TECHNICAL LEAFLET SERIES

Number 10:

Understanding the Record-Keeping Practices of Scientists

Kalpana Shankar

Mid-Atlantic Regional
Archives Conference

Mid Atlantic Regional
Archives Conference

Reviewers
Fynnette Eaton
Janet Linde
Lynne Humphries-Russ
Kathleen Williams

Editor's Note: The idea for a technical leaflet series originated with the New York Caucus of MARAC. The series is intended to provide brief, practical information about selected archival topics suitable especially for beginners in the profession.

The series editor welcomes proposals for future publications and comments from readers.

Technical Leaflet Series
Frank Serene, *Series Editor*
Susan Hamburger, *Chair, Publications Committee*

James Byers, *Chair, MARAC Steering Committee*

UNDERSTANDING THE RECORD-KEEPING PRACTICES OF SCIENTISTS

by
Kalpana Shankar

Doctoral Student
Department of Information Studies
University of California, Los Angeles

Introduction

The modern scientist does not operate in isolation. He or she is connected by complex mechanisms, formal and informal, to a network of colleagues, students, academic institutions, funding agencies, and governmental and private organizations. These private institutions can include scientific industry companies, industrial manufacturers, companies in the information technology sector, and others. Moreover, the scientist must integrate the roles of researcher, teacher, and administrator, among others. Documenting the progress of science requires some understanding of these roles and the ways in which they are intertwined. This pamphlet will describe some of the activities of scientists and the records associated with their work that are likely to be encountered by small repositories.

Most archival literature on the records of science focuses on records that have been preserved in the scientific archive. The aim of this pamphlet is to describe those records that might be accessioned by the non-scientific repository. The focus will be on records of post World War II scientific activity. Perhaps a scientist has donated his or her papers to such a repository, or such records were received as part of a larger acquisition.

Overview of the Scientific Environment

Scientific research can take place in any number of environments: universities, government laboratories (federal and state), research institutes, and private companies. Most such environments are divided into specialized research units, generally referred to as **laboratories** (not to be confused with the physical space in which experiments are conducted). Each laboratory is run by one or more senior scientists, or **principal investigators**. The principal investigator is generally responsible for overall administration and direction of research. He or she is also likely to be responsible for personnel recruitment, mentoring, publication, allocating resources and projects, and procuring funding.

Most university and research institute projects are funded by external granting bodies, which may be public, private institutions, or charitable organizations. Examples of the public institutions are the National Institutes of Health (the major United States granting body in the health and life sciences) and its subsidiaries (such as the National Cancer Institute), and the National Science Foundation. Pharmaceutical companies are examples of the many private firms that fund scientific research. Issues involving proprietary rights to laboratory notes can arise when a scientist receives simultaneous grants from several organizations.

For an academic scientist, the grant funds are allocated both to the university for "overhead" and to the laboratory for its work. University overhead covers infrastructure maintenance and use. Overhead can cover building use, grounds maintenance, storage facilities, and other indirect costs that the academic scientist incurs by virtue of his or her activities on the campus. Overhead costs can vary by discipline, by project, by university, and by granting body, but can take up more than half of the grant.

Much of the day-to-day research activities may fall to the staff in the laboratory. The staff includes four groups of individuals:

Post-doctoral fellows, who are scientists who have completed their doctoral degrees and are thus in a period of apprenticeship. Some may procure their own funding, following the same application process outlined above. Others may be paid through the principal investigator's budget.

Graduate students, who constitute a large part of the research work force.

Undergraduate students who may be employed or volunteer in a laboratory, in exchange for research experience and mentoring.

Technicians, who are staff who may be hired to perform routine laboratory maintenance chores and may conduct research projects of their own (or may assist others in the conduct of their work).

An important point to note is that these individuals are creating scientific records and documentation in the conduct of their research. While most of the administrative papers will be in the hands of the principal investigator, the responsibility for scientific research is dispersed through the network of individuals that comprise the working laboratory, resulting in multiple provenance of documentation. For this reason, a functional approach to appraisal may work better than provenance-based appraisal.

The Records Creation and Use Cycle

The purpose of presenting an overview of the operation and management of a laboratory is to acquaint archivists with procedures for the creation and use of records in the scientific environment. In most laboratories, the principal investigator or senior scientist obtains funds and establishes research priorities. He or she may assign these projects to junior faculty, post-doctoral fellows, graduate students, and others in the laboratory. In some situations, junior faculty and post-docs may have their own projects (and their own funding).

