Order From Chaos: The Poetics and Pragmatics of Scientific Recordkeeping

Kalpana Shankar
School of Informatics, Indiana University, 901 E. 10th Street, Bloomington, IN 47408.
E-mail: shankark@indiana.edu

Although the production of laboratory and field records is fundamental to the conduct of contemporary science, there has been little research into this topic in information studies. This article reports on a study in which, using ethnographic methods, the author studied recordkeeping as it is practiced in a basic research science laboratory. The process by which the record is created to reflect both personal need and professional norms is framed as a series of acts of selection, synthesis, and standardization. The article concludes with reflections on the role of deep understanding of scientific recordkeeping for other disciplines and the design of digital laboratory technologies.

Introduction

Writing and the production of texts are essential to the conduct of science—not just for the function of disseminating information but also for making "order from chaos" (Latour & Woolgar, 1986). Researchers in the information sciences have paid a great deal of attention to published writing in science, particularly the acts of scholarly writing and publication as they contribute to knowledge production and dissemination. Studies in the structure, function, and various other roles that documents, or laboratory "inscriptions," play in the daily practice of science have received less attention in information studies.

The act of daily recordkeeping and its relationship to the research enterprise is complex and worthy of analysis. First, the record plays multiple roles as a legally binding proof of discovery, historical artifact, intellectual property, a pedagogical tool, epistemological framework by which claims are made, organizational memory, and a mechanism of communication and collaboration. Historians, regulators, archivists, sociologists, and scientists themselves tend to agree that recordkeeping is essential to the conduct of science and plays a vital role in knowledge creation and dissemination. Information scientists may see records as constitutive of the sociotechnical environment and context and have suggested that records form an essential informational infrastructure (Botticelli, 2000). In the laboratory, these documentary products of daily activity represent tensions between standardization and flexibility, the collaborative nature of science and the practical and personal needs of the individual scientist, and individual learning and professional socialization. Recordkeeping is more than organizational memory; in science, it has profound implications for the production of knowledge and the development of professional identity.

Through extended ethnographic study, this article explores the processes by which the record is created and maintained in an academic research laboratory. It demonstrates how recordkeeping can be read as more than routine professionalism and less than infrastructure by casting it as a creative and personalized set of practices and acts to establish accountability and memory. By tracing three constituent acts of recording—information synthesis, formalization, and toolmaking—this article argues that these usually overlooked steps are components not only of creating records but also in personal and collaborative information management. The article then suggests avenues of exploration for "scaling up" to collaborative groups and implications for electronic laboratory notebooks and similar digital tools.

Methodology

This study is the result of 8 months of ethnographic fieldwork in an academic animal neuroscience laboratory. Ethnography as a methodology is particularly well suited to understanding phenomena that are emergent and undertheorized, and is of increasing use in contextualizing a wide variety of information practices. A fuller explanation of the methodology and the ethnography of recordkeeping has been described elsewhere (Shankar, 2004).

The research laboratory consisted of Philip, the principal investigator, two junior postdoctoral researchers, Michael and Susan, two graduate students, Yolanda and Veronica, and several undergraduates. The researchers tended to work on

Received April 27, 2006; revised November 21, 2006; November 28, 2006; accepted November 29, 2006

 ● Published online 12 June 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.20625

individual projects with minimal collaboration. Visitors, temporary researchers, and friends of the research group also were occasionally present. Quotes and illustrations are drawn from the working notes and daily notebooks of current researchers and previous researchers, fieldnotes, and open-ended interviews with laboratory personnel.

Review of the Literature

It is not surprising that researchers outside of the laboratory have focused their attention on the publication process and published artifacts. For them, the published article is the most physically, if not intellectually, accessible documentary evidence of scientific inquiry. For the researcher who is interested in such documentation, ownership of data, copyright, and other intellectual-property problems in accessing published documents are minimal. The scientific community itself may encourage the study of these finalized documents since the published reports portray the scientist as she or he wishes to be seen. As historian of science Frederic L. Holmes (1987) wrote, "[Publications] are stereotyped according to canons of form dictated by the authority structure of scientific disciplines" (p. 220). The complex series of negotiations between the scientist and the objects of study, the instruments used in discovery and analysis, and ultimately, multiple audiences tends to become obscured through the final publication process. Publications, therefore, are generally not considered an archival record because they are not themselves evidence of the scientific research process; rather, they serve as a distillation and stylized version of the processes by which the primary documents are created.

For the researcher who wishes to understand scientific discovery, there are several problems with making too many assumptions about the practice of science by looking at the published works on their own. First, those published works tend not to give us a picture of the ongoing practice of science since publishing is at best a semiregular activity. Second, the published document represents science in its most idealized form: as linear and straightforward, stripped of its particularity and idiosyncracies (Myers, 1990).

