

**NIDDK Central Database Repository**  
**NCBI dbGaP**  
**Data Use Certification**

**Version date 22 August 2008**  
**Contact: [niddk-dac@mail.nih.gov](mailto:niddk-dac@mail.nih.gov)**

Requestor: \_\_\_\_\_

Requesting Institution: \_\_\_\_\_

Requested Dataset(s) (“Dataset(s)”):  
\_\_\_\_\_

If Requestor is funded by NIDDK for this Research Project, the NIDDK grant number is:

\_\_\_\_\_.

**Introduction**

The National Institute of Diabetes and Digestive and Kidney Disease (NIDDK) has supported collection of Data and Samples from participants in numerous studies. In order to maximize the benefits of these resources collected with public funds and maximize their research value, it is important that these resources be made available, on appropriate terms and conditions, to the largest possible number of qualified investigators in a timely manner.

The NIDDK Central Database Repository and the NCBI dbGaP do not receive personally identifiable information. The Dataset(s) (genotype-phenotype datasets along with their pre-computed analysis) to be distributed by the NCBI DbGaP under this Data Use Certification (“Certification”) will be coded, meaning that the data are unidentified for research purposes by use of a random or arbitrary alphanumeric code. However, due to the phenotype and genotype data that may be included in the Dataset(s), the Dataset(s) from the NIDDK Central Database Repository will only be provided to investigators (Requestors) who, along with the Requestor’s Institutional Business Official (Requesting Institution) have agreed to the requirements and terms of access as detailed in this Certification.

Submission of the Data Use Certification indicates that the signers represent and warrant their qualifications for access to and use of the Dataset(s) from the NIDDK Central Database Repository and certify their agreement to the terms and conditions articulated herein and to the NIDDK principles, policies, and procedures for the use of the NIDDK Central Database Repository resources as articulated in this Certification. In the event that investigators from more than one institution will be collaborating on a project using the Dataset(s) transferred under this Certification, an investigator from each institution is required to complete a separate Data Use Certification request.

It is the intent of the NIDDK that Approved Users of NIDDK Central Repository Dataset(s)\_ recognize the limitations imposed by the original informed consent agreements of contributing studies as described in the NCBI dbGaP database and documents that accompany each request.

It is anticipated that, at least in some cases, the NCBI dbGaP will be updated with additional information and will be so identified by an appropriate date. Unless otherwise indicated, all statements herein are presumed to be true and applicable to the access and use of all versions of the NCBI dbGaP Dataset(s).

## Terms of Access

### *1. Definitions:*

**Approved Users:** Research investigators who submitted a request for access to NIDDK Central Database Repository Dataset(s) that have been approved by the NIDDK Data Access Committee and who have a fully-executed Data Use Certification for the requested Dataset(s).

**Contributing Study Investigators:** Research investigators who provided the phenotypic data and samples to the NIDDK Central Database Repository.

**Institutional Business Official:** A member of the Requestor's home organization with the authority to commit the organization to the terms of the request and certify the organization's agreement to the terms delineated in the Data Use Certification.

**NIDDK Central Database Repository Dataset:** contains the relevant data from a specific project. Access to this resource may be managed for NIDDK by the National Center for Information via dbGaP.

**Research Project:** includes the project title, the Requestor's name and Requesting Institution's name, the names of any independent collaborators and their institutions, a one to two paragraph Research Use Statement, a description of the research objectives and design, and an analysis plan.

**Research Use Statement:** Statement of proposed research to be conducted which may be made publicly available. The Research Use Statement is submitted by the Requestor as a part of Requestor's Research Project request for access to Data set(s).

**Study Subject:** means an individual who participated in the clinical research, either as a recipient of a test article or as a control. A Study Subject may be either a healthy human or a patient.

### *2. Research Project: Use of Dataset(s)*

- a) The NIDDK Central Database Repository Dataset(s) shall be used only for research purposes by the Requestor in his/her laboratory under suitable conditions as outlined in Article 5 for the research described with specificity in the attached Research Project. The Research Project will include the project title, the Requestor's name and Requesting Institution's name, the names of any independent collaborators and their institutions, a one to two paragraph Research Use Statement, a description of the research objectives and design, and an analysis plan. The Dataset(s) shall not be used in any research that is not disclosed and approved as part of the Research Project. The Requestor and Requesting Institution agree to retain control over the Dataset(s) and further agree not to transfer the Dataset(s) to other people not under the direct supervision of the Requestor.
- b) New uses of these Dataset(s) outside those described in the Research Project require submission of a new Data Use Certification request. Modification to an approved Research Project requires submission to the NIDDK Data Access Committee of an amendment to the Research Project. Appointment of another or different Principal Investigator to complete an approved Research Project is a modification and requires submission of an amendment to this Data Access Certification request.

