American Society for Investigative Pathology

Investigating the Pathogenesis of Disease

NIH Final Genomic Data Sharing Policy – Frequently Asked Questions

These FAQs have been developed by ASIP to assist researchers in interpreting NIH's new Genomic Data Sharing Policy (GDSP). The FAQs reflect ASIP's interpretation of the guidelines as of the date issued and provide a general framework for considering the impact of the new GDSP. ASIP does not agree with all components of the new GDSP; however, it is providing this interpretation to support scientific research and to facilitate compliance of ASIP members with the new requirements. Institutions may establish additional requirements beyond what is set forth in the GDSP. Consultation with your institution and/or your IRB may be appropriate. These FAQs should not be construed as establishing an official interpretation, legal opinion, or regulatory guidance regarding the GDSP. Researchers and their institutions are advised to review the specific facts in a given situation to determine an appropriate course of action, which might include consultation with an Institutional Review Board or with the funding Institute or Center.

INDEX of FAQs (with hyperlinks to FAQs in the document):

FAQs

I. SCOPE

1. What types of genomic data should be reported under the GDSP?

The GDSP applies to large-scale human or nonhuman genomic data including genome-wide association studies (GWAS), single nucleotide polymorphism (SNP) arrays, and genome sequence, transcriptomic, metagenomic, epigenomic, and gene expression data.

NIH has provided Supplemental Information on the GDSP.

Examples of data that must be submitted include:

- sequence data from more than one gene or region of comparable size in the genomes of more than 1,000 human research participants;
- sequence data from more than 100 genes or region of comparable size in the genomes of more than 100 research participants;
- data from 300,000 or more variant sites in more than 1,000 human research participants;
- sequence data from more than 100 isolates from infectious organisms;
- sequence data from more than 100 meta-genomes of human or model organism microbiomes;
- sequence data from more than 100 metatranscriptomes of human or model organism microbiomes;
- 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;
- comparisons of genome-wide methylated sites across more than 10 cell types; or
- 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).

Specifically excluded from this policy are:
- proof of concept or exploratory research;
- instrument calibration exercises;
- statistical or technical methods development; or
- use of genomic data for controls for purposes such as assay development.

If unsure whether the GDSP applies, contact the appropriate Program Official or Project Officer of the relevant Institute or Center to discuss. Individual Institutes or Centers have the ability to modify these standards, requiring inclusion of data or providing exemptions from data submission requirements.

2. What types of grants are impacted by the GDSP?

The GDSP applies to activities requesting support for research, including the following types of grants:
- Research project grants (Rs);
- Program projects (Ps) and SCORs (Ss);
- Cooperative agreements for research (Us); and
- Individual career development awards (Ks) that include a research component.

The GDSP specifically excludes the following activities:
- Institutional training grants (T32s, T34s, T35s & TL2s);
- K12 career development awards (KL2s);
- Individual fellowships (Fs); and
- Resource grants and contracts (Ss).

For specific questions regarding whether a project falls under the purview of the GDSP, please review both the Implementation Notice and the Supplemental Policy .

3. When does the GDSP go into effect?

The GDSP applies to competing grant applications submitted for the January 25, 2015 receipt date or subsequent receipt dates, as well as proposals for contracts that are submitted to the NIH on or after January 25, 2015. For NIH intramural research projects that generate genomic data, this policy applies to genomic data generated on or after January 25, 2015.

4. Can an Institute or Center request something different from what is in the GDSP?

Yes, NIH Institutes or Centers may expect submission of data from smaller scale research projects based on the state of the science, the programmatic priorities of those funding the research and the utility of

the data to the research community. They may also provide exemptions from data submission requirements. Each Institute or Center should have a Genomic Program Administrator (GPA) available to answer questions.

5. **NIH is not funding any of my research. Do I still need to submit my data? What if NIH is funding only part of my research?**

If research is not funded by the NIH, the GDSP does not apply. If NIH funds the part of a research project that generates genomic data, the GDSP applies.

II. IMPACT ON RESEARCH PROPOSALS

6. **When do I submit my data-sharing policy and will it be used to evaluate my research proposal?**

Funding applications and proposals should include, in the Resource Sharing Plan section, basic plans for following the GDSP. The minimum components of the data-sharing plan are outlined in: (a) the NIH Guidance for Institutions Submitting Grant Applications and Contract Proposals under the NIH Genomic Data Sharing Policy for Human and Non-Human Data; and (b) the NIH Guidance for Investigators in Developing Genomic Data Sharing Plans. A more detailed genomic data-sharing plan should be provided to the funding Institute or Center prior to award, along with any other just-in-time information.

