Responsible Conduct of Research: Data Acquisition, Ownership, Management, and Sharing

L. Gabriel Navar, Ph.D.

Department of Physiology
Renal and Hypertension Center of Excellence
Tulane University School of Medicine
New Orleans, Louisiana

Special thanks to Brian Weimer
“The institution of science involves an implicit social contract between scientists so that each can depend on the trustworthiness of the rest…the entire cognitive system of science is rooted in the moral integrity of aggregates of individual scientists.”

The Common Sense of Science
Jacob Bronowski
Objectives

1. Learn the broad principles of data acquisition, ownership, management and sharing and their effect on Tulane University (“TU”) researchers.

2. Acquire a basic understanding of federal and TU policies and practices governing these topics.

3. Learn where more detailed information can be found.
Ownership and Sharing of Data from Federally Funded Research

- The “Bayh-Dole Act” (The Patent and Trademark Amendment, PL96-517) of 1980 governs the ownership and transfer of technology developed under Federally supported grants and encourages the commercialization of new technologies.

- Section 36 of the Office of Management and Budget (OMB) Circular A-110, Uniform Administrative Requirements for Grants and Agreements with Institutions of Higher Education, Hospitals, and Other Non-Profit Organizations (Revised November 19, 1993, As Further Amended September 30, 1999) regulates the use, ownership, and sharing of data and other intellectual property developed under federal grants, including the rights of the public to gain access to research information under the Freedom of Information Act.
The “Bayh-Dole Act” established a national policy encouraging government, universities, and industry to work together to commercialize new technologies. It removed obstacles that previously blocked transfer of technologies developed with federal funding. Before the Bayh-Dole Act, funding agencies owned the intellectual property developed with their support, and fewer than 5% of the 28,000 patents held by the U.S. government were licensed to industry. Now, the universities usually retain title to intellectual property and are free to market it.
University Regulations and Policies

- Tulane University Policies and Procedures for its Human Research Protection Program (http://tulane.edu/asvpr/irb/upload/Tulane-HRPP-SOPs.pdf)

- Tulane University Institutional Animal Care and Use Committee Standard Operating Policies and Procedures (http://tulane.edu/asvpr/iacuc/hsc/sops.cfm)

- Tulane University Research Misconduct Policy, located in Tulane University Faculty Handbook, section III(H), p. 74 (http://tulane.edu/provost/upload/Faculty-Handbook-2010-11.pdf), which governs how allegations of research misconduct, including fabrication and falsification of data, are handled.
Freedom of Information Act ("FOIA")
Requests

• If you receive a FOIA request, notify the University Research Compliance Office at 504.988.1147 or researchcompliance@tulane.edu. Do not respond until advised to do so.

• Federally funded projects are subject to the FOIA under OMB Circular A-110, Section 36. Although it may be necessary to provide some data to the requester, other data may be exempt from disclosure.

• The rules are complex and change from time to time, and it is impractical for everyone to become an expert in these rules. Therefore, all TU personnel who receive a FOIA request should promptly notify the Research Compliance Office and defer any action until receiving their advice. TU’s FOIA policy is located on the Research Compliance Office’s website at http://tulane.edu/asvpr/research-compliance.cfm.
Principles of Data Acquisition

• Data acquisition is the process of obtaining and recording primary experimental information.

• Proper data acquisition and record keeping is an essential feature of experimental science and technology. It provides the foundation information on which subsequent data analysis and generalizations are based. Without good data collection and record keeping, all subsequent use of the data is tainted by questionable authenticity and accountability. Proper record keeping is of vital importance for patentable inventions.
PRELIMINARY EXPERIMENTATION:

• Develop experimental skills
• Confirm existing concepts
• Provide direct exposure to process or phenomenon
• Help develop an experimental plan or rationale

Data developed from such preliminary experimentation is primarily to help you; the responsibility is primarily to yourself i.e. don’t fool yourself! (Self-deception)
DATA HANDLING AND RECORD KEEPING

FORMAL DATA – documented observations used as basis for public disclosure of conclusions which may have important consequences!

• Further understanding or insight about specific question

• Call for change in lifestyle or therapeutic approach!

