts experiments, even though her laptop with her e-notebook is only two meters away, because it is *"always available"*[6].

MOBILITY: They can be moved around, brought to the bench or to meetings. For example, when Julie discusses analysis methods with her colleagues, she brings her paper notebook.

ROBUSTNESS: They do not require batteries, nor do they require regular backups. They resist to stains which means that most participants brought their notebooks at the bench.

OPENNESS: They let users write freely. Delphine explained that she could write down results in the order she wanted, add drawings but also easily paste outputs from machines in her notes.

5 Gel electrophoresis is one of the principal tools of Molecular Biology. This technique separates deoxyribonucleic acid (DNA), ribonucleic acid (RNA), or protein molecules using an electric current applied to a gel matrix and is usually performed for analytical purposes.

6 "toujours disponible"

LEAFING: They can easily be skimmed through. Furthermore, Delphine, Sabine and Christophe added physical marks (bookmarks, post-its) to retrieve information more quickly or leave cues about interesting information.

ARCHIVAL: Paper laboratory notebooks are known for their stability, they are the permanent record of work in the laboratory. They can easily be accessed and read years later, but are also quickly available on the shelves. Julie, who collaborates with different teams to help them analyze their data, explained that she wrote extensively in her notebook so that *"if one day [colleagues] need to get the results, I know where is the information, which microarray it is related to and I can get back to the results"*[7].

Electronic notebook use

Paper laboratory notebooks are not sufficient to handle all of Biologists' information. Most participants used e-notebooks side by side with their paper notebooks. E-notebooks are convenient to document Biologists' online activities. Sequence analysis, for example, would be *"very complicated to transcribe in a paper notebook"*[8]. However, research activities rarely happen solely in the laboratory or solely on the computer.

Biologists go back and forth between experiments in the laboratory and analyses on their computers. Figure 14 shows Carole explaining how she has information duplicated in her paper and electronic notebooks. Like many, she has a hard time deciding what should be kept on paper or on the computer. Biologists are thus experimenting with diverse types of e-notebooks. They are looking for e-notebooks they can adapt and which would let them handle the complex web of information distributed across different media.

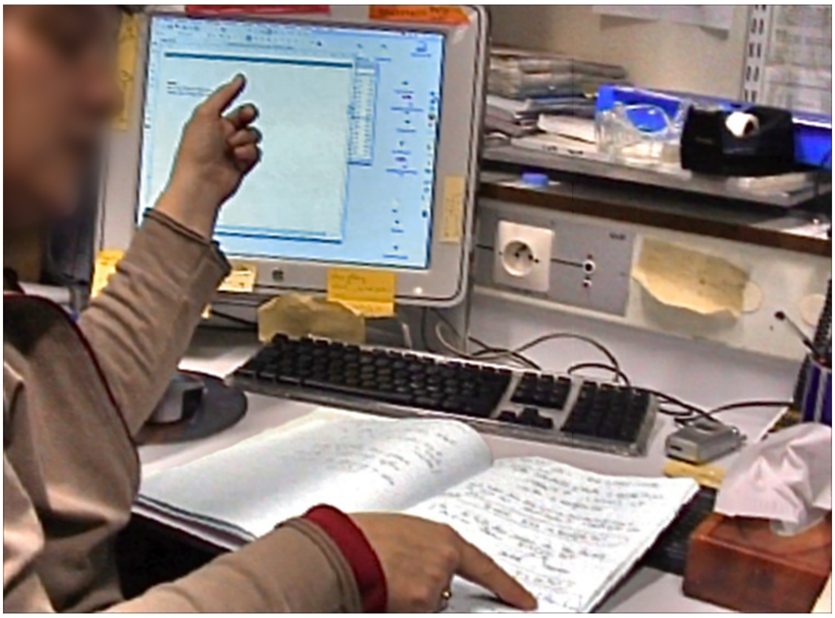


Figure 14: Carole explains how she has trouble managing information both in her paper and digital notebooks.

7 "Si un jour ils ont besoin de récupérer les résultats je sais que telle information, ça correspond à telle puce, et je peux re-sortir les résultats."
8 "C'est super compliqué sur un cahier de labo de la retranscrire." (Sabine)

Many participants tried commercial e-notebooks to manage their digital information, but none used them for more than a week. Commercial e-notebooks incorporate many constraints in the way users can input information, in order to be considered as official laboratory notebooks. For example, daily entries can only be edited during the day they are about and past entries are locked in order to provide temporal proofs. These notebooks are usually designed for the industry and carry a vision centered around repetitive workflows and well-defined procedures. Commercial e-notebooks are not suited to the exploratory aspects of the research carried at the Institut Pasteur. When typing a note, researchers do not want to specify up-front whether they are entering information about an experiment or about an analysis, or if there is an associated protocol already in the notebook database.

The Biologists we interviewed preferred to use text, Word or HTML files, that they can adapt. Daniel and Carole used Word files, named by date, where they copy and paste snippets of digital information. Christophe experimented with blogs and wikis before switching for simple text files. Pierre developed and used a simple HTML notebook that Delphine and Marie borrowed.

The file-based e-notebooks follow a chronological structure and support quick input. Users only have to open a file and can start typing. In order to quickly access their e-notebooks whenever they want to add notes, Biologists store their notebook files on the desktop, and usually leave their e-notebook open in order to access it faster.

The main motivation behind the use of e-notebooks is to re-use information. Participants copied the final version of a command line, of a script or of filtered data. Figure 15 illustrates a day of Carole's e-notebook: she copied command lines, wrote instructions on how to manipulate outputs, added references to another notebook-like file, copied command line parameters, file path to data and even interesting data snippets (a DNA sequence). Similarly to other use of electronic

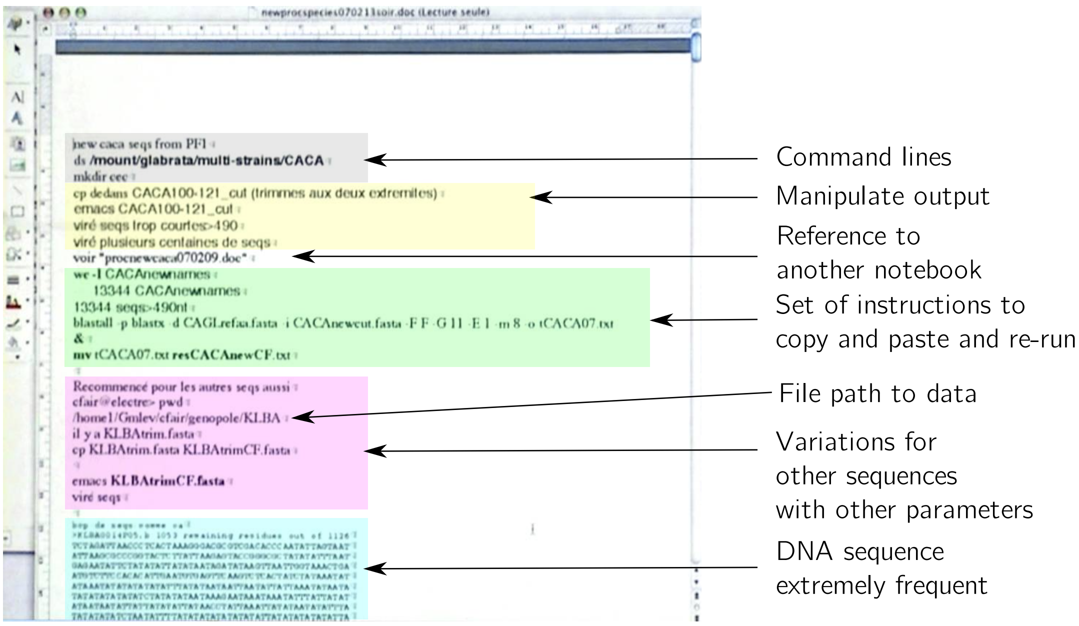


Figure 15: A Microsoft Word based notebook with different types of information pasted in.

notebooks we observed, Carole saves information from the computer with minimal transformations, in order to re-use it later in the most straightforward way.

Overall, Biologists use e-notebooks that provide the following advantages:

EASY UPDATING: E-notebooks can be updated at any time by editing the files. Carole described how she would develop an analysis in the file where she first described it. Delphine explained that she would update parameters of commands if she found more efficient ones or add references to new notes in past ones.

VERSIONNING: E-notebooks do not have size limitations and let users save different version of documents. Pierre, Christophe and Delphine copied scripts or functions that they were about to modify in order to keep a trace of the latest working version.

RE-USE: E-notebooks let Biologists copy and paste content in order to use it again. Pierre, Christophe and Delphine re-used command lines or functions they saved to analyze new data. Carole discovered she could copy and paste sets of commands from her notebook and execute them in batch. Since then, she saved the most efficient ones and pasted them, the ones after the other in her e-notebook.

SEARCH: E-notebooks can be searched with desktop search engines since they are based on files. Pierre used Google Destkop, Delphine used Apple Spotlight and Christophe used grep commands to retrieve notebook information.

SHARING: E-notebooks can be shared, the notebook files can be made available on servers and visible to colleagues. Marie used her HTML notebook to communicate with remote colleagues. At some point, Pierre experimented with sharing his notebook by putting it on a shared server.

MOBILITY: File-based e-notebooks can be copied and moved. Delphine, Carole or Pierre regularly copied their e-notebook to a USB key or a server to work on it from home or to backup a version.

Flexibility and constraints of notebooks

The training of Biologists influences the notebooks they use to save information. For example, they learn how to write systematically in their laboratory notebooks. However, Biologists also use different notebooks depending on their properties. We can distinguish notebooks' flexibilities and constraints at the following levels:

INPUT TECHNIQUE: *Paper notebooks* let users write and draw freely, paste printouts, but force them to input information manually.
Digital notebooks let users type information, accept copy and paste, insertion of documents, but make it difficult to illustrate notes with schematics or drawings.

FORM OF THE MEDIUM: *Paper notebooks* do not need any set-up and have no load time. Biologists can use them at the bench and they do not need batteries, but they are limited and fixed in size.

Digital notebooks do not have size limitations (at least it was not a problem for these Biologists). They can eventually be mobile but Biologists do not use them at the bench and rarely during meetings or conferences.

CONTENT: *Paper notebooks* can hold paper sheets and printouts. Biologists rarely modify earlier notes and do not delete content.
In *digital notebooks*, Biologists can easily insert text, or can expand past notes as an analysis goes on over a long period.

Paper is often praised for its flexibility. However, if paper notebooks persist after the advent of personal computers it is rather because of the discipline that paper lab notebooks impose. Biologists adhere to the temporal and spatial constraints of the notebook and rarely edit or delete information. The result is a definitive record, a snapshot in time of what Biologists, after reflection, found most important to record. This gives paper laboratory notebooks archival status: what is written is final and available for posterity. The disciplined writing in paper lab notebooks makes them more valuable than less well structured electronic logs or other forms of paper notes.

Since Biologists can easily edit the e-notebook files, it becomes difficult to find their 'definitive' state. This very flexibility in the revision process encourages a corresponding lack of discipline. What is useful at a particular moment, such as updating a to-do list or modifying a file, makes retrospective analysis of what happened far more difficult. Just because electronic and paper notebooks can, in principle, contain the same information, does not mean that they do.

Figure 16 illustrates how Biologists use their different notebooks depending on the expected life time of the information they save. Paper laboratory notebooks are for the long term. Electronic capture the ongoing activity on the computer. Scratch pads or informal paper notebooks let Biologists capture information on the fly, at the bench or during meetings.

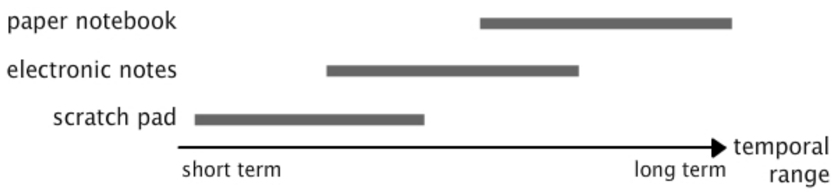


Figure 16: Notebook use, from throw-away scratch pads, to mid-term electronic notes, to archival laboratory notebooks.

3.3.2 *Writing and saving*

Biologists primarily write in their notebooks to save information for the future. However, writing and saving documents in notebooks provides additional benefits to Biologists. It forces them to decide what information they want to save which leads them to reflect on their activity.

Filtering information

When Biologists write or paste information in their notebook they must decide what is relevant and potentially useful in the future. Since

Biologists can hardly save an extensive account of their activity they filter the information they save. This filtering process takes many forms, from selective note taking to subtle management of documents in notebooks.

As researchers write, they must select information that is relevant in their current context. When Delphine conducted an experiment for the third time, she did not write the whole protocol in her notebook but only what she modified since the last time. By only writing down the changes, she highlighted the evolution of her experiment. Since she added references to her past experiments, when she looked back in her notebook she could reflect on the evolution of her hypotheses and results.

The physical nature of paper permits subtle distinctions in how information should be interpreted. For example, Carole and Sabine kept copies of gels in their notebooks. They only pasted the best one, and let the other loose between two pages. Biologists who program (Julie, Delphine, Daniel and Christophe) kept printouts of scripts in their notebooks (figure 17). However, they only pasted in the code if it was truly valuable and likely to be used in the future. Biologists keep samples loose knowing they will be lost at some point in the future. When they decide to paste content, they select (and filter) information.

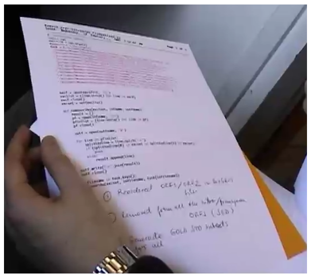


Figure 17: loose sheets from Christophe's notebook.

The tangible properties of paper let Biologists filter information and organize it in a dynamic way. Delphine explained how a sheet describing a protocol evolves over time, used independently at the bench, then inserted in her notebook and finally pasted in or written down with changes or comments. The same physical objects transmits different information depending of its context.

Processing information

As Biologists manipulate data in the laboratory or on the computer, they get familiar with the information it conveys. Rather than writing a script, Carole preferred to scan visually 30,000 to 60,000 sequences during three days. In the process, she discovered interesting sequences that she would not have noticed with a program. After 'seeing' the sequences (analyzing and filtering them), she was able to classify them but also created two new projects she would not have created had she used a script to filter her data.

Writing is another case in which Biologists process information and get familiar with it. Half the participants used informal notebooks which are more personal and do not have an official status. In these notebooks, they took notes they did not read later, especially during meetings or conferences. As Carole explained: *"I take them because I find it easier to understand by writing"*[9]. In a later interview, a Bioinformatician stated: *"when it's written, it's read"*[10]. The writing process appears necessary to incorporate the meaning of what is written.

On the computer, writing is somewhat different. Biologists value their digital notebooks because they can easily save heterogenous content through copy and paste or by linking to external data. The notes are more scarce, with minimal explanations about the information aggregated through copy and paste. They provide less context about the motivations and the development of activities. This could be explained

9 "Je les prends parce que je comprend mieux en écrivant"
10 "Quand tu l'as écrit, tu l'as lu"

by the fact that e-notebooks are not official notebooks. However, Pierre who only used an e-notebook and considered it as his official notebook was struggling with this problem. In order to force himself to provide context, he added to the start and end of his daily entries the questions *"What did I want to demonstrate?"* and *"What did I learn?"* even though he acknowledged that he did not answer the questions as often as he would like.

The value of notes does not only lie in the content that was recorded in order to be used later. Note taking also helps researchers articulate their activity as they pursue it: *"in such a way that everyone should be able to understand"*[11] (Christophe). The properties of paper notebooks support reflection, notebooks let users write, insert or glue documents according to the value they put on the information at hand. Paper notebooks help users accomodate the fact that information change over time: that one gel may be interesting today but of little value in a few months, when Biologists update notes, they leave traces. In E-notebooks, on the other hand, users can update their content as information changes or evolves, but traces about the changes are less visible or must be created explicitly.

3.3.3 *Organizing information*

Organization among the laboratory and computers

Beyond notebooks, the desk and the laboratory environment also play a role in holding and organizing information. At the bench, the experimental set-up provides distributed information about Biologists' activities. At the desk, the layout of post-it notes, articles, notebooks or protocols helps researchers resume their tasks or retrieve information.

On the computer, it becomes more complex to create these temporary and dynamic organizations of documents. Nevertheless, the desktop acts as a temporary place holder. Victor uses *"ongoing work"* icons on his desktop, Carole placed project folders on the bottom left part of her desktop, whereas Sabine uses a *"current project"* folder until she knows how to name the corresponding sequence (and thus can file it).

Chronology and projects

To bridge the gap between their physical and digital data, most participants rely on shared organization schemes across media. This leads to a constant back and forth between paper and digital information. Figure 18 illustrates how organization strategies carry from one medium to the other. Folder hierarchies focusing on projects on the computer (Figure 18-d) lead to project folders in the physical world (Figure 18-b). Whereas the chronological ordering of paper notebooks (Figure 18-a) leads Biologists to use folders organized by date on their computers (Figure 18-c).

The chronological organization of notebooks plays an important role in their efficiency, *"as we are sure to retrieve information"*[12] (Sabine). However, some activities do not fit in this chronological organization: conferences, projects data or long term experiments. *"It is also the problem of lab notebooks, when many experiments are going on in parallel, either you*

11 "de telle sorte que n'importe qui puisse comprendre"

12 " J'aime bien quand c'est fait de manière linéaire chronologique on est sur de pouvoir le retrouver."

have three notebooks or you have only one, but either things don't go at the same speed or they are mixed, and this is a problem"[13] (Sabine). This explains why chronology is not suitable for Victor: *"genetics is long, to characterize a gene it can take five to ten years [...] I'm currently writing two papers finalizing four years of work."*[14] In such cases, a fall back solution is folders associated with an analysis (as in figure 18-b).

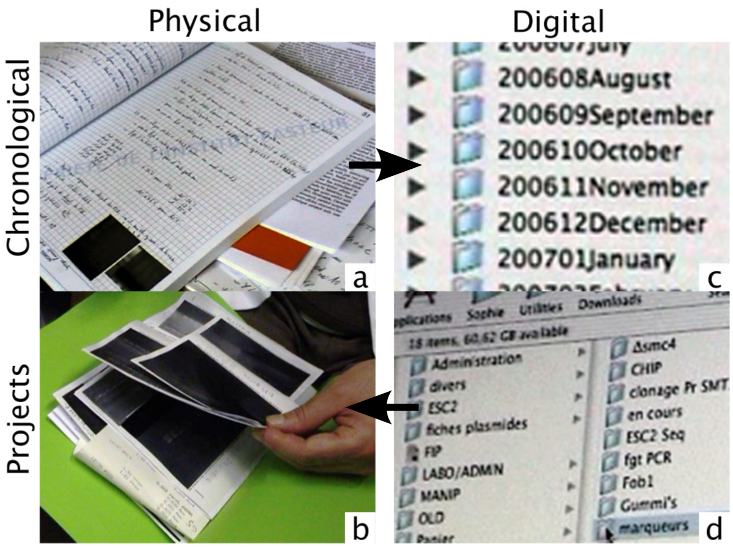


Figure 18: Organization schemes between the physical and digital space.

When Biologists collaborate, they rarely work on the same projects at the exact same time which leads them to organize information in a hierarchical way. Julie, Victor and Sabine rely on agreed upon folder hierarchies, when they work with colleagues. Figure 19 represents the folder hierarchy on a shared computer in Sabine's laboratory. With her colleagues, Sabine defined a hierarchical structure they all follow. The hierarchy starts with generic names of people or projects, the sub-folders are organized by data type, each containing projects sub-folders.

/ User name / Data type (e.g. lab, pdf, personal, references, sequences) / Project / Files

Figure 19: Folder hierarchy on a shared computer.

Shared structures force Biologists to manage information both at a group level and at a personal level. This leads them to create links from their notebooks to information they want to follow on computers. Participants created semantic codes to organize information across media. For example, Julie uses colors to encode information between her notebooks and folders, pink corresponds to a specific database work, yellow to another one.

13 "C'est aussi le problème des cahiers de manip, quand on plusieurs manips en parallèle soit on a trois cahiers de manips soit un seul mais tout ne va pas à la même vitesse et c'est un problème."

14 "La génétique c'est long, pour caractériser un gène ça peut prendre 5 à 10 ans [...] Je suis en train d'écrire deux papiers qui finalisent 4 ans de travail."

3.4 DISCUSSION

The field work at the Institut Pasteur provided two main benefits. First, it led us to better understand Biologists' information management practices and how they reflect on their activity. Second, we could identify problems Biologists face when they manage personal information, which helped us to draw implications for the design of new notebooks.

3.4.1 *Reflective practice*

Biologists use a variety of tools to manage their personal information. They filter, transform, save and organize information in notebooks but also in the laboratory space or on their computers. As they manage information, they must make sense of their activity, the tools they use help them reflect on their activity but also connect information spread across different media. To handle the diversity of information, Biologists also adapt their notebooks to better fit their needs and support their reflective practice.

Hypomnemata, tools supporting reflection

The study highlighted the wide set of tools Biologists use to manage information, from paper and electronic notebooks, to physical and digital folders. Biologists do not use these tools only to record their activity, but also to process information and make sense of their activity. Notebooks, in particular, let Biologists save information but also filter it and transform it. They are hypomnemata (Stiegler, 2001): tools supporting reflection whose *"role is to digest information, to let it go to intelligence, not memory."* (Seneca, 2002).

Unlike lifelogs or memory prostheses, the role of hypomnemata is not to supplement human memory but rather to help users select and synthesize the information they handle. When Biologists save information, they do not intend to provide an extensive account of their activity. They save what they consider to be important for the future or what they should sort out. Notebooks support the reflective practice of Biologists at two levels. They allow Biologists to *reflect-on-action* as Schön (1984) defined it. When Biologist look at their official notebooks, they face their past activities and the information they gathered and try to make sense of it. When they select and organize the information they save, they *reflect-in-action*, as they must articulate the elements of information they save.

Articulate information

When Biologists save information they transform a continuous activity into a discrete account of it. Saving information acts as a filtering process, leading Biologists to select elements of interest and discard the rest.

In this context, writing is not a simple translation of researchers' activity on paper, it forces them to define what they are working on. As they save information, they consider to which project it is related to, where to save it, what they already wrote and is not worth re-writing or what would be useful in the future. They articulate what they are doing to keep the information captured meaningful. Writing on paper or deciding what to save on the computer pushes Biologists to reflect on the information they manage.

The intent to make sense of ones' activity may explain why paper notebooks are still popular. Beyond the convenient properties of paper (reactiveness, mobility, robustness, openess), it is rather the discipline that paper notebooks impose that makes them useful by inviting researchers to write with care.

Adaptation

As Biologists' information gets distributed over many different people, locations and media, Biologists struggle with how best to structure what they save, either to find relevant information in the future or to help colleagues understand shared data.

In order to maintain a coherent information space, Biologists are changing with their environment. They explore new tools to capture information, such as alternative electronic notebooks, looking for the one that fits their needs the best. In order to sort out the complex inter-relationships among notebooks, we observed many strategies to customize or adapt notebooks. Some researchers preferred using unofficial notebooks, maintaining several notebooks in parallel, or adapting them more freely. They added layers of information inside their notebooks using color codes, physical notes and written annotations.

The way Biologists adapt their notebooks influence how they save information. Delphine who dedicates each notebook to different activities, does not save information in the same way as Carole who created color codes to link information among her notebooks.

Reflective framework

Biologists explore hypotheses through experimentations with iterated and refined protocols, new data, new analysis methods and new tools. Along the line, they constantly adapt their research and the tools they use to the problems they face. They are not only recording information for future use or as a proof of precedence; the way they handle their familiar information influences them. Selecting, capturing and writing down an account of activity forces researchers to look back at their activity and frame it in the perspective of their ongoing research.

Figure 20 describes the interaction between Biologists' research, the way they record their activity and manage their information, the tools they use to do so and how they adapted them. This corresponds to the selective coding phase of the Grounded Theory, each arrow of the figure is explained below.

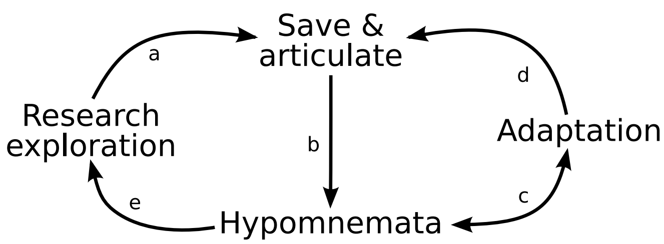


Figure 20: Managing familiar information, a reflective practice.

a. In the course of their *exploratory practice*, Biologists select information they want to save.

b. As Biologists save the information they selected, they must *articulate* their activity. They create an account of activity that connects the significant decisions, interesting snippets, modifications of protocols (or scripts), with their underlying explanations.

c. Biologists adapt their *hypomnemata* (notebooks, but also online applications), to the ever increasing wealth of information. We observed how they use personal codes in their notebooks, but also the new digital tools they create or adapt to handle their information online.

d. The way they *adapt* their hypomnemata influence how they filter, save and articulate information.

e. The information Biologists save influences their research orientations as they identify limits, repetitions or leverage colleagues' work.

