Supplementary Instructions For Principal Investigators Requesting Authorized Access to dbGaP Datasets

Principal investigators (PIs) seeking access to dbGaP datasets are urged to review the following information before initiating a data access request:

1. **Minimum Qualifications to Submit a dbGaP Project Request as a PI**. Investigators must be permanent employees of their institution at a level equivalent to a tenure-track professor or senior scientist with responsibilities that most likely include laboratory administration and oversight. Laboratory staff and trainees such as graduate students, and postdoctoral fellows are not permitted to submit project requests.

2. **Supplemental Documentation.** Some datasets require local Institutional Review Board (IRB) approval for use, as noted on the dbGaP study page. Other types of documentation may be required as described on the study page and/or Data Use Certification for the study. Evidence of IRB approval and other documentation can be uploaded as a PDF during the application process.

3. **Accessing Additional Datasets After Initial Approval.** Investigators who would like to access additional dataset(s) for use in an existing approved project should (1) revise the existing approved project request to include the new datasets and (2) update the Research Use Statement as appropriate. Investigators do not need to submit a new project request unless the dataset will be used for research outside of the scope of the approved Research Use Statement.

4. **Collection of Basic Institutional and Contact Information.** Contact information for PIs and their designated institutional signing official will be auto-filled in the project request based on information provided in eRA Commons; therefore, PIs should verify the accuracy of all eRA account information prior to initiating a project request in dbGaP.

5. **Small or Recently Established Institutions.** Investigators from a small or newly established organization may be asked for additional information and documentation, such as a curriculum vitae, basic institutional information, and an information technology (IT) security plan. These documents are used by NIH to gauge the requestor's ability to comply with the expectations outlined in the NIH Security Best Practices for Controlled-Access Data Subject to the Genomic Data Sharing (GDS) Policy and GDS Policy.

6. **Research Use Statement.** The approval of project requests depends primarily on a carefully written Research Use Statement (2,200 characters max). The statement should include the following components:
   - Objectives of the proposed research;
   - Study design;
   - Analysis plan, including the phenotypic characteristics that will be evaluated in association with genetic variants;
   - Explanation of how the proposed research is consistent with the data use limitations for the requested dataset(s); and
   - Brief description of any planned collaboration with researchers at other institutions, including the name of the collaborator(s) and their institution(s).

   The relevant Data Access Committee(s) (DAC(s)) will review the Research Use Statement to confirm that the proposed research is consistent with all applicable data use limitations for the requested dataset(s). An inconsistency between the proposed research and the applicable data use limitations, or insufficient detail to make this determination, is the most common reason for data requests to be rejected.

> **Important**!
> - Research should not deviate from the description in the approved Research Use Statement.
> - To expand the scope of a project after approval, investigators must submit a revised Research Use Statement for approval.
> - Conducting research not described in the Research Use Statement is a violation of the terms in the Data Use Certification agreement, which may lead to termination of data access and other penalties.

March 23, 2015

Supplementary Instructions For Principal Investigators Requesting
Authorized Access to dbGaP Datasets

Tips for drafting a successful Research Use Statement:

- For projects that aim to evaluate phenotypic characteristics that were not the primary focus of the original study (e.g., the project proposes creating a new outcome variable from existing variables in the original study), explain how the use of the phenotypic characteristics are consistent with the data use limitations for the requested dataset(s).

- For projects that intend to use the dataset(s) to develop or evaluate performance of new analytical methods, include a sentence indicating how this work is relevant to the specific disease dataset(s) being studied, such as "This methodological work will advance the understanding of the genetic bases of strabismus, cranial nerve abnormalities, and associated disorders."

- Special care should be taken when requesting datasets from more than one study in dbGaP.
  - Clarify whether datasets will be combined (with dbGaP or other sources) or analyzed separately for your analyses.
  - Describe the outcome variables proposed for the project and explain how their use is consistent with any data use limitations for <u>each</u> of the requested datasets.

7. **Cloud Use Statement and Cloud Servicer Provider Information (if applicable).** Investigators who wish to use cloud computing for storage and analysis of dbGaP data will need to indicate in their Data Access Request (DAR) that they are requesting permission to use cloud computing and identify the cloud service provider or providers that will be employed. They also will need to describe how the cloud computing service will be used to carry out their proposed research.

8. **Non-Technical Summary.** As part of the project request, investigators will provide a non-technical summary of their proposed research. If the project is approved, this statement will be publicly available for lay audiences to read the purpose and objectives of the research.

Non-technical summaries must be no more than 1,100 characters (including spaces).

9. **Staff and Collaborator Contact Information.** Provide the full legal names and contact information for internal collaborators (i.e., those employed at the PI's institution, but not directly supervised by the PI). Trainees and staff directly supervised by the PI, such as graduate students, postdoctoral fellows, and technicians, do not need to be listed on the project request. External collaborators should be listed in the external collaborator(s) section of the project request applications. Data exchange between all collaborators must be consistent with the NIH Security Best Practices for Controlled-Access Data Subject to the Genomic Data Sharing (GDS) Policy and GDS Policy.

10. **Information Technology (IT) Director Contact Information.** Provide the full legal name and contact information for the IT Director, who is expected to be a senior IT official with the necessary expertise and authority to affirm the IT capacities at an academic institution, company, or other research entity. The IT Director is expected to have the authority and capacity to ensure that the NIH Security Best Practices for Controlled-Access Data Subject to the NIH GDS Policy and the institution's IT security requirements and policies are followed by the Approved Users.