SOP#: RPS-23

Registering a Clinical Trial in dbGaP

Version #: 3.0

Next Review Date: 08/2025

Approved Date: 08/2023

Review Interval Period: Biennial

NCI Clinical Director Signature/ Effective Date:

James Gulley Digitally signed by James Gulley Date: 2023.08.30 12:07:52 -04'00'

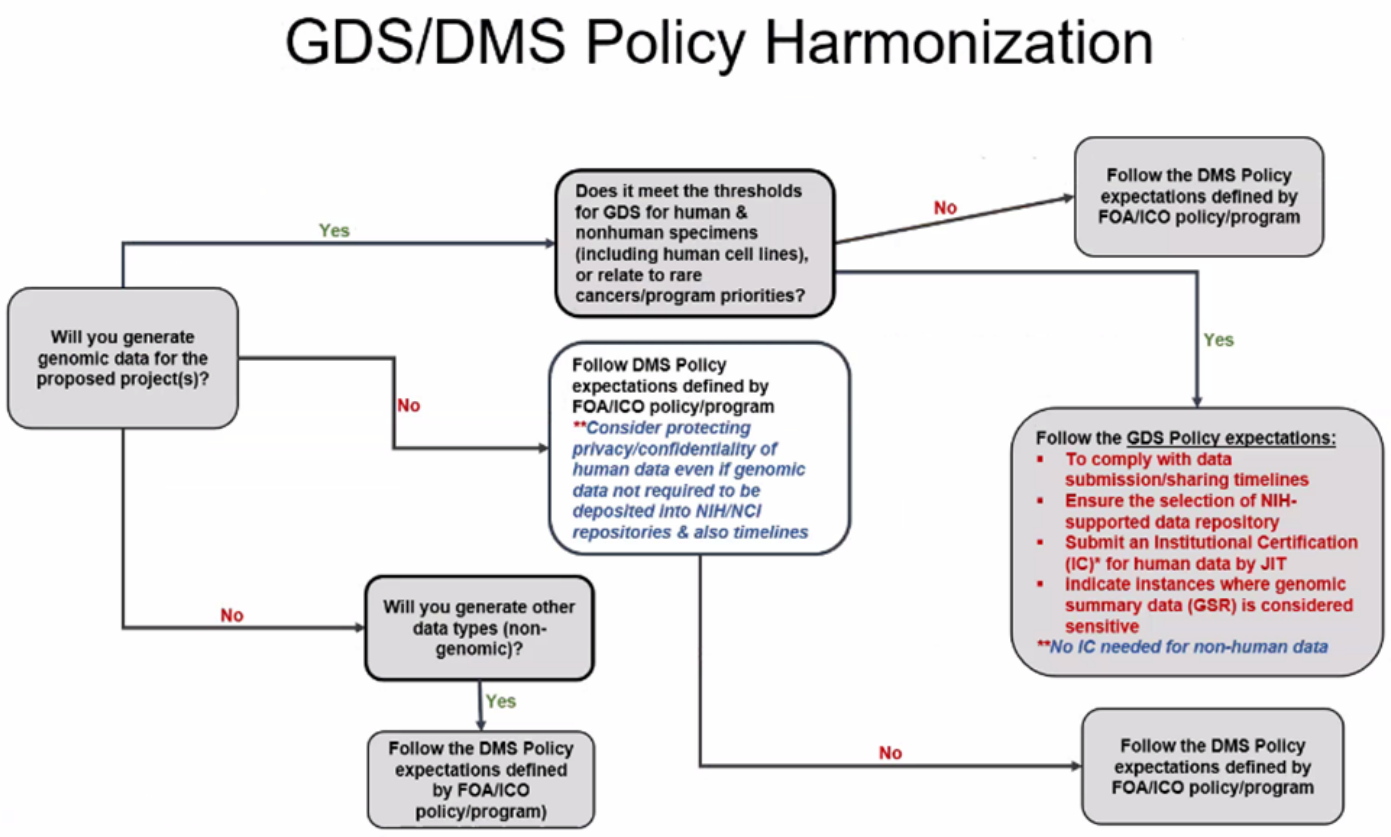
POLICY

The Data Management and Sharing (DMS) Policy applies to all NIH intramural research and was effective January 25, 2023. The Office of Intramural Research has specified that DMSP applies to research associated with a:

- ZIA (human and non-human research)
- **Clinical protocols that will undergo IC Initial Scientific Review**
- Genomic Data Sharing (GDS) project

The DMSP requires sharing *scientific data,* defined as data of sufficient quality to validate and replicate the research findings[1]. For additional information on the DMS Policy, please refer to the Center for Cancer Research (CCR) DMS Policy Home Page.