Grave moral responsibility to have reliable data which support your conclusions
DATA HANDLING AND RECORD KEEPING

SYSTEMATIC AND DESIGNED OBSERVATION:

Purely observational – requires no direct intervention

• Prospective study design
• Retrospective analysis of existing information

Experimentation – involves some intervention to observe process or phenomenon to

• Allow more precise measurement
• Gain access to information
• Test responses to a perturbation
• Determine limits, etc.
EXPERIMENTATION – the process of performing a deliberate controlled intervention in order to obtain detailed or more complete information

- to confirm or verify conclusions made by others
- to provide new insights leading to new conclusions, further developments to hypothesis, advance of concepts!
DATA HANDLING AND RECORD KEEPING

PRINCIPLES OF OBJECTIVITY:

• Impartial, objective, not biased
• Not motivated by personal gain
• Rigorous test of hypothesis
• Avoid becoming personally attached to a hypothesis or concept
• Willingness to modify concepts or position
• Accept responsibility for validity of report
Data Handling and Record Keeping

PERSONAL GAIN and HUMAN NATURE:

• Desire to please one’s supervisors or mentors
• Desire for promotions and advancements
• Desire for personal recognition (*Be Right!*)
• Improve chances for grant funding
• Facilitate acceptance of data and publication
What Do Scientists Recognize as Data?

- **Quantitative**: Recorded numbers, graphs and charts of numerical raw experimental results, and instrument output including photographs and digital images from which quantitative data can be derived.

- **Qualitative**: Notes of any type, some types of instrument output, photos, movies, and digital images

- **Original samples in unanalyzed form**: e.g., biological specimens

- **Research tools**: Protocols; computer software
Another Way of Classifying Data

- **Raw data**: Information obtained directly from experiments, surveys, etc. Includes information in laboratory notebooks and instrument output; may include information in computers.

- **Processed data**: Graphs, equations, tables, descriptions, summaries, and conclusions derived from raw data but not yet released to the public.

- **Published data**: Information distributed to people beyond those involved in data acquisition and project administration. Theses and dissertations are published when they become available to the public in a library.
A principal investigator (PI) outlines her theory for a certain effect to her graduate student who is involved in acquiring data to confirm or refute the theory. The PI explains her anticipation of finding experimental data having these values. The graduate student generates the experimental data with approximately the anticipated values. The data are published, the PI garners recognition, her grant is renewed, and the graduate student receives his Ph.D.

The next student working on this project, however, has difficulties reproducing the data. After further investigation, it is found that the first graduate student chose in cases of ambiguity of the data those values that came closer to the expected values, emphasizing a trend in the data that was not present in general.
Questions to Consider

1. Should the PI have restrained herself from mentioning her anticipations?

2. Should the PI have more closely supervised the acquisition of the data to verify the accuracy?

3. Should the PI have insisted on a more thorough check of data reproducibility?

4. What should the PI do now about the situation?
Discussion of Case Study 1

- The PI must supervise the design of experiments and the processes of acquiring, recording, examining, interpreting, and storing data. Accordingly, the PI must acknowledge negligence on her part.

- Scientists have a duty to avoid contaminating the literature with incorrect information. At this point, the PI should supervise the experiments necessary to resolve any uncertainties. If she concludes that the original publication was seriously flawed, she should publish a correction.
A PI acquires data that portrays a wonderful correlation explaining a physical effect. About 10% of the data, however, lie far removed from the predicted values. There are explanations as to why these data are different. For example, some experimental parameters have not been well controlled and may have been different for the experiments in which the results deviate. The PI chooses to ignore the outlying data in the publication.
Questions to Consider

1. Should the PI have published the outlying data with an explanation of the limited generality of its correlation?

2. Should the PI have repeated the experiments for these data and ignored them only once they constituted less than five percent of the total data?
The investigator almost certainly erred in throwing out the “outliers.” At minimum he should have performed a statistical analysis of the results. This analysis might have provided justification for disregarding the outliers, but this should have been explained in the publication. If the reason for the deviant results in the outliers was established with reasonable certainty, there might be good reason to throw them out. But that is not the situation described here.
Principles of Data Acquisition

*Laboratory Notebooks*

- Primary records are those set down contemporaneously in laboratory notebooks (datebooks). Suitable notebooks are bound and have sequentially numbered pages. A basic principle: Notebooks should be kept in a way that will enable someone else to repeat each experiment and obtain the same result.

CLEAR DOCUMENTATION is essential for scientific credibility!

• Written evidence of experimental plan (i.e. thesis proposal, grant application)

• Detailed records of specific experiments conducted (i.e. number of dates, raw data, strip recordings, computer printouts, assays, blots, prints, etc.)

• Results from individual experiments that led to conclusions

• Results are not “once in a lifetime” but can be reproduced by you and others
Electronic data notebooks should be validated in some way to assure that the data were actually recorded on a particular date and not changed at some later date. If you collect your data electronically, you must be able to demonstrate that they are valid and have not been changed.
Smith, a chemistry graduate student, begins a laboratory research project. At the start, Smith discusses the project thoroughly with his thesis advisor, Prof. Johnson, who also provides relevant references for the student to read. Johnson, however, does not mention laboratory notebook practices.

