National Institutes of Health
NIH Genomic Data Sharing Policy and Preamble

Preamble

NIH announces the final Genomic Data Sharing (GDS) Policy, which sets forth expectations that ensure the broad and responsible sharing of genomic research data. Sharing research data supports the NIH mission 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 a broad range of research data, in addition to genomic data, publicly available in a timely manner from the research activities that it funds.[1,2,3,4,5]

NIH published the *Draft NIH Genomic Data Sharing Policy Request for Public Comments* in the *Federal Register* on September 20, 2013,[6] and the *NIH Guide for Grants and Contracts* on September 27, 2013,[7] for a 60-day public comment period that ended November 20, 2013. NIH also used websites, listservs, and social media to disseminate the request for comments. On November 6, 2013, during the comment period, NIH held a public webinar on the draft GDS Policy that was attended by nearly 200 people and included a question and answer session.[8]

NIH received a total of 107 public comments on the draft GDS Policy. Comments were submitted by individuals, organizations, and entities affiliated with academic institutions, professional and scientific societies, disease and patient advocacy groups, research organizations, industry and commercial organizations, tribal organizations, state public health agencies, and private clinical practices. The public comments have been posted on the NIH GDS website.[9] Comments were supportive of the principles of sharing data to advance research. However, a number of questions and concerns, and calls for clarification, were raised about specific aspects of the draft Policy on particular topics. A summary of comments, organized by corresponding sections of the GDS Policy, is provided below.

Scope and Applicability

Several commenters stated that the draft Policy was unclear with regard to the types of research to which the Policy would apply. Some commenters suggested that the technology used in a research study (i.e., array-based or high-throughput genomic technologies) should not be the focus in determining

---

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

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

[3] NIH Policy on Sharing of Model Organisms for Biomedical Research. May 7, 2004. See http://grants.nih.gov/grants/guide/notice-files/NOT-OD-04-042.html.

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

[5] 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.

[6] Federal Register Notice. Draft NIH Genomic Data Sharing Policy Request for Public Comments. See http://www.federalregister.gov/a/2013-22941.

[7] The NIH Guide for Grants and Contracts. Request for Information: Input on the Draft NIH Genomic Data Sharing Policy. September 27, 2013. See http://grants.nih.gov/grants/guide/notice-files/NOT-OD-13-119.html.

[8] Public Consultation Webinar. Draft NIH Genomic Data Sharing Policy. November 6, 2013. See https://webmeeting.nih.gov/p7sqo6avp6j/.

[9] Compiled Public Comments on the Draft Genomic Data Sharing Policy. See http://gds.nih.gov/pdf/GDS_Policy_Public_Comments.PDF.

applicability of the Policy. They suggested instead that the information gained from the research should determine the applicability of the Policy. Many other commenters expressed the concern that the Policy was overly broad and would lead to the submission of large quantities of data with low utility for other investigators. Several other commenters suggested that the scope of the Policy was not broad enough. Additionally, some commenters were uncertain about whether the Policy would apply to research funded by multiple sources.

NIH has revised the Scope and Applicability section to help clarify the types of research to which the Policy is intended to apply, and the reference to specific technologies has been dropped. The list of examples of the types of research projects that are within the Policy's scope, which appeared in Appendix A of the draft GDS Policy (now referred to as "Supplemental Information to the NIH Genomic Data Sharing Policy" [10]), has been revised and expanded, and examples of research that are not within the scope have been added as well. Also, the final GDS Policy now explicitly states that smaller studies (e.g., sequencing the genomes of fewer than 100 human research participants) are generally not subject to this Policy. Smaller studies, however, may be subject to other NIH data sharing policies (e.g., the National Institute of Allergy and Infectious Diseases Data Sharing and Release Guidelines[11]) or program requirements. In addition, definitions of key terms used in the Policy (e.g., aggregate data) have been included and other terms have been clarified.

The statement of scope remains intentionally general enough to accommodate the evolving nature of genomic technologies and the broad range of research that generates genomic data. It also allows for the possibility that individual NIH Institutes or Centers (IC) 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. The Policy applies to research funded in part or in total by NIH if NIH funding supports the generation of the genomic data. Investigators with questions about whether the Policy applies to their current or proposed research should consult the relevant Program Official or Program Officer or the IC's Genomic Program Administrator (GPA). Names and contact information for GPAs are available through the GDS website.[12]

Some commenters expressed concern about the financial burden on investigators and institutions of validating and sharing large volumes of genomic data and the possibility that resources spent to support data sharing would redirect funds away from research. While the resources needed to support data sharing are not trivial, NIH maintains that the investments are warranted by the significant discoveries made possible through the secondary use of the data. In addition, NIH is taking steps to evaluate and monitor the impact of data sharing costs on the conduct of research, both programmatically through the Big Data to Knowledge Initiative[13] and organizationally through the creation of the Scientific Data Council, which will advise the agency on issues related to data science.[14]

Data Sharing Plans

Some commenters pointed out that the Policy was not clear enough about the conditions under which NIH would grant an exception to the submission of genomic data to NIH. Some also suggested that NIH

---

[10] Supplemental Information to the NIH Genomic Data Sharing Policy. See http://gds.nih.gov/pdf/supplemental_info_GDS_Policy.pdf.

[11] National Institute of Allergy and Infectious Diseases. Data Sharing and Release Plans. See http://www.niaid.nih.gov/labsandresources/resources/dmid/pages/data.aspx.

[12] Roster of NIH Genomic Program Administrators. See http://gds.nih.gov/04po2_2GPA.html.

[13] NIH Big Data to Knowledge. See http://bd2k.nih.gov.

[14] NIH Big Data to Knowledge. Scientific Data Council. See http://bd2k.nih.gov/about_bd2k.html#sdcmembership.

should allow limited sharing of human genomic data when the original consent or national, tribal, or state laws do not permit broad sharing.

While NIH encourages investigators to seek consent for broad sharing, and some ICs may establish program priorities that expect studies proposed for funding to include consent for broad sharing, exceptions may be made. The final Policy clarifies that exceptions may be requested in cases where the submission of genomic data would not meet the criteria for the Institutional Certification.

Some commenters expressed concern that it would be difficult to estimate the resources required to support data sharing plans before a study is completed. Others asked for additional guidance on resources that should be requested to support the data sharing plan. Several commenters suggested that NIH should allow certain elements of the data sharing plan, such as the Institutional Certification and associated documentation, to be submitted along with other "Just-in-Time" information. For multi-year awards, one commenter suggested that the data sharing plans should be periodically reviewed for consistency with contemporary ethical standards. Another suggested that data sharing plans should be made public. Under the GDS Policy, investigators are expected to outline in the budget section of their funding application the resources they will need to prepare the data for submission to appropriate repositories. NIH will provide additional guidance on these resources, as necessary. The final Policy clarifies that only a basic genomic data sharing plan, in the Resource Sharing Plan section of grant applications, needs to be submitted with the funding application and that a more detailed plan should be provided prior to award. The Institutional Certification also should be provided prior to award, along with any other Just-in-Time information. Guidance on genomic data sharing plans is available on the NIH GDS website.[15] Data sharing plans will undergo periodic review through annual progress reports or other appropriate scientific project reviews. Further consideration will be given to the suggestion that data sharing plans should be made public.

Non-human and Model Organism Genomic Data

The draft GDS Policy proposed timelines for data submission and data release (i.e., when data should be made available for sharing with other investigators). For non-human data, the draft Policy proposed that data should be submitted and made available for sharing no later than the date of initial publication, with the acknowledgement that the submission and release of data for certain projects may be expected earlier, mirroring data sharing expectations that have been in place under other policies.[3] Some commenters suggested that the data submission expectations for non-human data were unclear. One commenter suggested that NIH should consider a more rapid timeline than the date of first publication for releasing model organism data, while other comments supported the specified data release timeline. Other commenters were concerned that the specified timeline was too short.

The final GDS Policy does not change the timeline for the submission and release of non-human and model organism data. The timeline is based on the need to promote broad data sharing while also accommodating the investigators generating the data, who often must make a significant effort to prepare the data for sharing. The Policy points out that an NIH IC may choose to shorten the timeline for data submission and release for certain projects and expects investigators to work with NIH Program or Project Officials for specific guidance on the timelines and milestones for their projects.

There was broad support for the Policy's flexibility of allowing non-human and model organism data to be deposited in any widely used data repository. One commenter requested that a link or reference to

---

[15] Genomic Data Sharing Website. Resources for Investigators Submitting Data to dbGaP. See http://gds.nih.gov/06researchers1.html.

non-NIH-designated repositories be included in the Policy. Further information about NIH-designated repositories, including examples of such repositories, is available on the GDS website,[16] and additional information about non-NIH-designated data repositories will be incorporated in outreach and training materials for NIH staff and investigators and made available on the GDS website. NIH has clarified the final Policy to state that data types that were previously submitted to widely used repositories (e.g., gene expression data to the Gene Expression Omnibus or Array Express) should continue as before, while data types not previously submitted may go to these or other widely used repositories as agreed to by the funding IC.

