Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS)

Notice Number: NOT-OD-07-088

Key Dates

Release Date: August 28, 2007
Effective Date: January 25, 2008

Other Relevant Notices

- November 16, 2007 - See Notice (NOT-OD-08-013) Implementation Guidance and Instructions for Applicants.
- October 20, 2006 (NOT-OD-07-013) - NIH Town Hall Meeting on the Proposed Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS).
- October 20, 2006 (NOT-OD-07-012) - Extended Comment Period for the Proposed Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS).
- August 30, 2006 (NOT-OD-06-094) - Request for Information (RFI): Proposed Policy for Sharing of Data obtained in NIH supported or conducted Genome-Wide Association Studies (GWAS).
- May 15, 2006 (NOT-OD-06-071) - Notice to Applicants for NIH Genome-Wide Association Studies.

Issued by

National Institutes of Health (NIH) (http://www.nih.gov)

Table of Contents

Background

Preamble: Summary of Public Comments on Proposed Policy
I. Rationale for a Centralized Data Repository
II. Protection of Research Participants
III. Scientific Publication
IV. Intellectual Property

Policy for Genome-Wide Association Studies (GWAS)
I. Principles
II. Applicability
III. Data Management
IV. Publication
V. Intellectual Property

Inquiries

Background

The NIH is interested in advancing genome-wide association studies (GWAS) to identify common genetic factors that influence health and disease. For the purposes of this policy, a genome-wide association study is defined as any study of genetic variation across the entire human genome that is designed to identify genetic associations with observable traits (such as blood pressure or weight), or the presence or absence of a disease or condition. [1] Whole genome information,

when combined with clinical and other phenotype data, offers the potential for increased understanding of basic biological processes affecting human health, improvement in the prediction of disease and patient care, and ultimately the realization of the promise of personalized medicine. In addition, rapid advances in understanding the patterns of human genetic variation and maturing high-throughput, cost-effective methods for genotyping are providing powerful research tools for identifying genetic variants that contribute to health and disease.

For these reasons, the NIH announced in May 2006 that it planned to: (1) update NIH data sharing policies for research applications involving GWAS data; (2) initiate a public consultation process to inform policy development activities; and (3) track GWAS applications and awards at a central level (NOT-OD-06-071). A call for public comments on a proposed GWAS policy was issued on August 30, 2006 (NOT-OD-06-094). Between August 30 and November 30, 2006, the NIH solicited public comments from a range of public sectors (see Preamble below). Following the comment period, NIH convened a Town Hall Meeting in Bethesda, Maryland on December 14, 2006, to provide an opportunity for direct interaction with interested stakeholders on the important policy questions raised through the proposed policy (NOT-OD-07-022).

This Notice provides the NIH response to the public comments received during the public consultation activities and presents the revised GWAS policy developed by the NIH in response to the feedback received and further internal development of the issues.

The policy addresses (1) data sharing procedures, (2) data access principles, (3) intellectual property, and (4) issues regarding the protection of research participants through all phases of GWAS. Many of the principles contained in the policy reflect existing NIH polices and other NIH discussions.

The goal of the policy is to advance science for the benefit of the public through the creation of a centralized NIH GWAS data repository [2]. Maximizing the availability of resources facilitates research and enables medical science to better address the health needs of people based on their individual genetic information.

Protecting Research Participants

The potential for public benefit to be achieved through sharing GWAS data is significant. However, genotype and phenotype information generated about individuals, such as data related to the presence or risk of developing particular diseases or conditions and information regarding paternity or ancestry, may be sensitive. Therefore, protecting the privacy of the research participants and the confidentiality of their data is critically important. Risks to individuals, groups, or communities should be balanced carefully with potential benefits of the knowledge to be gained through GWAS. The sensitive nature of GWAS information about participants and the broad data distribution goals of the NIH GWAS data repository highlight the importance of the informed consent process to this research.

The NIH recognizes that scientific, ethical and societal issues relevant to this policy are evolving, and the agency has established on-going mechanisms to oversee GWAS policy implementation across the agency and to monitor whole genome association data use practices. The NIH will revisit and revise the policy and related practices as appropriate.

Preamble: Summary of Public Comments on Proposed Policy

On August 30, 2006, the NIH published the *Proposed Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS)* for public comment in the Federal Register and the *NIH Guide for Grants and Contracts*. The comment period ended with a Town Hall meeting held in Bethesda, Maryland on December 14, 2006, that was attended by a total of 374 people (on-site and via webcast).

