NHGRI-dbGaP Data Use Certification
Gene Environment Association Studies (GENEVA), part of the Genes, Environment, and Health Initiative (GEI)

Nurses' Health Study and Health Professionals' Follow-up Study Genes and Environment Initiatives in Type 2 Diabetes (NHS-HPFS Diabetes)

(7/23/2009 version)

Introduction and Statement of Policy

The **d**ata**b**ase of **G**enotypes **a**nd **P**henotypes (dbGaP) was developed to archive and distribute the results of studies provided by Contributing Investigators examining the relationship between genotype and phenotype. Such studies include genome-wide association studies, medical sequencing, and molecular diagnostic assays. Implicit in dbGaP is the view that scientific progress in this area will be greatly enhanced if the data produced by these studies are readily available to all investigators in the research community.

Access to controlled data in dbGaP will be granted by an NIH Data Access Committee (DAC). Users requesting access to controlled data must submit a Data Access Request (DAR) to the appropriate NIH DAC for approval. DAC approval for controlled data access will be dependent upon completion of the DAR, agreeing to the terms and conditions in the Data Use Certification (DUC), and confirmation that the proposed research use is consistent with any restrictions on data use identified by the institutions that submitted the dataset(s) to dbGaP.

Dataset access will only be provided to investigators who, along with their institutions, have certified their agreement with the requirements and terms of access detailed below. It is the intent of the NIH and NHGRI that Approved Users of NIH GWAS datasets recognize the restrictions on data use imposed by the original informed consent agreements of contributing studies, as identified by the submitting institutions and stated on dbGaP.

To promote the responsible use of the NIH GWAS datasets, all investigators and their institutions seeking access shall acknowledge their agreement with the policies and procedures articulated within this Data Use Certification. For collaborative projects, any independent Collaborating Requester from a separate institution shall complete a separate Data Access Request (DAR). If Collaborating Requesters are from the same institution, only one DAR and the appropriate DUC for each requested dataset need be submitted.

Definitions of terminology used in this document are found in the NIH GWAS Glossary at http://grants.nih.gov/grants/gwas/gwas_general_brochure.pdf .

The parties between whom this agreement is made include: the research investigator requesting access to the NHS-HPFS Diabetes GWAS dataset ("Requester"), his/her Institutional Signing Official as designated by the Institution through the eRA Commons system, and the NHGRI, National Institutes of Health.

Terms of Access

1. Research Use

The Requester agrees that if access is approved, he/she shall become an Approved User of the NHS-HPFS Diabetes GWAS dataset and research use will occur solely in connection with the research project described in the attached Data Access Request (DAR) submitted to the NIH, which includes the project title, the Requester's name and institution, the names of any independent collaborators at the same institution (named as "Senior/Key Persons"), and a 1-2 paragraph description of the research objectives and design including consistency with any data use limitations. New uses of these data outside those described in the DAR will require submission of a new DAR; modifications to the research project will require submission of an amendment to this application (e.g., the addition of new aims related to the approved project, adding or deleting collaborators from the same institution, and the potential addition of new NIH GWAS datasets to an approved project). The Requester further agrees that he/she shall use the dataset(s) only in accordance with the parameters described on the dbGaP Web site for the appropriate research use, and any limitations on such use, of the dataset(s) and as required by law.

Research data access to the NHS-HPFS Diabetes GWAS dataset is granted for a period of one year as defined below.

Contributing Investigators, or their direct collaborators, who provided the data or samples used to generate a particular NIH GWAS dataset and who have appropriate IRB approval, if applicable, for broader use of the data are exempt from the limitation on the scope of the research use as defined in the DAR.

2. Institutional and Approved User Responsibilities

The Requester agrees through the submission of the Data Access Request that he or she has reviewed and understands the principles for responsible research use and data handling of the genotype and phenotype data included within NIH GWAS datasets as defined in the NIH GWAS Data Sharing Policy and detailed in this Data Use Certification agreement. The Requester and his or her institution further acknowledge that they are responsible for ensuring that all uses of the data are consistent with federal, state, and local laws and regulations and any relevant institutional policies. Through submission of the DAR, the Requester also agrees to submit annual reports to the appropriate NIH Data Access Committee (DAC) describing his or her research use as an Approved User of the NHS-HPFS Diabetes GWAS dataset as described under *"Research Use Reporting"* below.