Reviewers will be asked to comment on the data-sharing plan during peer review. Currently, reviewers will not factor the plan into the Overall Impact score, unless specified in the Funding Opportunity Announcement. Institutes or Centers may take the breadth of data sharing permitted by the consent into consideration during program priority review.

7. **This is expensive. Can I include the cost of data sharing in my research budget?**

Yes, any appropriate resources needed to comply with the GDSP may be included in the budget.

8. **What is an Institutional Certification and how does it relate to the GDSP?**

A responsible official of the submitting research institution must certify, prior to award, that the data will be submitted to either an unrestricted or controlled access database. This certification includes providing assurance that data submission is consistent with the informed consent documents; in compliance with national, tribal, and state laws and regulations as well as institutional policies; and an Institutional Review Board (IRB) or equivalent body has overseen various aspects of compliance. For specific information regarding what will be contained in the Institutional Certification, please review the regulations and Section I.c. of Guidance for Institutions Submitting Grant Applications and Contract Proposals under the NIH Genomic Data Sharing Policy for Human and Non-human Data.

9. **I started my research before the GDSP went into effect. What impact might this policy have on my ongoing research?**

The GDSP does not apply to previously unfunded research proposals submitted prior to the Policy's effective date (January 25, 2015). When submitting the research performance progress report, investigators will likely be asked to provide an updated genomic data-sharing plan. If the current

consent being used in the research does not comply with the GDSP, investigators should plan to transition to a consent that allows for future research uses and broad data sharing, if possible.

III. DONOR CONSENT & BIOETHICAL ISSUES

10. How will this Policy impact donor consent forms?

For human studies submitted after the effective date of the Policy, investigators are expected to obtain participant consent allowing genomic and phenotypic data to be used for future research studies and to be shared broadly. The consent should specify whether the data will be held in an unrestricted repository or a controlled access repository. Even though data submitted to the repositories will be de-identified, NIH suggests that the risk of re-identification be explained to donors as part of the consent process.

For studies using genomic data from cell lines or clinical specimens that were created or collected after the effective date of the policy (January 25, 2015), consent for future research and broad data sharing should be obtained even if the cell lines or clinical specimens are de-identified. If there are compelling scientific reasons requiring the use of genomic data from cell lines or clinical specimens created or collected after the effective date of this policy and lacking consent for research use and data sharing, justification should be provided in the researcher's funding request.

11. The research that I do is considered exempt from human subjects research as I use de-identified clinical specimens and cell lines. How will this policy impact that research?

Donor consent should be obtained for research proposals using de-identified clinical specimens and cell lines created after the policy implementation date (January 25, 2015) and yielding results that should be submitted under the GDSP. NIH believes that as genomic technology advances, more sophisticated analytic methods raise the risk of re-identification. Specifically, NIH states that "given growing concerns about re- identification, it is no longer ethically tenable simply to de-identify clinical specimens or derived cell lines generate data for research use without an individual's consent" (Federal Register Vol. 79, No. 167, August 28, 2014, p. 51348).

According to the GDSP, consent for secondary research is generally not required for de-identified clinical specimens or cell lines created or collected before the effective date of the GDSP. Consider consulting your institution or IRB as either may have a more restrictive policy.

12. Do I have to re-consent donors donating specimens before the GDSP went into effect?

There is no current requirement to re-consent donors of specimens obtained prior to the implementation of the GDSP. If questions arise, IRBs may assist in determining the appropriate course of action.

13. What about research on populations that are unwilling to share data for secondary research?

The funding Institute or Center may consider exceptions from the GDSP on a case-by-case basis provided that a justification is submitted in the research proposal.

14. How will confidentiality of research participant data be honored? What is the difference between open and closed access and how can I determine which is appropriate?

The informed consent under which the samples were collected determines whether the data should be available through unrestricted or controlled access. Submitting institutions, through their IRBs, review informed consent materials to determine whether data can be shared for secondary research use and what level of access is appropriate. IRBs may consider the type of study and whether the data are obtained through prospective or retrospective data collection.

15. I am concerned about how incidental results will be handled and whether the genomic database raises additional concerns about incidental results that have not been clinically validated. What is NIH's stance on this issue?

Neither the NIH repository nor researchers using the data for secondary research will have access to the donor's identity. Therefore, neither NIH nor secondary researchers will be able to return individual research results directly to a participant. NIH encourages submitting institutions to work with their IRB to determine appropriate standards around the return of individual incidental findings from research studies where the principal investigator can determine the subject's identity.

IV. DATA SUBMISSION

16. Does NIH have different standards for the submission of nonhuman genomic data and human genomic data?

Yes, different standards apply to nonhuman and human genomic data. NIH has created an information table within a Supplement to clarify data submission and release (page 3 of the Supplemental Information).

17. When do I have to submit my data?

NIH has provided a timeline of data submission requirements within a supplement (page 3 of the Supplemental Information).

