Acquisition, Management, Sharing, and Ownership of Data

Responsible Conduct of Research Training

FRANK VAN BREUKELEN
UNIVERSITY OF NEVADA LAS VEGAS

PRESENTED BY: JOSEPH LOMBARDO
What Type of Data Do You Use/Generate? What Do you Think are the Issues?
Why are we here discussing data?

Data practices are central to all discussions of Responsible Conduct of Research (RCR)

- Fabrication and falsification (research misconduct) are typically data related.
- Sloppy data collection, management, or analysis practices can lead to biased results used by others.
- Incredible capacity for analysis and possibility of mistakes.
Data is a hot topic


- Authors point to the incredible amount of data now available and the opportunities and challenges related to data.
- The articles discuss many of the ethical dilemmas discussed in this presentation.
- **Big Data!**
What are data?

- NIH

“recorded information, regardless of the form or medium on which it may be recorded, and includes writings, films, sound recordings, pictorial reproductions, drawings, designs, or other graphic representations, procedural manuals, forms, diagrams, workflow charts, equipment descriptions, data files, data processing or computer programs (software), statistical records, and other research data.”
What are data?

- In the Office of Management and Budget’s (OMB) Circular A-110, research data are defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues.”
What are data?

- NSF

“What constitutes such data will be determined by the community of interest through the process of peer review and program management. This may include, but is not limited to: data, publications, samples, physical collections, software and models.”
Preliminary or “raw” data are not included for the purposes of access by the general public.

Investigators must retain this raw data in laboratory notebooks or records for purposes of validating research findings.

The raw data serves other purposes as well, such as patent applications, investigations of misconduct, or if the research results are used for public policy or regulatory purposes.

The definition provides the following exclusions: “This recorded material excludes physical objects (e.g., laboratory samples). Research data also do not include: (A) Trade secrets, commercial information, materials necessary to be held confidential by a researcher until they are published, or similar information which is protected under law; and (B) Personnel and medical information and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study.”

OMB definition applies across federal agencies
NSF now requires a data management plan for every award submitted. It must:

- Describe the data. Identify the data that will be measured, recorded, calculated, or modeled as well as the manner in which these tasks will be performed.
- Present the context of the data. Explain the rational for collecting or producing the data, as well as what insights could be gleaned from the data.
- Explain the nature of the data and identify the type of data collected.
- Describe the method for preserving and/or curating the data. Indicate how data will be backed up, as well as how it will be stored both on- and off-site.
NSF Data Management Plan Requirements Cont.:  
- Discuss the approach for accessing the data, if relevant. Be aware that when requested, data should be available for sharing within a reasonable period of time; understand that different communities may define the term “reasonable” in different ways.  
- State how long the data will be preserved and/or curated. NSF requires data to be preserved at a minimum three years beyond the end of the award.  
- Clarify ethical and/or privacy issues associated with the data, if relevant. Explain how these issues will be addressed.  
- Detail intellectual property concerns associated with the data, if relevant. Explain how these concerns will be addressed.  
- UNLV Sponsored Programs  
  (http://research.unlv.edu/osp/resources.html)
Federal requirements: Managing and Sharing Data

- NIH (2003) requires a data sharing plan only when grant is $500,000 or more per year in direct costs.
- NIH’s policy encourages timely release and sharing of final research results
  - “Timely release and sharing” is defined as no later than the acceptance for publication of the main findings of the final data set.
- NIH supports the sharing of unique research resources or research tools under reasonable terms and conditions for dissemination and acquiring the tools.
  - No funding amount limit
Federal requirements: Managing and Sharing Data

- Adequate management of federally-funded data
  - OMB stipulates 3 years or longer as required by granting agency (6 years by HIPAA, 2 years FDA)
  - If patent- life of patent- typically 20 years
  - If contested- until investigation is completed
  - Electronic copies are okay provided that there’s notification.
Who Owns the Data? (from HHS)

Case Vignette: Data Ownership

Dr. Smith works at The University and is the Principal Investigator on a large research project that is funded by the National Institutes of Health (NIH). However, while Dr. Smith wrote the original grant proposal, he does very little day-to-day work on the project. Instead, the Research Director, Betsy, oversees all aspects of the project, including staff supervision and all data management activities. In addition, Betsy has been lead author on several publications about the project’s research findings.

