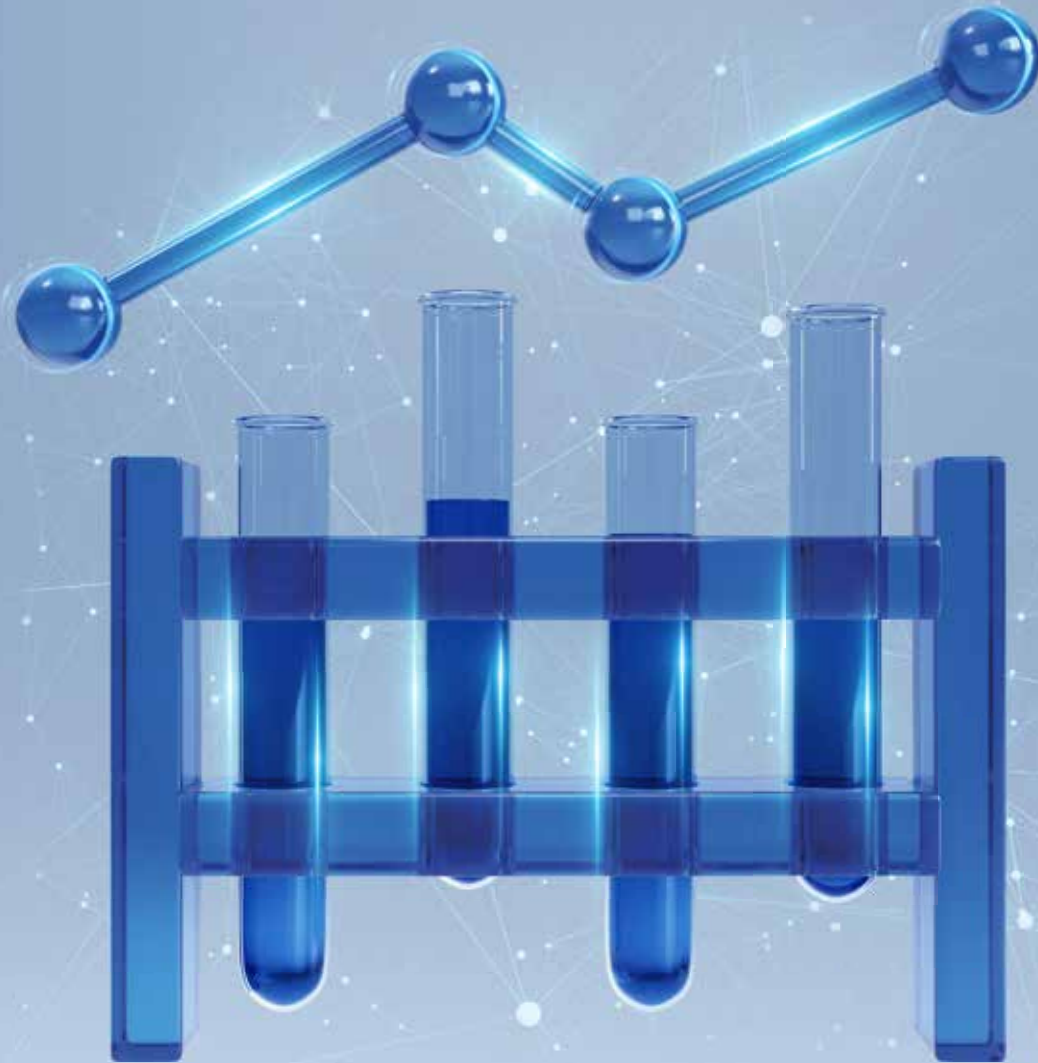


CATAPULT

Medicines Discovery



Data Access Quick Guide to



May 2020

The US National Institute of Health's (NIH) National Center for Biotechnology information (NCBI) is responsible for creating automated systems for storing and analysing knowledge about molecular biology, biochemistry, and genetics and their databases and software are used by research and medical communities world-wide.

