

DIVISION OF CANCER CONTROL AND POPULATION SCIENCES | DIVISION OF  
CANCER PREVENTION | NEW GRANTEE WORKSHOP 2019

# NIH DATA SHARING

**Charlisse Caga-anan JD**  
DCCPS Genomic Program  
Administrator (GPA)

November 21, 2019

# ASK YOURSELF THE FOLLOWING QUESTIONS

**What are the data (and resource) sharing policies/requirements at NIH?**

**How do I determine which sharing policies/requirements apply to my grant?**

**What do I have to do under each policy/requirement?**

- What/where/when/how/with whom do I share?
- What do I include in my grant application / provide prior to funding?
- What do I include in my progress reports?

**Where do I go for help?**

# WHAT ARE THE DATA (AND RESOURCE) SHARING POLICIES / REQUIREMENTS AT NIH?

## NIH GRANTS POLICIES

- NIH Sharing Policies and Related Guidance on NIH-Funded Research Resources (<https://grants.nih.gov/policy/sharing.htm>)

“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.”

## REQUIREMENTS SPECIFIC TO THE FUNDING OPPORTUNITY ANNOUNCEMENT (FOA)

- E.g. Moonshot FOAs, Cohort Infrastructure PAR, DCCPS Modular R01 PAR

# SELECT NIH SHARING POLICIES

Official Policy Title	a.k.a.	Resource Website
NIH Genomic Data Sharing Policy (2015)	<b>GDS Policy</b>	<a href="https://osp.od.nih.gov/scientific-sharing/genomic-data-sharing/">https://osp.od.nih.gov/scientific-sharing/genomic-data-sharing/</a>
Final NIH Statement on Sharing Research Data (2003)	<b>2003 Data Sharing Policy</b>	<a href="https://grants.nih.gov/grants/policy/data_sharing/">https://grants.nih.gov/grants/policy/data_sharing/</a>
NIH Policy on Sharing of Model Organisms for Biomedical Research (2004)	<b>Model Organism Policy</b>	<a href="https://grants.nih.gov/grants/policy/model_organism/">https://grants.nih.gov/grants/policy/model_organism/</a>

# HOW DO I DETERMINE WHICH NIH DATA SHARING POLICIES APPLY TO MY GRANT?

Policy	Look at...	Criteria
<b>GDS Policy</b>	Genomic data <i>generated</i> using grant funds	Large-scale genomic research data <u>or</u> smaller-scale high priority genomic data generated by grant (human or non-human)
<b>2003 Data Sharing Policy</b>	Budget	\$500k in direct costs in any given year
<b>Model Organism Policy</b>	Model organism(s) <i>generated</i> using grant funds	New, genetically modified variants of model organisms (non-human), and related resources, are generated by grant

Also consider policy's effective date

# EXAMPLE: DOES MY GRANT FALL UNDER THE GDS POLICY?

**Large scale genomic data** meeting NCI thresholds for

- *Technology/platform used*
- *Sample size*
- *Human or non-human*

OR

**Smaller scale high priority genomic data** depending on

- *State of the science*
- *Programmatic priorities of NCI*
- *Utility of the data for the research community*

E.g. projects examining rare cancers, rare-cancer-related outcomes, or rare cancer subtypes

Example NCI Thresholds

Type of Data	Data Examples	# of Human Specimens	# of Model Organisms
SNP array data from > 500k single nucleotide polymorphisms (SNPS)	GWAS Data	1,000	500
DNA or RNA sequence data from < 100 genes or regions of interest, or transcripts	Targeted Sequencing	1,000	500
DNA sequence data from ≥ 100 genes or regions of interest	Targeted Sequencing, Whole Exome Sequencing, Whole Genome Sequencing	100	50
Genome-wide RNA sequencing data	Transcriptomic Data	100	50