Beneath this edifice of publications, however, lies a bedrock of documents that yields a closer look at the daily practice of science: data files, laboratory notebooks, and printouts from instrumentation, and other primary sources. These kinds of documents, while they contribute to historical and contemporary studies of scientific practice, have tended to be understudied as a form of scientific documentation. There are some understandable reasons for this lapse, particularly with respect to laboratory notebooks as the working record. Obtaining the workbooks of research projects in progress is difficult since laboratory notebooks in many laboratory settings are not the private property of the laboratory or the scientist but that of the university or research institute. Many of the records never end up in an archive or other repository; even if they were to do so, looking at inactive records would not be useful for understanding the context of their creation. Laboratory notes often require a high level of understanding of the topic of study to make sense of them—the premises of the research, the conditions in the laboratory, and perhaps access to the researchers. And most importantly, the lab notebook can contain an enormous amount of uncontextualized numerical data that is impossible to understand without having access to publications that have resulted directly from that series of experiments. The sheer volume of detail that emerges could easily obscure the larger picture. As a result, almost all of the research that has been conducted on daily recordkeeping has been historical in nature (Holmes, 1990). Contemporary studies of scientific documentation and writing as process have emphasized scientific documents as information and ways of recruiting others to one's ways of thinking, not as evidence of activity (Latour & Woolgar, 1986; Myers, 1990). What is manifest, though, is that the existence of a laboratory notebook seems to be taken for granted in any discussion of scientific recordkeeping.

Given that the "record" is not the publication, and that finding it and understanding it may be problematic, it still becomes necessary to generate a working definition of that record. Scientific writing about records and recordkeeping has tended to focus on the pragmatic aspects of standards and the larger research community: how records should be kept and the ethical implications of recordkeeping for the communication of results, information technologies, and ramifications for the accountability of science; however, when one delves into literature on the creation, use, and management of records, one finds that the area of study is multidisciplinary and fragmented. There has been remarkably little said about recordkeeping, except when documents are called into question. It is interesting to speculate on why recordkeeping is given such short shrift when the ethical, evidential, and epistemological importance of the lab notebook is viewed as paramount by the scientific community.

Apart from the function of documenting the "who, what, where, when, why, and how" of research science, there is little discussion in the literature of the scientific record in context. Macrina (1995) implied that laboratory notebooks serve a normative function since the practice of good recordkeeping fosters the scientific norms of accuracy, replication, and reliability. In recent years, laboratory notebooks have received increased scrutiny as scientific documents since several highly publicized lapses in scientific ethics have hinged on the interpretation of data recorded in a notebook.

Beyond its importance as organizational memory, the daily record in both academic and corporate laboratories is an essential component of knowledge management. Data-ownership and intellectual-property worries are becoming prevalent in a world where the end products of science are increasingly seen as a commodity. The artifacts of science, including the records, have become closely held assets. Some in the scientific community have argued that this development is antithetical to the philosophy of the laboratory record, which some maintain should be a means of open communication (Macrina, 1995). Despite this spirit

DOI: 10.1002/asi

of commodification, the issues of property, ownership, evidentiary value, and control over data are still not clearly articulated by the entirety of the scientific community, and in many areas of science still do not pose controversy.

These issues point to conflict between the roles of the laboratory notebook as information commodity and as "recorded truth." Perhaps the problem is as Macrina (1995) opined: "[W]hen even fundamental expectations about scientific recordkeeping are not communicated, there is little hope for appreciation of the related deeper issues" (p. 46). One of those deeper issues is that the act of recording and writing is not just "writing things down." The process by which one learns to keep records, or create a record, itself can be conceptualized as an act of discovery.

Understanding science, an enterprise that is at the heart and soul of contemporary life in the West, means understanding how scientists inscribe their work and make order of it. Keeping records is the way scientists make order from chaos, and make the invisible visible (Latour & Woolgar, 1986). In his detailed work on scientific memory, Bowker (2005) expanded upon this theme, and challenged that pragmatic notion that the laboratory or bench scientist keeps a log of his or her work "merely" to remember details, to replicate experiments, to reflect on the results, and ultimately, to keep a repository of information that will result in formal publications. Instead, he argued, the primary laboratory records are the substrates that constitute our very notions of time and space.

In short, the normative explanation of recordkeeping proffered by the scientific community can be deeply unsatisfactory. So one asks other questions of it—what the title of this article calls the "poetics and pragmatics" of recordkeeping. Of the latter, we can ask: What constitutes the daily scientific record? Is it just the laboratory notebook, or the data files and documents that the laboratory notebook contextualizes? How are professional standards articulated in local context? For the former, one wants to know: What do those documents mean to the personal and professional identities of scientists? How do they figure in their acts of sensemaking in the world?

Recordkeeping standards articulated by professional societies, textbooks, and laboratory course instructors must eventually become part of the toolbox of the scientist at the lab bench: How that process happens, and the role that those records ultimately play in the creation of knowledge, becomes naturalized and not questioned as part of the socialization of the scientist. These issues are worth exploring because recordkeeping and records are like other practices and technologies in that they are not value-free tools to create order in the world (Winner, 1985). They, and the standards and assumptions imbricated in them, clearly have political, moral, ethical, and social implications. They are intentional uses of language and meaning.

However, standards and formally articulated best practices, how-to manuals, and the like do not tell the complete story. They call for dated complete volumes of records that can be verified, read easily, and called upon to produce a body of evidence in support of research results and aid in the discovery of new knowledge, but do not discuss the personal, organizational, and other aspects of the documentation process. Such manuals, like the formal scientific publications discussed earlier, are themselves at best an idealized manifestation of product, not process. These manuals suggest that the daily records are important for scientific conduct, but are often perceived as "less" important than the final publications and posters that result from them. Paradoxically, these records which are of grave import to the research enterprise become "nothing:" They are created day after day, "written up" in conference papers and posters, and ultimately shelved in countless laboratories, never to be consulted again. Granted, some of the notebooks of famous scientists end up in archives, and in still other cases, the recordkeeping activity is called into question for reasons of fabrication or falsification, perhaps most famously in what has come to be termed the "David Baltimore case." When Margot O'Toole, a postdoctoral researcher at Tufts University, was unable to replicate the published results of her supervisor, Dr. Tereza Imanishi-Kari, her subsequent actions resulted in a decade-long investigation into research and laboratories of, among others, Imanishi-Kari and her Nobel Laureate collaborator, David Baltimore. The investigators called in the U.S. Secret Service to the Massachusetts Institute of Technology and Tufts University to investigate these alleged frauds; the laboratory notebooks formed the basis of key forensic evidence in the investigation (Kevles, 1998).