To cope with the evolution of their practice and information spread over different media, researchers reflect on their activity. This reflective practice does not only happen after the fact, with Biologists looking back at their notes and trying to figure out alternatives. Reflection happens as they manage information everyday, when they decide that an event is worth reporting or when they change a folder hierarchy to better reflect the ongoing state of a project. In order to support this lightweight reflection we must design tools that help users save and articulate familiar information while staying out of their way.

3.4.2 *Implications for design*

This first study provided insights on what types of personal information management tools we could build for Biologists. I list here five key observations.

- Notebooks should let Biologists aggregate the *heterogenous information* they manage, so that they can use the notebook the most appropriate to the information they save.
- The capture should be very lightweight, support copy and paste in order to let users capture information from the sources they want.
- Notebooks should *invite users to write and articulate their activity*, not only log information coming from different sources.
- Notebooks should *support both chronological and project-oriented categorization* of information, so that Biologists can retrieve information from past notebook in the most efficient way.
- Notebooks should allow users to *re-use information*, by copy and paste or more evolved mechanisms.

3.5 CONCLUSION

Based on interviews of Biologists from the Institut Pasteur, we identified the various tools Biologists use to manage familiar information, from notebooks to computer files. These tools do not only serve to archive information, deciding what to save forces researchers to reflect on their activity: they filter, process and reframe the information they manage.

The next chapter describes the participatory design of a hybrid notebook, that takes into account our observations at the Institut Pasteur and complement them with observations of Bioinformaticians. The field study presented in this chapter and the participatory design study led to Prism, a hybrid notebook, presented in Chapter 5.

3.6 SYNTHESIS

The contributions of this chapter are:

- Comparison of the characteristics and use of paper and digital notebooks, of their respective flexibilities and constraints.
- Emphasizing the role of filtering and selecting when Biologists manage familiar information.
- Implications for the design of personal information management systems based on notebooks.
- Providing a minimal framework for analysing personal information management as a reflective practice for Biologists.

4 PARTICIPATORY DESIGN OF A HYBRID NOTEBOOK

This chapter presents the participatory design study of Prism, a hybrid (paper+digital) notebook. Based on observations of a team of Bioinformaticians and participatory design workshops, the chapter presents implications for the design of Prism.

4.1 INTRODUCTION

Based on the insights from the study at the Institut Pasteur, we focused on designing a hybrid notebook that would let Biologists filter, save and synthesize familiar information. As we conducted a first workshop at the Institut Pasteur, we realized that it would be difficult to experiment with Biologists' notebooks. The laboratory notebooks are an important part of the research process at the Institut Pasteur and cannot be modified lightly. A former researcher from the Institut Pasteur, who participated to this initial workshop, introduced us to her team at the French agricultural research institute (INRA), where she worked with Bioinformaticians.

The shift from the Institut Pasteur to INRA provided several benefits: first, INRA was less strict about notebooks; second, we were introduced by a member of the team which facilitated initial contacts; third, we could work with a team rather than with individuals (like we did at the Institut Pasteur); finally, managers were interested in better understanding their own information management practices and thus willing to explore new tools.

We conducted further interviews to understand Bioinformaticians practices, and worked with them in a participatory way on the design of Prism, a hybrid notebook which integrates personal information from notebooks and the computer.

4.2 A PARTICIPATORY DESIGN APPROACH

The participatory design study was motivated by what Bowker et al. (1997) describe as the *"great divide"* between the designers building tools and the social scientists analyzing their use. On the one hand, social scientists can study systems already used by organizations but may be reluctant to draw implications for design from their observations (Dourish, 2006). On the other hand, designers and developers create new systems which provide technical insights but may be hard to evaluate with users on the long term.

For designers, testing design ideas in workshops can help to generate systems suited to users and their situation of use, yet it is clearly no substitute for observing users in real situations. Greenbaum and Kyng (1991, p. 4) identify specifically four patterns that must be taken into account while designing for the workplace:

- *"the need for designers to take work practice seriously;*

- *the fact that we are dealing with human actors, rather than cut-and-dried human factors;*
- *the idea that work tasks must be seen within their context and are therefore situated actions;*
- *the recognition that work is fundamentally social, involving extensive cooperation and communication."*

These patterns come from the observation that good systems cannot be built by experts having a limited and idealized knowledge of users' work practices. Interviews are not enough to uncover the complex use of tools in the workplace but also workers' tacit knowledge. Many practices go unsaid and are even hard to put into words, like intuitions of *"looking right"* that can be found in most work places. For example, Biologists are able to recognize an *interesting* protein sequence at a glance, but have a hard time explaining why.

4.2.1 *Participatory Design*

Participatory design allows designers to bridge the gap between theoretical knowledge and work practices by letting users become actors during the various stages of the design process: from idea generation, critique, and evaluation. It provides an approach based on cooperative action rather than formal description, by insisting on:

- *"mutual learning between users and designers about their respective fields;*
- *use tools in the design process which are familiar to users;*
- *envisionment of future work situations to experience how emerging design may affect work practice rather than relying on seemingly esoteric language of system developers; and*
- *the importance of starting the design process in the practice of users."* (Greenbaum and Kyng, 1991, p. 5)

The goal of participatory design is not only to gather ideas and feedback from users, but also to reflect on what they do, to better understand the design context, and various aspects of the users' needs and desires which are unknown to the designers (Mumford, 1983). However, users do not become designers: they hold knowledge about their work while designers hold knowledge about design possibilities. Participatory design allows us to build a shared place where design communication can happen.

4.2.2 *Technology probes*

Participatory design workshops engage participants in creative activities in which they are invited to generate new ideas. Fostering this creativity requires designers to balance continuity of existing practices and innovation.

On the one hand, participants' imagination can be hindered by the tools they are actually using. On the other hand participants can fantasize about systems, while ignoring unexplored areas of the design space. Technology probes are systems designed to adjust the discussion between designers and users around concrete yet novel experiences. Hutchinson et al. (2003) describe them as *"simple, flexible, adaptable technologies with three interdisciplinary goals: the social science goal of understanding the needs and desires of users in a real-world setting, the engineering goal of field-testing the technology, and the design goal of inspiring users and researchers to think about new technologies"*. The participatory design of Prism presented in this chapter led to its implementation and longitudinal use as a technology probe, presented in the following chapter.

4.3 PARTICIPATORY DESIGN STUDY

We used a participatory design process that consisted of interviews, workshops (evaluation, brainstorming and prototyping) (Mackay, 2002) and the longitudinal use of a Technology Probe. The goal was to design a notebook that would help users manage familiar information and support reflection.

4.3.1 *Setting: URGI team at INRA*

INRA is one of the top agricultural institutes in Europe. 3500 researchers and students work at INRA, and a staff of 6000 engineers and technicians manage the experimental facilities and technology transfer. INRA campuses are spread all over France, mostly in non-urban areas. We worked with a team of Bioinformaticians located in Paris suburbs (Evry).

We focused on a specific team, the research unit in genomics and bioinformatics (URGI), and studied more closely how Bioinformaticians worked together. URGI is a laboratory dedicated to the study of plants and crops parasites, with an interest in genome dynamics. The team does not work with plants directly (there is no "bench work"), but rather runs experiments and analysis on the computer. They develop tools to help Biologists manage and leverage their data, host a bioinformatics platform and offers services including database design, data integration, genomics annotation and support to bioinformatics analysis.

In 2007, while doing the interviews, INRA coordinated a general effort to improve quality management[1]. This effort led researchers to reflect on their personal information management practices.

4.3.2 *Participants*

The research unit in genomics and bioinformatics (URGI) consists of both researchers and engineers. Most have a background in biology while their current interests lie in bioinformatics. Even though there is no bench work at URGI, participants work on biology related topics and improve the knowledge in this field. Still, they all share an expertise in computers science. They maintain databases and develop tools (mostly web platforms) allowing Biologists or Bioinformaticians to store, retrieve, and combine data in order to extract new knowledge.

1 Quality management is the set of management processess designed to improve the quality of care of services within an organization.

The participants hold diverse positions (Table 2): managers (2), researchers (4), post-docs(2) and engineers(4). Depending on their position in the team, members had different concerns and behaviors regarding information management. Whereas managers have a more global view of ongoing activities, researchers and engineers focus on ongoing projects and their short term future. They all have one or two computers, a Sun client to work on the servers, an Apple or Windows laptop for productivity activities (email, writing or presentations). All share information and have common concerns on how best to organize it.

4.3.3 *Procedure*

We interviewed eight Bioinformaticians and conducted five participatory workshops over the duration of the project (figure 21). We held the first workshop (W1) at the Institut Pasteur with three Biologists, a Bioinformatician (from INRA) and an archivist. The later workshops (W2-5) took place at INRA with between six and 12 participants. All participants were potential users of Prism, except the archivist, who was interested in the implications of such tools for institutional archiving.

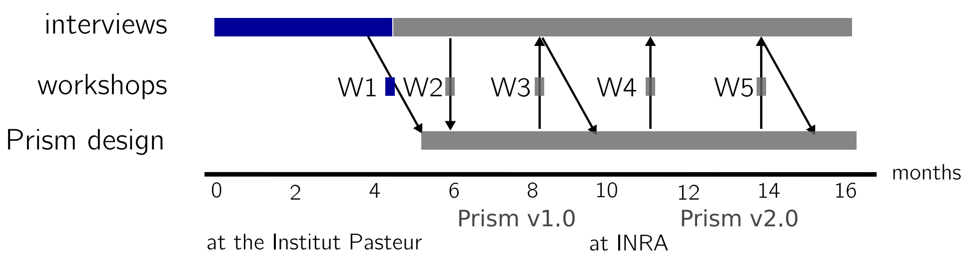


Figure 21: Evolution of the Prism project.

We assessed the conclusions of our observations in the first two workshops we conducted (W1 and W2), where we presented our conclusions from the interviews, scenarios of use and video prototypes. In the meantime we started to develop and study Prism as a Technology Probe (described in the next chapter). Prism is a hybrid notebook which integrates information streams. With Prism, users can browse and filter information they saved on their computer, on the Web or wrote in their paper and electronic notebooks. During subsequent workshops (W3, W4 and W5) we explored specific design issues raised through field use, which informed the design of Prism.

The primary goal of the interviews and the workshops was to identify design directions for notebooks that would support Bioinformaticians. I focused on identifying the similarities with the Biologists from the Institut Pasteur but also the differences. I then reviewed the field data (from interviews and workshops) using the iterative coding of Grounded Theory. This analysis complemented the study at the Institut Pasteur and revealed the relationship between the management of information at the individual and team levels.

4.4 RESULTS

This section presents the results from the preliminary field work at INRA, with a focus on personal and group information management. I first present the *interactions between the notebooks* used at INRA, then the difficulties Bioinformaticians face when they *collaborate and share information* and finally the problems they face when *organizing information for the group*.

4.4.1 *Interactions between notebooks*

Five out of eight participants used paper notebooks (see table 2), and five used some form of electronic notebooks. Like at the Institut Pasteur, there is no well-defined electronic notebook. Interestingly, the Bioinformaticians doing mostly research used paper notebooks (first five participants), whereas the ones who were mostly doing engineering work used digital notes (last four participants).

Participant	Paper Notebook	Electronic Notebook
Guillaume (Manager) R	Yes	No
Camille (Manager) R	Yes	No
Nadia (Post doc) R	Yes	HTML NB
Sarah (Post doc) R	Yes	HTML NB
Adeline R/Eng.	Yes	text files
Jeanne (Senior Eng.)	No	text file
Dominique (Senior Eng.)	No	text file
Pierre (Eng. - Sys Admin)	No	No

Table 2: Laboratory Notebooks' use at INRA (R: researcher, Eng.: engineer).

Participants used *experiment notebooks*, which are between scratch pads and officially registered notebooks. Since most Bioinformaticians' work happens online, they switch from one medium to another and maintain links between them. The paper notebooks contain references to URLs, emails, files (Figure 23), command or program names or even complete printouts of electronic notes written during a day (Figure 22).

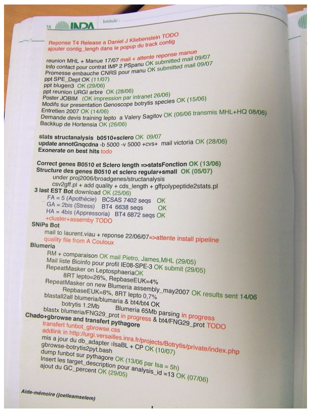


Figure 22: An electronic note pasted in a paper notebook.

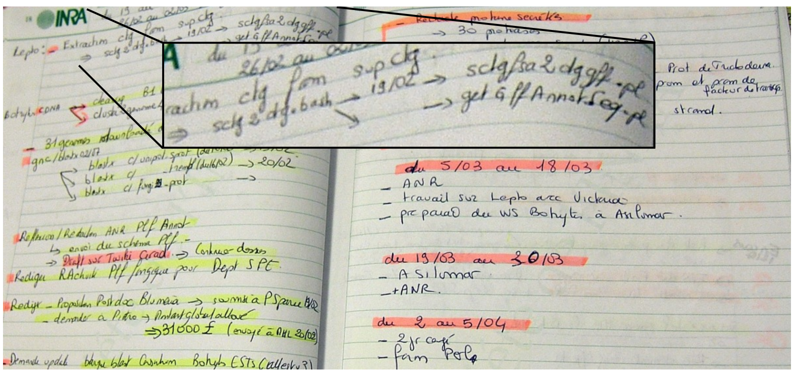


Figure 23: A Bioinformatician's paper notebook with references to the following files: sclg2dg.bash, sclgfsa2dggff.pl, getGffAnnotSeq.pl.

Adeline and Jeanne used text files to capture their ongoing activities on the computer, copying and pasting snippets that could be useful in the future. Nadia and Sarah used HTML notebooks similar to the ones of the Institut Pasteur described earlier.

The two post-docs (Nadia and Sarah), started to use an HTML notebook when they were working at the Institut Pasteur, and kept using it in their new team. However, their uses were different. Nadia used her e-notebook as a tool to share her activity with her manager. Sarah used it as a tool to capture her ongoing activities on the computer. Both also used paper notebooks, but in this case, their use of e-notebooks shaped how they used their paper notebooks. For Nadia, her paper notebook served to capture *not yet* shareable work. For Sarah, her paper notebook was more dedicated to thinking about her research and organizing thoughts (not capturing information).

Digital and paper notebooks are mostly personal. Some Bioinformaticians (Nadia or Adeline's student) use them to share accounts of their activity with someone else (an advisor or a very close collaborator), but they are not targeted toward the group.

4.4.2 *Personal and communal digital space*

For keeping and sharing information and procedures within the group, Bioinformaticians use different kinds of tools depending upon the information they want to share (figure 24). Paper and electronic notebooks are mostly personal (left part of the figure) even though Bioinformaticians can share specific notes with others. They share official documents and articles by email or print them on paper. The team hosts online information management tools (botton right) such as JIRA (for project management) or Alfresco (for document management). These applications do not hold personal information but are targeted to the group. On their main server, the team relies on README files to communicate with their colleagues.

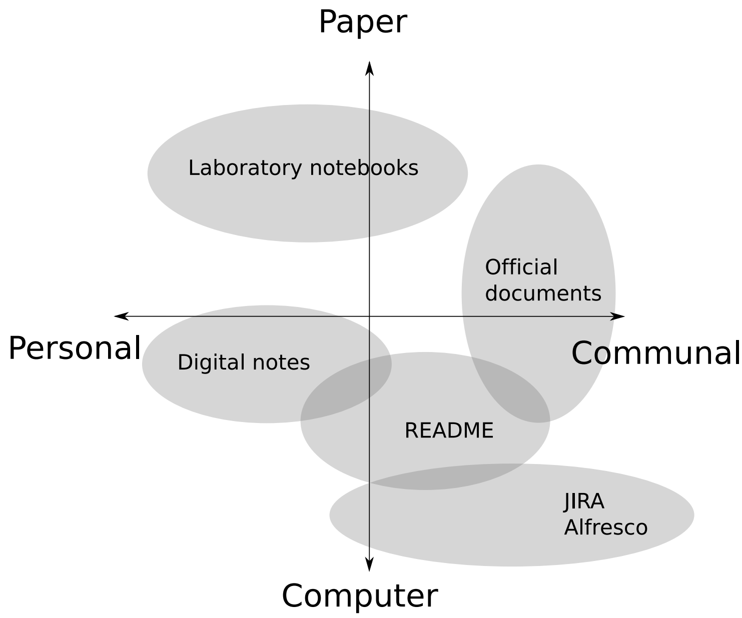


Figure 24: Information management tools used at URGI.

README files include comments on how to use a program but also change-logs[2] in case the programs or scripts are developed locally. They have one main characteristic that differentiates them from text files or e-notebooks: they are located with the content they are about. While it makes it easier to locate information while performing a task, it becomes harder for Bioinformaticians to follow their activity across the different files they edit in the course of their research. To organize information spread in different README files, participants such as Jeanne used a personal README file. This file acted as a chronological notebook, pointing to files she modified recently with personal comments. Her personal README file became as personal as a notebook. She placed it in her home directory and used UNIX file permissions[3] to be forbid others in the group to read or write in this file.

The team uses JIRA[4] to share information about their development activities. JIRA tracks bugs, issues and tasks to help developers to manage projects. Even though not all Bioinformaticians participate, it is central to the team's big development initiatives. Dominique and Nadia explained how they rely on the information held in JIRA (tasks to perform, blocking issues, comments about a problem) to perform their own work. The information in JIRA is not personal per se, since it is shared with colleagues. However, it is familiar to Bioinformaticians who need to know what others did recently to pursue their own development and research.

As we were working with the team, participants started to use Alfresco[5] for document management, after experimenting with WIKIs. Because information is spread over many different platforms (servers, project management systems and document management systems), the team has problems organizing the information in a way that makes sense for each person individually and for the group.

4.4.3 *Dynamic information and static structures*

Participants experimented with strategies to organize information in a way that made sense to them but also to their colleagues. As Guillaume explained, the chronological organization of notebooks makes sense when working alone on a few projects. However, participants worked on different projects in parallel, for different duration and with different collaborators. It made more sense for them to organize the information they share by projects.

Guillaume and his colleagues tried to build a coherent structure and minimize subsequent changes across their different information systems. However, project hierarchies are difficult to organize in advance since participants figure out the structure of their projects as they handle information.

Figure 25 describes the evolution of the conceptual hierarchy of projects on which Adeline worked. In genomics, researchers can work on different species sharing similarities. However, at the beginning of a project, the similarities are not always visible. In this case, two projects

2 A change-log is a log or record of changes made to a project, such as a website or software project.

3 There are three types of permissions on UNIX (what can be done with a file): read, write and execute. Permissions are defined for three types of users: the owner of the file, the group that the owner belongs to and others.

4 http://www.jira.com/

5 http://www.alfresco.com/

(*Lepido* and *Puceron* at the top of the figure) started separately. As Adeline worked on them, correspondences appeared. Parts of both projects had to be merged to avoid duplication, share scripts and common information that was necessary to both projects. As these projects went on, further refinements had to be done for new projects (*Abeille*) or to deepen in a common direction (the now joint *Lepido Puceron*).

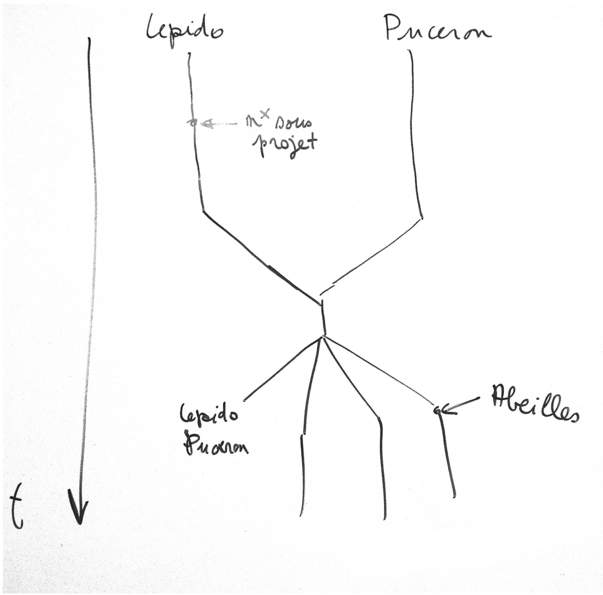


Figure 25: Evolution of projects hierarchies.

Hierarchies cannot change as quickly as researchers make sense of the data they manage. Team members leverage the fixed hierarchies to retrieve information and decide where to file information they created. When the hierarchies are reorganized, it becomes more difficult to know where to file data and some information gets lost. For example, Jeanne or Adeline could not find specific README files they knew existed but were lost in a re-organization.

The dynamic aspect of information explains why researchers are always looking for new tools (bug trackers, document managers, web 2.0 applications) and strategies to cope with the static organization of their shared files. Notebooks let them annotate and organize the information shared with the group in a personal way. Guillaume, Camille and Nadia used paper notebooks, Sarah her HTML notebooks and Jeanne her personal README file to this end.

Researchers want to become familiar with the information they manage. They want to retrieve information, but also put it in context in order to draw connections and make sense of their activity.

4.5 PARTICIPATORY DESIGN WORKSHOPS

The interviews at the Institut Pasteur and INRA modified our initial view of a mixed paper and electronic notebook. In the participatory design workshops, we focused on the design of Prism, a hybrid system that would not only integrate equally paper and electronic notebooks, but incorporate paper notebooks with the wealth of digital activities and information scientists manage on their computers.

We refined Prism through two main series of participatory design activities (see figure 26). We held a first series of workshops with potential users of Prism, to assess the understanding coming from the interviews (W1 and W2, top of the figure). We then held workshops with users of Prism (W3, W4 and W5, lower part of the figure) to explore particular design ideas through brainstorming and video prototyping.

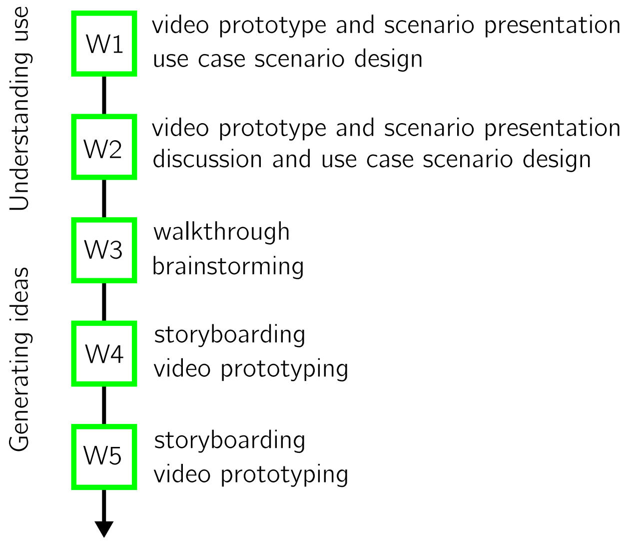


Figure 26: Participatory design process.

For three years I also co-taught a participatory design course for Biologists. In 2007, the course focused on the design of electronic notebooks. Students created three video prototypes that explored how to let users quickly input information in e-notebooks, how to let users create rich entries (mixing text, schematics and drawing) in e-notebooks and how to incorporate social features in notebooks?

Validating observations

During the first two workshops, our goal was to assess our understanding of the interviews and observations. We summarized our findings from the interviews and presented a video prototype, a combination of rapid prototyping on paper with Wizard of Oz video (Mackay, 2002). The video prototype relied on a scenario of use synthesized from the interviews and showed some technological directions we were investigating:

- parallel streams to integrate different sources of familiar information together,
- lightweight capture of information on the desktop,
- navigation and filtering of notebooks entries through keywords or temporal criteria.

Participants created personal scenarios based on what seemed relevant in our scenario of use and recent critical incidents (breakdowns in the way they handle information or use notebooks). The scenarios emphasized the problems of associating digital data belonging to different hierarchies, as well as problems of awareness about digital activities.

Design sketches

In the second series of workshops, we explored design ideas. We dedicated an initial workshop (W3) to idea generation. We presented the results from the field study and the first workshops and brainstormed with participants on how to support reflection, information aggregation, capture mechanisms or organization alternatives.

Ideas centered on how best to track personal activities from different sources. Participants explored alternative mechanisms for linking information from the web tools they were using to their notebooks. Some participants envisioned linking paper notes to the online tools and vice versa, while others seemed more interested in the integration of web information directly into the digital notebooks.

In subsequent workshops (W4, W5), we asked participants to turn specific problems brainstormed in the first phase into scenarios of use and to create video prototypes that illustrated how they would like to address the problems.