# WHAT DO I HAVE TO DO UNDER EACH POLICY/REQUIREMENT?

**What/where/when/how/with whom do  
I share?**

# GDS POLICY

<b>WHAT</b>	<p>The following data must be shared:</p> <ul style="list-style-type: none"><li>• <b>Genomic data (human or non-human):</b> All genomic/genetic data <i>generated</i> using grant funds, <u>AND</u></li><li>• <b>Phenotypic data:</b> Data pertinent to the interpretation of the expected genomic data (i.e. data need to reproduce primary analysis); regardless of how phenotype data collection was funded</li></ul>
<b>WHERE</b>	NIH-designated repositories (e.g. database of Genotype or Phenotypes [dbGaP])
<b>WHEN</b>	<ul style="list-style-type: none"><li>• <b>Submission:</b> Data must be submitted to the repository at point when data is generated and QA/QC is complete</li><li>• <b>Release:</b> The repository will release the data for access 6 months after QA/QC was completed <u>OR</u> <b>at the time of publication, whichever comes first [added to slide]</b></li></ul>
<b>HOW</b>	For data submitted under the GDS Policy, access is also governed by the GDS Policy (e.g. data in dbGaP is available for controlled access based on participant consent)
<b>WITH WHOM</b>	Data should be available to the broad research community

# 2003 DATA SHARING POLICY

**WHAT** Final research data (i.e. data used to address specific aims) must be shared; regardless of whether it was generated/collected within the grant

**WHERE**

- Maintained **locally/within project** (e.g. project database, archive, enclave), OR
- Submitted to an **NIH-supported database** (e.g., dbGaP)

**WHEN** Data should be available at whichever point of time comes first:

- At the time that primary manuscript is accepted for publication, OR
- Before the end of the project funding period

**HOW** When making data available, investigators

- Cannot require the requestor(s) to establish a collaboration or co-authorship as a condition of accessing the data
- Cannot place limits on research questions asked, as long as research is consistent with informed consent
- Cannot limit research based overlap with other analyses

**WITH WHOM** Data should be available to the broad research community

# MODEL ORGANISM POLICY

<b>WHAT</b>	<p>New, genetically modified model organisms (non-human) and related resources generated using NIH grant funding must be shared.</p> <ul style="list-style-type: none"><li>• “Model organism” includes non-human mammalian models, non-mammalian models (<a href="#">periodically updated list is available</a>)</li><li>• “Resources” includes what materials/data are necessary for the production and understanding of model organisms</li></ul>
<b>WHERE</b>	<ul style="list-style-type: none"><li>• Maintained by investigator <u>OR</u></li><li>• Submitted to a repository</li></ul>
<b>WHEN</b>	<p>Generally at least upon publication of the primary results announcing the development of the genetically modified model organisms</p>
<b>HOW</b>	<p>NIH expectations for transferring to not-for-profit entities, or signing agreements with third parties generally</p>
<b>WITH WHOM</b>	<p>Model organism and resources should be distributed and shared with the broad research community</p>

# DIFFERENCES IN NIH DATA SHARING POLICIES

	2003 Data Sharing Policy	GDS Policy
<b>Data type</b>	Final research data (i. e. data used to address specific aims).	<ul style="list-style-type: none"> <li>Genomic data: All genomic data <u>generated by the grant</u>; previously-generated genomic data is not expected</li> <li>Phenotype data: Data pertinent to the interpretation of the expected genomic data, i.e. data need to reproduce primary analysis (regardless of how phenotype data collection was funded)</li> </ul>
<b>Data repository</b>	<ul style="list-style-type: none"> <li>NIH-supported repository (e.g., dbGaP) <u>OR</u></li> <li>Locally/within project (e.g. project database, archive, enclave)</li> </ul>	dbGaP and other approved NIH repositories (e.g. Genomic Data Commons)
<b>Data submission timeline</b>	PI should make data available to secondary researchers at whichever time comes first (regardless of whether data is shared locally or at NIH): <ul style="list-style-type: none"> <li>Primary manuscript is accepted for publication <u>OR</u></li> <li>Before the end of the project funding</li> </ul>	PI should submit data when the QA/QC is complete (i.e. data are clean and analytic dataset is finalized)
<b>Data release timeline</b>		dbGaP/NIH will release the data at whichever time comes first: <ul style="list-style-type: none"> <li>Six months after QA/QC is complete (i.e. when data submission begins) <u>OR</u></li> <li>At the time of publication</li> </ul>
<b>Consent requirements and sharing agreements</b>	For local sharing: <ul style="list-style-type: none"> <li>Local IRB standards for consent and data sharing apply, as consistent with Common Rule</li> <li>Local institution requirements for sharing data outside of institution apply, e.g. DUA or other agreements</li> </ul> For dbGaP sharing, GDS Policy expectations apply	<ul style="list-style-type: none"> <li>GDS Policy expects informed consent for future research uses and broad sharing of data, assured by the Institutional Certification</li> <li>Secondary research use is governed by GDS Policy access requirements / dbGaP Data Use Certification</li> </ul>

