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

W T. Godbey, Ph.D.

Department of Chemical & Biomolecular Engineering
Tulane University
New Orleans, Louisiana

Special thanks to Brian Weimer and Gabriel Navar
“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.
I. Data Acquisition
II. Objectivity
III. The Laboratory Notebook
IV. Data Errors
V. Data Ownership
VI. Data Management
I. Data Acquisition
Principles of Data Acquisition

• **Data acquisition** – The process of obtaining and recording primary experimental information

  – Proper data acquisition and record keeping
    • Provides the foundation on which subsequent data analysis and generalizations are based.
    • Without good data collection and record keeping, all subsequent use of the data is tainted.
    • 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
Formal Data

- Documented observations
  - Further understanding or insight about specific question
  - Used as basis for public disclosure of conclusions

Grave moral responsibility to have reliable data which support your conclusions
Systematic and Designed Observation

• Purely observational aspects:
  ✪ Requires no direct intervention
  • Prospective study design
  • Retrospective analysis of existing information
Experimentation

• Involves some intervention to observe process or phenomenon
  – Allow more precise measurement
    • (relocate detector)
  – Gain access to information
    • (open up animal)
  – Determine limits, etc.
    • (Adjust gain, optimize resolution)
II. Objectivity
What Do Scientists Recognize as Data?

• Quantitative:
  – Recorded numbers, graphs and charts of raw experimental results
  – Instrument output from which quantitative data can be derived
    • Photographs, digital images
      – Cells can be counted, nanoparticles can be measured, …

• Qualitative
  – Notes (not including written measurements)
  – Some instrument output
    • Images and movies
      – Histology slides, movies of rolling neutrophils

• Potential
  – Original samples in unanalyzed form
    • biological specimens
Another Way of Classifying Data

- **Raw data**
  - Obtained directly from experiments
  - Information in laboratory notebooks and instrument output

- **Processed data**
  - Graphs, equations, tables, descriptions, summaries

- **Published data**
  - Information distributed to people beyond those involved directly in the project.
    - Includes theses and dissertations
Principles of Objectivity

• Impartial, 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
Personal Gain and Human Nature

• Desire to please one’s supervisors or mentors
• Desire for promotions and advancements
• Desire for personal recognition
• Improve chances for grant funding
• Facilitate acceptance of data and publication

“These are the data, we cannot change them”
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.
III. The Laboratory Notebook
Principles of Data Acquisition

**Laboratory Notebooks**

- Primary records are those set down, as they occur, in laboratory notebooks.
- Suitable notebooks:
  - Are bound
  - Have sequentially numbered pages.
  - Have dated entries
  - Have all entries written in ink
  - Inserts are signed on and off the insert
  - Have no erasures/obliterations
  - Have no pages torn out

Notebooks should be kept in a way that will enable someone else to repeat each experiment and obtain a similar result.
Clear Documentation

- Essential for scientific credibility
- More than just laboratory notebooks
  - Written evidence of experimental plan
    - thesis proposal, grant application
  - Detailed records of specific experiments conducted
    - raw data, strip recordings, computer printouts, assays, blots, etc.
  - Conclusions

Results are not “once in a lifetime” but can be reproduced by you and others
Electronic Laboratory Notebooks

• Should be validated in some way to assure that:
  
  – data were actually recorded on a particular date
  
  – data were not later changed

If you collect your data electronically, you must be able to demonstrate that they are valid and have not been altered.
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. What should be done?
2. Who is responsible?
Discussion of Case Study III

1. What can be done?
   Processed results unsupported by original, raw data are *unpublishable*. It is unfortunate, but all the experiments must be repeated if they are to be published (including in the thesis).

2. 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.

   While Johnson may be primarily responsible for the problem, Smith will bear the brunt of it; completion of his thesis will be delayed.
IV. Data Errors
Scientific Errors

- Design level
  - **Skew**: experimental design favors certain results
- Experimental Level
  - Undesirable or negative results are disregarded
- Analysis Level
  - Statistical treatment is not appropriate
  - Grouping is forced
- Interpretation Level
  - Personal bias leads to erroneous interpretation
- Fraud
  - Deliberate error with intent to deceive
Types of Errors

• Sampling Error
  • Due to chance variation in sample selection
    • Sample size may be too small

• Selection Bias
  • Distortion resulting from manner in which subjects were selected
    • Taking a poll in front of the Democrat National Headquarters
Types of Errors

• Information Bias
  • Measurement error
    • Machine not calibrated
  • Misclassification of subjects

• Confounding Error
  • Influence of uncontrolled variables that are linked with the independent and dependent variables under study
    • Incubator ran dry, cells responded differently
    • Mice were affected by researcher turning on the lights in the middle of the night for another experiment
    • Phase of the moon?
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
V. Data Ownership
Ownership and Sharing of Data from Federally Funded Research

• The Bayh-Dole Act
    • Governs ownership and transfer of technology developed under Federally supported grants
    • Encourages commercialization of new technologies.

• Office of Management and Budget (OMB) Circular A-110, (Sect. 36)
  — Uniform Administrative Requirements for Grants and Agreements with Institutions of Higher Education, Hospitals, and Other Non-Profit Organizations
    • Regulates use, ownership, and sharing of data and other intellectual property developed under federal grants,
    • Includes the rights of the public to gain access to research information under the Freedom of Information Act
Bayh - Dole Act

- Established a national policy encouraging government, universities, and industry to work together to commercialize new technologies.
- 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 Institutional Animal Care and Use Committee Standard Operating Policies and Procedures ([http://tulane.edu/asvpr/iacuc/hsc/sops.cfm](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](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.
Who Owns Research Done at TU?

• **Tulane!**
  – Raw data (including laboratory notebooks)
  – Processed data

(subject to conditions established by granting agencies or contracts with sponsors.)

• Management of research data
  – Implicitly delegated to the:
    • Principal Investigator
    • 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.
Question to Consider

• 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.
VI. Data Management
Access to data may be restricted

**Temporary restrictions**
- to allow investigators to
  - complete experimental protocols
  - repeat experiments to assure valid results
- to preserve intellectual property or copyright claims
- due to requirements of granting agencies or contracts

**Permanent restrictions**
- to ensure privacy of human subjects
Decisions to Release Data and Publish

The PI usually decides whether and how to release data.

- Considerations include:
  - Are the data
    - Accurate
    - Reliable
    - Significant enough for publication
      - This includes negative results
  - Have obligations to project sponsors been satisfied?
Data Retention and Storage

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.
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 a minimum, the PI should have ensured that the requirements of the Air Force contract for record retention were satisfied – three years after completion of the contract.

In this case, the PI is the principal victim of his failure to keep data longer. 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 sharing: Academia vs. Industry

• Academic tradition requires publication of all significant research results.
  – Papers
  – Conferences
  – Seminars
• Free exchange of ideas and information within the institution is also traditional
  – Discussions with colleagues
  – Student seminars on research in progress.
  – “Research Days” posters

• Industrial realities often require a level of secrecy surrounding new findings.
  – Aiding competitors
  – Jeopardizing patent rights.

It is a challenge for investigators to harmonize these traditions of openness with the requirements for protecting ownership rights to intellectual property. A degree of compromise is inevitable.
Formalizing the Compromise

• These differences are best addressed as the initial agreement is formulated between the industrial sponsor and TU’s Office of Research Administration (ORA).

• 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.

  – Often, industrial sponsors will want to have raw processed data secured at the university.