The NIH Genomic Data Sharing Policy:
Applicability to the Intramural Research Program

Kathleen Calzone, PhD, RN, AGN-BC, FAAN

Research Geneticist, Genetics Branch, Center for Cancer Research

GPA, Center for Cancer Research

NIH NATIONAL CANCER INSTITUTE
Center for Cancer Research

December 1, 2020

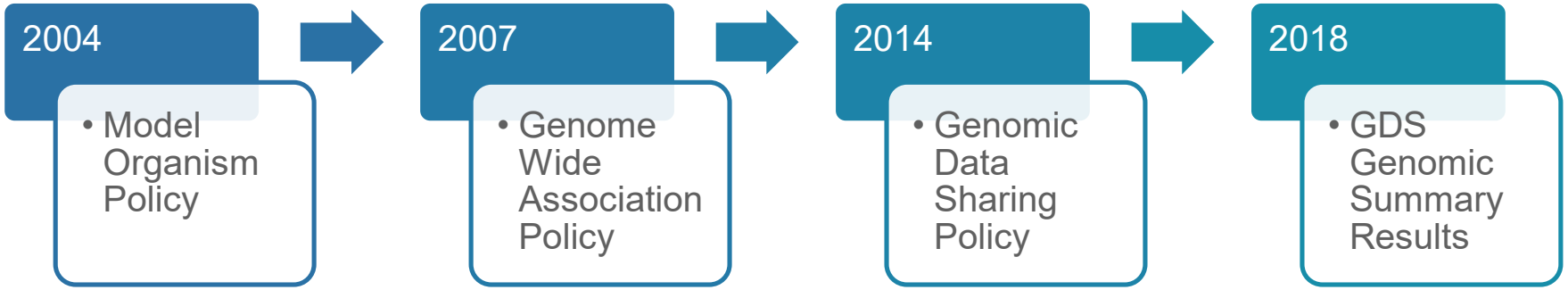
Objectives

- Describe the critical elements of National Institutes of Health (NIH) Genomic Data Sharing Policy.
- Recognize when the NIH Genomic Data Sharing Policy has applicability to intramural investigators generating genomic data (human and non-human).
- Discuss the critical components of adherence to the NIH Genomic Data Sharing Policy.
- Identify the protocol and consent specific content required by the NIH Genomic Data Sharing Policy.
- Illustrate how to identify your NIH Genomic Data Sharing Policy Genomic Program Administrator (GPA).

Critical Elements of the NIH Genomic Data Sharing Policy and Applicability to Intramural Investigators

Background

- NIH supports a culture of sharing
- Track record of continued support for data sharing policies that facilitate research that can improve human health since 1999
- Some examples include:



https://osp.od.nih.gov/wp-content/uploads/NIH_GDS_Policy_Overview.pdf

Benefits of Data Sharing

- Increases the capacity to answer research questions in the context of rare diseases
- Maximizes the utility of genomic data to explore additional research questions
- Enables higher statistical power as data from multiple studies can be combined to answer important questions
- Facilitates reproducibility
- Facilitates innovation of methods and tools for research

https://osp.od.nih.gov/wp-content/uploads/NIH_GDS_Policy_Overview.pdf

Genomic Data Sharing Policy Effective Dates

- The GDS Policy takes effect for competing grant applications and contract proposals submitted to NIH on or after **January 25, 2015**
- **The Intramural Research Programs were directed to comply with the GDS policy on data generated on or after August 31, 2015**
- On November 1, 2018, the GDS Policy was updated to allow unrestricted access to Genomic Summary Results

NIH Intramural Investigator Responsibilities Under the NIH GDS Policy https://osp.od.nih.gov/wp-content/uploads/NIH_Intramural_Investigator_Responsibilities_under_the_GDS_Policy.pdf

Genomic Summary Results

- Permits unrestricted access to genomic summary results
  - Aggregate genomic data
  - Genomic summary statistics
    - Genotype counts and frequencies; allele counts and frequencies; effect size estimates and standard errors; likelihoods; and p-values
- Useful to investigators seeking to determine which genomic variants may contribute to a disease or condition
  - These data are typically published in peer reviewed scientific literature

Green E., Wollinetz, C. (2018). Protecting Participants, Empowering Researchers: Providing Access
.gov/news/news-release/Protecting-Participants
-Empowering-Researchers-Providing-Access-to-Genomic-Summary-Results
https://grants.nih.gov/grants/guide/notice-files/NOT-OD-19-023.html

Genomic Summary Results, cont.