Smith begins laboratory work and soon begins to obtain interesting results. Smith and Johnson discuss the results periodically, and Johnson’s interest increases. After about six months, at Johnson’s request, Smith begins to write up the results in the form of a preliminary draft for publication. The publication can later be expanded to make a thesis. In Smith’s draft, the raw data have been processed into graphs, tables, and text.

Upon studying the draft, Johnson has a number of questions about the raw data and asks to see the Smith’s notebook. To his dismay, Johnson finds that no notebook exists; Smith has been keeping records on loose pieces of paper. The records are undated, and many can not be found at all.
Questions to Consider

1. Who is responsible?
2. What can be done?
Discussion of Case Study III

1. Who is responsible?
Experience shows that this sort of problem happens fairly often. While it may seem reasonable for Prof. Johnson to assume that Smith learned about laboratory notebook practices in undergraduate school or from his peers, Johnson cannot be excused. At the outset of the project, it was Johnson’s responsibility to make sure that Smith had a notebook and understood how to use it.

2. What can be done?
Processed results unsupported by original, raw data are unpublishable and unfit for use in a thesis. About all that can be done at this point is to repeat all the experiments that will become part of the publication or of Smith’s thesis. This time, proper notebook records must be kept. While Johnson may be primarily responsible for the problem, Smith will bear the brunt of it; thesis completion will be delayed.
Principles of Data Ownership: Who Owns Research Done at TU?

Raw data (including laboratory notebooks) and processed data are generally owned by the University.

University ownership is subject to conditions established by granting agencies or contracts with sponsors.

Management of research data according to these conditions is implicitly delegated to the Principal Investigator and the Administrator of the unit in which he/she works.
A graduate research assistant works on a sponsored project that financially supports 50 percent of her time while she pursues a Ph.D. degree. Her advisor, the principal investigator (PI), developed the original idea and has predicted some correlations. An undergraduate research assistant is involved in a crucial part of the data acquisition.

The research turns out to be scientifically successful, resulting in a new process that promises commercial revenue through licensing. The graduate student finishes her Ph.D. degree and obtains a copyright for her thesis in which the relevant correlations are reported. The University decides to file a patent application. The graduate student objects, however, because she wants to file a patent application herself, and has found a company to commercially exploit the idea. The graduate student argues that she conceived the idea while working on her own time and that she holds the copyright for an important part of the idea.
Discussion of Case Study

• Is the graduate student entitled to pursue the commercial interests of results derived from her thesis research?
Inventorship is reserved to describe the person, or people, who has created something new, or contributed intellectually to an invention. Only those individuals who meet the definition of inventor can be listed on a patent application. The contributions of the Principal Investigator, the graduate research assistant and the undergraduate research assistant must all be examined using the definition of “inventorship” before a determination can be made.

Ownership refers specifically to the person, company, or institution that holds title to a patent. It is the owner of the patent who governs the way in which the inventions and discoveries covered by the patent are used. Title to a patent is determined by many factors. If the research is externally sponsored, the research agreement will most often determine ownership rights. Where an invention is made while conducting federally sponsored research, the Bayh-Dole Act requires that the University retain title to the invention. For research that is sponsored by private companies, it is not uncommon for the University to assign, up-front, title to inventions.

Thus, even if this idea was developed on the graduate research assistant’s own time, and the PI and the undergraduate research assistant did not make an intellectual contribution, the University would still have title. The graduate research assistant was supported by the University and used University laboratories, supplies and other resources to conduct her research.
Access to raw data and processed data may be restricted for the following reasons:

- Temporary restrictions to allow investigators to complete experimental protocols and to repeat experiments as judged necessary to assure valid results
- Temporary restrictions to preserve intellectual property or copyright claims
- Temporary restrictions due to requirements of granting agencies or contracts with other sponsors
- Permanent restrictions to ensure privacy of human subjects
Decisions to Release Data and Publish

In practice, the PI usually decides whether and how to release data. Considerations include:

- Is the PI confident that the data are accurate and reliable?
- Are data significant enough to publish?
- Should negative results be published?
- Have obligations to granting agencies and project sponsors been satisfied?
There is no universal standard for how long raw and processed research data should be retained.

- Some federal agencies, such as NIH, require that data be retained for three years after completion of the project.

- In general, three years should be considered a minimum in academia. Some experts recommend retention of raw and processed data for five years. Many companies, however, have retention cycles.