Who owns the project and its data?

- The PI, Dr. Smith
- The Research Director, Betsy
- The University
- The National Institutes of Health
- No one person or organization
What are Some Ethical and Data Acquisition Issues?

- Various approvals may be needed prior to data collection (IRB, Institutional Animal Care and Use Committee, Land permits, etc).
- Dealing with unexpected data (outliers).
- Others need be able to reproduce/validate results, data may need to be available for this purpose (sharing).
- Proper data management takes a significant investment of time & effort.
Ethical and Data Acquisition Issues

- Must develop/use appropriate validated data collection methods
- Must have appropriate permissions (IRB, IACUC, Permits)
- Must protect data as it is collected
What are Some Data Management Concerns?

- If improperly managed, data may become invalid:
  - Human errors
  - Data transmission errors
  - Software malfunctions
  - Hardware malfunctions
  - Natural/structure disasters (fires, floods, etc)
- Data Storage: Original data stored safely; files backed up; samples saved so as not to degrade.
- Privacy and Confidentiality
- Retention: Generally must save data for 3 years after the completion of the project.
Data Management - Retention

Is there sufficient funding for your lab or research group for data curation?

There are many tales of early archaeologists burning wood from the ruins to make coffee. If we fail to curate the environmental archives we collect from nature at public expense, we essentially repeat those mistakes.

Science 11 February 2011:
Vol. 331 no. 6018 pp. 692-693
DOI: 10.1126/science.331.6018.692
INTRODUCTION: Challenges and Opportunities
Data Analysis Concerns

- How do you deal with outliers? If included in analysis may bias mean/variance, estimates, p-values, and conclusions. If removed should have good reason and describe when data is shared.
  - Check for recording error
  - “Rare” event syndrome (60 degree day in June in Las Vegas)
  - Transformation of data for analysis
  - Accommodation (nonparametric statistical methods)
- Statistical analyses
  - Proper analysis technique depends on type of data
  - Evaluation of sample size to ensure that it is valid
- Image manipulation
- Data selection
  - “Typical results”
INTRODUCTION: Challenges and Opportunities

Do you have the necessary expertise in your lab or group to analyze your data in the way you want?

26.5% YES
34.4% YES, through collaborators
23.0% NO
16.1% No special skills needed

“The next few years [particularly in medicine] the volume of data we need to analyze will expand exponentially.”

---

*Science* 11 February 2011:
*Vol. 331 no. 6018* pp. 692-693
*DOI: 10.1126/science.331.6018.692*
INTRODUCTION: Challenges and Opportunities
Case Vignette: Collecting Valid Data

Part of the data collection methodology for Dr. Smith's study includes distributing a 12-page self-administered questionnaire to participants; they must fill out and initial each page of the questionnaire to confirm completion.

One day on his way home from conducting an interview with a subject, the Research Assistant, Joel, needed to write directions for a friend and he reached in his bag and grabbed the first piece of paper that he could find. Joel accidentally ripped the back page off of one of the completed questionnaires to write the directions, which he then gave to his friend. He didn't realize this until a few hours later, when he was reviewing the data that he had collected that day.

Joel thought that he remembered the participant's answers on the last page of the survey, since they were mostly demographic questions.

What should Joel do?

☐ Staple on a new page and fill out the subject's responses, since he remembers them.
☐ Contact the subject and ask her to complete the last page of the questionnaire again.
☐ Omit the participant's questionnaire from the study, his/her partial data is invalid.
☐ Just pretend like he doesn't know what happened to the last page.
What are some Data Sharing Issues?

