Field Notes on Field Notes: Informing Technology Support for Biologists

Ron B. Yeh, Scott Klemmer
Stanford University HCI Group
Computer Science Department
Stanford, CA 94305-9035
{ronyeh, srk}@cs.stanford.edu

ABSTRACT
Biology fieldwork generates a wealth of qualitative and quantitative information, an unstructured "bag of data," requiring substantial labor to coordinate and distill. In the interest of furthering research, we present findings from interviews with 15 biologists and analyses of 13 notebooks. We learned that biologists rely on their notebooks as the primary record of observations, plans, measurements, and results. However, a gap exists between the field tools used to collect data and the laboratory tools used to analyze that data. Our goal is to bridge this gap, bringing the field to the laboratory and the laboratory to the field. We introduce ideas to augment notebooks with two related goals: presenting notes alongside contextual information and using the notebook itself as a query interface.

Author Keywords
Biology, augmented field/laboratory notebook, Anoto pen, contextual inquiry, interactive paper, sensor networks.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI).

INTRODUCTION

Biologists who work in the field and the laboratory face an increasingly difficult task of managing and searching through vast amounts of information. Technology exists to support data collection—digital still/video cameras, audio recorders, environmental sensors, and custom-built monitoring devices are used to collect information required for experiments. For example, sensors such as the iButton are beginning to replace analog displays for fine-grained temporal capture of light, humidity, temperature, *etc*.

While these technologies have improved the data collection process, first-hand fieldwork retains its primacy as a research method: for using these capture technologies and, just as importantly, for the production of qualitative written narratives. Once gathered, the wealth of information exists as an unstructured "bag of data," requiring substantial labor to coordinate and distill. In addition, the distilled results lack a tractable method for ascertaining the data's heritage.

Field Study

To understand biology practices and the potential for technology support, we conducted structured interviews with 15 biologists totaling 26 hours. The seven longer interviews (2-4 hour contextual inquiries per person) comprised five doctoral students, one post doctorate researcher, and one professor, all from the Department of Biological Sciences at our university. The eight shorter interviews included other biologists from our university, the California Academy of Sciences (CAS) and the Jasper Ridge Biological Preserve (JRBP). The interviewees span laboratory and field biologists, managers, and museum curators. We conducted each interview at the biologist's work place, beginning with a structured interview, and continuing with discussion of work practices and artifacts. We observed biologists using their field and lab notebooks, and also photographed and analyzed 471 pages from 13 notebooks of five biologists. To facilitate our analysis, we developed open-source software tools for a) transcribing and annotating spoken interviews and b) categorizing the content of handwritten notebooks; they are available on www.sourceforge.net.

We summarize the steps of biology research and describe our primary finding—that notebooks are the central organizing artifact in biology work, in both the lab and the field. We describe initial designs for an augmented notebook informed by this fieldwork and review related work.

PHASES OF BIOLOGY RESEARCH

Biology research frequently occurs in these phases:

Design: biologists (like most scientists) begin by reviewing relevant scientific literature, posing questions that interest them, and designing experiments to answer those questions.

Gather: biologists spend time in the field gathering specimens (*e.g.*, leaves and bees), measurements, and/or sensor data for later analysis (Fig. 1). Biologists collect physical specimens for a number of reasons. For systematics, the science of naming organisms, CAS owns 40,000 type specimens, acting as standards for millions of the other physical specimens CAS owns. Often, specimens are stored in containers with alpha-numeric identifiers. Their information is entered into databases, with the same identifier serving as the key. For some research, collection of actual specimens is not necessary, and measurements, photographs, audio, and/or video suffice. Measurements are written in notebooks, along with the necessary identifiers. Data collected via sensors are stored in computer files.

Experiment: Biologists then experiment on their data and specimens, recording results in notebooks.

Transcribe and Process: these can occur at the end of both *Gather* and *Experiment*, and may be repeated when new data is collected. Transcribing handwritten notes to computer tools is one of the most labor intensive tasks of biology research. One interviewee described an experiment in Costa Rica where he spent six hours each day processing and transcribing data from datasheets to his laptop. Another reports spending 12-hour sessions processing audio and transcribing field data into his Microsoft Access database.

Organize, Visualize, Analyze: biologists then explore the data to pose questions and find answers. They do this using statistical analysis tools to aggregate data and create charts. One interviewee used Microsoft Excel's pivot table feature, which allows him to create views of the data by dragging column headers onto different axes of a chart or table. He can easily break down his data (number of bees caught) first by location and then by trap type, or alternatively, by trap and then by collector's name. For advanced analysis and visualization, biologists export their data (from Excel or Access) to SYSTAT or JMP.

Share: finally, they share their results in papers, talks, web documents, and books. We have attended talks, examined literature, and viewed web content (e.g., CAS's AntWeb).

THE MANY FACETS OF BIOLOGY NOTEBOOKS

Our contextual inquiry results show that the notebook is the central organizing artifact for biology research in the field and in the lab. Biologists take notes to record and describe 1) data and results, 2) steps they took to get those results, and 3) context regarding the data or the experiment.

Many biologists keep separate field and lab notebooks. Field notebooks may be smaller, constructed out of rain-proof paper, and used to record observations of early phase work (e.g. *Gather*). Lab notebooks are larger, graph-ruled, and contain finished protocols and results from lab experiments. Despite this distinction, field notes are indeed brought back to the lab (for transcription), and lab notes are sometimes brought to the field (for reference).

Figure 1. A biologist counts germinated seeds in the field at JRBP. She records data in her field notebook. The inset shows plot construction for her field experiment.

Our observations echo Sellen and Harper's conclusion that while paper is often viewed as inefficient and passé, in actuality it is an effective technology: digital technologies *change* paper practices, but rarely make paper irrelevant [11]. For example, we learned that part of the success of paper field notebooks can be attributed to their resiliency to damage in rugged terrain—a sheet of paper torn in half becomes two; a tablet computer torn in half does not beget two computers. As evidence of rugged terrain, we have seen in field notes signs of dirt, minor rain damage (rain-proof coatings prevent extensive damage), and blood (one researcher cut his hand while climbing). This ruggedness of paper encourages biologists to treat their notebooks as *the* backup for data and procedures. Given corruption of processed digital media, a biologist can use his notebook to reconstruct his entire experiment. For example, one interviewee's notebook contains time stamps that tell her how to segment her digital sensor data into meaningful chunks. If she loses the processed data, she can reconstruct it by combining raw data with the information in her notes.

The Process

Grinnell, founding director of the Museum of Vertebrate Zoology (MVZ) at UC Berkeley and widely considered to be the originator of modern note-taking, emphasized frequent and careful documentation in a notebook with three sections: a journal, a species account, and a catalog [5]. *Journal entries* describe the day's work, and include name/date/title headers, the time, weather, participants, and pictorial annotations such as maps, or sketches of plants and animals. Entries are written in the field, while observations are still fresh in the biologist's mind. *Species accounts* include lists of species observed, and information learned, such as identifying voice/cues/behaviors, and number and location observed. The *catalog* includes entries regarding specimens added to the collection. Russell, a colleague of Grinnell, reminisced about Grinnell's meticulous style, explaining that "we never traveled one inch before he had that notebook open…. 'Put it all down!' he said. 'You may not think it is important, but somebody may'" [10].

However, our inquiry results show that the Grinnell emphasis on narrative is an ideal rarely achieved in practice. Our interviewees concentrate on quantitative data, and recording qualitative accounts takes time from the more "important work." For example, three interviewees print form-like data sheets to fill out, with one biologist's form organized as an abstract map of the plots in her experiment.

Notebooks act as an ideation surface: one biologist wrote his design decisions into his field notebook while evaluating sites for his experiment. Ideation seems to happen less frequently in lab notebooks, where results (data, printed graphs, and statistics) are more prevalent.

Composing notes is a means to learning: A biologist studying Starthistle seedlings diagrammed the wiring of her equipment as a way to learn the wiring for future trips to the field. Another biologist would copy and annotate her advisor's lab protocols to learn the procedures better. A third biologist used qualitative note-taking early in the experiment as a means to learning. He composed prose descriptions and sketched diagrams in his first year in the field, to learn the differences between difficult-to-distinguish species of bees. Now, in his third year (with much more expertise), he works in dense, quantitative data sheets, using names and numbers to identify species and encode information (*e.g.*, the collection site).

Notebooks are the definitive record: the notebook is the definitive reference of procedures, design decisions, measurements, and results. One biologist recorded a census of germinating seedlings in her field notebook. She also showed us calculations, saying that she could reconstruct corrupt digital data by reviewing tables and procedures from her notes. Another interviewee recorded transfers of bacteria culture, and said that she is able to backtrack through her experiment using her notebook, "debugging" multiple generations of *E. coli* culture to track any single outlier's data lineage.

Structure and Content

Using our analysis tool, we calculated the fraction of notebook area consumed by five different categories of content (the sixth category represents empty space):

Tables Unprocessed or aggregated data.