Overall the NIH received 196 written comments from professional societies, patient advocacy groups, privacy groups, individual scientists, and private citizens. The comments reflected a variety of interests and perspectives. In developing policies, the NIH strives to be respectful of the diversity of individual and group interests, incorporating appropriate protections while promoting maximum public benefit from the research it sponsors. The NIH GWAS policy and its implementation are expected to evolve in response to advances in scientific knowledge, available technologies, and the legal and ethical issues they raise.

I. Rationale for a Centralized Data Repository

Respondents asked for clarification of the rationale for creation of a central data repository instead of distributed repositories under the control of individual (and non-governmental) institutions and investigators. Concerns expressed about a central data repository included, for example, the resources required to maintain it and the extent to which it would duplicate efforts and resources already invested by multiple institutions.

The advantages and limitations of central versus distributed data repositories have been discussed extensively at the NIH. From a scientific standpoint, a central repository offers a number of important advantages: tighter and more consistent control over the standards and quality of the genotype and phenotype data included; the ability to standardize and update terminology and format as technology and methodology improve; consistent, defined and transparent security and standards for access to data; a long-term commitment to maintenance of data after studies have been completed; a common point of entry for all investigators who use the data; a consistent and defined approach to removal of data in the event of withdrawal of participant consent; facilitation of meta-analyses and analyses that use data from multiple studies; and the ability to implement consistent participant protections at the level of data submission and data access. Individual investigators and many institutions may lack sufficient resources to ensure consistency and quality control, or a long-term commitment to data storage and access. One of the potential disadvantages of a central repository residing at NIH is that the data may be accessible through the Federal Freedom of Information Act (FOIA), unless they are exempt from release under one of the FOIA exemptions. This is further discussed in the Protection of Research Participants section below.

As clinical and genomics research progresses, genotype and phenotype data are being collected into databases maintained by a variety of investigators, studies, and institutions. The NIH is concerned that the present situation may provide less consistent standards for the protection of research participants, data quality, and data access than would a central repository. However, the NIH recognizes that other databases will be designed to achieve different scientific aims or to integrate different analytic capacities, and the NIH GWAS policy is not intended to constrain the development of such databases or to curtail the deposition of NIH-supported GWAS data into other databases (as may be appropriate or required for some research programs). Among the on-going charges to the trans-NIH Technical Standards Steering Committee established through the GWAS governance structure (see Oversight and Governance section below) will be explicit consideration of the evolving technical capacities and interoperability needed to facilitate the submission of data into the NIH GWAS data repository [2] through other major database systems (e.g., the NCI caBIG network). This committee also will provide a forum for inter-IC coordination of data structures and standards to maintain interoperability of NIH databases.

II. Protection of Research Participants

Non-research Use of Data

Respondents noted that data held by the Government are subject to the FOIA, and thus could be obtained outside of the Controlled Access data request process described in the GWAS policy. Respondents expressed concern that data could be obtained for non-research purposes (e.g., by law enforcement agencies, employers, or insurance companies) or for purposes beyond the scope of the research uses envisioned within the GWAS policy.

As an agency of the Federal Government, the NIH is required to release Government records in response to a request under the FOIA, unless they are exempt from release under one of the FOIA exemptions. Although the NIH-held data will be coded and the NIH will not hold direct identifiers to individuals within the NIH GWAS data repository, the agency recognizes the personal and potentially sensitive nature of the genotype-phenotype data. Further, the NIH takes the position that technologies available within the public domain today, and technological advances expected over the next few years, make the identification of specific individuals from raw genotype-phenotype data feasible and increasingly straightforward.

The agency believes that release of unredacted GWAS datasets in response to a FOIA request would constitute an unreasonable invasion of personal privacy under FOIA Exemption 6, 5 U.S.C. § 552 (b)(6). Therefore, among the safeguards that the NIH foresees using to preserve the privacy of research participants and confidentiality of genomic data is the redaction of individual-level genotype and phenotype data from disclosures made in response to FOIA requests and the denial of requests for unredacted datasets.

In addition, the NIH acknowledges that legitimate requests for access to data made by law enforcement offices to the NIH may be fulfilled. The NIH will not possess direct identifiers within the NIH GWAS data repository, nor will the NIH have access to the link between the data keycode and the identifiable information that may reside with the primary investigators and institutions for particular studies. The release of identifiable information may be protected from compelled disclosure by the primary investigator's institution if a Certificate of Confidentiality is or was obtained for the original study. Within the final GWAS policy, the NIH explicitly encourages investigators to consider the potential appropriateness of obtaining a Certificate of Confidentiality as an added measure of protection against future compelled disclosure of identities for studies planning to collect genome-wide association data.