The DMS Policy has been harmonized with the Genomic Data Sharing (GDS) Policy (see Figure 1[2]).



The Genomic Data Sharing (GDS) Policy applies to all NIH intramural research that generates large-scale human or non-human genomic data and uses these data for subsequent research. Large-scale data includes but is not limited to the following:

- Genome-wide association studies (GWAS),
- Single nucleotide polymorphisms (SNP) arrays, and
- Genome sequence,
- Transcriptomic,
- Metagenomic,
- Epigenomic, and
- Gene expression data.

For additional information on the policy, please refer to the CCR DMS Home Page. To also assist you in determining which Data Sharing Policies apply to your research, the NIH Office of Science Policy has created an online tool, What Policies Apply to My Research. This approach involves answering a few questions about your research, which you submit to obtain information about which policy or policies apply to your research.

Investigators should work with their designated Genomic Program Administrator (GPA) to register all studies with human genomic data that fall within the scope of the GDS Policy in the database of Genotypes and Phenotypes (dbGaP).

****For clinical trials, the study registration process will start once data for the project have been generated, processed, and checked.**

PURPOSE

To provide instructions on registering a Clinical Trial in dbGaP.

RESOURCES

- Database of Genotypes and Phenotypes (dbGaP) portal
- CCR Data Management and Sharing Project Home (Wiki Page)
- NIH Genomic Data Sharing Policy
- NIH Genomic Data Sharing Website

CCR Genomic Program Administrator (GPA)
Kathleen Calzone, PhD, RN, AGN-BC, FAAN
240-760-6178
calzonek@mail.nih.gov

CCR Genomic Program Administrator (GPA)

Abid Al Reza, PhD
240-858-7909
abid.reza@nih.gov

PROCEDURES

STEP 1: Study Registration Process

Once data for your project have been generated, processed, and checked, contact the GPA via email, informing them that you would like to register your study with dbGaP. Attach an electronic copy of the signed Institutional Certification (IC) Memo.

NOTE: You should register your study even if you have applied for and have been granted a Data Management and Sharing Waiver.

- Complete the required information and email the following to the GPA:
  - dbGaP study configuration template and
  - dbGaP Basic Study Information Sheet

  Example templates of both are included in the Appendices.

NOTE: Please do not submit incomplete forms, which will delay your study registration process in dbGaP; the system does not allow for partially completed registrations to be saved.

- GPA will:
  - Ensure that forms are complete and accurately filled out
  - Register the study in the dbGaP portal
  - Invite PI to review the registered study
  - Invite PI/team-assigned data submitter to the study

STEP 2: Confirmation of Accuracy

- After the completion of the registration process, an email invitation from dbGaP will be sent to the PI to review the study information.
- PI reviews the information for completeness and accuracy. For any questions or concerns on any aspect of the registered study, or if there are any errors in the registration data, please email the GPA for help.

REFERENCES:

1. 2023 Data Management and Sharing Policy: Policy Scope https://sharing.nih.gov/faqs#/data-management-and-sharing-policy.htm?anchor=56772, accessed 6/23/2023.
2. NCI Office of Data Sharing Tactic meeting presentation held on May 2023. Author: Emily Boja.

Appendix A

Study Config Template for Web Form

See example: **https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000001.v3.p1**

Study Name: The Study Name will stay stable between study versions. Use title case. **https://www.ncbi.nlm.nih.gov/gap/** **Example: NEI Age-Related Eye Disease Study (AREDS)**

Study Description: Provide a brief but detailed original summary of the design and aims, population information, molecular technologies employed, principal findings of the study, and data that will be available through dbGaP. We prefer the description not be taken verbatim from a published article, but if it is, please submit copyright permission from the Journal. Summaries with copyrighted material must include the following within the description: "Reprinted from [Article Citation], with permission from [Publisher].

