I. Purpose

The National Institutes of Health (NIH) Genomic Data Sharing (GDS) Policy sets forth expectations that ensure the broad and responsible sharing of genomic[1] research data. Sharing research data supports the NIH mission[2] and is essential to facilitate the translation of research results into knowledge, products, and procedures that improve human health. NIH has longstanding policies to make data publicly available in a timely manner from the research activities that it funds.[3,4,5,6,7]

II. Scope and Applicability

The GDS Policy applies to all NIH-funded research that generates large-scale human or non-human genomic data as well as the use of these data for subsequent research. Large-scale data include genome-wide association studies (GWAS),[8] single nucleotide polymorphisms (SNP) arrays, and genome sequence,[1] transcriptomic, metagenomic, epigenomic, and gene expression data, irrespective of funding level and funding mechanism (e.g., grant, contract, cooperative agreement, or intramural support). The *Supplemental Information to the NIH Genomic Data Sharing Policy* (Supplemental Information)[9] provides examples of research projects involving large-scale genomic data that are subject to the Policy. NIH Institute or Centers (IC) may expect submission of data from smaller scale research projects based on the state of the science, the programmatic priorities of the IC funding the research, and the utility of the data for the research community.

At appropriate intervals, NIH will review the types of research to which this Policy may be applicable, and any changes to examples of research that are within the Policy's scope will be provided in the Supplemental Information. NIH will notify investigators and institutions of any changes through standard NIH communication channels (e.g., *NIH Guide for Grants and Contracts*).

---

[1] The genome is the entire set of genetic instructions found in a cell. See https://www.genome.gov/glossary/index.cfm?id=90.

[2] NIH's mission is to seek fundamental knowledge about the nature and behavior of living systems and the application of that knowledge to enhance health, lengthen life, and reduce illness and disability. See http://www.nih.gov/about/mission.htm.

[3] Final NIH Statement on Sharing Research Data. February 26, 2003. See http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html.

[4] NIH Intramural Policy on Large Database Sharing. April 5, 2002. See http://sourcebook.od.nih.gov/ethic-conduct/large-db-sharing.htm.

[5] Reaffirmation and Extension of NHGRI Rapid Data Release Policies: Large-scale Sequencing and Other Community Resource Projects. February 2003. See http://www.genome.gov/10506537.

[6] NIH Policy on Sharing of Model Organisms for Biomedical Research. Release Date May 7, 2004. See https://grants.nih.gov/grants/policy/model_organism/.

[7] NIH Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS). See http://grants.nih.gov/grants/guide/notice-files/NOT-OD-07-088.html.

[8] GWAS has the same definition in this policy as in the 2007 GWAS Policy: a study in which the density of genetic markers and the extent of linkage disequilibrium should be sufficient to capture (by the r^2 parameter) a large proportion of the common variation in the genome of the population under study, and the number of samples (in a case-control or trio design) should provide sufficient power to detect variants of modest effect.

[9] See https://osp.od.nih.gov/wp-content/uploads/Supplemental_Info_GDS_Policy.pdf.

NIH expects all funded investigators to adhere to the GDS Policy, and compliance with this Policy will become a special term and condition in the Notice of Award or the Contract Award. Failure to comply with the terms and conditions of the funding agreement could lead to enforcement actions, including the withholding of funding, consistent with 45 CFR 74.62[10] and/or other authorities, as appropriate.

III. Effective Date

This Policy applies to:

- Competing grant applications[11] that are submitted to NIH for the January 25, 2015, receipt date or subsequent receipt dates;
- Proposals for contracts that are submitted to NIH on or after January 25, 2015; and
- NIH intramural research projects generating genomic data on or after January 25, 2015.

IV. Responsibilities of Investigators Submitting Genomic Data

A. Genomic Data Sharing Plans

Investigators seeking NIH funding should contact appropriate IC Program Official or Project Officer[12] as early as possible to discuss data sharing expectations and timelines that would apply to their proposed studies. NIH expects investigators and their institutions to provide basic plans for following this Policy in the "Genomic Data Sharing Plan" located in the Resource Sharing Plan section of funding applications and proposals. 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. A more detailed genomic data sharing plan should be provided to the funding IC prior to award. The Institutional Certification (for sharing human data), should also be provided to the funding IC prior to award, along with any other Just-in-Time information. NIH expects intramural investigators to address compliance with genomic data sharing plans with their IC scientific leadership prior to initiating applicable research and are encouraged to contact their IC leadership or the Office of Intramural Research for guidance. The funding NIH IC will typically review compliance with genomic data sharing plans at the time of annual progress reports or other appropriate scientific project reviews, or at other times, depending on the reporting requirements specified by the IC for specific programs or projects.

[10] 45 CFR 74.62. Uniform Administrative Requirements for Awards and Subawards to Institutions of Higher Education, Hospitals, Other Nonprofit Organizations, and Commercial Organizations; Enforcement. See https://www.gpo.gov/fdsys/pkg/CFR-2000-title45-vol1/pdf/CFR-2000-title45-vol1-sec74-62.pdf

[11] Competing grant applications encompass all activities with a research component, including but not limited to the following: Research Grants (Rs), Program Projects (Ps), Cooperative Research Mechanisms (Us), Career Development Awards (Ks), and SCORs and other S grants with a research component.

[12] Investigators should refer to funding announcements or IC websites for contact information.

B. Non-human Genomic Data

1. Data Submission Expectations and Timeline

Large-scale non-human genomic data, including data from microbes, microbiomes, and model organisms, as well as relevant associated data (e.g., phenotype and exposure data), are to be shared in a timely manner. Genomic data undergo different levels of data processing, which provides the basis for NIH's expectations for data submission. These expectations are provided in the Supplemental Information. In general, investigators should make non-human genomic data publicly available no later than the date of initial publication. However, earlier availability (i.e., before publication) may be expected for certain data or IC-funded projects (e.g., data from projects with broad utility as a resource for the scientific community such as microbial population-based genomic studies).

2. Data Repositories

Non-human data may be made available through any widely used data repository, whether NIH-funded or not, such as the Gene Expression Omnibus (GEO)[13] Sequence Read Archive (SRA),[14] Trace Archive,[15] Array Express,[16] Mouse Genome Informatics (MGI),[17] WormBase,[18] the Zebrafish Model Organism Database (ZFIN),[19] GenBank,[20] European Nucleotide Archive (ENA),[21] or DNA Data Bank of Japan (DDBJ).[22] NIH expects investigators to continue submitting data types to the same repositories that they submitted the data to before the effective date of the GDS Policy (e.g., DNA sequence data to GenBank/ENA/DDBJ, expression data to GEO or Array Express). Data types not previously submitted to any repositories may be submitted to these or other widely used repositories as agreed to by the funding IC.

C. Human Genomic Data

1. Data Submission Expectations and Timeline

Investigators should submit large-scale human genomic data as well as relevant associated data (e.g., phenotype and exposure data) to an NIH-designated data repository[23] in a timely manner. Investigators should also submit any information necessary to interpret the submitted genomic data, such as study protocols, data instruments, and survey tools.

---

[13] Gene Expression Omnibus at http://www.ncbi.nlm.nih.gov/geo/.
[14] Sequence Read Archive at http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi.
[15] Trace Archive at http://www.ncbi.nlm.nih.gov/Traces/trace.cgi.
[16] Array Express at http://www.ebi.ac.uk/arrayexpress/.
[17] Mouse Genome Informatics at http://www.informatics.jax.org/.
[18] WormBase at http://www.wormbase.org.
[19] The Zebrafish Model Organism Database at http://zfin.org/.
[20] GenBank at http://www.ncbi.nlm.nih.gov/genbank/.
[21] European Nucleotide Archive at http://www.ebi.ac.uk/ena/.
[22] DNA Data Bank of Japan at http://www.ddbj.nig.ac.jp/.
[23] An NIH-designated data repository is any data repository maintained or supported by NIH either directly or through collaboration.

