HOW NCI IS LEADING ON DATA SHARING

Jaime Guidry Auvil, the Director of NCI's Office of Data Sharing
Webinar Summary, October 9, 2018
Slide and Recording can be found here.

The National Cancer Institute's (NCI) Office of Data Sharing (ODS), is in the Center for Biomedical Informatics and Information Technology (CBIIT). ODS is charged with creating a comprehensive data sharing vision and strategy for NCI and the cancer research community.

ODS Email: nciofficeofdatasharing@nih.gov
ODS must balance:

- Open-access and broad data sharing policies to enable reproducibility, secondary use, and knowledge sharing with
- Critical intellectual property concerns for individuals and organizations to support a healthy commercial marketplace

ODS Activities:

- Advises on ethical data access and sharing issues, policies, and practices
- Enhances the accessibility and utility of research data and metadata, in part by refining data and metadata standards (prevent data silos)!
- Develops sustainable, achievable, and meaningful incentives for data sharing
- Supports equitable sharing through a robust, sustainable data management ecosystem
- Encourages participation in major data-sharing initiatives, including contributing to NIH-supported data repositories
- Creates educational resources for the cancer community on the importance and processes of sharing data

TAKE HOME MESSAGE: IN DEVELOPING AND ENHANCING YOUR POLICY, IT'S IMPORTANT TO ANSWER THESE QUESTIONS:

- Do you allow your recipients to use funds to cover data-sharing costs? If yes, how much? is there a limit? Do you grant extra money on top of the grant?
- Do you check on compliance? What happens if they do not comply?
- What challenges did you face when developing and implementing the policy? How did you address those?
- What guidance can you give nonprofit nongovernmental funders (with fewer resources!) about how to develop a policy that works for "their" organization. Then how to market that to your scientists – or is mandating it enough?
- How does the community pay (funders pay, the scientists themselves??) for the myriad of costs incurred during the process of sharing data.

TALK GOAL 1: DEFINE "DATA SHARING" AND ESTABLISH THE IMPORTANCE OF RESPONSIBLE AND BROAD DATA SHARING TO ADVANCE DISEASE KNOWLEDGE AND IMPROVE CARE

Data Sharing Definition: The practice of making data & metadata used for scholarly research available to other investigators. Ability to replicate is key. Transparency and openness facilitate.

Benefits to Sharing Data:

- Data generated from one study can be used to explore a wide range of additional research questions
- Combining data from multiple studies Increases statistical power and scientific value
- Facilitates reproducibility and validation of research results
- Facilitates innovation of methods and tools for research
- Reduces duplication and saves time, valuable resources & experimental costs

Public Wants Broad Data Sharing: Evidence from a study in NJEM by Mello et al, 2018.

- Positives: Scientists were likely to allow their own data to be shared with university scientists and scientists in for-profit companies. Few felt that the potential negative consequences of data sharing outweighed the benefits. Potential for "great benefit" for improved healthcare.
- Concerns: Perception that there could be decreased willingness to enroll in clinical trials, and data would be used for marketing purposes or stolen.

TALK GOAL 2: OUTLINE CURRENT NIH DATA SHARING POLICIES: *Including Intramural, Clinical Trial, and Genomic Data Sharing Policies, and procedures for the database of Genotypes and Phenotypes (dbGaP)*

NIH's Data Sharing "View"

- As part of NIH's long-standing policy to share and make available to the public the results and accomplishments of the activities that it funds, NIH reaffirms its support for the concept of data sharing.
- We believe that data sharing is essential for expedited translation of research results into knowledge, products, and procedures to improve human health.
- The NIH endorses the sharing of final research data to serve these and other important scientific goals.
- The NIH expects and supports the timely release (by the time of publication) and sharing of final research data from NIH-supported studies for use by other researchers.
- Investigators submitting an NIH application seeking **$500,000*** or more in direct costs in any single year are expected to include a plan for data sharing or state why data sharing is not possible.

For HRA discussion:** ***Why limit this requirement to grants of $500,000 or more? In responding to NIH's RFI, we should note that the new NIH guidelines should not have this bar.

Setting NIH Policy Guidelines: NIH OSP focuses on 2 areas of Scientific Data Sharing:

1. Genomics and Health: Analyzes the scientific, ethical, and social implications of genetic and genomic research on health. Provides policy recommendations. Engages in trans-NIH and inter-agency collaborations to advance implementation of genomic medicine.

- NIH Genomic Data Sharing
- HeLa Cell Whole Genome Sequence Data Sharing
- Genomic Medicine
- Genetic Testing Registry

2. Scientific Data Management: Evaluates opportunities and challenges regarding sharing scientific research data. Provides policy recommendations associated with data sharing and management (ie. data science, public access, open science, big data).

- NIH Data Management and Sharing Activities Related to Public Access and Open Science
- NIH Data Science Policy Council
- Interagency and International Open Science Efforts

POLICY 1: NIH *INTRAMURAL* HUMAN DATA SHARING POLICY (08/2015)

- *The NIH's mission is to improve the health of the public through support of biomedical research and the training of biomedical scientists. To further advance and accelerate research to benefit the public health, data developed in the NIH Intramural Research Program (IRP) should be collected in a manner that permits and promotes the broadest sharing possible.*
- *Data sharing may be complicated or limited, in certain cases, by agreements with outside collaborators, e.g., Clinical Research and Development Agreements (CRADAs) or Clinical Trial Agreements, by Institutional Review Board (IRB) rules or by other constraints. NIH IRP investigators should share data broadly for secondary research purposes, in all cases consistent with applicable laws, regulations and policies.*

Key Points of Policy:

- Applies to ALL human data in the NIH Intramural Research Program (NIH Clinical Center & Institutes/Centers).
- A Data Sharing Plan must be developed for any research involving human data and will be included in the IC scientific review.
- Clinical investigators are expected to develop protocols and consent processes/forms to enable broad data sharing for secondary research consistent with this Policy.
- Sharing data for secondary research purposes shall comply with human subjects research regulations and procedures, if applicable.
- All IRP investigators are encouraged to deposit data in publicly accessible research repositories for sharing to the extent feasible and appropriate.

POLICY 2: NIH-FUNDED CLINICAL TRIAL INFORMATION – DISSEMINATION POLICY (01/2017)

The National Institutes of Health (NIH) is issuing this policy to promote broad and responsible dissemination of information from NIH-funded clinical trials through ClinicalTrials.gov. The policy establishes the expectation that all investigators conducting clinical trials funded in whole or in part by the NIH will ensure that these trials are registered at ClinicalTrials.gov, and that results information of these trials is submitted to ClinicalTrials.gov.

Key Points of Policy:

- Applies to all intramural and extramural clinical trials funded wholly or partially by NIH of FDA-regulated drug, biological, and device products and pediatric post-market surveillance studies of devices required by the FDA under the FD&C Act. (Does not apply to phase 1 trials or small feasibility device studies).
- Trials, including data elements, must be registered on ClinicalTrials.gov no later than 21 days after enrollment of the first participant.
- Results information is to be submitted to ClinicalTrials.gov no later than 12 months after primary completion date; possible delay of up to an additional 2 years for trials of unapproved products or of products for which initial FDA marketing approval or clearance is being sought, or approval or clearance of a new use is being sought.
- For federally funded trials, grant funding can be withheld if required reporting cannot be verified. Civil monetary penalties of up to $10,000/day (amount to be adjusted going forward)/ May lead to suspension or termination of grant or contract funding/ Can be considered in future funding decisions.

POLICY 3: GUIDING PRINCIPLE OF THE NIH GENOMIC DATA SHARING (GDS) POLICY

The greatest public benefit will be realized if largescale genomic data are made available in a timely manner to the largest possible number of investigators. For human data, data are made available under terms and conditions consistent with the informed consent provided by individual participants.

Benefits/Rationale for GDS:

- Enhance scientific progress and accelerate translation of genomic research into treatments, products and procedures that benefit public health
- Examine relationships between genomic data and phenotypes while
- respecting rights of research participants
- Unshared Information represents lost opportunity to improve public health
- Encourage data access and sharing unencumbered by intellectual property claims (discourage premature claims on pre-competitive information)
- Increased availability of data to a wide range of secondary data users not engaged in human subjects research

Key Points of Policy:

Purpose

- Sets forth expectations and responsibilities for investigators and their institutes that ensure the broad and responsible sharing of genomic research data in a timely manner (within 6 months of clean, QC'ed data)

Scope

- All NIH-funded research generating large-scale human or non-human genomic data and the use of these data for subsequent research
- Applies to all funding mechanisms (grants, contracts, intramural support) and there is no minimum threshold for cost

Data Sharing

- Non-human data: made available through current databases and resources remain standard mechanism; any widely used data repository (e.g., GenBank, SRA, ZFIN)
- Human data: studies with data derived from human specimens registered in dbGaP (the database of Genotypes and Phenotypes)

Access to Human Genomic Data: Unrestricted vs Controlled

Informed Consent: Determines the appropriateness of submitting human data to unrestricted or controlled-access NIH data repositories

- *Unrestricted/ Open-access* Tier: Data are publicly available to anyone (ie.The 1000 Genomes Project); includes study protocols, metadata, certain phenotype data, genomic summary results (sensitive populations may apply for all controlled-access).
- *Controlled-access* Tier: Investigators must obtain approval from NIH Data Access Committees to use the requested data (e.g., dbGaP); includes individual-level sequence data and potentially identifiable phenotype and analyzed data. Exception is when *informed consent* explicitly states unrestricted-access to individual-level human genomics data is appropriate.

KEY ELEMENTS OF NIH DATA SHARING PLANS (DSP)

- Data sharing is specific by funding announcement (IC program staff inform applicants whether to include DSP with funding application)
    - Program Announcements may request data sharing plans for applications that are less than $500,000 direct costs in any single year.
    - Reviewers will **not** factor the proposed data-sharing plan into the determination of scientific merit or priority score.
    - Program staff will be responsible for overseeing the data sharing policy and for assessing the appropriateness and adequacy of the proposed data-sharing plan.
- DSP should state clearly and include details for how data and metadata will be shared for secondary use *(data platforms, levels of data, associated clinical/ phenotype data)*
- Data release for public secondary use is expected to be no later than the acceptance for publication of the main findings from the final data set.
- NIH continues to expect that the initial investigators may benefit from first and continuing use but *not from prolonged exclusive use*.
- Grantees choose their own NIH Data Repository (Table of NIH Data Sharing Repositories), the decision process can be referenced in the DSP.
- If data cannot be shared in accordance with NIH policy, applicants must clearly outline the reasons and provide an alternate plan for program staff to consider.

PROCEDURES: FOR dbGaP (THE DATABASE OF GENOTYPES AND PHENOTYPES):

The dbGaP Data Request Process

1. Requester submits Data Access Request (DAR) to Institutional Signing Official
2. Signing Official (SO) approves and submits to Data Access Committee (DAC) staff
3. DAC Staff DAR pre-review
4. Full Data Access Committee (DAC) review
5. Data Access Request (DAR) is approved or disapproved: ***approval in days or even hours!***
6. Requestor is notified by email of DAC decision
7. Requestor downloads data
8. Requester completes Annual Report
9. Renew access or closeout

NIH Code of Conduct for Genomic Data Use

Investigators agree to:

- Use requested datasets only for the research described in their Data Access Request (DAR)
- Not distribute data to individuals not specified in their DAR
- Not attempt to contact or identify research participants
- Adhere to dbGaP Best Practices that ensures data security
- Report Data Management Incidents
- Ensure the proposed Research Use Statement is consistent with the Genomic Data Sharing Policy
- Not re-deposit data in public databases

*External collaborators must independently apply for data access.

See Addendum on page 8 for Complete NIH resources found on the NIH Office of Extramural Research page:

NIH Sharing Policies and Related Guidance on NIH-Funded Research Resources

Includes a Table of all NIH Data Sharing Policies and a link to NIH-accepted Data Sharing Repositories.

TALK GOAL 3: DISCUSS BARRIERS TO SHARING AND POTENTIAL WAYS TO OVERCOME THEM

- Lessons learned and potential solutions to barriers in broad & equitable sharing (Informed Consent, submission to public dBs)
- Clinical data/phenotype, open data and "coded" data sets

Jamie's slides (29-36) detail various Data Sharing needs at the NIH focusing on Precision Medicine, including The Cancer Genome Atlas (TCGA), The Genomic Data Commons (GDC), the NCI-Match Clinical Trial, TARGET and the importance in sharing Data in translating key discoveries.