Stigmatization

Respondents commented that some data to be included in the repository may be highly sensitive because they may suggest the existence either of individually identifiable or socially undesirable traits. These data have implications for both participants and family members.

Tools for analysis of genomic data increasingly are able to make inferences about some individual traits (e.g., height, weight, skin and hair and eye color) and to identify predilections for characteristics (e.g., risk of developing some diseases) and behaviors with social stigma. In recognition of these risks, the NIH policy includes steps to protect the interests and privacy concerns of individuals, families and identifiable groups who participate in GWAS research. The NIH is asking institutions submitting GWAS datasets to certify that an Institutional Review Board (IRB) and/or Privacy Board (as applicable) has considered such risks and that investigators have stripped the data of all identifiers before the data are submitted. The NIH Data Access Committees (DACs) will approve access only for research uses that are consistent with an individual's consent as defined by the submitting institution. In addition, in the event that requests raise questions or concerns related to privacy and confidentiality, risks to populations or groups, or other relevant topics, the DACs will consult with other experts as appropriate.

Informed Consent

Respondents asked for clarification regarding appropriate informed consent processes and consent documentation for individuals participating in studies for which data are to be submitted to the NIH GWAS data repository. Concern was raised that participants may not be aware of the potential privacy risks associated with placement of their genotype and phenotype data in a central repository at the NIH. Respondents also commented that adequate consent for data sharing requires participants to understand both the risks and potential benefits of the proposed sharing. Key stakeholders in these considerations are: research participants (both those who have participated in on-going or prior studies for which GWAS were not anticipated and those who may participate in prospective GWAS); investigators developing informed consent processes; institutions approving the submission of datasets to the NIH GWAS data repository; and IRBs asked to review studies proposing genome-wide association analysis. Respondents commented that additional institutional resources are likely to be required if additional consent is needed for data sharing.

As noted elsewhere and reflected in the GWAS oversight structure established to manage implementation of the GWAS policy (see Oversight and Governance section below), the NIH recognizes that the ethical considerations relevant to GWAS data sharing are complex and dynamic. Therefore, the NIH is developing informational materials [3] as a resource for IRBs and institutions for their consideration of the issues relevant to reviewing and approving individual studies proposing data submission to the NIH GWAS data repository. The NIH intends to continue to engage the Office for Human Research Protections, the research community, and the public to explore the participant protection issues related to GWAS and to identify best practices for the consideration and risk-benefit analysis of genotype and phenotype data sharing under this policy. These efforts will include discussion of the optimal methods for communicating with participants about relevant issues through the informed consent process for prospective studies, and discussion of issues to consider in the institutional review of consent materials for use of existing samples or data proposed for GWAS. Participant interests relevant to GWAS data sharing extend beyond individual participants to families, communities, and their respective cultural sensitivities. The NIH believes that institutional deliberations regarding data submission to the NIH GWAS data repository should include these broader interests. Further, especially complex issues exist with regard to GWAS where participant consent has been

provided by proxy (e.g., pediatric research or some studies involving mental health disorders). Discussion of this topic will be included in the informational materials [3] that the NIH is developing for submitting institutions and IRBs asked to review proposed GWAS.

The GWAS policy applies to genome-wide association research utilizing genetic materials and data collected both prospectively and retrospectively. For prospective studies, in which GWAS are conceived within the study designs at the time research participants provide their consent, the NIH expects specific discussion within the informed consent process and documentation that participants' genotype and phenotype data will be shared for research purposes through the NIH GWAS data repository. For retrospective studies performed using existing genetic materials and previously collected data, the NIH anticipates considerable variation in the extent to which data sharing and future genetic research have been addressed within the informed consent documents. As described in the policy, the submitting institution will determine whether a study is appropriate for submission to the NIH GWAS data repository (including an IRB and/or Privacy Board review of specific study elements, such as participant consent). The NIH anticipates that a number of GWAS proposing to include pre-existing data or samples may require additional consent of the research participants. The NIH may give programmatic consideration to requests for funds or other resources needed to conduct additional participant consent when appropriate.

In the event that participants withdraw consent for sharing of their individual-level genotype and phenotype data through the NIH GWAS data repository, the submitting institution will be responsible for alerting the NIH GWAS data repository and requesting that the specific record be removed from future data distributions. However, data that have been distributed to researchers will not be retracted.