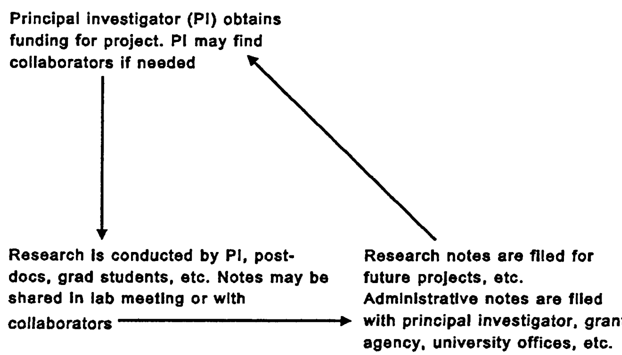
A project may require the attention of more than one staff member. Each individual is responsible for maintaining notes on the progress of the research. Ideally, individuals working on more than one research project will maintain separate notes for each project, but this may not happen in practice. Most laboratories have **lab meetings**, where all of the members of the laboratory meet once a week (or less often) to discuss routine administrative matters, supply and equipment issues, and research problems and ideas. One or more individuals in the laboratory may be asked to present the ongoing results of their research to receive feedback in this informal setting.

At some point in the research process, the researchers must use their notes to put together their results for publication. Usually, the members of the team compile their results and collaborate with the senior investigator to create the final publication. In some cases, the junior members are responsible for the bulk of the writing; in other cases, the senior investigator assumes this responsibility.

When projects are completed or when staff leave, the laboratory retains custody of the research notes. (Such notes are not the property of the research staff, but the individual conducting the research may be allowed to keep copies for his or her own use.) Other terms for research notes include **laboratory notebooks**, **logbooks**, and **data books**. They may be referred to for further research. These notes may also be needed for regulatory processes, especially in the cases of drug development, animal or human subject use, and toxic or radioactive waste disposal. The notes also may be useful in the writing of research-in-progress reports to granting agencies, university committees, and others. Granting bodies also have the right to ask for such notes to resolve legal disputes or to establish priority of research in case patents are filed. These research notes are generally not filed with the administrative papers, but may be maintained elsewhere on the laboratory premises.

Biotechnology companies, other high-technology industries, universities, and similar organizations claim proprietary rights to research notes and have formal procedures and rules for long-term maintenance of research notes in a centralized file. Non-academic institutions have stricter rules regarding ownership and the maintenance of records and have had more success than universities in requiring their staff to relinquish personal possession of these notes. Many scientists consider laboratory notes to be personal and are reluctant to relinquish custody of their research notes. Furthermore, they may be concerned about the ability of a non-scientist to handle scientific documentation appropriately. Also, scientists are always conscious of the need to reference their notes and are reluctant to surrender custody of them.

In summary, laboratory research notes tend to be: 1) dispersed among the personnel of a laboratory unit, at all levels of the hierarchy; 2) maintained by research project; 3) maintained in date order; and, 4) should be treated as lab files over which the scientist has custody and responsibility. The archivist's role in maintaining such records should be to treat such records as part of the work records of the scientists.



General Appraisal Principles

The years after World War II have been termed the era of "Big Science." Before the war, the staff of most laboratories was small, scientists seldom collaborated for the purpose of publication, and most equipment was relatively inexpensive. Since the end of the war, in contrast, there has been a growth in the numbers of personnel staffing scientific laboratories, research expenditures have increased dramatically, and collaborations are ubiquitous. Most scientific documents that find their way into a repository are likely to be only a minuscule amount of all available documentation. Some other sources of related records include university offices, granting agencies, collaborating laboratories in other universities, private companies engaged in collaboration, and government agencies who monitor various research activities (the Food and Drug Administration and the Environmental Protection Agency in particular). As a result, the archivist may not be able to collect all records pertaining to a specific research project.