A first group explored methods to manipulate written information on paper notebooks. Based on their existing experience of notebooks and symbols, they imagined semantic symbols to modify Prism's treatment of written notes. Participants sketched visualizations of their activity, based on the meta-data created by text annotations. Participants discussed these visualizations as they were interested in gaining an awareness of their activity and sharing it, but not confortable with the monitoring aspect of it.

Since the team used many web application to collaborate, a second group explored the integration of web information within Prism. The group prototyped techniques to avoid writing the same information twice, for example in the bug tracker and in the e-notebook. Figure 27 shows the participants prototyping mechanisms to link web snippets to the e-notebook. They imagined a browser extension that would directly extract selected information from collaborative sites and send it to the notebook.

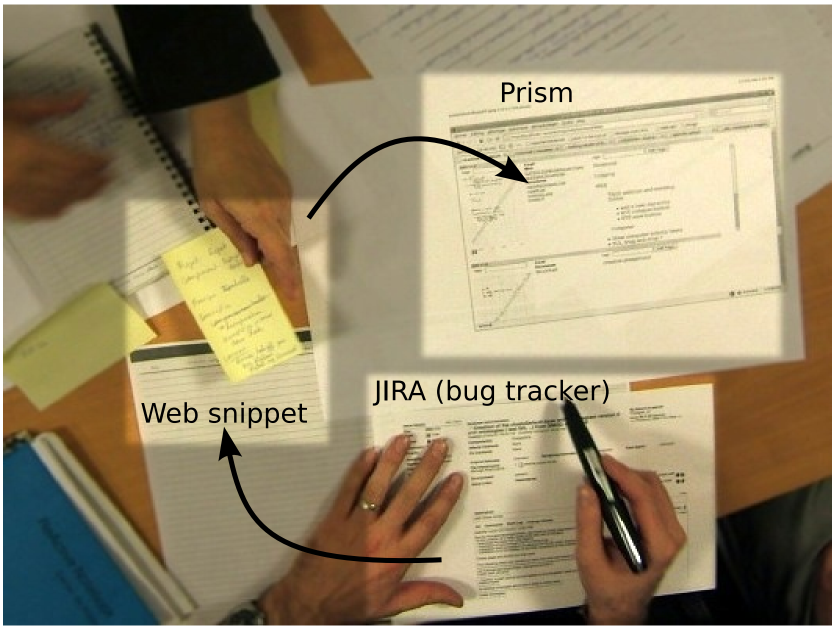


Figure 27: Prototyping linking mechanisms

4.6 DISCUSSION

The interviews highlighted three points we further explored in the workshops: the interaction between paper and digital notebooks (and digital information), the distribution of information over personal and communal spaces, the balance to find between dynamic information and stable structures. Participants were able to sketch adaptations of their existing notebooks, but also discuss their use of Prism hybrid notebook and explore how to make it evolve. The workshops allowed us to prioritize features in Prism and sort out what was doable, given time and development constraints.

4.6.1 *Differences between the Institut Pasteur and INRA*

At the Institut Pasteur, we interviewed individuals about their personal information management practices. At INRA, we observed how individuals manage familiar information within a team. At INRA, the Bioinformaticians are not preoccupied by proving the precedence of their work nor do they patent discoveries. They are interested in keeping the knowledge within the team. This is particularly important for them as software projects take years to build and must then be maintained.

At the Institut Pasteur, Biologists work both at the bench and on their computers. These different contexts influence how they use their paper and electronic notebooks (for example, Delphine wrote URLs in her paper notebook, rather than her e-notebook, when the URLs were related to a project described in her paper laboratory notebook.) At INRA, Bioinformaticians' work happens online. They use a wide variety of tools to manage information, from notebooks to web applications. Their digital activities influence their use of paper notebooks. What is difficult for them is to connect the different bits saved in different locations.

4.6.2 *Information ecology*

The field work at INRA led us to identify the continuum from personal information management to managing information with colleagues. The Bioinformaticians navigate between different machines and information management systems. They do not only rely on their notebooks but also on an information space composed of README files, emails and web applications.

In this context, they use their notebooks to save personal information but also to organize remote information in a personal way that makes sense to them. We could observe many inter-relationships among the tools Bioinformaticians use to manage information. This becomes a problem as information evolves over time, with hierarchies that must be reorganized and colleagues modifying information that was initially created by others.

4.6.3 *Communication*

Participants shared a common information space to communicate with their colleagues. Whereas at the Institut Pasteur information could be shared informally in the laboratory, at INRA, Bioinformaticians used many tools to share information online, such as READMEs on servers or web applications. Participants were particularly concerned with how to keep the knowledge within the team as new members come and others go. Indeed, many team members who primarily hold information are not permanent in the team. (Nadia left the team during the study, at the end of her one-year contract.)

Bioinformaticians are not only interested in their personal information but also in knowing if their colleagues are working on something that could have implications for themselves. Students and participants explored in workshops how to gain an awareness of the ongoing activities in the laboratory that could have implications for them.

Revisiting the reflective framework

The study at INRA confirmed the reflective practice of researchers as they manage information. Bioinformaticians use many tools to manage information from notebooks to online information systems. When they define project hierarchies or decide to add notes to a README file, they filter information and articulate it. They also adapt the tools they use to manage information, creating codes to follow information from one tool to another in a personal way.

Figure 28 revisits the reflective framework defined in Chapter 3. The new arrows describing the relationships among categories are described below.

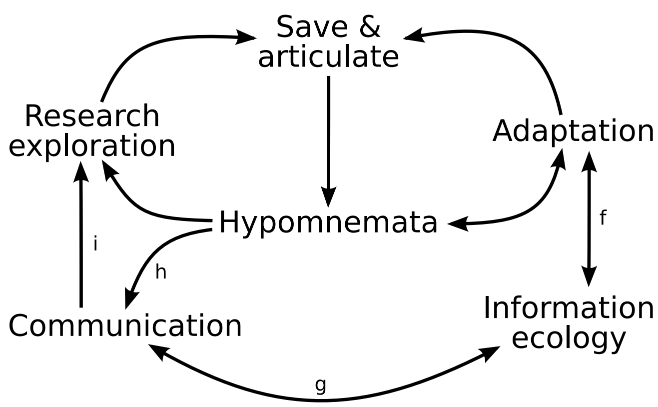


Figure 28: Managing information, a reflective practice.

f. Throughout their activity, researchers leverage their context to manage information, they adapt it to facilitate their ongoing tasks. Bioinformaticians re-organize folder hierarchies on shared servers to match the current state of a project.

g. Since the environment is shared within the team, it helps researchers *communicate* their activity and share information; Bioinformaticians share information on servers, the state of the bench provides information about Biologists' activity.

h. Bioinformaticians communicate with their colleagues through the tools they use, notebooks, README files or web applications.

i. In return researchers influence the *research* of their colleagues, through collaborations, comments or the information they share.

4.6.4 *Design implications*

The field work at the Institut Pasteur and the initial workshops influenced the first version of Prism. The participatory work at INRA led us to revisit Prism's initial design in new directions.

- Prism should let Bioinformaticians *integrate* the different streams of information they manage. This involves integrating paper notebooks and electronic notebooks but also digital information such as the Web applications that participants use.
- Prism should let users integrate colleagues' activities when they are relevant to the information they manage. Nevertheless, colleagues should be able to control the information they share.
- Prism should support a balance between the personal and communal aspects of information. Users should be able to organize shared information in a personal way.
- Prism should let users re-organize information they handle as they better understand their research problems.
- Prism should let users *create connections among the streams* of information they manage.

As participants started to use Prism or other digital notebooks, they faced shortcomings but also started to customize their notebooks to support their specific use. It is by using Prism as a technology probe that we could ground the workshops into concrete designs and realistic scenarios of use.

4.7 CONCLUSION

This chapter presented the participatory study that took place as Bioinformaticians tried Prism, a hybrid notebook integrating paper and digital information. I described the diverse interactions between paper and electronic notebooks, the continuity between personal information and information shared within the group, and how Bioinformaticians must balance personal views on information with the static organization of distributed information.

In the next chapter, I present the implementation and the results from the longitudinal use of Prism. By creating a tool participants could try early in the design process, we could narrow down the interviews and let participants gain an understanding of what a hybrid notebook could be.

4.8 SYNTHESIS

The contributions of this chapter are:

- Augmenting the reflective framework with the influence of the environment and communication.
- Description of the continuum between personal and communal information
- Description of the interactions between the different notebooks.
- Description of the dynamic nature of information confronted to the static organization structures.
- Implications for the design of hybrid notebooks.

5 PRISM AND MASTER NOTEBOOKS: A PLACE FOR REFLECTION

This chapter presents the implementation and longitudinal study of Prism: a hybrid notebook designed to support researchers' reflection. Prism relies on streams to integrate the diverse sources of information familiar to users: paper and digital notebooks as well as elements of their computer activity. Prism led users to create master notebooks, from all the information sources available they dedicated a particular notebook to synthesize and organize information.

The results from the field studies highlighted not only the reflective nature of researchers' practice but also the need for personal information management tools that specifically support their reflection. Using the participatory design approach described in the previous chapter, we[1] designed and tested Prism in collaboration with Bioinformaticians from INRA. The field work emphasized four key properties that Prism should support:

- Integration of researchers' paper and digital notebooks.
- Support for lightweight and selective capture in context.
- Balance of chronological and project-oriented organizations.
- Provision for open and shareable notebooks formats.

As we worked with users, Prism progressively evolved from a desktop application, to a Web-based application aggregating heterogeneous streams (Figure 29): Anoto paper notebook, e-notebook, desktop activity (email messages, websites and documents), and distributed activity (Web feeds or colleagues' streams).

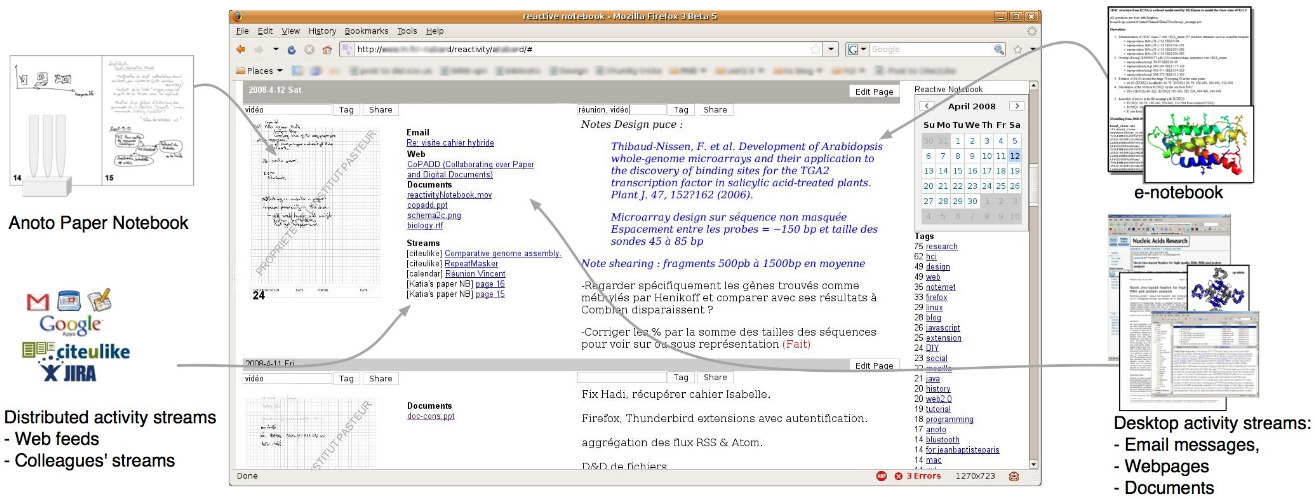


Figure 29: Streams available in Prism v2.0

1 Evelyn Eastmond participated to the participatory design and development of Prism v1.0

5.1 PRISM, A TECHNOLOGY PROBE STUDY

The goal of the participatory work was not only to design a tool suited to its users. We were also interested in studying the implications of Prism on users' activities and on the way they managed information. As part of the participatory design process, five users volunteered to experiment Prism for nine months. In this context, Prism acted as a Technology Probe (Hutchinson et al., 2003), an open tool built to explore design concerns appearing through field use.

Technology probes combine perspectives from social science, design and engineering. From a social science perspective, they allow researchers to collect information about use in a real-world setting. From a design perspective, they should inspire users and designers. By giving participants an open tool they use on a daily basis, they get a better understanding of the implications of its use and of the technological possibilities. From an engineering perspective, Technology Probes allow developers to verify if the architecture is appropriate, refine data models and test the tool in a realistic setting

Selecting a longitudinal and ecological approach had implications on the design of Prism. Modifying the tools Biologists or Bioinformaticians use to manage information is sensitive. Such tools are at the center of their activity, they represent hours of work and concentrate efforts to make sense of their research. Prism had to be reliable enough so that participants could try it without danger, but open enough so that it could evolve over time.

Setting

We tested Prism at INRA with participants of the participatory design study (see section 4.3). The URGI team did not have the same constraints as the Institut Pasteur did, which motivated our participatory work with them. They were also open to experiment with new forms of personal information management.

After an initial workshop at INRA, we installed Prism on the volunteers' main computer. The install lasted one hour including the set-up of the Anoto technology and a demonstration of the creation of one element for each stream (one paper note, one digital note, one web page, one email message and one document saved). For updating Prism, participants had to download a JAR[2] file and replace the existing one in their application folder.

Participants

Five bioinformaticians from INRA, who participated to the workshops presented in the previous chapter, volunteered to try Prism. Two were managers, two were senior bioinformaticians, and one was junior researcher. They all had an initial training in biology and additional training in computer science when their interest shifted to bioinformatics. Four out of the five participants already used paper notebooks before beginning the study.

2 A Java Archive (JAR) file bundles multiple files into a single archive file. Typically, a JAR contains the files necessary to run an application or a library (`http://java.sun.com`).

Procedure

Participants tried Prism for nine months, from August 2007 to April 2008. Prism evolved during this time in response to user's comments, with daily, weekly or monthly updates. This led to different types of use, depending on the stability of the new components and the overall system. Although Prism was not static, it was real, which gave participants sufficient experience to reflect about their use over a significant period of time.

When participants raised issues or when we observed that most of them shared common problems, we organized the workshops presented earlier to investigate design alternatives. This engagement in the design process made them more tolerant of the necessary exploratory status of the probe: it became a feature allowing them to control the design space.

During field use, we analyzed participants' notebooks, which parts they used and what they looked at. We also observed their on-going activities in the laboratory. We discussed their use throughout the iterative design process, in person and via email. Combined with the workshops it renewed participants' interest throughout the study.

5.2 PRISM FIRST IMPLEMENTATION

The initial Prism prototype was a desktop JAVA application (Figure 30). This first version focused on building a usable solution allowing users to quickly capture information. Prism integrated three streams of personal information (i.e. sets of information elements with associated timestamps): a paper laboratory notebook, an e-notebook and elements of computer activity (email, web pages and documents). We used the Anoto technology to capture paper notes, and developed addons for Firefox and Thunderbird, users could capture their activity on any platforms: Linux, Mac OS X and Windows. However, only Windows users had access to the Anoto paper notebook, since pen drivers were only available for that platform.

Paper Notebook

The paper notebook is an ordinary notebook, printed with the Anoto-pattern, a tiny dot pattern which encodes position on paper sheets, the page number and a notebook identifier. The dot pattern can uniquely encode around 70 trillion sheets of A4 sized paper. Users write as they would with an ordinary pen, but a micro-camera integrated into the Anoto-enabled pen records its movements. The resulting gestures are captured and stored on-line.

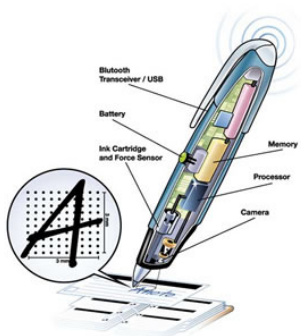


Figure 31: Anoto digital pen and the paper dot pattern.

We used Nokia SU-1B[3] and Logitech io2[4] pens for the study, both Anoto-enabled. As users leave ink on Anoto paper, the pen stores every stroke, its position along with a timestamp. When a user is ready to upload the handwritten data, he connects the pen to a USB cradle and the data is uploaded to his computer.

3 http://europe.nokia.com/A4471253
4 http://www.logitech.com/index.cfm/66/459

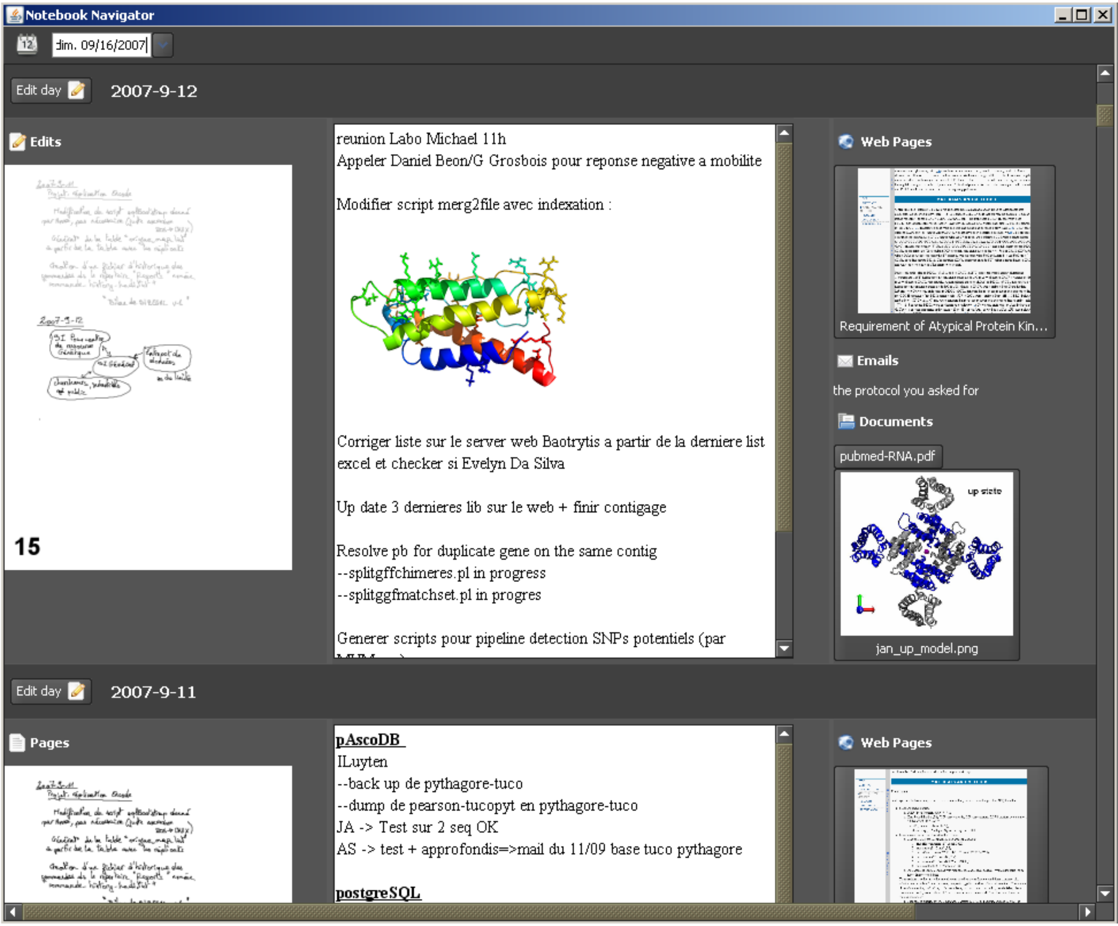


Figure 30: Prism v1.0 (Notebook Navigator) with 3 parallel streams organized chronologically, **Paper notebook** (left), **Electronic notebook** (center) and **desktop activity** (right).

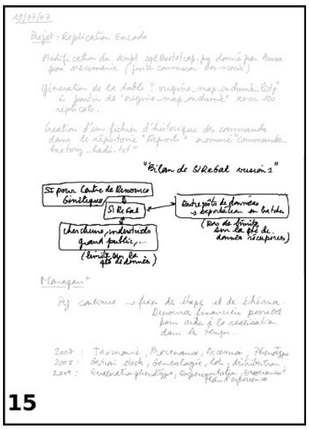


Figure 32: Digital version of an Anoto page

We use the Paper Toolkit (Yeh et al., 2008) to collect the data sent via the USB cradle, parse the pen strokes and store them in XML files. Prism transforms these files so that every file of strokes corresponds to a given page on a given day. This allows Prism to quickly display the content of a notebook page in different colors depending on the day it was written. In figure 32, the center of page 15 is darker because it was written during the day currently viewed in Prism, the top and bottom of the page are light-colored because they were not written the same day.

In order to improve load time, we generate images of the notebooks pages every time new strokes are written on a page and uploaded. Prism thus parses strokes only once when they are added and then displays pre-rendered images of the pages. To further improve load time, Prism caches thumbnails of pages.

Electronic Notebook

The electronic notebook is similar to an online journal where users write daily entries. Participants felt comfortable having a paper version and images of their paper notebook, but they were concerned about the portability of their digital data. Considering the popularity of the HTML notebook at the Institut Pasteur and INRA, we kept the electronic notebook as a visual HTML editor integrated into Prism (figure 33).

We modified the open-source editor ekit[5], to a notebook use. The editor accepts dragged images and documents, file URI handlers in addition to standard HTML. Users can thus create links to web pages but also to files on their computer which can be opened from Prism in one click. Prism stores daily e-notebook entries as separate HTML files. Users could thus read or edit their entries with any other application and move them from one computer to another.

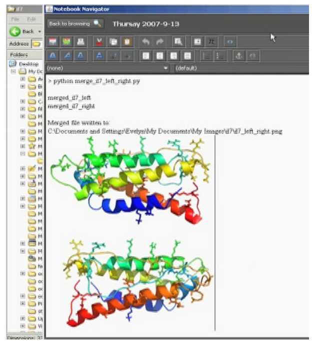


Figure 33: Editing an e-notebook entry.

Activity Logger

The activity logger aggregates key elements from users' streams of on-line activity, including email messages, web data and computer documents. The study highlighted users' need for a way to integrate computer information into notebooks. We investigated automatic desktop loggers such as wmtrace[6] (Chapuis, 2005) or PersonalVibe[7] (Hutchings et al., 2004). However, participants used a wide variety of platforms (Windows, Unix on Sun machines and Mac OSX), making it difficult to log at the desktop level. This would have required us to develop low level logging libraries for each system. Furthermore, we identified the importance of saving and writing in the field work and wanted to explore how we could support it on the computer. In order to simplify the capture mechanism and integrate it with the applications Biologists use daily, we identified three main common on-line elements that users wanted to save: web pages, email messages and documents.

WEB PAGES: Biologists browse the web looking for research articles, online databases and applications, and more generally for any kind of information they need. Bookmarks and history are not sufficient, they quickly become overloaded and users, rather than using these tools, prefer to navigate back to pages or to use search engines.

We observed that researchers created links from their notebooks to digital pages, either to remember a page, an application, or to re-run an analysis with the parameters already encoded in the URL. Prism thus provides a mechanism for sending web pages to the current entry of the notebook.

We developed a Firefox extension in XUL and JavaScript that displays a small button next to the URL address bar. When a Biologist visits a page of interest, she clicks the button to mark the page. The extension captures the URL, the title and a snapshot of the webpage, saves it in the web stream file which notifies Prism. The image of the page then appears in the online activity stream of Prism. The extension further allows users to send images, links or email adresses embedded in webpages to Prism. Users have to right click on any of these elements and select the type of link they want to send to Prism.

Figure 34: Button added to the Firefox browser and the Thunderbird email client to capture web pages or email and store them in Prism.

EMAIL: Biology researchers use email extensively to communicate and share information. We observed that they pasted email messages into their paper or digital notebooks, either to save them or to put them in context. Prism thus provides a mechanism to embed links

5 http://www.hexidec.com/ekit.php
6 http://insitu.lri.fr/~chapuis/software/wmtrace/
7 http://research.microsoft.com/en-us/downloads/0ea12e49-8b29-4930-b380-a5a00872d229/default.aspx

to email messages. On the model of the Firefox extension we built a Thunderbird extension displaying a button next to all messages. When the button is clicked, the corresponding message's subject line is added to Prism in the current day entry. For both the email and browser extension, subsequent versions allowed users to also tag the content they were saving.

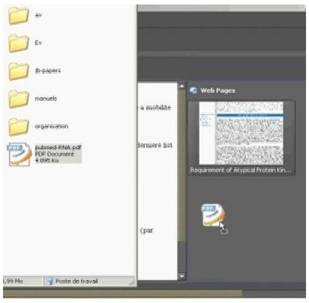


Figure 35: Dropping a pdf file into Prism.

DOCUMENTS: Biologists write, review and edit documents, from biological data to scientific articles. We observed the complexity of file management in the field studies and wanted to provide a lightweight way to follow the evolution of files. Users could save documents manually by dropping a file into Prism. If the file size was less than 50MB Prism created a backup of the file, which allowed participants to use Prism as a versioning system.