A list of all NCBI databases is available [here](#)

dbGaP

The database of Genotypes and Phenotypes (dbGaP) was developed to archive and distribute the data and results from studies that have investigated the interaction of genotype and phenotype in humans. It was created by the NCBI to support investigator access to data from large-scale cohort and clinical studies funded by the NIH, the primary agency of the United States government responsible for biomedical and public health research (and others).

A note about The Cancer Genome Atlas (TCGA)

The Cancer Genome Atlas (TCGA) data is hosted at the Genomics Data Commons website but dbGaP manages the controlled access approval process, so in order to obtain access to the controlled data available in the GDC, Principal Investigators must first obtain an NIH eRA Commons account and then obtain authorization to access the data through the NIH database of Genotypes and Phenotypes (dbGaP).

eRA Commons

eRA Commons is an online interface where grant applicants, grantees and federal staff at NIH and grantor agencies can access and share administrative information relating to research grants.

To do anything in eRA Commons you must have an account. The user's applicant organization must be registered in Commons. Only individuals with legal signing authority (the Signing Official, eg CEO) can register their organization in Commons. According to the eRA Commons website, the applicant

organization MUST be registered with the following before they can successfully submit an application to NIH:

1. **The System for Award Management (SAM)** - an official website of the US government which can be used for free to register and do business with the US government. To register in SAM firstly international registrants must have:
 - a) **NATO Commercial and Government Entity (NCAGE) code** (can be requested for free)
For UK companies - [NCAGE registrations can be created](#) (allow 5 working days to allow updates between UK and US CAGE databases)
 - b) **DUNS number** – a unique nine-digit identification number for each physical location of a business, required to register with the US federal government for contracts or grants. (Allow 5 working days)

To register on SAM you must generate a login.gov account. This can be done from the SAM portal. It requires an email and a working phone number as you have to set up second authentication method (text, phone call, authentication app, security key or back up codes).

According to the Federal Service Desk website, new registrations in SAM can take an average of 7-10 business days to process providing all information is complete and accurate. **It is very important that when registering with the different systems that the details (company name, address fields) match exactly. Otherwise there can be issues with rejection of the SAM application due to data discrepancies.**

To complete SAM registration, a letter formally appointing an Entity Administrator is required, signed by someone with signing authority (eg CEO). [Access guidance and a template letter.](#)

2. **Grants.gov** (allow 1 business day)
Once a company is registered in SAM and grants.gov, the Signing Official for the company can register the company in eRA Commons and set up accounts for administrators and Principal Investigators.

