From Individual to Collaborative: The Evolution of Prism, a Hybrid Laboratory Notebook

Aurélien Tabard[1,2]
tabard@lri.fr

Evelyn Eastmond[1]
evelyn@media.mit.edu

Wendy E. Mackay[1,2]
mackay@lri.fr

[1]INRIA
F-91405, Orsay, France

[2]LRI - Univ. Paris-Sud & CNRS
F-91405, Orsay, France

ABSTRACT

We report on our studies of the evolving work practices of biologists and the role paper and electronic lab notebooks play in supporting their individual and collaborative activity. We describe the participatory design and longitudinal field testing of *Prism*, a hybrid laboratory notebook that lets biologists capture, visualize and interact with cross-linked streams of physical and electronic data. We used Prism as a technology probe that users could adapt to integrate additional activity streams and share information from other biologists. Our key findings include the use of *master notebooks*, whether paper or electronic, which act as a reference point for handling and organizing the diverse strands of personal activity, and the importance of *redundancy,* which biologists use to make sense of their data. Prism provides a flexible, extensible tool that supports individual and collaborative reflection in creative work.

Author Keywords

Activity streams, Augmented Paper, E-science, Hybrid lab notebooks, Information management, Laboratory notebooks

ACM Classification Keywords

D.2.2 [Design Tools & Techniques]: User Interfaces. H.5.2 [User Interfaces]: Evaluation/methodology, User-centered design. H.3.3 [Information Search and Retrieval]: Information filtering. H.5.3 [Group and Organization Interfaces]: Computer-supported cooperative work

INTRODUCTION

Advances in computer technology have transformed how biologists work. Wouters [32] describes the emergence of *E-Science* as the result of three major developments: "the large-scale sharing of computational resources, the provision of access to massive, distributed and heterogeneous datasets (...tera to petabytes), and the use of digital platforms for collaboration and communication". It is now common for biologists to share their data and to communicate with researchers around the world, in a complex mix of competition and collaboration [14]. However, despite the undoubted advantages, the advent of networked computing has contributed greatly to biologists' information overload [25]. As lead users [28] or 'extreme knowledge workers', they need more powerful tools to cope with their increasing quantities of data and communication.

Historically, biologists, like other scientists, relied upon a simple, but powerful tool for recording procedures, data and results: the laboratory notebook. Although usually viewed as the personal record of an individual scientist, lab notebooks are written to be read by others, imposing a corresponding discipline in the style and choice of what is recorded. Together with published articles, they still serve as a primary archival and communication medium for academic biologists [4]. As research activities move on-line, with more remote collaboration, biologists are under increasing pressure to use fully electronic notebooks. At the same time, paper laboratory notebooks fill unique needs, for legal, archival and ease of use reasons [16].

We are interested in rethinking the design of laboratory notebooks so as to take advantage of both formats. We are particularly interested in supporting two aspects of biologists' work: *creative* generation of new hypotheses and experiments and making sense of the data, and *collaborative* verification, accountability and know-how sharing across local and remote teams.

This paper describes our work with biologists at the Institut Pasteur and at INRA. We examine how their work has changed, from primarily individual to increasingly collaborative, with a particular focus on the lab notebook and how it has evolved in the face of enormous change. We then describe *Prism*, a hybrid paper and electronic notebook created in collaboration with bioinformaticians and tested in their laboratory over a period of nine months. We present

CSCW'08, November 8–12, 2008, San Diego, California, USA.

our findings and conclude with a discussion and directions for future research.

RELATED WORK

In order to understand biologists' use of laboratory notebooks, it helps to look at personal information management and the problem of how to find documents that one has already used. Malone's [19] studies of how people organize their desks identified two key strategies: filing and piling. He noted 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. Barreau & Nardi [1] identified a similar phenomenon with computer documents, which they classified as ephemeral, working or archived. Bondarenko et al. [3] studied document use across the physical and digital divide and argued that documents should be embedded within meaningful context information, a particular problem for digital documents, and must be easily accessible for regrouping as the task progresses.

The problems of managing paper and digital documents increase in collaborative settings. In their study of a workgroup's migration from paper document collections to electronic repositories, Trigg et al. [27] highlighted the difficulty of maintaining alignment across the two media. Dourish et al. [5] argue that this is because organizations are inherently dynamic, making it difficult for people to categorize documents so that others can find them.

Traditional laboratory notebooks have specific properties that differentiate them from other work documents. They provide a long-term, chronological record of a biologist's activities. Biologists are trained to think of lab notebooks as formal documents that require disciplined writing and cannot be edited. Lab notebooks serve both as an archive of a scientist's work and as a legal document that establishes priority and intellectual property rights. They differ from other less formal notebooks that Biologists also use, such as disposable scratch pads or notes from meetings.

Many now regard paper lab notebooks as 'old fashioned' [4] and they are being replaced by electronic notebooks, particularly in large, commercial research labs. Some of these electronic notebooks are designed for personal use, e.g., *ipad*[1] and *ChemOffice*[2], while others are distributed and intended for remote work groups e.g., ELN [20]. Interestingly, many academic scientists are resisting the move to electronic lab notebooks [4]. This is unlikely to be due to their 'fear of computers', since many are also skilled computer users [10]. Some of this resistance is probably due to paper's characteristics: it is lightweight, ubiquitous, inexpensive and easy to use [17, 21]. We believe that it also serves as the final, authoritative version of the work [11].

