Ethics and "Omics"

Jeffrey Engler, Ph.D.
Dept of Biochemistry and Molecular Genetics
Associate Dean, UAB Graduate School

UAB GRADUATE SCHOOL

Office of Research Integrity/DHHS
http://ori.hhs.gov/education/products/clinicaltools/data.pdf

What is Scientific Misconduct?

ORI Definition (42 CFR Parts 50 & 93):

- "Fabrication, falsification, or plagiarism in proposing, performing, or reviewing research, or in reporting research results." (also known as FFP)

ORI research integrity

UAB Definition:

- "…fabrication, falsification, plagiarism, or other practices which seriously deviate from those that are commonly accepted within the scientific community for proposing, conducting, or reporting research

THE UNIVERSITY OF ALABAMA BIRMINGHAM

Slide adapted from a slide by Dr. Charles Prince and Mr. Joe Roberson, 2009

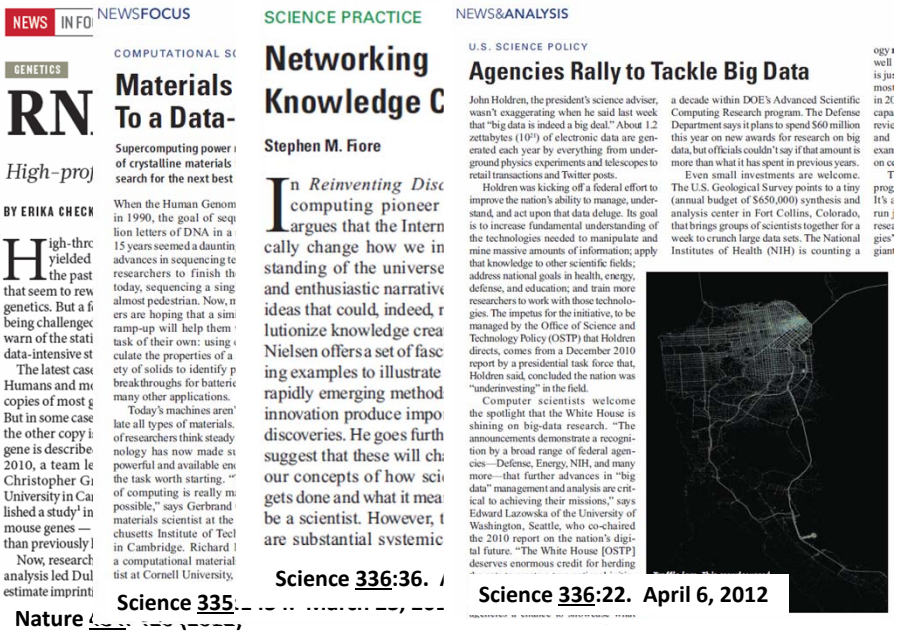
Nine Areas of Responsible Conduct

Responsible Conduct of Research (RCR)

ORI supports several programs designed to promote education and training in the responsible conduct of research (RCR) that covers the following nine instructional areas:

- Data Acquisition, Management, Sharing and Ownership **a.k.a. "Data Management"**
- Conflict of Interest and Commitment
- Human Subjects
- Animal Welfare
- Research Misconduct
- Publication Practices and Responsible Authorship
- Mentor / Trainee Responsibilities
- Peer Review
- Collaborative Science

Reference: http://ori.hhs.gov/education/. Accessed 6/13/11


NEWS IN FO NEWSFOCUS

COMPUTATIONAL SC

GENETICS

RN

Materials To a Data-

Supercomputing power of crystalline materials search for the next best

High-proj

BY ERIKA CHECK

When the Human Genom in 1990, the goal of seq lion letters of DNA in a 15 years seemed a dauntin advances in sequencing te researchers to finish th today, sequencing a sing almost pedestrian. Now, n ers are hoping that a sim ramp-up will help them task of their own: using culate the properties of a ety of solids to identify p breakthroughs for batteri many other applications.

Today's machines aren late all types of materials. of researchers think steady nology has now made st powerful and available en the task worth starting. " of computing is really m possible," says Gerbrand materials scientist at the chusetts Institute of Tech in Cambridge. Richard a computational material tist at Cornell University,

High-thro yielded the past that seem to rew genetics. But a fi being challenge warn of the stati data-intensive st

The latest case Humans and m copies of most g But in some case the other copy i gene is describe 2010, a team le Christopher Gi University in Ca lished a study[1] in mouse genes — than previously

Now, research analysis led Dul estimate imprint

Nature

SCIENCE PRACTICE

Networking Knowledge C

Stephen M. Fiore

In *Reinventing Disc* computing pioneer argues that the Intern cally change how we in standing of the universe and enthusiastic narrative ideas that could, indeed, r lutionize knowledge crea Nielsen offers a set of fasc ing examples to illustrate rapidly emerging method innovation produce impo discoveries. He goes furth suggest that these will ch our concepts of how sci gets done and what it mea be a scientist. However, t are substantial systemic

Science 336:36.

