Data Acquisition, Retention, Ownership & Sharing*

*Acknowledgments/Resources

The following slides are largely taken from Lecture 2 of GRAD544, shared with me by Dr. Kathy Partin

What are "data"?

- **Data are a collection of facts, measurements, or observations used to make inferences about the world we live in.**

- Examples of data:
  - Material created in a wet laboratory, such as an organic compound, electrophoresis gel or a DNA sequence
  - Information obtained in social-science research, such as a filled-out questionnaire, videotapes, and photographs
  - Microscope slides, cell lines, climate patterns, soil samples, astronomical measurements, and spectrographic analyses
  - Custom software or hardware and specialized methods
  - Observations & interpretations
  - Mathematical proofs (?)

Legal Definitions…

- What is a "data set"
  - An electronic device that provides an interface in the transmission of data to a remote station
  - A collection of related data records on a storage device
  - *Computing:* another name for file
  - A complete set of "x-y" data
- What is the important (IP, scientifically relevant) part of "data"
  - The data itself?
  - The interpretation?

And what about "metadata"?

- Data providing information about one or more aspects of the data, such as:
  - Means of creation of the data
  - Purpose of the data
  - Time and date of creation and creator/author of data
  - Placement on a computer where the data were created
  - Standards or methods used

- Metadata
  - Are data and can be stored/managed in a database/ registry or repository
  - Are a required element of NSF's requirements for datastorage

Who owns data collected at CSU*?

- In most cases, CSU owns it, but others have rights
  - If there is a federal sponsor, CSU owns it but the PI is the steward of the data.
- What are the PI's responsibilities?
  - The collection, recording, storage, retention, and disposal of data
    - Retain sufficient detail to allow replication of research or response to questions about interpretation, authenticity, or validity
  - Make decisions about disseminating data
- When data are published,
  - Copyright is retained by the PI, who then (usually) assigns it to the publisher of the journal.
  - Data and data books collected by undergraduates, graduates, and postdoctoral fellows on a research project belong to the grantee institution (CSU)

* Bayh-Dole Act of 1980 allows universities to have control of IP arising from federally-sponsored research

CSU AP Manual

SECTION J. RIGHTS AND RESPONSIBILITIES RELATED TO CREATIVE WORKS
(last revised June 14, 2000)

J.1 General Policy

Inventions, Academic Materials, Publications, and other creations, hereafter referred to as Works, are the natural outgrowth of activity in teaching, research, and service. University academic faculty members, administrative professionals, state classified staff, student employees, and anyone affiliated in a professional capacity with the University and using University Resources, who are inventors and creators (hereafter referred to collectively as "Members"), shall provide notification of their creation of Works to the appropriate administrator(s) and indicate their association with the University in all professional Publications.

In the interest of encouraging the development of new and useful scholarly material and the publication of such works, the University will continue the tradition of not claiming ownership or a share of the proceeds from scholarly works such as Academic Materials, textbooks or their equivalent not supported through the use of University Resources as defined below or expressly commissioned by the University or popular or scholarly nonfiction, novels, poems, sculpture, musical compositions, or other artistic works. A faculty member's general responsibility to produce scholarly and creative works does not constitute an express commission of Works. Subject to the use of such courses for continuing or distance education purposes, (see Section J.12.3) members are sole copyright holders in their own lectures and any publication, recording, or broadcasting of lectures must be authorized by the Members concerned.

Data Storage and Retention

How long do you keep data and where do you keep it?

- NIH (and NSF) rules:
  - Grant recipients must keep all data for 3 years beyond end of the grant.
  - Different agencies and academic societies have different requirements
  - Universities usually make PIs responsible for rules
- American Psychological Association
  - Retain data for a minimum of 5 years
  - Many universities: require 5 year minimum

<u>Why?</u>

- Patent protection
  - FDA: 10 years   U.S. Patent: 23 years
- Misconduct allegations based on data
  - In misconduct cases, the absence of data during the mandated retention period may be interpreted as evidence of guilt
- In practice:
  - Store data as long as it feels necessary to do so
  - Confidential data must be stored in such a way that access cannot be available
  - Audits may be necessary to determine whether data have been stored properly
  - Some store electronic data with archival resources.

CSU Sponsored Programs Principal Investigator's Manual

- "The University's policy on retention of documents can be found in Section J-7 of the Financial Policy and Procedure Manual. This policy states the general retention period for various University documents, e.g., journal entries, purchase requisitions and DPO's. As indicated in this policy, the University's period for retaining original documentation in support of expenditures on sponsored agreements may be less than required for the period of a sponsored agreement plus the subsequent three-year retention period. For example, a purchase requisition issued at the start of a three-year project, plus the three-year retention period, would need to be retained for a six-year period for audit purposes, while the University's retention period for that document could be only five years from date of issue."

- "By University policy, **departments and/or principal investigators are responsible for retaining pertinent documentation on individual sponsored projects. Such documentation** would include not only financial transactions and time and effort certifications, but statistical data, such as lab books, data tapes, graphs, case studies, field notes, original samples in unanalyzed form, and reports as well. In some situations, it may be possible to substitute microfiche copies in lieu of original records. On some agreements, particularly contracts, a sponsor may request transfer of certain records to its custody; in that situation, the three-year retention period does not apply. Contact your Sponsored Programs Research Administrator for information related to specific projects."

"You are completely free to carry out whatever research you want, so long as you come to these conclusions."

Avoiding Disputes About Data Ownership and Sharing

- Before a collaboration (or training period) begins:
  - Specify who is the "principal investigator"
    - Responsible for IRB or IACUC protocols
    - Responsible for stewardship of the research data
  - Specify the co-investigators
    - Are they technical consultants or collaborators?
    - Do they have full or conditional access to data?
    - What are the expectations for authorship?
  - Specify who is mentoring whom

Sharing of Research:
Preliminary Results, Materials, Ideas

- Science (research) is not an individual experience, it is shared knowledge
  - Social conventions of science play an important role in establishing the reliability of scientific knowledge
- Social conventions currently used arose from latter half of 17th century
  - Isaac Newton: fear that someone else would claim priority - frequently realized
  - Solution came from Henry Oldenberg (secretary of the Royal Society of London)
    - Making new discoveries public while assuring author's credit
    - Guaranteed rapid publication in Phil. Trans.
    - Official support of the society if the author's priority was brought into question
  - Birth of peer review system and scientific journals

Published Results

But in science, the credit goes to the man who convinces the world, not to the man to whom the idea first occurs.
                                                            -Sir Francis Darwin

- Once a result is published
  - Can be freely used to extend knowledge
- KEY: until it becomes common knowledge
  - People who use results are obliged to recognize the discoverer through citations
  - Scientists' "reward" for making results public
  - Citation practices
    - Whole other area of concern in ethical practices

Before publication...
different considerations apply

- Exploitation of unpublished material
  - Seen in privileged grant applications, unpublished papers
    - Essentially stealing intellectual property
- Scientists are generous in discussing preliminary theories or results
  - Provide copies of raw data
  - Not expected to make their data and thinking available to others at all times
  - "Privacy" period – allows one to get work to a point of confidence in both accuracy and meaning

After publication…

- Scientists expect data and other research materials to be shared
  - With qualified colleagues upon request
  - Federal agencies, journals and societies all have guidelines/policies
- Materials too voluminous, unwieldy, or costly to share freely
  - Ok, but if it's at all possible, it should be done
    - Risk of appearing untrustworthy or loss of trust
    - Professional isolation can damage scientist's work

Presenting (unpublished) Research Results

Ben, a third year graduate student, had been working on a research project that involved an important new experimental technique. For a national meeting in his discipline, Ben wrote an abstract and gave a brief presentation that mentioned the new technique. After his presentation, he was surprised and pleased when Dr. Freeman, a leading researcher from another University, engaged him in an extended conversation. Dr. Freeman asked Ben extensively about the new technique, and Ben described it fully.