SmartTea [22] and MyTea [23], designed for running chemistry experiments, offer an alternative that explicitly attempts to retain the affordances of paper while leveraging the advantages of computers. Another approach is to combine physical and electronic documents directly, e.g., Digital Desk [29]. Video Mosaic [13] also allowed authors to interact with each other, using paper documents as a method for accessing a media space. NoTime [8], Audio Notebook [24] and Dynomite [31] all linked paper-based note taking to video and/or audio recordings. Mackay et al. [18] developed a series augmented laboratory notebooks in which Biologists wrote on paper lab notebooks. Handwriting and coded images were captured with a portable or desktop graphics tablet and linked to a searchable electronic version of the lab notebook. Later, the A-book acted as a 'physical information lens', providing an interactive window on any page of the paper notebook, with bi-directional links between the paper and the computer.

These early technologies remained working prototypes, never practical enough to use on a daily basis. The advent of Anoto[3] technology, a video-enabled pen that reads a unique dot pattern on paper, makes it possible to create real-world applications. Toolkits designed to help designers and developers build pen-and-paper applications [34] or create interactive paper command systems [6], facilitated the design of systems such as PapierCraft [12] or PaperProof [30] which provide methods of annotating and updating paper documents. Yeh et al. [33] used Anoto to develop ButterflyNet, a mobile tool to help field biologists capture diverse data (handwritten notes, photographs, sensor readings, GPS track logs) from field research and link it to the computer. We are interested in exploring how biologists use lab notebooks, both paper and electronic, to support individual and collaborative creative activities in the lab. The next section describes our first study, a field study to observe how the use of lab notebooks has evolved.

STUDY 1: FIELD OBSERVATIONS AT TWO LABS

We have been working with biologists over the past decade [10], developing both tools [26] and lab notebooks [18]. We were interested in how both have changed during this time, so we initiated a new study to explore this question and to inform the subsequent design of a new Anoto-based tool.

Method

Participants: We interviewed 10 biologists at the Institut Pasteur, a leading biology research center in Paris, and eight bioinformaticians at INRA, the French government agricultural research institute in Evry. (The latter are trained in both biology and computer science.) Participants ranged from doctoral students to senior researchers and included biologists, bioinformaticians, managers and an archivist.

Procedure: We visited individual researchers in their labs or offices and asked to see specific examples of their uses

[1] http://ipadeln.com/

[2] http://www.cambridgesoft.com/

[3] http://www.Anoto.com/

of lab notebooks, paper or electronic. We also asked them to demonstrate their use of the computer for search, analysis and other common tasks. In each setting, we asked them to recall recent critical incidents, both positive and negative, relating to the management, reuse and sharing of on-line and off-line data. We also held several meetings in which small groups of researchers compared their experiences.

Results

We were struck by the pervasive use of computational tools and the transition from personal to collaborative work. In our earlier studies at the Institut Pasteur in the 1990s, few biologists had any formal computer training, personal computers were rare, the internet was new and most biologists involved in fundamental research considered patents irrelevant. Very few identified themselves as bioinformaticians. 'Bench' biologists worked individually on their experiments: they used paper laboratory notebooks and published research articles as the primary means of archiving and communicating their work.

Computers are now ubiquitous. No longer rare, we saw computers everywhere, tucked into the corners of almost every lab. In 1995, they were a scarce, shared resource; now everyone has a laptop. Biologists surf the web, often from home, and all use standard office software. Bioinformaticians often use multiple computers, some with multiple screens. All of the young biologists had studied computing as part of their biology degrees and the Institut Pasteur now has a diploma course in computer science for both junior and senior biologists.

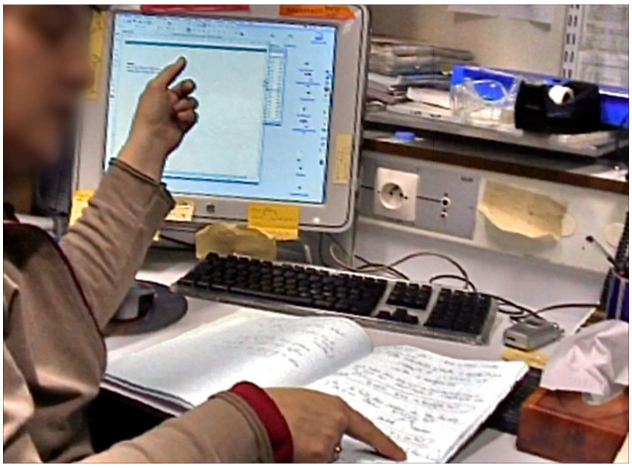

Figure 1. A biologist explains how she juggles a complex mix of paper and computer-based information

Managing information in both worlds. All of these researchers struggle with how best to structure their activities (Fig. 1), either to find relevant information in the future or to help colleagues understand shared data. Their work is a complex web of interrelated references, many kept in their heads, and does not map easily into a single organizational form or structure, either physical or digital. Some maintain personal lab notebooks (seven text-based lab notebooks, three doc files) while others use collaborative ones (four HTML electronic notebooks, one blog and one wiki). Some of the more technical users had also tried commercial electronic notebooks, e.g., *ipadeln*, but rapidly discarded them as too constraining.

Many have trouble keeping track of where their personal information is stored. Several said that re-implementing a script is faster than finding one: *"It takes me 10 times less time to redo the same analysis than to find the appropriate script."* Another knew she had created a detailed README file that explained a long protocol, but it was three months ago and she could not remember the name. She keeps searching for it, because it is too long to do over and she hopes she will stumble across one day. Another asked for help: *"I have lots of data, so if there is something that maintains links for me it would be much appreciated."*

Organization strategies. The chronological organization of paper lab notebooks makes it easy and efficient to find recently entered information, the most common use. Most electronic lab notebooks are also arranged chronologically. As one Biologist explained, *"Date is something primordial ... I like it when it's done in a linear chronological way, that way we're sure to be able to find it"*. However, over time, project-based organization becomes more efficient, particularly for senior researchers and managers who handle many projects in parallel. One geneticist, who had been working on two articles representing about four years of work, said: *"In a [time-based notebook] it's crazy, in two years. I will not remember when I did something from two months ago."* He uses both physical folders and project-based folder hierarchies on his computer. Fig. 2 shows their choices: chronological versus project-based organization, and physical versus computer-based media.