Applying for access to data held in or managed by dbGaP

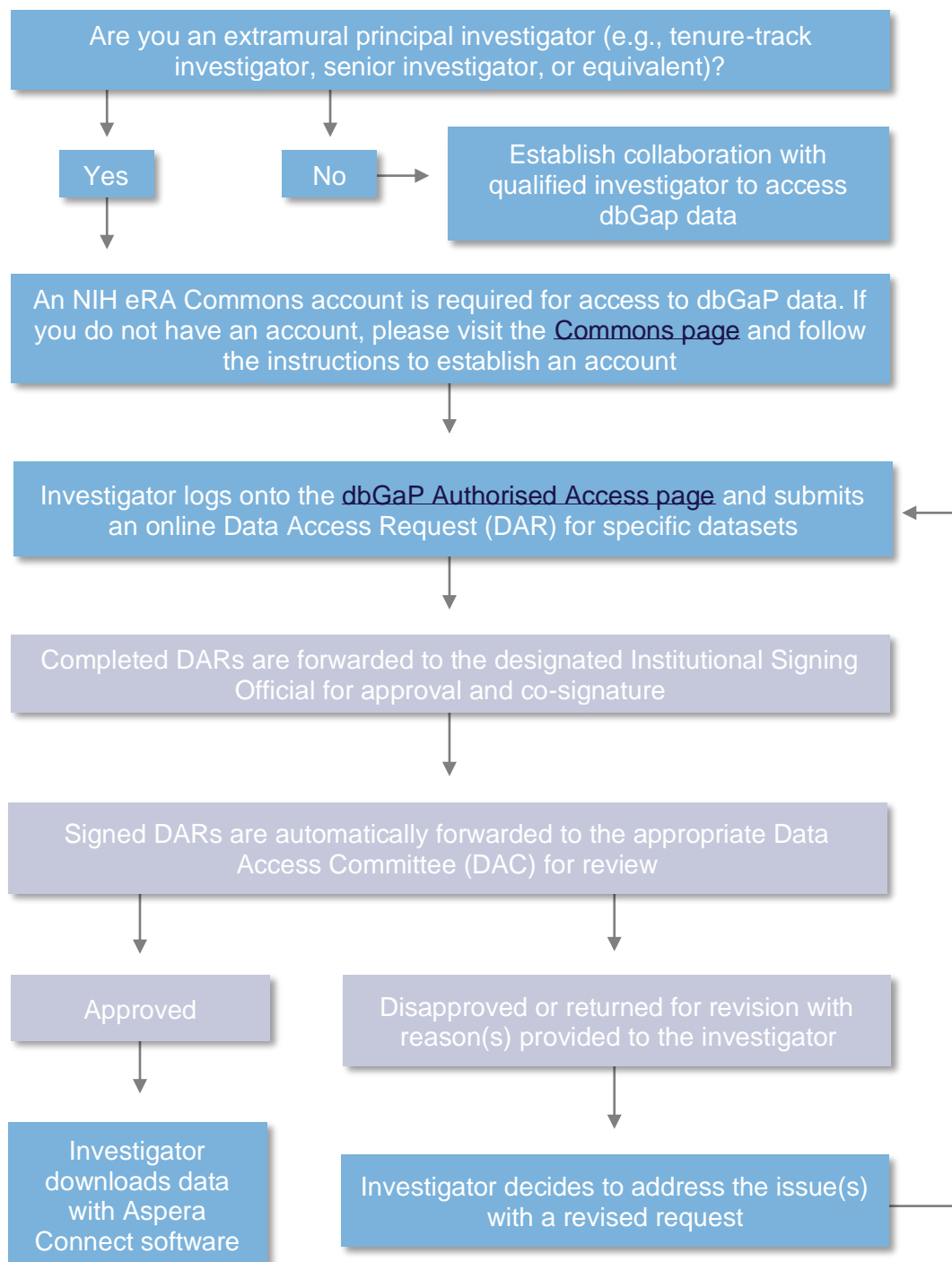
Some of the key requirements for dbGaP applications:

1. To submit an application, investigators must be permanent employees of their institution at a level equivalent to a tenure-track professor or senior scientist with laboratory administration and oversight responsibilities
2. Some datasets require local Institutional Review Board approval for use
3. Investigators from newly established or small organisations may be asked for supplementary information such as a CV and IT security plan
4. A clear Research Use Statement must be submitted outlining research objectives, study design, analysis plan, and collaborators This will be reviewed by a Data Access Committee
5. Investigators who wish to use cloud computing for storage and analysis of dbGaP data will need to indicate in the Data Access Request that they are requesting permission to use cloud computing and identify the cloud service provider that will be employed. They will also need to describe how the cloud computing service will be used to carry out their proposed research.
6. IT Director contact information must be given. They must have the necessary expertise and authority to confirm the IT capacities at the institution. The IT Director is expected to have the authority and capacity to ensure 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.

[Read the full guidance.](#)

The process for applying for access to data held in dbGaP is as follows (and available [here](#)):

Process for External Investigators to Access Datasets in dbGaP



Colour

- Steps for the investigator
- Administrative actions