National Institutes of Health
Preamble for the Genomic Data Sharing Policy

Introduction

NIH announces the final Genomic Data Sharing (GDS) Policy, which sets forth expectations that ensure the broad and responsible sharing of genomic[1] 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.[2,3,4,5,6]

NIH published the *Draft NIH Genomic Data Sharing Policy Request for Public Comments* in the *Federal Register* on September 20, 2013,[7] and the *NIH Guide for Grants and Contracts* on September 27, 2013,[8] 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.[9]

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.[10] Comments were supportive of the principles of sharing data to advance research. However, there were a number of questions and concerns, and calls for clarification about specific aspects of the draft Policy. A summary of comments, organized by corresponding sections of the GDS Policy, is provided below.

[1] The genome is the entire set of genetic instructions found in a cell. See http://ghr.nlm.nih.gov/glossary=genome.
[2] Final NIH Statement on Sharing Research Data. February 26, 2003. See http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html.
[3] NIH Intramural Policy on Large Database Sharing. April 5, 2002. See http://sourcebook.od.nih.gov/ethic-conduct/large-db-sharing.htm.
[4] 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.
[5] 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.
[6] 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.
[7] Federal Register Notice. Draft NIH Genomic Data Sharing Policy Request for Public Comments. See http://www.federalregister.gov/a/2013-22941.
[8] 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.
[9] Public Consultation Webinar. Draft NIH Genomic Data Sharing Policy. November 6, 2013. See https://webmeeting.nih.gov/p7sqo6avp6j/.
[10] Compiled Public Comments on the Draft Genomic Data Sharing Policy. See http://gds.nih.gov/pdf/GDS_Policy_Public_Comments.PDF.

National Institutes of Health
Preamble for the Genomic Data Sharing Policy

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 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" [11]), 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[12]) 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.[13]

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

[11] Supplemental Information to the NIH Genomic Data Sharing Policy. See http://gds.nih.gov/pdf/supplemental_info_GDS_Policy.pdf.
[12] National Institute of Allergy and Infectious Diseases. Data Sharing and Release Plans. See http://www.niaid.nih.gov/labsandresources/resources/dmid/pages/data.aspx.
[13] Roster of NIH Genomic Program Administrators. See http://gds.nih.gov/04po2_2GPA.html.

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[14] and organizationally through the creation of the Scientific Data Council, which will advise the agency on issues related to data science.[15]

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 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 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.[16] 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.

[14] NIH Big Data to Knowledge. See http://bd2k.nih.gov.
[15] NIH Big Data to Knowledge. Scientific Data Council. See http://bd2k.nih.gov/about_bd2k.html#sdcmembership.
[16] Genomic Data Sharing Website. Resources for Investigators Submitting Data to dbGaP. See http://gds.nih.gov/06researchers1.html.

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.[4] 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 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,[17] 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[11] 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

[17] Genomic Data Sharing Website. Data Repositories. See http://gds.nih.gov/02dr2.html.

the Bermuda and Ft. Lauderdale agreements.[5] 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,[6] 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.[18] NIH has revised the Supplemental Information by adding a section on resources for data 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

[18] 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.

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.[17]

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.[19] 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.[20,21,22] 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.

---

[19] Gymrek et al. Identifying Personal Genomes by Surname Inference. Science. 339(6117): 321-324. (2013).
[20] Kaufman et al. Public Opinion about the Importance of Privacy in Biobank Research. American Journal of Human Genetics. 85(5): 643-654. (2009).
[21] 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).
[22] Trinidad et al. Research Practice and Participant Preferences: the Growing Gulf. Science. 331(6015): 287-288. (2011).

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 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.[23] 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 from 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,

[23] 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.

"Anticipate and Communicate: Ethical Management of Incidental and Secondary Findings in the Clinical, Research, and Direct-to-Consumer Contexts."[24]

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.

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[25] and be stripped of the 18 identifiers listed in the HIPAA Privacy Rule.[26] 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*.[23]

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.

---

[24] 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.

[25] 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.

[26] 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.

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[27] and revisions proposed to the Common Rule in a 2011 Advance Notice of Proposed Rule Making (ANPRM).[28] 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. 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

[27] Federal Policy for the Protection of Human Subjects (Common Rule). 45 CFR Part 46. See http://www.hhs.gov/ohrp/humansubjects/commonrule/.
[28] ANPRM for Revision to Common Rule. See http://www.hhs.gov/ohrp/humansubjects/anprm2011page.html.

protect the confidentiality of de-identified data obtained through controlled access has been added by encouraging approved users to consider requesting a Certificate 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.[29] With regard to improving transparency on data access procedures, NIH plans to make statistics on access publicly available on the GDS website,[30] 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 genomic data available for general research use.[31,32] 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

[29] Genomic Data Sharing Website. Standard Data Use Limitations. See http://gds.nih.gov/pdf/standard_data_use_limitations.pdf.
[30] Genomic Data Sharing Website. See http://gds.nih.gov/.
[31] 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.
[32] 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.

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.[33] 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.

[33] 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.