- c) The Requestor and Requesting Institution acknowledge that other researchers have access to NIDDK Central Database Repository resources and that duplication of research is a distinct possibility.
- d) The Requestor and Requesting Institution further acknowledge that they are responsible for ensuring that all their uses of the Dataset(s) are consistent with Federal (including 45 CFR Part 46), state, and local laws and any relevant institutional policies.
- e) The Requestor and Requesting Institution shall use the Dataset(s) only for purposes related to biomedical research, cost-effectiveness, or other economic research. Purposes for which the Dataset(s) may not be used include but are not limited to:
  - identification and targeting of under- or over-served health service markets primarily for commercial benefit;
  - obtaining information about health care providers or facilities for commercial benefit;
  - insurance purposes such as redlining areas deemed to offer bad health insurance risks; and
  - adverse selection (e.g., identifying patients with high risk diagnoses).

It is anticipated that, at least in some cases, the Dataset(s) at the NIDDK Central Database Repository will be updated with additional information and will be so identified by an appropriate version number. Unless otherwise indicated, all statements herein are presumed to be true and applicable to the access and use of all versions of the Dataset(s) at and from the NIDDK Central Database Repository.

### ***3. Human Subject Protections: Compliance with IRB Requirements***

- a) Requestor and Requesting Institution acknowledge that the conditions for use of these Dataset(s) require the review and subsequent approval or waiver of approval by the Requestor's and Requesting Institution's Institutional Review Board (IRB) or other approval body operating under an Office of Human Research Protections (OHRP) - approved Assurance and in accordance with Department of Health and Human Services regulations at 45 CFR Part 46. Requestor agrees to comply fully with all such conditions.
- b) In order to respect the privacy of the Study Subjects, the Requestor and Requesting Institution agree not to contact or make any effort to identify individuals, families, communities, tribes or populations which are or may be the source of the Dataset(s). (This condition is not applicable to Contributing Study Investigators who provided the data used to generate the Dataset(s) at and from the NIDDK Central Database Repository, if they have appropriate IRB approval to retain the Study Subject identities or re-contact Study Subjects.)
- c) The Requestor shall not combine or link the Dataset(s) provided with any other collection or source of information that may contain information specific to individuals.
- d) Requestor and Requesting Institution agree to report promptly to the NIDDK, at the following e-mail address: [niddk-dac@mail.nih.gov](mailto:niddk-dac@mail.nih.gov), any proposed modifications in the Research Project and any unanticipated problems involving risk to Study Subjects or others. Requestor and Requesting Institution agree to this provision in addition to any of Requestor's and Requesting Institution's institutional policies or any local, State, and/or Federal laws and regulations which provide additional protections for human subjects. Such agreement does not supersede the applicable laws, regulations and policies.

#### **4. *Public Posting of Approved User's Research Use***

Requestor and Requesting Institution agree that if the attached Research Project is approved, information about the proposed research use may be posted on a public web site that describes the Dataset(s) from the NIDDK Central Database Repository. The information may include the Requestor's and Requesting Institution's names, project title, and Research Use Statement.

#### **5. *Data Security/Non-transferability***

The Requestor and Requesting Institution agree to store the Dataset(s) on a computer with adequate security controls, and to maintain appropriate control over the Dataset(s). Best practices for computer security and data control are available online at

<http://www.ncbi.nlm.nih.gov/projects/gap/pdf/dbGaPLevel2SecurityProcedures.pdf>. The Requestor and Requesting Institution agree to establish appropriate administrative, technical, procedural, and physical safeguards to protect the confidentiality of the Dataset(s) and to prevent unauthorized access to it. The Requestor and Requesting Institution agree to ensure that the Dataset(s) is protected by reasonable safeguards against loss, unauthorized access, use, modification or disclosure, and against any misuse and agree to notify the NIDDK Central Database Repository at the following e-mail address: [niddk-dac@mail.nih.gov](mailto:niddk-dac@mail.nih.gov) as soon as a security breach is discovered.