- Preliminary data usually should not be released.
- Confirmed or validated data - data should be confidential until accepted for publication.
- Published data - Once published there is an expectation that all the information about that experiment (including final data) should be freely available.
- MTA - Material Transfer Agreements (legal contracts) are often required. MTA’s are often required for sharing research materials also. Contact OSP for more information.
Data Sharing (from HHS)

Case Vignette: Data Sharing

After completing the first phase of data analysis, 1 of the 3 main hypotheses of Dr. Smith and the research team was proven correct. However, the team also found some results from another facet of the project that they were not expecting. While these secondary results do not directly impact Dr. Smith's primary research questions, they may affect at least 3 other investigators' research. The results appear to be pretty definitive, but data analysis is still being conducted on other parts of the project.

The 2 Research Associates working on the project, Samantha and Enrique, are insistent that the team should immediately publish their findings in a journal, since the results may have implications on other PIs' work. Dr. Smith and Betsy, the Research Director, do not intend to publish any results for at least another year, since the research is ongoing and some questions are still unanswered.

What should the research team do?

___ They should publish the results in a journal as soon as possible.
___ They should tell the funding agency about the findings, and let the agency disseminate the information if it wants.
___ They should contact the other researchers to let them know the preliminary results.
___ They should do nothing; they aren't legally allowed to share their results until all data have been fully validated.
Privacy and Confidentiality

• HIPAA – Health Insurance Portability and Accountability Act (1996)
  ○ Authorization prior to the use of a subject’s individually-identifiable health information for research purposes.
  ○ Specific security requirements for health data access and storage.

• As of May 6, 2004, NSHE, a hybrid covered entity, designated its health care components of UNLV as follows:
  - Dental School and any associated clinics
  - Student Health Center & Pharmacy and Laboratory, Counseling & Psychological Services (CAPS); and Faculty and Staff Treatment Center (FAST);
  - Athletic Training Department
  - Center for Individual & Family Counseling
  - Center for Health Information Analysis
  - National Supercomputing Institute
  - The PRACTICE (a community mental health training clinic)
  - Marriage and Family Therapy clinic
  - Center for Autism and Spectrum Disorders; & Academic Success Center’s Learning Specialist Program
Privacy and Confidentiality

- FDA, select agents, dual-use technologies (homeland security), patents, export controls (dissemination at foreign meeting), and classified research will affect these.

- State “sunshine” laws (e.g., open meeting) may grant access.

- Industry sponsored research - pharmaceutical trials

- Biobanking and genomic data
Case Vignette: Research Team Responsibilities

After collecting data for about a year, Dr. Smith's research team revisited their original research questions. They decided to investigate an additional hypothesis related to a new issue that arose during the study. This change required adding about a dozen new questions to the self-administered questionnaire.

One day, the Research Assistant, Joel, realized that they had been administering the revised survey to subjects, but the Institutional Review Board (IRB) had not yet approved the changes.

Whose responsibility was it to make sure that data collection did not continue until the IRB approved the changes?

- The PI, Dr. Smith
- The Research Director, Betsy
- The Research Associates, Samantha and Enrique
- The Research Assistant, Joel
Intellectual Property

Who owns it?

- Traditionally it was the investigator
- More recent - the grantee (institution)
  - NIH and NSF both state the property belongs to the grantee but that they retain access.
  - State may actually retain possession if it’s a state school
    - Not in Nevada - given to Regents who in turn have given the intellectual property to the university.
- Bayh Dole Act (or Patent and Trademark Law Amendments) Universities are given the right to retain title to inventions funded with federal $$ and right to control licensing – must share w/ inventor
- Rights of subjects to their “data”
Sources

- Office of Research Integrity: [http://ori.hhs.gov/](http://ori.hhs.gov/)
- University of Idaho, Office of Research Assurances
- NIH Data Sharing Plan Policy
  (http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm)