DNAnexus®

Security, Privacy & Quality for your Genomic Analysis

This white paper explores how the DNAnexus Platform enables researchers and clinicians to focus on their genomic analyses while complying with the ever-increasing regulations, laws and industry expectations for security, privacy and quality.

V3.3
06 MAY 2021

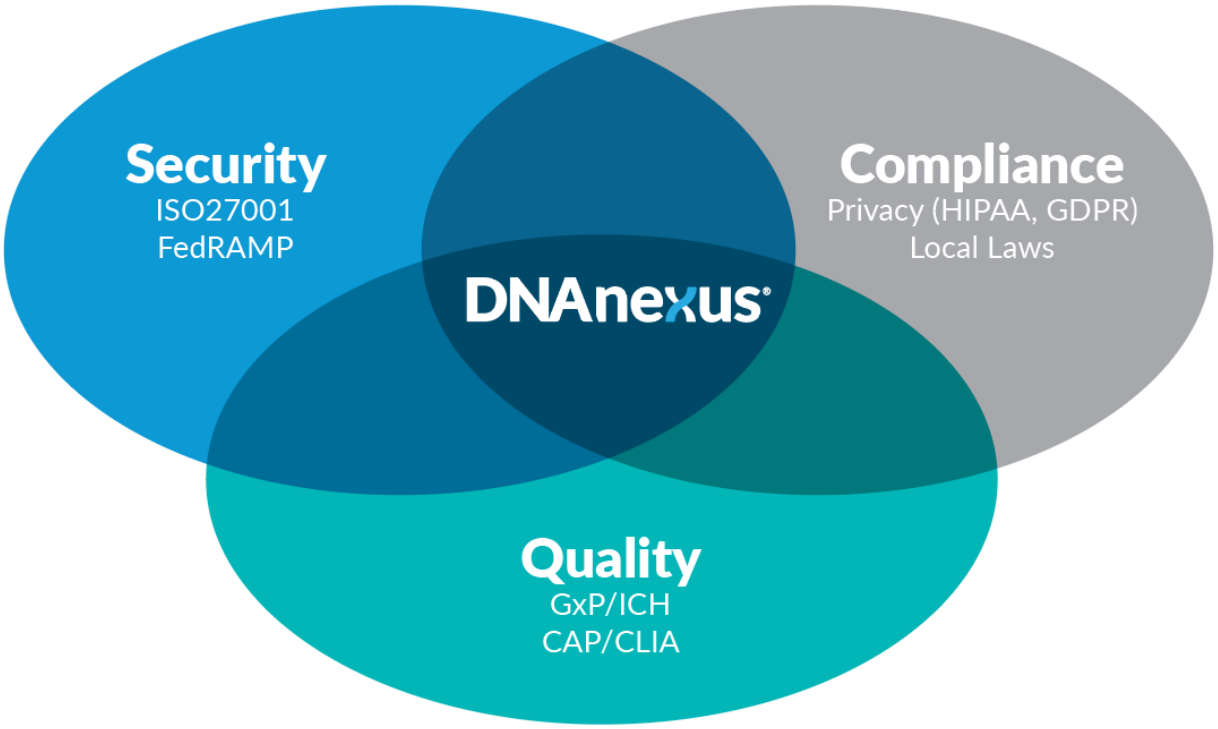
Contents

Overview

The **DNAnexus Platform** provides a consistent experience for the scientist to run their genomic analyses in a world with ever-increasing regulatory, security and industry expectations. This white paper examines three critical dimensions a global platform must have to provide value: security, privacy and quality. Each of these dimensions are interdependent and while their importance may vary depending on the use of the Platform, each is critical to sustain reproducible science.

DNAnexus combines these dimensions to provide a secure and collaborative environment to support both basic and clinical researchers as well as commercial development and production of pharmaceutical active ingredients and diagnostic tests that pass regulatory approval.

DNAnexus achieves compliance with current and future regulations, standards, and industry frameworks to understand the overlap of control points; the congruence of security, privacy, and quality design principles; and continuous monitoring of compliance to provide robust "guard rails" or forcing functions to keep scientists on a reliable Platform today and in the future.



When comparing the security, compliance and quality requirements of a "future-proof" system, there is a significant overlap of control points and design principles. Almost all depend upon robust authentication and authorization, track-and-trace functionality of people, objects, data, governance, and much more. DNAnexus' FedRAMP Moderate Authority to Operate (ATO) has over 1,600 control points with significant overlaps with the EU 2016/679 (GDPR), and ISO27001/2.