- a) These Dataset(s) represent a significant investment on the part of NIDDK. The Requestor and Requesting Institution therefore agree to retain control over the Dataset(s), and further agree not to transfer or distribute the Dataset(s) in any form to any entity or individual not under the Requestor's direct supervision. The Dataset(s) may be shared with independent collaborating investigators listed in the attached Research Project who are also Approved Users who have a fully-executed Data Use Certification for the purposes of this Research Project. The Requestor and Requesting Institution acknowledge responsibility for ensuring appropriate use of these Dataset(s) in accordance with the terms of this Certification.
- b) The Requestor agrees to retain control over the data as an Approved User, and further agrees not to distribute individual-level data in any form. No copies or derivatives shall be made of the Dataset(s) except as necessary for the purposes authorized in this Data Use Certification. Requestor and Requesting Institution acknowledge that if any copies of the Dataset(s) are generated, the terms and conditions of this Certification apply to such copies. The Requestor shall keep an accurate written account of all such copies and derivative files, which will be furnished to the NIDDK upon request. At the completion of the activities in the Research Project, all files containing Dataset(s) or any portion thereof will be destroyed or returned to the NIDDK at the Requestor's and Requesting Institution's expense, and any derivative files and copies shall be destroyed. Subject to Article 7, the Dataset(s) transferred under this Data Use Certification may be kept for no longer than five years from the date of receipt.
- c) The Requestor agrees that if he/she changes institutions, a new Data Use Certification in which the new Requesting Institution acknowledges and agrees to NIDDK principles, policies, procedures and the terms of access will be required in order for the Requestor to continue the Research Project at the new institution.

## 6. *Intellectual Property*

By requesting access to Dataset(s) from the NIDDK Central Database Repository, the Requestor and Requesting Institution acknowledge the intent of the NIDDK to see that Approved Users follow the NIH Intellectual Property (IP) policy as outlined below:

- Achieving maximum public benefit is the ultimate goal of data distribution through NIDDK Central Database Repository mechanisms. The NIDDK believes that NIDDK Central Database Repository data should be considered as pre-competitive, and urges Requestor's, Requesting Institutions and others accessing the Dataset(s) to avoid making intellectual property (IP) claims on the Dataset(s) and their analyses. However, the NIDDK also recognizes the importance of the later development of IP on downstream discoveries, especially in therapeutics, which will be necessary to support full investment in products that the public needs.
- In this spirit, it is expected that Dataset(s) from the NIDDK Central Database Repository and conclusions derived therefrom will remain freely available, without requirement for licensing, for applications such as, but not necessarily limited to, the following: the use of markers in developing assays or diagnostic tools; the use of combinations of markers in multiplex assays; and the use of markers as guides toward identification of new drug targets.
- The NIDDK encourages licensing practices consistent with the recommendations cited in NIH's [Best Practices for the Licensing of Genomic Inventions](#) and in the [NIH Research Tools Policy](#).

## 7. *Data Access Renewal Period*

Initial access is granted for one-year from the date of the Data Use Certification approval. Renewal of access may be granted if requested of the NIDDK Central Repositories by submission of a request to the following e-mail address: [niddk-dac@mail.nih.gov](mailto:niddk-dac@mail.nih.gov).

## 8. *Research Reporting and Dissemination of Research Results - Acknowledgment of NIDDK*

Prompt publication or any public disclosure of the results of the Research Project is encouraged. Requestors are strongly encouraged to publish their results in peer-reviewed journals.

- a) The Requestor and Requesting Institution agree to submit, one year from the date of the Data Use Certification approval or renewal of access being granted, a report to the NIDDK on the Research Project that includes all analyses conducted with the Dataset(s) and related manuscripts submitted for publication. The analyses conducted with the Dataset(s) may then be included in the NIDDK Central Repository Database at the discretion of NIDDK.
- b) Before the Requestor or the Requesting Institution submits a paper or abstract for publication or otherwise intends to publicly disclose information about the Dataset(s), Requestor and the Requesting Institution shall ensure that NIDDK has at least thirty (30) days to review the proposed publication or disclosure and agree to provide NIDDK a copy of any manuscript or other disclosure document thirty (30) days in advance of submission for publication, in order to ensure compliance with the confidentiality requirements set forth in this Data Use Certification. NIDDK reserves the right to delete

or modify information that might reasonably be viewed as offensive to the Study Subjects involved.

