Data Management and Sharing

January 19, 2023

Kacey Strickland and
Nicole Cobb, Research Compliance & Security

Dr. Deborah Lee, MSU Libraries

2023 - Year of Open Science

The White House Office of Science and Technology Policy launches Year of Open Science to advance national open science policies across the federal government in 2023.

Press release: https://www.whitehouse.gov/ostp/news-updates/2023/01/11/fact-sheet-biden-harris-administration-announces-new-actions-to-advance-open-and-equitable-research/

Commitment Statements

- "NASA commits to the full and open sharing of Earth science data obtained from NASA Earth observing satellites, sub-orbital platforms and field campaigns with all users as soon as such data become available."
  -- NASA Data & Information Policy

- "CDC believes that public health and scientific advancement are best served when data are released to, or shared with, other public health agencies, academic researchers, and appropriate private researchers in an open, timely, and appropriate way. The interests of the public—which include timely releases of data for further analysis—transcends whatever claim scientists may believe they have to ownership of data acquired or generated using federal funds."
  -- CDC/ATSDR Policy on Releasing and Sharing Data (pdf)

- "Data should be made as widely and freely available as possible while safeguarding the privacy of participants, and protecting confidential and proprietary data"
  -- National Institutes of Health Data Sharing Policy

Key Issues to Consider

- Maximize access
- Research security and data sharing
- Preserve IP rights
- Cost of data management
- Compliance

Maximize access

- Incomplete, incorrect, or messy data limit use and reuse
- Repositories 'curate' or enhance data to make it more usable for future researchers
  - Add descriptive labels
  - Correct coding errors
  - Gather documentation
  - Standardize final versions of files
- Finding data in repositories requires descriptive metadata (data about data)

What is Metadata?

- Metadata describe a dataset:
  - How they were collected
  - When they were collected
  - What assumptions were made in methodology
  - Geographic scope
  - How multiple files relate to one another
  - Definitions of variables
  - Calibration of equipment used in collection
  - Version of software used for analysis
- Metadata should be documented for research outputs of any kind (computer file to piece of art)

Metadata make data FAIR

- Findable – most searches are done using text (like a Google search)

- Accessible – once found, how can it be accessed (authentication and authorization)

- Interoperable – enables a data set to be integrated with other data

- Re-usable – if not well-described then future use is limited

Learn more about making data FAIR at **https://www.go-fair.org/fair-principles/**.

How to Share Data Securely

- The major benefits of sharing data also brings risks to security and privacy
- Cyber attacks on labs highlight the cost and impact
- Data classification is important (Export Controlled, PHI, CUI, FCI)
- Researchers and IT experts should jointly develop data security protocols
- Even Fundamental Research should be secured from hackers and ransomware

Preserve IP Rights

- Original research can be commercially valuable and proprietary
- To manage these interests:
  - Tailor copyright and patent licenses
  - Use an embargo period or delay dissemination
- Ultimately, all interests should be considered with an eye toward public release

Cost of Data Management

- Data management services carry real costs.

- Estimating and planning for these costs ensures long-term preservation and access.

- Appropriate costs for documenting, preparing, publishing, disseminating, and sharing research findings are allowable charges against the grant (NSF Grant Proposal Guide).

Compliance with DMSP

- Updates on data management and sharing should be included in the annual progress reports.
- If plans change, work proactively with your Program Officer to obtain review and approval of modifications.
- PIs are ultimately responsible.
- Failure to comply may result in termination of the award and factored into future funding decisions.