Return of Results

Respondents asked for clarification of plans *for* return of results to study participants.

The NIH does not anticipate that participants will be able to obtain individual results of secondary analyses on data obtained from their participation in primary studies. Because the NIH GWAS data repository and secondary data users will not have access to identifying information or to the link to the keycode within the data, neither will be able to return individual results directly to subjects. Secondary investigators may share their findings with primary investigators, who may determine whether it is appropriate to return individual or aggregate research results to participants whose health may be affected, following established institutional procedures (e.g., IRB approval) and specific parameters defined within the original study.

Oversight and Governance of the NIH GWAS Data Repository, Submission and Access

Some respondents commented on the importance of adequate oversight of policies for data submission and access, and on the details of the repository. A need for oversight of the quality control measures for genotype and phenotype data and of the security measures for the repository was noted by many respondents. Some respondents commented on the importance of the policies established by the Data Access Committees, and their function within the Institutes and Centers.

The NIH has developed a governance structure for GWAS that provides oversight tailored to the specific role involved. The NIH Director will oversee the GWAS policy and its implementation. In carrying out this responsibility, the NIH Director will be informed by a Senior Oversight Committee composed of Institute and Center (IC) Directors and appropriate leadership from within the Office of the Director. The Senior Oversight Committee will be responsible for the on-going management and stewardship of GWAS policy and operating implementation procedures across ICs. Reporting to the Senior Oversight Committee will be two Steering Committees charged with the implementation, communication, and development of specific procedures related to the conduct, submission and data release practices for GWAS supported by the NIH. One of these groups, the Research Participant Protection and Data Management Steering Committee, will include among its members the chairs of all Data Access Committees at the NIH as well as appropriate staff from NIH policy and oversight offices (e.g., the Office of Science Policy and the Office of Human Subjects Research). This committee will work to promote consistent and robust participant protections across relevant NIH programs. The second group, the Technical Standards Steering Committee, will include membership from scientific programs across the NIH as well as staff from the National Center for Biotechnology Information. This committee will focus on the challenges and needs associated with building and maintaining the NIH GWAS data repository and on formulating or stimulating the consideration of data standards (for genotype or

phenotype data) where appropriate. Critical input from individual genome-wide association research programs and studies will be provided to the two Steering Committees through the ICs' Data Access Committees or other project oversight bodies created for specific studies, e.g., community representative groups, scientific advisory boards.

In order to maintain GWAS policy consistent with evolving technological and ethical considerations, the NIH Director will solicit recommendations on the policy from external experts representing public and scientific stakeholders through the Advisory Committee to the Director.

III. Scientific Publication

Some respondents commented on the considerable logistical difficulties posed by limiting the period of publication exclusivity, particularly considering the complexity of many of the studies and the lag time between submission and publication of peer-reviewed scientific papers. Some respondents were concerned that submitting investigators would not receive appropriate credit for their work, and would have insufficient control over use of their data. Concern was expressed about enforcing compliance with publication policies. Some respondents commented that the limited period of exclusivity could stimulate a rush to publish initial analyses prematurely, deterring subsequent studies and reducing the overall quality of the reports.

The NIH initially proposed that GWAS datasets be made available as soon as appropriate quality control measures (as defined for a given NIH program) were complete, and that a 9-month period of exclusivity would exist for primary investigators to submit analyses of GWAS datasets for publication. The NIH believes that an extended period of exclusivity would undermine the potential benefits of data sharing. However, in response to concerns raised through the public comment process, the NIH has lengthened this exclusivity period to 12-months in the final policy. The publication exclusivity period will commence on the date that a GWAS dataset is first made available through the NIH GWAS data repository, and the expiration date of this time period will be featured prominently in all descriptions and overviews of the dataset provided through both the public and controlled access pathways of the NIH GWAS data repository. The policy now is explicit on the inclusion within this exclusivity period of electronic and other means of information dissemination beyond peer-reviewed publications. As part of an overarching desire for transparency in the use of GWAS datasets, the names, institutional affiliations, and Data Access Committee-approved research uses for all GWAS data users will be available to the public within the NIH GWAS data repository. GWAS data users will be encouraged to collaborate with the primary investigators for GWAS as appropriate. The period of exclusivity is consistent with existing practices for other genome-wide association programs already available or in the pipeline for deposition into the NIH GWAS data repository, and is intended only as an upper limit as some NIH programs may stipulate shorter (or no) publication exclusivity timelines. The NIH anticipates that over time investigators will become more comfortable with the GWAS data sharing policy as the benefits of greater research access to the data are realized.