Human Genomic Data

The Supplemental Information to the NIH GDS Policy[10] establishes timelines for the submission and subsequent release of data for access by secondary investigators based on the level of processing that the data have undergone. A number of commenters expressed concern about these timelines, suggesting that they were too short and could limit an investigator's ability to perform adequate quality control and publish results within the provided timeline. Many commenters proposed that the timeline for data release be extended to 12 or 18 months or be the date of publication, whichever comes first. Others were concerned that the timelines were too long and that they should reflect the longstanding principle of rapid data release as articulated in the Bermuda and Ft. Lauderdale agreements.[4] Some commenters were concerned that the elimination of the embargo period, (i.e., the period between when a study is released for secondary research and when the submitting investigator first publishes on the findings of the study) would adversely affect the goal of rapid data release. One commenter was concerned that data would be released before investigators could discuss consequential findings with participants.

NIH has modified the Supplemental Information to clarify that the 6-month deferral for the release of Level 2 and Level 3 human genomic data does not start until the data have been cleaned and submission to NIH has been initiated, which is typically about three months after the data have been generated. Because there will be significant variation in research projects generating Level 2 and Level 3 human genomic data, the timeline for submission is project-specific and will be determined in each case by the funding NIH IC through consultation with the investigator, and the Supplemental Information has been clarified accordingly. Under the GWAS Policy,[5] a publication embargo period was used as a way of making data more rapidly available. In exchange for immediate data access, secondary users were not permitted to publish or present research findings until 12 months after the data were released. NIH did not adopt this approach for the GDS Policy because in practice, the publication embargo dates were difficult for secondary users to track, especially for datasets that had multiple embargo periods for certain types of data, raising the risk of unintentional embargo violations. Regarding the concern that human genomic data will be made available before investigators can notify participants of consequential findings, such data would be considered Level 4 data and would not be expected to be released before publication, which NIH believes will provide sufficient time to discuss consequential findings with participants.

Many commenters called for the Policy to include technical data standards for the submission of human genomic data, such as platform information, controlled vocabulary, normalization algorithms, data quality standards, and metadata standards. NIH agrees with the importance of developing and using standards for genomic data and is aware that there are numerous initiatives underway to develop and promote such standards.[17] NIH has revised the Supplemental Information by adding a section on resources for data

---

[16] Genomic Data Sharing Website. Data Repositories. See http://gds.nih.gov/02dr2.html.
[17] See for example the Genomic Standards Consortium, http://gensc.org/; the Global Alliance, http://www.broadinstitute.org/news/globalalliance; and the NIH Big Data to Knowledge focus on community-based data and metadata standards, http://bd2k.nih.gov/about_bd2k.html#areas.

standards. It provides references to instructions for data submission to specific NIH-designated data repositories, which include data standards. Additional resources for data standards will be incorporated in the Supplemental Information as they are developed and become appropriate for broad use.

Several commenters asked for a definition of an NIH-designated data repository and for guidance on determining which non-NIH repositories are acceptable as well as examples of such repositories. Commenters also expressed interest in additional details regarding the use of Trusted Partners, which are third-party partnerships established through a contract mechanism, to provide infrastructure needs for data storage and/or tools that are useful for genomic data analyses. A definition of an NIH-designated repository is now included in the final Policy. Additionally, further information about non-NIH-designated repositories that accept human genomic data will be made available on the GDS website and incorporated in outreach and training materials for NIH staff and NIH-funded investigators. Additional information about Trusted Partners, including the standards required for trusted partnerships, is also available on the NIH GDS website.[16]

Regarding informed consent, the GDS Policy expects investigators generating genomic data to seek consent from participants for future research uses and the broadest possible sharing. A number of commenters were concerned that participants would not agree to consent for broad sharing and that enrollment in research studies may decline, potentially biasing studies if certain populations were less likely to consent to broad use of their data. Some commenters also raised a concern about the competitiveness of an application that proposed to obtain consent for more limited sharing of data. Several commenters suggested that NIH permit alternative forms of informed consent other than broad consent, such as dynamic consent or tiered consent.