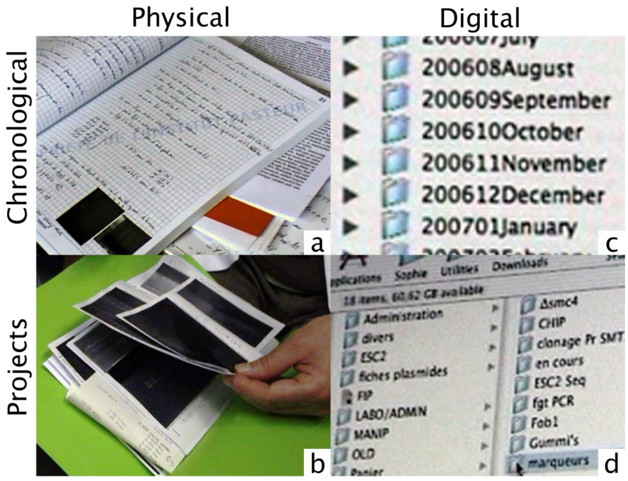

Figure 2. Biologists' organizational strategies: physical vs. computer-based and chronological vs. project-based.

Beyond notebooks. In our earlier studies, the paper lab notebook was the central point for keeping track of physical or printed data. Today's biologists, in addition to their paper or electronic notebooks, must also manage and find information in email, personal data files, shared data servers and on the web. They use various tools, ranging from simple README files to JIRA, an online project management tool and Alfresco, online document manager, as well as the ubiquitous Google and its applications. All write README files, but in different ways, for different audiences. One senior researcher writes for her colleagues: *"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."* Another insists that her READMEs are *"just for me, although I still forget them after a couple of months. I keep my files organized [for my own use], but it won't work for others."*

Sharing within the group. Within research groups, the chronological organization of information sometimes breaks down. *"When we do things in parallel, we don't go at the same speed and a temporal organization makes no sense"*. Instead, they organize things by projects: *"It's easier to have all the 'oligo' in the same place"* so everyone knows where the data is stored. INRA has many remote data repositories, "*but we try to keep the same structure; otherwise it's a loss of time for everybody"*. However this is difficult to maintain. One manager described how, after he set up a hierarchy of README files, the structure broke down quickly because the project evolved and they did not have a common set of keywords.

Global data repositories have transformed their work: The internet has fundamentally changed the nature of their work. Whenever they publish an article, they also submit raw data to a global data repository, such as the human genome project. In turn, these world-wide networks provide them with access to the latest raw data, algorithms and experimental results. Although this has magnified the scope and increased the speed of their research, with enormous scientific benefits, it has also raised major challenges in how to operate effectively within this new environment.

One bioinformatician described how even the replication of research results, fundamental to the scientific method, has changed. Instead of comparing her results to data published a journal article, she now makes an on-line comparison to data she finds in a global database. If she or anyone else wants to replicate her results, they must compare the new data to the previous dataset, which has of course changed. Although she could also preserve a 'snapshot' of the state of the global database at the time of each comparison, this is usually too much data for her laptop and often even for the central server. As she says, "The reference points keep changing" and any result is always subject to change. This problem is particularly true for bioinformaticians, who conduct all of their experiments *in silico* rather than *in vivo*.

From personal to collaborative work: This globalization of science has increased the level of collaboration, particularly for the bioinformaticians. When asked who her competitors were, one laughed and said "*None! We collaborate with them!*" Although her reaction may be less true in commercial research, in the academic labs we observed, biologists sought to balance being first to discover a phenomenon and collaborating with the best possible group of colleagues, to share costs and resources. They are struggling to find new ways of collaborating "*We need some kind of interconnection.*" They want reliable, accessible repositories of knowledge, not just data. They want to share the know-how and informal details that may be captured in a personal notebook, but are rarely shared.

Discussion

The prevalence of computers in research labs has opened new possibilities in term of access to data, analysis power and collaborations. However, biologists have difficulty managing this information, whether physical or digital.

Discipline in paper and electronic notebooks: Biologists use paper in a variety of forms (Fig. 3): sheets of scratch paper hold transient data and ideas, spiral notebooks track meetings and non-archival information, and formally registered laboratory notebooks record the current state of the biologist's research activity. To complicate matters, computer files are also used for any of the above, making it an on-going challenge for biologists to keep track of what is located where, in which medium.

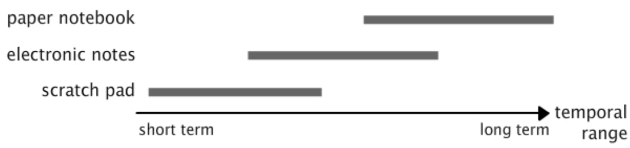

Figure 3. Temporal range: from throw-away scratch paper, to long-term online files, to archival laboratory notebooks.

Even though all use the computer extensively, 75% of biologists and 60% of bioinformaticians continue to use paper lab notebooks. Biologists are trained to write systematically, including what they consider to be the key important results, as well as any additional information necessary for them or someone else to replicate the experiment. Interestingly, it is not simply its inherent flexibility that has caused paper to persist after the advent of personal computers. Rather, it is 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 the biologist, 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.

Paper lab notebooks serve as a formal repository for information that can be treated as a stable point in an evolving world and we remain convinced that they remain a key method for helping users manage their time and work. From our perspective, visions of a purely "digital lab" are often founded on incorrect assumptions about the nature of scientific work and what it means to interact with an 'electronic notebook'. When computer files can be easily edited, it becomes difficult to find the 'definitive' state of a particular document. 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.