Standards and "how-to" manuals are not enough if we wish to understand the activity of information creation and the ways in which human beings make sense of their information worlds vis-à-vis externally articulated, often vague, "standards." Records ultimately give us insight into science as practice which, like all human activity, is grounded, embodied, and active.

While science is understood to be a communal activity, belonging to such a community does not imply that kind of homogeneity. Traweek (1988) defined a community as a "group of people who have a shared past, hope to have a shared future, have some means of acquiring new members, and have some means of recognizing and maintaining differences between themselves and other communities" (p. 6). How individuals develop a professional voice and identity, and negotiate them amidst the expectations of their research discipline and broader profession, is reified in the artifacts of their study, the practices they engage in, and for the purposes of this work, the texts they create. Scientists form multiple communities of practice (Lave & Wenger, 1991)—in the laboratories they work in, with geographically dispersed colleagues who share their research interests, and with collaborators on projects that form and break apart. Documentary practices may be renegotiated within each community of practice. This idea that people participate in and belong to multiple communities of practice is a useful one because it suggests that identities are shaped by numerous social contexts. Scientists, for example, may arguably belong to one community of practice

DOI: 10.1002/asi

consisting of those in their own laboratories, but also may be collaborating with several other groups, be part of a professional society, and still retain ties to the lab in which she or he received prior training. Asking how professional and personal identities form and under what circumstances is to ask how people engage with their communities of practice, and the multiple roles of information in creating those identities.

In modern organizational settings of all kinds, not just the scientific laboratory, records and recordkeeping are a kind of information infrastructure—embedded in other information practices. The localized implementation of an information infrastructure and the ways in which it becomes normalized, and the relations between those implementations, always must be embedded in a larger framework of universal standards and norms (Star & Ruhleder, 1996). For example, a working scientist must develop his or her daily recordkeeping practices in ways so that the resulting documents are convenient, useful, and organized for his or her work. This process is learned as part of the membership in a scientific community of practice, to some extent by trial and error. At the same time, that scientist and his or her records must adhere to external norms of institutional accountability as determined by professional standards, funding agencies, research institutes, universities, federal and state law, and so on. Epistemological and methodological assumptions and considerations, such as veridicality, crucial to the sensemaking process that these acts of record represent, are encoded as well.

The following case study illustrates this negotiation by looking at several constituent processes: how the scientist conceptualizes the record and creates it, the processes of learning and practice by which the record as evidence of professional and knowledge production activity is created.

Setting the Scene

A young electrical engineer looking for a research project in communications chanced upon the work of a biologist at Cornell who was working on North American bullfrogs. This biologist electronically synthesized the mating call of a bullfrog and played it to a captive male, who responded instantly. But when he played croaks from other species, the frog did not respond at all. How did a frog living in a rain forest, which supports any number of signaling creatures, distinguish the signals from individuals of its own species from the noise of others? It was as much a problem of biology as it was of engineering.

The engineer, inspired by this problem, went on to obtain an interdisciplinary doctorate in neurobiology and behavior and to build a successful academic scientific career studying the communication of frogs as well as other animals. Perhaps because of his own multidisciplinary background and approach to his subject, as well as the subject itself, his research lab now draws students, postdoctoral fellows, and collaborators from a wide array of disciplinary perspectives and backgrounds. As a prospective postdoctoral fellow told me, "When I was looking for a postdoc lab, my advisor said there were only three frog labs worth working in—and Philip's was one of them."

Philip's status as a popular researcher with whom to work is evident: There are currently two postdoctoral researchers, a graduate student, and several undergraduates conducting ongoing research with him. Texts and documents of one sort or another are everywhere, constituting evidence of work and play. The whiteboard on the wall contains several jottings: a reminder for someone in the lab to feed a batch of frogs because their usual custodian is going on vacation, a diagram or equation that someone else was puzzling through, and a pithy quote. For several weeks, the board said "Oreos are the biscuits of Satan" (attributed to Michael, a British postdoctoral researcher, who did not like Oreos). The winter holiday season brought "God rest ye, merry scientists." The two refrigerators in the room (one for human food, the other for experimental materials) are covered with frog cartoons and postcards from all over the world. The bulletin board just outside the lab door is an assemblage of current events: sound- and frog-related newspaper articles, calls for papers, and fellowship announcements. The cabinets contain folders of publications, instrumentation manuals, and similar papers in labeled folders. Across the room, binder after binder, blue and hastily labeled, contain the lab notes of many of the graduate students and other researchers who have long since left the laboratory. And the central table in the laboratory is covered with black-and-white composition books, a cardboard model of the frog middle ear, a joke printed from someone's e-mail, yesterday's crossword puzzle from the university newspaper, and conference posters and papers in progress.

