Acquisition, Management, Sharing, and Ownership of Data

Responsible Conduct of Research Training

FRANK VAN BREUKELEN
UNIVERSITY OF NEVADA LAS VEGAS

Why are we here discussing data?

Data practices are central to all discussions of 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

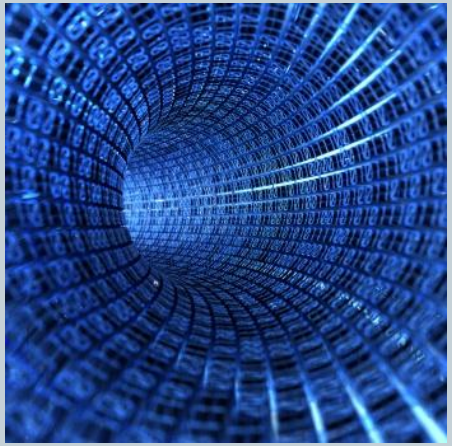
Entire Special Collection in *Science* called "Dealing with Data" (11 February 2011, *Vol. 331 no. 6018* : http://www.sciencemag.org/site/special/data/)

- 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.

What are data?

No universally-accepted federal definition of 'research or scientific data.'

- NIH Grants Policy Statement defines "data" as "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, work flow 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, *Uniform Administrative Requirements for Grants and Agreements With Institutions of Higher Education, Hospitals, and Other Non-Profit Organizations*, 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.**"

OMB definition applies across federal agencies

- 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.”

Federal requirements: Managing and Sharing Data

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.

Federal requirements: Managing and Sharing Data

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.

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.

Ethical Issues

- Various approvals may be needed prior to data collection (IRB, IACUC, Land permits, etc).
- Dealing with unexpected data (outliers).
- Scientific research requires replicability, data may need to be available for this purpose (sharing).
- Proper data management takes a significant investment of time & effort.