Today's electronic notebooks focus on experimental procedures, facilitating the capture and search for information. Yet this is only the first step: research involves more than just experiments. We are interested in supporting the full range of research activities, from forming hypotheses to interpreting results. In the second study, we investigated how we could help researchers by taking advantage of their existing knowledge and experience, without overwhelming them with cumbersome technology.

STUDY 2: PARTICIPATORY DESIGN OF PRISM

Based on our initial observations, we used a participatory design approach to design *Prism,* a hybrid paper-electronic notebook for to help biologists manage their daily research. *Prism* integrates cross-linked streams of activity: an Anoto-based paper notebook, an electronic notebook, and digital documents (Fig. 4). Like most lab notebooks, *Prism* is organized chronologically. Users can also freely tag any entry from any activity stream and associate it with any other document or project entity, which they can filter later. Users can also search among the digital entries and share their activity streams with other scientists.

Method

We designed *Prism* as a technology probe [7], with three parallel goals: to collect data about its use *in situ*, to understand the technical implications and limitations of hybrid (paper-electronic) documents, and to inspire participants and designers to think about new possibilities and needs that emerge only through everyday use. Although testing design ideas in workshops is useful, it is clearly no substitute for observing users in real situations. This however involves a trade-off. On one hand, any real-world technology must be reliable enough that busy users can try it without danger. On the other hand, building a working system commits us to a particular design and architecture: details may be easy to change, but the overall structure is not. We thus tested ideas for *Prism*, in the form of scenarios and video prototypes, in two workshops, before developing the final working version that we tested with biologists.

Participants: Five bioinformaticians from INRA volunteered to try *Prism*. All are biologists with additional training in computer science, including two managers, two senior bioinformaticians, and a junior researcher. Unlike 'bench' biologists at the Institut Pasteur, their primary

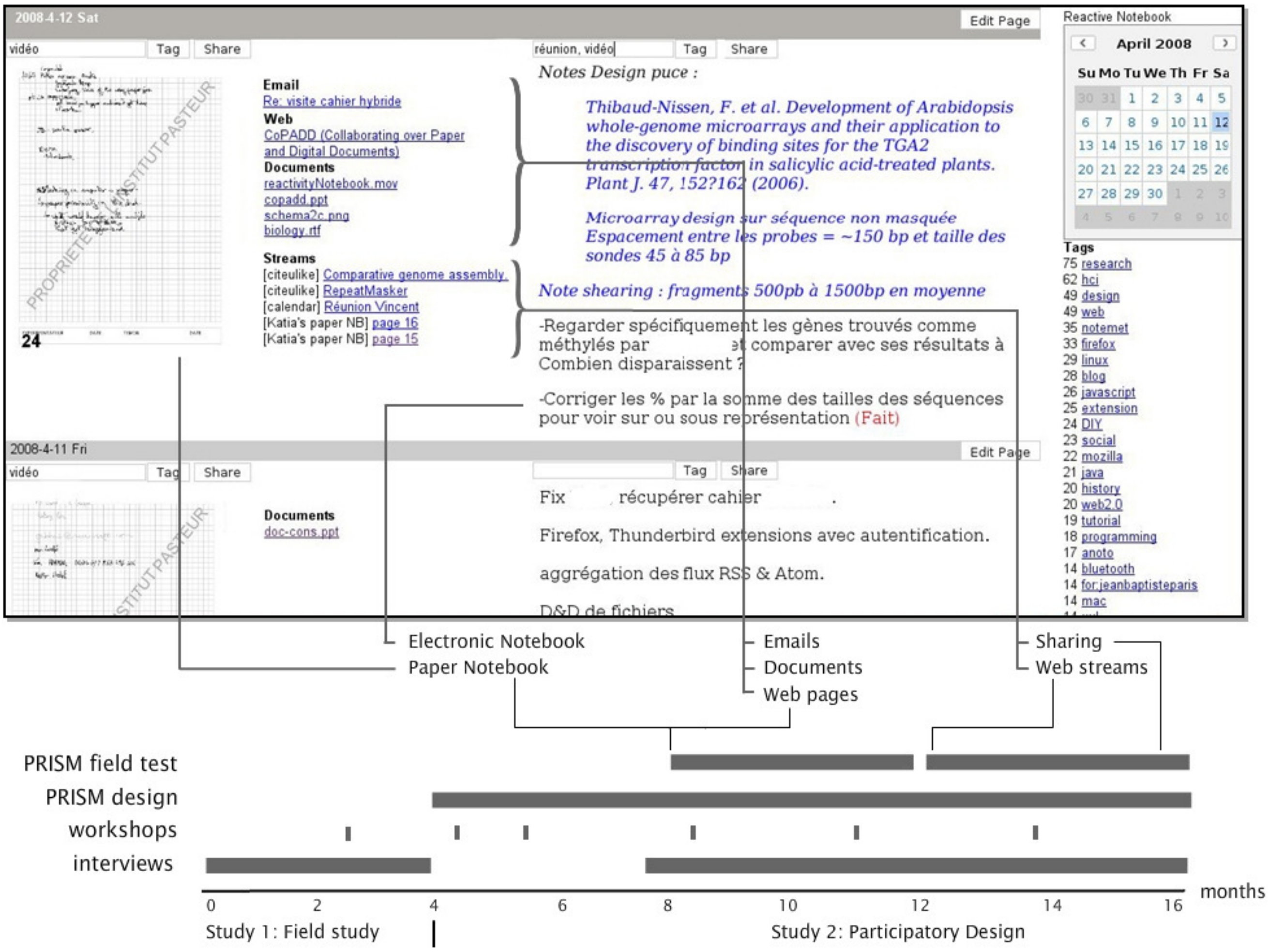

Figure 4: Prism includes multiple cross-linked activity streams: hand-written entries from an Anoto-based paper notebook, typed entries from an electronic notebook, other computer activity, e.g., relevant emails and web results, and shared entries from others.

research activities occur on-line: they collect and analyze data on the computer, not in the lab. INRA provides three levels of notebook: from what are effectively scratch pads, to experimental notebooks for recording information about experiments, to formal lab notebooks. They have the choice and most use the mid-level experiment notebooks.