Haas, Samuels, and Simmons (see Appendix) suggest that the archivist consider where related records might be, what other records could be added to the collection, and how those records affect the overall documentary picture. For example, the papers of an individual scientist may be donated to a private repository. Related records may be properly filed in the grants office of that scientist's university, in the files of affiliated personnel (students, collaborators, and postdoctoral researchers), and in the funding agencies which supported the scientist's work. The archivist must keep in mind that the records with which he or she is concerned will be only part of a complicated trail of scholarly and personal communication and make appraisal decisions accordingly. The archivist may not be able to collect all pertinent records.

Because contemporary (post-World War II) scientific practice has been so convoluted in nature, there are multiple approaches to the appraisal of scientific records. The archivist may wish to consider appraising such records not only by provenance (an approach some have argued may inaccurately reflect the complexity of day-to-day practice of science), but also by the function that those records played in the conduct of scientific affairs. Haas, Samuels, and Simmons recommend this "functional analytic" approach to preserve more accurately the complex environment of science and discuss this approach in more detail in their book. They suggest that scientific records be appraised in the context of personal/professional activities and scientific/technological activities. In the latter category, they include administrative, research/development, and dissemination activities.

Functional Analysis

Samuels uses the term "functional analysis" to describe the process by which an archivist or records manager analyzes the tasks of the organization or institution being documented, then assesses the documentation produced as by-products of that function. For example, in the scientific environment, the principal scientist will serve as researcher, teacher, mentor, colleague, and administrator, among other functions. The archivist, using functional analysis, will be sure to preserve the appropriate papers to reflect this range of roles. Samuels goes into great depth describing functional analysis in her book *Varsity Letters* (see Appendix), using the research university as an example of an institution whose scientific records could be appraised with this approach.

Functional analysis, as opposed to provenance-based appraisal, may be particularly effective in managing post-World War II scientific records. As mentioned above, scientific records tend to be scattered among various institutions, making provenance difficult to establish in many cases. By focusing on the activities that are part of the scientific endeavor rather than the origins of the records, archivists may have more success in documenting science as it is being practiced.

Personal and Professional Papers

Historians, sociologists, and others who endeavor to understand science as a social and cultural practice place great value on understanding the personal and professional lives of scientists. The records that fall into this category are very similar in nature to the majority of papers and artifacts that form collections focused on individuals. Such papers, while perhaps not directly related to the intellectual development of the scientific field in which the person was engaged, still may provide insight into the factors that influenced the person's scientific persona. Some records to look for include:

- Correspondence
- Journals, diaries, scrapbooks
- Educational papers (term papers, class notebooks, syllabi, teaching preparation materials)
- Evidence of professional activity (professional society activity records, consulting records, lecture notes, invitations to present at a conference or publish in a book or journal, grant reviews and reports)
- Financial documents (bills, receipts, contracts)

Scientific Activities

The papers that document scientific activities will probably form the bulk of most scientists' papers. The non-scientific repository is likely to acquire those records that focus primarily on the planning and conduct of research. Administrative and financial papers may be retained by the institution at which the scientist worked. Keeping these points in mind, the records that are most likely to appear document such activities as:

Research planning
Conduct of research
Dissemination of results
Evidence of collaborative and other peer-related activity
Administrative activities

Research planning can run the gamut from the most informal of notes to formal contracts and documents. Archivists should look for sketches, drawings, and jottings, as well as letters and drafts of proposals that plan activities of collaboration with other individuals, universities, research institutions, and private industry. This phase of scientific research may not be isolated from the others; separating these records from records of other scientific activities may result in artificial distinctions.

The most important component of scientific documentation, at least with respect to understanding the day-to-day conduct of research, is the laboratory notebook. In private industry, there are usually very strict guidelines and procedures that must be followed when recording the results of experiments. However, this is usually not the case in the academic environment. While there are suggested formats for maintaining the lab notebook, often such notebooks take on a highly idiosyncratic character. Lab notebooks are generally dated and kept in date order. However, research notes can also be kept in loose-leaf format. Other research documentation including printouts from instruments, lists of references, and photographs may also form part of the laboratory notebook.

An important distinction is the difference between scientific data and scientific records. Raw data in the form of printouts from computers or other instrumentation, without the context of the conditions under which the experiments were performed, is generally not very useful as records of activity. However, if these data are included in the laboratory notebook, they form part of the record of the conduct of research.