NIH recognizes that consent for future research uses and broad sharing may not be appropriate or obtainable in all circumstances. ICs may continue to accept data from studies with consents that stipulate limitations on future uses and sharing, and NIH will maintain the data access system that enables more limited sharing and secondary use. With regard to the competitiveness of grant applications that do not propose to utilize consent for broad sharing, this Policy does not propose that applications be assessed on this point during the merit review, but investigators are nonetheless expected to seek consent for broad sharing to the greatest extent possible. The breadth of the sharing permitted by the consent may be taken into consideration during program priority review by the ICs. Regarding the alternative forms of consent, the Policy does not prohibit the use of dynamic or tiered consents. It promotes the use of consent for broad sharing to enable the greatest potential public benefit. However, NIH recognizes that changing technology may enable more dynamic consent processes that improve tracking and oversight and more closely reflect participant preferences. NIH will continue to monitor developments in this area.

Several commenters were unsure whether the GDS Policy would apply to research in clinical settings or research involving data from deceased individuals. Research that falls within the scope of the GDS Policy will be subject to the Policy, regardless of whether it occurs in a clinical setting or involves data generated from deceased individuals.

Several commenters also expressed concern that the Policy is unclear about the ability of groups, in addition to participants, to opt-out or withdraw informed consent for research and whether the ability to withdraw could be transferred or inherited. The Policy states that investigators and institutions may request that NIH withdraw data in the event that individual participants or groups withdraw consent for secondary research, although some data that have been distributed for research cannot be retrieved. Institutions submitting the data should determine whether data should be withdrawn from NIH repositories and notify NIH accordingly.

Many commenters urged NIH to develop standard text or templates for informed consent documents so that investigators would be assured that their consent material would be consistent with the Policy's expectations for informed consent and data sharing. One of these commenters noted the challenge of conveying the necessary information (e.g., broad future research uses) without adding to the complexity of consent forms. Developing educational materials or tools to guide the process for obtaining informed consent was also suggested. Other commenters expressed concern about the burden of rewriting and harmonizing existing informed consent documents. NIH appreciates the suggestion to develop template consent documents and plans to provide guidance to assist investigators and institutions in developing informed consent documents.

Many comments questioned the proposal to require explicit consent for research that is not considered humans subjects research under 45 CFR Part 46 (e.g., research that involves de-identified specimens or cell lines). There were also several comments about the draft GDS Policy proposal to grandfather data from de-identified clinical specimens and cell lines collected or generated before the effective date of the GDS Policy. The reason the Policy expects consent for research for the use of data generated from de-identified clinical specimens and cell lines created after the effective date of the Policy is because the evolution of genomic technology and analytical methods raises the risk of re-identification.[18] Moreover, requiring that consent be obtained is respectful of research participants, and it is increasingly clear that participants expect to be asked for their permission to use and share their de-identified specimens for research.[19,20,21] The Policy does not require consent to be obtained for research with data generated from de-identified clinical specimens and cell lines that were created or collected before the effective date of the Policy because of the practical and ethical limitations in recontacting participants to obtain new consent for existing collections and the fact that such data may have already been widely used in research.

The draft GDS Policy included an exception for "compelling scientific reasons" to allow the research use of data from de-identified, clinical specimens or cell lines collected or created after the effective date of the Policy and for which research consent was not obtained. Commenters did not object to the need for such an exception, but they asked for clarification on what constitutes a "compelling scientific reason," and the process through which investigators' justifications would be determined to be appropriate.

The funding IC will determine whether the investigators' justifications for the use of clinical specimens or cell lines for which no consent for research was obtained are acceptable, as provided in their funding application and Institutional Certification. Further guidance on what constitutes compelling scientific reasons will be made available on the GDS website and will likely evolve over time as NIH ICs, the NIH GDS governance system, and program and project staff acquire greater experience with requests for research with such specimens.

For clinical specimens and cell lines lacking consent for research and collected before the effective date of the Policy, several commenters were concerned that the Policy was unclear about whether data from such specimens can be deposited in NIH repositories. This provision of the Policy is intended to allow the research use of genomic data derived from de-identified clinical specimens or cell lines collected or created after the Policy's effective date in exceptional situations where the proposed research has the potential to advance scientific or medical knowledge significantly and could not be conducted with

---

[18] Gymrek et al. Identifying Personal Genomes by Surname Inference. Science. 339(6117): 321-324. (2013).
[19] Kaufman et al. Public Opinion about the Importance of Privacy in Biobank Research. American Journal of Human Genetics. 85(5): 643-654. (2009).
[20] Vermeulen et al. A Trial of Consent Procedures for Future Research with Clinically Derived Biological Samples. British Journal of Cancer. 101(9): 1505-1512. (2009).
[21] Trinidad et al. Research Practice and Participant Preferences: the Growing Gulf. Science. 331(6015): 287-288. (2011).

consented specimens or cell lines. The draft GDS Policy stated that NIH will accept data from clinical specimens and cell lines lacking consent for research use that were collected before the effective date of the Policy, and this remains unchanged in the final Policy.

A concern shared by several commenters was that the risks posed to the privacy of individuals with rare diseases, populations with higher risk of re-identification by the broad sharing of data, or populations at risk of greater potential harm from re-identification were not adequately addressed. Several commenters were particularly concerned that no additional protections were specified for these populations, and a subset suggested that research subject to the GDS Policy that involves these populations should be entirely exempt from the Policy's expectations for data sharing.

Currently, NIH requests IRBs to consider ethical concerns related to groups or populations when determining whether a study's consent documents are consistent with NIH policy.[22] In addition, NIH has clarified in the final GDS Policy that exceptions may be requested for the submission and subsequent sharing of data if the criteria in the Institutional Certification cannot be met (e.g., an IRB or equivalent body cannot assure that submission of data and subsequent sharing for research purposes are consistent with the informed consent of study participants). If a submitting institution determines that the criteria can be met but has additional concerns related to the sharing of the data, the institution can indicate additional stipulations for the use of the data through the data use limitations submitted with the study.

Several commenters suggested that return of medically actionable incidental findings should be included in the consent or that re-identification of participants should be allowed in order to return such incidental results. NIH recognizes that, as in any research study, harms may result if individual research findings that have not been clinically validated are returned to subjects or are used prematurely for clinical decision-making. The return of individual findings from studies using data obtained NIH-designated repositories is expected to be rare, because investigators will not be able to return individual research results directly to a participant as neither they nor the repository will have access to the identities of participants. Submitting institutions and their IRBs may wish to establish policies for determining when it is appropriate to return individual findings from research studies. Further guidance on the return of results is available from the Presidential Commission for the Study of Bioethical Issues' report, "Anticipate and Communicate: Ethical Management of Incidental and Secondary Findings in the Clinical, Research, and Direct-to-Consumer Contexts."[23]

Several commenters were concerned that the draft GDS Policy was unclear about which standard should be used to ensure the de-identification of data. Another issue raised by a number of comments related to identifiability of genomic data. Several commenters were concerned that de-identified genotype data could be re-identified, even if these data are de-identified according to Health Insurance Portability and Accountability Act (HIPAA) and the Common Rule. Others asserted that genomic data could not be fully de-identified. A number of commenters suggested that the GDS Policy should explicitly state that risks exist for participant privacy despite the de-identification of genomic data and should require informed consent documents to include such a statement. Others suggested that the Policy should state that genomic information cannot be de-identified. Commenters suggested that the risks of re-identification were not adequately addressed in the draft Policy.

---

[22] NIH Points to Consider for IRBs and Institutions in their Review of Data Submission Plans for Institutional Certifications Under NIH's Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS). See http://gds.nih.gov/pdf/PTC_for_IRBs_and_Institutions_revised5-31-11.pdf.

[23] Presidential Commission for the Study of Bioethical Issues. Anticipate and Communicate: Ethical Management of Incidental and Secondary Findings in the Clinical, Research, and Direct-to-Consumer Contexts. December 2013. See http://bioethics.gov/node/3183.

The final GDS Policy has been clarified to state that for the purpose of the Policy, data should be de-identified to meet the definition for de-identified data in the HHS Regulations for Protection of Human Subjects[24] and be stripped of the 18 identifiers listed in the HIPAA Privacy Rule.[25] NIH agrees that the risks of re-identification should be conveyed to prospective subjects in the consent process. This is one of the reasons why NIH expects explicit consent after the effective date of the Policy for broad sharing and for data that will be submitted to unrestricted-access data repositories (i.e., openly accessible data repositories, previously referred to as "open access"). NIH will provide further guidance on informing participants about the risks of re-identification through revisions to guidance documents such as the *NIH Points to Consider for IRBs and Institutions in their Review of Data Submission Plans for Institutional Certifications Under NIH's Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies*.[22]

Several commenters were particularly concerned about the cost and burden of obtaining informed consent for the research use of data generated from clinical specimens and cell lines collected or created after the effective date of the GDS Policy. NIH recognizes that these consent expectations for data from de-identified clinical specimens collected after the effective date will require additional resources. Given growing concerns about re-identification, it is no longer ethically tenable simply to de-identify clinical specimens or derived cell lines to generate data for research use without an individual's consent. In addition, NIH anticipates that obtaining consent for broad future research uses will facilitate access to greater volumes of data and ultimately will reduce the costs and burdens associated with sharing research data.