IV. Intellectual Property



Respondents raised concerns that the policy might diminish the intellectual property rights of the submitting investigators, as well as their ability to obtain patents. Some respondents questioned whether the proposed policy text is a violation of the Bayh-Dole Act.



The NIH believes that the intellectual property section of the policy presents no conflict with, or infringement upon, rights granted by the Bayh-Dole Act or any other federally-created intellectual property rights. Funding recipients are still able to elect title to any inventions or discoveries developed under the respective federal funding agreements that are or may be patentable, consistent with the Bayh-Dole Act and NIH policies. The NIH expects that intellectual property issues or questions that may occur will be resolvable through appropriate negotiations under the rubrics provided previously in NIH guidance to the research community within the Research Tools Policy and the Best Practices for the Licensing of Genomic Inventions. The NIH encourages development of new diagnostics, therapeutics, or other interventions building on basic discoveries, and believes they will be enabled through the NIH GWAS data repository. The NIH anticipates that downstream technology development opportunities will increase as a result of broad research access to genotype-phenotype associations provided through the GWAS policy. The NIH has engaged in informal discussions with academic and private sector experts in intellectual property, these interactions, as well as formal responses received from stakeholders through the GWAS public

consultation process, have suggested that the GWAS policy is consistent with existing practices and can be expected to better promote the development of exciting new discoveries for the public benefit.

Policy for Genome-Wide Association Studies (GWAS)

I. Principles

The NIH is interested in advancing genome-wide association studies (GWAS) to identify common genetic factors that influence health and disease. For the purposes of this policy, a genome-wide association study is defined as any study of genetic variation across the entire human genome that is designed to identify genetic associations with observable traits (such as blood pressure or weight), or the presence or absence of a disease or condition [1]. Whole genome information, when combined with clinical and other phenotype data, offers the potential for increased understanding of basic biological processes affecting human health, improvement in the prediction of disease and patient care, and ultimately the realization of the promise of personalized medicine. In addition, rapid advances in understanding the patterns of human genetic variation and maturing high-throughput, cost-effective methods for genotyping are providing powerful research tools for identifying genetic variants that contribute to health and disease.

Consistent with the NIH mission to improve public health through research, the NIH believes that the full value of GWAS to the public can be realized only if the genotype and phenotype datasets are made available as rapidly as possible to a wide range of scientific investigators. Rapid and broad data access is particularly important for GWAS because of the significant resources they require; the challenges of analyzing large datasets; and the extraordinary opportunities for making comparisons across multiple studies.

Protection of research participants is a fundamental principle underlying biomedical research. The NIH is committed to responsible stewardship of data throughout the research process, which is essential to protecting the interests of study participants and to maintaining public trust in biomedical research.

In consideration of the evolving scientific, ethical, and societal issues related to this policy, the NIH is establishing a governance structure for NIH GWAS activities that will:

- Ensure ongoing, high-level agency oversight; and
- Obtain regular input from public representatives, including those with expertise in bioethics, privacy, data security, and appropriate scientific and clinical disciplines; and
- Revisit and revise the policy as appropriate.

II. Applicability

This NIH policy applies to:

- Competing grant applications that include GWAS and are submitted to the NIH for the January 25, 2008, and subsequent receipt dates;
- Proposals for contracts that include GWAS and are submitted to the NIH on or after January 25, 2008; and
- NIH intramural research projects that include GWAS and are approved on or after January 25, 2008.

An application or proposal will be identified as GWAS by applicants and/or NIH staff (see NOT-OD-06-071).

III. Data Management

Data Repository

To facilitate broad and consistent access to NIH-supported GWAS datasets, the NIH has developed a central NIH GWAS data repository [2] at the National Center for Biotechnology Information (NCBI), National Library of Medicine. The repository

will provide a single-point of access to basic information about NIH-supported GWAS and to available genotype-phenotype datasets for GWAS. Although the NIH envisions that access to all NIH-supported GWAS datasets will be possible through this repository, it does not intend the repository to become the exclusive point of data submission for these data, nor does it intend the central database to delimit the structures or tools that may be appropriate for other similar databases. The repository also will accept GWAS datasets contributed from other sources.

To ensure the security of the data held by the repository, the NCBI will employ multiple tiers of data security (such as sequential firewalls and independent networks) based on the content and level of risk associated with the data. The NIH will establish and maintain operating policies and procedures for the repository to address issues including, but not limited to, the privacy and confidentiality of GWAS research participants, the interests of individuals and groups, data access procedures, and data security mechanisms. These will be reviewed periodically by the GWAS oversight bodies.

