

The Bionimbus PDC:

Obtaining Access FAQ



Center for
Data Intensive Science



OPEN COMMONS CONSORTIUM



OPEN SCIENCE DATA CLOUD

TABLE OF CONTENTS

PREREQUISITES	3
LEGAL DOCUMENTS	3
SECURITY TRAINING	3
GENERAL GUIDELINES	4
<u>AUTH METHOD 1: USING AN ERA TO GAIN ACCESS TO A DBGAP DATASET</u>	5
GETTING AN ERA COMMONS ACCOUNT	5
GETTING DBGAP ACCESS	5
DATA ACCESS REQUEST BOILERPLATE	5
REQUESTING PROJECT ACCESS TO A DBGAP DATASET	6
GRANTING PROJECT ACCESS TO INDIVIDUALS IN YOUR LAB	8
<u>AUTH METHOD 2: ERA, SHIBBOLETH OR OAUTH ACCESS TO A SUPPORTED PDC PROJECT</u>	9

Prerequisites

The Bionimbus PDC is a HIPAA compliant cloud for analyzing and sharing protected data. The Bionimbus PDC is an OpenStack cluster utilizing ephemeral storage in VMs with access to a separate S3-compatible storage system for persistent data storage. Allocations to all users and projects are given at the “tenant” level. To learn more about how the PDC operates and the user experience, visit:

<https://www.opensciencedatacloud.org/support/pdc.html>

For a researcher to gain access to the PDC, a researcher must:

- Be associated with a project the PDC authenticates
 - Method 1: a project with access to dbGaP controlled data
 - Method 2: an existing PDC project with contract supported authentication
- Show proof of completion of security training
- Sign some legal documents
- For researchers requiring ssh access to VMs, they need to complete the application at: https://bionimbus-pdc.opensciencedatacloud.org/pre_apply/?next=/apply/

Legal documents

All applicants are required to have signed:

- Acceptable Use Policy (AUP)
- Code of Conduct (COC)
- Non-Disclosure Agreement (NDA)

For user convenience we have compiled these documents into a single .pdf that can easily be e-signed. For most applicants the Bionimbus PDC account shepherd will send this using Adobe’s Echosign. These generally appear in an inbox under the ‘echosign’ domain.

For special cases that bypass the traditional account shepherd process, the legal documents are available at: https://bionimbus-pdc.opensciencedatacloud.org/static/BionimbusPDC_AUP_COC_NDAv1.1.pdf

Security Training

The Protected Data Cloud provides a computing environment designed to support research with protected health information and other sensitive data, but the computing environment is only one component of what is required to properly secure protected health information or other sensitive data. Another important requirement is that those using a computing infrastructure like the PDC have proper training and have received from their institution all the required approvals for working in secure environments.

We require the following documentation:

1. For each project, that the PI of the project send us a copy of the IRB approved protocol for the study using the Protected Data Cloud or a letter showing the study is exempt from needing an IRB Protocol. The IRB Protocol or Exemption should be from the home institution of the PI for the project. For new protocols or renewals of protocols, please refer explicitly to the PDC environment.
2. For each researcher with a Protected Data Cloud account, a certificate proving that the researcher has completed CITI training appropriate for working in secure environments. For international researchers, we can accept NIH security training as an alternative. A copy of a certificate indicating that you have completed the required training will be requested each year.

For more information visit:

<https://www.opensciencedatacloud.org/support/intro.html#pdc-handling-sensitive-data-training>

Once complete, users should submit this certificate to accounts@occ-data.org if they are not already carrying on an email conversation around their application.

General Guidelines

It's very important that a PDC user is the only one who logs in with an account. If you have members of your lab that would like to participate in this research, we ask that they each get separate accounts on the PDC. Having an account used by a single prescribed user is an important step in keeping the PDC secure and protected.

Before your account is created on the Bionimbus Protected Data Cloud (PDC), you'll need the following credentials:

- 1) If Method 1: eRA Commons through dbGaP
 - i. an eRA Commons account
 - ii. dbGaP access
 - iii. dbGaP access to a project with approved access to a TCGA, TARGET or CGCI data
- 2) If Method 2: eRA, Shibboleth, or Oauth addition to a supported project
 - i. an eRA Commons account OR a Shibboleth id OR an Oauth ID (only as a last resort)
 - ii. addition to the maintained authorization list for the project the user is associated with

For all users we authenticate their credentials against lists provided by dbGaP for selected datasets (TCGA, TARGET and CGCI) OR authentication lists maintained through special contracts.

What follows are detailed instructions for gaining the credentials required for secure authentication.