Members of this lab are not particularly unusual in the ways in which they create, manage, and use documents. As the aforementioned depiction suggests, some documents are transient and public, meant for immediate use or a quick communication. Others are reports and files generated by external entities such as publications, instrument warranties and instruction manuals. Some, such as the frog anesthesia log locked up in the drug cabinet, the animal experimentation logs, or the grant and personnel files in Philip's office, are required by policy and law. Still other documents such as the group crossword puzzle, the shared joke, and the postcards demonstrate the members' mutual engagement in a community of practice (Wenger, 1999). Some of these texts are called upon to perform multiple activities.

Of course, these documents do not carry the same epistemic weight. The scratch pads, models, graphs, and laboratory notebooks and fieldnotes eventually become the substrate upon which other texts and representations of scientific knowledge are constructed. But how does this happen? The next sections explore how these records embed and embody memory and knowledge, the processes by which this happens, and the many tensions and ambiguities that are generated by these processes.

DOI: 10.1002/asi

Nature of the Record

One way to illustrate these tensions is to explore the multiple and ambiguous usage of the words "data" and "record," and how the scientists creates the latter from the former. The concept of a "scientific record" seems at first glance to imply that the scientist uses some set of rules to act upon "raw"[1] data, disparate pieces of information are pulled together, and layers of narrative are added to create a reliable document. Each succeeding layer of translation and transformation removes the scientist and the record further and further from the stream of numbers that constitutes his or her "raw" data, but never fully expunges the underlying traces. Thus, pulling together a computer printout, writing text on it, and the numerical data from an instrument becomes another mechanism of managing the trajectory of the record. Amann and Knorr-Cetina (1988) called this act the creation of a montage, by which data become evidence that can be directly deployed in making one's scientific case. The creation of this montage foregrounds the important data elements, brackets them as one wishes to see them to make one's claims, and endeavors to put any "noise" in the system into the background.

Integration through synthesis becomes a sanitizing mechanism by which "inscription overload" is managed. The unwieldy mass of data is manipulated and displayed through graphs and other condensed representations of work (Bowers, 1992); however, not all traces of work are integrated into the built record. For the scientist, the acts of adding layers of meaning, combining and subtracting, construct a reliable record to which the scientist can return because the he or she can make judgment calls as to what is transferable and manipulable and what is not. In short, the acts of synthesis and selection make such documents more of a record than would the simple annotation of reams of numerical data. Susan, a young postdoctoral researcher, commented on this when I asked her if it would not have been easier to document her work on the computer, given that she had to spend so much time transcribing her handwritten notes:

> Yeah, so why do I want to do this on paper? Mostly because it's a combination … first of all, I created it by hand, so that's why I'm able to combine different kinds of information—it's all in the same format.

Michael, another postdoctoral researcher, made this point about selection and synthesis even more directly. In his words:

> The yellow books are sort of scratch books. Where I write, um, I do equations in there sometimes before I'm sure that I've got them right. And I write … during the experiment I scrawl down where it's all going. But the end result, especially during an experiment is extremely messy and difficult to read and if I went back to it myself many days later I'd probably have difficulty interpreting it. So I very shortly write it up into my main laboratory books, um … because it's much clearer after that and I can put everything in … in order, I can insert the pictures and the graphs and things like that, that I produce after the experiment. The end result is that the main laboratory books have got all the information which is correct in it.

So which are data and which is record? From these quotes, it would appear that the record is a "value-added" document, value added by the scientist's post hoc processes of information management. Michael adds:

> The experiments usually take a full day so at the end of the day I've got it all stored in my computer file and I've got scrawled notes in my yellow pad as to what I did and then the next morning my job the next morning is to write all of that down into my own notebook and process the data and print up the graphs so I have the full experimental record in my notebook.

Michael's comments, and the kinds of documents they refer to, suggest that the synthesis and annotation of research notes are natural steps along a pathway from "raw" data to "formal publication." This is not to suggest that any potential stage of the process results in any notion of a "true record," but instead that *all* of these steps increase the sense of reliability, authenticity, and other qualities that Michael attributes to his laboratory notebooks.

In the first instance, what I examined were examples from two collectivities of records, generated by two different people for the same purpose (capturing the results of their research), with some similar elements (date, order, and regular form for data input), and yet the same documents can be styled "data" or "record" depending upon who is asked. In the quote mentioned earlier, Michael used the word "record" to refer to both his daily log of activity and the resultant publication. He and Susan "record" the processes of transforming data to logbook and logbook to publication. To transform "raw" data into a reliable trace of scientific activity requires tidying up the "untidy;" in other words, transforming the liminal state of "scratchness" of the notes into an easily categorized genre of document, one that can be easily recognized as a lab notebook or peer-reviewed paper by form and usage. The documents become a kind of palimpsest, where the previous traces of recording or writing are not expunged but are instead added onto, condensed, annotated, and otherwise "managed." This process of personal documentation management requires that the scientist apply layers of meaning and organization to ensure that his or her records capture his or her intent, action, and result. Botticelli (2000) noted that organizational records undergo transformations that make them more and more a part of the environment in which they are created. Each subsequent translation—from data to research notes to publication—increases, for the records creator, the value of the resulting document because each step "is linked to the set of shared, tacit meanings that bind together communities of practice" (Botticelli, 2000, p. 175).

[1] I use the term "raw" cautiously; most scholars of science studies would agree that there is no such thing since the act of selection is always present. The scientist's laboratory records should reflect these acts at all stages.

