Study of an Electronic Lab Notebook Design and Practices that Emerged in a Collaborative Scientific Environment

Gerard Oleksik
Dovetailed Ltd
23 Kings St., Cambridge, UK
gerard@dovetailed.com

Natasa Milic-Frayling
Microsoft Research
21 Station Rd., Cambridge, UK
natasamf@microsoft.com

Rachel Jones
Instrata Ltd
12 Warkworth St., Cambridge, UK
racheljones@instrata.com

ABSTRACT
Prolific adoption of digital media across scientific fields has led to inevitable transformation of a traditional lab book into an *electronic lab notebook* (ELN). Research so far has focussed on designing ELN prototypes and learning from their limited deployments. At the same time, a variety of commercially available ELNs have been adopted by industrial and academic laboratories. That provides opportunities for situated research and a deeper understanding of the role that ELNs assumes as an integral part of a scientific environment. In this paper we present a study of ELN design that has emerged as scientists appropriated commercial off-the-shelf note-taking software and adapted it to their work. Through in-situ observations we analysed the interplay between the technology and emerging practices. Our study reveals a tension that is intrinsic to the digital nature of ELNs: a conflict between the flexibility, fluidity, and low threshold for modifying digital records and the requirement for persistence and consistency. This led to refined requirements and design considerations for ELNs.

Author Keywords
Electronic lab notebook; content organization; practice; linking; tagging; pivots

ACM Classification Keywords
H5.m. Information interfaces and presentation: Misc.

INTRODUCTION
A scientist's lab notebook is a lynchpin of scientific record keeping. Scientists use it to keep daily records of the research progress, the developing thoughts, the findings, and the experiments they have run [15]. Traditionally a paper artifact, the lab notebook has been destined to undergo considerable transformations as the digital technology penetrated many aspects of the scientific work.

CSCW '14, February 15–19, 2014, Baltimore, Maryland, USA.
Copyright 2014 ACM 978-1-4503-2540-0/14/02...$15.00.
http://dx.doi.org/10.1145/2531602.2531709

Designers and HCI researchers have produced prototypes of the *electronic lab notebook* (ELN) to study implications of these transformations [7, 12, 15]. Kolkmose and Zander [3] gathered an exhaustive list of requirements for the future ELNs based on a series of workshops and discussions with a community of physicists. Studies of the developed ELN prototypes have considered the ways the electronic version of the lab notebooks can foster collaboration [15], create affinities between electronic and paper resources [5, 17], introduce semantics [16], and streamline data gathering and analysis workflows [12]. In many instances these studies involved a limited deployment of the prototype. While they led to valuable insights about the potential use of the prototypes, they also ascertained the need to extend the research inquiry through observation studies of ELNs that have been adopted as an integral part of a scientific environment.

In recent years, we have seen a steady rise in a variety of ELN solutions offered by commercial vendors. Rubacha et al. [10] provide a market survey of 35 commercially available ELNs grouped into 5 categories based on the primary audience: two categories related to the scientific fields of biology and chemistry, two categories focused on the functional objectives such as support for research and development (R&D) and support for quality assurance and compliance (QA/QC), and a class of ELNs with cross-disciplinary offerings.

Several publications from the practitioners describe the advantages of deployed solutions measured by the productivity, compliance, and quality assurance metrics [9, 18]. Recent work by Iyer and Kudrle [2] reports on the use of ELN to support a course curriculum in a bioanalytical laboratory. Although these papers provide accounts of adopted commercial ELN solutions, none of them follow observational and situated research methodology and, thus, offer no deeper insights into the role that ELNs play in the work of individuals and teams.

Our research fills this gap by observing the ELN design and practices that emerged within a fully operational scientific environment as a means of supporting cutting edge research and graduate level education in nanotechnology. We conducted a two-week observation study of a team of 8 scientists at the Nanophotonics Research Centre (NRC) who created and used ELNs in their everyday work. The NRC scientists adopted tablet PCs and generic, commercial,

Figure 1: Use of the ELN in the NRC lab (above). The laboratory environment that ELN is interfacing with (below).

off-the-shelf software for note taking and turned them into ELNs. By closely following scientists' activities, we analyzed the use of ELNs for conducting experiments, processing data, sharing insights, and coordinating work. We captured interactions with ELNs and organizational structure of the created records. Based on the collected data, we gained valuable insights about the emerging practices and the influence of the organizational culture and the technology infrastructure on the use of ELNs.

Our research confirms most of the findings from the studies of ELN prototypes such as Prism by Tabard et al. [15], and the ELN design requirements that were solicited from the physics scientists by Klokmose & Zander [3]. However, situated research of adopted ELNs enables us to take a step further and understand the issues that arise in practice from the intrinsically digital nature of ELNs. First, we observe the tension between the flexibility, fluidity, and low threshold for modifying digital artifacts and the requirements for persistence and consistency of scientific records across individuals. Second, we uncover how these two aspects have a ripple effect on the collaborative work. The flexibility of digital media enables personalization and optimal support for needs of individual scientists while the conformity and consistency in the scientific record increases the capacity for sharing and leveraging each other's work.

Designing ELNs to achieve a desired balance between these opposing forces is critical for the practical deployment of ELNs and a sustained use to support scientific work. Constraining the format of data representation and record keeping in ELNs is commonly promoted by the commercially available solutions through features such as templates for data entries and fixed workflows [10]. That may be suitable for routine and well defined experimental work. However, for a wider spectrum of usage scenarios, where flexibility is preferred, we propose an alternative approach that applies computing technologies, such as text mining and natural language processing techniques, to extract information from free-form or semi-structured scientific records and present it into a consistent and unified view.

In the rest of the paper we first provide an overview of the research on ELNs and scientific record keeping. We then follow with a detailed description of the study method and observations. We discuss the implications of the findings for the design and conclude with a summary of presented research and future directions.

RELATED WORK

From a broader perspective, our research can be situated within two distinct strands of investigation, both concerned with the role of computing technologies as a support and a catalyst for discovery and innovation.

First, our methodology is inspired by the ethnographic studies of science and technology that contextualize science and consider "institutional circumstances of scientific work" as promoted by Latour and Woolgar [4] and Lynch [5]. Our analysis of the ELNs adoption was part of a broader effort to understand the NRC work environment and characterize the ways the computing technologies support scientific activities. Through in-situ observations, we captured the social interactions and the intricate digital ecology that evolved from the practices and supported the scientific discovery [6]. We used this deeper understanding as a backdrop for the analysis of the practices around the ELNs use.