Auth Method 1: Using an eRA to Gain Access to a dbGaP Dataset

Getting an eRA Commons Account

You can learn more generally about eRA Commons and how to register your organization or institution at: https://era.nih.gov/commons/faq_commons.cfm#II2

A list of other helpful links to learn more about the eRA Commons process include:

- Main eRA Commons access: <https://public.era.nih.gov/>
- Helpful Steps for Navigating eRA Commons: https://era.nih.gov/commons/faq_commons.cfm or https://era.nih.gov/Docs/COM_UGV2630.pdf
- NIH / eRA Helpdesk/Ticketing System: <http://grants.nih.gov/support/index.html>

NOTE: If you're working with an academic institution, they can often help navigate this process.

Getting dbGaP Access

- 1) To get authorized access start at this site: <https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=login>
- 2) Login to the webpage above and follow the instructions.
- 3) You will probably complete form SF 424, which will have the section listed in 4. below.
- 4) Below is some sample text you can use to fill out certain sections of the application for access.

Data Access Request Boilerplate

Research use statement:

[ENTER SHORT DESCRIPTION OF PROJECT/ANALYSIS INTENDED]

The **[ENTER PROJECT NAME]** will be utilizing the technical capabilities of the Bionimbus PDC to analyze multiple protected datasets.

The PDC is a secure FISMA compliant biomedical computing environment with an NIH trusted partner status that includes approved access to select dbGaP data for distribution to approved authorized users. It is a project that is building a computing resource for the research community so that authorized researchers can manage, analyze and share large quantities of protected human genomic data in a secure and compliant fashion. The Bionimbus PDC is designed to hold both controlled access data (e.g. Level 1 BAM files from CGHub) and open access data (e.g. de-identified clinical data from the TCGA Data Coordination Center). This computing model allows researchers immediate access to data and computational resources without the burden of maintaining numerous servers and large repositories of sensitive data in a local facility. This is especially important for projects such as The Cancer Genome Atlas (TCGA), which currently has about 500 terabytes of Level 1 data and is projected to grow to multiple petabytes. The PDC

provides the capability to complete the proposed analysis on an expedited, compressed schedule.

Non-technical summary of Research use statement:

The proposed analysis of [ENTER PROJECT NAME AND POTENTIAL RESULTS SOUGHT] will be expedited using the Bionimbus PDC.

The Bionimbus PDC is a project that is building a computing resource for the research community so that authorized researchers can manage, analyze and share large quantities of protected human genomic data in a secure and compliant fashion. With the Bionimbus PDC, users can login and have immediate access to select datasets that they are authorized to view as well as computational resources so that they can easily analyze the data without first downloading it.

Cloud Use Statement:

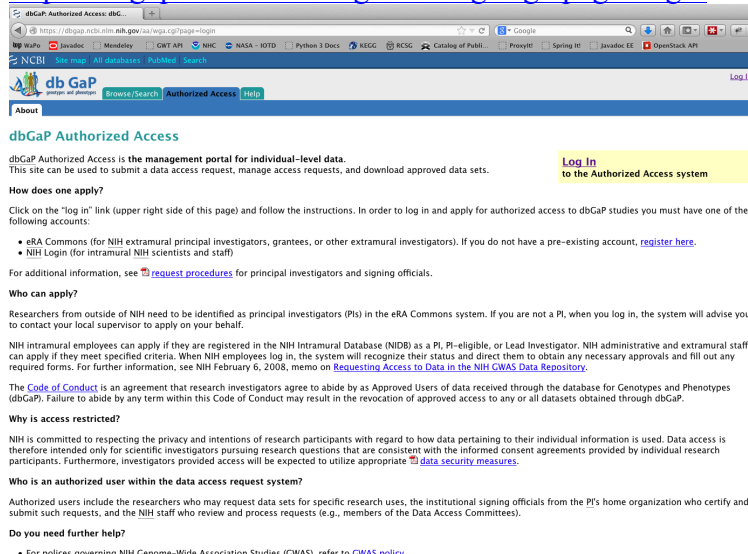
The Bionimbus PDC is a secure biomedical cloud operated at FISMA moderate as IaaS with an NIH Trusted Partner status for analyzing and sharing protected datasets. The Bionimbus PDC is an OpenStack cluster utilizing ephemeral storage in VMs with access to a separate S3-compatible storage system for persistent data storage. Allocations to all users and projects are given at the “tenant” level.

The Bionimbus PDC provides self-managed Linux virtual machines with the most recent long term support (LTS) release. The resource allows the user to determine the tools they wish to use and install on those images. Installation is accomplished by manually whitelisting requested repositories and having the user run all external requests from their VMs through a proxy server.