Science 335

NEWS&ANALYSIS

U.S. SCIENCE POLICY

Agencies Rally to Tackle Big Data

John Holdren, the president's science adviser, wasn't exaggerating when he said last week that "big data is indeed a big deal." About 1.2 zettabytes (10^{21}) of electronic data are generated each year by everything from underground physics experiments and telescopes to retail transactions and Twitter posts.

Holdren was kicking off a federal effort to improve the nation's ability to manage, understand, and act upon that data deluge. Its goal is to increase fundamental understanding of the technologies needed to manipulate and mine massive amounts of information; apply that knowledge to other scientific fields; address national goals in health, energy, defense, and education; and train more researchers to work with those technologies. The impetus for the initiative, to be managed by the Office of Science and Technology Policy (OSTP) that Holdren directs, comes from a December 2010 report by a presidential task force that, Holdren said, concluded the nation was "underinvesting" in the field.

Computer scientists welcome the spotlight that the White House is shining on big-data research. "The announcements demonstrate a recognition by a broad range of federal agencies—Defense, Energy, NIH, and many more—that further advances in "big data" management and analysis are critical to achieving their missions," says Edward Lazowska of the University of Washington, Seattle, who co-chaired the 2010 report on the nation's digital future. "The White House [OSTP] deserves enormous credit for herding

a decade within DOE's Advanced Scientific Computing Research program. The Defense Department says it plans to spend $60 million this year on new awards for research on big data, but officials couldn't say if that amount is more than what it has spent in previous years.

Even small investments are welcome. The U.S. Geological Survey points to a tiny (annual budget of $650,000) synthesis and analysis center in Fort Collins, Colorado, that brings groups of scientists together for a week to crunch large data sets. The National Institutes of Health (NIH) is counting a

Science 336:22. April 6, 2012

What is Data?

True or False?

In research, only the information and observations that are made as part of the inquiry are considered data.

What is Data?

T/F?: In research, only the information and observations that are made as part of the inquiry are considered data

False!

Data also includes anything related to understanding the data generated by the project

- **Samples collected, survey instruments, cell lines, informed consent documents, procedures, products generated, online content.**

Key Concepts of Data Management

- Data Ownership
  - Who owns it?
- Data Collection
  - Systemic and reliable
- Data Storage
  - What should be retained?
- Data Protection
  - Safe storage
  - Prevent tampering
- Data Retention
  - How long to keep original data?
- Data Analysis
  - Rubric for analysis and interpretation
- Data Sharing
  - Dissemination plan
- Data Reporting
  - Publication and authorship

Adapted from Steneck, N.H. (2007), Introduction to the responsible conduct of research. http://ori.dhhs.gov/documents/rcrintro.pdf. Accessed 6/11/11.

Who "Owns" Your Data?

- The Principal Investigator?
- All members of the research team?
- The research subjects?
- Your home institution?
- The sponsor?
  - Grants ("assistance funding")
  - Contracts ("procurement funding")

Data Collection

What is the role of Data Collection in completing successful research?

1. Ensuring the validity of data is key
2. Ensuring reliability is key
3. Ensuring both validity reliability is key
4. It doesn't matter.

Data Collection – Upholds the Integrity of the Project

- **Details the rationale for the project and its design**
- **Yields reliable and valid results**
- **Allow accurate analysis and assessment**
- **Allows others to replicate the process and evaluate the results**
- **Provides justification to sponsors for costs and expenditures.**

Planning for Data Collection

A well thought out plan will help assure that all members of the project team collect data consistently.

Questions that should be addressed:

- **Purpose of the research project**
- **Rationale for methodologies chosen**
- **Implementation of methodologies**
- **What worked and failed to work**
- **How data will be collected or analyzed**
- **How to report unexpected findings or errors**
- **Implications of the research and future directions.**

Data Storage

- **What data to retain?**
  - **Everything necessary to reconstruct the findings.**
- **How long to retain it?**
  - **3 years beyond the end of funding for the project.**
  - **Longer if there are patents to be filed.**
  - **Can establish precedence of the work**

Data Protection

In the following list, which is the most effective way to protect project data?

1. Strip identifiers from human subjects data
2. Limit who has access
3. Use an encrypted password system
4. Destroy written data after transferring it to an electronic database

Review: Key Concepts of Data Management

- **Data Ownership**
  - **Who owns it?**
- **Data Collection**
  - **Systemic and reliable**
- **Data Storage**
  - **What should be retained?**
- **Data Protection**
  - **Safe storage**
  - **Prevent tampering**
- **Data Retention**
  - **How long to keep original data?**
- **Data Analysis**
  - **Rubric for analysis and interpretation**
- **Data Sharing**
  - **Dissemination plan**
- **Data Reporting**
  - **Publication and authorship**

Adapted from Steneck, N.H. (2007), Introduction to the responsible conduct of research. http://ori.dhhs.gov/documents/rcrintro.pdf. Accessed 6/11/11.