Second, our attention was particularly drawn to the design of ELNs adopted at the NRC by the fact that the scientists appropriated and adapted generic note-taking software for their purposes in novel and innovative ways. This provided a unique opportunity to confirm in practice the position by Silverstone and Haddon [14] that "innovation requires to be seen as a process which involves both producers and consumers in a complex interweaving of activities". The observed behavior and attitude towards note-taking technology by the scientists are aligned with the notion of technology *domestication* where consumers of technology are involved "in appropriation, in taking technologies and objects home, and in making, or not making, them acceptable and familiar" [14]. More importantly, this frame of thought led us to reason about the interrelationship of software design and software appropriation in a way that enable us to identify specific features that give rise to the extended use. In effect, our work illustrates and confirms that the innovation does not stop with the technology design; instead, it evolves through appropriation and use over time.

In order to gain deeper insights about the characteristics of ELNs designs and practices, we focus our literature review on two research areas. First, we consider the studies of scientific record keeping and information management that investigate the role of the paper lab notebooks in the scientific practices. We then discuss the literature about research prototypes, ranging from augmented paper lab notebooks to full ELN applications, and the insights provided by that body of work. At the end, we reflect upon the state of the commercial solutions and outline the research questions that underpin our study.

Paper lab notebook and scientific record keeping

The lab notebook is a cornerstone of the scientific record keeping and often serves as a legal document and proof of discovery in patent claims [3, 16]. The work by Yeh et al. [17] and Mackay et al. [7] draws attention to the heterogeneous information types that scientists physically paste into lab notebooks, in some instances including X-ray images, photographs and computer printouts. In particular, Mackay et al. [7] describe the paper lab notebooks created by biologists as 'extremely multi-media documents'.

Schraefel et al. [10] observed the use of paper lab notebooks by chemists in their day to day work within chemistry labs and devised an ELN that was practical and useful for data entry and access. Tabard et al. [15] studied the use of paper notebooks in a biology lab, alongside electronic folders. Despite the ubiquity of electronic files, many participants continued to use paper lab notebooks. Indeed, due to the diligence required to complete these paper records and their fixed, un-editable nature, these paper artifacts continue to play an essential role in record keeping: 'paper lab notebooks serve as a formal repository for information that can be treated as a stable point in an evolving world' [15].

The chronological structure of a lab notebook enabled participants to find easily recently entered information. However, over time, the reliance on the chronological memory broke down and a project level categorization became more desirable. Investigating models for organizing the data and records throughout the lab, Tabard et al. [15] found that chronology and project based organization are, indeed, the two dominant strategies.

Augmented paper and electronic lab notebooks

While paper notebooks provide a rich body of contextual information, they are, unfortunately, not broadly accessible and shareable [16]. Furthermore, a highly personal style of information keeping in some instances, and a lack of *a priori* information about the content they include, may hinder search and reuse of paper notebooks by others [11]. In order to address these and other issues, a body of research has sought to augment the paper notebooks or replace them entirely with electronic versions.

Schraefel et al. [10] identified four main approaches to designing ELNs, i.e., those seeking to replicate, supplement, replace, or augment paper lab notebooks. Systems and applications that aim to replicate paper notebooks typically enable notebook pages to be scanned directly into searchable databases (e.g., software systems by SCRIP-SAFE deployed to support intellectual property claims [13]). Those that supplement the paper notebooks may focus on supporting some parts of the scientific process, such as entry of experiment planning data. For example, the software may produce a printed output which is then pasted into the paper notebooks (e.g., Labscape Lab Assistant by Borriello et al., [1]). The a-book (short for augmented notebook), designed by Mackay et al. [5] extends the reach of the traditional paper notebook by allowing the user to associate digital artifacts with handwritten notes via magic lens interaction.

Tabard et al. [15] conducted a longer term study of Prism, a hybrid paper electronic notebook aimed at gathering data from distributed information stores that biologists use in their work. The Prism prototype consists of Anoto paper notebook which tracks user handwriting and stores digital entries into an electronic lab notebook. The users can then tag ELN pages, send email and Web content, and link to documents and images. The digital pages can be shared through RSS feeds, thus increasing the visibility and reach of information. The prototype was used by 6 biologists in their daily work for 9 months. The authors found that tagging, although well used, was of limited benefit since scientists found it hard to recall the meaning of tags as time passed. On the other hand, the use of the tool was extended with sharing bibliographic references and calendar events.

While Mackay et al. [7], Yeh et al. [17], and Tabard et al. [15] pursued the augmentation of paper notebook, Schraefel et al. [10] have produced a wholly electronic lab notebook prototype. They developed a tablet based ELN and designed the software to mimic various affordances of the paper notebooks. Talbott et al. [16] sought to enhance ELNs with semantic features. They implemented the information architecture and described the integration of the lab notebook software with semantic services. However, they did not report on the system performance and user evaluation.

Research objectives

The discussed research literature illustrates a great interest in providing scientists with an effective means of creating rich lab records. The available commercial solutions now include a broad range of features, from general to domain specific support for record keeping activities [10].

For example, Kalabie ELN, a product of Agilent Technologies, supports scientists in creating experiment records through predefined experiment templates, copying and pasting of data between experiment records, providing alerts for signatories, and enabling rich text and media management. CERF-Notebook by Rescentris promotes both general features for R&D and specific support for biology research. It has built-in ontologies to support customization of the scientific data models, templates, and business policies. As R&D functions, it enables creating, viewing, analyzing, and annotating research records, enforcing the intellectual property protection, and integrating scientific records with existing systems and data. For biology research, it offers content management, search using full-text index, metadata, and controlled vocabularies, and features to enforce compliance with templates and forms for protocols and data capture. These two examples of ELN products are illustrative of features that are typically offered by bespoke ELN solutions.

The NRC has made a conscious decision not to use specialized ELN products for two reasons. First, to reduce the risk of software obsolescence since the market is volatile and the providers may not be able to guarantee software support for a long time. Second, most of the commercial ELNs' functionality is supporting mainstream processes and operations which are not relevant for the academic environments where the focus is on investigative research and innovative approaches. Interestingly, during the study the NRC researchers had to consider the upgrade from OneNote 2007 to OneNote 2009 which caused concerns about the backward compatibility and possible implications for their work.

The opportunity to conduct in-situ study of NRC has proven invaluable because we could study the emergent ELN design as the participants adopted the technology to support their scientific work. That enabled us to compare the observed ELN functions with the bespoke ELN prototypes and commercial solutions. More importantly, we began to understand the impact of ELNs on the individual's and collective work. Our research covered three main areas:

- Properties and practices enabled by the ELN in use at NRC
- Constraints of the NRC design of ELNs, as observed through workarounds and interviews with the scientists
- Design requirements and principles that can address the observed issues.

STUDY PLAN

Over a period of two months, we conducted three one-week observations within the scientific lab at the Nanophotonics Research Centre (NRC). The broad objective was to understand the entire work environment and characterize the ways the computing technologies support scientific activities [6]. Our observation centered on the work practices and movement of 8 study participants (P1–P8, Table 1). However, we also observed and investigated other aspects of the lab life such as group meetings and general interaction among the staff. During the first week we conducted initial interviews with all the participants and started observations of the meetings and experiment procedures. This provided a detailed overview of the Centre, its organization, spatial layout and IT infrastructure for storing data and managing information.