DOI: 10.1002/asi

In short, a nuanced understanding of the scientific record would suggest that we think beyond the daily accretion of notes and notations and look more closely at the management of these notes as the true process of "recording." Through choice of recording medium, selection based on context, integration of relevant documents and written traces into a coherent whole that tells a memorable story (at least to the records creator), annotating the collectivity, and placing the whole in the larger context of the literature and previous experimentation, the scientist creates meaning and embeds himself or herself in the final result. The result is a record that is reliable, ethical, replicable, and personally valuable.

In this light, it is not surprising that many scientists consider the formal publication as the "true" record of science. It can be argued that the final publication is as fully embedded in webs of meaning and community (in style, content, reference to other work, and purpose) as any document can be. Michael noted:

> In a way of course, the ultimate form of recordkeeping in science is the published paper at the end. I think for many people, that's it, that's the record. And the intermediate stuff … is not so important. It's like taking these eighteen notebooks and boiling them down into one fully referenced, several page article.

If making data into record is an act of personal information synthesis (and a highly idiosyncratic one), creating the appropriate tools and environment to make this process "fit" the larger structures and strictures of science also is part of the act of records creation through which the creator learns and creates. While the previously mentioned illustrations suggest at least some ways in which scientists negotiate the development of the record as *aide-mémoire* and as the infrastructural glue of the laboratory, it only suggests one suite of standardized mechanisms by which scientists make records of data, where records are enmeshed in larger meanings that dovetail with expectations of the broader scientific community. Standardization is one way of increasing replicability, and thus is essential to this process.

The creation of standards is an act of formalization. Formalization is another means by which individuals personalize recordkeeping while still keeping to the broader mandates of their research disciplines. My reading of formalization as a way of increasing the reliability of information is a somewhat unorthodox projection of formalization and formalism, as the concept has generally been reserved for much larger scale designed and implemented systems (Bowers, 1992). A working definition of a formal tool, Bowers wrote, would encompass the following: circumscribed input rules (delineating "legal" data entry from "illegal" data entry) and rules that specify workflow. Nevertheless, these micro-examples of the formalization of working tools express similar tensions. To some extent, we can read the researchers' attempts to routinize and make more convenient some of the repetitive aspects of their records creation process as a move toward formalization and increasing the standardization of the resulting document. Berg and Bowker (1997) referred to such forms as "formal tools." Computer systems that are configured to enforce a particular sequence of actions or expert systems generally fall into this category. However, scientists can choose to create other systems that standardize data, and make it more recordlike.

Susan showed me the notes she had generated during the course of her current research project. Most of Susan's documents are similar in layout and structure. Her binder is divided by tabs with dates delimiting each section, and one section at the end that reads, "Earlier experiments." There also is a tab marked "Protocols." All of the documents in this binder are typed. The date is entered in the upper left hand corner of the document. The next line is labeled "Frog," and is followed by the frog number and its species, weight in grams, and sex. The next line is labeled "Experiment" and followed by a short description of the purpose of the experiment. Below that is a section Susan called "Results." She writes a short textual description of the results of the procedures that she is testing, the results of the surgery, and the disposition of the frogs. A separate section for "Anesthesia" follows and contains information relevant to the anesthetizing procedure: the amount and kind of anesthetic administered, the method of administration ("IM" indicates "intramuscularly"), the time, and observations of the frog under anesthetic. Another section is marked "Surgery," which contains a paragraph describing her technique, problems she encountered, her methods for resolving them, and her feelings with respect to her execution of the technique. The next section, marked "Rig set up," indicates the settings with which she worked and some short commentary. The observations she made are laid out, from the first electrode to the last. The observations for each electrode include the electrode's resistance in mega ohms (MΩ), numerical results, the ultimate fate of the electrodes (and her fears for them), and sometimes, textual commentary.

On another day, I look over the blue, nondescript binders which contain some of the notes and documents of James, a former postdoctoral researcher in the laboratory. The binders are labeled by animal number (e.g., the designation TW1–49 would indicate that the binder holds data related to frogs TW1 through TW49). The preponderance of these files follows a similar pattern of order and layout. The first sheet consists of a form created by hand, with blanks to fill in for the number of the animal, the date of the experiment, the species, the weight of the organism, the kind of anesthetic applied and particulars related to the anesthetizing process, and a series of "fields" for entering the results of tests done to the frog. The last field is a comments field. The second page is an annotated printout from an instrument. The document is extensively annotated by hand, with the frog number and date written on the left margin. The numbers appear to be some variables important to the researcher's results.

Both recordkeepers have employed some common elements of structure and order: Both creators used dates to order their documents and store their documents in temporal

DOI: 10.1002/asi

order. In both cases, the person documenting the activity has created a standardized form for data entry that forces them to enter the record and create it in a particular flow of activity. "Demographic" information about the frog is recorded first, followed by a closer delineation of the procedures performed upon that frog.

Similarities end there, however. The results of Susan's experiments continue to be organized linearly, with appropriate tags or markers to indicate the kinds of data that the numbers represent; however, James does not mark his data with any kind of contextual clues. Nevertheless, they both employ formalization to manage their work. The implementation of formalism and its modifications are evident in the practices of both Susan and James, as they have used free text annotations to supplement, or in some cases, work around the formal tools that they have set up for themselves to fulfill earlier goals that demanded standardization.

The scientist, is not, however, working around the system so much as modifying the system itself since she or he is the one responsible for creating and using it. If the formal system created does not accurately capture all that she or he wants captured, it is easy enough to improvise around it. A subsequent redesign of the form (in Susan's case, a computer data-entry sheet; in James' case, a handwritten form of which he made many photocopies) is not necessary because the scientist in these cases I have described has the ultimate authority to create and transcend the limitations of his or her formal system of recording.

