



Applying for Controlled Access Data

Overview

How does the dbGaP data access request (DAR) process work? Could you give an overview?

The following is a general overview of how the dbGaP data access request (DAR) process works:

1. Applicants log in to the dbGaP [Authorized Access System](#) as a Principal Investigator (PI) using the username and password of the applicant's eRA Commons account. The applicant creates a project and follows multiple steps to complete and submit the online application. Note: requests for several consent groups from different dbGaP studies can be included in the same project. The application then is forwarded to an Institutional Signing Official (SO) for approval.
2. The SO approves the requests and they move into the queue of a NIH Data Access Committee (DAC) to review. The review is to confirm that the proposed research use(s) is consistent with the data use limitations (DULs) of the requested dataset(s). These restrictions are established by the language in each study's informed consent materials, and therefore vary across studies in dbGaP.
3. Once the DAC review is complete, the PI and SO is notified by email of approval or disapproval of the data access request (DAR).
4. Once the approval email notification is received, the PI can download the data by logging in to the same dbGaP account used for making the DAR.

A more detailed explanation of the DAR process can be found on the "How does one apply" section in the [dbGaP Authorized Access system](#). It can also be found on this archive here.

Additional information about NIH policies and procedures for Genomic Data Sharing (GDS) Policy can be found on the [NIH Office of Science Policy \(OSP\) website](#) or you can send an email directly to gds@mail.nih.gov.

(07/2272017)

Information of Data Available for Download

Where can I find the information about the data files available for download prior to making a data access request (DAR)?

Phenotype variables and datasets can be browsed through the public dbGaP web pages. [Here](#) is how to find the study report and data manifest that include the information of the phenotype and genotype data files.

Please note that the information of sequencing data (including Next-Gen sequence raw data, BAM files, or other high throughput data.) is not included in the study report and data manifest available from the public FTP site. That information, however, is available inside of the dbGaP [Authorized Access System](#). Please see [here](#) for more information.

(08/30/2012)

SRA Data Availability

I've been approved to access all the data of the study, but I find that only the phenotype data are available for download. I know that the study is supposed to have SRA (Short Read Archive) data. Could you explain why the SRA data are not available?

The dbGaP phenotype and genomic data are submitted and processed separately from the SRA data. The SRA data may include the Next-Gen sequence raw data, BAM files, or other high throughput data. It is quite common that the data submitter submits phenotype and genomic data of a study while publishing the paper, but holds on to the submission of SRA data due to either technical or other reasons.

The dbGaP often goes ahead and releases the study as soon as the phenotype and genomic data are ready independent of the SRA data. Once the study is released, the study page is published on the [dbGaP website](#) and the individual level phenotype and genomic data are made available through the dbGaP [Authorized Access System](#). The SRA data release may fall behind either because the Principal Investigator (PI) may not have submitted the SRA data or the National Center for Biotechnology Information (NCBI) has not completed the data processing.. In cases where the SRA data lags as a result of the PI, the delay of SRA data release could be significant and sometime indefinite.

The SRA data submission and processing are handled by NCBI's SRA group. If the dbGaP study has been released and the SRA data have not yet been made, it is suggested to directly contact the sra@ncbi.nlm.nih for any questions regarding the status of SRA data release. They are in a better position to give you an update.

As the last resort, you may want to directly contact the PI of the study for more information. For most of studies, the PI's contact information is posted on the "Study Attribution" section at the bottom of the study page on the dbGaP website.

(08/29/2017)

A Brief Step-by-Step for Filling Out the Data Access Request

Would you give me a step-by-step instruction of how to apply for dbGaP Controlled-Access data?

There are several videos available on YouTube that demonstrate the dbGaP data access request (DAR) process. The links to them are listed [here](#).

The following is a brief step-by step guide to the dbGaP Authorized Access application process:

Visit the dbGaP [Authorized Access System](#). When you get there, you will find that you will need an NIH eRA Commons account username and password to log into the system ().

To get an eRA account and password, go to [eRA Commons](#), and follow the directions for online account registration. Please also see [here](#) for more about an eRA Commons account; For NIH researchers, in order to obtain access to genomic data in dbGaP that is available through controlled-access, eligible NIH Institute and Center (IC) intramural scientists and IC extramural program scientific staff must first obtain the approval of their IC. Please see [here](#) on additional information.

When you login for the first time, you will be asked to provide and save your contact information. This contact information will be used by the system to complete the cover page forms of your DAR. Once this is complete, you will not have to fill it out again in order to access the system in the future. Should your contact information change in the future, you can click on the "My Profile" tab located near the top of the page and update the information accordingly.

Click on the "My Projects" tab. Once you are on the "My Projects" page, click the "Create a new project" button to start a data access request. **Please be aware that the name of a new project cannot be the same as that of any existing projects, including closed projects.** After entering the information, move on to the next step using the "Begin New Research Project" button.

Choosing Datasets

All datasets available for download are listed under "Choose Datasets" sub-tab.

(Note: For a newly created project, this is the first sub-tab. After the first time, you will see this sub-tab before the "Confirm Dataset" sub-tab)

With hundreds of datasets listed on this page, it is difficult to find anything specific by browsing through the page. It is suggested to search the page using the web browser's search function (Control-F) to find related datasets. For example you will find Framingham datasets by searching for "phs000007" (no quotes) or by "Framingham".

Please note that study participants are partitioned according to their informed consent restrictions on use. Select only those participant sets that have consent restrictions consistent with your proposed research.

Please also be aware that if there is more than one consent group in a study, the participants of different consent groups do not overlap, which means that you would have to get the data of all consent groups to cover all participants of the study.

Before including the datasets in the project you are creating, please read the "Data Use Limitations" of each dataset to make sure it matches the Research Use Statement that is required in a later step.

Select all the datasets you have decided to include using the checkbox and move to the next step by clicking on "Add Selected and Continue" button at the bottom of the page.

Research Project

On this page you will enter the following information

- a. A title for the research activity
- b. Type of research
- c. A research use statement (limited to 2000 characters)
- d. A non-technical summary (limited to 1100 characters)
- e. The name of your Institutional Signing Official (SO)
- f. The organization information of data requester or the Principal Investigator (PI).
- g. Create decryption password for the project. The password is used to create dbGaP repository key that is required for configuring the NCBI decryption tool and SRA-toolkit. The decryption password and dbGaP repository key are shared among all download packages created under the project.

Note: after being created for the first time, the decryption password can be reset from the page under the "Data Security" sub-tab. **Reset password will only affect the download packages created after the reset**, which means that the download package created before the reset will still use the old dbGaP repository key for decryption.

Move to the next step by clicking on "Save and Continue" button.

Collaborators

On this page, you will enter the names and contact information of your collaborator(s). The collaborators within your institution (if any) should be provided. The degree of detail for your list of collaborators is

decided by your DAC (Data Access Committee) reviewers, but generally speaking, a “collaborator” is meant to include staff with an official appointment at your institution, and not supervised students and technical staff. Use the “add another collaborator” button if you have more than one local collaborator. The data downloaded from the dbGaP can be shared between listed internal collaborators through a secured computer system.

If investigators plan to collaborate with investigators outside their own institution, the investigators at each external site must submit an independent DAR using the same project title and Research Use Statement, and if using the cloud, Cloud Computing Use Statement. **The “Research Use Statements” section of each application should also mention the respective external collaborators and the fact that this is a joint research project** so that the Data Access Committee (DAC) can review the requests together. These are important for an expedited review of separately submitted requests from all external collaborators.

Information Technology (IT) Directors

On this page you will enter the name(s) and contact information for your IT Director(s). Generally, 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](http://www.ncbi.nlm.nih.gov/projects/gap/pdf/dbgap_2b_security_procedures.pdf) and the institution’s IT security requirements and policies are followed by the Approved Users.

Choose Datasets

This is the same sub-tab mentioned early in this instruction. When the dataset is selected for the first time, this is the first sub-tab. After that it is placed to the middle before the “Confirm Datasets” sub-tab. Please see the early “**Choose Datasets**” section in this instruction for details.