Data Submission

All investigators who receive NIH support to conduct genome-wide analysis of genetic variation in a study population are expected to submit to the NIH GWAS data repository descriptive information about their studies for inclusion in an open access portion of the NIH GWAS data repository. All data and information will be submitted to a high security network within the NCBI through a secure transmission process. Submissions should include the following:

- the protocol,
- questionnaires,
- study manuals,
- variables measured, and
- other supporting documentation.

In addition, the NIH strongly encourages the submission of curated and coded phenotype, exposure, genotype, and pedigree data, as appropriate, to the NIH GWAS data repository as soon as quality control procedures have been completed at the local institution. These detailed data will be made available through a controlled access process according to the GWAS Data Access procedures (described in Data Access section below). Investigators who elect to submit their GWAS data to additional data repositories or networks should verify that appropriate data security, confidentiality, and privacy measures are in place for protection of GWAS participants. Irrespective of where the data are submitted, researchers submitting GWAS data are encouraged to consider whether a Certificate of Confidentiality might be appropriate for their data as an additional safeguard with regard to involuntary disclosure of the research participant identities. Further information about Certificates of Confidentiality is available at the following website: http://grants.nih.gov/grants/policy/coc/.

In order to minimize risks to study participants, data submitted to the NIH GWAS data repository will be de-identified and coded using a random, unique code. Data should be de-identified according to the following criteria: the identities of data subjects cannot be readily ascertained or otherwise associated with the data by the repository staff or secondary data users (45 C.F.R. 46.102(f)); the 18 identifiers enumerated at section 45 C.F.R. 164.514(b)(2) (the HIPAA Privacy Rule) are removed; [4] and the submitting institution has no actual knowledge that the remaining information could be used alone or in combination with other information to identify the subject of the data. Keys to codes will be held by submitting institutions. Submissions of GWAS data should be accompanied by a written certification (detailed below) stating that the identities of research participants will not be disclosed to the NIH GWAS data repository. Therefore, the NIH GWAS data repository will be unable to provide individual research results derived from analyses of submitted data to participants. General information regarding known publications analyzing GWAS datasets will be made available through the repository.

All submissions to the NIH GWAS data repository should be accompanied by a certification by the responsible Institutional Official(s) of the submitting institution that they approve submission to the NIH GWAS data repository [2]

The certification should assure that:

- The data submission is consistent with all applicable laws and regulations [5], as well as institutional policies ;
- The appropriate research uses of the data and the uses that are specifically excluded by the informed consent

documents are delineated;

- The identities of research participants will not be disclosed to the NIH GWAS data repository; and
- An IRB and/or Privacy Board, as applicable, reviewed and verified that:
    - The submission of data to the NIH GWAS data repository and subsequent sharing for research purposes are consistent with the informed consent of study participants from whom the data were obtained;
    - The investigator's plan for de-identifying datasets is consistent with the standards outlined above;
    - It has considered the risks to individuals, their families, and groups or populations associated with data submitted to the NIH GWAS data repository; and
    - The genotype and phenotype data to be submitted were collected in a manner consistent with 45 C.F.R. Part 46.

While the NIH encourages data sharing through this policy, circumstances beyond the control of investigators may preclude submission of GWAS data to the NIH GWAS data repository. Applications submitted to the NIH for support of GWAS in which the above expectations for data submission cannot be met will be considered for funding on a case-by-case basis by the appropriate IC.

Submitting investigators and their institutions may request removal of data on individual participants from the NIH GWAS data repository in the event that a research participant withdraws his or her consent. However, data that have been distributed for approved research use will not be retrieved.

Data Access

The basic descriptive and aggregate summary information submitted to the NIH GWAS data repository for each NIH-supported or conducted GWAS will be available publicly through the NIH GWAS data repository. Access to the genotype and phenotype datasets submitted and stored in the NIH GWAS data repository, along with appropriate automated calculations (e.g., quality control measures, simple genotype-phenotype associations, or a listing of all variants known to be in linkage disequilibrium [6] with variants measured in the genotype), will be provided for research purposes through an NIH Data Access Committee (DAC). Membership of the DACs will include Federal staff with relevant expertise in areas such as the relevant particular scientific disciplines, research participant protection, and privacy. The NIH anticipates that individual DACs may be established based on programmatic areas of interest and the relevant needs for technical and ethics expertise. All DACs will operate according to common principles and follow similar procedures to ensure the consistency and transparency of the GWAS data access process.