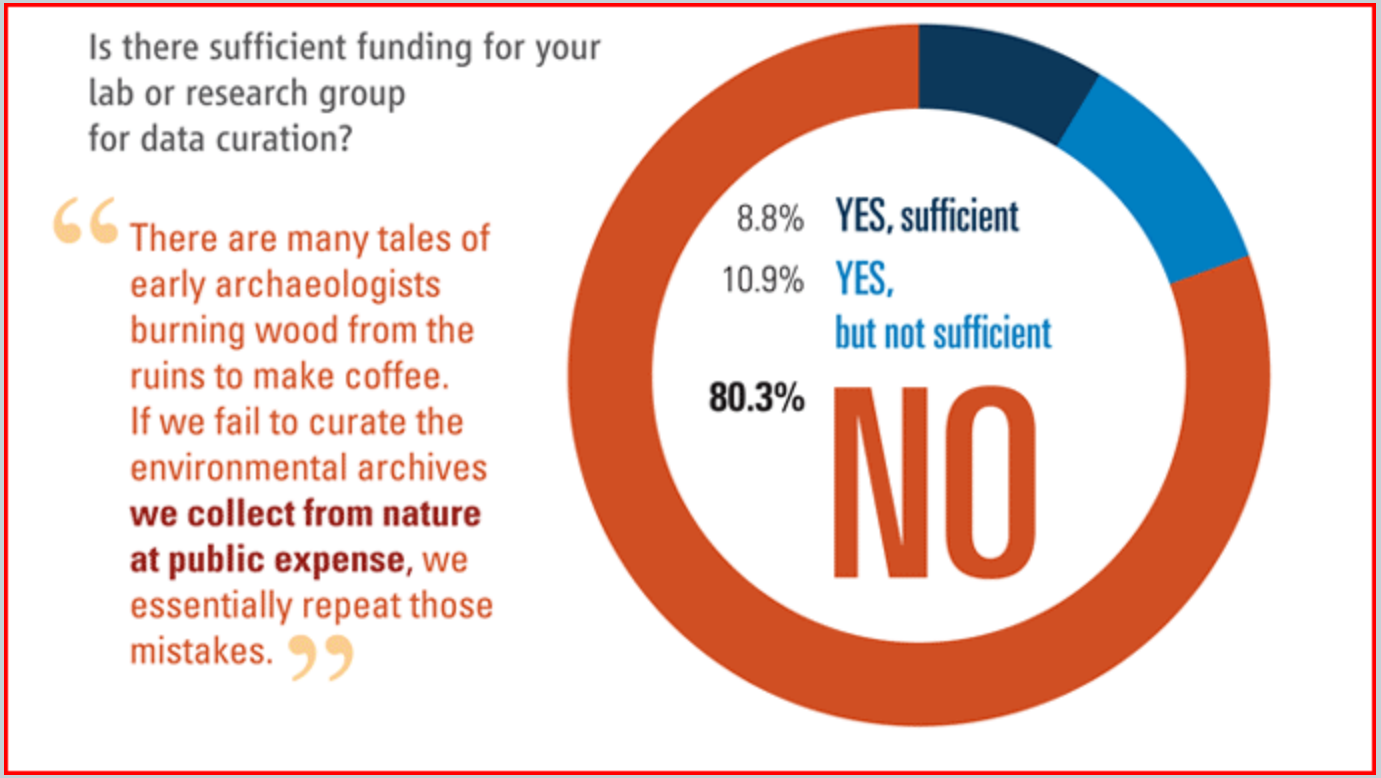
Data Acquisition Concerns

- Must develop/use appropriate validated data collection methods
- Must have appropriate permissions (IRB, IACUC, Permits)
- Must protect data as it is collected