Prose Free-form text describing procedure, insights, and other qualitative observations.

Lists Lists of procedures or steps taken, or todos.

Images Hand-drawn diagrams or printed images.

Graphs Printed statistics and data visualizations.

Empty Margins, padding, and other blank areas, including pages pre-allocated for future work.

We did this for each of 471 notebook pages (Fig. 2). While pages vary (no one page reflects all the types in their relative proportions) the data we present reveals where biologists spend most of their time and effort, most notably in composing tabular data.

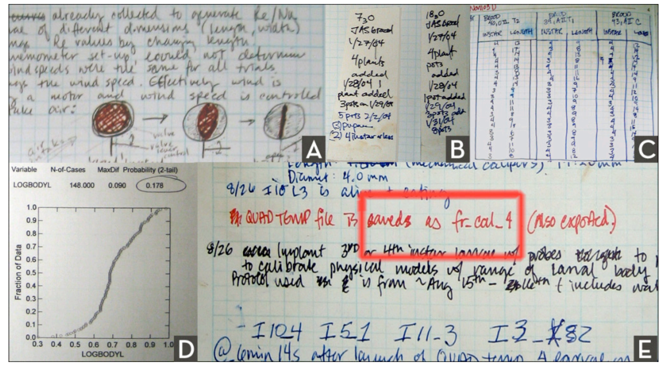
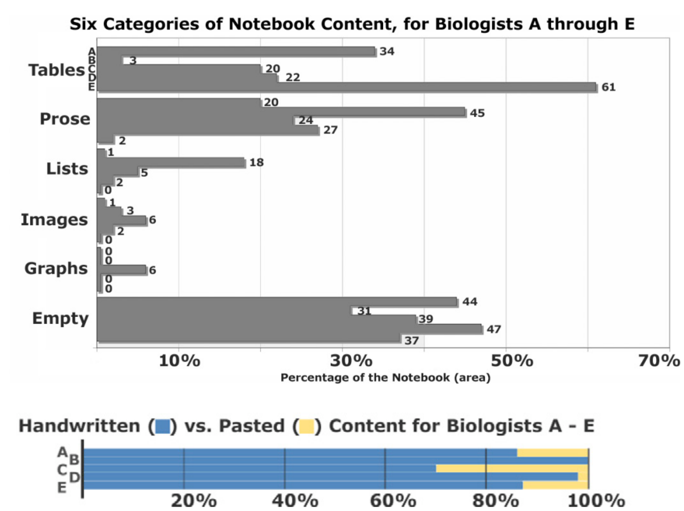


Figure 2. Content from one notebook, including a) observations and drawings, b) artifacts from the field experiment, c) handwritten tabular data, d) data visualizations, and e) logs of procedure. Note also the "link" to a data file ("saved as fr_cal_4"), which we boxed.

Some content is printed and pasted into the notebook, also reported in [8]. This happens for several reasons. For some content, printing and pasting is an act of "finalizing" it. Whereas handwritten data is unprocessed, needing to be transcribed for further analysis, printed data may contain fully "debugged" lab procedures that have been used with success. Computer graphics (e.g., data visualizations or photographs) are also printed and pasted. Sometimes, printing/pasting is used to incorporate others' protocols or publications. One interviewee pasted in pages photocopied from her advisor's notebook, annotated this page to customize the protocol to her own experiments. She also photocopies (shrinking while maintaining readability) and pastes publications she needs to reread and reference.



Graph 1. Breakdown of content in biologists' notebooks. Top graph depicts the percentage area of a notebook that different types of content consume for different biologists (A to E). Mean values are displayed. Individual pages vary: a biologist may put 85% tables on one page, and then 80% text on the next, to describe the previous page's data. The bottom graph depicts the percentage breakdown between handwritten notes and notes that were pasted in.

TECHNOLOGY IMPLICATIONS

Our findings that the notebook is the primary organizing artifact and that processing tabular data is labor intensive suggests a number of opportunities for technology support. Paper notebooks and electronic tablet computers each have important merits [11]. To explore the relative benefits of each technology for field biologists, we are developing the Data Pac-man notebook as both an electronic tablet notebook and an augmented paper notebook using the Anoto system, which comprises paper printed with a dot pattern and an inking pen augmented with ink capture and storage. We are also developing a sensor network infrastructure that will integrate with the Data Pac-man notebook. The Data Pac-man will provide dynamic document generation; facilitate the transfer of field notes into analysis tools; enable context-aware, cross-modal queries; and support data lineage. Dynamic notes will use contextual information such as location, time, and environmental information to augment written notes with a sidebar of photographs, audio/video recordings, maps, and sensor readings. Cross-modal queries enable scientists to ask questions such as "what have I written in the past about a sudden drop in temperature?" As written notes are transferred to electronic analysis tools, we plan to make readily available the lineage back to the original recording. Among other benefits, this helps double-check results and avoid errors.