- Excludes public release of genomic summary results in sensitive populations
  - Isolated geographic regions
  - With rare or potentially stigmatizing traits
    - i.e. HIV
  - Populations with data restrictions
    - i.e. Native American/Alaska Native or other indigenous populations
  - Study population small enough that participants privacy could be at risk

Green E., Wollinetz, C. (2018). Protecting Participants, Empowering Researchers: Providing Access to Genomic Summary Results https://www.genome.gov/news/news-release/Protecting-Participants-Empowering-Researchers-Providing-Access-to-Genomic-Summary-Results
https://grants.nih.gov/grants/guide/notice-files/NOT-OD-19-023.html

Intramural GDS Policy Applicability

- All intramural NIH-funded research that generates large-scale genomic data
  - Human or non-human genomic data
    - Non-human consists of Model organisms, Microbes, Microbiome
  - Genome-Wide Association Studies (GWAS)
  - Single Nucleotide Polymorphisms (SNP) arrays
  - Genome Sequence
  - Transcriptomic
  - Metagenomic
  - Epigenomic
  - Gene Expression

https://osp.od.nih.gov/wp-content/uploads/NIH_GDS_Policy.pdf

What is “Large Scale”: Examples of Data Submission Thresholds

Human Only

- Sequence data from more than one gene or region of comparable size in the genomes of >1,000 human research participants
- Sequence data from >100 genes or region of comparable size in the genomes of >100 human research participants
- Data from 300,000 or more variant sites in >1,000 human research participants
- Comparisons of differentially methylated sites genome-wide at single-base resolution within a given sample (e.g., within the same subject over time or across cell types within the same subject).

What is “Large Scale”: Examples of Data Submission Thresholds

Model Organisms, Non-Human Cell Lines, or Infectious Organisms

- Sequence data from more than 100 isolates from infectious organisms.
- Whole genome or exome sequence data of more than one model organism species or strain
- Comprehensive catalog of transcripts and non-coding RNA from one or more model organism species or strains
- Catalog of more than 100,000 single nucleotide polymorphisms (SNPs) from one or more model organism species or strains

What is “Large Scale”: Examples of Data Submission Thresholds

Human and Model Organisms, Non-Human Cell Lines, or Infectious Organisms

- Sequence data from more than 100 metagenomes of human or model organism microbiomes.
- Sequence data from more than 100 metatranscriptomes of human or model organism microbiomes.
- Comparisons of genome-wide methylated sites across more than 10 cell types.

What is “Large Scale”: Trans **NCI** Data Submission Framework

	Number of Specimens	
	Human (Including human cell lines)	Model Organisms, Non-Human Cell Lines, Infectious Organisms
SNP array data from >500K single nucleotide polymorphisms (SNPs) (e.g., GWAS data)	1,000	500
DNA sequence data from < 100 genes or regions of interest (e.g. targeted sequencing)	1,000	500
DNA sequence data from ≥ 100 genes or regions of interest (e.g., targeted sequencing, whole exome sequencing, whole genome sequencing)	100	50
Genome-wide RNA sequencing (RNA-seq) data (e.g., transcriptomic data)	100	50
Genome-wide DNA methylation data (e.g., bisulfite sequencing data)	100	50
Genome-wide chromatin immunoprecipitation sequencing (ChIP-seq) data (e.g. transcription factor ChIP-seq, histone modification ChIP-seq)	100	50
Metagenome (or microbiome) sequencing data (e.g., 16S rRNA sequencing, shotgun metagenomics, whole-genome microbial sequencing)	100	50
Metatranscriptome sequencing data (e.g., microbial/microbiome transcriptomics)	100	50

Genomic Data Sharing Applicability

- If you are unsure whether your research falls into Genomic Data Sharing Policy required sharing, contact your Genomic Program Administrator (GPA)
- NIH Institutes or Centers may choose on a case-by-case basis to apply the Policy to projects generating data on a smaller scale depending on the state of the science, the needs of the research community, and the programmatic priorities of the IC
  - For example, the Scientific Director of the Center for Cancer Research, issued guidance on sharing of genomic data for projects examining rare diseases and rare cancers with **no minimum thresholds**
    - Definition of rare disease – a disease that affects **less than 200,000 persons** in the United States (U.S Food and Drug Administration (FDA)).

https://ccrod.cancer.gov/confluence/display/CCRGDS/CCR+Genomic+Data+Sharing+Project+Home

Components Necessary to Adhere to the NIH Genomic Data Sharing Policy

Genomic Data Sharing Plan

- "All investigators doing research covered by GDS Policy must develop and have in place an approved data sharing plan **PRIOR to start of the research**."
  - What data will be shared
  - How and where data will be shared
    - All human studies must be registered in dbGaP regardless of where the data will ultimately be shared
  - Timeframe for sharing

https://osp.od.nih.gov/wp-content/uploads/NIH_Intramural_Investigator_Responsibilities_under_the_GDS_Policy.pdf

Institutional Certification

Confirms that data submission is:

- Consistent with national, tribal, and state laws and regulations as well as relevant institutional policies
- Consistent with the informed consent
- Consideration was given to the to the study participants from whom the data were obtained, including access to genomic summary results
- Any limitations on the research use of the data as specified in the informed consent are reflected on the Institutional Certification (Table 3)
- The investigators plan for de-identifying datasets is consistent with the GDS policy
  - Assigned random unique codes by investigator(s)

https://osp.od.nih.gov/scientific-sharing/institutional-certifications/

Institutional Certification

- For Studies Using Data Generated from Cell Lines Created or Clinical Specimens Collected AFTER August 31, 2015
  - Consent is required
- For Studies Using Data Generated from Cell Lines Created or Clinical Specimens Collected BEFORE August 31,2015
  - Lack Consent
  - Have Consent
- The Institutional Certification is required **PRIOR to the start of the research**
  - OSP-Scientific Review
    - Study Initiation

https://osp.od.nih.gov/scientific-sharing/institutional-certifications/

Institutional Certification-Multi-Site Studies

- Provided for **all** institutions contributing samples
- Two options
  - Primary study site may submit one* Multi-Center Institutional Certification indicating that it is providing certification on behalf of all collaborating sites
    - *Documentation of consent and data use limitations from all collaborating sites
  - Each site providing samples may provide their own Single-Site Institutional Certification.

Data Use Limitations

What are Data Use Limitations?

- Description of limitations, if any, on the distribution and use of human data submitted to controlled-access NIH designated data repositories
- Based on the terms included in the informed consent
- Stipulated by the submitting institution
- Many Data Use Limitations are General Research Use
  - Data would generally be made available to any qualified investigator, irrespective of the specific research purpose

Exceptions

- When Institutional Certification criteria cannot be met
  - Provide a justification for any data submission exceptions
  - Investigator expected to develop an alternate plan to share data through other mechanisms
  - IF an exceptions is granted, the study must still be registered in dbGaP
    - The reason for the exception will be included
    - A reference will be provided to an alternative data-sharing plan or resource

Study Registration

Human Data

- All human genomic data studies covered by the GDS Policy must be registered in dbGaP
- This is regardless of which NIH-designated data repository (e.g., dbGaP, GEO, SRA, etc.) will ultimately house and distribute the data

https://osp.od.nih.gov/wp-content/uploads/NIH_Intramural_Investigator_Responsibilities_under_the_GDS_Policy.pdf

Data Submission Timeline

Level	General Description of Data Processing	Example Data Types	Data Submission Expectation	
			Human	Non-Human
0	Raw data generated directly from the instrument platform	Instrument image data	Not expected	Not expected
1	Initial sequence reads, the most fundamental form of the data after the basic translation of raw input	DNA sequencing reads, ChIP-Seq reads, RNA-Seq reads, SNP arrays, Array CGH	Not expected	Not expected ***except for de novo sequence data***

https://osp.od.nih.gov/wp-content/uploads/Supplemental_Info_GDS_Policy.pdf

Data Submission Timeline

Level	General Description of Data Processing	Example Data Types	Data Submission Expectation	
			Human	Non-Human
2	Data after an initial round of analysis or computation to clean the data and assess basic quality measures	DNA sequence alignments to a reference sequence or de novo assembly, RNA expression profiling	Project specific; after data cleaning and quality control, which is generally **within 3 months after data have been generated**	Data submission is expected **no later than the time of initial publication**; an earlier submission date may be designated for certain data types or NIH projects

https://osp.od.nih.gov/wp-content/uploads/Supplemental_Info_GDS_Policy.pdf

Data Submission Timeline

Level	General Description of Data Processing	Example Data Types	Data Submission Expectation	
			Human	Non-Human
3	Analysis to identify genetic variants, gene expression patterns, or other features of the dataset	SNP or structural variant calls, expression peaks, epigenomic features	Project specific; after cleaning and quality control, **which is generally within 3 months after data have been generated**	Data submission is expected **no later than the time of initial publication**; an earlier release date may be designated for certain data types or NIH projects

https://osp.od.nih.gov/wp-content/uploads/Supplemental_Info_GDS_Policy.pdf

Data Submission Timeline

Level	General Description of Data Processing	Example Data Types	Data Submission Expectation	
			Human	Non-Human
4	Final analysis that relates the genomic data to phenotype or other biological states	Genotype-phenotype relationships, relationships of RNA expression or epigenomic patterns to biological state	Data submitted **as analyses are completed**	Data submission is expected **no later than the time of initial publication**

https://osp.od.nih.gov/wp-content/uploads/Supplemental_Info_GDS_Policy.pdf

Protocol and Consent Content Required by the NIH Genomic Data Sharing Policy

https://osp.od.nih.gov/wp-content/uploads/Supplemental_Info_GDS_Policy.pdf

Consent, the GDS Policy and the revised Common Rule

- The revised Common Rule (effective January 19, 2018) does not require informed consent for research with de-identified biospecimens or cell lines
- **HOWEVER**, NIH GDS Policy requires that informed consent be obtained for the future research use and broad data sharing
  - Rationale is that genomic technology and analytical methods may increase the risk for identification

https://osp.od.nih.gov/scientific-sharing/genomic-data-sharing-faqs/

Consent

- **Studies using data or specimens collected after August 31, 2015**
  - Consent required for genomic data sharing for future research
    - Consent data sharing template language is available in the IRBO consent templates https://irbo.nih.gov/confluence/display/ohsrp/Templates+and+Forms
  - "and even when data is **from cell lines or clinical specimens if they are created or collected after August 31, 2015**"
    - An Exception request should be submitted if planning to use cell lines or clinical specimens without consent
- **Study population Sensitivity Determination**
  - Established **BEFORE** submission to the IRB
  - Determines the consent language on Genomic Summary Results

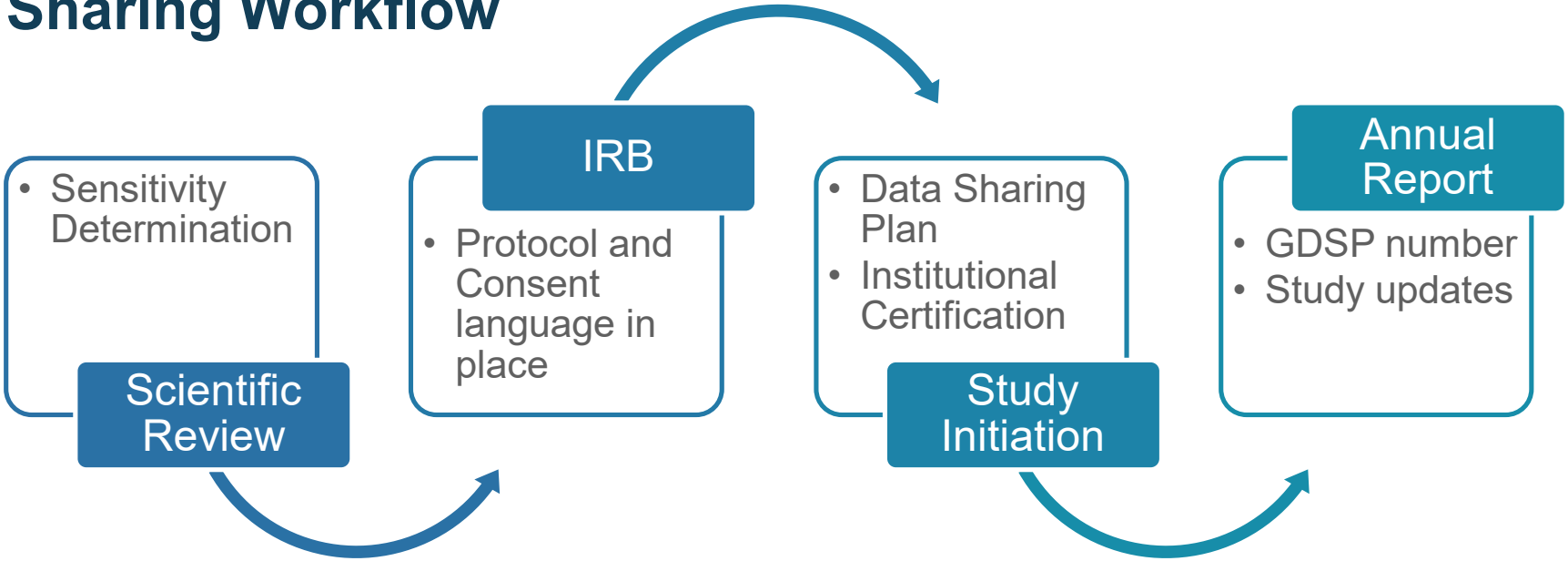
https://osp.od.nih.gov/wp_content/uploads/NIH_Intramural_Investigator_Responsibilities_under_the_GDS_Policy.pdf

Clinical Protocol and the GDS Policy

- Include in the Confidentiality and Privacy section of the protocol whether the population is considered Sensitive or Not Sensitive
  - Confirms the correct consent language is used for Genomic Summary Results
  - ****Note-Repository studies could enroll Sensitive and Not-Sensitive populations****
- Genomic analyses should be sufficiently described in the protocol
  - A description of how samples/data will be labeled should be included
    - Will identifiers be attached to data or will samples/data be coded or unlinked
- Include whether Genomic Data Sharing is expected
  - Protocol content is **NOT** your Data Sharing Plan
    - A Data Sharing Plan consists of what, how, when and where data will be shared

Model Intramural Workflow: NCI/CCR Genomic Data Sharing Process

Exemplar Center for Cancer Research Genomic Data Sharing Workflow



Infrastructure Underpinning Workflow

- CCR Wiki page
- Standard Operating Procedures
- Genomic Data Sharing Portal

https://ccrod.cancer.gov/confluence/display/CCRGDS/CCR+Genomic+Data+Sharing+Project+Home

Finding your NIH Genomic Data Sharing Policy Genomic Program Administrator (GPA)

Role of the Genomic Program Administrator (GPA)