Current Barriers to Data Sharing

- Inability to integrate data due to disparate consenting language and processes
    - NIH strongly encourages patient consent language that allows for future use and broad sharing of data without additional use restriction placed on it.
    - Ideally "Unrestricted" or "General Research Uses" should be used to allow data to be combined/analyzed with any other dataset.
    - Avoid Disease-specific references particularly if the language could be interpreted as exclusive use *(ie. Instead of Permission for "kidney cancer" and related disorders use "must include disease….")*
- Large volumes of data generated quickly without consistent formats or data/metadata standards
- Lack of searchable and interconnected data repositories with associated tools and services
- Lack of agreed upon ontologies, vocabularies, and data models that severely impacts interoperability, integration, and analysis across multiple datasets

- Policy and procedural obstacles preventing patients and researchers from contributing their data to certain databases
    - Mandates and legal issues from funding sources (GDPR)
    - Lack of resources to format data and metadata files, and further submit them to databases
    - Difficulty choosing the best database to house the data
- Consent and data-use agreements (Electronic, trackable, machine-readable consents and terms-of-use agreements for data and other services to enable monitoring, computationally enforcing, and updating these agreements)

The Beau Biden Cancer Moonshot *(Slide 40)*

Overarching goals

- Accelerate progress in cancer, including prevention & screening
- From cutting edge basic research to wider uptake of standard of care
- Encourage greater cooperation and collaboration within and between academia, government, and private
- sector
- Enhance data sharing

Recommendations from Blue Ribbon Panel – October, 2016

- Build a National Cancer Data Ecosystem – MUCH broader than NCI.
    - Will include ***Cancer Research Data Commons***
- Enhanced cloud-computing platforms
- Services that link disparate information, including clinical, image, and molecular data
- Essential underlying data science infrastructure, standards, methods, and portals for the Cancer Data
- Ecosystem

To Build Ecosystem: NCI is Creating Partnerships

- Administrative supplements for Cancer Centers in GENIE and GA4GH coordination.
- Coordination with and support of Moonshot Programs
    - Assistance for U24 programs, e.g., Human Tumor Atlas & Immuno-oncology Data Coordinating Centers
- Work across related initiatives/programs
    - NCI, other NIH Institutes, NIH Data Commons Pilot Phase Consortium, All of Us, Chan Zuckerberg Initiative, GA4GH
- Workshops and RFIs to gather community input, feedback, and participation
- Establish CRDC governance process, including Scientific and Technical Advisory Board and Steering Committee.
- Establishing NCI Office of Data Sharing as a resource to NCI staff, external investigators and the broader research and participant communities.

TALK GOAL 4: INTRODUCE WAYS THAT THE NEW NCI OFFICE OF DATA SHARING IS ADVOCATING FOR THE PROPER BALANCE OF BROAD AND OPEN DATA SHARING/ACCESS WHILE RESPECTING THE RIGHT OF PATIENTS TO PARTICIPATE IN AND BENEFIT FROM RESEARCH AS THEY SEE FIT. *Jaime did not have time in this webinar to address this point, but her slides contain valuable information relating to this effort.*

ADDENDUM

Complete NIH resources can be found on the NIH Office of Extramural Research page:
NIH Sharing Policies and Related Guidance on NIH-Funded Research Resources

It is NIH policy that the results and accomplishments of the activities that it funds should be made available to the public. PIs and funding recipient institutions are expected to make the results and accomplishments of their activities available to the research community and to the public at large. The following links highlight selected NIH policies and related guidance on sharing of research resources developed with NIH funding.

Policy on Dissemination of NIH-Funded Clinical Trial Information (8/2016)

NIH Intramural Human Data Sharing Policy (8/2016)

NIH Public Access Plan for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Funded Scientific Research (02/2015) (PDF - 474 KB) – This document describes NIH's plans to build upon and enhance its longstanding efforts to increase access to scholarly publications and digital data resulting from NIH-funded research.

Genomic Data Sharing (GDS) (8/2014) Final Genomic Data Sharing (GDS) Policy that provides for the sharing, for research purposes, of large-scale human and non-human genomic data generated from NIH-funded research. Effective for grant applications and contract proposals submitted for January 25, 2015 due date and thereafter.