Prism also integrated a cross-platform monitoring system which tracked the evolution of the *recent documents* folder. This tracker filled the document stream of Prism with links to documents recently opened. The tracker allowed us to explore the advantages and drawbacks of automated logging over manual capture of users' documents.

Data Management

Since participants fed Prism with real data, and invested their time, we had to guarantee that their notes would survive after the study was finished. Based on discussions with participants Prism had to comply with the following properties:

- durability of the data,
- stability of the system,
- confidentiality and privacy control.

For *durability*, the electronic notebook entries are HTML files. The Paper Notebook is always accessible through paper and pages are also available as PNG files. We chose an XML format to collect the data generated by the users from all three activity streams. Figure 36 illustrates the stream data model. The data are divided into daily entries with fields indicating the content and date. The content part differs from one stream to the other. It holds links either toward the content of paper or digital notebooks entries, or toward email messages, web pages and documents.

```
<activity stream>
  <entry id="" date=""><content/></entry>
  <entry id="" date=""><content/></entry>
  ...
</activity stream>
```

Figure 36: Initial stream data model.

We tested updates internally for a week before distributing them to users, to be confident of the *stability*. To ensure *confidentiality*, we stored the data locally at the INRA lab, where we performed the study.

Notebook Navigator

All three streams are aggregated and visualized via the *Notebook Navigator*, Prism's information browser (figure 30). The Notebook Navigator organizes the streams chronologically with the most recent entries at the top. Users scroll through time to access their stored traces, or alternatively they use a calendar to jump to a specific date. They can create entries in the future to set reminders or create to-do lists.

Users edit the e-notebook through the Navigator by scrolling to the relevant day and clicking on the edit button. They can then switch from the stream view, to a day view, where one stream is emphasized and the others are minimized on the right side.

The Notebook Navigator application is written in Java. At launch, it loads the XML files describing the streams (paper notebook, electronic notebook and on-line activity) and creates a DOM[8] representation so that the resources can be loaded quickly while scrolling.

5.3 Results of Prism's initial use

Figure 3 summarizes the use of Prism v1.0 over two months. Guillaume mainly used the Anoto paper notebook and resisted the e-notebook since he worked on different machines and could not share his notes among them. Sarah and Adeline who used Mac machines kept using their paper notebooks because Anoto-pen drivers were only available for Windows. They created digital notes regularly and integrated webpages into their notebooks either to save temporary results or to capture and share them with others by email. Nadia kept her original HTML notebook but switched to the Anoto notebook. Jeanne who did not use dedicated paper and electronic notebooks but text files, and never knew where to look for the information she saved, embraced both notebooks and saved documents related to her entries in Prism.

Participant	platform	paper pages	e-notebook entries	web pages	email	documents
Guillaume (M)	PC/Unix	42	7	2	1	1
Adeline (M/R)	Mac/Unix	N/A	28	13	1	7
Sarah (R)	Mac/Unix	N/A	10	26	1	5
Nadia (R)	PC/Unix	16	4	2	1	3
Jeanne (R)	PC/Unix	21	22	6	1	3

Table 3: Prism use after two months (45 business days).
M: manager, R:researcher.

We observed how participants adapted Prism to fit their research activities and identified three types of behaviours we considered important and common enough to explore in workshops and integrate into the design iterations:

- *organization strategies:* participants developed to balance chronological ordering of entries with the projects they worked on.
- *alternatives to filing:* participants used Prism when dealing with information they did not know how to save.

8 Document Object Model (DOM) is an interface that exposes an XML document as a tree structure usually loaded in memory.

- *personal and shared information on the web:* participants used online applications on a daily basis.

5.3.1 *Organization strategies*

For managers who handle many projects at once, chronological organisation artificially scatters information, making it difficult to handle the wide range of activities they engage in. The bioinformaticians also work on multiple projects, but are able to focus more on their own research.

In order to help overcome the limitations of the purely chronological structure of notebooks, everyone used keywords either to track activities, e.g., *meeting* or *submission*, projects, e.g., *pascoDB* or *blumeria*, or data type, e.g., *Perl scripts*. Jeanne even used a file name as a keyword, *splitgffchimeres.pl*, since that file was evolving from one day to the next and she wanted to follow the progression in her notebook.

Nevertheless, time remained an important organizing principle for participants. Both managers and researchers used Prism's electronic notebook to *set reminders for upcoming tasks*. They marked webpages, documents or email messages about meetings or items to complete, such as an unfinished README file. The notebook acted as a task manager. Bioinformaticians wrote actions items in electronic notes set in the future and progressively transformed the action items into paragraphs as how they completed the task.

Finally, researchers were reluctant to use the automated document logger. They felt it was too intrusive and added irrelevant entries into Prism. They preferred to mark documents by themselves, to limit the number of entries in the notebook and to control the data they published.

5.3.2 *Prism as an alternative to filing and distributed data*

Jeanne and Adeline used Prism as an *alternative to standard filing or bookmarking*, especially when they had doubts about where to file a document or when information seemed useful for a short period of time. Most researchers in the group used at least two computers with different operating systems. As Nadia explained: *"being cross-platform is a problem when we want to follow data"*[9]. Jeanne gave an example: *"This file should probably stay here [points to a folder on the Solaris machine], but I'll need it for a presentation [done on a Windows laptop] so I don't really know where to keep it."*[10] Participants wanted to adapt Prism to manage their information transparently on their different computers.

Three participants used Prism as a *versioning system* to track the progress of research articles or interesting web pages. Since Prism stores snapshots of web pages, users discovered they could *capture the parameters of web forms* and later re-examine versions of pages that were no longer available online. Online bioinformatics applications only store analysis results for a limited period of time in order to save disk space on servers. Saving a version of the page let users avoid re-running costly analyses.

9 "être multi-plateforme c'est un problème quant on veut suivre des données."
10 Cet fichier devrait probablement rester là, mais j'en ai besoin pour faire une présentation, alors je ne sais pas trop où le ranger.

5.3.3 *Personal and shared information on the web*

The interviews presented in the previous chapter combined with the use of Prism highlighted participants' use of Web applications to share information. We were surprised by all the personal data that was only stored remotely, on the Web. In addition to Google Mail, participants shared documents, bibliographic references[11], and calendars using web-based applications. They also hosted JIRA on their servers, a Bug and issue tracker and Alfresco, a document manager.

Participants shared information, captured with Prism, with colleagues. We initially envisioned the capture of web pages as a way to keep track of useful information from the web, e.g., research articles. Adeline explained how Prism helped her follow the evolution of a website she was developing because of the snapshots it took when she marked the page, she could then share her progress with her collaborators.

Participants also created links to web applications, such as ongoing tasks in JIRA. Jeanne created an ecosystem of links from her electronic notebook to README files, the scripts she modified on her machines and related JIRA tasks. Because of the dynamic nature of these applications, the URLs participants captured were not always linked directly to the desired content. Participants often had to log-in or navigate to the desired information: a particular email, event or biological application result. These observations led us to modify Prism further, to better integrate it with the web.

Prism v1.0 limitations

After two month of use, we identified missing features that made some participants reluctant to use Prism. Prism did not provide any *search* functionality. It seemed acceptable at the beginning, but after two months of use, it became a problem that limited participants' use of Prism.

When editing notebook entries, participants wanted to check past entries or copy information from another stream. However, Prism editing was modal, which made it complicated to copy content from a past entry while already typing text in the daily entry. Users could either be in editing mode (restricted to one day) or browsing mode. Prism also started to suffer from longer load times as users added more and more multimedia content (such as paper pages and web content).

Finally based on our observations and the participatory workshops we led, we decided to create a distributed version of Prism to enable users to integrate information from any computer platform, including web activity. In the following section, I present how we shifted to a web application, added tagging capabilities, and integrated Web streams into Prism.

11 CiteULike: http://www.citeulike.com

5.4 PRISM DESIGN ITERATION

Figure 37 shows Prism v2.0 as an online application. I re-designed Prism for the Web, turning it into a distributed application that participants could use on their multiple computers. Prism v2.0 still integrated Anoto notebooks (left column of the figure) and the desktop streams (in the center). Users accessed Prism v2.0 in their web browser, and could benefit from the following new features:

- *Search*, users could benefit from the page search function of the browser.
- *Redesigned e-notebook*, users could edit, navigate and copy and paste among entries in parallel, thanks to a re-implemented, non modal, e-notebook (right of the figure).
- *Extended tagging*, users could tag notebooks entries, not just web pages and email, and could use tags to filter entries (far right column).
- *Feed subscription*, users could follow their web activity or those of colleagues who also used Prism.
- *Feed broadcasting*, users could share their activity among their different computers or with colleagues.

Prism v2.0 was hosted on a server from INRA for security and privacy reasons. The web server uses Apache and PHP, while the client can be any modern browser. JavaScript and asynchronous requests to the server in an AJAX style make Prism sufficiently interactive and avoid page reloads, while the Yahoo UI library made it easier to build Prism around HTML widgets.

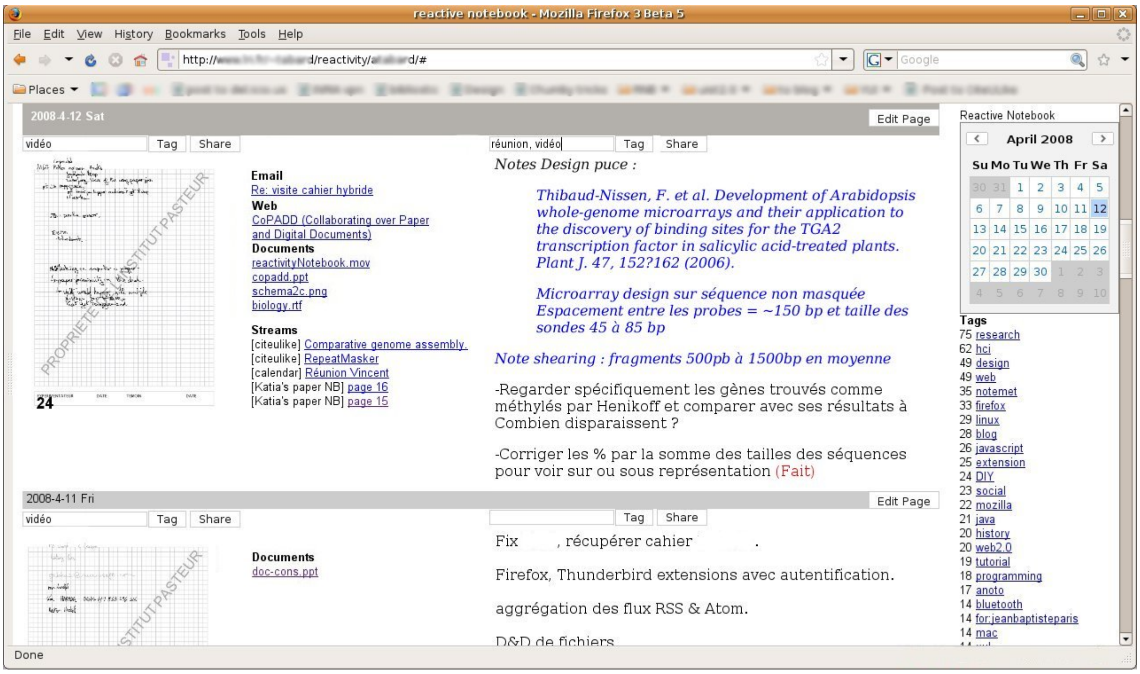


Figure 37: Prism v2.0, web application, Anoto notebook (left), online activity (center), e-notebook (right), tags (right column).

5.4.1 *Distributed information streams*

Prism v2.0 integrates familiar information scattered on the Web as a new type of activity stream and lets participants decide which information to follow and share. While many web 2.0 applications provide APIs to access their content, using APIs implies coding a different module for every application. As a lightweight alternative, I decided to use web feeds[12] to incorporate *web streams* into Prism. Prism supports RSS and Atom feeds which are the two main web feed formats used to publish information in a computer readable way. I modified the current stream model to meet the Atom format[13]. Figure 38 shows how the stream model followed the Atom syntax, with a header providing meta-data about the stream, indicating the last update of the feed and of every entry:

```
<feed xmlns="http://www.w3.org/2005/Atom">
    <title> Paper notebook feed </title>
    <subtitle>Prism - hybrid notebook</subtitle>
    <link href="http://www.lri.fr/~tabard/reactivity/user"/>
    <id>urn:uuid:60a76c80-d399-11d9-b91C-0003939e0d32</id>
    <updated>2009-05-13T18:30:02Z</updated>
    <author>
        <name>User name</name>
    </author>

    <entry>
        <title>2009-05-13 Paper notebook</title>
        <link href="http://www.lri.fr/~tabard/reactivity/user/paperNB/page15-1.png"/>
        <id>urn:uuid:1225c695-cfb8-4ebb-aaaa-80da344efa6a</id>
        <updated>2009-05-13T18:30:02Z</updated>
        <summary>Page 15 edit 1</summary>
    </entry>
</feed>
```

Figure 38: Revised stream data model based on Atom.

Figure 39 shows the overall architecture of Prism v2.0. Prism v2.0 aggregates two types of streams: Web feeds outside of its control (in the cloud) and the activity streams of Prism v1.0 (coming from the desktop and Prism's online application).

The desktop part of Prism v2.0 consists of Firefox and Thunderbird extensions and a JAVA daemon monitoring Anoto activity. The Anoto server uploads new pages when users connect the pen to the computer. The e-notebook is directly available online as well as the tagging facility.

Prism aggregates and caches Web feeds so that users can access them later, even when they expired on the original server. For example, Google Calendar only publishes the last 25 events in its feed. In order to display the whole series of events, Prism must save the feed items on the server. Prism uses simplePIE[14] for parsing feeds.

12 "A web feed is a data format used for providing users with frequently updated content. Content distributors syndicate a web feed, thereby allowing users to subscribe to it. Making a collection of web feeds accessible in one spot is known as aggregation, which is performed by an Internet aggregator." http://en.wikipedia.org/wiki/Web_feed

13 Atom is an XML-based document format that describes lists of related information known as "feeds". Feeds are composed of a number of items, known as "entries", each with an extensible set of attached metadata. For example, each entry has a title. [RFC 4287]

14 http://simplepie.org/

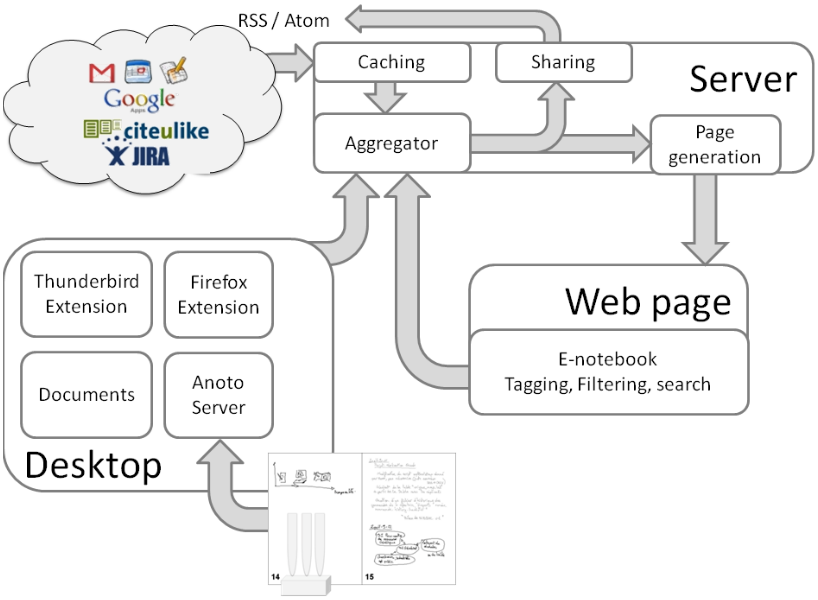


Figure 39: Prism v2.0 stream architecture.

The new architecture around feeds made it easier to aggregate information from different sources without requiring further programming. It also opened up new opportunities for customizing information sources and sharing information.

5.4.2 *Shared broadcast feeds*

Basing streams on the Atom feed format improved data portability by making entries readable by any feed aggregator[15]. It also made it much easier to build a sharing mechanism on top. I thus designed a broadcast and subscription mechanism for the desktop streams (notebooks, documents, web pages and email), based on the Atom feed format (Figure 40).

Users could customize their streams of activity and subscribe only to the ones they were interested in (top part of the figure: **Subscribe**). They could also share their notebook entries by making streams available to others. Publishers could attach tags to entries and only publish a feed containing the tagged entries (like *public*). To protect privacy, users had to use the feed generator (lower part of the figure: **Share**) which encodes the feed parameters (type of stream and tags), so that no one can guess feed URLs. This forces users to send the URL feed directly to the people interested.

15 Wikipedia lists more than 60 desktop feed aggregator clients http://en.wikipedia.org/wiki/List_of_feed_aggregators

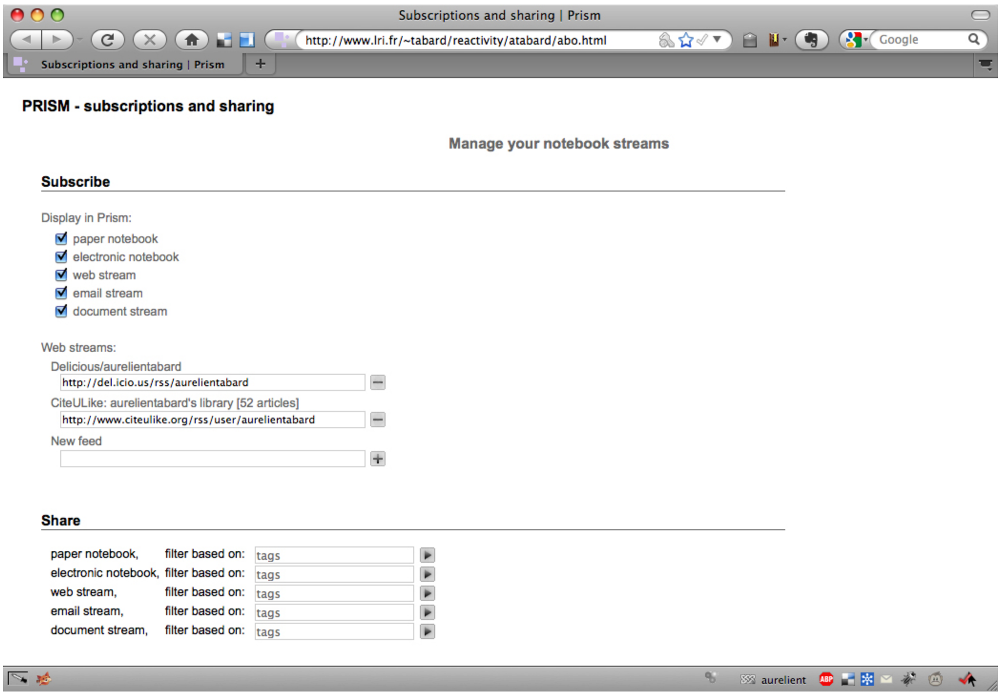


Figure 40: Broadcasting and subscribing to information streams in Prism

5.5 Longitudinal study results

As a technology probe, Prism provided insights about how scientists manage information over time. In addition, it helped us to better frame what a hybrid notebook should be.

5.5.1 *Shared and distributed information*

Participants shared information for future use or with colleagues. We identified different sharing relationships:

- Sharing from junior to advisor. For example, one post-doc shared her notebook with her supervisor, via the web.
- Sharing with remote colleagues. A researcher captured web pages as she developed them. The web stream became a way to share a series of images, showing how her work had evolved, which she could share with her collaborators.
- Sharing with anyone that may be interested. Within the team, participants broadcast information without knowing who will see it, leaving README files in the program directories of the central server to record history and share details about how to launch programs.
- Team sharing with equal participation. Some participants engaged in reciprocal sharing of bibliographic references and calendar events with each other.

Participants did not only share information with colleagues but also with themselves, either to follow the evolution of their work over time, or to share personal data across computers and platforms. One of their main problems was to keep track of data available on different machines, on paper and on the web. Jeanne and Adeline used activity logs and the electronic notebook to keep track of data across Unix, Mac and Windows computers. Keeping a version of the file on the server facilitated transfer from one machine to the other and reassured them that they would be able to access it later.

Sharing is a motivation for reflection, *"I always think about how I did something and write how we did it, to help the others. It's more work, but it's well used"*[16] (Adeline). As researchers decide what to save and how to write it down, they must explain their activity so that others will be able to understand it. They reflect upon their activity as they document it. However, researchers do not always articulate activity so clearly, for reasons of time, motivation or efficiency. They often have an idea of who it is for and how it will be used in the future, and adapt their notes to the context of use.

5.5.2 *Organizing information*

Enabling Prism for the web allowed it to become a central notification point. With its ability to aggregate and distribute information, Prism concentrated information from heterogeneous sources into a central location. Some information is repeated or related from one stream to the other, other information is repeated over time.

We initially envisioned stream integration as a way to avoid duplication of information, letting users have the right information at the right place. Yet information is dynamic and evolves over time. One participant explained how the e-notebook content followed the progression of her work: *"I mark things I have to do in future days and I develop them, copy or move depending on what I did. [. . .] This way, when a task is done, I have a complete description of it"*. In such a case, snippets are repeated from one day to the next and this serves as a record of the evolution of the tasks.

Participants referred to information from one stream to the other and developed a common vocabulary to support future reflection on their work. In addition to common tags such as TODO, IMPORTANT, and DONE, they also color-coded content, surrounded activities within the notebook and added meta-notes to comment on and synthesize existing notes. These meta-notes are easy to distinguish: text may be shifted, colored, or pointed to with arrows.

Organizing information does not only mean filing or categorizing it, but rather *treating it*. As users decide what to save, they organize it mentally and process it in a deeper way than would have happened otherwise. We tried to design Prism's capture tools as lightweight as possible, allowing users to process information without being distracted.

5.5.3 *Master Notebook*

Prism's hybrid design, by linking heterogeneous information, provided new opportunities for associating information. But it also resulted in a wealth of information across different media, platforms,

16 Je réfléchis toujours à comment j'ai fait quelque chose et à comment le décrire pour aider les autres. C'est plus de travail mais c'est mérité.

and organizational structures. Managing information coming from different locations and updated at different rates was a challenge for everyone we spoke to.

By offering highlights of users' activity, Prism allowed users to build a reasoned account of their activity. We found that participants created a reference point, or *master notebook*, to organize the diverse strands of their personal activity. As Guillaume stated: *"We need a master document that is the source of the different things"*.

The master notebook is where researchers reflect upon their work: they comment on an interesting web page that was marked (Figure 41-A), use hand-drawn sketches in the paper notebook to explain how some tasks are related to each other (Figure 41-B) and add electronic notes to track how a project is evolving (Figure 41-C).

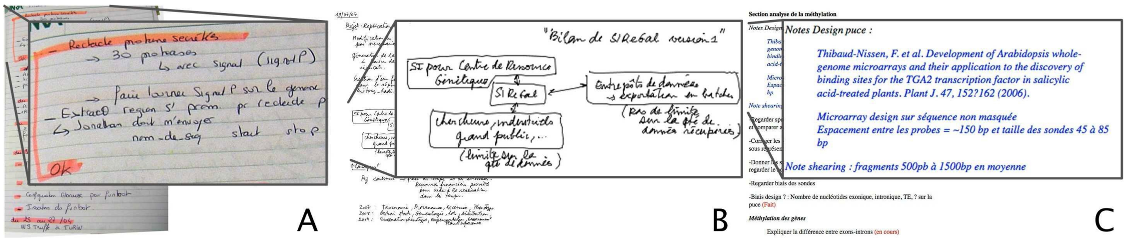


Figure 41: Reflection strategies in Master Notebooks,
A: Color coding and post-hoc remarks (paper notebook),
B: Project organization (Anoto notebook),
C: Meta-notes shifted from the body and colored (e-notebook).

Interestingly, the bioinformaticians who could use Prism's paper notebook, turned it into their master notebook (Figure 41-B). The two Mac users who did not have access to a Prism-based paper notebook still referred to one of their notebook as the source of their activity. Adeline used a standard paper notebook not connected to Prism (Figure 41-A) and Sarah used her electronic notebook (Figure 41-C).

5.5.4 *Feedback, limitations and Prism v3*

The overall feedback on Prism was positive. One participant commented that it allowed her to keep the discipline of paper lab notebooks, with the old flexibility of paper binders that researchers used before the patent era. One of the major benefits of Prism they reported was the integration of streams at different levels of granularity, from an email to a reflection about analysis results. They felt that it helped them manage information in the course of their activity. However, information organization remains a problem and tags alone are no panacea. As organization schemes change, and information evolves, the initial tags users created may not be relevant anymore.