# WHAT DO I HAVE TO DO UNDER EACH POLICY/REQUIREMENT?

What do I include in my **grant application** / provide **prior to funding**?

What do I include in my **progress reports**?

# GRANT APPLICATION: RESOURCE SHARING PLAN

In grant's **Resource Sharing Plan**, investigators are expected to

- Affirm their intention to share consistent with all relevant NIH data and resource sharing policies/requirements
- State all NIH data/resource sharing policies relevant to the proposed grant application
- State whether they may need to request an exception to any NIH resource/data sharing policy, and the rationale for the request

Exceptions are rare, may require program leadership approval; GDS exception requires NCI Director approval

## ➤ Describe

- Extent and type of data that will be shared
- Method of sharing (e.g. data repository to which the data be submitted or how data will otherwise be made available, means for requesting data, conditions for accessing data, etc.)
- Timeline for the data to be shared
- Other policy- or FOA-specific requirements

Consistent with applicable policies and requirements

# GENOMIC DATA SHARING PLAN (GDSP)

## Elements of Genomic Data Sharing Plan:

1. Data Type (i.e. extent and type of data that will be shared)
2. Data Repository (i.e. method of sharing)
3. Data Submission and Release (Timeline)
4. IRB Assurance of the Genomic Data Sharing Plan (Policy-specific requirement)
5. Appropriate Uses of the Data

NCI GDSP template is available at <https://datascience.cancer.gov/sites/default/files/2019-02/nci-dsp.pdf>

NIH Guidance for Developing GDSP: [https://osp.od.nih.gov/wp-content/uploads/NIH\\_Guidance\\_Developing-GDS\\_Plans.pdf](https://osp.od.nih.gov/wp-content/uploads/NIH_Guidance_Developing-GDS_Plans.pdf)

NCI GENOMIC DATA SHARING PLAN

Data produced through this award will be shared in a manner consistent with data-sharing under the NIH Genomic Data Sharing Policy (NOT-OD-14-124).

**Intramural Project (Z01), Grant, or Contract number (if available):**  Enter project, grant or contract #

**Project Title:**  Enter project title

**Principal Investigator:**  Enter Principal Investigator

**Investigators Affiliation (Institution/Division/Program/Branch):**  Enter Institution

**1. Data to be Shared (mark all that apply):**

**Species:**  Human  Rat  
 Mouse  C. elegans  
 Drosophila  Yeast:  Indicate Species  
 Bacteria:  Indicate Species  Other:  Indicate Species

**Sample Type:**  Tumor Tissue  Normal Tissue  Blood  
 Buccal  Urine  Other:  Indicate sample type

**Analyte Type**  DNA  RNA  Other:  Indicate analyte type

**Genomic data (See NCI GDS framework for data sharing thresholds for each data type):**

<input type="checkbox"/> SNP array data from >500K single nucleotide polymorphisms (SNPs) <i>e.g., GWAS data</i>	<input type="text"/> Enter # of Samples
<input type="checkbox"/> DNA sequence data from < 100 genes or regions of interest (in general < 34Mb) <i>e.g., targeted sequencing</i>	<input type="text"/> Enter # of Samples
<input type="checkbox"/> DNA sequence data from ≥ 100 genes or regions of interest (in general > 34Mb) <i>e.g., targeted sequencing, whole exome sequencing, whole genome sequencing</i>	<input type="text"/> Enter # of Samples
<input type="checkbox"/> Genome-wide RNA sequencing (RNA-seq) data <i>e.g., transcriptomic data</i>	<input type="text"/> Enter # of Samples
<input type="checkbox"/> Genome-wide DNA methylation data <i>e.g., Illumina 450k or other platforms, bisulfite sequencing data</i>	<input type="text"/> Enter # of Samples
<input type="checkbox"/> Genome-wide chromatin immunoprecipitation sequencing (ChIP-seq) data <i>e.g., transcription factor ChIP-seq, histone modification ChIP-seq</i>	<input type="text"/> Enter # of Samples
<input type="checkbox"/> Metagenome (or microbiome) sequencing data <i>e.g., 16S rRNA sequencing, shotgun metagenomics, whole-genome microbial sequencing</i>	<input type="text"/> Enter # of Samples
<input type="checkbox"/> Metatranscriptome sequencing data <i>e.g., microbial/microbiome transcriptomics</i>	<input type="text"/> Enter # of Samples
<input type="checkbox"/> Other: <input type="text"/> List data types	<input type="text"/> Enter # of Samples