Confirm Datasets

On this page, the datasets you selected in the “Choose Dataset” step are listed. This is your first chance to review the “Data Use Limitations” (DUL) of each selected datasets. Please make sure it matches the data uses described in the Research Use Statement.

For each listed dataset, if you want to remove it from the list, you can use the dropbox on the right side of the dataset name to select “Remove”, and then use the “Remove Selected” button to remove it from the list. Please note that a datasets is no longer removable if it has ever been approved.

Review the Data Use Certification

On this page, you are given a chance to review the Data Use Certificate (DUC) and the Dataset Manifest (look for PDF icons in the right most column) for each of the datasets you have requested.

- At this point you can use the “Back” button to go back to the previous step if you want to remove datasets, or go back to the “Choose Datasets” step if you want to add datasets.
- Review the DUCs and the Dataset Manifests to make sure everything is correct.
- If any of the datasets included requires IRB approval, go to the “Program-specific required attachments” section at the bottom of the page

(Please note that the “Upload” or “Delete” button may not be shown if no dataset is selected at the “Choose Dataset” Step).

The IRB requirement is imposed by the data submitter. For the datasets marked as IRB required, the IRB approval document has to be provided in order for the system to go through. There are no other alternatives.

Review the Data Use Limitation

On this page, you are given a final opportunity to review the Data Use Limitation (DUL) of each selected dataset and make sure it matches the data uses described in the Research Use Statement. If the DUL of any of

selected dataset doesn't match intended data use, you would need to go back to previous "Confirm Datasets" step to remove it. If you agree the DUL are consistent with your intended use as described in your Research Use Statement, make sure all the checkboxes are checked and move to the next step by "I agree the Data Use Limitation(s)" button.

Review Applications (& Electronic Signature)

On this page, you can find links to your completed application(s). You may want to review your application(s) as a whole before the final submission. If you need to revise any portion of your applications, use one of the sub-tabs to return to a previous step and make changes.

Once you have thoroughly reviewed the application, you will need to click the checkboxes labeled "I agree ...". **Checking these two checkboxes constitutes an electronic signature**, which affirms the accuracy of the application. After clicking the two check boxes, click the "Submit Application to Signing Official" button to route the request to your Institutional Signing Official (SO) for their signature.

After the application being submitted, under "My Requests" tab, you can see that selected data requests of the project are under "SO review" status. In the meanwhile, an automatically generated email notification will be sent to the designated SO for reviewing the application. Once being approved by the SO, each DAR under the project will be in queue for review by the respective Data Access Committee (DAC). The status of the request is changed to "DAC review".

After the application is submitted and before being pick up from the SO queue, you can use the "Recall Application" button to withdraw the submission. After the recall, you can revise and submit it again.

Please see here for how to check the status of submitted DARs.

You will be notified once DAC review is done. You will be allowed to download requested data once the DAR gets approved by the DAC.

(7/11/2017)

Check Status of Submitted Data Requests

How do I check the status and review my data access requests (DARs)?

You can track the progress of your submitted data access requests (DARs) through the dbGaP [Authorized Access System](#). After logging in, click on "My Requests" tab located near the top of the page, which takes you to the "My Requests" page.

The progress of your DAR is indicated in the "Status" column of the request table. The following are a few commonly seen statuses:

- a. **SO review:** After the dbGaP project is submitted, the DAR(s) under the project will be under "SO review" status. In the meanwhile, an automatically generated email will be sent to notify the Signing Official (SO) to review the data requests. You may want to directly contact the SO to expedite this process.
- b. **DAC review:** After being approved by the SO, the DAR is put in the queue for an NIH Data Access Committee (DAC) to review. The time required for DAC review varies depending on many factors, however, average DAC review is two weeks. Please directly contact the respective DAC for an update if necessary.
- c. **Approved:** Once the request is approved by the DAC, you will receive an email notification about the status of the request. The status of the request will be shown as "Approved". Once approved, you can immediately start to assemble download packages by going to the DAR page through "My Requests" tab and download the data through "Downloads" tab.