Some commenters expressed concern that the draft Policy's standards for consent are more restrictive than other rules governing human subjects protections including the Common Rule[26] and revisions proposed to the Common Rule in a 2011 Advance Notice of Proposed Rule Making (ANPRM).[27] Some commenters sought greater clarification regarding regulatory differences or the regulatory basis for the draft Policy's protections.

NIH has the authority to establish additional policies with expectations that are not required by laws or regulations but advance the agency's mission to enhance health, lengthen life, and reduce illness and disability. The GDS Policy builds on the GWAS Policy, which established additional expectations that were not required by the Common Rule for obtaining consent for, handling, sharing, and using human genotype and phenotype data in NIH-funded research. NIH expects that in addition to adhering to the GDS Policy, investigators and institutions will also comply with the Common Rule and any other applicable federal regulations or laws. In response to the concern that the draft Policy is inconsistent with the ANPRM for revisions to the Common Rule, NIH will evaluate any inconsistencies between the GDS Policy and the Common Rule when the Common Rule revisions are final.

Responsibilities of Investigators Accessing and Using Genomic Data

Commenters asserted that the draft GDS Policy did not do enough to protect against the misuse of the data by investigators accessing the data. They suggested that the Policy state that responsibilities outlined in the Policy for data users should be "required" rather than "expected" and should state that there will be penalties for noncompliance with the Policy and rigorous sanctions for the intentional misuse of data.

---

[24] Code of Federal Regulations. Protection of Human Subjects. Definitions. See 45 CFR 46.102(f) at http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html#46.102.
[25] The list of HIPAA identifiers that must be removed is available at 45 CFR 164.514(b)(2). See: http://www.gpo.gov/fdsys/pkg/CFR-2002-title45-vol1/pdf/CFR-2002-title45-vol1-sec164-514.pdf.
[26] Federal Policy for the Protection of Human Subjects (Common Rule). 45 CFR Part 46. See http://www.hhs.gov/ohrp/humansubjects/commonrule/.
[27] ANPRM for Revision to Common Rule. See http://www.hhs.gov/ohrp/humansubjects/anprm2011page.html.

There was also a comment proposing that a submitting institution should be able to review and comment on all data access requests (DARs) to NIH before NIH completes its internal review process and proposed that NIH notify submitting institutions and research participants of any policy violations reported by users of genomic data.

NIH Data Access Committees (DACs) review DARs on behalf of submitting institutions by using the data use limitations provided by the institutions to determine whether the DAR is consistent with the limitations to ensure that participants' wishes are respected. As part of its ongoing oversight process, NIH reviews notifications of data mismanagement or misuse, such as errors in the assignment of data use limitations during data submission, investigators sharing controlled-access data with unapproved investigators, and investigators using the data for research that was not described in their research use statement. To date, violations have been discovered before the completion of the research, and no participants have been harmed. When NIH becomes aware of any problems, the relevant institution and investigators are notified, and NIH takes appropriate steps to address the violation and prevent it from recurring. To ensure that the penalties for the misuse of data are clear for all data submitters, users, and research participants, the GDS Policy has been revised to clarify that secondary users in violation of the Policy or the Data Use Certification may face enforcement actions. In addition, a measure to protect the confidentiality of de-identified data obtained through controlled access has been added by encouraging approved users to consider requesting a Certificates of Confidentiality.

Several comments were submitted by representatives or members of tribal organizations about data access. Tribal groups expressed concerns about the ability of DACs to represent tribal preferences in the review of requests for tribal data. They also proposed new provisions for the protection of participant data, for example, including de-identification of tribal membership in participant de-identification and revision of the Genomic Data User Code of Conduct to reference protocols for accessing, sharing, and using tribal data, such as de-identification of participants' tribal affiliation.

The final Policy has been modified to reference explicitly that tribal law, in addition to other factors such as limitations in the original informed consents or concerns about harms to individuals or groups, should be considered in assessing the secondary use of some genomic data.