The most scripted and finalized forms of dissemination are final publications and conference talks; the original records upon which publications and talks are based may be difficult to find. However, some correspondence the archivist finds may deal with this aspect of scientific research. Preliminary drafts of technical reports and publications may also be included as supporting documentation. Since non-scientific repositories will not have access to esoteric conference proceedings and journals, final copies of publications should be retained. Certainly, any outlines or drafts should be kept as evidence of the creative process.

Some documents to look for include:

- Informal notes on research planning
- Lab notebooks, research notes
- Data books from the operation and maintenance of laboratory equipment

Logs of animal, radioactive material, and toxic material use
Preliminary versions of articles, grant reports, technical reports, and conference papers

Administrative Activities

While the archivist may have difficulty deciding which of these records are necessary to understanding the context of scientific research, there are some administrative documents that should be kept. Evidence of administrative activities may include grant proposals and contracts, working reports, bills for equipment, copies of transactions, and correspondence between the individual scientist and his or her research institute or university. Patents, contracts, grants, litigation records, and technical licenses are important evidence of research activity, and have legal value as well.

Equipment and Tools

While most repositories will not have the expertise or desire to handle scientific artifacts, laboratory equipment and tools should be considered to some extent in the appraisal of a collection. Most laboratory equipment is standardized and, thus, is not needed for the sake of the collection. Some equipment is too large. However, if the development or modification of a particular tool or piece of equipment is the direct result of the work of the scientist, it should be considered for inclusion. Ascertaining the instrument's contribution to the completeness of the records may be difficult; the archivist will have to rely on the rest of the documentation and perhaps the advice of experts in the field to make an informed choice. Other possibilities include conducting oral histories on the development of the instrument and photographing the instrument. Including scientific artifacts in the collection can add to the documentation of science, since the artifacts may themselves be the most important products of research.

Information Files of Scientists

Scientists rely on the work of others to inform their own research. Most scientists collect pre-prints, reprints, review articles, review copies, and journals. Haas, Samuels, and Simmons suggest that historians of pre-World War II science value these files to learn more about the information networks of the scientists they are studying. These papers can be instrumental in understanding who influenced a particular scientist, the other scientists and scholars with whom he or she was in contact, and the research networks in which he or she played a role. However, because of the ease with which one can now obtain scientific publications, scientists tend to collect more of these and it is harder to ascertain their social networks from their files. Thus, such information files may not play the role they once did in documenting scientific work, but should be retained as part of the collection.

Electronic Records

The role of digital records in modern scientific institutions, particularly academic ones, is still in its infancy and not fully understood. Moreover, digital records of scientific activity remain a problem for the archivist, and some of the larger scientific archives are reluctant to handle electronic records. Archivists, however, need to keep in mind that the paper records they receive are increasingly only one part of the way in which scientists conduct their activities. The collection of raw data is often performed solely by computer, so that analyses of results can be manipulated easily with software packages. Laboratory or field notebooks are routinely maintained in electronic form. Revisions to papers and reports are generally made with word processing software. Dissemination of research results and correspondence is increasingly conducted via electronic mail.

To maintain (preserve) and make information available in electronic formats, archivists need knowledge and equipment to deal with the two major problems associated with electronic records: 1) the tendency of the electronic medium to deteriorate in a relatively short period of time, and 2) rapid technological changes, which in a few years renders obsolete the original equipment used for preserving and reading information in electronic format. Whether the archivist has accessioned electronic mail or word processor documents current problems are encountered by the archivist. To access the information in these records, the archive will have to retain the original software and hardware or be prepared to periodically migrate the information to current technology.

Thus, if an archivist accessions electronic records, he or she should plan to move the information from its storage media at designated intervals, depending on the life expectancy of that media. Rapid rate of technological change in both hardware and software compounds the problem. As technological changes occur in hardware and software, the archivist must move (migrate) the information to newer software or be prepared for indefinite maintenance problems on outdated hardware, for which parts and technicians will become scarce as time goes on.

Sometimes, however, the software is itself part of the primary documentation and, therefore, must be maintained as part of the permanent record. This situation can happen, for example, if the electronic "collaboratories" have taken part in several projects and the participants designed software for the express purpose of conducting a research project. The archivist would, in this case, accession the software with which the collaboration was created and a computer that can run the software.