- c) The Requestor and Requesting Institution agree not to publish or otherwise disclose the Dataset(s) to any person or organization unless the Dataset(s) have been aggregated (that is, combined into groupings of data such that the data are no longer specific to any individuals within each grouping), and no cells (aggregates of data) contain information on fewer than ten individuals or fewer than five providers or facilities. The Requestor and Requesting Institution shall not publish or otherwise disclose Dataset(s) that identify individual providers or facilities, or from which such identities could be inferred.
- d) The Requestor and Requesting Institution agree to acknowledge the contribution of the Contributing Study Investigators and the NIDDK Central Database Repository in any and all oral and written presentations, disclosures, and publications resulting from any and all analyses of the Dataset(s). The Requestor and Requesting Institution further agree that the acknowledgment shall include a reference to the specific version of the NIDDK Central Database Repository Dataset(s) downloaded from NCBI dbGaP.

A sample statement to be used in acknowledgments is:

"The *[insert name of contributing study]* was conducted by the *[insert name of contributing study]* Investigators and supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). This manuscript was not prepared in collaboration with Investigators of the *[insert name of contributing study]* study and does not necessarily reflect the opinions or views of the *[insert name of contributing study]* study or the NIDDK."

#### **9. *Non-Endorsement, Indemnification***

The Requestor and Requesting Institution acknowledge that although all reasonable efforts have been taken to ensure the accuracy and reliability of the Dataset(s) in the NIDDK Central Database Repository, the Dataset(s) are provided as a service to the research community. The Dataset(s) are supplied to Requestor and Requesting Institution with NO WARRANTIES, EXPRESS OR IMPLIED, INCLUDING ANY WARRANTY OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. NIDDK AND THE NIDDK CENTRAL DATABASE REPOSITORY make no representation that the use of the Dataset(s) will not infringe any patent or proprietary rights of third parties.

The Requestor and Requesting Institution agree not to claim, infer, or imply endorsement by the US Government of the Research Project, the entity, or personnel conducting the Research Project or any resulting commercial product(s).

No indemnification for any loss, claim, damage or liability is intended or provided by any party to this agreement. Each party shall be liable for any loss, claim, damage, or liability that the party incurs as a result of its activities under this agreement, except that the NIDDK, 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.

#### **10. *Disqualification, Enforcement***

Failure to comply with any of the terms specified herein may result in disqualification of Requestor or Requesting Institution from receiving additional Dataset(s). from the NIDDK

Central Database Repository. The United States Government shall have the right to institute and prosecute any proceeding at law or in equity against the Requestor and Requesting Institution for violating or threatening to violate the confidentiality requirements of this Certification, the limitations on the use of the Dataset(s) provided, or both. Proceedings may be initiated against the violating party, legal representatives, and assigns, for a restraining injunction, compensatory and punitive damages, mandamus, and/or any other proceeding in law or equity, including obtaining the proceeds from any intellectual property or other rights that are derived in whole or in part from the breach of the confidentiality requirements or use limitations of this Certification. In addition, Requestor and Requesting Institution acknowledge and agree that a breach or threatened breach of the confidentiality requirements or use limitations of this Certification may subject Requestor and/or Requesting Institution to legal action on the part of Study Subjects, their families, or both.

## ***11. Acknowledgement***

By signing and dating this Data Use Certification, the Requestor and Requesting Institution represent and warrant the Requestor's qualifications for access to and use of Dataset(s) from the NIDDK Central Database Repository and certify their agreement to the NIDDK principles, policies and procedures for the use of the Dataset(s) from the NIDDK Central Database Repository, as articulated in this document. Before any person is permitted access to the Dataset(s) or participation in the Research Project in accordance with this Certification, Requestor and Requesting Institution agree to share this document and the NIDDK policies and procedures with such person and obtain his/her agreement to abide by its terms. Institutional Business Officials also acknowledge that they have shared this document and the relevant NIDDK policies and procedures with appropriate institutional organizations, such as the Office of Technology Transfer and the Office for 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 NIDDK 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 NIDDK DAC notification, the Requester, through the PI and the Institutional Signing Official, agree to submit to the NIDDK 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:**

NIDDK Data Access Committee URGENT: [niddk-dac@mail.nih.gov](mailto:niddk-dac@mail.nih.gov)

GDS mailbox: [gds@mail.nih.gov](mailto: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 NIDDK 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.