Study Inclusion and Exclusion Criteria: if applicable.

Study History: if applicable.

Study Design: One study design using the standardized list below:

- Case Set
- Case-Control
- Clinical Genetic Testing
- Clinical Trial
- Collection
- Control Set
- Cross-Sectional
- Family/Twin/Trios
- Interventional
- Mendelian
- Metagenomics
- Prospective Longitudinal Cohort
- Tumor vs. Matched-Normal
- Xenograft
- Methods

Study Types: One or more study types. For example: Longitudinal, Case-Control, Case Set, Control Set, Parent-Offspring Trios, Cohort, etc.

Study Web Links: if applicable.

Clinical Trials: ClinicalTrials.gov Identifier for this study if applicable. For example, NCT01274598. The final link that will be on the public page will be https://clinicaltrials.gov/ct2/show/NCT01274598.

Primary Phenotypic Term and Additional Phenotypic MeSH Terms: Any number of phenotype terms and/or disease names associated with this study. The phenotype term and disease name must be a MeSH term. To check, type query in the search box at https://www.ncbi.nlm.nih.gov/mesh/. Disease names will be ordered as submitted below. Please mark one MeSH term as the primary phenotype term with a star (*). Please note the MeSH term may not be not be an exact match to the term you have in mind, but is primarily used as a filter feature in dbGaP Advanced Search. If you would like a specific term to be indexed and searchable, we suggest including the term in the Study Description, Inclusion/Exclusion, or history sections. Also, variable names and descriptions are indexed and searchable.

1.
2.
3.

Gene Names: List human genes with significant results for your study. For example, APOE. To find gene symbols, please type query in the search box at NCBI Gene database, https://www.ncbi.nlm.nih.gov/gene. Genes will be ordered as submitted below.

References with PubMed IDs (PMIDs) Section: Provide the PMIDs using https://pubmed.ncbi.nlm.nih.gov/ for publications describing the study data submitted. For example, 17903304.

References without PMIDs Section: If the article, abstract, or book does not have a PMID or is waiting for a PMID, provide the reference in the following formats below. Please include one of the following [Submitted, In Press, Accepted, In Revision] in the "Journal or Book" section if a PMID is pending.

For Journals:
Authors Authors (Last Name FM)
Title Title of Article or Abstract
Journal or Book Journal Name. Year;Volume:Start page-End page.

For Books:
Authors Authors (Last Name FM)
Title Chapter or Section Name
Journal or Book Book Name. Editors. Publication City: Publisher, Year.

Authors	Title	Journal or Book

Study Attributions: will appear as submitted.
Column1 Role in Study (i.e., Principal Investigator, Co-Investigator, Sequencing Center)
Column2 Name of the person or Name of Grant
Column3 Affiliation or Institute (include City, State, Country)
Example: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000001.v3.p1

For updates to the **Acknowledgement Statement, please contact your GPA directly, who will update the dbGaP Submission System for your study. The curator does not have control of that section.

Role in Study	Name	Affiliation or Institute
Principal Investigators	John Doe, PhD	National Institutes of Health, Bethesda, MD, USA
Principal Investigators	Jane B. Doe, MD, PhD	National Institutes of Health, Bethesda, MD, USA
Funding Source	R01#####	National Institutes of Health, Bethesda, MD, USA

Appendix B

NCI dbGaP Data Submission Information

(Basic Study Information form)

In order for the **NIH NATIONAL CANCER INSTITUTE** to register your data into the dbGaP Submission System, please provide the information listed below and return to your NIH Program Officer (PO), or intramural Genomic Program Administrator (GPA). You may use the sample documents or any other format.