- When patent or other legal issues are involved, advice of an attorney should be sought before any records are destroyed.

Proper data retention and storage is the responsibility of the Principal Investigator and the Administrator of the unit in which he/she works.
Principles of Data Management

Case Study

A PI and a graduate student working on an Air Force-sponsored project obtained data that led to a rethinking of some fundamental aspects of High Tc superconductivity. The resulting publications stimulated a significant amount of discussion and enhanced the investigators’ careers. About three years after the original publication date, however, other investigators suggested a different interpretation of the data. Thus, the principal investigator (PI) had an urgent need to reevaluate the raw data, taken four to five years earlier. Unfortunately, the laboratory notebook had vanished after the graduate student left, and the computer files were thrown out with an old computer one year earlier.
Questions to Consider

1. To what extent should the PI ensure that important raw data are retained?

2. What procedures are in place at TU to store laboratory notebooks and computer files?
1. To what extent should a PI ensure that important data are retained?

At minimum, the PI should have ensured that the requirements of the Air Force contract for record retention were satisfied. Presumably, these requirements would have been similar to those of other government agencies – three years after completion of the contract.

In this case, the PI is the principal victim of his failure to do so. In other cases, TU or its contractors might suffer by loss of raw data supporting patents.
2. What procedures are in place at TU to store laboratory notebooks and computer files?

At present there is no centralized system. Each PI and each unit are responsible for proper record retention.
DATA HANDLING AND RECORD KEEPING

SCIENTIFIC ERRORS:

- Design level – experimental design favors desired results
- Experimental Level – undesirable or negative results are discarded or disregarded
- Analysis Level – statistical treatment is not appropriate or grouping is forced
- Interpretation Level – personal bias leads to interpretation not consistent with data
- Fraud – deliberate error with intent to deceive
TYPES OF ERRORS OR BIAS

• Sampling Error – due to chance variation in sample selection, usually because of very small sample size
• Selection Bias – distortion resulting from manner in which subjects were selected
• Information Bias – distortion due to measurement error or misclassification of subjects
• Confounding Error – influence of uncontrolled other variables that are linked with studied independent and dependent variables
DATA HANDLING AND RECORD KEEPING

FRAUD

There are many different reasons why erroneous data may result or incorrect interpretations may be made. Mistakes of this nature can be costly and may have serious consequences….but they are NOT FRAUD

Fraud involves the deliberate intent to deceive

• Saying you ran experiments that you did not
• Saying you ran more experiments than you did
• Changing the data to fit your bias
• Intentional misinterpretation of data
Research Misconduct

Fabrication, Falsification or Plagiarism in proposing, performing, reporting, or reviewing research results

• Fabrication – making up results and recording or reporting them

• Falsification – Manipulating materials, equipment, processes; changing or omitting data such that research is not accurately represented

• Plagiarism – Appropriations of another’s ideas, processes, results or works without giving credit to source
RESEARCH WITH HUMAN SUBJECTS

CLINICAL/TRANSLATIONAL RESEARCH involves unique problems:

- Conflicts of interest related to outcome of treatment
- Conflicts of interest related to financial support from drug companies
- Problems related to placing human subjects on placebo or less effective treatment (?)
- Difficulties in ensuring compliance by patients
- Risks to human subjects must be given very serious consideration because consequences may be irreversible
Human Subjects

Special data-management practices are necessary for research projects involving human subjects. The most common types of projects involving human subjects are:

- Medical research that may involve personal medical information, including DNA information, about individuals
- Psychological testing and interviews
- Opinion surveys, including questionnaires and interviews

Each discipline has well-established protocols for protecting the types of information it handles. Knowledge of these protocols is essential for investigators in a particular discipline. In general, all records should be kept confidential, and written records should be carefully secured. Failure to adhere to these protocols could result in serious harm or embarrassment to not only the subjects but also the investigators. In extreme cases, it could subject the investigators and TU to legal action.

"In the computer model the only side effect was a dry mouth."
Industrial Contracts

Industrial sponsors of academic research often want to have raw processed data secured at the university. They have legitimate reasons for wanting security, since premature distribution of the information could help competitors and could jeopardize patent rights.
Industrial Contracts (continued)

While legitimate, the sponsors' desire for information security does not harmonize well with normal practice in many university laboratories. Universities generally operate in an open way, including

- information flowing informally and in seminars among investigators working on different projects
- shared work-spaces, instruments, computers, and networks
- absence of strict practices for securing raw and processed data within laboratories (except where human subjects are involved)
- relatively open reporting to the university administration
Industrial Contracts (continued)

These differences are best addressed as the initial agreement is formulated between the sponsor and TU’s Office of Research Administration (ORA). Compromises may be necessary from both sides. Once an agreement is reached, all University personnel involved in a project should be made aware of the contract provisions. The Principal Investigator and the administration should make a good-faith effort to carry out the agreed practices.
Academic tradition requires publication of all significant research results. Free exchange of ideas and information within the institution is also traditional, for example, in student seminars on research in progress.