Discussion

The extended examples mentioned earlier illustrate two complementary requirements that the record must fulfill: Documents first must be useful and "accountable" to the records creator, but also must be accessible to the larger community of practice. Both are creative acts, not routine ones, even though the scientists may actively work to make such acts more routine. In the first case, I argue that the record evolves through the accretion of layers of meaning and personal-knowledge management, accomplished through multiple acts of selection, integration, synthesis, and annotation. While these acts make the record personally meaningful, they are not guaranteed to make the record broadly reliable. Thus, the scientist then must actively work to make the record less idiosyncratic, often accomplished through standardization of process and product. None of these processes is rote. They may need to be reinvented for each new project and environment, but the results are the same: a document that is paradoxically wholly personal and yet intrinsically professional.

This kind of authority and agency is afforded to few recordkeepers. Academic freedom, a certain amount of institutional autonomy, and science's privileged role in knowledge creation give scientists great leeway and power to keep records in ways that suit them, without needing to address greater organizational mandates. This situation is unusual in organizational recordkeeping. The content of the records of the academic scientist and the recordkeeping systems they use privilege and reinforce their autonomy and relative power, even those of the undergraduate science major conducting a study for a term project. The academic scientist is afforded great latitude of choice in recording technology, the ordering of data elements, the integration of external documents, and so on.

Even so, the work of the scientist to design a system (paper or digital) that fulfills one's recordkeeping needs both for the short term and the group's/discipline's needs beyond requires a form of active learning in recordkeeping that reinforces that which works while allowing one to move past that which does not. Learning to record is an essential part of learning to do science, learning which is manifested in ways in which James, Susan, and the other denizens of the laboratory foreground some data elements, routinize others, and integrate still others to create new patterns of information which are used to make decisions. In short, the basic form of the laboratory record can be construed as an inherited/socialized part of professional training, but making the record personally meaningful—which is necessary for successful learning and scientific practice—is an act of creativity and formalization.

On the other hand, one could argue that the forms created by James and Susan are not formal tools that they have learned to create at all—they are just conveniences and contrivances designed to make the repetitive kinds of information with which they work easier to capture. One could ask how formal a system laboratory notebooks are, when they are clearly so easy to override and tinker with. It could always be *more* formalized, of course—electronic laboratory notebooks are increasing in use in scientific research (although their use is still highly limited in academic research). Such notebooks or laboratory data-management systems constitute traditional formalisms and force the scientist into a particular workflow and adherence to certain disciplinary norms; these are encoded by those in charge of systems development and design. But this line of argumentation forces us to make dichotomies where none need to exist, and to tinker with the discourse of formal versus informal (Bowers, 1992). Instead of pursuing it, I would suggest that learning to create and use a recordkeeping system, and in the process increasing the reliability of one's technical and scientific skills in using that system, is a fundamental part of learning to work with formalisms in systems of all kinds.

The flexibility and autonomy of the academic scientist is greater than that of other professionals who create organizational records: the police officer (Meehan, 1986), the nurse (Ngin, 1994), or radiologists (Yakel, 2001). Formalization of documentary practice, even in the fairly rudimentary ways exemplified by the selections from Susan's and James' records, is considered "best practice" in the creation of scientific records because in their cases, formalization makes a more coherent, annotated document out of numerical, decontextualized data. Numbers on a piece of paper or graphical notations are meaningless if they are not given some status of the formal. Data become formalized only through annotation

DOI: 10.1002/asi

(with the notations of the date of the procedures enacted and under what experimental parameters) and by placement (stored sequentially, with related documents). The formalization makes the storage of "raw" data unnecessary. As Susan noted:

> If we keep good records, we can throw away the data. That happened once … we had thrown away the data, but had reused the programs that we used to generate the data but forgot to change the variables. But we had kept track of the calibration files. So we were able to recreate 10,000 samples.

The kinds of formalization I have described are sometimes dependent, but not always, upon an increasing reliance on information technology, which comes with its own problems in terms of flexibility and usability through the imposition of expected workflows and internal formalizations. For example, while Susan created a computerized form to serve as her laboratory notebook, the structure of the word-processing program she used to create it dictated what files she could incorporate (i.e., only those that were compatible in format) and allowed her to incorporate these files only with some awkwardness. She resorted to a combination of handwritten notes and computer files because this combination afforded her the most flexibility. Which formalizations are used depends to some extent upon the intellectual tradition, generation, and personality of the records creator. Susan pointed out that her fiancé, who was 4 years younger than she but who did his doctoral work in the same laboratory that she did, was "almost completely paperless. But his background is in computer science."

For people to interact in a community of practice, Wenger (1999) argued, there needs to be what he terms a "regime of mutual accountability" (p. 81). This kind of accountability goes beyond the idea that one needs to make one's work visible to others in some predefined fashion (Yakel, 2001). Wenger suggested that mutual accountability includes:

> What matters and what does not, what is important and why it is important, what to do and not to do, what to pay attention and to what to ignore, what to talk about and what to leave unsaid, what to justify and what to take for granted, what to display and what to withhold, when actions and artifacts are good enough and when they need improvement or refinement (p. 81).

