NIH Policies and Activities: Increasing Access to Digital Scientific Data

Jennie Larkin, PhD
Senior Advisor
Office of the Associate Director for Data Science (ADDS)

▶ BRDI Open Session ▶ October 27, 2016 ▶

FAIR Principles

- NIH is adopting principles to make data "FAIR" (Findable, Accessible, Interoperable, and Reusable; http://www.nature.com/articles/sdata201618)

- Increased access to digital scientific data is expected to accelerate scientific discovery and progress.

- Effective data sharing relies upon:
  - Identification, adoption, and crediting of good data management and sharing practices
  - Recognition of effective data sharing and re-use of high quality data which can be achieved through the citation of such data.

NIH Considerations for Implementation of the NIH Public Access Plan (Data)

- **NIH Scientific Data Council (Internal)**
  - Created in 2013 as part of a multifaceted expansion of NIH programs and governance related to data science
  - Addresses the growing challenges and opportunities associated with 'big data' and data science in biomedical research
  - Identified need to assess value, benefits, and costs of sharing different types of data (i.e., priorities for data sharing)

- **Changes to policies, systems, forms, procedures.** *For example:*
  - Requirement for a DMP in funding application or proposal
  - Data Management and Sharing Plan (DMP) form (machine-readable)
  - Evaluation of DMPs during Peer Review
  - Timelines for policy implementation

- **Other:**
  - For human data: privacy, confidentiality, informed consent issues
  - Costs and benefits, e.g., establishing suitable repositories
  - Implementation costs within constraints of existing budgets and resources

Current NIH Activities

- NIH Data Commons
- NIH Director's Task Force on Data Management and Stewardship
- Biomedical Informatics Community (BMIC) Coordination for Common Data Elements
- BD2K Initiative
  - Resource Indexing (e.g., BioCADDIE, Data Discovery Index)
  - Funding Opportunities (e.g., curation, community-based standards, technologies)
- Global Alliance for Genomics and Health
- Precision Medicine Initiative (PMI)
- Cancer Moonshot

Next Steps for NIH

- **Release NIH Requests for Information**:
  - *Metrics to Assess Value of Biomedical Digital Data Repositories* – Comments due October 17, 2016 (NOT-OD-16-133)
  - *Including Pre-Prints and Interim Research Products in NIH Applications and Reports* – Comments due November 29, 2016 (NOT-OD-17-006)
  - Future RFI to Assess Priorities for Data Management and Sharing
    - Data sharing strategy development
    - Data and software citation
    - Will inform development of a new NIH policy

- **BD2K Initiative – Future Funding Opportunities**

- **Issue draft policy for public comment**
  - Outreach and communication
  - Guidance and implementation for systems, forms, procedures, etc.
  - Identify and address challenges as appropriate

- **Issue final policy and implementation guidance, along with education, training, tools, etc.**

Priorities and Needs Regarding Research Data *(wish list)*

- Easier and better (yet secure) access to data and compute to support science
  - More facile access, but with security, citation, and provenance to support the digital ecosystem.

- Improved solutions to support data/software archiving
  - What? Where? How?

- Improved solutions to making digital resources (archived or in the Commons) FAIR.
  - Find it in the cloud, or find it in an archive.

Data Science at NIH

▶ https://datascience.nih.gov

▶ bd2k@nih.gov

▶ @NIH_BD2K

▶ #BD2K

Jennie.Larkin@nih.gov

National Institutes of Health
Office of the Director
Data Science at NIH