NIH Grants Policy Statement (Availability of Research Results) (11/2015) - Section of the NIH Grants Policy Statement discussing the availability of research results developed with NIH funding, including publications, data, unique research resources, and intellectual property (inventions and patents).

Common Data Element (CDE) Resource Portal (03/2013) - The Common Data Element (CDE) Resource Portal provides access to NIH-supported CDE initiatives and other tools and resources which can help researchers use common data elements (CDEs) in clinical research, patient registries, and other human subject research in order to improve data quality and opportunities for comparison and combination of data from multiple studies and with electronic health records.

Table of NIH Data Sharing Policies (03/2013) - This table lists additional data sharing policies in effect at NIH at the NIH, IC, division, and program levels that apply to broad sets of investigators and data.

Table of NIH Data Sharing Repositories (03/2013) - This table lists various NIH-supported data repositories that accept submissions of appropriate data from NIH-funded investigators and others, as well as including resources that aggregate information about biomedical data and information sharing systems.

Data Repositories Resource Guide (09/2012) - (MS Word - 30 KB) - This resource guide document is designed to assist the NIH extramural community by identifying examples of data repositories which may be used for sharing data developed under NIH funding programs, consistent with NIH sharing policies.

Data Standards and Common Data Elements Resource Guide (09/2011) - (MS Word - 29 KB) - This resource guide document is designed to assist the NIH extramural community in identifying and utilizing certain data standards and common data elements in NIH programs.

Example Plan addressing Key Elements for a Data Sharing Plan under NIH Extramural Support (08/2010) - (MS Word - 55 KB) - This resource document is designed to assist the NIH extramural applicant community in preparing data sharing plans by providing an example that shows how a sharing plan addresses the key elements for a data sharing plan.

Key Elements to Consider in Preparing a Data Sharing Plan under NIH Extramural Support (12/2009) – (PDF - 32 KB) - This resource document is designed to assist the NIH extramural applicant community in preparing data sharing plans by identifying key elements that should be addressed in the plan.

NIH Genome-Wide Association Studies (GWAS) Policy (Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS)) (08/2007) - Policy concerning sharing of GWAS data obtained in NIH supported or conducted research. (Please refer to Genomic Data Sharing (GDS) Policy webpage.)

Data Sharing Regulations/Policy/Guidance Chart for NIH Awards (08/30/2006) - (MS Word - 62 KB) - This chart is designed as a quick guide only for the purpose of identifying various data sharing regulation/policy/guidance documents applicable to NIH funding.

NIH Public Access Policy (02/2005) - Policy on Enhancing Public Access to Archived Publications Resulting from NIH-Funded Research.

NIH Model Organism Sharing Policy (NIH Policy on Sharing of Model Organisms for Biomedical Research)(05/2004) - Policy concerning the sharing and distributing of model organisms and related research resources generated using NIH funding.

NIH Data Sharing Policy (Final NIH Statement on Sharing Research Data) (02/2003) - Policy concerning the sharing of research data for funding applications seeking $500,000 or more in direct costs in any year of the project period.

NIH Research Tools Policy (Principles and Guidelines for Recipients of NIH Research Grants and Contracts on Obtaining and Disseminating Biomedical Research Resources) (12/1999) - (PDF – 150 KB) - Policy designed to provide NIH funding recipients with guidance concerning appropriate terms for disseminating and acquiring unique research resources developed with federal funds, and intended to assist recipients in complying with their obligations under the Bayh-Dole Act and NIH funding policy.

Biological Materials Policy (NIH Procedures for Handling Non-Election of Title to Patentable Biological Materials) (05/1996) - NIH policy for allowing NIH funding recipients to retain and license biological materials for which patent protection might not be pursued.

Developing Sponsored Research Agreements (Considerations for Recipients of NIH Research Grants and Contracts) (11/1994) - Issues and points to consider in developing sponsored research agreements with commercial entities, where such agreements may include research activities which are fully or partially funded by NIH, in order to assist funding recipients ensure such agreements comply with the requirements of the Bayh-Dole Act and NIH funding agreements while upholding basic principles of academic freedom.

For any questions or further guidance or assistance, please email sharing@nih.gov.
For any questions or further guidance or assistance for extramural invention reporting and related intellectual property policy issues, please refer to http://inventions.nih.gov or email Inventions@nih.gov.