**NIH** U.S. National Library of Medicine
National Center for Biotechnology Information

# Downloading Data

Created: October 22, 2008; Updated: August 12, 2013.

## Aspera Connect

**What is Aspera software, where to get it?**

The dbGaP Authorized Access System uses Aspera, a high-speed file transfer system, to facilitate client download. It requires Aspera Connect to be installed on client's download machine. Aspera Connect is an install-on-demand browser plugin. It is available for free on the Aspera website. From the software download page, please make sure to select and install Aspera Connect instead of any other Aspera client products. Aspera Connect is available for Linux, Mac, and Windows platforms. In addition to the web user interface, Aspera Connect also includes a command line ASCP executable utility. (**04/21/2015**)

## Download Procedure

**Download using prefetch command-line utility with the cart file or SRA accession**

The principal investigator (PI) of the project or downloaders designated by the PI can download the data as soon as the data access request is approved. The recommended way of downloading dbGaP data is using the "prefetch" utility available in the NCBI SRA toolkit.

The prefetch utility can download dbGaP non-SRA and SRA data files in bulk when a cart file is provided as an argument. It can also download the data of individual SRA run when individual SRR accession is provided as an argument. The documentation of prefetch can be found from here.

The following are main steps of downloading with prefetch.

1. Download and install Aspera Connect (see here for more information).

2. Select and save data files information in a cart (.kart) file

   (For SRA data download, in addition to bulk download with cart file, the prefetch can also run with individual SRA accession, which is often preferred method for program/script directed automatic download. See the section 5 for more about this.)

   - Login to the dbGaP Authorized Access System using the eRA account login credentials. (Intramural NIH scientists and staff need their NIH email username and password).
   - Click on "My Requests" tab. The list of Approved Requests is under "Approved" sub-tab.
   - Find the table row of approved dataset, click on the link named "Request Files" in the "Actions" column.
   - On the "Access Request" page, different types of data files available for download are shown separately under different sub-tabs. To download non-SRA data, go to the "Phenotype and Genotype files" sub-tab and click on the "dbGaP File Selector" link. To download SRA data, go

to the "SRA data (reads and reference alignments)" sub-tab and click on the "SRA RUN Selector" link.

- Wait until the page loading is complete. Click on the "Help" icon on top of the page to see instruction/information about the selector).
- Add/remove files using the facets listed in the left panel facet manager. From the right panel file list, select/unselect files by checking/unchecking checkboxes in front of the file names.
- Once the files are selected (checked), click on the "Cart File" button (on the upper part of the page) and save the cart file (.kart).

3. Download and decrypt dbGaP data files

- Download the **latest** version of the NCBI SRA Toolkit. Untar or unzip downloaded toolkit file.
- Follow this dbGaP Download Guide to download dbGaP data using the sratoolkit.

4. Specific steps and commands

Before running the download commands below, make sure the dbGaP repository key (.ngc) and the cart files are ready.

- Download a fresh dbGaP repository key (.ngc) file and re-config the toolkit with the command below.
  *$ /path-to-your-sratoolkit-installation-dir/bin/vdb-config -i*
- From the sratoolkit GUI interface, import the repository key
- Download dbGaP data files
- Run the command below to download the files specified in the cart file.
  *$ /path-to-your-sratoolkit-installation-dir/bin/prefetch --ngc /path-to-ngc-file-dir/xxxxx.ngc /path-to-your-cart-file/xxxxx.krt*
  Please make sure the sratoolkit, ngc, and cart files are on the same disk drive.
- Decrypt downloaded files
- The downloaded dbGaP non-SRA files need to be decrypted before use. Run the command below to decrypt the files.
  *$ /path-to-your-sratoolkit-installation-dir/bin/vdb-decrypt --ngc /path-to-ngc-file-dir/xxxxx.ngc /path-to-top-level-download-dir/*

5. Compatibility issue with older versions of sratoolkit

If 2.9.6 or older version of the sratoolkit had been installed and used on the machine, before running above commands, the old toolkit settings need to be disabled by renaming the settings file as below.
*$ cd ~/.ncbi*
*$ mv user-settings.mkfg user-settings.mkfg.old*

**(12/08/2020)**

## How to Add Downloaders to Projects?

**I am a principal investigator (PI). Is it possible to allow my lab staff or collaborator to download data without sharing my eRA login credentials?**

Here is a video related to this topic. Recently improved user-interface of the dbGaP Authorized Access System allows principal investigator (PI) to designate one or more downloaders within PI's institution. A Downloader is an individual assigned by the PI to perform the time-consuming task of retrieving large data files. The downloaders can login to the dbGaP system through their own account and make download. The download is limited to the data sets approved to access and specified for downloader by primary PI.

The following is how to assign downloaders to approved datasets within all or specific projects:

1. Login to the dbGaP Authorized Access System as a PI using the eRA login credentials; If respective project hasn't yet been created, create the project and follow multiple steps to complete and submit the online application.
2. Navigate to "Downloader" page through "Downloaders" tab. Search for the name of intended downloader by the first name and last name using the search boxes.

**Note**: A downloader needs to have a valid NIH eRA Commons account or a NIH email account, and have successfully logged into the dbGaP Authorized Access System at least once. Downloader's eRA account does not need to have a PI role, but it does need to be affiliated with PI's institution.

1. Confirm to make sure the resulting user name is correct; Click on the name; select all or a specific project from the pull-down manual, and finally click on "Set downloader" button to make the assignment. The downloader's name and the projects accessible to the downloader will be displayed on the page.
2. The PI can use the "X" buttons in "Remove Role" column of downloader table to remove any downloaders or downloader's projects.

(**07/13/2011**)

## How to Become a Downloader?

**I am a data analyst working for a principal investigator (PI) who has multiple approved data access requests. How can I download PI's datasets without logging into his account?**

Here is a video related to this topic. Downloader has to be designated by the PI through the dbGaP system. Please see here for more details. Prior to be chosen as a downloader, the individual must

1. Have a valid NIH eRA Commons account affiliated with the same organization as the PI, or has an NIH email account. The eRA account does not need to have a PI role.
2. Have already completed at least one successful login to the dbGaP Authorized Access System.

(**07/12/2011**)

## Download Procedure for Downloader

**I am a downloader designated by the principal investigator (PI). How do I make download?**

The download procedure is nearly the same for PI and for downloaders. Please see here for more details. (**06/30/2011**)

## Expired Download Package

**My download package is expired. What can I do with it?**

In most of cases, the expiration interval of a download package is set to two months. You can always delete expired package and order a new one if you need to download the same data again. The new download package can include some or all of the previously downloaded files. Please see here for more details. (**06/30/2011**)

## FTP Site Availability for Downloads

**Can I use FTP instead of Aspera to download dbGaP data? I don't have large file to download.**

No, the FTP interface is no longer available for downloading dbGaP data. The Aspera Connect is the only choice. (**06/21/2011**)