White Paper: NCBI Database of Genotypes and Phenotypes (dbGaP) Security Best Practices – Compliance Overview for the New DNAnexus Platform

April 18, 2013

Overview

This White Paper summarizes how the new DNAnexus platform enables compliance with the Security Best Practices established by the NCBI for data sets included in its Database of Genotypes and Phenotypes (dbGaP), such as The Cancer Genome Atlas, that are subject to "controlled access."

For more information regarding security and compliance using the New DNAnexus platform, please see the additional security and compliance White Papers available at wiki.dnanexus.com/home.

What are dbGaP Security Best Practices? Who has to comply with them?

The NCBI established dbGaP "to archive and distribute the results of studies that have investigated the interaction of genotype and phenotype."[1] dbGaP datasets are organized into two tiers: Open Access and Controlled Access data.[2]

The Open Access data tier includes data that cannot be attributed to an individual research study participant. In contrast, Controlled Access data consist of individual-level data that are unique to an individual, even though the individual study participant's personal identifiers have been removed. These data include the following:

- Individual germline variant data (SNP .cel files)
- Primary sequence data (.bam files)
- Clinical free text files
- Exon Array files[3]

The NCBI explains the controlled access requirement, and the Security Best Practices that researchers must implement as a condition to their access to these data, as follows:

> NIH is committed to respecting the privacy and intentions of research participants with regard to how data pertaining to their individual information is used. Data access is therefore intended only for scientific investigators pursuing research questions that are consistent with the informed consent agreements provided by individual research participants. Furthermore, investigators provided access will be expected to utilize appropriate security measures.[4]

Consistent with this approach, the application for access to controlled data requires that investigators agree to adhere to specified security best practices.[5] To obtain access to TCGA, an investigator must

[1] www.ncbi.nlm.nih.gov/gap

[2] www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/about.html

[3] http://www.ncbi.nlm.nih.gove/projects/gap/cgi-bin/study.cgi?study_id=phs000178.v7.p6

[4] https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

[5] Id.

similarly agree to a Data Use Certification Agreement, which includes a provision requiring compliance with dbGaP Security Best Practices.[6]

How does the *new* DNAnexus Platform enable compliance with dbGaP Security Best Practices?

The balance of this White Paper explains how the use of the *new* DNAnexus Platform to analyze, manage, and share genomic information enables compliance with dbGaP Security Best Practices.[7]

"Think Electronic Security"

1. Requirement: "Download data to a secure computer or server and not to unsecured network drives or servers."
    a. Compliance with DNAnexus: DNAnexus provides secure servers for downloaded and stored data.
2. Requirement: "Make sure these files are never exposed to the Internet."
    a. Compliance with DNAnexus: Data stored with DNAnexus can be shared selectively using DNAnexus sharing controls, rather than posting, ftp'ing, or emailing insecurely.
3. Requirement: "Have a strong password for file access and never share it."
    a. Compliance with DNAnexus: DNAnexus enforces strong passwords.
4. Requirement: "If you leave your office, close out of data files or lock your computer."
    a. Compliance with DNAnexus: DNAnexus automatically locks sessions after 15 minutes of inactivity.
5. [no item 5]
6. Requirement: "Data stored on laptops must be encrypted."
    a. Compliance with DNAnexus: Data stored with DNAnexus does not require downloading to laptops for processing.

.

"Think Physical Security"

1. Requirement: If the data are in hard copy or reside on portable media, treat it as though it were cash.
    a. Compliance with DNAnexus: DNAnexus obviates the need for hard copies or copies on removable media, which are easy to lose.
2. Requirement: Don't leave it unattended or in an unlocked room.
    a. Compliance with DNAnexus: DNAnexus data are stored in highly secure data centers.

6 https://tcga-data.nci.nih.gov/tcga/tcgaAccessTiers.jsp

7 www.ncbi.nlm.nih.gov/projects/gap/pdf/dbgap_2b_security_procedures.pdf

3. Requirement: Consider locking it up.
    a. Compliance with DNAnexus: DNAnexus' data centers are secured with locks, video surveillance, and other high-security controls.
4. Requirement: Exercise caution when traveling with portable media, i.e., take extra precautions to avoid the possibility of loss or theft.
    a. Compliance with DNAnexus: Data stored with DNAnexus remains in its secure location, but you can access it securely, regardless of where you are.