Participants used Prism alongside multiple existing tools, notably online applications for managing information, which motivated our shift to an online stream architecture. But participants kept relying on other tools to manage part of their information, such as many desktop applications, or even a notebook they preferred. Because Prism was designed to be cross-platform, it provided limited integration with existing desktop applications (we only focused on Firefox, Thunderbird and the file system). By focusing on a specific operating system, we could have integrated lighter and more pervasive capture mechanisms such as in Tagctivity (Oleksik et al., 2009) or Giornata (Voida and Mynatt, 2009).

Prism's evolution toward a Web application made capture more automated. When participants subscribe to external feeds, all the entries are pushed into the notebooks; in the current version of Prism there is no functionality to remove an uninteresting item from external feeds. With Prism, participants had either to save information manually (by uploading a document or marking an email or web page) or automatically (by subscribing to feeds). Prism did not support a middle ground, for example automatically saving information for short periods of time in order to let users decide if they want to keep it, or to let users mark an email as potentially interesting (for a limited duration).

Finally, Prism was designed as a technology probe, not a final prototype. It allowed us to explore design directions and served as a basis for discussing participants' use of a hybrid notebook. There are a few shortcomings that I did not adress, preferring to focus on directions discussed with participants. One shortcoming is the limited visualization and navigation capabilities, which made it annoying to browse past entries. Another is Prism's search functionality which relied on the page search functionality of the browser (meaning that users could only search in one month at a time.)

The next version of Prism takes these problems into account, in the context of the ReActivity[17] project which explores how to support scientists' reflection on activity. To solve the limited visualization and search capabilities, facilitate user management and enhance security, Prism is being re-designed around Wordpress[18], an open source web publishing platform.

5.6 DISCUSSION

Participants adapted Prism, actively filtering, saving and synthesizing familiar information they deemed important to reflect on. Prism stream architecture facilitated the integration of sources and provided a general way to publish and reuse information across applications. The use of master notebooks revealed that scientists' needs go beyond personal information management. As they manage information, they reflect on it, which influences how they think about it.

5.6.1 *Stream architecture, adaptable integration of information*

Bellotti et al. (2003) and (Boardman, 2004) stress the importance of integrating information management tools directly into the applications handling the information rather than creating dedicated personal information management tools. However, the information familiar to scientists is distributed over paper and digital notebooks, many machines, and the Web. Re-designing all the applications scientists use to manage their personal information would be costly. With Prism, we thus focused on integrating information, in a spirit closer to Lifestreams (Freeman and Gelernter, 1996).

To support the diversity of personal information management, we had to provide an adaptable tool capable of integrating the diverse sources of familiar information. Prism supports publication streams

17 http://www.msr-inria.inria.fr/Projects/reactivity
18 http://wordpress.org

adapted to the type of information users manage. Based on existing standards, Prism uses the Atom feed format, which is convenient for distributing information:

- For users, feeds are already available on the web. They can be read by people who do not use Prism, either to subscribe to or to publish information.
- For developers, feeds are easy to support, with many libraries available. If applications broadcast feeds, they can be interpreted by Prism and by many other applications.

The use of standards allows users to quickly visualize data, and answers concerns regarding durability expressed by participants. Streams can be browsed using any feed aggregator, and Atom is an open format that could be extended to support particular event types. Finally, the stream approach is relatively decentralized, every application can produce one or many streams of activity.

5.6.2 *Saving: to process and offload information*

A key benefit of Prism is the ability to capture information from diverse sources, either manually, through writing or copy and paste, or automatically, through file monitoring and web feed aggregation. There is a challenge here, to balance the fluidity of ongoing activities and the conscious capture of information.

On the one hand, automated logging is invisible and captures large quantities of information. However, participants felt that monitoring their desktop documents' modifications was too intrusive. They used their laptops for professional and personal activities and did not want to have to control the monitoring applications. Furthermore, the captured content was noisy, it contained many documents that were not related to the daily tasks and which made it more difficult to locate the relevant documents.

On the other hand, conscious capture let participants control the information they save. It forced participants to make the conscious decision of what information to save and required a shift of attention which may disrupt their work-flow. However, because it is conscious, the participants are also more likely to remember that it occurred and that it is available for subsequent retrieval. We thus tried to design Prism's marking and saving mechanisms as lightweight as possible so that users could offload cognitive charge on Prism without distraction.

The writing process played an important role in helping participants to filter and organize information. With Master notebooks, participants built a definitive record of activity, they pruned information to build a narrative of their ongoing activity, that could later be used by them or others.

5.6.3 *Redundancy as a resource for reflection*

While building their master notebooks, we found that the participants valued redundant information because it helped them to reflect on their previous activity. Redundancy helps scientists understand how their thoughts have evolved over time. When an item appears again and again in different forms, it indicates that it is probably important. When redundant events are linked to each other, scientists can make better sense of them, and in some fortunate instances, redundancy allows them to discover new insights and move in creative new directions.

This is counter intuitive from a computer science perspective, since allowing redundancy might be viewed as adding complexity and wasting space. Repeated writing may be seen as inefficient. However, Biologists find value in copying and transforming information manually from one tool to another, since it helps them think about it and become familiar with it.

Redundancy is at some point irreducible: a reference to a project can take many forms. No 'intelligent' algorithm could ever be able to recognize all the relevant connections and repetitions. Users take advantage of their context and experience when reflecting on their activity. By providing multiple possible paths for finding information, users can revisit previous contexts and draw new connections.

5.7 CONCLUSION

This chapter presented Prism, a hybrid notebook, integrating heterogeneous streams of familiar information: paper notebooks, e-notebooks and elements of users' online activity. Prism relies on both automated and manual capture of information. Users preferred to manually save information as it let them control, transform and annotate it. Users' active selection of information helps them make sense of the information they manage. To this end, they used *master notebooks*, a notebook dedicated to reflect on the information they have at hand.

In the next chapter, I explore a middle ground between manual and automated capture which did not exist in Prism. How can we let researchers actively save information without disrupting their workflow?

5.8 SYNTHESIS

The contributions of this chapter are:

- The description of **Prism's stream architecture** to integrate heterogeneous information.
- A discussion of the **benefits of active saving** by users over automated logging.
- A **longitudinal study** of Prism's use and evolution led to consider **redundancy** of information in PIM systems as **beneficial**.
- The concept of **master notebook**, users adapted Prism to reflect on their activity in a dedicated notebook.

6 PAGELINKER: PUTTING TRACES IN CONTEXT

This chapter presents PageLinker, a contextual bookmarking tool. We designed PageLinker with Biologists based on a study of their browsing habits. PageLinker field evaluation shows that it reduces significantly the number of pages loaded to complete tasks, allowing users to navigate from one task to the next more efficiently.
PageLinker has two main implications for design: implicit interaction as a method to balance automated and manual capture of activity; and interaction traces, which group visited web pages together without requiring organization efforts from users.

This chapters describes PageLinker, a contextual bookmarking tool, that addresses two issues raised by Prism:

- How to design a history tool that lets users easily navigate within its entries, without relying on a chronological log or active organization of hierarchies.
- How to balance manual and automated capture of activity, so that users can focus on their ongoing workflow and actively save information at the same time.

PageLinker supports users' exploratory but repetitive Web browsing tasks. When Biologists repeat an analysis with different data, they do not repeat the exact same tasks but could benefit from information about their past actions. To this end, when users copy and paste content between two pages, PageLinker creates links between them. The next time users come back to one of these pages, they can easily jump from one page to the next via the contextual bookmarks (zoomed-in part of figure 42).

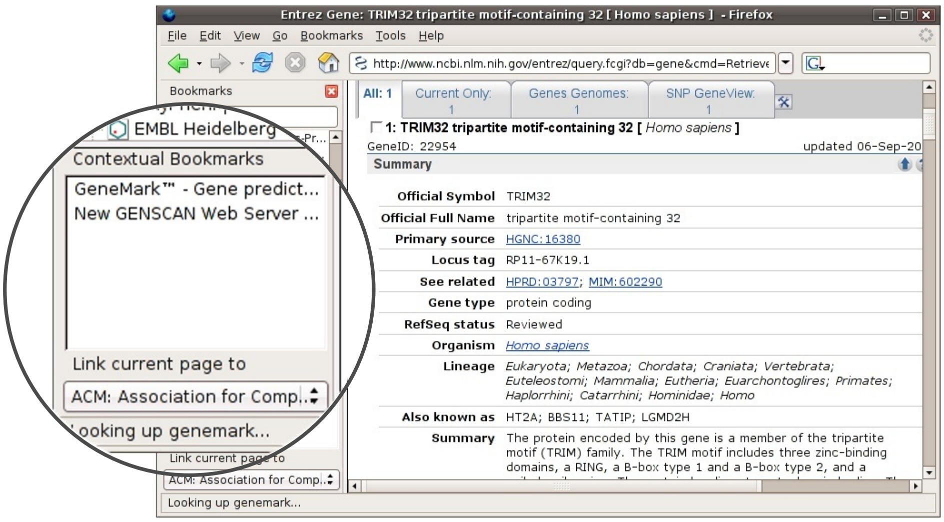


Figure 42: PageLinker's contextual bookmarks provide links to relevant web pages previously visited by users.

The next section (6.1) motivates our choice of focusing on researchers' web browsing as a subset of information management practices. Then I present the results of a study highlighting the main problems researchers face when browsing or analyzing data on the Web (section 6.2). The key problems we identified led to the design of PageLinker (section 6.3) which I evaluated in the field over a period of one month (section 6.4). I finally discuss implications of PageLinker for the design of information management systems (section 6.6).

6.1 MOTIVATIONS

6.1.1 *Biologists and the Web*

Biologists constitute a particularly web-intensive group of users. Aside from classical Web tasks such as looking for information or collaboration, they treat the Web as an enormous, constantly-searched database and as an analysis tool. They repeat collections of tasks, revisiting the same sets of pages, browsing sequentially and in parallel as they analyze data sets and pursue hypotheses.

Such activities are not specific to scientists, but are also representative of the use of the Web as a major source of information for knowledge workers (Sellen et al., 2002). We can even encounter these complex browsing instances in mundane situation, while planning a trip or buying a product online.

Aside from having results that apply to a wide range of the population, studying web browsing as a subset of information management proved to be popular in HCI as it presents two advantages for researchers:

1. Web navigation happens in one application: the Web browser. Studies can thus focus on the use of one tool within the information management process.

2. Web page visits can be easily monitored, through proxies or browser add-ons. Studies can thus leverage logging and quantitative analysis.

6.1.2 *Current browser limitations*

Very few Biologists use bookmarks or history pages, relying instead on Post-It notes, e-mails or search engines to find the sites they visited earlier. Web browsers have not kept pace at the level of user interaction with the evolution of users practices.

Bookmarks and history pages become overloaded and are not sufficient anymore. These early user aids for finding previously visited pages, have little evolved since their introduction in the early 1990's with Mosaic (Baker, 1994). More generally, even though studies showed that revisitation accounts for more than half of the visited pages (a decrease from 80% to 50%) (Catledge and Pitkow, 1995; Tauscher and Greenberg, 1997; Cockburn and McKenzie, 2001; Weinreich et al., 2006) it appears that the tools designed to help users find again and re-visit web pages, e.g., bookmarks and history, are rarely used (Catledge and Pitkow, 1995; Tauscher and Greenberg, 1997; Obendorf et al., 2007).

Capturing information with bookmarks breaks users' flow due to the associated filing task. Users must decide to bookmark the page they visit and where to file the bookmark within the hierarchy (Abrams et al., 1998). The changing nature of the web and users' changing interests (Dix and Marshall, 2003) cause classification and relevance problems. Bookmark lists tend to grow over time as users add new pages without removing unused ones (Cockburn and McKenzie, 2001), providing *"neither a reminding function nor a context of relevance"* (Jones et al., 2001). If users do not constantly edit and prune their lists, they end up with inappropriate and uninteresting URLs, little better than no bookmarks at all (Tauscher and Greenberg, 1997). Histories, on the other hand, are difficult to review and provide a very poor description of the pages visited (only date and title). Users thus lose pages that are useful to them, take more time to retrieve information or to achieve repetitive actions.

6.1.3 *Visualization tools*

Graphs of navigation history provide an alternative to history lists (Hightower et al., 1998), situating current activity within previously used paths. However, graphs require additional screen space and force users to shift between their primary browsing tasks and a secondary orienteeting task. An interesting alternative is WebView (Cockburn et al., 1999), a browser enhancement that integrates several revisitation capabilities into a single display, resulting in a compact revisitation tool. While WebView goes in the right direction, it focuses mainly on providing a better interaction with the global history. Ringel et al. (2003) or Dumais et al. (2003) proposed putting the history into perspective with past activities by intertwining public landmark events (news or weather) or personal ones (other pages visited or emails received).

6.1.4 *Search and automation*

An alternative proposal for supporting better revisiting of web pages is to provide search tools in the history and a means of automating current navigation based on past experience. By defining users' interests or modelling their behavior, one might predict their navigation path. Morrison et al. (2001) and Adar et al. (2008) worked on defining taxonomies that classifies types of revisits. Based on the idea that users follow information scents, information foraging (Pirolli, 2007) provides methods of modelling users' actions and predict their navigation paths.

However, the fluctuating workflows of Biologists make it difficult to predict navigation based only on the information captured by the browser. Biologists rethink their workflow at each navigation step. They base their choices upon multiple situated factors, including the complexity of the analyses they run, balanced by the time they have, knowledge of server loads or difficult-to-articulate factors such as ones' intuitions about whether certain results are *'normal'*. Tools developed to support navigation should thus take into account and try to support the situated nature of browsing, information seeking, filtering and retrieval.

Other common problems with automating complex workflows are the lack of transparency, of automation tools. Particularly when users search for the cause of unexpected results, the lack of interaction prevents them to trace back to the source of their problems, or explore different alternatives in parallel. A biologist commented that he *"needs to redo the protocol step by step because there is no convenient way to access the problem source directly"*. Also, the instability of the Web introduces many practical problems: changes in page content, URLs, and data formats can break formerly correct automation sequences.

Finally, while search is useful, it does not answer all of users' needs. Teevan (2007) showed that people perform most directed searches by *orienteering* via small, local steps using their contextual knowledge as a guide, rather than by *teleporting*, or jumping directly to their target using a keyword-search utility. Re-finding information (or looking for information you already looked for) is different from finding new information. She notes that a *"distinguishing feature of re-finding is that the searcher often knows a lot of meta-information about the target".*

The process of navigating through various websites acquaints Biologists with changes on the servers, new programs, and new layouts that might provide easier access to some pages, helping them to gather knowledge about their virtual environment. In a desktop context, Bergman et al. (2008) found that users preferred navigating rather than searching on their file system; Teevan et al. (2004) argue that directed situated navigation reduces the quantity of information that users need to specify and provides the context they need to help them understand the results they obtain.

6.2 WEB BROWSING STUDY

To further understand the problems Biologists face when navigating the web, we conducted a study at the Institut Pasteur. Biologists are highly experienced Internet users who have modified their work practices to take advantage of the wealth of biological data and analysis tools available on the Web.

With Prism, I discussed the advantages of manual saving over automated recording. On one hand, manual saving appears to be beneficial for the reflective process but breaks users' ongoing workflow. On the other hand, automation induces history overload and users lack involvement in the recording process.

The underlying goal of the study was to figure out how can we facilitate page re-visitation tasks. How can we design a tool that helps users to remember the pages visited in a way that does not interfere with their ongoing activities?

6.2.1 *Participants and procedure*

We selected 20 Biologists who had recently used on-line biological data and analysis programs as an integral part of their research. We conducted videotaped interviews in their laboratories and looked for specific examples of history and revisitations problems. After each interview, we wrote detailed transcripts.

During the interviews, we asked biologits to replay, with their own data, the most recent web analyses that they had run. We also asked them to search for specified information on the Institut Pasteur's online documentation. We used a talk-aloud protocol, asking them to explain their choices and what bothered them, as they performed these tasks.

We also organized a video-brainstorming (Mackay and Fayard, 1999) workshop that focused on refining PageLinker design. In 2006, during a computer science course for biology at the Institut Pasteur, students used participatory design methods to investigate tools facilitating the revisitation of web pages. I also presented PageLinker during a bioinformatics seminar at Institut Pasteur which was followed by a discussion on web browsing experiences.

We reviewed interviews transcripts and video prototypes, looking for repeated patterns. The goal of the analysis was to find as many specific examples as possible of browsing problems related to revisiting web pages. The following scenario illustrates typical problems a biologist studying a protein would face. It is a composite derived from real examples from the interviews designed to highlight the critical design issues such as the complexity of online analyses, the different data to handle or the lack of support from existing browsers.

6.2.2 *Illustrating the navigation problem*

> Anne wants to explore alternative hypotheses before conducting a time-consuming lab experiment. She begins by collecting data: From the Biology department's homepage, she follows links to the protein database page. Unfortunately, it does not offer links to relevant analysis tools and she must browse a huge, hard-to-navigate hierarchical directory with hundreds of links spread over many pages. She eventually finds the relevant page and checks the research literature to see if similar forms of the protein appear. She then looks for the protein sequence in two different databases to find out if different DNA sequences are associated with the protein. She encounters incompatible data formats, forcing her to transform the data before using her chosen analysis protocol.
>
> The lack of relevant links in the data pages makes it difficult for Anne to move from one step to the next. Even when she does find appropriate online resources, she has trouble keeping track of them. Several weeks later, when she decides to analyze a new set of data, she has to recreate her initial search process in order to find the same pages again. Like others in our study, Anne rarely uses bookmarks or history pages, and instead relies on Post-it notes, her notebook, emails and search engines to find previously visited sites.

6.2.3 *Observations*

Researchers use the web for a variety of activities, some very common to web users, others more original such as data analyses. The online analysis process illustrates several problems faced by Biologists. It requires them to use appropriate tools to go from one analysis page to the next as they pursue complex analyses. The data generated by one analysis tool is not always compatible with the next. Pages change and are not stable over time. Because of all the exploratory aspects of these analyses tasks, Biologists may have a goal, but not necessarily predefined plans.

Complex digital protocols

The analyses Biologists run online are complex and difficult to automate. Biologists analyze data through series of tools, often having to cope with different formats or versions of the same data. While running these complex protocols, Biologists must interact with the data, get a feeling for it and adapt the protocol steps accordingly.

exploration: Biologists' purposes and procedures change rapidly. Constructing an online biological protocol is not fully algorithmic and requires human judgment along the way. Biologists check the accuracy or significance of results and decide whether and how to carry out an analysis using complex criteria that would be difficult to automate. The reasons are difficult to articulate, Biologists rely on a sense of "rightness" (Latour, 1988) rather than only theories. A biologist might decide to use different processes, proceed with full data or extract subsets depending upon the characteristics of the data and her current research goal.

data flow: During this exploratory process, Biologists create diverse data flows, piping the output of one program into another as well as reformatting, transforming, filtering and extracting data. With the tools disseminated on different servers on the Web and on their desktop, Biologists rely on the simplest cross application communication mechanism: *copy-paste* to chain their protocols steps. Copy-and-paste may require reformatting but it is robust and the most efficient mechanism to pipe information from one tool into another.

Pipelining and automation tools can support Biologists when they want to repeat similar analyses. However, Biologists do not use them widely. The learning curve may be too steep and without sufficient rewards to attract Biologists. Furthermore, automated workflows may be efficient when they work well, but tracing back where errors or anomalies came from is usually difficult. When Biologists go from one step to the next manually, they can react to the results they get and adapt their analysis protocols.

As researchers, Biologists are specifically looking for exploration rather than streamlined processes. If they apply well-known protocols, and use predefined workflows, it is most likely for them to reuse their previous knowledge than for automating their task.

EQUIVALENT OBJECTS: Data formats are often incompatible: the output of one tool may not be interpretable as input by another tool. This forces Biologists to store the same data in the different formats needed by different tools.

Such incompatibilities prevent users from piping results from one analysis tool into another, which forces Biologists to edit intermediate results. It also makes replicability of experiments more complex, forcing Biologists to manage collections of *"equivalent"* data objects such as:

- same data in different formats, e.g., different tools may produce similar analyses but accept different data formats;
- different versions of the same data, e.g., two versions of an annotated genome; or
- name changes, e.g., the name of a gene can change after the genome is fully sequenced.

Relationship to software: exploration and intuitions

Biologists use techniques they already know rather than learning a new, potentially better one. Most stay with a single Web server if it provides all the tools they need, even though better tools might exist on other servers. Far from irrationality or pure conservatism, like other experienced computer users (Mackay, 1990b), most Biologists want to use a stable and predictable set of known tools.

When facing a problem, they rely on the team, close colleagues or the institutional infrastructure to gain knowledge of new tools. We observed in many teams, a person trusted for his or her technical knowledge who acts as a filter and translators (Mackay, 1990a) between new tools and less technically oriented researchers. Less technically saavy members of the team referred to them as *"like Google"*, and consulted them when looking for something.

We interviewed a few of these translators: users who monitor new applications, and are constantly trying new tools. They have a reflective posture, and are interested in the implication of information tools. They experiment with new tools, constantly looking for the *right* one to do the job.

Most Biologists simply do not have the time or the patience to cope with experimental software, especially if they already have something they know will give results in a way they are familiar with. They need this familiarity to leverage their experience with *their* data, and decide based on intuitions which may be hard to articulate during the exploratory process.

When dealing with this exploratory analyses, Biologists rely on their practice and working context. They rarely use the options available in the online analysis form, nor do they customize them. They rely on the default settings, then refine and filter the results. The visual aspect is thus important as it lets them gain a feeling of *'rightness'* that can be challenged by comparing the results from different platforms. Many consider these results as *"only theoretical"*, to get an idea. As one Biologist explained: *"if it's really important I'll do it at the bench"*.

Use of existing tools (history and bookmarks)

Few analysis results are kept as is; they are saved only when they seem interesting, after which they are printed and pasted in a notebook or filed on the computer. Few Biologists actively use bookmarks. We observed many huge bookmarks lists, yet Biologists rely on emails, physical Post-it notes, or comments in notebooks to record and re-visit interesting web pages. When they next want to reach previously bookmarked sites, they rely on search engines. Bookmarks lists are difficult to browse and Biologists rarely remember where they actually saved their URLs.

We found that Biologists preferred to replay a path rather than using history to access a specific page. We did not observe a single use of history and many users were not even aware of its existence. Rather, participants often repeated long navigation paths to access well known web pages. Either when looking for a specific page, or during a protocol that required going from one distant web page to another.

Summary

Biologists use the Web to gather information and explore hypotheses. They use scientific portals to gather data, run analyses or validate results. As Biologists run complex analyses, they must make decisions about the steps to do next. Such decisions, to a large extent, rely on experience and intuitions that are difficult to formalize. For this reason, researchers are quite conservative regarding the tools they use. They also did not find convenient tools for optimizing their navigation. Existing browser tools such as bookmarks or history are insufficient, and automation workflow systems require too much advanced planning.

With PageLinker, we focused on supporting Biologists workflow by taking into account the following issues:

- Diversity of data: Biologists must constantly transform data to fit the tools they are using, switching among data versions, formats or names.
- Lack of transparency: Biologists using workflow systems have difficulties with identifying the source of errors in the analyses.
- Uncertainty about the value of an analysis: Biologists do not know the underlying algorithms used by the analysis tools, making it hard to compare the values of different results.
- Instability of the Web: Biologists must handle changes in web pages, error pages or heavy server loads.
- Concerns about unknown methods: Biologists never know whether learning how to use a new tool will provide results that are worth the time invested.
- Uncertainty about the importance of an analysis: Biologists have a hard time deciding whether a result is worth keeping and will have value in the future or not.
- Revisiting pages: Biologists have problems re-locating pages they visited and rely on search engines or long navigation trails rather than on history or bookmarks.

- Deciding where to keep a web page: Biologists must decide how to save a page and where to store it when they decide to keep it, however at this particular moment they may not know the best way to file this information.
- Organizing the stored pages and results: when dealing with bookmarks or saved files, Biologists must regularly re-organize and prune the information they save to avoid overload and be able to find the valuable information in the future.

6.3 designing pagelinker

Based on the interviews, we decided to focus on supporting Biologists' web analysis activities. We wanted to create a tool that fit within their existing work patterns, so they could use familiar work practices and their own data without being forced to pursue additional tasks. We based the design on our observation that Biologists use *copy* to extract data from one web page and *paste* it into an analysis form[1], thus identifying which pages should be associated together.

PageLinker helped us explore how to balance automated and manual capture. Can manual capture be more lightweight? Can automated logging be closer to human tasks?

6.3.1 *Initial design choices*

We selected the Firefox web browser because it offers an extension architecture, it is available on Mac OS X, Linux, and Windows and was already used by half the Biologists in the study. Installing a Firefox extension is easy: users only need to click on the link of the extension they want to install.

PageLinker contextualizes Biologists' browsing experience by providing, in a sidebar, links that have proven useful after visiting the current web page (Figure 43). PageLinker automatically generates links by tracking copy and paste events between pages. Later versions of PageLinker also allowed users to create these contextual bookmarks manually, with better feedback and control.