- d. **Approved Expired:** Through submission of the DAR, the Principal Investigator (PI) agrees to submit either a project renewal or close-out request prior to the expiration date of the 1-year data access period. Failure to submit a renewal or to complete the close-out process, including confirmation of data destruction by the Institutional Signing Official, may result in termination of all current data access and/or suspension of the PI and all associated personnel and collaborators from submitting new DARs for a period of time..
- e. **Rev. Requested:** If the DAR is under “Rev Request” status, please go to the data request page through “My Request” tab and read DAC comment. You may need to revise the project accordingly and resubmit it through your dbGaP account. Please see [here](#) for more about how to revise and resubmit an existing dbGaP project.
- f. **Rejected:** If your DAR is rejected by the DAC, it often means that stated data use does not match the data use limitation of selected datasets. Please go to the DAR page through “My Request” tab and read the DAC comment for more details. You may need to reselect datasets, change your Research Use Statement, or make other necessary changes, then resubmit the project through your dbGaP account. Please see [here](#) for more about how to revise and resubmit an existing dbGaP project.
- g. **To review:** After a project is submitted for “SO Review,” if by some reasons the application is returned back to the PI by the SO for revision, or the PI has chosen to withdraw the application using “recall from SO” link, or “Recall Application” button, all the DARs under the project will be under “to review” status. This gives the PI an opportunity to revise and resubmit the project for SO review again.

(7/13/2017)

Time Estimate for Data Access Committee (DAC) Review

I've applied for two studies. Can you give me an estimate for how long it will take for a decision to be reached after the application is submitted?

The length of time for processing an application request depends on:

1. How fast the Signing Official (SO) signs the application.
Through an automatically generated email, the SO of the dbGaP project should be notified to review the data access requests (DARs) when a project is submitted. You may want to directly contact the SO to expedite this process.
2. How fast the Data Access committee(s) DAC approves the application.
The time required for DAC review varies depending on many factors, however, average DAC review is two weeks

Since dbGaP is not directly involved in the review and approval process of DARs, you may have to directly contact respective DACs to check on the progress of your application. You will find that your application includes the email address for the DAC of each study. The DAC email address can also be found from the “Authorized Access” section of respective study page on dbGaP public website (see [here](#) for the screenshot).

(07/05/2017)

Signing Official (SO) and IT Director

How can I find SO and IT Director for my dbGaP project?

A list of the Signing Official (SO) at a Principal Investigator (PI's) institution registered with the NIH eRA Common is available under the “Project details” page. To find the page, go to “My Projects” tab, click on the

project name, and click on “Project details” sub-tab. The SO registration with the eRA is often handled by institution’s eRA administrator. Please see [here](#) for more about institutional Signing Official (SO).

The IT Director is generally, 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](http://www.ncbi.nlm.nih.gov/projects/gap/pdf/dbgap_2b_security_procedures.pdf) and the institution’s IT security requirements and policies are followed by the Approved Users..

(02/06/2013)

Select Datasets by Consent Groups

How to select dbGaP datasets? Where can I get the consent information of a study prior to submitting a data access request?

The dbGaP data are organized into consent groups which consist of all of the data from study participants who have agreed to the same data use limitations as specified in the informed consent for the study. When applying for dbGaP data, the data available are grouped in unit of consent groups.

Data access is only approved in unit of consent group, the data requesters therefore should understand the Data Use Limitations of a consent group prior to applying for dbGaP data access. Data Use Restriction of a dbGaP study is listed in the “Authorized Access” on study page. [Here](#) is how to find it.

Please note that, **for a given study, participants in different consent groups do not overlap**, which means that you would have to get the data of all consent groups to cover all participants of the study.

(7/27/2017)

No Need to Complete the Request in a Single Session

Is it possible to begin a data access request; complete some sections, save the work and come back later to complete further sections?

Yes, you can complete some sections of the application and return to finish in one or more additional sessions, as the data access request application system will save your application at each step when you click the green "Next Step" button at the bottom. Note: the text of the green button may be slightly different depending on the context of the button. (08/11/2017)

How to Write the Research Use Statement

What kind of research use statement is necessary for a data access request? Does my project need to be closely related to the dbGaP study of interest to qualify for access?