Legal Issues Surrounding Scientific Records

Universities, research institutes, and funding bodies have policies surrounding the ownership and retention of scientific records. Private research and development firms generally have even stricter guidelines. Most such organizations require scientists to sign technology transfer agreements that give the host institution extensive rights with respect to the ownership of intellectual property. Nevertheless, ownership and jurisdiction over records funded by multiple agencies and organizations can present many problems to the archivist.

The archivist may need to determine whether or not the records in the repository's possession have evidentiary value. Research records may be used to establish investigative priority or to provide evidence (important to consider in the context of our litigious society and the establishment of patent or other intellectual property claims), for auditing by granting agencies, by the Environmental Protection Agency, or the Federal Drug Administration for regulatory purposes, and for fraud investigations. Current federal regulations (and many research institutes and university regulations follow suit) stipulate that records must be retained for three to five years after the last expenditure report is filed with the granting agency.

The retention of primary research documentation may be governed by the statute of limitations and, therefore, be much longer in cases dealing with drugs or medical devices, human or animal subjects, radioactive or toxic materials, or classified research. Retention periods and ownership regulations for research records may differ among the various public agencies and private institutions; the archivist may have to consult these organizations to resolve doubts about retention periods and ownership rights. Untangling the morass of legal responsibility of these records may need to be negotiated with the technology transfer office at the university or institute where the records were created.

Health Concerns

Many laboratory notebooks can contain materials that should be handled with caution. Radioactive isotopes, chemical toxins, virus samples, and asbestos may be present on the paper. Archivists should ascertain what kind of research had been conducted before processing scientific collections and they should use latex gloves when appropriate. Research universities have offices that will know more about handling hazardous materials; questions can be directed to them as well.

Science as a Social and Cultural Practice

In the last thirty years, sociologists and anthropologists of science have begun to examine research science as a set of practices built upon definitions of community, language, understanding, norms, and unspoken assumptions. The published scientific paper is a highly formalized document; the practitioners in the appropriate discipline determine its language and components. These assumptions also form the bedrock of daily scientific activity in the laboratory and in the classroom. Anthropologists and sociologists have used their traditional tools of analysis to examine science as a culture and society. For the archivist, understanding some of this literature can be invaluable in thinking about scientific practice in two ways. One, research science is as filled with assumptions, disputes, and negotiations as any other human activity; the archivist would be mistaken in assuming that the progress of science is always a rational process and not a human one. Secondly, modern science is an expensive public endeavor; as such, public institutions demand a great deal of accountability from the records of scientific activity.

Scientific Records and the Work of an Archival Institution

This pamphlet has discussed the general aspects of the scientific record-keeping practice. Less attention has been focused on the role of the archival repository and its traditional functions in handling scientific records collections. This gap is based to some extent on the nebulous nature of scientific collections; handling these collections will depend a great deal on the resources of the repository. In building scientific collections, the repository must consider what other related collections exist and which organizations may hold those collections. Many of the research papers of scientists can be treated as personal papers documenting work. However, as noted above, tackling a larger acquisition of a laboratory or research unit, for example, will require thinking more broadly about the extent of the collection and how such a collection would fit into the repository's mission.

The functions of accessioning, appraisal, and description must be performed in the context of understanding how the particular collection fits into the larger acquisition of documents as well as the administrative value of the records. As mentioned earlier, some scientific records may need to be preserved for auditing and accountability to granting agencies and other government or university bodies. These issues also need to be considered in scheduling and disposal. However, if the collection is to be treated as an historical resource and not as a collection of administrative or legal value, issues of accountability are clearly less important to consider.

Conclusions

Several aspects of scientific research and records of its conduct are worth noting. The records of most modern scientific projects are multi-provenancial. The archivist may almost have no choice but to split collections; records may be dispersed among universities, industries, government, and many individuals. The variety of record formats in science is a function of the large-scale, collaborative, and money-intensive nature of modern science. While the laboratory notebooks may be the best indicators of scientific activity, the administrative documentation is necessary to understand the complex climate of scientific research. Lastly, the authors of the JCAST report (see Elliot, Appendix) suggest that many scientists view published articles as the archival record of their activity. However, these articles are not archival in the sense that historians and archivists use the term, since their format is highly stylized and they do not "tell what happened." Therefore, it may be necessary to convince many scientists that their records form a more accurate picture of their work than their published articles do.