Part.	Profile	At the Lab
P1	Professor. Founder of the Centre	18 months
P2	Post Doc – multiple projects	18 months
P3	Post doc – multiple projects	18 months
P4	PhD Student – Nano group	10 months
P5	PhD Student – Nano group	6 months
P6	PhD Student – Nano group	10 months
P7	PhD Student – Engineering group	10 months
P8	PhD Student – Nano group	18 months

Table 1: Study participants

During observations of meetings, the researcher video recorded the proceedings with a static camera and took notes throughout. The lab observation of experiments (Figure 1) was also video recorded, encouraging the participants to 'speak aloud' about their tasks and reasoning behind their activities. The researcher also asked the participants occasional questions to clarify the context or motivation for a particular activity, e.g., a reason for recording a particular piece of information into the ELN

Alongside observations, we conducted up to 3 in-depth interviews with each participant. Interviews lasted between 60 and 90 minutes and covered a broad range of aspects of the participants' work. We asked participants how and where they stored information, what collaborative and communication methods they used, how they ran the experiments in the lab, and how they used and managed their ELNs. We also gathered entire electronic lab notebooks from the participants and analyzed their structure and content.

We analyzed the collected data by applying the Thematic Coding method (Miles & Hubermann [8]) and using the software package Atlas.ti to aid the data processing. The analysis involved 15 hours of recorded and transcribed interviews and 5 hours of video footage of meetings and lab observations. Throughout the coding process, two researches discussed and refined the emerging codes. Over several sessions, the codes were clustered to identify the emerging research themes, concepts, and categories. During the analysis, samples of OneNote workbooks were viewed in order to understand the information structuring practices of the participants.

THE FIELD SITE AND STUDY PARTICIPANTS

The Nanophotonics Research Centre (NRC) is a nimble and active research environment, hosting a wide range of inter-connected and collaborative projects. The Center is oriented towards high impact science and exploration of new approaches in the field. The group is headed by a professor with many years of experience in industry and academia (P1). He is the founder of the Centre. At the time of the study the Centre employed 5 Post Docs, 11 PhD research students, and 5 support staff who were involved in 5 major research initiatives, comprising 12 sub-projects in total.

The staff applies highly sophisticated lab equipment to material samples that have been specially produced to facilitate scientific experiments. Furthermore, each scientist is provided with a tablet PC running Microsoft Operating System and Microsoft Office applications. MS OneNote 2007 was adopted as a lab notebook to create records of research activities. However, it also served as a means of collaboration and general information management. The use of Tablet PC provided the required mobility because the researchers conducted work both in the Lab and at their desk in the common research room. Furthermore, they needed access to their ELNs in meetings that took place in the professor's office and during discussions with their peers.

While OneNote is not bespoke lab notebook software, its functions have been co-opted to this purpose. Each scientist maintains a personal ELN, daily recording notes on the experiments and analyses, and importing supporting materials such as key graphs and Web resources to produce an account of progress for themselves and for discussions with others. As a collaboration tool, the lab notebook pages are shared back and forth between research staff to gather feedback and comments. In addition, the professor creates shared MS OneNote files to support discussion and note taking during meetings and project level notebooks to store information from sub-projects into a central repository. The latter are intended as a detailed record of the scientific work conducted in the Centre.

Both the experiment data and all OneNote files are stored on the networked servers in the lab and all the research files are accessible to everyone in the Centre. Most of the data analysis is performed by using Igor, a specialized signal processing and visualization software.

Microsoft OneNote. MS OneNote 2007 software is designed based on a notebook paradigm. The top level concept is a binder of notebooks. Each notebook supports a three level organization (Figure 2): section list (on the left), sub-sections tabs (on the top) and list of individual pages (on the right). Unfolding the hierarchy enables the user to switch between pages, sections, and notebooks.

MS OneNote on the Tablet PC supports both handwritten and typed input and maintains editing history. Most of the content formats can be imported by the standard drag&drop and copy&paste facilities. This includes the embedded files, represented as file icons or displayed fully in the pages, and files referenced by the file paths or URLs in case of Web pages. The user can annotate and tag pages and parts of pages through handwriting as well as recorded voice. The content within each notebook can be accessed by browsing, searching for tags, or through keyword search.

OneNote is integrated with other Office applications, allowing the user to direct content to OneNote from MS Word, MS Outlook (email client) and MS Internet Explorer (Web browser) through a 'send to' command. Furthermore, the user can link notebooks and individual pages from different notebooks. Most of these technical aspects of OneNote were articulated in the ELN requirements collected from physics scientists by Klokmose & Zander [3] and can be found in the specifications of the commercially available ELNs.

FINDINGS

We group the study findings into two broad categories: (1) the frameworks that participants applied to organize content within their personal ELNs and how effectively that organization supports their work, and (2) the ways that the participants use their ELNs throughout their research practice. Our analysis did include the usage of specific application features by the participants but we report on the intentions of the participants and achieved results rather than the mechanics of the technical features. We will, however, refer to them when discussing the users' workarounds. As it turned out, some of the OneNote features fell short in supporting the user. For example, the search is focused on the page content and not specifically on annotations. Second, the structure of section headings provides a fixed organization and does not support pivoting around specified headings to enable alternative views.

Information management

With the encouragement of the research leader (P1), the staff has adopted a naming convention of ELN records that includes the date, material sample, and experiment name. However, the organization of ELNs is largely under the control of individual scientists and varies considerably.

ELN Record Organization

From the interview data and analysis of the lab notebooks, we observe a fundamental tension between the content organization based on chronology, as suggested by the lab-wide convention, and the organization by semantic sections that some of the staff found useful.

Topical organization. Participants used OneNote sections to create areas in their notebook for recording notes and data about specific subjects, sub-projects, meetings, and other aspects of their work. While semantically consistent sections provide a structure for recording data, this approach has proven problematic due to the nature of the collected data. In their initial months at the Centre, the participants P2, P3 and P6 attempted to organize ELN content solely by subject such as sub-project, meeting notes, material samples or equipment used in the experiments.