Some commenters proposed changes to controlled access for human genomic data. Some commenters thought controlled access unnecessarily limited research, and many provided a range of suggestions on how to improve the process of accessing the data, such as: allowing unrestricted access to de-identified data; developing standard data use limitations for controlled-access data; streamlining and increasing transparency of data access procedures and processing time; and modifying the database of genotypes and phenotypes (dbGaP) to facilitate peer-review and collaboration.

The final GDS Policy permits unrestricted access to de-identified data, but only if participants have explicitly consented to sharing their data through unrestricted-access mechanisms. Standard data use limitations have been developed by NIH and are available through the GDS website.[28] With regard to improving transparency on data access procedures, NIH plans to make statistics on access publicly available on the GDS website,[29] including the average processing time for NIH to review data access requests. From its inception, dbGaP has solicited feedback from users and worked to improve data submission and access procedures, for example, the creation of a study compilation that allows investigators to submit a single request for access to all controlled-access aggregate and individual-level

---

[28] Genomic Data Sharing Website. Standard Data Use Limitations. See http://gds.nih.gov/pdf/standard_data_use_limitations.pdf.

[29] Genomic Data Sharing Website. See http://gds.nih.gov/.

genomic data available for general research use.[30,31] NIH will continue to seek user feedback and track the performance of the dbGaP system.

Several comments expressed concern that the GDS Policy will increase administrative burden for NIH DACs, potentially resulting in longer timeframes to obtain data maintained under controlled access. NIH is aware of the burden that may be imposed on DACs by additional data access requests and will continue to monitor this possibility and, as needed, develop methods to decrease DAC burden and improve performance for investigators, institutions, and NIH ICs.

Intellectual Property

The GDS Policy expects that basic sequence and certain related data made available through NIH-designated data repositories and all conclusions derived from them will be freely available. It discourages patenting of "upstream" discoveries, which are considered pre-competitive, while it encourages the patenting of "downstream" applications appropriate for intellectual property. Of the several comments received on intellectual property, many supported the draft Policy's provisions. However, a few commenters opposed patenting in general, and one suggested that the Policy should explicitly prohibit rather than discourage the use of patents for inventions that result from research undertaken with data from NIH-designated repositories.

As noted above, NIH encourages the appropriate patenting of "downstream" applications. NIH will continue to encourage the broadest possible use of products, technologies, and information resulting from NIH funding or developed using data obtained from NIH data repositories to the extent permitted by applicable NIH policies, federal regulations and laws, while encouraging the patenting of technology suitable for private investment that addresses public needs. As is well known, the Supreme Court decision in *Association for Molecular Pathology et al. v. Myriad Genetics, Inc. et al.* prohibits the patenting of naturally occurring DNA sequences.[32] Consistent with this decision, the NIH expects that patents directed to naturally occurring sequences will not be filed.

Conclusion

NIH appreciates the time and effort taken by commenters to respond to the Request for Comments. The responses were helpful in revising the draft GDS Policy and enhanced the understanding of additional guidance materials that may be necessary.

---

[30] dbGaP Compilation of Aggregate Genomic Data for General Research Use. See http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000501.v1.p1.
[31] dbGaP Collection: Compilation of Individual-Level Genomic Data for General Research Use. See http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/collection.cgi?study_id=phs000688.v1.p1.
[32] 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.

NIH Genomic Data Sharing Policy

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 research data. Sharing research data supports the NIH mission 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.[1,2,3,4,5]

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),[33] single nucleotide polymorphisms (SNP) arrays, and genome sequence[Error! Bookmark not defined.], 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)[10] 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*).

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[34] and/or other authorities, as appropriate.

III. Effective Date

This Policy applies to:

---

[33] 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.

[34] 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 http://www.gpo.gov/fdsys/pkg/CFR-2011-title45-vol1/xml/CFR-2011-title45-vol1-part74.xml#seqnum74.62.

- Competing grant applications[35] 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[36] 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.

B. Non-human Genomic Data
1. Data Submission Expectations and Timeline

Large-scale non-human genomic data, including data from microbes, microbiomes, and model organism, 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),[37] Sequence Read Archive (SRA),[38] Trace Archive,[39] Array Express,[40] Mouse Genome Informatics (MGI),[41] WormBase,[42] the Zebrafish

---

[35] 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.