- Institute specific GDS Policy implementation leaders
  - Intramural or extramural or serve in both capacities
- Answer questions about study applicability
- Answer questions about GDS Policy requirements and paperwork
  - Data Sharing Plan
  - Institutional Certification
  - Exception Requests
- Register the study in dbGaP
- Facilitate data submission processes for human NIH-designated controlled-access data repositories

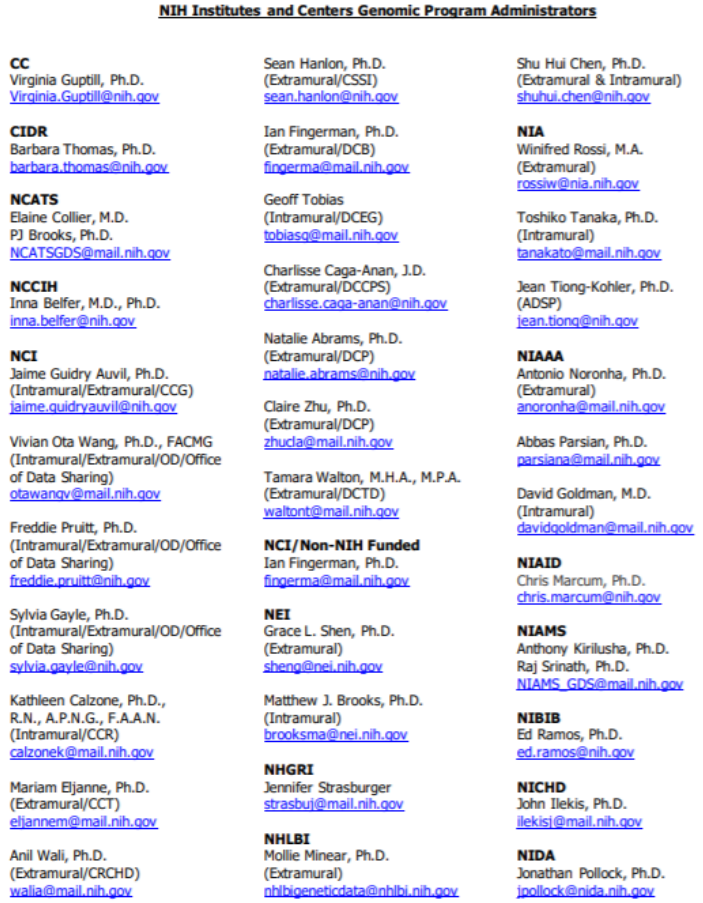
Finding Your Genomic Program Administrator

- Step One: Go to the NIH Office of Intramural Genomic Data Sharing Home Page

https://oir.nih.gov/sourcebook/intramural-program-oversight/intramural-data-sharing/genomic-data-sharing

Finding Your Genomic Program Administrator

- Step Two: Find your Institute
- Each GPA listed includes
  - Name
  - Email
  - Additional details if applicable, i.e. Intramural or Extramural
- This list was updated **Nov. 20, 2020**



NIH Institutes and Centers Genomic Program Administrators

CC
Virginia Guptill, Ph.D.
Virginia.Guptill@nih.gov

CIDR
Barbara Thomas, Ph.D.
barbara.thomas@nih.gov

NCATS
Elaine Collier, M.D.
PJ Brooks, Ph.D.
NCATSGDS@mail.nih.gov

NCCIH
Inna Belfer, M.D., Ph.D.
inna.belfer@nih.gov

NCI
Jaime Guidry Auvil, Ph.D.
(Intramural/Extramural/CCG)
jaime.guidryauvil@nih.gov

Vivian Ota Wang, Ph.D., FACMG
(Intramural/Extramural/OD/Office of Data Sharing)
otawangv@mail.nih.gov

Freddie Pruitt, Ph.D.
(Intramural/Extramural/OD/Office of Data Sharing)
freddie.pruitt@nih.gov

Sylvia Gayle, Ph.D.
(Intramural/Extramural/OD/Office of Data Sharing)
sylvia.gayle@nih.gov