Genomic data undergo different levels of data processing, which provides the basis for NIH's expectations for data submission and timelines for the release of the data for access by investigators. These expectations and timelines are provided in the Supplemental Information. In general, NIH will release data submitted to NIH-designated data repositories no later than six months after the initial data submission begins, or at the time of acceptance of the first publication, whichever occurs first, without restrictions on publication or other dissemination.[24]

Investigators should de-identify[25] human genomic data that they submit to NIH-designated data repositories according to the standards set forth in the HHS Regulations for the Protection of Human Subjects[26] to ensure that the identities of research subjects cannot be readily ascertained with the data. Investigators should also strip the data of identifiers according to the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule.[27] The de-identified data should be assigned random, unique codes by the investigator, and the key to other study identifiers held by the submitting institution.

Although the data in the NIH database of Genotypes and Phenotypes (dbGaP) are de-identified by both the HHS Regulations for Protection of Human Subjects and HIPAA Privacy Rule standards, NIH has obtained a Certificate of Confidentiality for dbGaP as an additional precaution because genomic data can be re-identified.[28] NIH encourages investigators and institutions submitting large-scale human genomic datasets to NIH-designated data repositories to seek a Certificate of Confidentiality as an additional safeguard to prevent compelled disclosure of any personally identifiable information they may hold.[29]

2. Data Repositories

Investigators should register all studies with human genomic data that fall within the scope of the GDS Policy in dbGaP[30] by the time that data cleaning and quality control measures begin, regardless of which NIH-designated data 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[31]). NIH-designated data repositories need not be the exclusive source for facilitating the sharing of genomic data, that is, investigators may also elect to submit data to a non-NIH-designated data repository in addition to an NIH-designated data repository. However, investigators should ensure that appropriate data security measures

---

[24] A period for data preparation is anticipated prior to data submission to NIH, and the appropriate time intervals for that data preparation (or data cleaning) will be subject to the particular data type and project plans (see Supplemental Information). Investigators should work with NIH Program or Project Officials for specific guidance.

[25] De-identified refers to removing information that could be used to associate a dataset or record with a human individual.

[26] See 45 CFR 46.102(f) at http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html#46.102.

[27] See 45 CFR 164.514(b)(2). The list of HIPAA identifiers that must be removed is available at: http://www.gpo.gov/fdsys/pkg/CFR-2002-title45-vol1/pdf/CFR-2002-title45-vol1-sec164-514.pdf.

[28] Confidentiality Certificate. HG-2009-01. Issued to the National Center for Biotechnology Information, National Library of Medicine, NIH. See http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/GetPdf.cgi?document_name=ConfidentialityCertificate.pdf.

[29] For additional information about Certificates of Confidentiality, see http://grants.nih.gov/grants/policy/coc/.

[30] Database of Genotypes and Phenotypes at http://www.ncbi.nlm.nih.gov/gap.

[31] Cancer Genomics Hub at https://cghub.ucsc.edu/.

are in place,[32] and that confidentiality, privacy, and data use measures are consistent with the GDS Policy.

3. Tiered System for the Distribution of Human Data

Respect for, and protection of the interests of, research participants are fundamental to NIH's stewardship of human genomic data. The informed consent under which the data or samples were collected is the basis for the submitting institution to determine the appropriateness of data submission to NIH-designated data repositories, and whether the data should be available through unrestricted or controlled-access. Controlled-access data in NIH-designated data repositories are made available for secondary research only after investigators have obtained approval from NIH to use the requested data for a particular project. Data in unrestricted-access repositories are publicly available to anyone (e.g., The 1000 Genomes Project[33]).