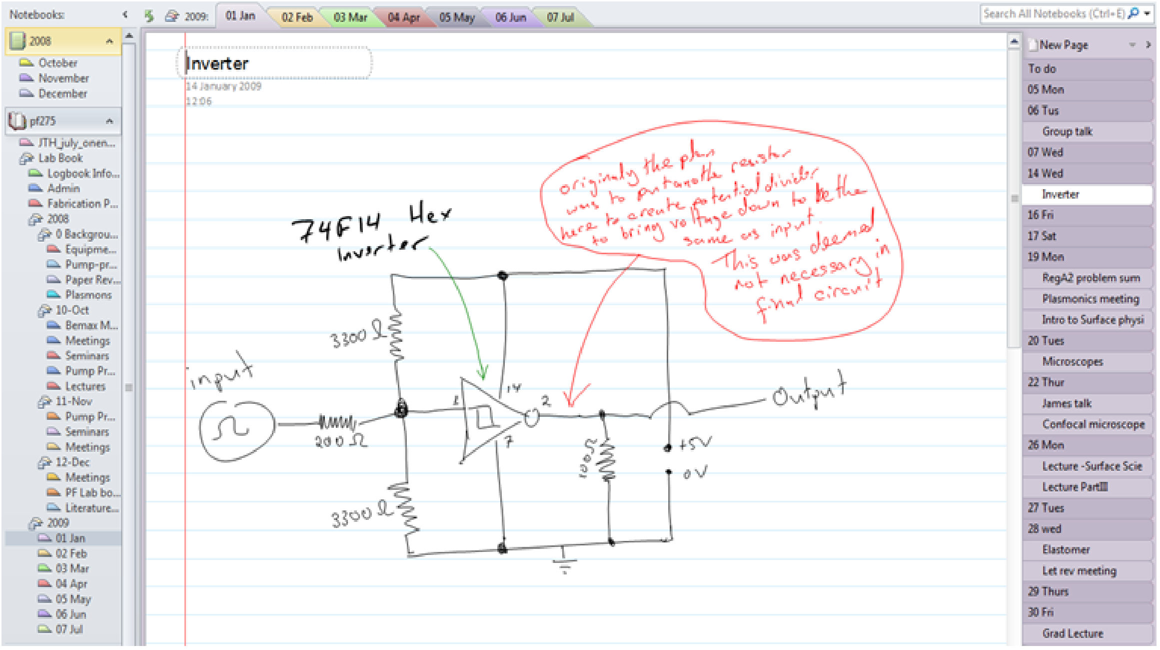

Figure 2: Example OneNote ELN page. Left pane displays open notebooks; upper tabbed pane displays sections; right hand pane displays individual pages.

While that helped with finding pages, each of them abandoned this approach due to the overhead it imposes when gathering information. The diversity of information grew and its classification became more difficult with a larger number of categories. The only participant who operated a fully topical organization of the ELN records at the time of the study was P5. He organized his data according to the material samples and experimental procedures. Significantly, he was the newest PhD student, having been with the group for only 6 month, and was already considering a change to the chronological structure as his data became more diverse.

Chronological organization. At the time of the study, chronology was the most prevalent organizational principle used by the participants (Figure 3). Chronology imposes a low overhead, allowing the users to record information without concern for categorization and, thus, without disrupting the experimental or analytic work.

Mixed approach. Similarly to the findings by Tabard et al. [15], we found that chronology was effective for searching recent information but not so for revisiting information that was recorded long ago. P2 stated that the transition point occurs around 2 weeks after which project and topic based information would be more useful for searching. Participants P3, P4, P6 and P7 attempted to overcome the issues of the sole chronological approach by replacing the date with keywords that describe the content, e.g., the name of the material sample, the name of the experiment, etc. (Figure 3). During the interview, participants stated that such a method helped when revisiting information since they preserved the chronological order of pages and more easily ascertained their contents based on the title. Furthermore, the date of creation is revealed on mouse hover, providing more specific chronological information where necessary.

This was very important to P8 who was applying a very loosely structured chronological approach, recording only the date of each page. P8's research was highly unstructured, engendering high degrees of uncertainty, and that was mirrored in the organization of his ELN:

"Everything I do is just exploratory. It's trying, 'oh, I wonder what will happen is I take this measurement or that measurement'… So it's a lot harder to keep a rigid structure of how you store everything… I don't really produce large amounts of data yet, I don't do systematic scans… When this experiment starts working, then it will be more systematic and you'll get a nice method of doing it."

Tagging Strategies

Tags were used to resolve a tension between organizing data by chronology and subject categories. OneNote's tagging functionality enables the user to tag their notes using a list of predefined tags or icons to which they can assign labels. While tags can be added at various levels of granularity, from individual sentences to paragraphs or entire pages, the majority of tags were added at the sentence level. Of the two participants who used tags, P2 was more prolific, creating a total of 21 different tags and applying them to over 200 objects within his ELN. Nearly half (45%) of tagged items were related to specific projects while just over a quarter (27%) referred to issues around laboratory equipment or computing. About 18% of tags were used to highlight key ideas and issues to return to and the remaining 10% were assigned to meetings or miscellaneous items.