Recordkeeping allows one to "forget" because it serves as a form of external memory, but the act of creating the record also is an act of remembering relations among documents and learning, relations that become negotiated through membership in that community of practice. Some of these relations may be reified in the form of standard operating procedures, rules of conduct, and formal policies of one sort or another. But as numerous ethnographers of science and recordkeeping have noted, those that are not explicitly articulated are often just as important to understand. In effect, this reinforces the infrastructural, taken-for-grantedness of recordkeeping because these relations become pervasive by being negotiated into practice. For example, the "mental" records of Meehan's (1986) police officers are an essential, if unarticulated, aspect of police practice because the officers take it for granted as part of their source of information.

Wenger (1999) wrote that objects, symbols, and activities become meaningful in a community of practice because their importance to that community is a negotiated, learned, and taught by the community's members. They develop a shared repertoire of these things, including, in Wenger's words, "routines, words, tools, ways of doing things, stories, gestures, symbols, genres, actions, or concepts that the community has produced or adopted in the course of its existence and which have become part of its practice" (p. 83). Recordkeeping is situated in multiple contexts: juridical, administrative, institutional, and perhaps others. Recordkeeping reflects all of these contexts (Duranti, 1998), even though it is likely that the records creators will not articulate all of these contexts during their daily work; however, it is these contexts that ultimately contribute to the ways in which a community of practice keeps records because those records account for activities understandable only within those contexts. For example, a particular form or document may be created in an office, but the use of that form and its role in business affairs is coherent only when considered as part of the particular administrative context of that office. This point may seem obvious, but it ties together a number of important themes that converge to imply that understanding the particular recordkeeping contexts of an organization is one of the prerequisites to membership in a particular community of practice.

Conclusion

This article was written to serve two functions. The first was to report on an empirical study of recordkeeping in a research laboratory and to present an exploration of the multifarious meanings, roles, and processes by which scientific records are created, used, and conceptualized. This exploration represents a return to basics, in some sense. By deliberately focusing on paper records and individual records creators, we return to a kind of "ground state" in understanding records, especially scientific ones: What can we learn about records creation and the role of records in knowledge production when we look closely at how scientists first learn to create and manage their experiment results and account for them, at least to themselves, on a daily, almost mundane, basis?

I use the record in ways drawn from multiple areas of practice as well as the usages of my respondents. This should not indicate conceptual confusion but instead point to a need for focus on process. As noted, in each discipline and even to the members of the laboratory, the record has different meanings, and some of these meanings conflict with each other. Is a record an immutable text, or is it the evidence-bearing document? My goal is not to privilege one or the other in an uncritical way, or even advance a third view. Instead, I wished to ask "How are trust, reliability, professional

DOI: 10.1002/asi

identity, learning, and meaning infused into the daily documents of a laboratory, and what is the relationship of those processes to scientific practice and product?"

Philip's laboratory is somewhat "typical:" It is an academic laboratory engaged in "normal science," to borrow Thomas Kuhn's (1962) phrase. Philip, Susan, Michael, and the others are engaged in accumulating data to fit within the broad theoretical frameworks that underpin their respective disciplines. Their work is not highly collaborative, and most of the members of the laboratory are at the early stages of their scientific careers. They are working with what might seem to be a surprising amount of paper, at least for their documentation. While this case is not unusual in science, it is perhaps somewhat antithetical to recent trends in the information studies literature. Continuing to think about the paper document, and individual relationships with it, is still an essential component of knowledge production and personal information management. Ignoring that relationship obscures the many processes that can still yield insight into digital data management and the use of electronic technologies in laboratory science.

A cursory reading of this article would suggest to some that the scientists in Philip's research group are engaged in "ad hoc" work, reinventing process and product, and that introducing an electronic laboratory notebook or other digital tool would simplify repetitive data capture and integration of external documents and datasets. Although this was not an issue that came up in the context of Philip and his colleagues, or at least was not mentioned as a concern, the alteration or erasure of previously gathered data can be more easily prevented, given the right kinds of digital tools. Indeed, many private scientific enterprises are moving in just that direction to standardize data entry and make available to the laboratory scientists a more complex suite of tools for data manipulation. While this is a laudable goal, it is perhaps not universally a valid one. The experiences of the group of scientists studied in this article suggest that standardization of data entry is not the only aim of scientists when creating records—again, that there is a poetics to the act of records creation that encompasses and perhaps even exemplifies rich and powerful interactions of the social, personal, and professional.

The second aim of this article was to serve as a more general call for further discussion in the information studies community: to consider recordkeeping and the process by which records creators learn to keep those records as a key, if understudied, component of professionalization and learning and to reconsider recordkeeping not in administrative, managerial, or bureaucratic terms but about as knowledge production. By examining recordkeeping in an arena where the creators have a wide degree of latitude in crafting the process and product, we get some sense of the learning and socialization required to make order from chaos, convert "data" to "record" through the integrated processes of synthesis, selection, and standardization. As noted earlier, other work in this arena has opened the door for our thinking about scientific documents as constitutive of memory (Bowker, 2005) and epistemic authority (Myers, 1990). Even more broadly, information creation and management—and the resulting documentary artifacts—can and should be viewed in a larger conceptual framework of policy (local as well as institutional), pedagogy (formal and informal), and practice.

Reflection

Conducting ethnographies of information, as opposed to information creators, is challenging. As my own respondents asked on more than one occasion, "What are you getting out of watching us write things down?" The ethnographer is forced to contextualize the solitary act of routine information management in the larger context of pedagogy and practice, which begs the question: What does this work suggest about the documentary practices of other scientists (including the author, who counts herself as a social scientist)?