Investigators and institutions seeking data from the NIH GWAS data repository will be expected to meet data security measures (such as physical security, information technology security, and user training) and will be asked to submit a data access request, including a Data Use Certification, that is co-signed by the investigator and the designated Institutional Official(s). Data access requests should include a brief description of the proposed research use of the requested GWAS dataset(s). Within a Data Use Certification investigators will agree, among other things [7], to:

- Use the data only for the approved research;
- Protect data confidentiality;
- Follow appropriate data security protections;
- Follow all applicable laws, regulations and local institutional policies and procedures for handling GWAS data;
- Not attempt to identify individual participants from whom data within a dataset were obtained;
- Not sell any of the data elements from datasets obtained from the NIH GWAS data repository;
- Not share with individuals other than those listed in the request any of the data elements from datasets obtained from the NIH GWAS data repository;
- Agree to the listing of a summary of approved research uses within the NIH GWAS data repository along with his or her name and organizational affiliation;
- Agree to report, in real time, violations of the GWAS policy to the appropriate DAC;
- Acknowledge the GWAS policy with regard to publication and intellectual property; and
- Provide annual progress reports on research using the GWAS dataset.

Data Access Committees or their designees will review requests for access to determine whether the proposed use of the dataset is scientifically and ethically appropriate and does not conflict with constraints or informed consent limitations identified by the institutions that submitted the dataset to the NIH GWAS data repository. In the event that requests raise concerns related to privacy and confidentiality, risks to populations or groups, or other concerns, the DAC will consult with other experts as appropriate.

IV. Publication

The NIH expects that investigators who contribute data to the NIH GWAS data repository will retain the exclusive right to publish analyses of the dataset for a defined period of time following the release of a given genotype-phenotype dataset through the NIH GWAS data repository (including the pre-computed analyses of the data). During this period of exclusivity, the NIH will grant access through the DACs to other investigators, who may analyze the data, but are expected not to submit their analyses or conclusions for publication during the exclusivity period. The maximum period of exclusivity is twelve months from the date that the GWAS dataset is made available for access through the NIH GWAS data repository, although a shorter period of exclusivity may be determined by the NIH funding IC. Contributing investigators are encouraged to shorten the period of publication exclusivity at their own discretion. Publication exclusivity is expected to extend to all forms of public disclosure, including meeting abstracts, oral presentations, and publicly accessible electronic submissions (e.g., websites, web blogs). Following expiration of the exclusive publication period for a given GWAS dataset, the NIH expects that all investigators with access to the data may submit publications or present analyses for any purpose consistent with the practices and policies of their institutions and the NIH.

The NIH also expects all investigators who access GWAS datasets to acknowledge the Contributing Investigator(s) who conducted the original study, the funding organization(s) that supported the work, and the NIH GWAS data repository in all resulting oral or written presentations, disclosures, or publications of the analyses.

V. Intellectual Property

It is the hope of the NIH that genotype-phenotype associations identified through NIH-supported and NIH-maintained GWAS datasets and their obvious implications will remain available to all investigators, unencumbered by intellectual property claims. The NIH discourages premature claims on pre-competitive information that may impede research, though it encourages patenting of technology suitable for subsequent private investment that may lead to the development of products that address public needs.

The NIH will provide approved GWAS data users with certain automated calculations (described under the Data Access section) as a component of the GWAS datasets distributed through the NIH GWAS data repository.

The NIH expects that NIH-supported genotype-phenotype data made available through the NIH GWAS data repository and all conclusions derived directly from them will remain freely available, without any licensing requirements, for uses such as, but not necessarily limited to, markers for developing assays and guides for identifying new potential targets for drugs, therapeutics, and diagnostics. The intent is to discourage the use of patents to prevent the use of or block access to any genotype-phenotype data developed with NIH support. The NIH encourages broad use of NIH-supported genotype-phenotype data that is consistent with a responsible approach to management of intellectual property derived from downstream discoveries, as outlined in the NIH's Best Practices for the Licensing of Genomic Inventions and its Research Tools Policy.

The filing of patent applications and/or the enforcement of resultant patents in a manner that might restrict use of NIH-supported genotype-phenotype data could diminish the potential public benefit they could provide. Approved users and their institutions, through the execution of an NIH Data Use Certification, will acknowledge the goal of ensuring the greatest possible public benefit from NIH-supported GWAS.