Requesting Project Access to a dbGaP Dataset

1) Once you have access to dbGaP, go to the dbGaP login page and login in the upper right with your eRA Commons username and password:

<https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=login>

The screenshot shows the 'dbGaP Authorized Access' web page. At the top, there's a navigation bar with links for 'About', 'Browse/Search', 'Authorized Access', and 'Help'. The main heading is 'dbGaP Authorized Access'. Below it, a sub-heading states: 'dbGaP Authorized Access is the management portal for individual-level data. This site can be used to submit a data access request, manage access requests, and download approved data sets.' To the right of this text is a yellow button labeled 'Log In to the Authorized Access system'. The page then asks 'How does one apply?' and provides instructions to click the 'log in' link. It lists two account types: 'eRA Commons (for NIH extramural principal investigators, grantees, or other extramural investigators)' and 'NIH Login (for intramural NIH scientists and staff)'. It also mentions a 'request procedures' link for principal investigators. A section titled 'Who can apply?' explains that researchers from outside NIH need to be identified as principal investigators (PIs) in the eRA Commons system. Another section, 'Why is access restricted?', states that NIH is committed to respecting privacy and that data access is for scientific purposes only. A final section, 'Who is an authorized user within the data access request system?', mentions that authorized users include researchers and institutional signing officials. At the bottom, there's a 'Do you need further help?' section with a link to 'GWAS policy'.

2) This will take you to the NIH Secure Identity Solutions Login Page:

NIH Login User Name and Pass...

https://trustauth.nih.gov/CertAuth/forms/NIHPwOrFormLogin.aspx?TYPE=33554433&REALMOID=06-1fe501d6-778a-4dec-8af...

Wapo Javadoc Mendeley GWT API NHC NASA - IOTD Python 3 Docs KEGG RCSG Catalog of Publi... Proxy! Spring It! Javadoc EE OpenStack API

aTrust
NIH SECURE IDENTITY SOLUTIONS

User Name: grossmanr

Password: ***** [Change Password](#)

OR

Insert your PIV card into your smart card reader before attempting to login.

For assistance, read the instructions for [using smart cards and certificates with NIH Login \(PDF, 21 pages, 726 KB\)](#).

[Log in](#)

Warning Notice

This is a U.S. Government computer system, which may be accessed and used only for authorized Government business by authorized personnel. Unauthorized access or use of this computer system may subject violators to criminal, civil, and/or administrative action.

All information on this computer system may be intercepted, recorded, read, copied, and disclosed by and to authorized personnel for official purposes, including criminal investigations. Such information includes sensitive data encrypted to comply with confidentiality and privacy requirements. Access or use of this computer system by any person, whether authorized or unauthorized, constitutes consent to these terms. There is no right of privacy in this system.

If you need assistance - Please call the NIH IT Service Desk call 301-496-4357 (local), 866-319-4357 (toll-free), or 301-496-8294 (TTY), or [Submit NIH IT Service Desk Ticket](#)

3) From here, you'll see the "My Projects" page. Here you can create a new research project (follow the required steps after hitting "Create New Research Project") or revise your project to request TCGA, TARGET or CGCI access ("Revise Project" link on the right):

NCBI Site map All databases PubMed Search

dbGaP
genotypes and phenotypes

Browse/Search Authorized Access Help

Logged in as Robert Grossman | [Log out](#)

My Projects My Requests Downloaders My Profile

My Research Projects

[Create New Research Project](#)

#	Project	Actions
5296	Bionimbus Protected Data Cloud <i>Bionimbus PDC Pilot</i> SO: Carol Zuchies, UNIVERSITY OF CHICAGO	revise project close out project get embargo report get dbGaP repository key

NIH Genotype and Phenotype database is a service of NCBI. Please [contact us](#) with any questions.
National Center for Biotechnology Information | U.S. National Library of Medicine
[Privacy Notice](#) | [Disclaimer](#) | [Accessibility](#)

4) You can check and see if your Research Project has the desired dataset access by going to the "My Requests" tab. You should see "Approved" under the status.

dbGaP
genotypes and phenotypes

Browse/Search Authorized Access Help

Logged in as Robert Grossman | [Log out](#)

My Projects **My Requests** Downloaders My Profile

Request List

Approved (1)

#	Study, Consent	Status	Expiration	Actions
Project #5296: Bionimbus Protected Data Cloud 22135-2	TCGA - The Cancer Genome Atlas (phs000178.v8.p7) General Research Use (phs000178.v8.p7.c1), TCGA	✓ Approved GRANTED	2014-08-06	Request Files Processing History Renew application

NIH Genotype and Phenotype database is a service of NCBI. Please [contact us](#) with any questions.
National Center for Biotechnology Information | U.S. National Library of Medicine
[Privacy Notice](#) | [Disclaimer](#) | [Accessibility](#)

5) If you do not see the desired dataset as approved under your research project, choose the “Datasets” Tab, select the desired dataset click “Add Selected and Continue” at the bottom of the page, and walk through the required steps.

The screenshot shows the dbGaP Project Request page for the Bionimbus Protected Data Cloud. The page is titled "Project Request" and includes a sub-header "#5296: Bionimbus Protected Data Cloud". Below this, there is a navigation bar with tabs: "Project Details", "Research Project", "Collaborators", "IT Director", "Choose Datasets", "Confirm Datasets", "Review DUC", "Review DUL", "Review Applications", "Data Security", and "Feedback". The "Choose Datasets" tab is currently selected. The page contains a table with columns: "Consent Group", "Data Use Limitations", "Participants", and "DAR Status". The table lists two datasets: "A Genome Wide Scan of Lung Cancer and Smoking" and "A Genome-Wide Association Study of Peripheral Arterial Disease".

Consent Group	Data Use Limitations	Participants	DAR Status
A Genome Wide Scan of Lung Cancer and Smoking (phs000093.v2.p2)			
<input type="checkbox"/> Research related to adult diseases and methods (phs000093.v2.p2.c1), CCEMS	The informed consent document signed by the PLCO study participants allows use of these data by investigators for discovery and hypothesis generation in the investigation of the genetic contributions to cancer and other adult diseases as well as development of novel analytical approaches for GWAS.	1629	
<input type="checkbox"/> Research related to smoking or lung disease (phs000093.v2.p2.c2), CCEMS	Use of the dataset is limited to scientific genetic research related to the etiology, molecular basis, and outcome of lung disease and smoking.	3937	
A Genome-Wide Association Study of Peripheral Arterial Disease (phs000203.v1.p1)			

Granting Project Access to Individuals in your Lab

1) Once you have an accepted research project and dataset access, you can give access to other members of your lab who will require PDC accounts. To do so, go to the “Downloaders” tab:

The screenshot shows the dbGaP Downloaders page. The page is titled "Downloaders" and includes a search bar with "Find user: First name" and "Last name" fields, and a "Find" button. Below the search bar, there is a section titled "The prerequisites to be chosen as a downloader:" with two bullet points: "1. Have an eRA Commons account or a NIH email account. The eRA account doesn't need to have a PI role." and "2. Successfully logged into the dbGaP Authorized Access System at least once."

Downloaders on your projects:

Project	Remove Role
Raymond Powell	<input type="button" value="X"/>
#5296: Bionimbus Protected Data Cloud	<input type="button" value="X"/>
Allison heath	<input type="button" value="X"/>
#5296: Bionimbus Protected Data Cloud	<input type="button" value="X"/>
Leonie Misquitta	<input type="button" value="X"/>
#5296: Bionimbus Protected Data Cloud	<input type="button" value="X"/>
Matthew Greenway	<input type="button" value="X"/>
#5296: Bionimbus Protected Data Cloud	<input type="button" value="X"/>

2) The person you are giving access to must a) have an eRA Commons account and b) have successfully logged into the dbGaP Authorization System at least once using their eRA (<https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=login>)

Use the “Find User” fields for First Name and Last Name. The Lab member should come up in the list. NOTE: There’s a small scroll bar on the right that you can easily miss.

dbGaP: Authorized Access: Dow...

https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=downloaders_manage&fname=raymond&lname=&find_req_user=Find

NCBI Site map All databases PubMed Search

db GaP genotypes and phenotypes Browse/Search Authorized Access Help

Logged in as Robert Grossman | Log out

My Projects My Requests Downloaders My Profile

Downloaders

Find user: First name Last name

The prerequisites to be chosen as a downloader:

- 1. Have an eRA Commons account or a NIH email account. The eRA account doesn't need to have a PI role.
- 2. Successfully logged into the dbGaP Authorized Access System at least once.

Select user:

- Raymond Kim, Geneprium, Inc.;
- Raymond Cho, University Of California San Francisco;
- Raymond Cho, University Of California San Francisco;
- Raymond Hovey, Medical College Of Wisconsin;

3) Select their name, and add them to the project, hit “Set Downloader”

dbGaP: Authorized Access: Dow...

https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=downloaders_manage&pid=8181&filter=pid

NCBI Site map All databases PubMed Search

db GaP genotypes and phenotypes Browse/Search Authorized Access Help

Logged in as Robert Grossman | Log out

My Projects My Requests Downloaders My Profile

Downloaders

Find user: First name Last name

The prerequisites to be chosen as a downloader:

- 1. Have an eRA Commons account or a NIH email account. The eRA account doesn't need to have a PI role.
- 2. Successfully logged into the dbGaP Authorized Access System at least once.

Give downloader role to:

Renuka Arya

Email: rarya@medicine.bsd.uchicago.edu

Phone: 2242504293

Organization(s): University Of Chicago;

For project(s):

Auth Method 2: eRA, Shibboleth or Oauth Access to a Supported PDC Project

For some projects with special contracts to use the Bionimbus PDC, users will gain access by working with their project shepherds to add their eRA Commons, Shibboleth or Oauth credentials to the access list maintained by their project.

For this they will be required to have an identifier of either eRA Commons, Shibboleth, or Oauth (only as a last resort).