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

[37] Gene Expression Omnibus at http://www.ncbi.nlm.nih.gov/geo/.

[38] Sequence Read Archive at http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi.

[39] Trace Archive at http://www.ncbi.nlm.nih.gov/Traces/trace.cgi.

Model Organism Database (ZFIN),[43] GenBank,[44] European Nucleotide Archive (ENA),[45] or DNA Data Bank of Japan (DDBJ).[46] 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[47] 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.

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.[48]

Investigators should de-identify[49] 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[24] 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.[25] 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

---

[40] Array Express at http://www.ebi.ac.uk/arrayexpress/.
[41] Mouse Genome Informatics at http://www.informatics.jax.org/.
[42] WormBase at http://www.wormbase.org.
[43] The Zebrafish Model Organism Database at http://zfin.org/.
[44] GenBank at http://www.ncbi.nlm.nih.gov/genbank/.
[45] European Nucleotide Archive at http://www.ebi.ac.uk/ena/.
[46] DNA Data Bank of Japan at http://www.ddbj.nig.ac.jp/.
[47] An NIH-designated data repository is any data repository maintained or supported by NIH either directly or through collaboration.
[48] 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.
[49] De-identified refers to removing information that could be used to associate a dataset or record with a human individual.

data can be re-identified.[50] 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.[51]

2. Data Repositories

Investigators should register all studies with human genomic data that fall within the scope of the GDS Policy in dbGaP[52] 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[53]). 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 are in place,[54] 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[55]).

4. Informed Consent

For research that falls within the scope of the GDS Policy, submitting institutions, through their Institutional Review Boards[24] (IRBs), privacy boards,[56] or equivalent bodies,[57] 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*.[22]

---

[50] 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.

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

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

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

[54] dbGaP Security Best Practices. See http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/GetPdf.cgi?document_name=dbgap_2b_security_procedures.pdf.

[55] The 1000 Genomes Project at http://www.1000genomes.org/.

[56] 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.

[57] 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.

For studies initiated <u>after</u> 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[58] that were created or collected <u>after</u> 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 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 <u>before</u> 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 <u>before</u> 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-[59] 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[60] 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;[61]

- Any limitations on the research use of the data, as expressed in the informed consent documents, are delineated;[62]

---

[58] Clinical specimens are specimens that have been obtained through clinical practice.
[59] Aggregate data are summary statistics compiled from multiple sources of individual-level data.
[60] 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.
[61] 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.

- 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:

  o The protocol for the collection of genomic and phenotypic data is consistent with 45 CFR Part 46;[26]

  o Data submission and subsequent data sharing for research purposes are consistent with the informed consent of study participants from whom the data were obtained;[63]

  o Consideration was given to risks to individual participants and their families associated with data submitted to NIH-designated data repositories and subsequent sharing;

  o 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

  o 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 GDS website.[15]

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

---

[62] For guidance on clearly communicating inappropriate data uses, see NIH Points to Consider in Drafting Effective Data Use Limitation Statements, http://gwas.nih.gov/pdf/NIH_PTC_in_Drafting_DUL_Statements.pdf.

[63] 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.

Access to human data is through a tiered model involving unrestricted- and controlled-data access mechanisms. Requests for controlled-access data[64] are reviewed by NIH Data Access Committees (DACs).[65] 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-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.[51]

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[66] through their agreement to the Data Use Certification.[67] 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

---

[64] dbGaP Authorized Access. See https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=login.
[65] For a list of NIH Data Access Committees, see http://gwas.nih.gov/04po2_1DAC.html.
[66] Genomic Data User Code of Conduct. See http://gds.nih.gov/pdf/Genomic_Data_User_Code_of_Conduct.pdf.
[67] Model Data Use Certification Agreement. See http://gwas.nih.gov/pdf/Model_DUC_7-26-13.pdf.

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[54] 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.

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;[68]
- 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.[32] 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*[69] and Section 8.2.3, Sharing Research Resources, of the NIH Grants Policy Statement.[70] 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.

---

[68] In certain cases, NIH may consider approving research intended to enhance genomic data privacy protection procedures.
[69] NIH Best Practices for the Licensing of Genomic Inventions. See
http://www.ott.nih.gov/sites/default/files/documents/pdfs/70fr18413.pdf.
[70] NIH Grants Policy Statement. 8.2.3, Sharing Research Resources. See
http://grants.nih.gov/grants/policy/nihgps_2012/nihgps_ch8.htm#_Toc271264950.