18. What are the standards to achieving data de-identification?

Prior to submitting information to the data repository, researchers should delete data identifiers according to the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. The now de-identified data should be assigned random, unique codes with the key to be held by the submitting institution.

19. What data repositories may I use to submit nonhuman data?

For nonhuman genomic data, NIH allows the use of many widely used data repositories, whether NIH funded or not. Examples of approved entities include: Gene Expression Omnibus (GEO); Sequence Read Archive (SRA); Trace Archive; Array Express; Mouse Genome Informatics (MGI); WormBase; the Zebrafish Model Organism Database (ZFIN); GenBank; European Nucleotide Archive (ENA); or DNA Data Bank of Japan (DDBJ).

20. What data repositories may I use to submit human data?

All studies with human genomic data and falling under the GDSP should be registered in dbGaP by the time data cleaning and quality control measures begin, regardless of when the dbGaP repository will receive the data. After registration in dbGaP, investigators should submit the data to the relevant NIH designated data repository (e.g., dbGaP, GEO, SRA, the Cancer Genomics Hub). Investigators may choose to submit data to a non-NIH designated data repository in addition to an NIH designated data repository.

Funding Institutes or Centers may grant exceptions to submitting relevant data and require that the investigator develop an alternative plan to share data through other mechanisms. If exceptions are granted, studies must still be registered with dbGaP and the reason for the exception should be provided.

21. What if I'm going to use specimens gathered before the effective date of the policy and want to submit data to a repository?

For studies using data from specimens collected before the effective date of the GDSP, an assessment by an IRB, privacy board or equivalent body is needed to ensure that data submission is consistent with the informed consent provided by the research participant. Guidance from an IRB may also offer assistance in determining whether data will be submitted under controlled or open access.

NIH will accept (but not require) data derived from de-identified cell lines or clinical specimens lacking consent for research created or collected before the effective date of this policy.

22. What if some specimens were gathered earlier and some gathered after January 25, 2015?

If there is a mix of specimens, some gathered before and some after the January 25, 2015 date, any concerns regarding consent should be clearly outlined for the granting Institute or Center in advance of award. An IRB, privacy board or other entity may provide guidance.

23. What if some of the specimens are from donors that are now deceased?

The GDSP applies regardless of whether the research is conducted on data generated from deceased individuals.

24. What are the consequences if data are not submitted?

Adherence to the GDSP is a special term and condition in the Notice of Award or the Contract Award. Failure to comply may lead to enforcement actions, including the withholding of funding.

V. ACCESSING THE DATA FOR SECONDARY RESEARCH

25. When will NIH release my data for access?

Timing for release of data varies by the type of data. Please see page 3 of the Supplemental Information

for additional information regarding data availability.

26. How can I access this database to conduct secondary research?

The repository holding the data will establish its policy for data access. At a minimum, researchers wishing to download unrestricted access data should agree that they will:

- not attempt to identify individual human research participants from whom the data were obtained; and
- provide acknowledgment in all presentations or publications of the specific data set and the data repository through which the investigator obtained the information.

Researchers wishing to access controlled information agree to abide by the NIH Genomic Data User Code of Conduct. Researchers submit a data access request and the inquiry for controlled access data will be reviewed by one of NIH's Data Access Committees. Researchers must sign a Data Use Certificate agreeing to comply with the NIH Genomic Data User Code of Conduct, including the following:

- using the data only for approved research;
- protecting data confidentiality;
- not attempting to identify individual participants from whom the data were obtained;
- not sharing any of the data obtained through controlled access repository; and
- providing specific acknowledgments in all presentations or publications.

Access to a data set is typically limited to one year, with the possibility of extension.

27. How will making data available for secondary use impact my intellectual property rights?

Basic sequence data and related information (e.g., genotypes, haplotypes, *p*-values, allele frequencies) are considered pre-competitive and thus available for release to NIH designated data repositories. NIH requires that data made available through NIH designated data repositories, and all conclusions derived directly from them, should remain freely available without any licensing requirements.

VI. RESOURCES USED TO PREPARE THESE FAQs:

1. Final NIH Genomic Data Sharing Policy. Also available in the Fed Reg. Vol. 79, No. 167, August 28, 2014. Pgs. 51345-51354.

2. Summary of comments on the draft NIH Genomic Data Sharing Policy.

3. Supplemental information to the National Institutes of Health Genomic Data Sharing Policy.

4. Implementation of the NIH Genomic Data Sharing Policy for NIH grant applications and awards (NOT-OD-14-11).

5. Guidance for Institutions Submitting Grant Applications and Contract Proposals under the NIH Genomic Data Sharing Policy for Human and Non-human Data

6. NIH Guidance for Investigators in Developing Genomic Data Sharing Plans