Requesters who may have access to personal identifying information for research participants in the original study, at their institution or through their collaborators, may be required to have IRB approval for their research protocol. By approving and submitting the attached Data Access Request, the Institutional Signing Official provides assurance that relevant institutional policies and applicable federal, state, or local laws and regulations (if any) have been followed, including IRB approval if required. The Institutional Signing Official also assures through the approval of the Data Access Request that other organizations within the institution with relevant authorities (e.g., the Office of Technology Transfer, the Chief Information Officer, the Office of Human Subjects Research, etc.) have reviewed the relevant sections of the NIH GWAS Data Sharing Policy and the associated procedures and are in agreement with the principles defined.

It is anticipated that, at least in some cases, these datasets will be updated with additional information. Unless otherwise indicated, all statements herein are presumed to be true and applicable to the access and use of all versions of these datasets.

3. Public Posting of Approved User's Research Use Statement

The Requester agrees that, if he or she becomes an Approved User, information about the proposed research use can be posted on a public web site that describes the projects. The information may include the Approved User's name and institution, project name, Research Use Statement, and a Non-technical Summary of the Research Use Statement. In addition, citations resulting from the use of NIH GWAS datasets will be posted on the NIH GWAS Data Repository website.

4. Non-Identification

The Requester, if he or she becomes an Approved User, agrees not to use NIH GWAS datasets, either alone or in concert with any other information, to identify or contact individual participants from whom phenotype data and DNA samples were collected. This agreement does not apply to research investigators operating with specific IRB approval to contact individuals within datasets under an approved IRB research protocol.

5. Non-Transferability

The Requester agrees to retain control over the data as an Approved User, and further agrees not to distribute individual-level data in any form to any entity or individual other than his/her research staff or trainees or independent collaborating investigators listed in the attached Data Access Request under "Research & Related Senior/Key Person Profile." If Approved Users are provided access to NIH GWAS datasets for inter-institutional collaborative research described in the research use statement of the Data Access Request, and all members of the collaboration are also Approved Users, derivatives of the dataset may be securely transmitted within the collaborative group for the purpose of quality control or other research requirements. All data security practices and other terms of use defined in this agreement and the *dbGaP Security Best Practices* for the original data are expected to be followed for the derived data, including any transmission of the data.

The Requester/Approved User and his/her institution acknowledge responsibility for ensuring the review and agreement to the terms within this Data Use Certification and the appropriate research use of NIH GWAS data by research staff associated with any approved GWAS Project, subject to applicable laws and regulations. NIH GWAS Datasets containing individual-level information, in whole or in part, may not be sold to any individual at any point in time for any purpose.

The Requester agrees that if he or she changes institutions during the period as an Approved User, a new Data Access Request and Data Use Certification from the new institution will be submitted and approved before data use resumes. Any versions of data stored at the prior institution will be destroyed and documented through a final Data Use Report as described below. However, if advance written notice and approval by the NHGRI Data Access Committee is obtained to transfer responsibility for the approved research project to another Approved User within the same institution, the data may not need to be destroyed.

6. Data Security and Data Release Reporting

The Requester and his/her institution acknowledge that they, the Information Technology Director or his/her designee, and all other individuals named in the Data Access Request have reviewed and agree to handle the NHS-HPFS Diabetes GWAS dataset(s) according to the current *dbGaP Security Best Practices*, including its detailed description of requirements for security and encryption. This includes, but is not limited to:

- all approved users have completed all required institutional computer security training, e.g. http://irtsectraining.nih.gov/, or the equivalent;

- the data will always be physically secured (e.g. through camera surveillance, locks on doors/computers, security guard);
- servers must not be accessible directly from the internet (i.e. must be behind a firewall or not connected to a larger network), and unnecessary services disabled;
- use of portable media, e.g., on a CD, flash drive or laptop, is discouraged; if necessary then they must be encrypted;
- updated anti-virus/anti-spyware software;
- security auditing/intrusion detection software, detection and regular scans of potential data intrusions;
- strong password for file access and never share it.
- All copies of the dataset must be destroyed whenever any of the following occurs:
  - the DUC expires;
  - the NHGRI requests destruction of the dataset;
  - the continued use of the data would no longer be consistent with the DUC.