Procedure: We initially ran participatory design workshops, at the Insitut Pasteur and INRA, to gather ideas and scenarios of use, as well as to engage participants in the study (Fig 5). The INRA bioinformaticians were especially eager to experiment with new forms of information management. At the time, several, including the manager, were specifically reflecting about how they manage information individually and as a group. This led us to focus on the collaborative elements of the notebook, an issue of less concern for Pasteur's 'bench' biologists.

Participants used *Prism* for nine months, from August 2007 to April 2008. We also ran a series of workshops to investigate design alternatives. We interviewed them about their use of *Prism* throughout the field test, in person and via email. We also analyzed their notebooks, what parts they used and what they looked at. *Prism* evolved during this time, in response to user's requests, with daily, weekly or monthly updates. This lead to different types of use, depending upon the stability of the new components and the overall system. Although *Prism* was not static, it was real, and gave users a sufficiently long record that we could begin exploring what it means for them to reflect on their own and others' activity over significant periods of time.

Prism and notebooks: Design and implementation

We designed *Prism* to integrate multiple streams of activity. It evolved from a desktop application to an online collaborative tool for capture, visualization and sharing of users' activity. The key components include:

Paper notebook:	Hand-written text and drawings
Electronic notebook:	Typed or copied text and images
Desktop activity:	Email, websites and documents
Web activity:	Feeds from the web
Shared activity:	Notebooks entries from others

Prism Initial architecture

Prism was initially developed as a desktop JAVA application and two add-ons for Firefox and Thunderbird, so that it could run on the different platforms used by researchers: Linux, Mac OSX and Windows. However, only Windows users could the Paper Notebook, since Anoto pen drivers were only available for that platform. Users worked with the following components:

Paper Notebook: Pages are printed with the Anoto dot-pattern. Users write with the Anoto pen, which leaves ink on the paper and saves a digital copy. We use the Paper Toolkit [33] to process Anoto marks into images, which are then stored on-line with associated meta-data (timestamp, revisions, tags). We generate images for the different versions of the page, so that if a user writes on the same page on two different days, it will appear as two images.

Electronic Notebook: Similar to an on-line journal, users write daily entries with an HTML editor. Each day is stored as a separate HTML file and its metadata (timestamp, revisions, tags) is stored in an XML file.

Activity Log: Aggregated computer activity includes:

Web use: A Firefox extension displays a small button next to the URL address bar. Clicking this button sends the current web page to the notebook. Users can also tag a page, which saves its image and displays it on the appropriate day in the electronic notebook.

Email use: A similar Thunderbird extension displays a button next to each message. Clicking this button sends the email message to the notebook. Users can also tag emails, and display them in the notebook.

Document publishing: Biologists write, review and edit a variety of on-line documents. *Prism* uses the operating system's list of recent documents to link these documents to the notebook. Users can also explicitly create links by dragging and dropping files or images into the notebook.

Since these biologists are using their real data, we had to guarantee that their notes would survive after the study was finished. We thus chose standard data formats: electronic notebook entries are HTML files, paper notebook entries are PNG files (and paper) and pointers to email, web pages and documents are XML files. Data was stored locally at the INRA lab where we ran the study, for security reasons.

Results: initial use

Even with only six participants, use of *Prism* varied greatly, influenced by personality, job type and number of projects.

Organization strategies. The two managers manage both people and research activities. Both attend many meetings and have lost count of how many projects they handle at once. The three senior bioinformaticians also work on multiple projects, but are able to focus more on their own research. Everyone used tags: to track *activities*, e.g., meeting or submission, *projects*, e.g., pascoDB or blumeria, or *data type,* e.g., Perl scripts. One senior researcher used a file name, `splitgffchimeres.pl`, as a tag because the file was evolving from one day to the next and she wanted find it again, and also associate it with comments in her notebook.

Time remains an important organizing principle for this group. Both managers and a senior researcher used *Prism*'s electronic notebook to set reminders for upcoming tasks. They marked emails about meetings or items to complete, such as an unfinished README file. Two researchers also used *Prism* as an alternative to standard filing or bookmarking, especially when they had doubts about file category or when the information was useful but temporary. Three researchers used *Prism* as a versioning system to track the progress of research articles or interesting web pages. *Prism* stores snapshots of web pages and users discovered they could capture the parameters of web forms and later reexamine versions of pages that were no longer available.

Distributed personal data. All the researchers in this group use at least two computers and two different operating systems. As one researcher explained: *"being cross platform is a problem when we want to follow data"*. Another 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."* Two researchers used activity logs and an electronic notebook to keep track of data across Unix, Mac and Windows computers. They also used *Prism* as an alternative when they did not know how to file a document. 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.

Personal and shared information on the web. All team members share information via the web. They use the JIRA on-line task manager and the Alfresco document manager on a local server, as well as web-based applications such as Google Mail, Calendar and Docs to work collaboratively and CiteULike.com to share bibliographic references. Because of the dynamic nature of these applications, their URLs do not necessarily link directly to the desired content, so users often need to log in and navigate to get to the information.

We initially envisioned the capture of web pages as a way to keep track of useful information from the web, e.g., research articles. But participants also made links from web applications, such as to ongoing tasks in JIRA. One senior researcher created an ecosystem of links from her electronic notebook to README files, the scripts she modified on her machines and related JIRA tasks. We decided to modify *Prism* further, to better integrate it with the web.