NIH Data Access Committees (DACs) review research use statements to make sure the proposed research is consistent with the data use limitations on the requested datasets in the data access request(s) (DAR). The RUS 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 a brief description of any planned collaboration with researchers at other institutions,

including the name of the collaborator(s) and their institution(s). Additional guidance to how to write a research use statement can be found on the Office of Science Policy [website](#)..

Some studies have additional instructions that are required in their applications for controlled-access. If you have study-specific questions, you may wish to contact the Data Access Committee (DAC) for that study. The email address for each study's DAC is provided in the "Authorized Access" section of the study page (see screenshot [here](#)).

(10/19/2017)

Examples of Research Use Statements

I am working on the online application forms. Are there examples of research statement of approved data requests that I can take a look?

All research statements of approved data access requests (DARs) are posted on the respective study page of the dbGaP public website. There is a link named "Authorized Requests" on top of the study page (see screenshot below).



The link leads to "Authorized Data Access Requests" section that contains research use statements of approved datasets. You may need to use "Show/Hide" next to "Research Use Statement" on each requester's section to view text. The "Public Research Use Statement" and "Technical Research User Statement" are both included.

Authorized Data Access Requests

[Download list of all approved requestors](#)

There are 175 authorized requests associated with this study.

1 2 3 4 5 6 7 8 ... 18 Next

- Requestor:** ALLISON, DAVID
Affiliation: UNIVERSITY OF ALABAMA AT BIRMINGHAM
Project: Prediction of Human Phenotypes Using Whole Genome Markers from the Framingham Population Data Set.
Date of approval: Oct 01, 2010
Request status: approved
Research use statements ([Show](#))

- Requestor:** ALLISON, DAVID

(07/03/2012)

IRB Requirement

What is IRB approval? How do I know if a specific dataset needs IRB approval to access?

An Institutional Review Board (IRB), privacy board, and/or equivalent body is a committee that reviews and monitors research involving human subjects. The group serves an important role in the protection of the rights and welfare of human subjects research. Some datasets require IRB, privacy board, and/or equivalent

body approval for use, as noted on the dbGaP study page displayed in the “Authorized Access” section.. Please see [here](#) for more about how to locate it.. 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.

(08/01/2013)

For Institutional Review Board (IRB)-Related Questions Contact the Data Access Committee (DAC)

The requested dbGaP dataset requires IRB approval, however, we do not belong to a university, and thus do not have a local IRB. Will approval from a non-local IRB be acceptable?

Any questions about IRB approval should be directed to the Data Access Committee (DAC) for the study in question. The email address of the DAC for a specific study is located in the “Authorized Access” section of the Study page. Please see the screenshot at [here](#). (06/15/2011)

Add or Update Institutional Review Board (IRB) Approval Documents

How to add or update IRB approval documents for my data requests?

If any of the datasets included in a dbGaP project have an IRB requirement, the respective IRB approval documents in a PDF format will have to be uploaded in order to complete the application process. This needs to be done by the Principal Investigator (PI) of the data access request (DAR) through the dbGaP [Authorized Access System](#). The following is how to add or update IRB documents:

After logging in to the dbGaP system, please do the following:

1. Click on the “My Projects” tab, and click on the project name.
2. For new projects or for existing projects that wish to add datasets, make sure datasets are selected under “Choose Datasets”.
3. Confirm selected datasets by hitting “Next” button until you reach, “Review DUC”.
4. On the “Review DUC” page, if any of the datasets included in the project have an IRB requirement, you will see the “Program-specific Required Attachments” section at bottom of the page. Use the “Upload” button to upload IRB approval documents for the datasets shown in the table.
5. To update previously uploaded IRB approval documents for an existing project, go to the “Review DUC” page, in the “Program-specific Required Attachments” section at bottom of the page. Use the “Delete” button to remove existing IRB approval documents, and use the “Upload” button to upload new ones.

If you don't see the “Upload” button, please be aware that the **“Upload” button is visible only if the dataset that requires an IRB approval document is selected**. You may want to go back to the “Choose datasets” page to select the dataset first.

(10/19/2011)