Kathleen Calzone, Ph.D., R.N., A.P.N.G., F.A.A.N.
(Intramural/CCR)
calzonek@mail.nih.gov

Mariam Eljanne, Ph.D.
(Extramural/CCT)
eljannem@mail.nih.gov

Anil Wali, Ph.D.
(Extramural/CRCHD)
walia@mail.nih.gov

Sean Hanlon, Ph.D.
(Extramural/CSSI)
sean.hanlon@nih.gov

Ian Fingerman, Ph.D.
(Extramural/DCB)
fingerma@mail.nih.gov

Geoff Tobias
(Intramural/DCEG)
tobiasg@mail.nih.gov

Charlisse Caga-Anan, J.D.
(Extramural/DCCPS)
charlisse.caga-anan@nih.gov

Natalie Abrams, Ph.D.
(Extramural/DCP)
natalie.abrams@nih.gov

Claire Zhu, Ph.D.
(Extramural/DCP)
zhucla@mail.nih.gov

Tamara Walton, M.H.A., M.P.A.
(Extramural/DCTD)
waltont@mail.nih.gov

NCI/Non-NIH Funded
Ian Fingerman, Ph.D.
fingerma@mail.nih.gov

NEI
Grace L. Shen, Ph.D.
(Extramural)
sheng@nei.nih.gov

Matthew J. Brooks, Ph.D.
(Intramural)
brooksma@nei.nih.gov

NHGRI
Jennifer Strasburger
strasbuj@mail.nih.gov

NHLBI
Mollie Minear, Ph.D.
(Extramural)
nhlbigeneticdata@nhlbi.nih.gov

Shu Hui Chen, Ph.D.
(Extramural & Intramural)
shuhui.chen@nih.gov

NIA
Winifred Rossi, M.A.
(Extramural)
rossiw@nia.nih.gov

Toshiko Tanaka, Ph.D.
(Intramural)
tanakato@mail.nih.gov

Jean Tiong-Kohler, Ph.D.
(ADSP)
jean.tiong@nih.gov

NIAAA
Antonio Noronha, Ph.D.
(Extramural)
anoronha@mail.nih.gov

Abbas Parsian, Ph.D.
parsiana@mail.nih.gov

David Goldman, M.D.
(Intramural)
davidgoldman@mail.nih.gov

NIAID
Chris Marcum, Ph.D.
chris.marcum@nih.gov

NIAMS
Anthony Kirilusha, Ph.D.
Raj Srinath, Ph.D.
NIAMS_GDS@mail.nih.gov

NIBIB
Ed Ramos, Ph.D.
ed.ramos@nih.gov

NICHD
John Ilekis, Ph.D.
ilekisj@mail.nih.gov

NIDA
Jonathan Pollock, Ph.D.
jpollock@nida.nih.gov

11/20/20

https://osp.od.nih.gov/wp-content/uploads/IC_GPAs.pdf

Considerations

Considerations

- Genomic Data Sharing is part of the NIH mission
- Consider the Genomic Data Sharing Policy requirements when you are **planning** your research project
- Work with your GPA from the start of planning for your research project through data submission to help you with what is required
- For Human Studies, the GDS Policy requires consent
  - The Institutional Certification **cannot** be signed unless there is consent
- When your research plan evolves, how does this impact genomic data sharing? (i.e. population, accrual, types of planned analyses, etc.)

Publishers Mandating Data Sharing

- Publishers are increasingly mandating genomic data sharing regardless of whether the study meets a GDS specified threshold
  - Examples include:
    - American Association for Cancer Research Journals
    - American Journal of Human Genetics
    - Cell Press
    - Journal of the American Medical Association
    - The Lancet
    - New England Journal of Medicine
    - Science

https://osp.od.nih.gov/wp-content/uploads/Journals_Requiring_Broad_Sharing_of_Data.pdf

Questions/Discussion

calzonek@mail.nih.gov
240-760-6178

Genomic Data Sharing FAQ's
https://osp.od.nih.gov/scientific-sharing/genomic-data-sharing-faqs/