Prism design iteration

We transformed *Prism* from a single-user desktop application into a distributed web application. The web server uses Apache and PHP, while the client is a conventional browser that uses JavaScript and the Yahoo UI library for more rapid interaction. We integrated two new features: distributed streams and broadcast feeds.

Distributed information streams. In one workshop, several participants video brainstormed ideas about how to manage and share their personal information scattered on the web. We decided to integrate this as a new type of activity stream, letting individuals decide which information they access and share. Many web 2.0 applications provide an API to access their content, but this requires additional programming. As a lightweight alternative, we decided to use the RSS and Atom feeds used by web sites to access recent information and incorporate them into the notebook as web streams. This new architecture allowed us to aggregate information from different sources without requiring further programming and also opened up new opportunities for sharing information.

Shared broadcast feeds. We designed a broadcast mechanism, based on Atom feeds, for participants to notify each other of different activities. It allows users to share notebook entries and to make streams available to others. Once a feed is created, users send its URL to subscribers by email. Publishers can attach tags to the feeds and subscribers can filter them according to these tags.

Longitudinal study results

Prism has been an effective technology probe, providing us with insights about how this particular group manages personal and shared information over time, as well as helping us to better frame the concept of *Prism* itself.

Sharing. All the participants worked actively with others, both in the team and outside. As they capture and reflect upon their activity, they often have an idea of who it is for and how it will be used in the future. We observed two broad categories of sharing: over time, to preserve traces of their activity for their successors (as well as themselves) and among colleagues, to share acquired knowledge and knowhow. As in the earlier study, some record activities only for themselves: "I never share my README files". However, even this researcher believes that her files might be useful for her successors: *"These notes are not clean, I can't share them with colleagues, but they could be used by people after me."* Others try to make their notes easier for others to decipher. Non-permanent staff, particularly students and post-docs, are especially encouraged to leave archived traces of their activity. For example, one post-doc shared her paper notebook with her supervisor, via the web. 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.

Enabling *Prism* for the web allowed it to become a central notification point. Some participants engaged in reciprocal sharing of bibliographic references and calendar events with each other. They also collaborated with remote colleagues. For example, one senior researcher set up a wiki for a conference subscription and used the feed to keep track of the modifications by subscribing to the change feed. Another captured web pages as she developed them, which created a series of images showing how her work had evolved, which she could share with her collaborators.

Organizing information: Prism concentrates information from heterogeneous sources into one central location. Some information is repeated or related from one stream to another, other information is repeated over time. *Prism* users 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. Participants also organized their data dynamically. One researcher explained:

"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".

Master Lab Notebook: However these strategies only help with finding information again, they do not help researchers or other possible readers to understand why things were done. As one manager explained: *"We need a master document that is the source of the different things"*. We found that all participants created such a reference point, or *master notebook,* to organize the diverse strands of their personal activity. Interestingly, the bioinformaticians who could use *Prism*'s paper notebook, turned that into their master notebook. The two Mac users who did not have access to a paper notebook still created a master notebook. One used a standard paper notebook not connected to *Prism* and the other used the electronic notebook. This master notebook is where researchers reflect upon their work: they comment on an interesting web page that was marked, use a hand-drawn sketch in the paper notebook to explain how some tasks are related to each other and add electronic notes to track how a project is evolving.

Feedback. 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". Several would have liked *Prism* to become the official record of patentable findings but this conflicts with the ease of sharing of information. We would need to find a better balance between the sharing and protection of information. 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 an experiment. They felt that this helped them to capture their activity without disturbing their activity. Even so, information organization remains a problem. Tags alone are no panacea. As organization schemes change, they become inappropriate or out of sync with user's current mental state.

DISCUSSION

Our studies of *Prism* demonstrate that: a stream architecture offers a useful, general way to publish and reuse information across applications; scientists' needs for reflection go beyond personal information management; master notebooks are a common reference point that can help researchers to manage and reflect upon their activities; and redundant information is a key element for reflection.

Stream architecture: Adaptable information sharing

When trying to complete a task, users clearly do not want to focus on the process of capturing information. On the other hand, automatic monitoring of personal activity can be intrusive. For example, although an early version of *Prism* included "Recent Documents" in the computer activity stream, participants preferred to mark relevant documents themselves, to limit the number of entries in the notebook and to control which data they published.

In an environment in which researchers share data, protocols and other forms of know-how with strangers who are both competitors and collaborators, it is important to have fine-grained control over what is shared with whom, when. The scientists in our study kept personal information in diverse forms, including paper, different local computers and the web. We needed adaptable tools that would allow them to integrate and share these diverse sources of information. We used two broadcast formats, Atom and RSS feeds, which can be read by people who do not use *Prism*, either to read or to publish information. This is convenient for users, who can find the information they need on the web, and for developers, who can easily support it.

Reflection: Beyond Personal Information Management

According to Latour [9] 'science in action' is not just about individual experiments unconnected to the environment and other activity, nor is it a clean process with clear starting and end points. The idea that scientists create a hypothesis, test it experimentally and publish the results is an over-simplification. Biologists also explore and reflect. For example, some biologists would launch an analysis on a large set of data, expecting to find something interesting, but not really knowing what. Experiments are only one part of a complex process that includes extensive collaboration, diverse forms of data analysis, the creation of specific analysis tools and, above all, reflection. *Prism* provides a first step, making it easier for researchers to review information from diverse sources and comment on it, providing a series of ongoing reflections over time (fig. 6).