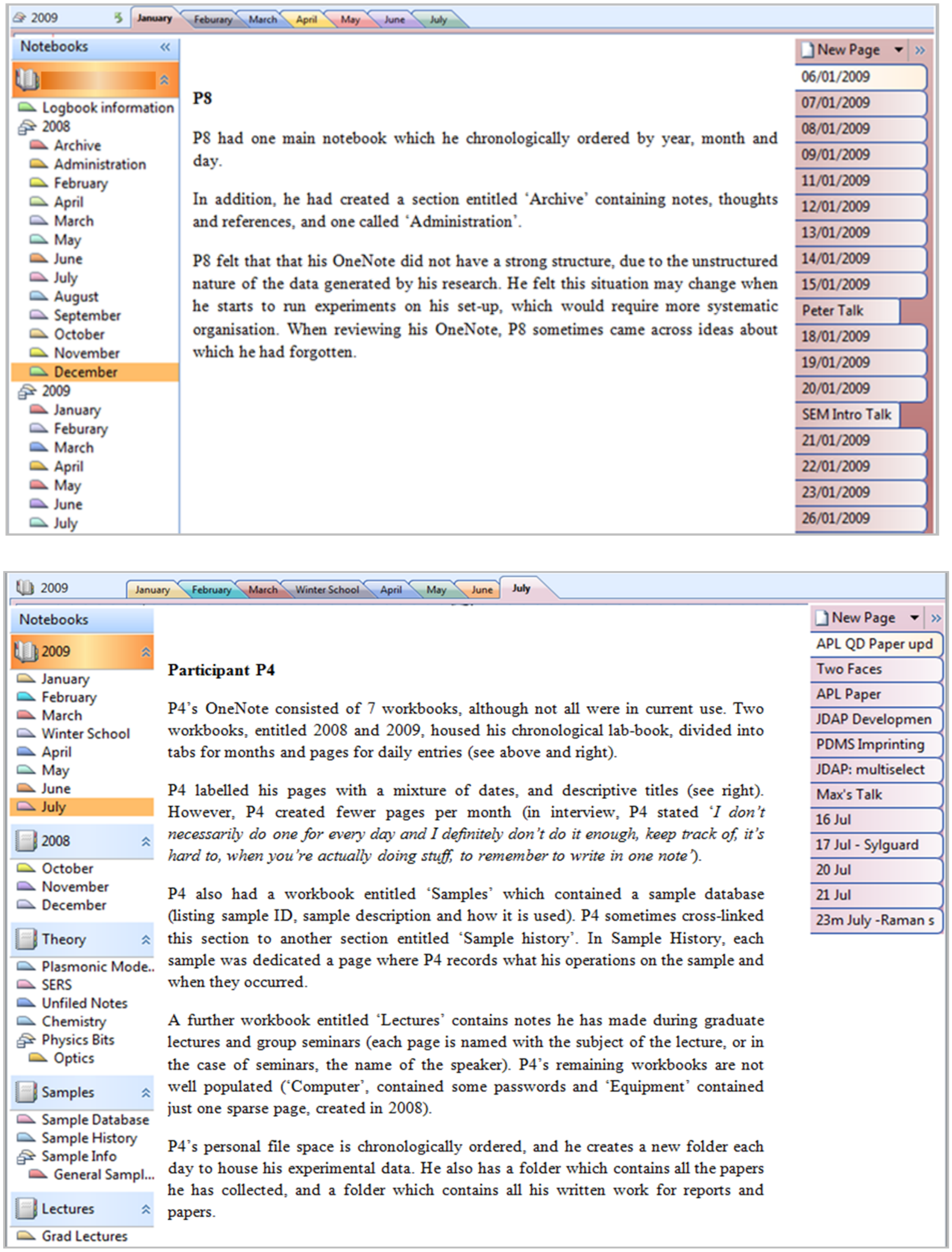

Figure 3: Examples of the chronological (top) and mixed chronological and topical organization of the ELN (bottom).

The high proportion of project tags tallies with P2's comments about his use of tags. He felt the need to split the chronological flow of records into projects categorization to help with revisitation after a longer time period, when the time based clues fade from his memory. While discussing his ELN structure, P2 stated:

"In OneNote one of the problems we have when you're doing it day by day if you're working in multiple projects its quite tough... I'll go back in time to last year when I was working on multiple projects... One of the problems you have is that you really don't know what, ... you know what you've done on a certain day, but you don't know when was the last time you were working on a certain type of project."

In the end, P2 was let down by the tag search features of OneNote which do not scale well nor offer sufficient options to filter results. Consequently, P2 had ceased to use tags except for the 'to do' tag to indicate required or planned actions.

P7's use of tags was quite different. He introduced them to increase the 'readability' of his lab notebook. Tags enable him to scan pages quickly and see what types of

Tag category	Tag name	Symbol	Applied
Actions	To do		500+
Highlighting / surfacing	Idea		23
	Important		140
	Question		32
	Critical		1
	Highlight		6
Contacts	Contacts		99
Other	Address		3
	Topic		1

Table 2: Tags applied by the participant P7.

information they contain (see Table 2). He used 9 different OneNote tags. Eight of these tags were applied in 305 instances. In two third (66%) of cases, the tags were used to highlight information such as research questions raised by the experiments, or to surface important details, such as password information. The remaining third (33%) designated contacts. The ninth, most frequent, was 'to do' tag, applied over 500 times to denote tasks and to keep track of progress.

Linking Strategies

Participants also used links to add structure to ELNs and connect to external resources. They created links to ELN pages, web pages, experimental data, local resources, and pages from another researcher's ELN. Table 3 shows the frequency and types of links included in OneNote by 6 participants. For each reviewed ELN we extracted and categorized all the links from OneNote pages. The analysis included OneNote sections with high concentrations of links, such as lists of links to material samples.

Half of the considered notebooks contained links within or between ELN pages. In the majority of instances, these links pointed to notes on previous experiments or operations relevant to the experiment currently being documented. This enabled researchers to reflect on previous relevant findings in the context of current research. In some instances, links were grouped and nested into tight clumps to pull together related information and provide an overview of a particular area. For example, P3 produced an ELN section entitled 'Summary' which, akin to bookmarks, contained a collection of links to key pages in the previous year. In the interview, P3 remarked that having a central access to these links saves him an effort of continually re-finding these important pages. This Summary page contained links to 36 ELN pages and 5 key data sets on the lab's servers.

P7 maintained an ELN section titled "Sample List" which contained a list of material samples that he had used in experiments. Each time P7 used a material sample, he created a link from the ELN record to the sample list and vice versa. This enabled him to keep track of all his activities related to a particular material sample.

Area	Link to	P2	P3	P4	P5	P7	P8	Tot.
Web	Web info sources	17	13	2	6	4	4	**46**
	Web info on equip.	18	5	4	2		8	**37**
Samples	Samples list			52		20		**72**
Data	Raw data on server	12	23		3	10		**48**
	From other researcher		6					**6**
OneNote	Other pages	16	14			10		**40**
Meta-summary	Multiple pages		41					**41**
Shared resources	Manuals					2		**2**
	Shared document	11			1			**12**
	Scripts, commands					6		**6**
	Other's document				2			**2**
	Embedded program	6						**6**
Own docs	Locally stored PDF	1			1	6		**8**
	Own document	1	4					**5**
	Embedded Word doc	4	1					**5**
Summary	Own summary	5			1			**6**
	Other's summary	3						**3**
	Total	**94**	**107**	**58**	**16**	**58**	**12**	**358**

Table 3: Linking practices of six study participants.

The most frequently linked information sources within ELNs were webpages. Some links were auto-generated by OneNote when the participants pasted text or images from a Web site. Others were directly dropped from the Browser into the page by the user. Four participants created links between ELN pages and raw data such as data files, images, videos, etc., stored on the lab's servers. Linking to another researcher's ELN was rare; P3's ELN provided the only example of this.

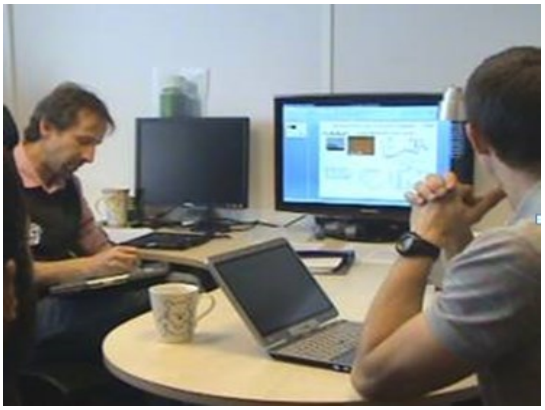

Figure 4: A meeting setting where the shared ELN is shown on the large screen and accessed from the shared network drive. Participants can interact with the shared OneNote from their Tablet PCs.

Desire for Structure

While the participants already applied structure to record keeping through ELN using section tabs, tags, and links, they all wanted more means of structuring the data. Ability to pivot on specific entities or attributes was a desire consistently expressed by the participants. For example, due to inter-related nature of projects, material samples were often reused across projects, thus creating a natural association among experiments that shared the same sample but considered different aspects.