"Protecting the Security of Controlled Data on Servers"

1. Requirement: Servers must not be accessible directly from the Internet, (i.e. must be behind a firewall or not connected to a larger network) and unnecessary services disabled.
    a. Compliance with DNAnexus: All DNAnexus servers are protected by stateful packet inspection firewalls, with only necessary services allowed.
2. Requirement: Keep systems up to date with security patches.
    a. Compliance with DNAnexus: DNAnexus applies all relevant security patches within 30 days.
3. Requirement: dbGaP data on the systems must be secured from others and if exported via file sharing, ensure limited access to remote systems.
    a. Compliance with DNAnexus: DNAnexus offers tight control of sharing - only users you specify can access your data.
4. Requirement: If accessing system remotely, encrypted data access must be used.
    a. Compliance with DNAnexus: All DNAnexus data are transmitted using HTTPS, which provides encrypted data access.
5. Requirement: Ensure that all users of this data have IT security training suitable for this data access and understand the restrictions and responsibilities involved in access to this data.
    a. Compliance with DNAnexus: You can designate "project" administrators who control access to the data. Ensure all your users know to leave data in DNAnexus and with whom it can and cannot be shared.
6. Requirement: If data is used on multiple systems (such as a compute cluster), ensure that data access policies are retained throughout the processing of the data on all the other systems. If data is cached on local systems, directory protection must be kept, and data must be removed when processing is complete.
    a. Compliance with DNAnexus: Once you have set sharing policies on a project, these data access policies are automatically retained for all data, servers, processing, and outputs associated with that project. Data do not need to be cached on local systems.

"Use Data by Approved Users on Secure Systems"

1. Requirement: The requesting investigator must retain the original version of the encrypted data. The requesting investigator must track any copies or extracts made of the data and shall

make no copy or extract of the subject data available to anyone except an authorized staff member for the purpose of the research for which the subject data were made available.

   a. Compliance with DNAnexus: Retain your uploaded data on DNAnexus, and considering using the DNAnexus feature to disable data deletion. All copies and processing are automatically tracked by the system. Only share project data with an authorized staff member for the purpose of the research for which the subject data were made available.

2. Requirement: Collaborating investigators from other institutions must complete an independent data use certification to gain access to the data.
   a. Compliance with DNAnexus: DNAnexus access controls allow you to verify that collaborating investigators from other institutions have completed an independent data use certification before you share data with them.

"When use of the dataset is complete—destroy all individually identifiable data"

1. Requirement: Shred hard copies.
   a. Compliance with DNAnexus: Storing data in DNAnexus makes hard copies unnecessary.
2. Requirement: Delete electronic files securely.
   a. Compliance with DNAnexus: Delete files or projects when completed with use. DNAnexus automatically deletes electronic files securely.
3. Requirement: At minimum, delete the files and then empty your recycle bin.
   a. Compliance with DNAnexus: All files deleted from DNAnexus are not recoverable, there is no recycle bin.
4. Requirement: Optimally, use a secure method, e.g., an electronic "shredder" program that performs a permanent delete and overwrite.
   a. Compliance with DNAnexus: Media that contained DNAnexus data are securely electronically "shredded" or physically destroyed when no longer used.

"Appendix A: CIS checklist for Linux Variants":

- DNAnexus processes data on secured Linux servers in a highly secure data center behind a strict firewall. The DNAnexus Linux configuration is as- or more-secure than the TCGA Linux Configuration best practices.

"TCGA User Certification Agreement: 6. Data Security and Data Release Reporting":

The TCGA User Certification Agreement requires some specific security and data reporting controls. These requirements include the above dbGaP Security Best Practices, and detail the following requirements. This is how DNAnexus facilitates compliance with these requirements:

- Requirement: All Approved Users have completed all required computer security training required by their institution, for example, the http://irtsectraining.nih.gov/, or the equivalent.

    a. Compliance with DNAnexus: Approved Users must complete computer security training required by their institution.

- Requirement: The data will always be physically secured (for example, through camera surveillance, locks on doors/computers, security guard).
    a. Compliance with DNAnexus: Data stored with DNAnexus are always physically secured in highly secure data centers with locks, guards, and surveillance.
- Requirement: Servers must not be accessible directly from the internet, (for example, they must be behind a firewall or not connected to a larger network) and unnecessary services should be disabled.
    a. Compliance with DNAnexus: DNAnexus servers are not accessible directly from the Internet, and are behind a stateful packet inspection firewall.
- Requirement: Use of portable media, e.g., on a CD, flash drive or laptop, is discouraged, but if necessary then they should be encrypted consistent with applicable law.
    a. Compliance with DNAnexus: Portable media and laptops are not needed for data stored on DNAnexus.
- Requirement: Use of updated anti-virus/anti-spyware software.
    a. Compliance with DNAnexus: Approved Users should have updated anti-virus and anti-spyware software on the machines they use to access DNAnexus.
- Requirement: Security auditing/intrusion detection software, detection and regular scans of potential data intrusions.
    a. Compliance with DNAnexus: DNAnexus performs regular scans, audits, and intrusion detection on its systems.
- Requirement: Use of strong password policies for file access.
    a. Compliance with DNAnexus: DNAnexus enforces a strong password policy.
- Requirement: All copies of the dataset should be destroyed, as permitted by law, whenever any of the following occurs:
    - the DUC expires and renewal is not sought;
    - access renewal is not granted;
    - the NCI/NHGRI TCGA Data Access Committee requests destruction of the dataset;
    - the continued use of the data would no longer be consistent with the DUC.
    a. Compliance with DNAnexus: Users are able to delete their DNAnexus projects and/or files when any of the above occur.