Master Notebook

Although the web provides new opportunities for collaboration and data sharing, tracking the resulting flood of information across different media, platforms, and organizational structures has proven a major challenge for everyone we spoke to. Users need to manage this confusing mass of information, hidden in different locations and updated at different rates. *Prism*'s hybrid design links heterogeneous physical and digital information, making it ideal for what we call a *master notebook*. All participants constructed one, providing an organized account of what they did and plan to do. Interestingly, when given the choice, people preferred to use the paper format as the master notebook, even those bioinformaticians whose work was almost solely on the computer. The physical process of writing played an important role in this externalization of memory and helped them filter and organize information.

Redundancy as a resource for reflection

We found that the study participants valued redundant information because it helped them to reflect creatively about 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, a sort of reverse triangulation. When redundant events are linked to each other, scientists can often make better sense of them, and in

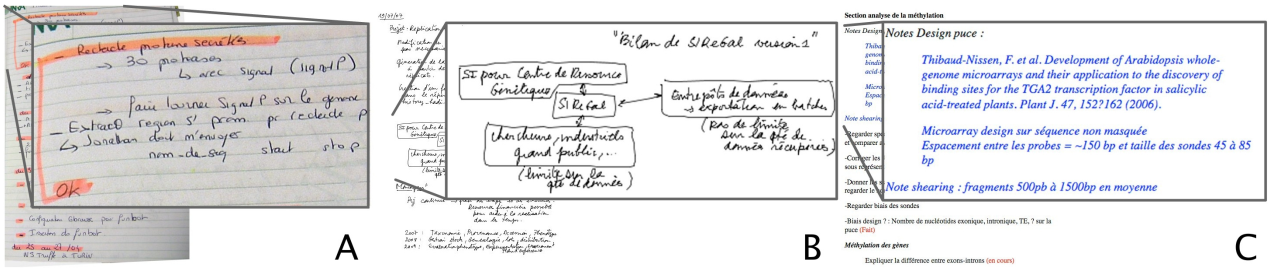

Figure 6: Reflection strategies in a Master Notebook A: Color coding and post-hoc remarks (paper notebook). B: Project organization (augmented paper notebook). C: Meta-notes shifted from the body and colored (electronic notebook).

some fortunate instances, redundancy allows them to discover new insights and move in creative new directions.

From a computer science perspective, permitting redundancy might be viewed as adding complexity and wasting space. Some electronic notebooks, such as ipad, force users to write information only once, in the 'right' place, and to use structured meta-data to annotate their data. However, from the user's perspective, such restrictions waste their time and is one of the main reasons that many users cited for rejecting such digital notebooks.

Real-world events are situated in a complex and rich context that users take advantage of when trying to reconstruct previous events. *A priori* filing is not always possible, since structures evolve over time and knowing which data to capture depends upon on both context and the user. In our experience, electronic notebooks should avoid restraining how users input information or deciding what users should capture. By providing multiple possible paths for finding information, users can revisit previous contexts and draw new connections. 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.

CONCLUSION AND FUTURE WORK

Our initial study highlighted the rapid changes in biologists' and bioinformaticians' work practices, as computers became an essential part of the laboratory. These changes created new opportunities for accessing, analyzing and sharing data, but also exacerbated information management problems. Rapid changes in how science is conducted require new thinking about the kinds and characteristics of the tools we create for scientists.

Even though an increasing number of research activities are conducted online, the majority of researchers in our study still use paper notebooks. This is not only because of the inherent flexibility of pen and paper, but also because they help researchers maintain a disciplined approach to recording and structuring their data.

In our second study, we used a participatory design approach to develop *Prism*, a hybrid laboratory notebook with both paper and electronic elements. *Prism* integrates multiple heterogeneous streams of user activity and web information and offers researchers an effective balance between discipline and flexibility as they record their activities.

We created *Prism* to be a technology probe, to help us learn about how biologists manage and share information and to inspire new ideas about future information management tools. We field-tested *Prism* over a period of nine months with a team of bioinformaticians. They transformed *Prism* into the central focus for all of their activities, both on- and off-line, organized around what we call a *master notebook*. They adapted *Prism* to support sharing of a variety of types of information, both within the team and with external colleagues. *Prism*'s stream architecture reifies [2] temporal events, allowing users to cross-link and interact with different types of activity. Participants developed new strategies, not only for organizing this information, but also for reflecting upon it. They used redundancy as a clue when reviewing and interpreting previous comments, which helped them derive insights and explore creative new research directions.

Future work will focus on enhancing the paper master notebook with richer forms of interaction, such as creating links and sharing notes directly on paper. We will also explore how to help users make better trade-offs when organizing information, whether chronological or project based. Finally, we plan to investigate alternative types of organization based on relationships among elements from different streams, to further support reflection and creativity.

ACKNOWLEGMENTS

We would like to thank the biologists and bioinformaticians at the Institut Pasteur and INRA, who generously gave their time and shared their insights with us. We would also like to thank Catherine Letondal and Michel Beaudouin-Lafon for fruitful discussions and comments on earlier drafts of the paper.

REFERENCES

1. Barreau, D. & Nardi, B.A. (1995) Finding and reminding: File organization from the desktop. In *ACM SIGCHI Bulletin*, **27**(3), pp. 39-43.
2. Beaudouin-Lafon, M. & Mackay, W. (2000) Reification, Polymorphism and Reuse: Three Principles for Designing Visual Interfaces. *Proc. AVI'00, Advanced Visual Interfaces,* Palermo, Italy. pp. 102-109.
3. Bondarenko, O. & Janssen, R. (2005) Documents at Hand: Learning from Paper to Improve Digital Technologies, In *Proc. CHI'05 Human Factors in Computing Systems*, ACM, pp. 121-130.

