Challenges of Coping with Funding and Data Management in a Changing World

Rick Lyons
Director Infectious Disease Research Center

Colorado State University

BIG DATA-
A collection of digital data with sufficient volume that, when viewed in aggregate does not fit on one computer screen and has a bunch of columns

NIH Requirements

While NIH does not require a data management plan as the NSF does, for most grants over $500,000, a **data sharing plan** must be included in applications and incorporated as a part of the award. Final research data "should be made as widely and freely available as possible while safeguarding the privacy of participants, and protecting confidential and proprietary data."

Possible Components of a Data Sharing Plan

1. Describe briefly the expected schedule for data sharing

2. Describe the format of the final dataset

3. Describe the documentation to be provided

4. Identify whether or not any analytic tools also will be provided

5. Identify whether or not a data-sharing agreement will be required and, if so, a brief description of such an agreement (including the criteria for deciding who can receive the data and whether or not any conditions will be placed on their use)

6. Describe the mode of data sharing (See list below.)

The NIH outlines several ways to share data to satisfy this requirement:

1. Under the auspices of the prinicpal investigator.

2. Saving the data on a data archive (A place where machine-readable data are acquired, manipulated, documented, and finally distributed to the scientific community for further analysis.)

3. Saving the data on a data enclave (A controlled, secure environment in which eligible researchers can perform analyses using restricted data resources.)

4. A combination of these ways

Federal sponsor's goals

- BD2 K
  - Appropriate access to shareable biomedical data through technologies, approaches, and policies that enable and facilitate widespread data sharing, discoverability, management, curation, and meaningful re-use; - See more at: http://bd2k.nih.gov/about_bd2k.html#sthash.0dfzGoHg.dpuf
- Strategies
  - Develop a catalog of data received from funded projects
  - Develop "standardized data sets for testing and developing new technologies

Landscape of data management in sponsor's eyes

Local lab impact

Collaborations

New Funding

New Interpretation / Information from old data

Archiving/ Compliance

Local lab Impact

•Historically maintained in lab notebooks /paper occasionally picture files . Based on personal experience not well catalogued or tracked

•Technical advances in tools commonly lead to giga- and tera byte level data collections within individuals labs. Storing with integrity and seamlessly retrieving data critical for efficient lab

•Development of this infrastructure allowing for state of the art approaches required for CSU competitiveness for funding

Sponsors pushing for Collaborative Activities

- Application of big data to problems requires integrated approaches from different disciplines and many large contracts from the federal government require multiple institutions and big data.

- Multiple eyes must be able to see and handle the same data in a facile iterative fashion for information generation as well as for publication

New Funding

- Improvements in infrastructure are needed to stay in the research game. This type of infrastructure is becoming as important as bricks and mortar.

OBAMA ADMINISTRATION UNVEILS "BIG DATA" INITIATIVE: ANNOUNCES $200 MILLION IN NEW R&D INVESTMENTS

New Funding

NSF Core Techniques and technologies for Advancing Big Data Science & Engineering

DoD Data to Decisions

DOE Scientific Advances Through Advanced Computing

NIH Big data to Knowledge (BD2K)
1000 genomes Project Data

US Geological Survey Big data for Earth Science

Generating New Information from Old data

- We cannot anticipate when the next "transformational" advances will be made in analytical tools . Preserving data will allow us to go back to well archived data and reanalyze

Archiving /Compliance

- **8.2 Availability of Research Results: Publications, Intellectual Property Rights, and Sharing Research Resources**
  - **Continued increase in culture of sharing from Fed sponsors for funded work**

Colorado State University

Archiving / Compliance

US Federal Funding Agency	Policy and Guideline Status	Additional Information
Department of Energy	DOE's CIO has primary responsibility to ensure that Information Technology (IT) is acquired and information resources are managed in a manner consistent with statutory, regulatory, and Departmental requirements and priorities. With this responsibility, the CIO provides information resources management advice and assistance to the Secretary of Energy and to other senior managers.	DOE Policies Standard Research Terms and Conditions ARM Data Sharing and Distribution Policy Developing Data Management Policy and Guidance Documents for your NARSTO Program
Environmental Protection Agency (EPA)	EPA does have an official data management and sharing policy.	Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by the Environmental Protection Agency Policy, Regulation and Guidance Survey of EPA and Other Federal Agency Scientific Data Management Policies and Guidance 2010