4. Informed Consent

For research that falls within the scope of the GDS Policy, submitting institutions, through their Institutional Review Boards[34] (IRBs), privacy boards,[35] or equivalent bodies,[36] are to review the informed consent materials to determine whether it is appropriate for data to be shared for secondary research use. Specific considerations may vary with the type of study and whether the data are obtained through prospective or retrospective data collections. NIH provides additional information on issues related to the respect for research participant interests in its *Points to Consider for IRBs and Institutions in their Review of Data Submission Plans for Institutional Certifications*.[37]

For studies initiated after the effective date of the GDS Policy, NIH expects investigators to obtain participants' consent for their genomic and phenotypic data to be used for future research purposes and to be shared broadly. The consent should include an explanation about whether participants' individual-level data will be shared through unrestricted- or controlled-access repositories.

For studies proposing to use genomic data from cell lines or clinical specimens[38] that were created or collected after the effective date of the Policy, NIH expects that informed consent for future research use and broad data sharing will have been obtained even if the cell lines or clinical specimens are de-identified. If there are compelling scientific reasons that necessitate the use of genomic data from cell lines or clinical specimens that were created or collected after

---

[32] dbGaP Security Best Practices. See https://osp.od.nih.gov/wp-content/uploads/NIH_Best_Practices_for_Controlled-Access_Data_Subject_to_the_NIH_GDS_Policy.pdf.
[33] The 1000 Genomes Project at http://www.1000genomes.org/.
[34] See 45 CFR 46.102(g) at http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html#46.102.
[35] See the roles of Privacy Boards as elaborated in 45 CFR 164 at http://www.gpo.gov/fdsys/pkg/CFR-2011-title45-vol1/pdf/CFR-2011-title45-vol1-part164.pdf.
[36] Equivalent body is used here to acknowledge that some primary studies may be conducted abroad and in such cases the expectation is that an analogous review committee to an IRB or privacy board (e.g., Research Ethics Committees) may be asked to participate in the presubmission review of proposed genomic projects.
[37] Points to Consider for IRBs and Institutions in their Review of Data Submission Plans for Institutional Certifications. See https://osp.od.nih.gov/wp-content/uploads/GDS_Points_to_Consider_for_Institutions_and_IRBs.pdf.
[38] Clinical specimens are specimens that have been obtained through clinical practice.

the effective date of this Policy and that lack consent for research use and data sharing, investigators should provide a justification in the funding request for their use. The funding IC will review the justification and decide whether to make an exception to the consent expectation.

For studies using data from specimens collected before the effective date of the GDS Policy, there may be considerable variation in the extent to which future genomic research and broad sharing were addressed in the informed consent materials for the primary research. In these cases, an assessment by an IRB, privacy board, or equivalent body is needed to ensure that data submission is not inconsistent with the informed consent provided by the research participant. NIH will accept data derived from de-identified cell lines or clinical specimens lacking consent for research use that were created or collected before the effective date of this Policy.

NIH recognizes that in some circumstances broad sharing may not be consistent with the informed consent of the research participants whose data are included in the dataset. In such circumstances, institutions planning to submit aggregate-[39] or individual-level data to NIH for controlled-access should note any data use limitations in the data sharing plan submitted as part of the funding request. These data use limitations should be specified in the Institutional Certification submitted to NIH prior to award.

5. Institutional Certification

The responsible Institutional Signing Official[40] of the submitting institution should provide an Institutional Certification to the funding IC prior to award consistent with the genomic data sharing plan submitted with the request for funding. The Institutional Certification should state whether the data will be submitted to an unrestricted- or controlled-access database. For submissions to controlled-access and, as appropriate for unrestricted access, the Institutional Certification should assure that:

- The data submission is consistent, as appropriate, with applicable national, tribal, and state laws and regulations as well as relevant institutional policies;[41]
- Any limitations on the research use of the data, as expressed in the informed consent documents, are delineated;[42]
- The identities of research participants will not be disclosed to NIH-designated data repositories; and
- An IRB, privacy board, and/or equivalent body, as applicable, has reviewed the investigator's proposal for data submission and assures that:
  - The protocol for the collection of genomic and phenotypic data is consistent with 45 CFR Part 46;[43]