RELATED WORK

Khan interviewed professionals about their note-taking practices, discovering that problems arose from missed facts, illegible names, and degraded recall over time [6]. Our study complements that work by distilling insights present in biology work practices. Khan's research informed the design of Filochat, one of a class of capture and access systems for personal and shared notes [12, 13].

Our interest in merging paper and electronic books is inspired by earlier designs in this area [2-4, 7]. Most notably, Mackay *et al.*'s a-book is a notebook for laboratory biologists integrating a traditional paper notebook with a PDA [8]. The PDA is used to create a table of contents, links between pages, and links to external data sources. Whereas Mackay concentrated on the use of the notebook in the laboratory, we concentrate on how a biologist might access context sensitive sensor data while in the field. However, our fieldwork results do corroborate many of her findings—especially that notebooks are "multimedia" documents (with various types of content), and that the notebook is the central organizing tool for supporting the design and execution of biologists' experiments. Our ideas are also informed by work on technology support for biologists. Labscape is a ubiquitous computing environment allowing cell biologists to express and track experiments via a graph representing experimental phases [1]. Low-powered, wireless sensor networks have been deployed in the environment to monitor locations that are difficult to manage (*e.g.*, they are sensitive to human disturbance) [9].

FUTURE WORK

We have summarized the work practices of biologists, with the intention of informing future designs in this domain. In our research, we classified notebook content into six types, supporting our key design points. In the future, we plan to continue conducting contextual inquiries to refine our designs of the Data Pac-man. One of us has been accepted as an apprentice biologist, and will assist biologists in conducting experiments at JRBP. We are currently developing a prototype using sensor networks, Anoto pens, and a pipeline to integrate the streams of data. From there, we hope to identify and solve the tougher HCI research questions that will help to further biology research.

ACKNOWLEDGMENTS

We thank the biologists who participated in our studies, and Andreas Paepcke, Terry Winograd, and Hector Garcia-Molina for providing support and insightful comments.

REFERENCES

1. Arnstein, L., Hung, C.-Y., Franza, R., Zhou, Q.H., Borriello, G., Consolvo, S. and Su, J., Labscape: A Smart Environment for the Cell Biology Laboratory. IEEE Pervasive Computing, (2002), IEEE Educational Activities Department, 13-21.
2. Cohen, P.R. and McGee, D. Tangible multimodal interfaces for safety-critical applications. Commun. ACM, 47 (1). 41-46.
3. Guimbretière, F., Paper augmented digital documents. UIST 2003, (Vancouver, Canada), ACM Press, 51-60.
4. Heiner, J.M., Hudson, S.E. and Tanaka, K. Linking and Messaging from Real Paper in the Paper PDA. CHI Letters, 1 (1). 179–186.
5. Herman, S.G. The Naturalist's Field Journal: A Manual of Instruction Based on a System Established by Joseph Grinnell. Harrell Books, 1986.
6. Khan, F. A survey of note-taking practices HP Labs Technical Reports, 1994.
7. Klemmer, S.R., Graham, J., Wolff, G.J. and Landay, J.A. Books with Voices: Paper Transcripts as a Tangible Interface to Oral Histories. CHI Letters, 5 (1). 89–96.
8. Mackay, W.E., Pothier, G., Letondal, C., Bøegh, K. and Sørensen, H.E. The Missing Link: Augmenting Biology Laboratory Notebooks. CHI Letters, 4 (2). 41-50.
9. Mainwaring, A., Culler, D., Polastre, J., Szewczyk, R. and Anderson, J., Wireless sensor networks for habitat monitoring. Proc. of 1st ACM international workshop on Wireless sensor networks and applications, (2002), ACM Press, 88-97.
10. MVZ. http://mvz.berkeley.edu/Grinnell_Method.html
11. Sellen, A.J. and Harper, R.H.R. The Myth of the Paperless Office. MIT Press, 2003.
12. Stifelman, L., Arons, B. and Schmandt, C. The Audio Notebook: Paper and Pen Interaction with Structured Speech. CHI Letters, 3 (1). 182–189.
13. Whittaker, S., Hyland, P. and Wiley, M., FILOCHAT: handwritten notes provide access to recorded conversations. in CHI '94, ACM Press, 271-277.