Each participant spoke of the wish to be able to access all the information related to a particular lab item, for example a specific material sample, and immediately gather all the information relevant to it—experiments that had been run with it, by whom, when, and with what results. The value of aggregating this information in meta-summaries was highly praised and widely supported amongst the participants.

Interestingly, the aspect of alternative representations and pivot views emerged in another context of the study. Reflecting on the flexibility of the ELN and the ease with which data may be moved within, P6 stated that the software negatively affected the diligence with which he kept his notebook and his overall attitude towards the ELN records that were meant to be comprehensive and fixed snapshots of the research process. In effect, the ELN's flexibility interfered with the lab notebook as an ongoing and stable chronicle of work and disturbed the intrinsic context which arises from it. When asked what may help to reinstate this context, P6 replied:

"I guess if each piece of information was related to something, so if someone saw this and immediately knew that this was linked to this experiment on this sample, then that would bring all the context you need basically, because if someone looks at this and has no idea what I was doing here.. But if this was linked to the experiment, the sample and the original data, then someone can completely make sense of it, you'd get all the context you need."

ELN use in collaborative scientific work

By observing lab activities, we identified how participants used the ELN during data collection, analysis and group interpretation of data. Here we illustrate the benefits and shortcomings of the adopted ELN in supporting collaborative work.

Management of collective information resources

The common practice was to create a shared OneNote workbook for each project. A project workbook contains the notes that the research leader P1 keeps for all meetings on behalf of the group members, plus the sections for collaborative activities, such as drafts of papers that were written as a group effort. The Plasmonics workbook, for example, also had sections for individual experiments, which had been part of an effort *to pool information by experiment across the group. However, these sections were not well populated and had become defunct. P8 explains some of the reasons for this:*

"*I think we did try... a shared logbook for each experiment... It seems like a very good idea on paper but it didn't catch on. I think it's because people felt that they're writing all the information in their personal logbook then why is there a need to reproduce that information again just to write in the general logbook*?"

Thus, while there was a collective desire to capture, as a team, all the information related to their experiment, the practicalities of recording information in two places proved too costly relative to the perceived benefit of individuals.

Group communication and feedback

Analysis of the group meetings and ELNs contents revealed that researchers used ELNs to collaborate synchronously and asynchronously with group members. Participants reported cases when they wished to gain rapid feedback on a particular analysis and had sent links to note pages to the research leader P1 who then made additional notes and annotations. This was easily facilitated since both the experiments and the ELNs are stored on the lab servers that are used as common computational resources. Coupled with the open access policy, this meant that the ELN pages become shared collaborative documents. They enable rapid circulation of information, improving the speed at which ideas are circulated and assistance and feedback are provided between a researcher and the lab leader. Similar ideas were promoted by the ELN architecture implemented by Talbot et al. [16].

A shared OneNote notebook was used for synchronous collaboration during group meetings. Tablets with ELNs and a large display for viewing the shared notebooks enabled efficient running of the meetings (Figure 4). All the participants had an opportunity to contribute to the meeting and share notes while updating their individual ELNs. The common ELN notes served as a record of discussions, conclusions, and plans for further actions. Participants used this resource to remind themselves of the tasks they needed to perform between meetings.

Group tagging and summarization

P4 described OneNote tags as being '*pretty useless... for what we want to do*', but felt that the notion of tagging could be highly beneficial to the collective organization of information and collaboration within the lab if correctly implemented. He felt that tagging would be best served through a '*central tag system*' which would support standardization in information management and, thus, efficient pooling of information. Reflecting on the usefulness of this, P4 stated that, when working with a new chemical or substance, he would like to see immediately who had worked on the chemical, what results they had derived, which experiments the chemical had been used in, etc. He felt that the time and date would also be important information to turn into tags since that information was relevant to the everyday work in the Centre.

Tagging could be a possible facilitator of meetings. We noted a recurrent desire for easy access to a summary of information related to an entity, e.g., a project, a process, and similar. All the participants stated that it would be beneficial to have a quick access to the entirety of information related to a particular item, for example a material sample used in the experiments. P4 spoke very clearly about this issue and wished to be able to access summaries for samples which pull together experimental results, prepared notes, images, people who had worked on it, relevant papers, associated projects, etc. This information could then be viewed on any of these pivots.

Alongside group level summaries, there was also a desire for 'personal' summaries. For example, P4 stated that he would like to have a daily summary of all the information he had used during a day:

"*What would be nice is something which helps bring all this together at the end of the day because we use lots of different machines in a day and lots of different programs and what would be nice is to have for One Note page which automatically collects things you've done*."

DISCUSSION

We compare the observations of the ELNs created by NCR researchers with the design recommendations by Klokmose & Zander [3] that were based on the input from physicists. Table A, in the Appendix, provides details of the requirements that were relevant and fulfilled by the ELNs implemented at the NRC. In summary, most of the requirements related to interlinking, structuring, and aggregation of data, accessing and sharing of content, browsing, and history search were achieved through the adopted ELN. At the same time, the requirements for automation and control of instruments were not so prominent in the NRC lab. As noted in the previous sections, the NRC researchers dealt with very sensitive equipment that required careful manual calibration. Thus, automated calibration and control of parameters was not considered. However, the automation of data capture and generation of graphs were archived through the adoption of Igor software. Furthermore, the lack of the privacy concerns is an artifact of the NRC primary mission as a teaching research organization where knowledge sharing and collaboration is the key.

With regards to the observed practices, we compare them with the usage patterns of the Prism ELN which was designed and deployed by Tabard et al. [15] as a technology probe to investigate support for individual and collaborative work of biologists. In Table B (Appendix) we list the main aspects of the user practices in the Prism study and show how they relate to the observed behavior of the NRC researchers. Overall, the NRC researchers and the Prism users were supported in similar activities, from sharing, summarization, to review and reflection. While the Prism study participants practiced conformant data entries in the paper form, such requirements were not imposed in the NRC case. In fact, the researchers were encouraged to set exploratory experiments as they are inspired, without being constrained by formal planning and recording of the experiments. This practice was supported by the facilities to capture experiment data and metadata automatically or copy them easily into the notebooks from other software.

Two aspects that were not mentioned in the requirements from Klokmose & Zandar [3] and Tabard et al. [15] pertain to the exploitation of semantics and pro-active exposure of data by the system, both of which were articulated by the NRC study participants. In their work Talbot et al. [16] explore integration of ELN with SAM, a layered set of components and services for managing annotations and semantic relationships between data. NRC researchers, on the other hand, repeatedly expressed the desire for tools to drill and slice through the space of topics, projects, ideas, and experiments. At the moment, the search support is restricted to OneNote's keyword and tag search and desktop search.