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

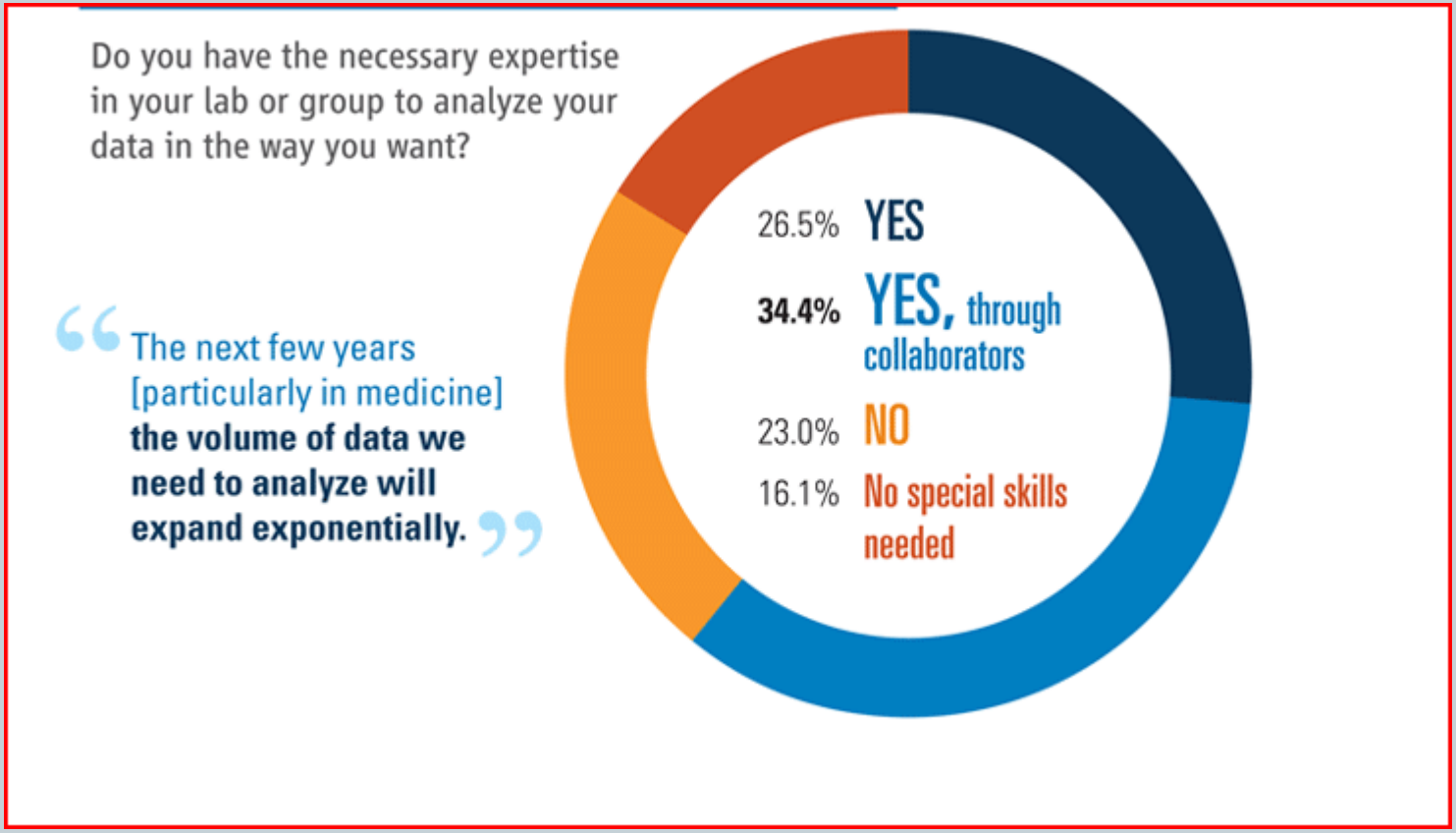


Science *11 February 2011:*
Vol. 331 no. 6018 *pp. 692-693*
DOI: 10.1126/science.331.6018.692
INTRODUCTION: Challenges and Opportunities

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 validity of effect
- Image manipulation
- Data selection
  - "Typical results"

Data Analysis



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

Accuracy in Reporting Data

- "Cleaning" images
- Journal considerations
  - IACUC, IRB, electronic image integrity issues, etc.
- Ethical publication and presentation of data
  - Data belong to institution and so some institutions will insist on oversight of publication including copyright assignment
  - Industry-sponsored research
- Investigators should review copyright assignments
  - life of author plus 70 years
- These assignments generally give the publisher all rights to the article or manuscript – not the data – which will limit the author's ability to use the publication in future works.

Data Sharing

- Preliminary data usually should not be released.
- Confirmed or validated data - keeping data 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 are often required. MTA's are often required for sharing research materials also. Contact OSP for more information.

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
  - Student Wellness
  - Center for Individual & Family Counseling
  - Center for Health Information Analysis
  - National Supercomputing Center for Energy and the Environment

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 may grant access
- Industry sponsored research- pharmaceutical trials
- Biobanking and genomic data

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 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"

Specific example of data misconduct

Scott J. Brodie, DVM, PhD, University of Washington

- 15 findings of research misconduct by the NIH
- He "knowingly and intentionally" fabricated and falsified data on 9 PHS grant applications and progress reposts, several published papers, manuscripts and PowerPoint presentation related to AIDS research.
  - Falsified figures
  - Falsified images
  - Deliberately mislabeled images
- He has been debared for 7 years from contracting or subcontracting with any agency of the US government and from involvement in any "covered transactions"
- He has also been prohibited for 7 years from serving in any advisory capacity to the PHS.
- His name and details of the case are public record.
- He is an author on approximately 90 articles.

Case Studies

- Leaving a laboratory- who owns the samples, the data, the rights to publication?
- The next step?
- Tenure denials?
  - Institution is responsible for the data

Case Studies

Privacy- leak of data
- Grad student leaks info to press about a politician (HIPAA)
- Pharmaceutical studies
- Sunshine laws- access to data for patent

Sharing data
- Hoppe story
- Protocols

Appropriate measures

Accuracy in reporting data
- sloppy notebooks
- ‘typically’

Sources

- Office of Research Integrity: http://ori.hhs.gov/
- National Institutes of Health. *Research Conduct and Ethics Instruction Materials:* http://www1.od.nih.gov/oir/sourcebook/ResEthicsCases/cases-toc.htm
- University of Idaho, Office of Research Assurances
- NIH Data Sharing Plan Policy
- (http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm
- NIH Data Sharing Workbook (pdf and MS word samples):
- http://grants.nih.gov/grants/policy/data_sharing/
- NSF Overview of the Dissemination and Sharing of Research Results (including Directoratelevel
- guidance): http://www.nsf.gov/bfa/dias/policy/dmp.jsp
- NSF Data Sharing Policy:
- http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp#VID4
- NSF Data Management Plan Requirements:
- http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp
- NSF Data Management Plan “Frequently Asked Questions”:
- http://www.nsf.gov/bfa/dias/policy/dmpfaqs.jsp
- MIT Library Guide to Data Management Planning:
- http://libraries.mit.edu/guides/subjects/data-management/