Six months later, Ben was leafing through a journal when he noticed an article by Dr. Freeman. The article described an experiment that clearly depended on the technique that Ben had developed. He didn't mind, in fact, he was again, somewhat flattered that his technique had so strongly influenced Dr. Freeman's work. But when he turned to the citations, expecting to see a reference to his abstract or presentation, his name was nowhere to be found.

1. Does Ben have any way of receiving credit for his work?
2. Should he contact Dr. Freeman in an effort to have his work recognized?
3. Is Ben's faculty advisor mistaken in encouraging her students to be so open about their work?

Are there rules about openness?

- Would it have made any difference if the situation were the same except
  - (a) Ben's work had already been published?
  - (b) Dr. Freeman was the presenter and Ben was a member of the audience?
- Is it prudent to orally present work that has not yet been published?
  - As a junior researcher, how do you know what's ok?
- What responsibilities, if any does one have when using ideas presented by someone making an oral presentation?
  - What about ideas obtained in a private conversation?

NIH Policy on Sharing

"NIH considers the sharing of such unique research resources (also called research tools) an important means to enhance the value of NIH-sponsored research. Restricting the availability of unique resources can impede the advancement of further research. Therefore, when these resources developed with NIH funds and the associated research findings have been published or after they have been provided to NIH, it is important that they be made readily available for research purposes to qualified individuals within the scientific community."

"In addition to sharing research resources with the research community, upon request of the NIH awarding office, the grantee also must provide a copy of documents or a sample of any material developed under an NIH grant award. The grantee may charge a nominal fee to cover shipping costs for providing this material. Income earned from these charges must be treated as program income..."

http://www.ott.nih.gov/policy/rt_guide_final.html

More on NIH's policy

In 2003, the NIH instituted this new policy on data sharing. The new policy applies to investigator-initiated one-year $500,000 grants and may have an impact on smaller grants too. The goal of the policy is to expedite the timely release and sharing of final data to enhance the research enterprise.

Release of data can be complex. Intellectual-property considerations, non-governmental sponsorship issues, and human-subject confidentiality protection must be considered before data are released.

The NIH is requiring that investigators include with their grant applications information about how they plan to share the data generated from their research. If a grant is awarded, the data-sharing plan must be enacted.

Sharing Data: National Security

U.S. Patriot Act:

The current U.S. Patriot Act attempts to balance the ability of researchers to share data with national-security interests. Among its many provisions, the act creates restrictions on the transport of potentially dangerous biological specimens. It also characterized a type of research, called "sensitive but unclassified," which requires review by the federal funding agency before publication.

More Federal Guidelines

Council on Governmental Relations:

The council describes the definition of data as wide, at least as seen through the lens of different governmental agencies, such as the NSF, the PHS, and the EPA. Each has different requirements for what they consider data and, therefore, for data retention and sharing. The council also addressed data custody, pointing out that universities usually give custody of data to researchers, who then are responsible for keeping the data in trust, "not moving or destroying it without appropriate advance notice and permission from the legal owner" - which, typically, is the university. Who has access to data is also taken into account. Universities claim access to researcher data, especially in cases of scientific misconduct, but who else should have access is subject to ethical, intellectual-property, and research-based considerations. Concerns are time of access (before or after publication), level of access (raw, transformed, or summarized data), and cost of sharing the data.

More Federal Guidelines

"The NSF expects significant findings from research and activities it supports to be promptly submitted for publication, with authorship that reflects the contributions of those involved. It expects investigators <u>to share with others</u> at no more than incremental cost and within a reasonable time, the data, the 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 to make them useful and usable. <u>Exceptions may be allowed to safeguard the rights of individuals and subjects, the validity of results or the integrity of collections</u>."

NSF: Data Management Plans Now Required

CSU template:

"This Data Management Plan addresses the NSF policy that primary data "commonly accepted in the scientific community as necessary to validate research findings" be made available at little or no cost to the PI or project. In accordance with this policy and guidance from the Office of Management and Budget, this plan does not include preliminary analyses (including raw data), drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. Data that must be withheld long enough to enable peer review and publication/dissemination or protection of intellectual property are subject to this plan only after those steps have taken place."