Furthermore, Sarini et al. [11] propose to capture work patterns and provide suggestions during scientists' work, to lead them towards the fulfillment of their goals. Researchers at NRC did not express a need for automated guidance. They acquire their skills through communication with peers and the professor. However, they expressed a desire for a system that recommends relevant documents, experiments, data, and material samples as they use the ELN and other applications to plan, analyze, and summarize experiments.

RECOMMENDATIONS

The observed workarounds and expressed needs of the participants emphasize several fundamental issues of ELNs, some of which were hinted in the requirements and recommendations of previous studies but fully revealed in the practices of NRC. The very flexibility of the digital media raises concerns and calls for design considerations to

- Achieve a level of expected fixity of the digital records and consistency in practice.
- Support transitions in the organization of the project records over time as the role and activities of scientists evolve
- Enable integration of multiple ELNs with potentially diverse organizational structures.

The last two points speak eloquently to the conflicting requirements that arise in practice. They demonstrate the need to balance the support for diversity that is essential to achieve productivity of individuals in the team and the support for uniform and integrated view of the collective work that enables collaboration. We here reflect on techniques and design principles that could be suitable to achieve the outlined objectives.

Fixity vs. flexibility of the digital records and practices. Flexibility and fluidity are the assets of the digital media. However, the practice shows that in content organization, they can cause disconnect between the individual preference and needs and the institutional policies for record keeping that aim to ensure long term value to the team. In practice, that is reflected through rather simple and apparently inconsequential features, such as the ability to rename records and add information at a later time. However, this has a knock of effect by leading to much less thought and commitment to precision during record keeping. The consequences are often not foreseeable or fully appreciated at the time. This may need to be addresses with features and practices that go beyond the simple management of the edit history. It may require the ability to "freeze" specific views of the record so that they can be replicated at any time with full authenticity. The use of templates and data entry prompts, as suggested by Schraefel et al. [12], could be useful in counteracting the lack of diligence that some users felt arose from the use of the digital medium. Most of the commercial products, in fact, offer capabilities for creating data entry templates, with compliance enforcement, and devising workflow support to streamline the practices.

Evolving and heterogeneous structures vs. unified views. The study revealed different approaches to content organization across individuals and a practice of changing the organizational structure of the captured content over time. This was illustrated through the tension between topical and chronological record keeping and through the observed use of tagging and linking among ELN pages to support access to relevant information.

Making an up-front decision about the content organization is difficult in its own right. But, the situation is far more complex. The nature of the work changes over time, e.g., from initial experiments where chronological records may be suitable, to aggregation of findings where maintaining topical records is more natural. Early stages of research may require participants to work in unstructured ways, which may benefit from more open ended 'thinking spaces'. Similar conclusions were arrived by observing the use of 'master lab notebook' in the deployment of Prism by Tabard et al. [15]. Thus, the organization needs to change over time and it is of critical importance to provide flexibility in restructuring the content.

In order to fully embrace this diversity we recommend enhancing the ELNs with technologies that can deal with unstructured and multi-structured data. The ELN environments need to incorporate automatic or semi-automatic features that are supported by sophisticated technologies from text mining, natural language processing, information retrieval, schema integration, and similar in order to extract relevant entities and relations from the ELNs content. The extracted structure can be then be used by various applications and services to provide flexible search and pivot views.

CONCLUDING REMARKS

Through a detailed observation study, we arrive at insights into practices that naturally emerged around the ELN system adopted by a scientific research laboratory. Our work provided the means of comparing commercially available and prototype systems that were subjects of previous studies and the ELNs adopted by the NRC as fully operational and essential for the scientific work. Based on the workarounds and needs of the study participants, we expanded the set of requirements and design guidelines to mediate seamlessly conflicting aspects that arise from the digital nature of ELNs: the flexibility, fluidity, and low threshold for modifying digital records and the requirements for persistence and consistency. Addressing these issues is critical. That was reflected in a number of related requirements, including the migration of the content structure and providing alternative views to support specific analysis through pivoting, filtering, and aggregation. We suggest that this apparent tension can be mediated through the use of sophisticated digital technologies to deal with unstructured data and, thus, reconcile the need for both the personalization of record keeping practices and the conformity to enable collaborative work.

While our study is limited to the observation of a single research laboratory, we expect that our findings will inform future situated research of ELNs and expand the scope of inquiry to practices that emerge across various scientific domains as a result of ELN adoption.

ACKNOWLEDGMENTS

We thank Professor Jeremy Baumberg, Director of the Nanophotonics Centre at the University of Cambridge, England, and his team of researchers for their invaluable input and generous help with conducting the study.

REFERENCES

1. Borriello, G., Consolvo, S., and Su, J., Labscape: A Smart Environment for the Cell Biology Laboratory, *Intel Research Tech. Report*, IRS-TR-02-016, (2002).
2. Iyer, R. and Kudrle, W. Implementation of an Electronic Lab Notebook to Integrate Research and Education in an Undergraduate Biotechnology Program. *Technology Interface Int. J.* 12 2, (2012).
3. Klokmose, C. N. and Zander, P.O. Rethinking Laboratory Notebooks. In *Proc. Int. Conf. on Cooperative Systems* (2010).
4. Latour, B. and Woolgar, S. *Laboratory Life: The Construction of Scientific Facts*, Princeton University Press, (1979, 1986).
5. Lynch, M. *Art and artifact in laboratory science: A study of shop work and shop talk in a research laboratory*, Routledge and Kegan Paul (1985).
6. Oleksik, G., Milic-Frayling, N., and Jones, R. Beyond data sharing: artifact ecology of a collaborative nanophotonics research centre. In *Proc. of CSCW 2012*. ACM Press (2012), 1165-1174.
7. Mackay, W. E., Pothier, G., Letondal, C., Begh, K. and Srensen, H. E. The Missing Link: Augmenting Biology Laboratory Notebooks. In *Proc. UIST 2002*. ACM Press (2002), 41-50.
8. Miles, M. B., and Huberman, M. A. Qualitative Data Analysis, Second Edition. *SAGE,* (1994).
9. Rajarao, J. and Weiss S. Using E-WorkBook Suite to implement quality control in real time: expanding the role of electronic laboratory notebooks within a bioanalysis laboratory. *Bioanalysis* 3, 13, (2011), 1513-1519.
10. Rubacha M, Rattan A. K. and Hosselet S.C. A review of electronic laboratory notebooks available in the market today. *J. Lab Automation* 16, 1, (2010) 90–98.
11. Sarini, M., Blanzieri, E., Giorgini, P. and Moser, C. From actions to suggestions: Supporting the work of biologists through laboratory notebooks. In *Proc. Int. Conf. on Design of Cooperative Systems*, (COOP2004), (2004), 131-146.
12. Schraefel, M. C., Hughes, V. G., Mills H. R., Smith, G., Payne, R. T. and Frey, J G. Breaking the book: translating the chemistry lab book into a pervasive computing lab environment. In *Proc. CHI 2004*, ACM Press (2004), 25-32.
13. SCRIP-SAFE, PatentPad. http://www.scrip-safe.com/laboratory_notebooks.htm.
14. Silverstone, R. and Haddon, L. Design and the Domestication of ICTs: Technical Change and Everyday Life. In Silverstone, R. and Mansell, R (eds), *Communication by Design. The Politics of Information and Comm. Tech.*, Oxford Univ. Press (1996).
15. Tabard, A., Eastmond, E., and Mackay, W. E. From Individual to Collaborative: The Evolution of Prism, a Hybrid Laboratory Notebook. In *Proc. CSCW 2008*, ACM Press (2008), 569-578.
16. Talbott, T., Peterson, M., Schwidder, J., Myers, J.D. Adapting the electronic laboratory notebook for the semantic era. In *Proc. Int. Symp. on Collaborative Tech. and Systems,* (2005), 136-143.
17. Yeh, R., Liao, C., Klemmer, S., Guimbretière, F., Lee, B., Kakaradov, B., Stamberger, J., Paepcke, A. ButterflyNet: a mobile capture and access system for field biology research. In *Proc. SIGCHI 2006,* ACM Press (2006), 571-580.
18. Zeng, J., Hillman, M. and Arnold, M. Impact of the implementation of a well-designed electronic laboratory notebook on bioanalytical laboratory function. *Bioanalysis* 3, 13, (2011), 1501-1511.