In the information studies literature, there has been little room for reflection on documentary practices. As an ethnographer, I would be remiss in not including my own reflections on this subject. While ethnographic observation had made theoretical sense to me, only by fully participating in the process (and for me, that meant recording "subjective" information) could I internalize how and why it worked as a way of understanding the world. In general, there has been far less written about social scientists and their recordkeeping practices than about scientists. The few writings on anthropologists and their fieldnotes (Jackson, 1990; Sanjek, 1991) have suggested that this community is perhaps more straightforward in accepting that affect, identity, and a surprising amount of "ad hocery" are integral to the documentary process of making fieldnotes. For this particular study, and this particular author, these qualities proved to be essential to the process and product being studied—the most important of which was that the quality and completeness of documentation are essential to one's success as a scientist. Not sufficient alone, certainly, but necessary.

Recordkeeping is of profound practical and theoretical import. The processes by which individuals and communities of practice transform "raw" data to record of activity for which they are held professionally accountable should be of great interest to the information community as a transformative act of professional and even personal identity. Understanding how broader professional standards of accountability and reliability become personalized and reified in documentary practices and artifacts is essential to understanding the role of information creation in documentary practice.

Pragmatic concerns enter the picture as well. Scientific knowledge management demands initial control over the records of daily activity before tools can be created, or deployed, to extract meaningful patterns. The development and increasing use of electronic laboratory notebooks, particularly in industrial settings, suggest that this topic is not without its importance, and yet there has been almost no research on what scientists do with "basic" records. This study suggests that we need to return to "first principles:" understanding

DOI: 10.1002/asi

recordkeeping as an act of information creation, not just in academic laboratories by individual scientists and groups, but in other organizational settings. We can do this by placing information creation in conversation with regimes of mutual accountability and value, and thinking carefully about the role of recordkeeping in the professional development of not just scientists but others who are "trained" to keep records of one sort or another. Finally, studies such as this suggest we revisit our thinking about scientific knowledge management, and suggest that there is more theorizing we can and should be doing if development and evaluation of information technologies for creating, sharing, and managing records and data are to be truly useful.

Acknowledgments

The author thanks Ronald E. Day, Josh Tenenberg, and the anonymous reviewers for their helpful suggestions and comments on an earlier draft of this article.

References

Amann, K., & Knorr Cetina, K. (1988). The fixation of (visual) evidence. Human Studies, 11, 133–169.

Berg, M., & Bowker, G. (1997). The multiple bodies of the medical record: Towards a sociology of an artifact. Sociological Quarterly, 38, 513–537.

Botticelli, P. (2000). Records appraisal in network organizations. Archivaria, 49, 161–191.

Bowers, J. (1992). The politics of formalism. In M. Lea (Ed.), Contexts of computer-mediated communication (pp. 232–261). New York: Harvester Wheatsheaf.

Bowker, G. (2005). Memory practices in the sciences. Cambridge, MA: MIT Press.

Duranti, L. (1998). Diplomatics: New uses for an old science. Lanham, MD: Scarecrow Press.

Holmes, F.L. (1987). Scientific writing and scientific discovery. Isis, 78, 220–235.

Holmes, F.L. (1990). Laboratory notebooks: Can the daily record illuminate the broader picture? Proceedings of the American Philosophical Society, 134, 349–366.

Jackson, J.E. (1990). Deja entendu: The liminal quality of anthropological fieldnotes. Journal of Contemporary Ethnography, 19, 8–43.

Kevles, D.J. (1998). The Baltimore case: A trial of politics, science, and character. New York and London: Norton.

Kuhn, T. (1962). The structure of scientific revolutions. Chicago: University of Chicago Press.

Latour, B., & Woolgar, S. (1986). Laboratory life: The construction of scientific facts. Princeton, NJ: Princeton University Press.

Lave, J., & Wenger, E. (1991). Situated learning: Legitimate peripheral participation. Cambridge, UK: Cambridge University Press.

Macrina, F.L. (1995). Scientific integrity: Introductory text with cases. Washington, DC: American Society of Microbiology Press.

Meehan, A.J. (1986). Record-keeping practices in the policing of juveniles. Urban Life, 15, 70–102.

Myers, G. (1990). Writing biology: Texts in the social construction of scientific knowledge. Madison: University of Wisconsin Press.

Ngin, P.M. (1994). Recordkeeping practices of nurses in hospitals. American Archivist, 57, 616–630.

Sanjek, R. (1990). Fieldnotes. Ithaca, NY: Cornell University Press.

Shankar, K. (2004). Recordkeeping in the production of scientific knowledge: An ethnographic study. Archival Science, 4, 367–382.

Star, S.L., & Ruhleder, K. (1996). Steps towards an ecology of infrastructure: Design and access for large information spaces. Information Systems Research, 7, 111–134.

Traweek, S. (1988). Beamtimes and lifetimes: The world of high energy physics. Cambridge, MA: Harvard University Press.

Wenger, E. (1999). Communities of practice: Learning, meaning, and identity. Cambridge, UK: Cambridge University Press.

Winner, L. (1985). Do artifacts have politics? In J. Wajcman & E. Mackenzie (Eds.), The social shaping of technology (pp. 28–38). Milton Keynes, United Kingdom; Philadelphia: Open University Press.

Yakel, E. (2001). The social construction of accountability: Radiologists and their record-keeping practices. Information Society, 17, 233–245.

DOI: 10.1002/asi