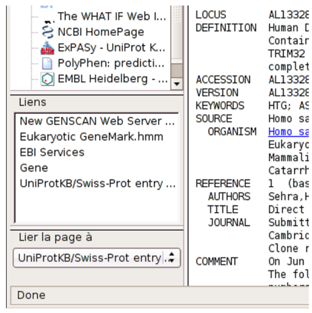


Figure 43: PageLinker's contextual bookmarks.

6.3.2 *Iterative Design of PageLinker*

Phase 1: Initial implementation

The first version of PageLinker focused on creating links automatically, based on the user's cut, copy and paste actions. PageLinker overrides copy, cut and paste events: When a copy or cut event is detected, it records the page (title, URL, and date) and, as soon as a paste event is detected, creates a link between the two pages. The copy (or cut) page thus points to the page where the paste occurred.

Our interviews and workshops indicated that Biologists rarely use the exact output data from one page when they need to fill out a new form. Instead, they usually edit the data, either to address incompatible data formats or to refine their request. We link the page of the most recent copy event to the current paste page, without considering the

1 I use the term "form" to refer to pages that require the user to enter data. Some of these forms also generate data.

contents of the clipboard. We can thus accommodate the *equivalent objects* mentioned earlier, where the data formats are different but, from the biologist's perspective, the content is the same.

PageLinker uses XUL[2], JavaScript[2] and RDF[2]. The new definitions of copy, cut and paste items from the menus are implemented with XUL, an XML-based language used to define interfaces. JavaScript handles user interface actions and manages data. PageLinker overrides the clipboard shortcuts events by grabbing Ctrl-C/X/V on Windows and Linux or Cmd-C/X/V on Mac OS.

PageLinker uses RDF to represent the network of contextual bookmarks. A collection of RDF statements represents a labeled, directed graph. Figure 44 shows the graph illustrating a link between two pages, Genscan[3] and Blast[4]. Each page is a node pointing to the pages it is related to. Since RDF allows only simple oriented graphs, our structure is redundant for bi-directional links.

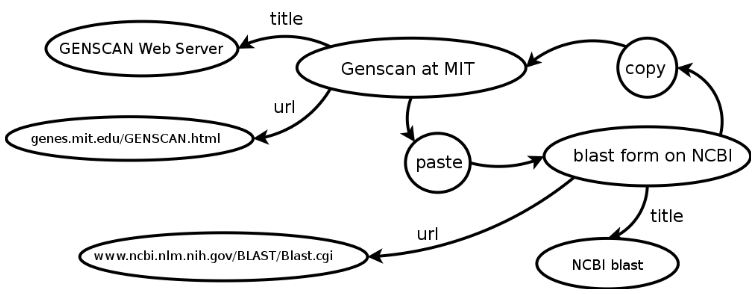


Figure 44: RDF graph outline of a bi-directional link between Genscan results (copy) and a Blast form (paste).

Each page points to its descriptors, e.g., title and URL, as well as a copy node and a paste node. The copy node points to the list of pages where data was copied from the current page and the paste node points to the list of pages from where data was pasted into the current page. The RDF is queried through a template-based request language supported by XUL in order to map the contextual bookmarks display and the RDF file. When the RDF is modified, its corresponding UI component is automatically updated.

Linking

Copy and paste events are good descriptors of connections between web pages. However, they do not describe how we should associate the pages together. Figure 45 illustrates three different elements of a URL we could use to link pages together.

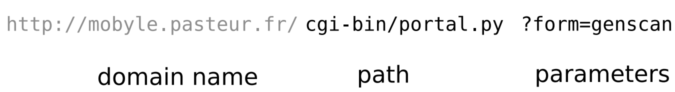


Figure 45: A simple URL schema.

---
2 See: http://developer.mozilla.org/
3 Genscan identifies gene structure in DNA sequences.
4 BLAST finds regions of similarity between biological sequences.

If we use the entire URL, the result is too restrictive: we get a large number of pages with only minor variations among them. If we use the domain name, i.e., the main site at the top of a hierarchy of web pages, we only get the main site and lose all of the interim browsing or searching the user has done. PageLinker uses the full URL, minus the query string (or parameters). The resulting contextual bookmarks are specific to a particular web form, rather than a particular result or the whole server.

Iterative design based on user feedback

We worked with Biologists from the Institut Pasteur using participatory design techniques to generate ideas for PageLinker. We tested PageLinker v0.1, with six Biologists who installed it and provided feedback via interviews and direct observation. The design was as simple as possible: links were based on automatically captured copy-paste events and users interacted with PageLinker via the Shortcuts menu (zoomed in figure 46).

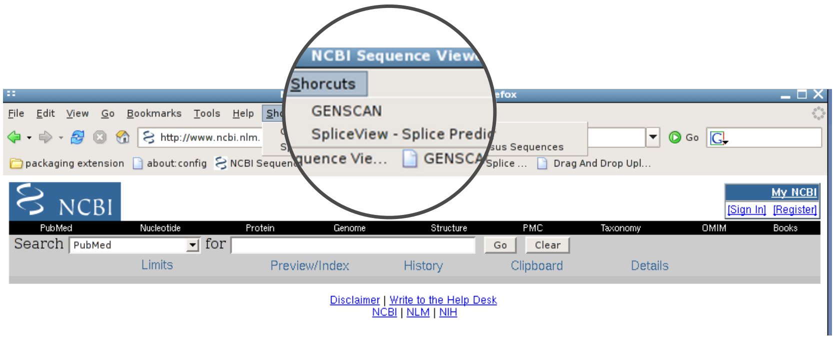


Figure 46: Shortcuts contextual menu, PageLinker v0.1.

At this point, some users discovered how to use PageLinker to manually add links between pages. This is an example of co-adaptive behavior where *"users both adapt to the available technology and appropriate the technology"* (Mackay, 1990b). They used the *Ctrl-C* and *Ctrl-V* shortcuts on pages without entry forms, e.g., between one page with press reviews and another with the referenced newspaper articles or between an application form and the relevant documentation page.

Although using control-keys was fine for some users, it adds unnecessary steps. Users would usually decide to link back to a previous page only after they had successfully identified an interesting subsequent page. Using the copy-paste required returning to the previous page to copy and then come back to the target page to paste.

Based on this feedback, we conducted a participatory design workshop (Figure 47) to explore simpler ways to create links between pages. We worked together with the Biologists to create video prototypes that envisioned scenarios for linking to a desired destination from a previous page. We created prototypes of three linking strategies: via open pages or tabs, via the last visited page and via the global history.

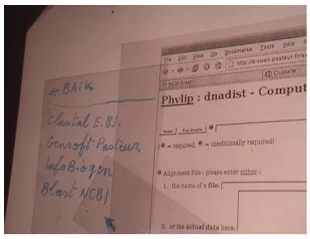


Figure 47: Video prototype of link creation.

PageLinker v0.2 implemented all three methods. We added a *"link to"* menu to the toolbar (Figure 48) that presents a list of all the browser's open web pages (both on tabs and in other windows) and the seven most recently visited websites from the global history. Links are sorted by time, similar to Firefox's *Go* menu. Selecting any of these creates a link from that page to the current page.

PageLinker v0.2 also created a reverse link, from the current page to the one just selected. We reworked the *Shortcuts* contextual bookmarks menu to separate links by direction. One list presents links *to* the current page (either via copy-paste or direct selection). The other list presents links *from* the page. Links on both menus were ordered by recency. Based on user comments, we also added the ability to delete a contextual bookmark by right clicking on the corresponding menu item. After one week of use, we observed that the menu appeared too complex to Biologists and was redundant. It was hard to navigate within the bidirectional links presented in two different menus and users did not notice they that could delete them.

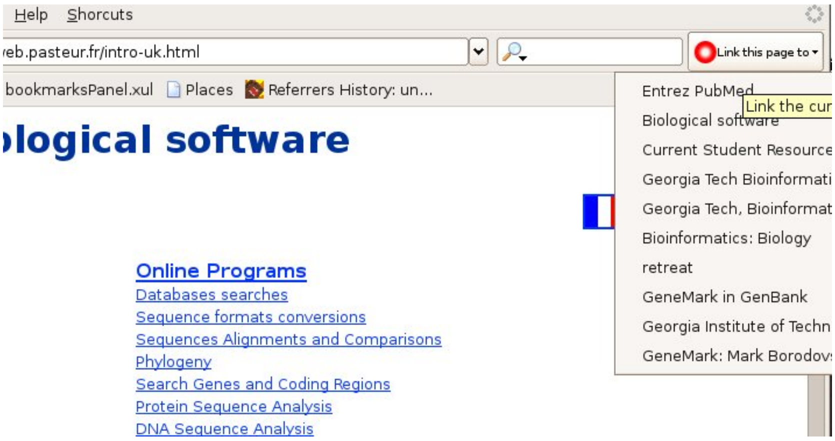


Figure 48: Adding explicit linking (v0.2)

PageLinker v0.3 simplified the linking menu to include just the last visited pages. We also classified bidirectional Shortcuts by order of recency. Finally, we integrated contextual bookmarks and linking via the bookmarks sidebar (bottom left of Figure 49). Most users quickly be-

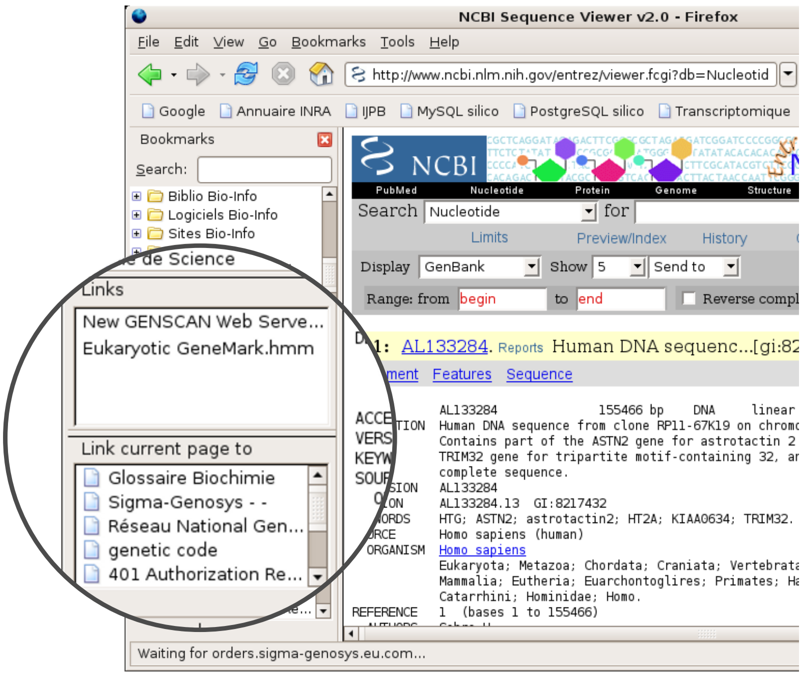


Figure 49: Making contextual bookmarks visible (v0.3)

gan using the bookmark sidebar instead of the menu. They found it useful to have their contextual bookmarks visible immediately upon changing pages, without needing to click on the menu list, since contextual bookmarks change from one page to the next.

Table 4 summarizes the four versions of PageLinker, including the types of links, how contextual bookmarks are created and how to access PageLinker. We used PageLinker v0.3 in the field evaluation presented in the following section. We then released PageLinker v1.0 which included a minor modification: To avoid confusion between the contextual bookmarks list and the linking list, we converted the *"link to"* list into a menu.

Release	**Link type**	**Creation**	**Access**
v0.1	directed, insuppressible	copy/cut-paste	popup accessible from the menu bar
v0.2	bidirectional, suppressible	copy/cut-paste; popup accessible from the menu bar showing the opened pages and the last seven visited	popup accessible from the menu bar
v0.3	bidirectional, suppressible	copy/cut-paste; list of the last visited pages in the *Bookmarks* side-bar	list in the *Bookmarks* side-bar
v1.0	bidirectional, suppressible	copy/cut-paste; popup accessible from the *Bookmarks* side-bar showing the last visited pages	list in the *Bookmarks* side-bar

Table 4: Design choices associated with successive versions of PageLinker.

6.4 pagelinker evaluation

Evaluating history-based tools such as PageLinker poses interesting methodological challenges with respect to validity (Cook and Campbell, 1979). We considered the following possibilities:

1. A laboratory experiment is easier to control but poses external validity problems. Our fieldwork indicated that Biologists' navigation and bookmarking behavior on unfamiliar tasks with artificial data might differ greatly from their behavior with familiar data and resources, making the results potentially meaningless. Also, users cannot fully leverage their personal knowledge in a laboratory experiment nor can they take advantage of their episodic memory. We were also interested in gathering realistic adoption and usage data for PageLinker: not only measuring performance advantages, if any, but also observing how user behavior evolves over time and whether users make the tool part of their repertoire.

2. An uncontrolled field study has greater external validity but is very difficult to control. Longitudinal field studies require extensive logging and extensive data analysis, especially if the participants' environment is not modified. Long-term monitoring also raises serious privacy issues and risks interfering with Biologists' confidentiality agreements. For example, some Biologists asked us to stop recording during the interview if they thought

we might see confidential data. These Biologists would not have been willing to participate in long-term automatic recording of their activities.

Biologists also alternate between periods of intense on-line data analysis and periods of laboratory research. At any point in time, individuals may be out of phase with each other, depending upon who is writing a paper, running an experiment, or analysing data. This diversity complicates any comparisons and analysis of activity logs. For example, it would be difficult to tell, for any one subject, whether a decrease in pages visited was due to PageLinker or an overall change in research activity. It would also be difficult to compare people who were at different phases in their work.

3. A limited time-series field experiment, or quasi-experiment (Cook and Campbell, 1979), offers the optimal compromise, with the external validity of a field study and most of the control offered by a laboratory experiment. Because we wished to compare PageLinker's navigation performance to existing browsers, it made sense to alternate PageLinker with the user's usual browser. This allowed us to track changes in use over time, based on realistic tasks performed in the user's real work setting, together with their existing bookmarks and other revisitation techniques. We chose this third option to evaluate PageLinker.

6.4.1 *Method*

Participants

Twelve Biologists or Bioinformaticians (9 men and 3 women between 20 and 40 years old) working in four research institutes (the Institut Pasteur, Génopole, Université Paris 5 and INRA) participated in the study. All were Firefox users with browsing and bookmarking experience. Two had also participated during the participatory design phases (Post-hoc analysis did not show significant differences. between their results and those of the other participants.)

Apparatus

HARDWARE: Participants used their usual browser with their own bookmarks and history, on their own system: 5 Mac OS X users, 4 Windows users and 3 Linux users.

SOFTWARE AND LOGGING: We used PageLinker v0.3 and developed Navtracer[5] (Roussel et al., 2006), a Firefox extension that logs user interactions with the browser. Navtracer ran on existing versions of Firefox (from 1.0 to 2.0) and could be installed and disabled rapidly in each user's browser without requiring special knowledge. This allowed us to minimize disruptions and let participants continue using their standard bookmarks, history and other Firefox extensions in both conditions in the evaluation.

To protect privacy, Navtracer does not begin logging automatically. Rather, users press a start button added to the Firefox window and

5 http://navtracer.mozdev.org/

fills out a form describing the experimental condition. This gives users full control of logging: they can pause, resume or stop at any time.

When Navtracer was first installed, we showed users how to enable and disable logging and where the CSV log file was stored. They were invited to delete the file or modify its contents if they had concerns about what had been logged. The extension registers various event handlers to detect the opening or closing of tabs and windows and the acquiring or loss of focus. It also tracks web-page changes and the relations between them via the page referrer. Switches between windows or tabs are also recorded.

Event handlers append log data to a plain text file stored in the user's profile folder. Timestamps are systematically added to every record. Navtracer also logged PageLinker events such as link creation and usage of created links.

Scenario design

We based the experiment scenario on our observations of common tasks and navigation patterns, including:

- **Search**: web search engines, biological databases, directories;
- **Parallel exploration**: same analysis with different programs;
- **Results comparison**: comparison of results;
- **Analysis**: visual scan of results to check validity and pertinence;
- Biological **links directory**: scanning for options; and
- **Repeated path**: access the same page.

We created a scenario with five related subtasks (Figure 50) with the aid of two Biologists from the Institut Pasteur. The scenario had to be short enough (between 15 and 20 minutes) so that it would not be too time consuming for participants, but still be representative of their tasks and understandable for every specialty. The five tasks illustrated aspects of web navigation presented above. The scenario was open and we encouraged participants to use their usual websites to perform the tasks. The websites presented here were the most commonly used, taken from different servers to illustrate the resource diversity faced by Biologists. The five tasks were:

1. Database search: Find the gene corresponding to human muscular dystrophy and choose the nucleotide sequence attached to the TRIM32 gene (most participants used NCBI[6]).

2-3. Parallel exploration: Analyze the nucleotidic sequence with two different tools, e.g., Genscan[7] and Genemark[8], to predict what the peptide sequence would be.

4. Comparison: Compare the two predicted sequences, e.g., using bl2seq[9] to check if predictions are reliable (result R1).

6 http://www.ncbi.nlm.nih.gov/
7 http://genes.mit.edu/GENSCAN.html
8 http://exon.gatech.edu/GeneMark
9 http://bioweb.pasteur.fr/seqanal/interfaces/bl2seq

5. Analysis and visual scan: Analyze one of the predicted peptide sequences to find regions of local similarity with other sequences with Blast[10] (result R2). The goal is to find species other than homo sapiens that express the same protein with a high degree of confidence and are interesting for researchers looking for a related analysis or literature.

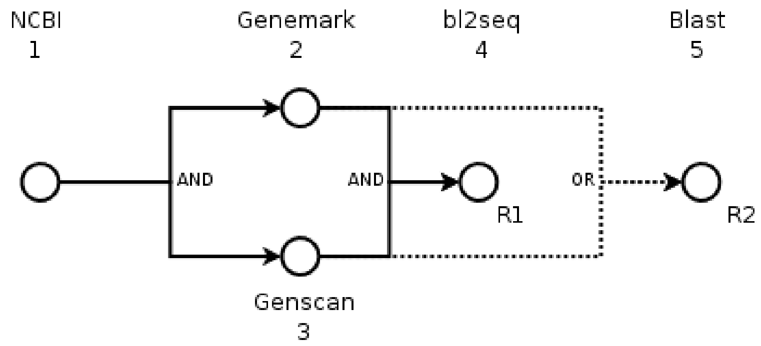


Figure 50: Scenario structure: Task 1 is performed first, followed by tasks 2 and 3 which are often performed in parallel. Task 4 is possible only after tasks 1-3 are complete and produces R1. Task 5 may be conducted independently after tasks 2 or 3 and produces R2.

Procedure

We used an ABAB within-subjects design, with one factor:

- **FireFox**: Firefox browser with logging,
- **PageLinker**: Firefox browser with logging and PageLinker.

Users alternated between the PageLinker and the unmodified Firefox conditions at one-week intervals. Users kept their history, standard bookmarks and other Firefox extensions when changing conditions. This allowed them to work with their own real data settings instead of an empty initialized browser or one with artificial bookmarks and history the user was not familiar with.

Our goal was to collect data over long periods without extensive logging, so we asked them to follow the five-task scenario described above. Full counterbalancing of tasks across subjects would have been difficult, because PageLinker requires a first visit to websites to create the contextual links. (In other words, the unmodified Firefox condition must run first, for all subjects.) We used an ABAB procedure, repeating each condition twice, to dissociate learning effects as much as possible from improvements due to PageLinker.

During the evaluation, each session was separated from the next by an interval of at least a week. Based on our previous observations, it appeared that seven days, including a week-end, should be long enough for participants to partially forget the exact details of what they had done during the previous session. This reduced the learning effect and is also representative of Biologists' typical behavior: They

10 http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi

frequently perform a series of tasks for one purpose and then repeat it after days or weeks of performing other tasks.

I visited each of the participants in their lab once a week for a month. During each visit, participants were asked to perform the same scenario. In the first session, I introduced PageLinker and invited the Biologists to use it freely until they felt comfortable with link creation and use. This training period lasted between 10 and 15 minutes. I first showed participants how to create links either by copy/paste or the menu list. They were then free to try creating lists between any pages they liked. I finally asked them to determine pages they thought were related to each other and to create links between them using the two techniques. In case they had no idea of what to link, I suggested that they create links between pages they had visited during a recent break so as to avoid conflicts with our scenario. (Note: this occurred primarily during the first session, with a few Biologists who had not done this type of analysis for a long time.)

I presented the standard scenario, explaining its biological purpose and the necessary steps to achieve it. During this phase, I avoided mentioning any particular online tools and encouraged participants to use their favorite applications, portals or search engines. Our only guidance consisted of reminding them of the next task after they completed the previous one. Tools and portals were only suggested if they did not know what software was appropriate for a task or if their usual application server was down. (Note: a server went down twice in the course of the month-long study and ran very slowly approximately once per participant.)

The PageLinker extension remained installed during all phases of the study, but was invisible to users during the Firefox-only conditions. In the latter case, it simply logged the creation of links between pages via copy/paste, as a conventional history tool. To protect privacy, we disabled the logging extension after each session. I also asked users if they wanted PageLinker to be disabled between sessions: All decided to keep it. To avoid interference between contextual bookmarks created during the experiment and non-experiment phases, we stored the contextual bookmarks in different files.

6.4.2 *Predictions and Hypotheses*

Based on feedback from our first field release and our personal use of the extension, we predicted the following results: PageLinker would generate fewer page loads, fewer clicks per task and reduce time spent on each task. We also predicted that with PageLinker, the majority of links would be created on the first visit to each relevant website. Since we had designed Prism iteratively with users and responded to their requests during the design of PageLinker, we also expected our participants to be mostly satisfied with the design and interaction techniques used in the experiment.

6.5 Results

This section presents both the results from the experiment and its limitations. We gathered quantitative results through logging, and qualitative results by observing users and sending a questionnaire three months after the experiment.

6.5.1 *Quantitative Results*

PageLinker performed significantly better than the unmodified Firefox browser with respect to the following dependent variables:

- task completion time was 28% shorter ($p < 0.01$)
- 22% fewer clicks occurred ($p < 0.01$)
- 38% fewer pages loaded ($p < 0.01$)

If we focus more specifically on the limited time series (Figure 51), we observe the same pattern for clicks and page loads, although the difference is only significant for the number of page loads. The decreased number of page loads corresponds to the biologist seeing 38% ($p < 0.05$) fewer pages during a typical task. Although there is an overall learning effect, i.e. Biologists become more efficient running the tasks in the scenario over time, there is also a significant effect of PageLinker. Columns two and four (PageLinker conditions) are always more efficient than columns one and three (Firefox-only conditions).

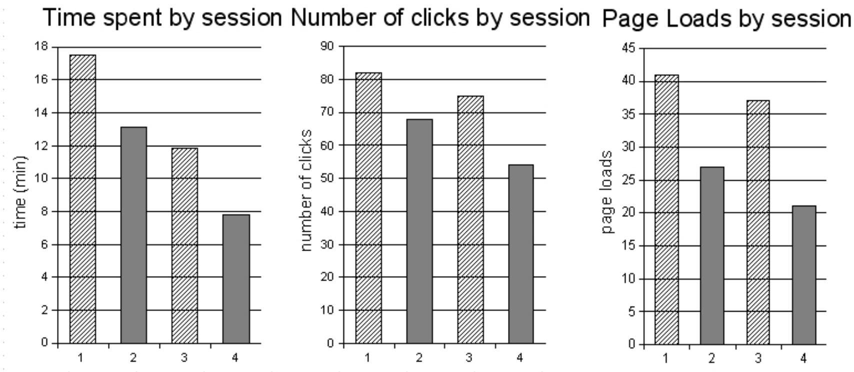


Figure 51: Evolution of time, clicks and page loads over sessions. Columns 1 & 3 are Firefox only, columns 2 & 4 are PageLinker.

The overall number of links created is not significantly different over the four sessions: A mean of 20 contextual bookmarks were created during the first session and 12 during each other session. Participants never had too many contextual bookmarks, with the corresponding risk of overload since the use of contextual bookmarks increased linearly with the number of created links $F_{1,11} = 8.73$, ($p < 0.05$). In case of overload, we should have observed a decrease in use as more bookmarks were available. In summary, these results suggest that PageLinker actively facilitates page revisitation:

- fewer page loads implies that users visited fewer search websites and transition pages,
- fewer clicks implies they used fewer transition pages, and
- fewer time per session implies that participants took less time to complete the five tasks of the scenario.

6.5.2 *Limitations of the experiment*

Dissociating PageLinker effects from learning effects is complex when interpreting the time spent on the scenario and the number of clicks. Time is highly correlated with external factors, such as the current server load. For example, users may wait more than five minutes for a Blast result from NCBI, if the servers are heavily loaded.

Another potential problem is assessing the correlation between the number of contextual bookmark links and their use. A month-long evaluation may be too short to overload the contextual bookmarks menu. We expect that the recency classification we use, which only shows the most recently used links, should reduce the overload effect, but we would need a much longer study to find out.