Definitions of Relevant Terms

Laboratory Both the physical setting in which research is conducted and the informal term used to denote a working group, comprised of students, staff, and junior faculty under a principal investigator or investigators

Laboratory meeting Regularly scheduled meeting of project staff to discuss administrative issues and on-going research

Laboratory notebooks Bound books in which a scientist keeps his or her notes during the conduct of research. Also called research notes or data books

Logbook Records kept on the maintenance and use of a scientific instrument, laboratory animals, or radioactive or other hazardous materials. May also refer to the laboratory notebook

Post-doctoral fellows Junior researchers who have completed their doctoral degree and are working as research apprentices

Principal investigator The senior scientist on a project or in a research group

Protocol Specific procedures used to conduct an experiment

Standard Operating Procedure (SOP) The established method for accomplishing a routine task or handling an administrative matter

Technicians Staff who are hired to perform routine tasks and assist with the day-to-day conduct of research. Some technicians may be responsible for their own research projects

APPENDIX

For Further Information

AIP Study of Multi-Institutional Collaborations. Phase I, High-energy Physics. New York, NY: Center for History of Physics, American Institute of Physics, 1992; *Phase II, Space Science and Geophysics.* College Park, MD: Center for History of Physics, American Institute of Physics, 1995.

> The AIP conducted a multiphase study of the organizational structures of large research collaborations in contemporary physics research. Phase I focused on high-energy physics; Phase II focused on space physics. These reports are recommended reading for individuals and institutions responsible for selecting research records of archival value as well as scholars interested in the social context of scientific practice. Reports can be ordered from the American Institute of Physics.
>
> Phone: (301) 209-3165
> Email: chp@aip.org
> Web: http://www.aip.org/history/pubslst.htm

Elliot, Clarke A., ed. *Understanding Progress as Process: Documentation of the History of Post-War Science and Technology in the United States. Final Report of the Joint Committee on Archives of Science*

and Technology (HSS-SHOT-SAA-ARMA). Chicago, IL: Society of American Archivists, 1983.

This JCAST report discusses many of the topics addressed in this pamphlet at greater length. The authors explain some of the characteristics of scientific activity, the unique nature of scientific documentation, and the stakeholders in scientific activity.

Haas, Joan K., Helen Willa Samuels, and Barbara Trippel Simmons. *Appraising the Records of Modern Science and Technology: A Guide.* Cambridge: Massachusetts Institute of Technology, 1985.

This book provides an excellent introduction to the appraisal of scientific records. While it adheres to general principles of archival appraisal, it is laid out in an idealized sequence of scientific activities. Records of each activity and their significance to an overall collection are discussed.

Hefner, Loretta L. "Lawrence Berkeley Laboratory Records: Who Should Collect Them?" *American Archivist* 59 (1996): 62-87.

The article is a case study that illustrates the challenges of documenting contemporary scientific institutions. The author discusses alienated federal records, the integrity of the

record collectivity, and tensions between preserving records for administrative value and historical value.

The **Howard Hughes Medical Institute** publishes pamphlets on archival activities and advice to scientists who wish to document their own careers and donate their records to a repository. For more information, contact:

Records Administrator: (301) 215-8952
Archivist: (301) 215-8658
Fax: (301) 215-8957
E-mail: archives@hhmi.org

Kanare, Howard M. *Writing the Laboratory Notebook.* Washington, DC: American Chemical Society, 1985.

This book introduces the scientist to standards and practices of laboratory record-keeping. The author discusses regulatory requirements, standard format, content, paper quality, and the role of record-keeping in ethical scientific conduct.

Latour, Bruno and Steve Woolgar. *Laboratory Life: the Construction of Scientific Facts.* Princeton, NJ: Princeton University Press, 1986.

The first and most well-known study of the sociology of a scientific laboratory. Latour and Woolgar spent two years studying a laboratory at the Salk Institute in La Jolla, California. An excellent introduction to understanding the practice of science from an "outsider's" perspective.

Samuels, Helen W. *Varsity Letters: Documenting Modern Colleges and Universities.* [Chicago, IL]: Society of American Archivists; Metuchen, NJ: Scarecrow Press, Inc., 1992.

Samuels discusses the concept of functional analysis at great length and suggests some guidelines for function-based appraisal. However, most of the text is concerned with using functional analysis as the appraisal technique for documenting the university setting.

11/99