Checklist for required documents:
☐Institutional Certification
☐dbGaP Data Submission Information

PART I – Study Registration Information

Study name:		
Is this a multi-center study? **If YES, please list participating sites:**	◯ Yes ◉ No	
Target data delivery date: (YYYY-MM-DD)	Target public release date: (YYYY-MM-DD)	
Estimated number of study participants:		
Is data submission expected to the following repositories? (Description of repositories)	Sequence Read Archive (SRA): (Y/N)	Trusted Partner (e.g. Bionimbus, GDC) (Y/N/NA)

PART II – Principal Investigator (PI) and Funding Information

PI name:	PI e-mail:
PI institution:	
PI assistant/submitter name:	PI assistant/submitter e-mail:
Do you have an eRA Commons or an NIH account? (Y/N) ◯ Yes ◉ No If **YES**, go to next question. If **NO**, please register at https://commons.era.nih.gov/commons/registration/registrationInstructions.jsp.	
NIH Intramural Project (Z01), Grant or Contract Number:	NIH PO (if applicable):
NIH Institutes/Centers supporting the study:	

PART III – Policy

Do you have Institutional Certification (IC) to submit these data? ◯ Yes ◉ No

The IC should include the **Data Use Limitations (DUL)**, which are based on the informed consent given by each research subject. For every research subject, his/her corresponding data will be tagged with the appropriate DUL. Each study may have multiple DULs, based on the informed consent in the study.

If **YES**, send attachment to NIH PO or intramural GPA, along with this form.

If **NO**, please obtain the Institutional Certification from your Institutional Official. dbGaP requires that the sponsoring IC verifies that this certification has been met. A description of the requirements for the Institutional Certification and an example may be found in the accompanying "Submission into the NIH Database of Genotypes and Phenotypes (dbGaP)" guide.

PART IV – Study Description

Study type(s) (e.g., longitudinal, case-control, case set, control set, parent-offspring trios, cohort):

Is aggregate-level data appropriate for General Research Use? ◯ Yes ◉ No

If **YES**, aggregate data will be included in the Compilation of Aggregate Genomic Data, a collection of analyses across many dbGaP studies that can be accessed with a single Data Access Request.
NOTE: This should be consistent with the Institutional Certification

Samples genotyped/sequenced:

Please check all data types expected for this study:

General	**Sample Types**	**Array Data**
☐Individual Phenotype	☐Germline	☐SNP Array
☐Individual Genotype	☐Tumor/Normal	☐Expression Array
☐Individual Sequencing	☐DNA	☐Methylation Array
☐Supporting Documents	☐RNA	
☐Metagenomic	☐Mitochondria	
☐Protomic/Metabolomic	☐Microbiome	
☐Images	☐From Repository	
Genotypes	**Sequencing**	**Analyses**
☐Array derived Genotypes	☐Whole Genome	☐Association/Linkage Results
☐CNV calls from miroarray	☐Whole Exome	☐Array derived Expression
☐CNV calls derived from Sequencing	☐Targeted Genome	☐RNA Seq derived Expression
☐Genotype calls derived from Sequence	☐Targeted Exome	☐Array derived Methylation
☐Somatic SNV (.MAF)	☐Whole Transcriptome	
☐Array CGH CNVs	☐Targeted Transcriptome	
	☐Epigenomic Marks	
	☐Sanger	
	☐16S rRNA	

Genotype/Sequence platform information

Name and version	**Vendor**	**# Probes**	**URL**	**Description (optional)**
Example: [GenomeWideSNP_6] Affymetrix Genome-Wide Human SNP 6.0 Array	*Affymetrix*	*1880794*	*http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL6801*	

PART V – Acknowledgement Statement(s)* **
The submitting PI should provide specific points that should be included in an acknowledgement, such as sources of support or collaborators who have made subjects or samples available. Any NIH support must be specifically acknowledged by including the grant number. Consider citing a specific publication that comprehensively describes the origin of the dataset. The suggested Acknowledgement Statement to accompany the dataset is: Click here to enter text.
PART VI – Original Summary of Study
Please provide an original description of the study.[1] Click here to enter text.

[1] If the submitting institution certifies that aggregate data from a project can be included in the Compilation, then a study description for the aggregate data should be provided in addition to a description for the individual data.