PageLinker can only reduce hyperlinks clicks, not the clicks needed to fill in forms. Nevertheless, the logger counted all clicks indiscriminately, whether they occurred on links or on forms. PageLinker thus accounted for only a small percentage of the overall number of clicks and the reduction was indistinguishable from noise.

Finally, from what I could observe during the experiment, participants mostly used copy and paste to create links. However, this observation may be biased since the scenario was designed around tasks where copy and paste happen.

6.5.3 *Longitudinal use*

The study demonstrated that PageLinker's contextual bookmarks improve web page revisitation and that, unlike history and bookmarks mechanisms, they are less prone to information overload. After the evaluation, we released PageLinker v1.0 which modified how contextual bookmarks are created. Figure 52 shows that the *link to* list has been changed into a menu.

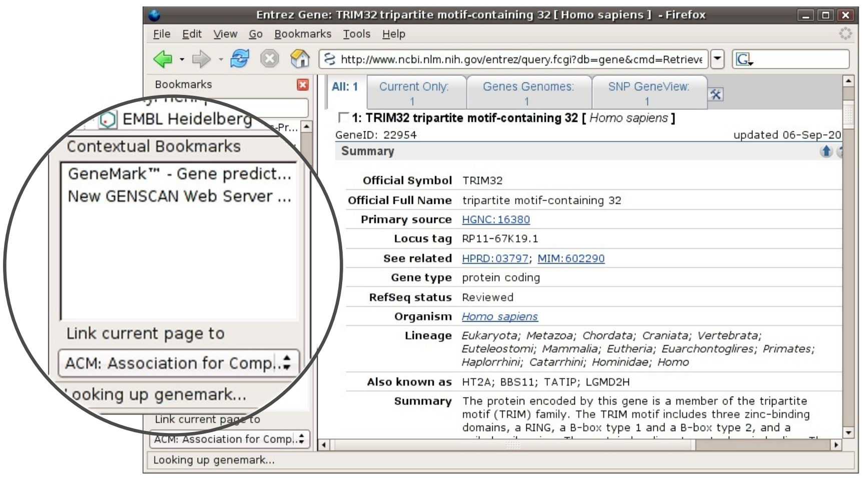


Figure 52: Current PageLinker version (v1.0)

Three months later, we sent the participants a questionnaire. Of the 12 participants, two had changed institution and did not answer, two had changed browser or workstation without re-installing PageLinker and eight still used PageLinker. The questions in the questionnaire are presented in Table 5. Participants found PageLinker easy to use (to create and use links). Their opinion on the usefulness of links varied more but was mostly positive.

Question	Mean	SD
How usable is the link creation?	4.33	0.87
How usable are the created links?	4.44	0.73
How useful are the links created?	3.56	1.24

Table 5: Responses to the questionnaire using a five point Lickert scale: From 1 = not at usable all to 5 = very usable.

Our observations of PageLinker use highlighted several ways in which it improved users' workflow and how they manage their links. When interruptions occurred, such as people asking questions or phone calls, PageLinker helped users reorient themselves when they returned to their task. By seeing the links to and from the pages they were currently visiting, participants could more easily remember what they were doing and what their goals had been. We also observed that it helped users in case of server slowdown or breakdown. They began to keep alternate links to the same program on different servers, something they never did with standard bookmarks because it would have generated an unacceptably large number of bookmarks.

6.6 DISCUSSION

We began by addressing a very specific problem that Biologists faced, linking data output pages to data analysis pages. After we released the first version of PageLinker, users appropriated it, revealing the need for a more general contextual bookmark tool. Users sought ways to associate pairs of web pages in order to facilitate future navigation within groups of previously visited pages.

6.6.1 *Implicit interactions*

Our observations highlighted Biologists' use of copy-and-paste to transfer information during online analyses. Copy-and-paste events are meaningful from the users' perspective, they signal key transitions in the analysis process, but they are also easy to interpret for computers (or developers). We thus leveraged these events to create implicit interactions, interactions which have a primary goal but can generate side-effects, in this case, linking web pages.

Ju (2008) describes implicit interactions along two axes, *"the level of attentional demand the system places on the user and the balance of initiative taken by system on behalf of the user"*. We can extend this description to capture mechanisms that stay in the background (*Ready-to-hand*) and come to the foreground (*Present-at-hand*) in the case of breakdown: a mis-created link, or when users want to create links between pages where copy-and-paste would not otherwise happen.

Copy-and-paste interaction makes the capture mechanism transparent to users: it is easy for them to create a mental model of what is happening. A copy action selects the source page of the link, a paste action selects the target page of the link and creates the link. This led users to adapt their use: they created links between pages without forms. Users knew that they created links during copy-and-paste, but they did not have to focus on capture since it was a side-effect of their actions, *it was implied.* Such implicit interactions offer a balance between automated and manual capture.

6.6.2 *Contextual traces*

In his classic article, "As We May Think", Bush (1945) argues that the human mind operates by association, connecting items into a web of trails. In the spirit of his Memex idea, PageLinker allows users to *"build a trail of interest through the maze of materials available"*.

PageLinker lets users create contextual traces: web pages are related to each other without the hierarchical order imposed by regular bookmarks. Linking web resources while navigating is a powerful way to retrieve information and to let users reflect on their activity as they explore alternatives. Contextual bookmarks provide two main benefits: users avoid "history overload" and the links users created can easily evolve which makes them more robust to changes in the structure of the web sites involved.

As users visit more pages, they create more contextual bookmarks, and thus the potential for "history overload". PageLinker avoids this by distributing bookmarks over the whole set of pages that users have visited, rather than concentrating them in a central place. Unlike classic bookmarks, with PageLinker, when users do not visit a page anymore, they are not bothered by the contextual bookmarks associated to this particular page. With PageLinker, when users do not visit a page anymore, they are not bothered by the contextual bookmarks associated to this particular page. Unused contextual bookmarks are not lost but they do not get in users' way as classic bookmarks do. If a user comes back to a page with contextual bookmarks, they will still be available.

The simple model behind PageLinker allows users to easily re-link pages together whenever the structure of a website changes. Unlike automation tools, which would require users to redefine a whole chain of actions, PageLinker allows users to simply re-create links as they revisit web pages. Furthermore, if users' areas of interest change slightly as they visit new sites, they do not have to prune their bookmarks and explicitly create new ones. With PageLinker, they simply create a few additional links implicitly and forget about obsolete ones. These contextual traces associate related pages to each other without the hierarchical order imposed with regular bookmarks or other link organisers.

6.7 Conclusion

In this chapter, I discussed the design, evaluation and implications of PageLinker, a contextual bookmarking system. PageLinker provides a solution to balance automated and manual capture mechanisms. It significantly improves Web browsing by letting users save their navigation paths and re-use them later. These contextual bookmarks do not follow chronological or hierarchical organization, but are grouped relative to each other, allowing users to jump from one page of interest to a related one.

PageLinker helps users manage the websites they visit regularly and reflect on their past browsing experience. The benefits of PageLinker are not limited to Biologists analyzing data. Such browsing patterns, e.g., exploration, comparisons and multiple visits over time, are similar to those of users organizing trips or exploring potential purchases on e-commerce web-sites.

6.8 SYNTHESIS

The contributions of this chapter are:

- **Study of Biologists' web browsing** highlighted the repetitive yet different workflows of Biologists as they analyze data online.
- Design of **PageLinker** a contextual bookmarking tool.
- **Controlled field experiments** over time as a method to balance internal and external validity of evaluations.
- **Evaluation of PageLinker**, a field-experiment showed that contextual bookmarks significantly improve browsing.
- Identification of **implicit interaction** as a mechanism balancing automated and manual capture.
- A discussion of the **benefits of contextual traces**, which provide a distributed model of history.

7 CONCLUSION AND PERSPECTIVES

This chapter revisits the contributions of the thesis and highlights future directions they inspire. Ultimately, this research should help consider personal information management practices from a different point of view. Designers should not only consider how to best record information, but also how knowledge workers filter, save and synthesize information, in order to support their reflective practice.

7.1 THESIS OVERVIEW

The dissertation explored how to support Biologists' information management practices in a context of social and technological changes. Rather than providing access to large quantities of captured information, I focused on supporting the ways in which Biologists reflect on information as they manage it. Personal information management tools should not only support retrieval but also how users engage in managing information, i.e., save it, transform it and articulate it.

I investigated how users manage familiar information along three axes:

1. *information capture*, i.e., capture can be selective or exhaustive,
2. *users' involvement* in managing information, i.e., users can be active or passive,
3. *users' attention* when saving information, i.e., users can focus on saving information or the capture can happen in the background.

The field studies described in chapter 3 and 4 highlight the benefits of selective capture. It is not that tools capturing detailed logs of users' activity are useless, but rather that filtered and transformed information helps researchers reflect on their activity.

With the design of Prism, we identified active saving by users as an important element of information management that helps users make sense of the information they manipulate. When Biologists manage information, they filter, synthesize and frame it in the context of their other activities.

Finally, with PageLinker, we focused on selective capture by active users. The capture can happen in the background of attention, when biologists do copy-and-paste, or in the foreground as they consciously create links between two pages. This mechanism let users focus on their ongoing activity but still control the information saved.

When designing tools supporting lightweight reflection on familiar information, one should consider the three dimensions. What information to capture, how to engage knowledge workers in managing information and how to incorporate reflection in their ongoing activities without disrupting their workflow.

7.2 CONTRIBUTIONS

The thesis provides contributions at three levels, with empirical findings, adaptable technologies and some theoretical perspectives. I summarize them here, following the dissertation plan.

7.2.1 *Information Management as a reflective practice*

The results from the field studies at the Institut Pasteur and at INRA highlighted the reflective practice of researchers as they manage familiar information. When writing in notebooks or organizing documents, researchers select the information they save and articulate it in the context of their ongoing activities.

Depending on the context, researchers save information over different media, and the media they use influence how they save information. If paper notebooks persist it is not only due to their flexibility but because their chronological structure and physical constraints impose a disciplined writing.

7.2.2 *Prism, supporting reflection*

With Prism, a hybrid notebook acting as a technology probe, I explored how to support reflection. Prism integrates paper and electronic notebooks as well as the users' desktop and web activity. Once the information was at hand, participants created master notebooks, to think and reflect about their practice, not necessarily capture information. Indeed, unlike to hypotheses of total recall, participants preferred building their own account of activity rather than taking advantage of automated logging.

While it initially seemed that participants could leverage Prism to save information in the 'right' place, it appeared that related information coming from different sources helped participants to frame the relevance of information within the wealth of traces available. In this context of rich information coming from different sources, Prism evolved from a desktop to a Web application aggregating heterogeneous streams of familiar information.

7.2.3 *PageLinker, supporting lightweight capture*

With PageLinker, I explored how to build traces mechanisms that balance manual and automated capture. PageLinker provided contextual bookmarks that are specific to the page currently visited. Its coadaptation demonstrated the efficiency of implicit interactions: interactions having underlying meaning or effects. Users, aware of interactions implications, modified their behaviors to leave traces or not, while using their every-day tools.

PageLinker's field evaluation showed that contextual traces improve Web browsing and can limit history overload by presenting history in context. Similar patterns, repetitions, stand out but do not overload or interfere with users as they are only presented in similar contexts of use.

7.3 Limitations and perspectives

I emphasized in the dissertation the role of lightweight reflection when Biologists manage familiar information. I hope it may help to reconsider reflection as a continuous activity that happens in the course of action, i.e., as knowledge workers manage information, not only as a post-hoc activity when knowledge workers step back and think about their past actions.

However, the results from the initial field studies are mainly qualitative and may be limited to the studied settings. A quantitative study across a larger population could have helped compare participants behaviors, but it would have been difficult to assess the significance of the results. I rather chose to use Grounded Theory to analyze the qualitative data within a clear frame and seek the emergence of phenomena to confirm or infirm the analysis. I am looking forward to conduct further studies in order to compare the observations from the Institut Pasteur and INRA with observations of information management practices in other settings.

The qualitative approach may have limited the results to the context described, but it provided insights that could be applied to other populations, and should raise questions regarding the way people interact with familiar information. I highlight here a few perspectives based on the limitations of the thesis research.

7.3.1 *Selective traces in applications*

Since Prism collected some information automatically, after a few months of use, the navigation became less fluid as Prism contained more information. PageLinker limited history overload by capturing a limited amount of information and displaying history traces in context. An alternative solution would be to implement progressive loss of history in applications.

While some predicted that by 2010 the evolution of storage space would make it possible to store extensive amounts of data, such as the log of ones digital life (Bell and Gemmell, 2007), there is still a limit to what can be stored. In September 2009, my two external hard drives and two laptop drives amount for approximately 1TB[1] and are full, notwithstanding that a lot of my data is now in the clouds.

It seems more realistic to assume that digital space will always be limited. Furthermore, most of us do not want everything to be recorded for ever. One approach would be to lengthen the time frame of logs but make them more sparse. Santry et al. (1999) devised a file system which versions changes with some intelligence, so that only main versions of files are kept. It would be interesting to explore similar systems for emails, web browser applications or other information management systems. In such a case, the question becomes how should we design traces mechanisms that incorporate decay?

1 TB: terabyte, 1 terabyte equals 1000 gigabytes.

7.3.2 *Paper-based management of digital information*

Paper is a good candidate to incorporate subtle information management techniques. The studies presented in chapter 3 and 4 described many strategies that participants developed to manage information on paper, in their notebooks, on their desk or in the laboratory. However, the way Prism integrates paper and digital information is still limited. The paper notes are simply uploaded and displayed on the computer.

Richer forms of interplay between paper and digital information could let users manage information over one medium or the other depending on the situation. An illustration would be the ability to create paper bookmarks of digital information. Whereas digital data must be filed immediately, tangible items offer more progressive and adaptable organizations. Tangible items can easily move from ephemeral to working or archived status and be grouped at will. Augmented paper scraps could help users offload cognitive charge and organize digital information on paper.

7.3.3 *Virtual Patina, designing for implicit traces*

On a longer time frame, I am interested in supporting lightweight reflection with contextual visualizations of interaction traces. Whereas we leave visible traces of our activities in the physical world, very few of our traces are visible online. Rather than creating dedicated applications to navigate in interaction histories, we could augment existing digital documents with a virtual patina, i.e., layers of interaction traces.

Virtual patina should be integrated in the existing systems people use and reveal the interaction they have with digital objects. In the spirit of Read Wear Edit Wear (Hill et al., 1992) or PageLinker, such traces could reveal users' past activity in context. They would add value to the digital information people manipulate by creating unique layers of personal information on top of digital objects. Designers could leverage these traces created implicitly to build an awareness of colleagues activities without intruding into their personal information space.

BIBLIOGRAPHY

Abrams, D., Baecker, R., and Chignell, M. (1998). Information archiving with bookmarks: personal web space construction and organization. In *CHI '98: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 41–48, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co. (Cited on page 77.)

Adar, E., Teevan, J., and Dumais, S. T. (2008). Large scale analysis of web revisitation patterns. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 1197–1206, New York, NY, USA. ACM. (Cited on page 77.)

Arnstein, L., Borriello, G., Consolvo, S., Hung, C., and Su, J. (2002). Labscape: A smart environment for the cell biology laboratory. *IEEE Pervasive Computing Magazine*, 1(3). (Cited on page 19.)

Baker, S. (1994). Mosaic-surfing at home and abroad. In *SIGUCCS '94: Proceedings of the 22nd annual ACM SIGUCCS conference on User services*, pages 159–163, New York, NY, USA. ACM Press. (Cited on page 76.)

Bardram, J., Bunde-Pedersen, J., and Soegaard, M. (2006). Support for activity-based computing in a personal computing operating system. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 211–220, New York, NY, USA. ACM. (Cited on page 18.)

Barreau, D. and Nardi, B. A. (1995). Finding and reminding: file organization from the desktop. *SIGCHI Bull.*, 27(3):39–43. (Cited on pages 3, 13, and 14.)

Beale, R. (2009). Usability and e-science. *International Journal of Human-Computer Studies*, 67(4):279 – 280. (Cited on page 19.)

Bell, G. and Gemmell, J. (2007). A. digital life. *Scientific American*. (Cited on pages 15 and 101.)

Bellotti, V., Ducheneaut, N., Howard, M., and Smith, I. (2003). Taking email to task: the design and evaluation of a task management centered email tool. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 345–352, New York, NY, USA. ACM. (Cited on page 72.)

Bellotti, V., Ducheneaut, N., Howard, M., Smith, I., and Neuwirth, C. (2002). Innovation in extremis: evolving an application for the critical work of email and information management. In *DIS '02: Proceedings of the 4th conference on Designing interactive systems*, pages 181–192, New York, NY, USA. ACM. (Cited on page 11.)

Bergman, O., Beyth-Marom, R., and Nachmias, R. (2003). The user-subjective approach to personal information management systems. *J. Am. Soc. Inf. Sci. Technol.*, 54(9):872–878. (Cited on page 12.)

Bergman, O., Beyth-Marom, R., Nachmias, R., Gradovitch, N., and Whittaker, S. (2008). Improved search engines and navigation preference in personal information management. *ACM Trans. Inf. Syst.*, 26(4):1–24. (Cited on pages 14 and 78.)

Bernstein, M., Kleek, M. V., Karger, D., and Schraefel, M. C. (2008). Information scraps: How and why information eludes our personal information management tools. *ACM Trans. Inf. Syst.*, 26(4):1–46. (Cited on page 13.)

Block, F., Haller, M., Gellersen, H., Gutwin, C., and Billinghurst, M. (2008). Voodoosketch: extending interactive surfaces with adaptable interface palettes. In *TEI '08: Proceedings of the 2nd international conference on Tangible and embedded interaction*, pages 55–58, New York, NY, USA. ACM. (Cited on page 23.)

Boardman, R. (2004). *Improving Tool Support for Personal Information Management*. PhD thesis, Imperial College London University of London. (Cited on pages 3, 11, 12, and 72.)

Boardman, R. and Sasse, M. A. (2004). "stuff goes into the computer and doesn't come out": a cross-tool study of personal information management. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 583–590, New York, NY, USA. ACM. (Cited on pages 14 and 16.)

Bondarenko, O. and Janssen, R. (2005). Documents at hand: Learning from paper to improve digital technologies. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 121–130, New York, NY, USA. ACM. (Cited on page 14.)

Bowker, G., Star, S. L., Turner, W., and Gasser, L., editors (1997). *Social Science, Technical Systems and Cooperative Work: Beyond The Great Divide*. Lawrence Erlbaum Associates, Inc., Mahwah, NJ, USA. (Cited on page 43.)

Buckland, M. K. (1991). Information as thing. *JASIS*, 42(5):351–360. (Cited on page 11.)

Bush, V. (1945). As we may think. *The Atlantic Monthly*, 176(1):101–108. (Cited on pages 13, 15, and 95.)

Butler, D. (2005). A new leaf. *Nature*, 436:20–21. (Cited on page 19.)

Catledge, L. D. and Pitkow, J. E. (1995). Characterizing browsing strategies in the world-wide web. *Comput. Netw. ISDN Syst.*, 27(6):1065–1073. (Cited on page 76.)

Chalmers (2004). A historical view of context. *Computer Supported Cooperative Work (CSCW)*, 13(3):223–247. (Cited on page 15.)

Chapuis, O. (2005). Gestion des fenêtres: enregistrement et visualisation de l'interaction. In *IHM 2005: Proceedings of the 17th international conference on Francophone sur l'Interaction Homme-Machine*, pages 255–258, New York, NY, USA. ACM. (Cited on page 61.)

Cockburn, A., Greenberg, S., McKenzie, B., Smith, M., and Kaasten, S. (1999). Webview: A graphical aid for revisiting web pages. In *OZCHI'99*. (Cited on page 77.)

Cockburn, A. and McKenzie, B. (2001). What do web user do: an empirical analysis of web use. *International Journal of human-computer studies*, 54:903–922. (Cited on pages 76 and 77.)

Cook, T. D. and Campbell, D. T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Houghton Mifflin Company. (Cited on pages 87 and 88.)

Costa-Cunha, P. and Mackay, W. (2003). Augmented paper and anoto stylus. In *IHM 2003: Proceedings of the 15th French-speaking conference on human-computer interaction on 15eme Conference Francophone sur l'Interaction Homme-Machine*, pages 232–235, New York, NY, USA. ACM. (Cited on page 22.)

Cubaud, P., Stokowski, P., and Topol, A. (2002). Binding browsing and reading activities in a 3d digital library. In *JCDL '02: Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, pages 281–282, New York, NY, USA. ACM. (Cited on page 16.)

Dirks, L. and Hey, T. (2007). The coming revolution in scholarly communications & cyberinfrastructure. *CTWatch Quarterly*, 3(3). (Cited on pages 3, 18, and 26.)

Dix, A. and Marshall, J. (2003). At the right time: when to sort web history and bookmarks. In Jacko, J. and Stephandis, C., editors, *Proceedings of HCI International*, volume 1, pages 758–762. Lawrence Erlbaum Associates. (Cited on page 77.)

Dourish, P. (2004). What we talk about when we talk about context. *Personal and Ubiquitous Computing*, 8. (Cited on page 15.)

Dourish, P. (2006). Implications for design. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 541–550, New York, NY, USA. ACM. (Cited on page 43.)

Dourish, P., Edwards, W. K., Lamarca, A., and Salisbury, M. (1999). Presto: An experimental architecture for fluid interactive document spaces. *ACM Transactions on Computer-Human Interaction*, 6:133–161. (Cited on page 16.)

Ducheneaut, N. and Bellotti, V. (2001). E-mail as habitat: an exploration of embedded personal information management. *interactions*, 8(5):30–38. (Cited on page 14.)

Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R., and Robbins, D. C. (2003). Stuff i've seen: a system for personal information retrieval and re-use. In *SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 72–79, New York, NY, USA. ACM. (Cited on pages 17 and 77.)

Engelbart, D. C. (1962). Augmenting human intellect: A conceptual framework. AFOSR 3223, Stanford Research Institute, SRI Project 3578, Air Force Office of Scientific Re- search, Menlo Park, CA,. Contract AF49(638)-1024. (Cited on page 15.)

Fass, A., Forlizzi, J., and Pausch, R. (2002). Messydesk and messyboard: two designs inspired by the goal of improving human memory. In *DIS '02: Proceedings of the 4th conference on Designing interactive systems*, pages 303–311, New York, NY, USA. ACM. (Cited on page 16.)

Fertig, S., Freeman, E., and Gelernter, D. (1996). Lifestreams: an alternative to the desktop metaphor. In *CHI '96: Conference companion on Human factors in computing systems*, pages 410–411, New York, NY, USA. ACM. (Cited on page 17.)

Flanagan, J. C. (1954). The critical incident technique. In *Psychological Bulletin*, volume 5(4), pages 327–358. (Cited on page 27.)

Forsythe, D. E. (1999). "it's just a matter of common sense": Ethnography as invisible work. *Computer Supported Cooperative Work (CSCW)*, 8(1):127–145. (Cited on page 28.)

Freeman, E. and Gelernter, D. (1996). Lifestreams: a storage model for personal data. *SIGMOD Rec.*, 25(1):80–86. (Cited on page 72.)

Gemmell, J., Bell, G., and Lueder, R. (2006). Mylifebits: a personal database for everything. *Communications of the ACM*, 49(1):88–95. (Cited on page 15.)

Greenbaum, J. and Kyng, M., editors (1991). *Design at work: cooperative design of computer systems*. L. Erlbaum Associates Inc., Hillsdale, NJ, USA. (Cited on pages 43 and 44.)

Guimbretière, F. (2003). Paper augmented digital documents. In *UIST '03: Proceedings of the 16th annual ACM symposium on User interface software and technology*, pages 51–60, New York, NY, USA. ACM. (Cited on page 22.)

Hey, T. and Trefethen, A. (2003). e-science and its implications. *Philosophical Transactions of the Royal Society*, pages 1809–1825. (Cited on page 18.)

Hightower, R. R., Ring, L. T., Helfman, J. I., Bederson, B. B., and Hollan, J. D. (1998). Graphical multiscale web histories: a study of padprints. In *HYPERTEXT '98: Proceedings of the ninth ACM conference on Hypertext and hypermedia : links, objects, time and space—structure in hypermedia systems*, pages 58–65, New York, NY, USA. ACM. (Cited on page 77.)

Hill, W. C., Hollan, J. D., Wroblewski, D., and McCandless, T. (1992). Edit wear and read wear. In *CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 3–9, New York, NY, USA. ACM. (Cited on page 102.)

HPCwire (2007). escience: It's really about people. (Cited on page 19.)

Hutchings, D. R., Smith, G., Meyers, B., Czerwinski, M., and Robertson, G. (2004). Display space usage and window management operation comparisons between single monitor and multiple monitor users. In *AVI '04: Proceedings of the working conference on Advanced visual interfaces*, pages 32–39, New York, NY, USA. ACM. (Cited on page 61.)