In addition, if the Requester becomes an Approved User, he/she agrees along with the institution to keep the data secure and confidential at all times and to adhere to information technology practices in all aspects of data management to ensure that only authorized individuals can gain access to NIH GWAS datasets. This agreement includes the maintenance of appropriate controls over any copies of the data.

Approved Users and his/her institution agree to notify the NHGRI Data Access Committee of any unauthorized data sharing, breaches of data security, or inadvertent data releases that may compromise data confidentiality within 24 hours of when the incident is identified. Notifications should include the known information regarding the incident and a general description of the activities or process in place to fully define and remediate the situation. Within 3 business days of the NIH notification, the Institution, through the Approved User and the Institutional Signing Official, agree to submit to the NHGRI DAC a more detailed written report including the date and nature of the event, actions taken or to be taken to remediate the issue(s), and plans or processes developed to prevent further problems, including specific information on timelines anticipated for action.

All notifications and written reports of data security incidents should be sent to:
The NHGRI Data Access Committee's urgent e-mail address:
Urgent_NHGRIDAC@mail.nih.gov

NOTE: The urgent email address should be used ONLY for data security incidents or other Terms of Access. Please email NHGRIDAC@mail.nih.gov to communicate with the NHGRI Data Access Committee on all other matters.

The NHGRI, the NIH, or another entity designated by the NIH may also investigate any data security incident. Approved Users and their associates agree to support such investigations and provide information, within the limits of applicable local, state and federal laws and regulations. In addition, Approved Users and his/her institution agree to work with the NHGRI and the NIH to assure that plans and procedures developed to address identified problems are mutually acceptable.

7. Intellectual Property

By requesting access to the NHS-HPFS Diabetes GWAS dataset(s), the Requester and his or her Institution acknowledge the intent of the NIH to see that Approved Users, and anyone else authorized for research access through the Data Access Request, follow the intellectual property principles within the NIH GWAS Policy for Data Sharing as summarized below:

Achieving maximum public benefit is the ultimate goal of data distribution through the NIH GWAS Data Repository. The NIH believes that these data should be considered as pre-competitive, and urges Approved Users to avoid making IP claims derived directly from the GWAS dataset(s). However, the NIH also recognizes the importance of the subsequent development of IP on downstream discoveries, especially in therapeutics, which will be necessary to support full investment in products to benefit the public.

It is expected that these NIH-provided data, and conclusions derived therefrom, will remain freely available, without requirement for licensing. The NIH encourages broad use of GWAS datasets coupled with a responsible approach to management of intellectual property derived from downstream discoveries in a manner consistent with the NIH's Best Practices for the Licensing of Genomic Inventions and the NIH Research Tools Policy.

8. Research Dissemination and Acknowledgement of NIH GWAS Datasets

It is the intent of the NIH to promote the dissemination of research findings from NIH GWAS dataset(s) as widely as possible through scientific publication or other appropriate public dissemination mechanisms. Approved Users are strongly encouraged to publish their results in peer-reviewed journals and to present research findings at scientific meetings, etc. In accord with the NIH GWAS Policy for Data Sharing, and as expressed through the Requester's submission of the DAR, Approved Users agree not to submit findings using the NHS-HPFS Diabetes GWAS dataset(s), or updated versions thereof, for publication or presentation for a period of exclusivity for Contributing Investigators concluding with the Embargo Date identified on the NIH GWAS Data Repository.