4. Butler, D. (2005) Electronic Notebooks: A new leaf. *Nature*, **436**, pp. 20-21.
5. Dourish, P., Edwards, W.K., LaMarca, A. & Salisbury, M. (1999) Presto: an experimental architecture for fluid interactive document spaces, *ACM Trans. Computer-Human Interaction*, **6** (2):133-161.
6. Guimbretière, F. (2003). Paper augmented digital documents. In *Proc. UIST '03 User Interface Software and Technology*, ACM, pp. 51-60.
7. Hutchinson, H., Mackay, W., Westerlund, B., Beder-son, B. B., Druin, A., Plaisant, C., Beaudouin-Lafon, M., Conversy, S., Evans, H., Hansen, H., Roussel, N. & Eiderbeck, B. (2003) Technology probes: inspiring design for and with families. *In Proc. CH '03 Human Factors in Computing Systems*, ACM. pp. 17-24.
8. Lamming, M. (1991) NoTime – A Tool for Notetakers. *Xerox EuroPARC Technical Report*. 91-135.
9. Latour, B. (1987) *Science in Action*. Cambridge: Harvard University Press.
10. Letondal, C. & Mackay, W. (2004) Participatory Programming and the Scope of Mutual Responsibility: Balancing Scientific, Design & Software Commitment. In *Proc. PDC'04 Participatory Design*, pp. 31-41.
11. Letondal C. & Mackay, W. (2007) Paperoles: The Paperoles Project: An analysis of paper use by music composers. In *Proc. CoPADD'07 Collaborating over Paper and Digital Documents.*
12. Liao C., Guimbretiere, F. & Hinckley, K. (2005). PapierCraft: a command system for interactive paper. In *Proc. UIST'05 User Interface Software & Technology*, ACM, pp. 241-244.
13. Mackay, W. (1994) Video Mosaic: Laying out Time in Space. In *Proc. Multimedia*, ACM, pp. 165-172.
14. Mackay, W.E. (2000) Responding to cognitive overload: Co-adaptation between users and technology. *Intellectica*. **30**(1):177-193.
15. Mackay, W. (1989) Strange Bedfellows: Research Alliances among Competitive Firms. *MIT Technical Report.*
16. Mackay, W. (1988) More than just a communication system: Diversity in the use of electronic mail. *ACM Trans. Office Information Systems* **6**(4):380-397.
17. Mackay, W. (1998) Augmented Reality: linking real and virtual worlds. In *Proc. AVI '98 Advanced Visual Interfaces*, ACM, pp. 13-21.
18. Mackay, W., Letondal, C., Pothier, G., Bøegh K. and Sørensen, H.. (2002). The missing link: augmenting biology laboratory notebooks. In *Proc. UIST '02 User Interface Software & Technology*, ACM, pp. 41-50.
19. Malone, T. (1983) How do people organize their desks? Implications for the design of office information sys-tems, *ACM Trans. Information Systems*, **1**(1):99-112.
20. Myers, J. (2003) Collaborative Electronic Notebooks as Electronic Records: Design Issues for the Secure Electronic Laboratory Notebook (ELN). In *Proc. CTS'03*.
21. Sellen, A. & Harper, R. (2003) *The Myth of the Paperless Office*. Cambridge, MA: MIT Press.
22. Schraefel, M., Hughes, G., Mills, H., Smith, G., Payne, T. & Frey, J. (2004). Breaking the book: translating the chemistry lab book into a pervasive computing lab environment. In *Proc. CHI'04 Human Factors in Computing System*, ACM, pp. 25-32.
23. Schraefel, M. & Dix, A. (2007) Within Bounds and Between Domains: Reflecting on Making Tea within the Context of Design Elicitation Methods. *Int. Journal Human Computer Studies*. in press.
24. Stifelman, L., Arons, B. & Schmandt, C. (2001) The audio notebook: Paper and pen interaction with structured speech. In *Proc. CHI'01 Human Factors in Computing Systems*, ACM, pp. 182-189.
25. Stokstad, E. (2001) Information overload hampers biology reforms. *Science*. **31**(293):1609.
26. Tabard, A., Mackay, W., Roussel, N. & Letondal, C. (2007). PageLinker: integrating contextual bookmarks within a browser. In *Proc. CHI'07 Human Factors in Computing Systems,* ACM, pp. 337-346.
27. Trigg, R., Blomberg, J. & Suchman, L. (1999). Moving document collections online: the evolution of a shared repository. *In Proc ECSCW'99 European Computer-Supported Cooperative Work*, pp. 331-350.
28. Von Hippel, E. (1986) Lead users: a source of novel product concepts. *Management Science* **32**(7):791-805.
29. Wellner, P. (1993) Interacting with paper on the DigitalDesk. *Comm. ACM* **36**(7):87-96.
30. Weibel, N., Ponti, P., Signer, B. & Norrie, M. (1997) PaperProof: A Paper-Digital Proof-Editing System. In *Proc. CoPADD'07 Collaborating over Paper and Digital Documents.*
31. Wilcox, L., Schilit, B. & Sawhney, N. (1997) Dynomite: A dynamically organized ink and audio notebook. In *Proc. CHI'97 Human Factors in Computing Systems*, ACM, pp. 186-193.
32. Wouters, P. (2004) Public proofs, science, technology and democracy: What is the matter with e-Science? In *4S & EASST Conference*, Paris.
33. Yeh, R., Liao, C., Klemmer, S., Guimbretière, F., Lee, B., Kakaradov, B., Stamberger, J. & Paepcke, A. (2006) ButterflyNet: A Mobile Capture and Access System for Field Biology Research. In *Proc. CHI'06, Human Factors in Computing Systems*, ACM, pp. 571-580.
34. Yeh, R., Paepcke A. & Klemmer, S.R. (2008) Design and Evaluation of an Event Architecture for Paper UIs: Developers Create by Copying and Combining. In *Proc. UIST '08 User Interface Software & Technology*, ACM, in press.