**Phenotype data:**

Data pertinent to the interpretation of genomic data, including the minimal phenotype information needed to reproduce the primary analysis —such as associated phenotype data (e.g., clinical information), exposure data, relevant metadata, and descriptive information (e.g., protocols or methodologies used)—will be shared. Individual-level Phenotype data will include, at minimum:

Describe the covariate information you will include with your data submission

NOTE: Investigators are not required to use the GDSP template as long as they provide a plan that covers all required elements

# 2003 DATA SHARING PLAN

*Content and level of detail to be included in a data-sharing plan depends on several factors, such as whether or not the investigator is planning to share data, the size and complexity of the dataset*

Further information is available at

[https://grants.nih.gov/grants/policy/data\\_sharing/data\\_sharing\\_guidance.htm#inc](https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#inc)

## Extent and type of data to be shared

- Format of the final dataset
- Documentation to be provided
- Whether or not any analytic tools also will be provided

## Method of sharing

- Mode of data sharing (e.g. data repository to which the data be submitted or how data will be made available locally or via project, means for requesting data)
- Whether or not a data-sharing agreement will be required; if so, a brief description (including criteria for determining who can access the data and whether any conditions will be placed on data use)

## Timeline

- Expected schedule for data sharing

# MODEL ORGANISM SHARING PLAN

*Sharing plans may vary depending on the organism, the nature of the resources that will be shared, the extent to which intellectual property issues may be considered in sharing, and the plans for distributing the resources*

Further information and samples of model organism sharing plans may be found at

[http://grants.nih.gov/grants/policy/model\\_organism](http://grants.nih.gov/grants/policy/model_organism)

**Specify how novel strains will be made available to the scientific community, including**

## Extent and type of resource to be shared

- Form in which the organisms will be provided
- Related research resources and data

## Timeline

- Reasonable time frame for periodic deposition of material and associated data

## Method of sharing

- Whether a repository will be used (if so, which one)

## Policy-specific requirements

- If relevant, how risks of infection or contamination will be minimized

**Specify how technology transfer and intellectual property issues will be handled, including:**

- How institution plans to make organisms and resources widely available to the research community
- How institution plans to ensure that any rights or obligations to 3rd parties are consistent with the terms and conditions of the NIH award to ensure appropriate dissemination of model organisms or reagents under the NIH award
- Description of mechanisms to distribute organisms and related research resources (e.g. MTAs).

# [NEW SLIDE] GRANT APPLICATION: BUDGET & BUDGET JUSTIFICATION

Preparing the data/resources for sharing may have associated costs.

The GDS Policy, 2003 Data Sharing Policy, and Model Organism Policy all allow investigators to include the cost of data/resource sharing in their budget.

Beyond these three policies, the investigators should check the policy or the FOA to confirm that costs for data sharing are permitted.

## Policy language permitting data sharing costs in budget

**GDS Policy:** Any resources that may be needed to support a proposed genomic data sharing plan (e.g., preparation of data for submission) should be included in the project's budget.

## 2003 Data Sharing Policy – Implementation

**Guidance:** Applicants may request funds in their application for data sharing. If funds are being sought, the applicant should address the financial issues in the budget and budget justification sections.

**Model Organism Policy – Brochure:** Investigators may request funds in their applications to defray reasonable costs associated with the sharing of model organisms and related research resources.