The Requester agrees to acknowledge the NIH GWAS Data Repository, the Contributing Investigator(s) who contributed the phenotype data and DNA samples from his/her original study, and the primary funding organization that supported the contributing study in all oral and written presentations, disclosures, and publications resulting from any analyses of the data. The Requester further agrees that the acknowledgment shall include the dbGaP accession number to the specific version of the dataset(s) analyzed.

A sample statement for the acknowledgment of the NHS-HPFS Diabetes GWAS dataset(s) follows:

Funding support for the GWAS of Gene and Environment Initiatives in Type 2 Diabetes was provided through the NIH Genes, Environment and Health Initiative [GEI] (U01HG004399). The human subjects participating in the GWAS derive from The Nurses' Health Study and Health Professionals' Follow-up Study and these studies are supported by National Institutes of Health grants CA87969, CA55075, and DK58845. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the Gene Environment Association Studies, GENEVA Coordinating Center (U01 HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Funding support for genotyping, which was performed at the Broad Institute of MIT and Harvard, was provided by the NIH GEI (U01HG004424). The datasets used for the analyses described in this manuscript were obtained from dbGaP at [http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gap] through dbGaP accession number [phs000091].

9. Research Use Reporting

To ensure that NIH policies and procedures for participant protection and other elements of GWAS data use are adhered to, Approved Users agree to provide annual feedback on how these data have been used and any results that have been generated as a result of access

to the data, including patents and publications. This information will be used by the NHGRI DAC staff as required for program oversight activities, and by the NIH GWAS Governance committees as part of the NIH effort to provide ongoing oversight and management of all NIH GWAS activities.

Approved Users thus agree to provide a brief Annual Data Use Report on the research specified within the DAR attached to this DUC. Requesters who are previously Approved Users (that is, those seeking renewal) will be asked to provide specific information in a renewal DAR. Those not seeking renewal will be asked to provide specific information to the Data Access Committee via the contact information below. Annual Data Use Reports will provide information regarding potentially significant findings and publications or presentations that resulted from the use of the NHS-HPFS Diabetes GWAS dataset(s), a summary of any plans for future research use, any violations of the terms of access described within this Data Use Certification and the implemented remediation, and information on any downstream intellectual property generated as a result of the data. Approved Users also may be asked to provide general comments regarding topics such as the effectiveness of the NIH GWAS data access process, such as ease of access and use, appropriateness of data format, challenges in following the policies, and suggestions for improving data access or the program in general.

Annual Data Use Reports will be submitted no later than the first day of the month following the anniversary of the Approved Access Date assigned by the DAC and specified within the complete DAR provided to Requester by the NIH GWAS Data Repository at the time that the data access request is submitted. The final Data Use Report within a given approval period will be submitted as within the context of a renewal Data Access Request, or via a letter signed by the Institutional Signing Official and the Approved User.

Annual Data Use Reports should be submitted to:

The NHGRI Data Access Committee by e-mail at NHGRIDAC@mail.nih.gov, unless otherwise indicated in automated reminder messages from NCBI/dbGaP. Requests for continued data access should be made through dbGaP.

Note that any inadvertent or inappropriate data release incidents should be reported to the NHGRI DAC according to the agreements and instructions under Term 6.

10. Non-Endorsement, Indemnification

The Requester and his/her institution acknowledge that although all reasonable efforts have been taken to ensure the accuracy and reliability of NIH GWAS data, the NIH, the NHGRI, and Contributing Investigators do not and cannot warrant the results that may be obtained by using any data included therein. The NIH, the NHGRI, and all contributors to these datasets disclaim all warranties as to performance or fitness of the data for any particular purpose.

No indemnification for any loss, claim, damage or liability is intended or provided by any party under this agreement. Each party shall be liable for any loss, claim, damage, or liability that said party incurs as a result of its activities under this agreement, except that the NIH, as an agency of the United States, assumes liability only to the extent provided under the Federal Tort Claims Act, 28 U.S.C. 2671 et seq.