Hutchins, E. (1996). *Cognition in the Wild*. The MIT Press. (Cited on page 13.)

Hutchinson, H., Mackay, W., Westerlund, B., Bederson, B. B., Druin, A., Plaisant, C., Beaudouin-Lafon, M., Conversy, S., Evans, H., Hansen, H., Roussel, N., and Eiderbäck, B. (2003). Technology probes: inspiring design for and with families. In *CHI '03*, New York, NY, USA. ACM. (Cited on pages 45 and 58.)

Jones, W. (2007). *Keeping Found Things Found, The Study and Practice of Personal Information Management*. Morgan Kaufmann. (Cited on pages 3 and 12.)

Jones, W., Bruce, H., and Dumais, S. (2001). Keeping found things found on the web. In *CIKM '01: Proceedings of the tenth international conference on Information and knowledge management*, pages 119–126, New York, NY, USA. ACM. (Cited on page 77.)

Jones, W. P. and Teevan, J., editors (2007). *Personal information management*. University of Washington Press. (Cited on page 3.)

Ju, W. (2008). *The Design of Implicit Interactions*. PhD thesis, Stanford University. (Cited on page 94.)

Kaptelinin, V. (2003). Umea: translating interaction histories into project contexts. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 353–360, New York, NY, USA. ACM Press. (Cited on page 17.)

Kaye, J. J., Vertesi, J., Avery, S., Dafoe, A., David, S., Onaga, L., Rosero, I., and Pinch, T. (2006). To have and to hold: exploring the personal archive. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 275–284, New York, NY, USA. ACM. (Cited on page 13.)

Kidd, A. (1994). The marks are on the knowledge worker. In *CHI '94: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 186–191, New York, NY, USA. ACM. (Cited on pages 11 and 13.)

Kwasnik, B. H. (1991). The importance of factors that are not document attributes in the organisation of personal documents. *Journal of documentation*, 41:389–398. (Cited on page 13.)

Lamming, M. (1991). Notime - a tool for notetakers. Technical report, Xerox EuroPARC Technical Report. 91-135. (Cited on page 19.)

Lamming, M., Brown, P., Carter, K., Eldridge, M., Flynn, M., Louie, G., Robinson, P., and Sellen, A. (1994). The design of a human memory prosthesis. *The Computer Journal*, 37(3):153–163. (Cited on page 15.)

Lamming, M. and Flynn, M. (1994). Forget-me-not: intimate computing in support of human memory. In *Proceedings FRIEND21 Symposium on Next Generation Human Interfaces*. (Cited on page 16.)

Lamming, M. G. and Newman, W. M. (1992). Activity-based information retrieval: Technology in support of personal memory. In *Information Processing*, pages 68–81. (Cited on page 16.)

Lansdale, M. (1988). The psychology of personal information management. *Applied Ergonomics*, 19(1):55–66.

Latour, B. (1988). *Science in Action: How to Follow Scientists and Engineers through Society*. Harvard University Press. (Cited on pages 19, 25, and 80.)

Lessig, L. (2001). *The Future of Ideas. The Fate of the Commons in a Connected World*. Random House. (Cited on page 12.)

Liao, C., Guimbretière, F., Hinckley, K., and Hollan, J. (2008). Papiercraft: A gesture-based command system for interactive paper. *ACM Trans. Comput.-Hum. Interact.*, 14(4):1–27. (Cited on page 22.)

Luff, P., Heath, C., Norrie, M., Signer, B., and Herdman, P. (2004). Only touching the surface: creating affinities between digital content and paper. In *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 523–532, New York, NY, USA. ACM. (Cited on page 23.)

Lysakowski, R. (1997). Comparing paper and electronic laboratory notebooks. Scientific Computing and Automation. (Cited on page 19.)

Mackay, W., Letondal, C., Pothier, G., Bøegh, K., and Sørensen, H. (2002). The missing link: augmenting biology laboratory notebooks. In *UIST '02: Proceedings of the 15th annual ACM symposium on User interface software and technology*, pages 41–50, New York, NY, USA. ACM Press. (Cited on pages 21 and 27.)

Mackay, W. and Pagani, D. (1994). Video mosaic: laying out time in a physical space. In *MULTIMEDIA '94: Proceedings of the second ACM international conference on Multimedia*, pages 165–172, New York, NY, USA. ACM. (Cited on page 20.)

Mackay, W. E. (1988). More than just a communication system: diversity in the use of electronic mail. In *CSCW '88: Proceedings of the 1988 ACM conference on Computer-supported cooperative work*, pages 344–353, New York, NY, USA. ACM. (Cited on page 14.)

Mackay, W. E. (1990a). Patterns of sharing customizable software. In *CSCW '90: Proceedings of the 1990 ACM conference on Computer-supported cooperative work*, pages 209–221, New York, NY, USA. ACM. (Cited on page 81.)

Mackay, W. E. (1990b). *Users and Customizable Software: A Co-Adaptive Phenomenon*. PhD thesis, M.I.T. (Cited on pages 81 and 85.)

Mackay, W. E. (1998). Augmented reality: linking real and virtual worlds: a new paradigm for interacting with computers. In *AVI '98: Proceedings of the working conference on Advanced visual interfaces*, pages 13–21, New York, NY, USA. ACM. (Cited on page 20.)

Mackay, W. E. (2002). Using video to support interaction design. DVD Tutorial, INRIA and ACM/SIGCHI. http://stream.cc.gt.atl.ga.us/hccvideos/viddesign.php. (Cited on pages 27, 45, and 51.)

Mackay, W. E. and Fayard, A.-L. (1997). Hci, natural science and design: a framework for triangulation across disciplines. In *DIS '97: Proceedings of the conference on Designing interactive systems*, pages 223–234, New York, NY, USA. ACM Press. (Cited on page 6.)

Mackay, W. E. and Fayard, A. L. (1999). Video brainstorming and prototyping: techniques for participatory design. In *CHI '99: CHI '99 extended abstracts on Human factors in computing systems*, pages 118–119, New York, NY, USA. ACM Press. (Cited on pages 8 and 79.)

Mackay, W. E., Fayard, A.-L., Frobert, L., and Médini, L. (1998). Reinventing the familiar: exploring an augmented reality design space for air traffic control. In *CHI '98: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 558–565, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co. (Cited on page 20.)

Malone, T. W. (1983). How do people organize their desks?: Implications for the design of office information systems. *ACM Trans. Inf. Syst.*, 1(1):99–112. (Cited on page 13.)

Mann, S. (2004). Continuous lifelong capture of personal experience with eyetap. In *CARPE'04: Proceedings of the the 1st ACM workshop on Continuous archival and retrieval of personal experiences*, pages 1–21, New York, NY, USA. ACM. (Cited on page 15.)

Marshall, C. C. and Jones, W. (2006). Keeping encountered information. *Commun. ACM*, 49(1):66–67. (Cited on page 13.)

Mistry, P. and Maes, P. (2008). Intelligent sticky notes that can be searched, located and can send reminders and messages. In *IUI '08: Proceedings of the 13th international conference on Intelligent user interfaces*, pages 425–426, New York, NY, USA. ACM. (Cited on page 23.)

Morrison, J. B., Pirolli, P., and Card, S. K. (2001). A taxonomic analysis of what world wide web activities significantly impact people's decisions and actions. In *CHI '01: CHI '01 extended abstracts on Human factors in computing systems*, pages 163–164, New York, NY, USA. ACM. (Cited on page 77.)

Muller, M. J. (2003). *Participatory design: the third space in HCI*. Lawrence Erlbaum Associates, Inc., Mahwah, NJ, USA. (Cited on page 8.)

Mumford, E. (1983). Participative systems design: Practice and theory. *Journal of Occupational Behaviour*, 4(1):47–57. (Cited on page 44.)

Norman, D. A. (1993). *Things that make us smart: defending human attributes in the age of the machine*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA. (Cited on page 13.)

Norrie, C., Signer, B., and Weibel, N. (2006). General framework for the rapid development of interactive paper applications. In *CoPADD 2006, 1st Workshop on Collaborating over Paper and Digital Documents*. (Cited on page 23.)

Obendorf, H., Weinreich, H., Herder, E., and Mayer, M. (2007). Web page revisitation revisited: implications of a long-term click-stream study of browser usage. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 597–606, New York, NY, USA. ACM. (Cited on page 76.)

O'Hara, K. . S. A. (1997). A comparison of reading paper and on-line documents. In *CHI 1997*, pages 335–342. (Cited on page 20.)

Oinn, T., Greenwood, M., Addis, M., Alpdemir, M. N., Ferris, J., Glover, K., Goble, C., Goderis, A., Hull, D., Marvin, D., Li, P., Lord, P., Pocock, M. R., Senger, M., Stevens, R., Wipat, A., and Wroe, C. (2006). Taverna: lessons in creating a workflow environment for the life sciences: Research articles. *Concurr. Comput. : Pract. Exper.*, 18(10):1067–1100. (Cited on page 19.)

Oleksik, G., Wilson, M. L., Tashman, C., Mendes Rodrigues, E., Kazai, G., Smyth, G., Milic-Frayling, N., and Jones, R. (2009). Lightweight tagging expands information and activity management practices. In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*, pages 279–288, New York, NY, USA. ACM. (Cited on pages 18 and 72.)

Paszko, C., Turner, E., and Hinton, M. D. (2002). *Laboratory information management systems*. CRC Press. (Cited on page 19.)

Petrelli, D., van den Hoven, E., and Whittaker, S. (2009). Making history: intentional capture of future memories. In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*, pages 1723–1732, New York, NY, USA. ACM. (Cited on page 15.)

Pirolli, P. (2007). *Information Foraging Theory: Adaptive Interaction with Information*. Oxford University Press. (Cited on page 77.)

Polonsky, A. (2006). Semantic laboratory notebook: Managing biomedical research notes and experimental data. In Proceedings, C. W., editor, *Third European Semantic Web Conference (Industry Forum)*, volume 194. (Cited on page 19.)

Rekimoto, J. (1999). Time-machine computing: a time-centric approach for the information environment. In *UIST '99: Proceedings of the 12th annual ACM symposium on User interface software and technology*, pages 45–54, New York, NY, USA. ACM. (Cited on page 17.)

Ringel, M., Cutrell, E., Dumais, S., and Horvitz, E. (2003). Milestones in Time: The Value of Landmarks in Retrieving Information from Personal Stores. In *Proceedings of Interact 2003, the 9th IFIP TC13 International Conference on HCI*, pages 184–191. IOS Press, Amsterdam. (Cited on pages 17 and 77.)

Robertson, G., Czerwinski, M., Larson, K., Robbins, D. C., Thiel, D., and van Dantzich, M. (1998). Data mountain: using spatial memory for document management. In *UIST '98: Proceedings of the 11th annual ACM symposium on User interface software and technology*, pages 153–162, New York, NY, USA. ACM. (Cited on page 16.)

Rogers, Y. (2006). Moving on from weiser's vision of calm computing: Engaging ubicomp experiences. In *UbiComp 2006*, pages 404–421. Springer. (Cited on page 15.)

Roure, D. D., Goble, C., and Stevens, R. (2009). The design and realisation of the virtual research environment for social sharing of workflows. *Future Generation Computer Systems*, 25(5):561 – 567. (Cited on page 19.)

Roussel, N., Fekete, J.-D., and Langet, M. (2005). Vers l'utilisation de la mémoire épisodique pour la gestion de données familières. In *IHM 2005: Proceedings of the 17th international conference on Francophone sur l'Interaction Homme-Machine*, pages 247–250, New York, NY, USA. ACM. (Cited on page 12.)

Roussel, N., Tabard, A., and Letondal, C. (2006). All you need is log. In *WWW 2006 Workshop on Logging Traces of Web Activity: The Mechanics of Data Collection*. 4 pages. (Cited on page 88.)

Santry, D. S., Feeley, M. J., Hutchinson, N. C., Veitch, A. C., Carton, R. W., and Ofir, J. (1999). Deciding when to forget in the elephant file system. *SIGOPS Oper. Syst. Rev.*, 33(5):110–123. (Cited on page 101.)

Schön, D. (1984). *The Reflective Practitioner: How Professionals Think In Action*. Basic Books. (Cited on pages 19, 25, and 38.)

Schraefel, M. C., Hughes, G. V., Mills, H. R., Smith, G., Payne, T. R., and Frey, J. (2004). Breaking the book: translating the chemistry lab book into a pervasive computing lab environment. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 25–32, New York, NY, USA. ACM. (Cited on page 19.)

Seifried, T., Jervis, M., Haller, M., Masoodian, M., and Villar, N. (2008). Integration of virtual and real document organization. In *TEI '08: Proceedings of the 2nd international conference on Tangible and embedded interaction*, pages 81–88, New York, NY, USA. ACM. (Cited on page 23.)

Sellen, A. J., Fogg, A., Aitken, M., Hodges, S., Rother, C., and Wood, K. (2007). Do life-logging technologies support memory for the past?: an experimental study using sensecam. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 81–90, New York, NY, USA. ACM. (Cited on page 15.)

Sellen, A. J. and Harper, R. H. (2001). *The Myth of the Paperless Office*. MIT Press, Cambridge, MA, USA. (Cited on pages 3 and 20.)

Sellen, A. J., Murphy, R., and Shaw, K. L. (2002). How knowledge workers use the web. In *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 227–234, New York, NY, USA. ACM. (Cited on page 76.)

Seneca, L. A. (2002). *Lettres a Lucilius : anthologie*. Mille et une nuits. (Cited on page 38.)

Signer, B. (2008). *Fundamental Concepts for Interactive Paper and Cross-Media Information Spaces*. Verlag. (Cited on page 23.)

Signer, B. and Norrie, M. C. (2007). Paperpoint: a paper-based presentation and interactive paper prototyping tool. In *TEI '07: Proceedings of the 1st international conference on Tangible and embedded interaction*, pages 57–64, New York, NY, USA. ACM. (Cited on page 23.)

Stallman, R. M. (2002). *Free Software, Free Society: Selected Essays*. GNU Press. (Cited on page 12.)

Stiegler, B. (2001). *La technique et le temps. Tome 3: Le temps du cinéma et la question du mal-être*. Editions Galilee. (Cited on page 38.)

Stifelman, L., Arons, B., and Schmandt, C. (2001). The audio notebook: paper and pen interaction with structured speech. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 182–189, New York, NY, USA. ACM. (Cited on page 21.)

Strauss, A. C. and Corbin, J. (1990). *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Sage Publications, Inc. (Cited on pages 7, 28, 29, and 121.)

Suchman, L. (1987). *Plans and situated actions: the problem of human-machine communication*. Cambridge University Press. (Cited on page 12.)

Tabard, A., Mackay, W. E., and Eastmond, E. (2008). From individual to collaborative: the evolution of prism, a hybrid laboratory notebook. In *CSCW '08: Proceedings of the ACM 2008 conference on Computer supported cooperative work*, pages 569–578, New York, NY, USA. ACM. (Cited on page 26.)

Tauscher, L. and Greenberg, S. (1997). Revisitation patterns in world wide web navigation. In *CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 399–406, New York, NY, USA. ACM. (Cited on pages 76 and 77.)

Teevan, J. (2007). *Supporting Finding and Re-Finding Through Personalization*. PhD thesis, Massachusetts Institute of Technology. (Cited on page 78.)

Teevan, J., Alvarado, C., Ackerman, M. S., and Karger, D. R. (2004). The perfect search engine is not enough: a study of orienteering behavior in directed search. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 415–422, New York, NY, USA. ACM. (Cited on pages 14 and 78.)

Tsandilas, T., Letondal, C., and Mackay, W. E. (2009). Musink: composing music through augmented drawing. In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*, pages 819–828, New York, NY, USA. ACM. (Cited on page 22.)

Tulving, E. (1983). *Elements of Episodic Memory*. Oxford University Press, Oxford, UK; New York. (Cited on page 16.)

Voida, S. and Mynatt, E. D. (2009). It feels better than filing: everyday work experiences in an activity-based computing system. In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*, pages 259–268, New York, NY, USA. ACM. (Cited on pages 18 and 72.)

von Hippel, E. (1986). Lead users: a source of novel product concepts. *Management Science*, 32(7):791–805. (Cited on page 18.)

von Hippel, E. (1988). *The Sources of Innovation*. Oxford University Press. (Cited on page 3.)

Weibel, N., Ponti, P., Signer, B., and Norrie, M. (2007). Paperproof: Paperproof: A paper-digital proof-editing system. In *CoPADD'07 Collaborating over Paper and Digital Documents*. (Cited on page 22.)

Weinreich, H., Obendorf, H., Herder, E., and Mayer, M. (2006). Off the beaten tracks: exploring three aspects of web navigation. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, pages 133–142, New York, NY, USA. ACM. (Cited on page 76.)

Weiser, M. (1991). The computer for the 21st century. *Scientific American Special Issue on Communications, Computers, and Networks,*. (Cited on page 15.)

Wellner, P. (1993). Interacting with paper on the digitaldesk. *Commun. ACM*, 36(7):87–96. (Cited on page 20.)

Whittaker, S. and Hirschberg, J. (2001). The character, value, and management of personal paper archives. *ACM Trans. Comput.-Hum. Interact.*, 8(2):150–170. (Cited on page 13.)

Whittaker, S., Hyland, P., and Wiley, M. (1994). Filochat: handwritten notes provide access to recorded conversations. In *CHI '94: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 271–277, New York, NY, USA. ACM. (Cited on page 19.)

Whittaker, S. and Sidner, C. (1996). Email overload: exploring personal information management of email. In *CHI '96: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 276–283, New York, NY, USA. ACM. (Cited on page 14.)

Wilcox, L. D., Schilit, B. N., and Sawhney, N. (1997). Dynomite: a dynamically organized ink and audio notebook. In *CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 186–193, New York, NY, USA. ACM. (Cited on page 19.)

Winograd, T. and Flores, F. (1985). *Understanding computers and cognition*. Ablex Publishing Corp., Norwood, NJ, USA. (Cited on page 12.)

Wouters, P. (2004). Public proofs, science, technology and democracy: What is the matter with e-science? In *4S & EASST Conference*. (Cited on page 18.)

Yeh, R., Liao, C., Klemmer, S., Guimbretière, F., Lee, B., Kakaradov, B., Stamberger, J., and Paepcke, A. (2006). Butterflynet: a mobile capture and access system for field biology research. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 571–580, New York, NY, USA. ACM. (Cited on page 21.)

Yeh, R. B. (2007). *Designing Interactions that Combine Pen, Paper, and Computer*. PhD thesis, Stanford University. (Cited on page 23.)

Yeh, R. B., Paepcke, A., and Klemmer, S. R. (2008). Iterative design and evaluation of an event architecture for pen-and-paper interfaces. In *UIST '08: Proceedings of the 21st annual ACM symposium on User interface software and technology*, pages 111–120, New York, NY, USA. ACM. (Cited on pages 23 and 60.)

Part II

APPENDICES

A INTERVIEWS

A.1 QUESTIONS

We prepared the following questions as a basis for the interview. For most interviews the discussion started right away as interviewees were aware that we were interested in notebooks use and information management practices. The original questions were in French or English.

Online and offline information management:

- Do you use a laboratory notebook?
- When was the last time you wrote something in it?
- Could you show us a recent page with references to your computer activity?
- Is this your only notebook? (do you also have a paper/digital notebook?)

Finding information again:

- When did you start this notebook?
- When was the last time that you came back to notes in paper notebook?
- Where are your past notebooks?
- When was the last time you searched for something on your desktop?
- Do you have backups?

Research:

- Are you running an experiment? Where is the last protocol you used?
- Do you run analysis on Pasteur central server?
- Did you run any analysis on the web recently?
  - How did you access the page?
  - How did you learn about this page?
- How and where did you save the output data?
- Did you come back to it?

B CODINGS

B.1 OPEN CODING

After the investigation at the Institut Pasteur and INRA, I organized in a hierarchy the observed micro-phenomena. I grouped them by names defining them (concepts according to Grounded Theory). The concepts were then refined to remove similar ones and reorganized in the hierarchy. Finally, I grouped the concepts into categories according to the open coding phase of Grounded Theory.

Tables 6 and 7 present the open coding of the field observations. The tables are organized around the different categories identified, with their respective properties and the dimensions of every property.

Categories	Properties	Dimensions
Notes	Media	Scraps – Electronic – Paper (NB)
	Structure	Constrained – Flexible
	Status	Scratch – Transient – Official
Writing/ and saving	Storing, remembering	Time (temporary – long term)
	Filter, selection	Degree of information overload
	Discard	Intentionality: decay, loss, deletion
	Cognitive offload	Cognitive load, environment richness
Organization	Re-finding	Convenience, Success
	Reminding	Cues
	Schemes (Topical)	Date/manipulation/project/person
	Media	Physical – digital (space)
Online work	Repetition	Make – Reuse – Automate
	Alternatives	Degree of complexity
	Routine	Degree of familiarity
Sharing/ Awareness	Implicit	Physical space (where it is in the lab)
	Explicit	Social mediation, agreed upon rules
	Privacy	Public – private
	Embedded cues	Marks in the notebook, post-its...
	Media status	Personal – communal

Table 6: Study 1, Institut Pasteur, Open coding

Categories	Properties	Dimensions
Memory tools	Quality management	Quality of information transfer, Degree of discipline
	Cognitive offload	Temporary – long term
	Media (paper, digital)	Personal – shared – up-to-date
	Status	Scratch – transient – official
Organization	Digital	Project – person – time
	Dynamism	Speed of changes
	Adaptation	Degree of flexibility
Analyzing	Repetition	Make – reuse – automate
	Software development	Target (biologists, bioinformaticians)
	Stability, Efficiency	Speed, data quality
Sharing	Virtual placement	Space (where it is on the server)
	Collaboration, group work	Team, remote, 1to1, PhD – teacher
	Privacy	Public – private

Table 7: Study 2, INRA, Open coding

B.2 AXIAL CODING

Following the Grounded Theory approach, the concepts coming from the open coding are grouped together to identify emerging themes. The *Axial coding* phase consists in putting together the data from the open coding in new ways. The goal here is to develop a coding system that seeks to identify causal relationships between categories.

> *'The axial coding is the process of relating categories to their subcategories, termed "axial" because coding occurs around the axis of a category, linking categories at the level of properties and dimensions.* (Strauss and Corbin, 1990)

The axial coding led us to find concepts which are common to our different settings. Figure 53 presents the hierarchies of categories in a concept map.

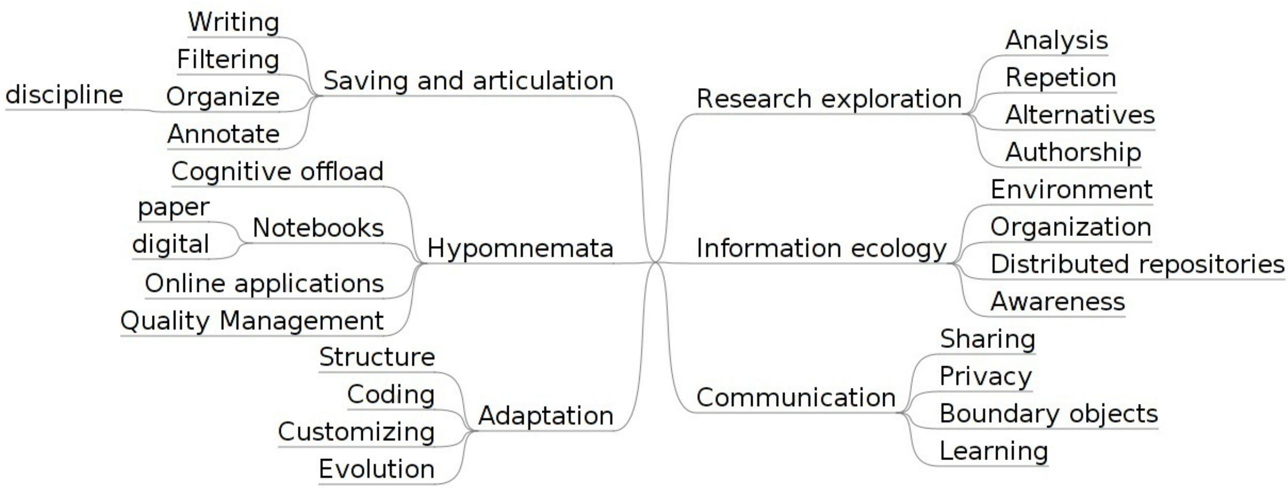


Figure 53: Axial Coding

The categories of figure 53 are organized around six axis: The *exploratory aspect of research* and how the changes in science modify research habits. A striking aspect of these changes are the new tools and strategies developed to *communicate* between researchers. As information moves to the digital world, it becomes both easier and more difficult to organize it and become aware of its changes (*information ecology*). In order to handle this information shifting between two worlds, Biologists and Bioinformaticians use notebooks and online tools supporting reflection, *hypomnemata*. As they capture information they must *articulate their activity*.