APPENDIX

VISION OF FUTURE ELN (Klokmose & Zander, 2010)

Aim: Laboratory notebook system to support fusion of documentation, equipment control and data analysis in the same user interface.

Broken work cycle is a consequence of using the paper-based laboratory notebook together with computer applications for data collection and analysis.

CONFIGURATION		**NRC**
Automatic logging	Notebook should be configurable to document relevant information automatically and to support (or force) the researchers to document what had to be observed manually.	YES
Automatic pre-analysis of data	Produce intermediate graphs from data directly in the notebook.	N/A

ORGANIZATION		**NRC**
Interlinked data	Support structures that resemble Wiki, where all data could be interlinked. Abandon the forced chronology of the bound notebooks and have something that resembles the loose-leaf system.	YES
INTERACTION		**NRC**
Access and sharing	Access to notebook by multiple people from multiple devices – including mobile devices.	YES
Natural interaction	Multi-modal interaction and support for handwriting	YES
ARCHITECTURE		**NRC**
Common programming assets	Instruments and governors—interaction logic and application logic—could easily be programmed and shared. Support for programming through dynamic scripting language or visual programming language to enable automation of simple tasks.	N/A
Automation	Collecting and processing data in the laboratory notebook could be handled through data-flow programming. Should be able to continuously work from configuring data collection and data analysis to configuring governors to translate the data into publication quality figures and graph in the documentation part of the notebook environment.	YES
Instrumental interaction	Software architecture that enables ubiquitous instrumental interaction where there is no distinction between object for documentation (text and graphics) and objects for control (widgets). E.g., to document the angle of a laser, copy the control widget into the documentation part of the laboratory notebook and potentially use these objects to re-load the settings.	N/A
Cross platform use of ENL	Bring parts of the laboratory notebook interface to mobile devices, e.g., PDA.	N/A
DATA FLOW SUPPORT		**NRC**
Summaries	System should support either manual or automatic generation of summary pages that could be loaded on an interactive whiteboard in the meeting room.	YES
Structure	Structure, categorize, and access entries in ELN in multiple ways, e.g., by activity, chronology, or particular piece of equipment.	YES
Privacy	Support personal version of entries, where notes and annotations are not intended to be read by others, but could be made public through interaction with a simple instrument.	N/A
Forms for data input	Support for generating templates comprising fields and layout for documentation, comments, etc., that could include controls relevant for a given experiment.	N/A
History	Enable browsing the editing history (without cluttering the interface). History backed up in a secure way.	YES
Access to related content & controls	Hypermedia ways of clicking content to see associated data and controls.	YES
Aggregation	Need to support integration of digital content and data in different formats. Special facilities may be needed to enable integration or imbedding.	YES

Table A: Comparison of the recommendations for ELN by the physics scientists and the properties of the OneNote ELNs in use by the NRC scientists.

PRISM (Tabard et al., 2008)	
Aim: Support evolving work practices that involve paper and electronic lab notebooks. Support biologists in integrating activity streams and sharing information with colleagues.	
NRC: All the research assets are in the digital form, except for the material samples used in the Lab. There is no use and no concern about paper artifacts.	
SHARING	
Time factor	Two types of sharing: (1) over time, to preserve traces of activity for successors and (2) among colleagues, to share acquired knowledge and knowhow. **NRC:** Sharing among colleagues is achieved through open access to ELNs and regular meetings and presentations of findings. Projects level ELNs, created and curated by the professor, support archiving and preservation of knowledge.
Control	Adaptable information sharing. A fine grain control over what is shared, with whom, and when. **NRC:** n/a
Access	Enabling Prism for the Web made it into a central notification point, and the means of reciprocal sharing of bibliographic references and calendar events with each other. **NRC:** Shared OneNote ELNs support real time and asynchronous collaboration. Email is used for communication and calendar.
ORGANIZATION	
Common vocabulary	Development of the common vocabulary to support future reflection on work. Common tags: to do, important, and done; color coded content surrounded activities within the notebook and added meta-notes to comment on and synthesize existing notes. **NRC:** Common tags and annotations in OneNote facilitate planning and distribution of work.
Access and sharing	Access to notebook by multiple people from multiple devices—including mobile devices. **NRC**: The Center practices an open data access policy. All the data is stored on networked desktop PCs, Tablet PCs, and servers.
Dynamic organization	Organize data dynamically—mark things to do, develop, copy or move depending on what is done. **NRC**: OneNote ELN provided flexibility for adding new pages as the researcher decided to investigate a new direction. Re-assembling information can be easily done through copy&paste.
OVERVIEWS AND REFLECTIONS	
Master Lab Notebooks	All the participants created reference points, or master notebook, to organize the diverse strands of their personal activity. **NRC:** Summary and overview pages were used by a number of participants. The professor created master documents for projects that involve multiple team members.
Review and reflection	Prism enables review of information from diverse sources and comment on it, providing a series of ongoing reflections over time. **NRC:** OneNote ELN enables hybrid content to be collected and viewed within pages, allowing for annotations and comments.
Redundancy	Prism users welcome redundancy as a resource of reflection is welcome. Redundant information helps to reflect creatively about previous activities. Re-appearance of the item, in a different form means it is important. **NRC**: At NRC, redundancy in observations of the same phenomena is perceived as important signal for discovery. Redundancy in repeated experiments was undesirable if done because the students were not aware of each other's work.

Table B: Comparison of the recommendations for ELN by the physics scientists and the properties of the OneNote ELNs in use by the NRC scientists.