11. Termination and Violations

This Data Use Certification will be in effect for a period of one year from the date the data

were made accessible to the Approved User ("Approved Access Date"). At the end of the one-year period, the Approved User agrees to destroy all copies of the NIH GWAS dataset(s), and all derivatives that contain individual-level information.

Consideration will be given to a renewal of this agreement upon submission of a new DAR. Copies of NIH GWAS dataset(s) may not need to be destroyed if, with advance notice and approval by the NHGRI Data Access Committee, documentation is provided that other Approved Users at the same institution are using the dataset(s) under an active DAC approved research project.

The NIH or the NHGRI may terminate this agreement and immediately revoke access to all NIH GWAS datasets at any time if the Requester is found to be no longer in agreement with the policies, principles and procedures of the NIH and the NHGRI.

****

By submission of the attached Data Access Request, the Requester and his/her Institutional Signing Official attest to the Requester's qualifications for access to and use of NIH GWAS dataset(s) and certify their agreement to the NIH principles, policies and procedures for the use of the NHS-HPFS Diabetes GWAS datasets as articulated in this document.

Requesters further acknowledge that they have shared this document and the NIH GWAS policies and procedures for access and use of GWAS datasets with any research staff, and Key Personnel identified in the DAR, who will participate in the proposed research project defined in the Data Access Request.

Institutional Signing Officials acknowledge that they have considered the relevant NIH GWAS policies and procedures, that they have shared this document and the relevant policies and procedures with appropriate institutional organizations, and have assured compliance with local institutional policies related to technology transfer, information technology, privacy, and human subjects research.

Addendum to the Data Use Certification Agreement
Modification of Data Security Terms and Best Practices

Effective for all dbGaP Data Access Requests submitted on or after March 23, 2015,
Section 6 of the Data Use Certification Agreement is replaced in its entirety by the following:

6. Data Security and Data Release Reporting

The Requester and Approved Users, including the institutional IT Director, acknowledge NIH's expectation that they have reviewed and agree to manage the requested dataset(s) according to the current NIH Security Best Practices for Controlled-Access Data Subject to the GDS Policy and the institutional IT security requirements and policies, and that the institution's IT security requirements and policies are sufficient to protect the confidentiality and integrity of the NIH controlled-access data entrusted to the Requester.

If approved by NIH to use cloud computing for the proposed research project, as outlined in the Research and Cloud Computing Use Statements of the Data Access Request, the Requester acknowledges that the IT Director has reviewed and understands the cloud computing guidelines in the NIH Security Best Practices for Controlled-Access Data Subject to the GDS Policy.

Requesters and PIs agree to notify the NHGRI DAC of any unauthorized data sharing, breaches of data security, or inadvertent data releases that may compromise data confidentiality within 24 hours of when the incident is identified. As permitted by law, notifications should include any known information regarding the incident and a general description of the activities or process in place to define and remediate the situation fully. Within 3 business days of the NHGRI DAC notification, the Requester, through the PI and the Institutional Signing Official, agree to submit to the NHGRI Data Access Committee a detailed written report including the date and nature of the event, actions taken or to be taken to remediate the issue(s), and plans or processes developed to prevent further problems, including specific information on timelines anticipated for action.

All notifications and written reports of data security incidents should be sent to:
NHGRI Data Access Committee URGENT: nhgridac@mail.nih.gov
GDS mailbox: gds@mail.nih.gov

NIH, or another entity designated by NIH may, as permitted by law, also investigate any data security incident. Approved Users and their associates agree to support such investigations and provide information, within the limits of applicable local, state, and federal laws and regulations. In addition, Requesters and Approved Users agree to work with the NHGRI and NIH to assure that plans and procedures that are developed to address identified problems are mutually acceptable and consistent with applicable law.

Addendum to the Data Use Certification

Effective for all dbGaP study datasets registered prior to July 31, 2013

- Annual Data Use Reports will no longer be submitted by email to the relevant Data Access Committee(s) (DACs), as stated in the Data Use Certification(s) associated with the datasets you are requesting.

- Principal Investigators are now expected to submit online research progress updates through the dbGaP system, as part of the annual renewal process or close-out process.