Expectations Defined in the Policy for Investigators

The detailed expectations are enumerated in the individual sections of this policy, and summarized as follows:

Investigators submitting GWAS data are expected to:

- Provide descriptive information about their studies;
- Submit coded genotypic and phenotypic data to the NIH GWAS data repository; and
- Submit certification by the Institutional Official(s) of the responsible submitting institution that it has reviewed and approved submission to the NIH, noting any limitations on data use based on the relevant informed consents and providing assurance that all data are submitted to the NIH in accord with applicable laws and regulations and that the identities of research participants will not be disclosed to the NIH GWAS data repository.

Investigators requesting and receiving GWAS data are expected to:

- Submit a description of the proposed research project;
- Submit a data access request, including a Data Use Certification co-signed by the designated Institutional Official(s) at their sponsoring institution;
- Protect data confidentiality;
- Ensure that data security measures are in place;
- Notify the appropriate Data Access Committee of policy violations; and
- Submit annual progress reports detailing significant research findings.

Inquiries

Specific questions about this Notice should be directed to:

Laura Lyman Rodriguez, Ph.D.
Special Advisor to the Director
National Human Genome Research Institute
31 Center Drive, Room 4B09
Bethesda, MD 20892
Phone: 301-496-0844

Sam Shekar, M.D., M.P.H.
Assistant Surgeon General and
Director, Office of Extramural Programs
Office of Extramural Research
1 Center Drive
Bethesda, MD 20892
Phone: 301-435-3492

Email inquiries should be directed to GWAS@nih.gov

Additional information and detailed implementation guidance related to the NIH GWAS Policy will be provided at http://grants.nih.gov/grants/gwas/index.htm.

Endnotes

---

[1] To meet the definition of a GWAS, the density of genetic markers and the extent of linkage disequilibrium should be sufficient to capture (by the r2 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.

[2] Currently named the NIH database of Genotypes and Phenotypes (dbGaP) (http://www.ncbi.nlm.nih.gov/entrez/query/Gap/gap_tmpl/about.html)

[3]The NIH anticipates releasing additional GWAS implementation documents in the next few months, including a *Points to Consider* document on issues related to the submission of data to the repository.

[4]The identities of data subjects cannot be readily ascertained or otherwise associated with the data by the repository staff or secondary data users (Common Rule); and the following data elements have been removed (HIPAA Privacy Rule).

1. Names.
2. All geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP Code, and their equivalent geographical codes, except for the initial three digits of a ZIP Code if, according to the current publicly available data from the Bureau of the Census: a. The geographic unit formed by combining all ZIP Codes with the same three initial digits contains more than 20,000 people. b. The initial three digits of a ZIP Code for all such geographic units containing 20,000 or fewer people are changed to 000.
3. All elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death; and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older.
4. Telephone numbers.
5. Facsimile numbers.
6. Electronic mail addresses.
7. Social security numbers.
8. Medical record numbers.
9. Health plan beneficiary numbers.
10. Account numbers.
11. Certificate/license numbers.
12. Vehicle identifiers and serial numbers, including license plate numbers.
13. Device identifiers and serial numbers.
14. Web universal resource locators (URLs).
15. Internet protocol (IP) addresses numbers.
16. Biometric identifiers, including fingerprints and voiceprints.
17. Full-face photographic images and any comparable images.
18. Any other unique identifying number, characteristic, or code, unless otherwise permitted by the Privacy Rule for re-identification

In addition, the submitting institution should have no actual knowledge that the remaining information could be used alone or in combination with other information to identify the individual who is the subject of the information.

[5]Applicable Federal regulations may include HHS human subjects regulations (45 CFR Part 46), FDA human subjects regulations (21 CFR Parts 50 and 56), and the Health Insurance Portability and Accountability Act Privacy Rule (45 CFR Part 160 and Part 164, Subparts A and E).

[6] Linkage disequilibrium information will be based on data from the International HapMap Project (http://www.hapmap.org/).

[7] Investigators requesting access to GWAS datasets who also have access to identifying information for the individuals within the dataset will require IRB approval.

Weekly TOC for this Announcement
NIH Funding Opportunities and Notices

Office of Extramural Research (OER)

National Institutes of Health (NIH) 9000 Rockville Pike Bethesda, Maryland 20892

Department of Health and Human Services (HHS)

Note: For help accessing PDF, RTF, MS Word, Excel, PowerPoint, RealPlayer or Video files, see Help Downloading Files.