It is a challenge for investigators to harmonize these traditions of openness with the requirements for protecting ownership rights to intellectual property, described previously. A degree of compromise is inevitable.
The academic tradition of openness is supported by federal funding agencies, but with qualifications. An example is the NSF policy statement on “Sharing of Findings, Data, and Other Research Products” (http://www.nsf.gov/pubs/gpg/nsf04_23/6.jsp):
Principles of Data Sharing
Government Funding (continued)

• “NSF expects significant findings from research and education activities it supports to be promptly submitted for publication, with authorship that accurately reflects the contributions of those involved. It expects investigators to share with other researchers, at no more than incremental cost and within a reasonable time, the data, samples, physical collections, and other supporting materials created or gathered in the course of the work. It also encourages awardees to share software and inventions or otherwise act to make the innovations they embody widely useful and usable.”

• “Adjustments and, where essential, exceptions may be allowed to safeguard the rights of individuals and subjects, the validity of results, or the integrity of collections or to accommodate legitimate interests of investigators.”
Determining authorship and inventorship are often sensitive issues, and they have led to many heated controversies. The NSF says that papers should be published with “authorship that accurately reflects the contributions of those involved.” A good guideline, but who decides?
Usually, the PI has the first-line responsibility for determining who is listed as author of each publication or presentation from his/her group and for determining the order in which the authors are listed (a significant matter, since the order is often considered to rank the relative contributions of each coauthor). Investigators are advised to:

- Follow the practices and customs of his/her field. In some fields, the PI's name traditionally comes first, while in others the custom is to list it last, but with an asterisk indicating that this is the author “to whom inquiries may be addressed.”

- Include, rather than exclude, co-authors whose contributions appear to be of borderline significance. Researchers who believe they have been slighted may appeal the PI’s authorship decisions to the university administration.
The co-inventors of a patent may not be the same people as the co-authors of a paper describing the patent. For example, the invention may have occurred before some of the co-authors even started working on the project. Since incorrectly listing inventors can jeopardize validity of a patent, and since inventorship is the subject of a complex body of patent law,* it is recommended that a patent attorney be consulted about inventorship at the time each patent is being prepared for filing.

Check Out the Data First!
L. Gabriel Navar, Ph.D.
Useful Notebooks Explain:

- Why you did it
- Where ideas came from and who was involved
- How you did it
- What materials you used and where they can be obtained
- What happened—and didn’t happen
- Your interpretation
- What’s next
The Conditions for Carrying Out Federal Research Grants

• Set by the Office Management and Budget
• Set by the granting agency
• Codified in the Code of Federal Regulations
The National Society of Professional Engineers Code of Ethics for Engineers states:

- Section II.3.a: "Engineers shall be objective and truthful in professional reports, statements, or testimony. They shall include all relevant and pertinent information in such reports, statements, or testimony...."

- Section III.3.a: "Engineers shall avoid the use of statements containing a material misrepresentation of fact or omitting a material fact necessary to keep statements from being misleading...."
Good Notebook Records:

• Are made at the time the work was performed
• Are in bound notebooks with sequentially numbered pages
• Are the work of one person—each researcher keeps his/her own notebook
• Are written legibly in ink
• Are written in sequence (page 1, 2, 3, etc.) with no blank pages; large blank areas on finished pages are canceled; no pages are torn out
• Are dated and signed by the investigator on each page
• Are not altered later unless corrections are dated and signed
• Ideally are witnessed and dated by a neutral party (vital to establish inventorship date if patent action is planned)
• Are secured in locked files or desks when not in use
• Have photos, drawings, instrument data, and printouts attached and referenced
• May be copied for convenience
Principles of Data Management (continued)

• Decisions to Release Data
• Human Subjects
• Industrial Contracts
CARDINAL NORMS

• Data are ultimately judged in terms of intellectual criteria, rationality i.e. do they make sense?

• Disinterestedness (emotional neutrality) – objective is extension of scientific knowledge not personal gain

• Skepticism – healthy degree of uncertainty, unwillingness to accept on basis of authority

• Communalism – openness, collegiality, unselfish sharing of information and ideas