# PRIOR TO FUNDING — SPECIFIC TO GDS POLICY: JUST IN TIME & INSTITUTIONAL CERTIFICATION

## **GDS Policy: Institutional Certification (IC) submitted with Just-in-Time (JIT)**

IC provides official institutional assurance that data submission is consistent with requirements of GDS Policy

NIH GDS Institutional Certification templates are available at <https://osp.od.nih.gov/scientific-sharing/institutional-certifications/>

### Elements of Institutional Certification:

- *Consent*
- *IRB review*
- *Data submission is consistent with applicable national, tribal, and state law and regulations, and with relevant institutional policies*
- *Identities of research participants will not be disclosed to NIH-designated data repositories*
- *Data Use Limitation / Consent group: Permitted future research use of the data (as consistent with the informed consent)*

# PROGRESS REPORTS: DATA SHARING UPDATE

Research Performance Progress Report (RPPR) has section  
**C.5 OTHER PRODUCTS AND RESOURCE SHARING**

## Annual RPPR

Investigator should include any grant activities that indicate they are following data sharing plan, consistent applicable data sharing policies

- E.g. For GDS Policy, progress report should include data generation and QA/QC completion, whether data submission has begun, etc.

## Final RPPR

Investigator should describe how they have completed data sharing, in compliance with applicable data sharing policies

- E.g. For 2003 Data Sharing Policy, final report should include how data is made available to other researchers, process for obtaining data, etc.

# WHERE DO I GO FOR HELP?

## **PROGRAM DIRECTOR / PROGRAM OFFICIAL**

- Works with investigators throughout grant's project period
- Reviews grant application and RPPRs for consistency with data sharing policies and requirements
- Provides guidance or points investigators to appropriate resources

## **Specific to GDS Policy**

- Genomic Program Administrator (GPA) (Charlisse Caga-anan for DCCPS, Natalie Abrams and Claire Zhu for DCP)
  - Point of contact for GDS policy within division
  - Provides guidance to PD/POs, registers projects in dbGaP in preparation for data submission
  - Works with investigator and dbGaP staff
- dbGaP Staff
  - Works with investigators to process submitted data for controlled access release

# ONLINE RESOURCES

<b>GDS POLICY</b>	NIH Office of Science Policy Main Page: <a href="https://osp.od.nih.gov/scientific-sharing/genomic-data-sharing/">https://osp.od.nih.gov/scientific-sharing/genomic-data-sharing/</a> NIH OSP FAQs: <a href="https://osp.od.nih.gov/scientific-sharing/genomic-data-sharing-faqs/">https://osp.od.nih.gov/scientific-sharing/genomic-data-sharing-faqs/</a> NCI Office of Data Sharing: <a href="https://datascience.cancer.gov/data-sharing/genomic-data-sharing/about-the-genomic-data-sharing-policy">https://datascience.cancer.gov/data-sharing/genomic-data-sharing/about-the-genomic-data-sharing-policy</a>
<b>2003 DATA SHARING POLICY</b>	NIH Main Page: <a href="https://grants.nih.gov/grants/policy/data_sharing/">https://grants.nih.gov/grants/policy/data_sharing/</a> NIH FAQ: <a href="https://grants.nih.gov/grants/policy/data_sharing/">https://grants.nih.gov/grants/policy/data_sharing/</a> NIH Data Sharing Policy and Implementation Guidance: <a href="https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm">https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm</a>
<b>MODEL ORGANISM POLICY</b>	NIH Main Page: <a href="https://grants.nih.gov/grants/policy/model_organism/">https://grants.nih.gov/grants/policy/model_organism/</a> NIH FAQ: <a href="https://grants.nih.gov/grants/policy/model_organism/model_organisms_faqs.htm">https://grants.nih.gov/grants/policy/model_organism/model_organisms_faqs.htm</a> NIH Brochure: <a href="https://grants.nih.gov/grants/policy/model_organism/model_organism_brochure.pdf">https://grants.nih.gov/grants/policy/model_organism/model_organism_brochure.pdf</a>

**QUESTIONS?**

---

[charlisse.caga-anan@nih.gov](mailto:charlisse.caga-anan@nih.gov)