Data Retention

How minimum length of time does NIH/NSF require awardees to retain the raw data for a funded project after it ends?

1. **1 year**
2. **2 years**
3. **3 years**
4. **5 years**
5. **Until the PI leaves the institution.**

Data Retention

There is no set amount of time for which data must be stored.

- **NSF and NIH = minimum 3 years after the end of funding**
- **If there are intellectual property considerations, data may retained longer.**

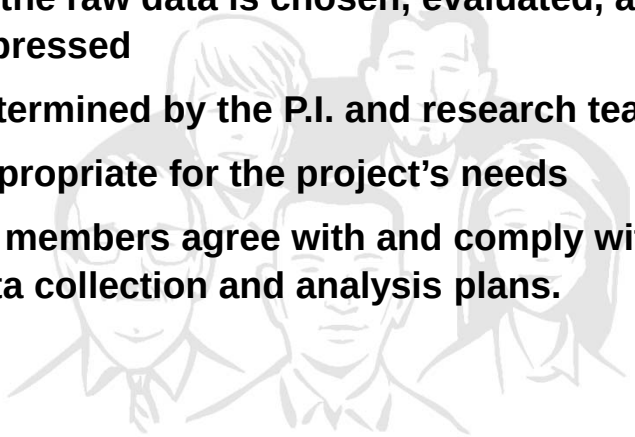
Weigh the long term benefits and risks of retention.

Data destruction should be thorough and complete.

Data Analysis

How the raw data is chosen, evaluated, and expressed

- **Determined by the P.I. and research team**
- **Appropriate for the project's needs**
- **All members agree with and comply with the data collection and analysis plans.**



Data Analysis

Questions that the P.I. should consider:

- **What are the accepted standards of the field?**
- **What data should be included?**
- **Include or exclude outliers?**
- **Dealing with missing or incomplete data**
- **Responsible conduct: falsifying, fabricating data**

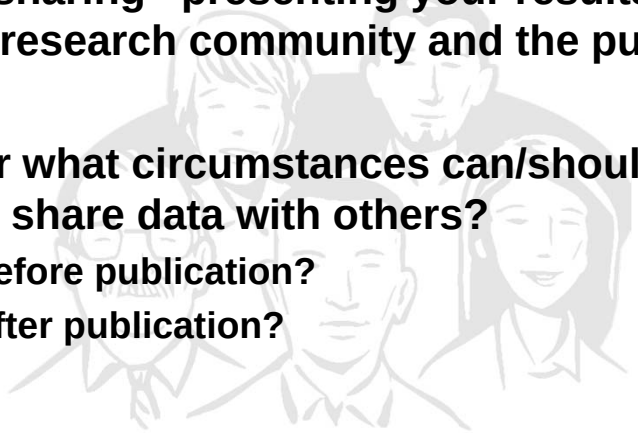
Communicating expectations to the research team is critical!

Data Sharing and Reporting

Data sharing –presenting your results to the research community and the public

Under what circumstances can/should you share data with others?

- **Before publication?**
- **After publication?**

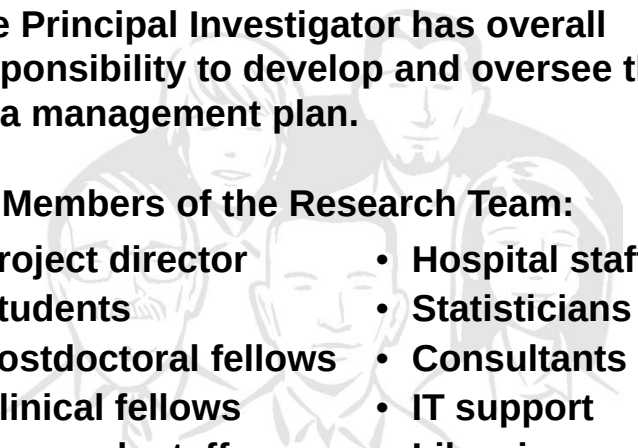


Who is Responsible for Data Management?

The Principal Investigator has overall responsibility to develop and oversee the data management plan.

All Members of the Research Team:

- **Project director**
- **Students**
- **Postdoctoral fellows**
- **Clinical fellows**
- **Research staff**
- **Hospital staff**
- **Statisticians**
- **Consultants**
- **IT support**
- **Librarians**



For More Information

Reference: http://ori.hhs.gov/general-resources-0. Accessed 3/7/12

Thank you!

UAB GRADUATE SCHOOL

Office of Research Integrity/DHHS
http://ori.hhs.gov/education/products/clinicaltools/data.pdf