[39] Aggregate data are summary statistics compiled from multiple sources of individual-level data.
[40] An Institutional Signing Official is generally a senior official at an institution who is credentialed through NIH eRA Commons system and is authorized to enter the institution into a legally binding contract and sign on behalf of an investigator who has submitted data or a data access request to NIH.
[41] For the submission of data derived from cell lines or clinical specimens lacking research consent that were created or collected before the effective date of this Policy, the Institutional Certification needs to address only this item.
[42] For guidance on clearly communicating inappropriate data uses, see NIH Points to Consider in Drafting Effective Data Use Limitation Statements, https://osp.od.nih.gov/wp-content/uploads/NIH_PTC_in_Developing_DUL_Statements.pdf.
[43] 45 CFR Part 46. Protection of Human Subjects. See http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html.

- Data submission and subsequent data sharing for research purposes are consistent with the informed consent of study participants from whom the data were obtained;[44]
- Consideration was given to risks to individual participants and their families associated with data submitted to NIH-designated data repositories and subsequent sharing;
- To the extent relevant and possible, consideration was given to risks to groups or populations associated with submitting data to NIH-designated data repositories and subsequent sharing; and
- The investigator's plan for de-identifying datasets is consistent with the standards outlined in this Policy (see section IV.C.1.).

6. Exceptions to Data Submission Expectations

In cases where data submission to an NIH-designated data repository is not appropriate, that is, the Institutional Certification criteria cannot be met, investigators should provide a justification for any data submission exceptions requested in the funding application or proposal. The funding IC may grant an exception to submitting relevant data to NIH, and the investigator would be expected to develop an alternate plan to share data through other mechanisms. For transparency purposes, when exceptions are granted, studies will still be registered in dbGaP, the reason for the exception will be included in the registration record, and a reference will be provided to an alternative data-sharing plan or resource, if available. More information about requesting exceptions is available on the NIH Office of Science Policy website.[45]

7. Data Withdrawal

Submitting investigators and their institutions may request removal of data on individual participants from NIH-designated data repositories, in the event that a research participant withdraws or changes his or her consent. However, some data that have been distributed for approved research use cannot be retrieved.

V. Responsibilities of Investigators Accessing and Using Genomic Data

A. Requests for Controlled-Access Data

Access to human data is through a tiered model involving unrestricted- and controlled-data access mechanisms. Requests for controlled-access data[46] are reviewed by NIH Data Access Committees (DACs).[47] DAC decisions are based primarily upon conformance of the proposed research as described in the access request to the data use limitations established by the submitting institution through the Institutional Certification. NIH DACs will accept requests for proposed research uses beginning one month prior to the anticipated data release date. The access period for all controlled-access data is one year; at the end of each approved period, data users can request an additional year of access or close out the project. Although data are de-

[44] As noted earlier, for studies using data or specimens collected before the effective date of this Policy, the IRB, privacy board, or equivalent body should review informed consent materials to ensure that data submission is not inconsistent with the informed consent provided by the research participants.
[45] See https://osp.od.nih.gov/scientific-sharing/genomic-data-sharing/.
[46] dbGaP Authorized Access. See https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=login.
[47] For a list of NIH Data Access Committees, see https://osp.od.nih.gov/wp-content/uploads/NIH_DACs_Chairs.pdf.

identified, approved users of controlled-access data are encouraged to consider whether a Certificate of Confidentiality could serve as an additional safeguard to prevent compelled disclosure of any genomic data they may hold.[29]

B. Terms and Conditions for Research Use of Controlled-Access Data

Investigators approved to download controlled-access data from NIH-designated data repositories and their institutions are expected to abide by the NIH Genomic Data User Code of Conduct[48] through their agreement to the Data Use Certification.[49] The Data Use Certification, co-signed by the investigators requesting the data and their Institutional Signing Official, specifies the conditions for the secondary research use of controlled-access data, including:

- Using the data only for the approved research;
- Protecting data confidentiality;
- Following, as appropriate, all applicable national, tribal, and state laws and regulations, as well as relevant institutional policies and procedures for handling genomic data;
- Not attempting to identify individual participants from whom the data were obtained;
- Not selling any of the data obtained from NIH-designated data repositories;
- Not sharing any of the data obtained from controlled-access NIH-designated data repositories with individuals other than those listed in the data access request;
- Agreeing to the listing of a summary of approved research uses in dbGaP along with the investigator's name and organizational affiliation;
- Agreeing to report any violation of the GDS Policy to the appropriate DAC(s) as soon as it is discovered;
- Reporting research progress using controlled-access datasets through annual access renewal requests or project close-out reports;
- Acknowledging in all oral or written presentations, disclosures, or publications the contributing investigator(s) who conducted the original study, the funding organization(s) that supported the work, the specific dataset(s) and applicable accession number(s), and the NIH-designated data repositories through which the investigator accessed any data.

NIH expects that investigators who are approved to use controlled-access data will follow guidance on security best practices[32] that outlines expected data security protections (e.g., physical security measures and user training) to ensure that the data are kept secure and not released to any person not permitted to access the data.

If investigators violate the terms and conditions for secondary research use, NIH will take appropriate action. Further information is available in the Data Use Certification.

[48] Genomic Data User Code of Conduct. See https://osp.od.nih.gov/wp-content/uploads/Genomic_Data_User_Code_of_Conduct.pdf..
[49] Model Data Use Certification Agreement. See https://osp.od.nih.gov/wp-content/uploads/Model_DUC.pdf.

C. Conditions for Use of Unrestricted-Access Data

Investigators who download unrestricted-access data from NIH-designated data repositories should:

- Not attempt to identify individual human research participants from whom the data were obtained;[50]
- Acknowledge in all oral or written presentations, disclosures, or publications the specific dataset(s) or applicable accession number(s) and the NIH-designated data repositories through which the investigator accessed any data.

VI. Intellectual Property

NIH encourages patenting of technology suitable for subsequent private investment that may lead to the development of products that address public needs without impeding research. However, it is important to note that naturally occurring DNA sequences are not patentable in the United States.[51] Therefore, basic sequence data and certain related information (e.g., genotypes, haplotypes, *p*-values, allele frequencies) are pre-competitive. Such data made available through NIH-designated data repositories, and all conclusions derived directly from them, should remain freely available, without any licensing requirements.

NIH encourages broad use of NIH-funded genomic data that is consistent with a responsible approach to management of intellectual property derived from downstream discoveries, as outlined in the NIH *Best Practices for the Licensing of Genomic Inventions*[52] and Section 8.2.3, Sharing Research Resources, of the NIH Grants Policy Statement.[53] NIH discourages the use of patents to prevent the use of or to block access to genomic or genotype-phenotype data developed with NIH support.

By signing this document you agree that you have read and understand the policies described herein and that you agree to abide by these policies.

Signature	Date Signed

[50] In certain cases, NIH may consider approving research intended to enhance genomic data privacy protection procedures.
[51] Association for Molecular Pathology v. Myriad Genetics, Inc., 569 U.S. ___ (2013) (slip opinion 12-398). See http://www.supremecourt.gov/opinions/12pdf/12-398_1b7d.pdf.
[52] NIH Best Practices for the Licensing of Genomic Inventions. See http://www.ott.nih.gov/sites/default/files/documents/pdfs/70fr18413.pdf.
[53] NIH Grants Policy Statement. 8.2.3, Sharing Research Resources. See http://grants.nih.gov/grants/policy/nihgps_2012/nihgps_ch8.htm#_Toc271264950.