Entrusting DNAnexus with your key business processes and data as they relate to genomics cause many to pause. The purpose of this paper is to show how we manage your risk. First, we will examine the security foundation of the DNAnexus Platform and its alignment with ISO, and NIST frameworks. Then we will address the privacy of your information stored in this Platform. We will finish with an exploration of the third dimension, how the Platform provides the Quality (GxP) foundation necessary for gaining regulatory approval of new pharmaceutical active ingredients (APIs), new diagnostic assays, and in the case where genomic analysis is used to sustain regulated manufacturing, cGMP compliance. As you conclude this White Paper, you should gain an understanding of each dimension, and like the legs of a three-legged stool, understand how they are interdependent and vital for the success of your research.

The following table provides a snapshot of how the DNAnexus Platform aligns with common standards and regulations.

DNAnexus Compliance Reference

	DNAnexus	ISO 27001/2 NIST-800 FedRAMP	Privacy (GDPR)	HIPAA PIPEDA	GxP/ICH	CLIA
Security Architecture	X	X	X	X		
Access Control	X	X	X	X	X	
Consistency of Results	X				X	X
Auditability	X	X	X	X	X	X
Availability	X	X	X	X	X	X
Consent	X		X	X	X	X
Compliance and Assessment	X	X			X	X

DNAnexus Platform Security

Platform Security Architecture

To ensure data integrity and confidentiality, DNAnexus has implemented the following physical and logical security features throughout the fabric of the DNAnexus Platform:

- **Overall Security Framework.** DNAnexus uses the ISO 27002 international security standard and the NIST 800 family of regulations, as required by FedRAMP[1], to manage and monitor security in a predictable and consistent fashion. These frameworks are risk-based and cover people, process, and technology dimensions, with security control objectives that are extensible to the regulations where DNAnexus does business across the globe.
- **Security by Design.** Security is part of the fundamental design, implementation and operation of the Platform. This approach provides "guard rails" to focus users on their science while operating with forcing functions that cannot be overridden. This approach establishes a reliable operation of auditable security control and enables continuous and real-time auditing and technical scripting of the DNAnexus governance policies. This automated environment for security, assurance, governance and compliance provides a functional and reliable model enabling secure execution of science.
- **Cloud Architecture.** The DNAnexus Platform provides secure and audited access to genomic information via a web browser, without the necessity of downloading the information, which remains in the cloud. A recent report of the Presidential Commission for the Study of Bioethical Issues identifies computer architectures that provide "computational access" to query genomic information *without* giving the user possession of the information as a best-practice privacy protection. This approach simplifies the technical requirements on the scientists by placing the onus of security and privacy on the DNAnexus Platform.
- **Physical Security.** DNAnexus restricts confidential user data to high-security facilities with SAS 70/SSAE 16, PCI Level 1, SOC-2 compliant facilities.
- **Encryption in Transit and at Rest.** The DNAnexus Platform user data are encrypted when in transit (SSL/TLS), both over the Internet and internally in the cloud, and at rest – stored with AES 256 encryption or better in this homomorphic encryption approach. Operating on the principle of "Least Trust," this approach minimizes information exposure in the event of unauthorized access to systems, storage, or networks.

[1] The Federal Risk and Authorization Management Program (FedRAMP) is a U.S. Federal Government-wide program that proves a standardized approach to security assessment, authorization and continuous monitoring for cloud products and services. If you do business with the U.S. Federal Government, use of FedRAMP certified cloud service providers is a common expectation. See: https://www.fedramp.gov/

- **Monitoring.** DNAnexus conducts regular system scanning and continuous monitoring to track potential vulnerabilities and both actual and potential intrusions. This continuous monitoring of actionable information can replace static annual security audits.

Access Control

- **Authorization.** The DNAnexus Platform allows the administrator of a project to control and audit access to data, and to specify appropriate privilege levels, including Viewer, Uploader, Contributor, or Administrator roles. Project administrators may also specify "Copy Not Allowed" to prevent non-administrators from moving data to other projects or other geographic regions.
- **Authentication.** DNAnexus has implemented 2-factor authentication for DNAnexus administrative and user access, password complexity requirements, password change requirements, and session timeout features for customers to protect against unauthorized access to confidential user data.
- **Firewalls.** While the DNAnexus Platform infrastructure uses strict stateful network firewalls to protect all servers, including those processing confidential user data, it also embraces Zero-Trust, where every component within the Platform is independently secured and audited.
- **Provenance.** The DNAnexus Platform architecture provides controls to trace the source, consumption and information integrity of the data, with provisions to address incorrect, fabricated or cloned data before the information is shared among ecosystem partners.
- **Collaboration**. Much of the value of genomic research is via collaboration. This collaboration may be between healthcare providers and researchers, commercial, government and academic researchers. The DNAnexus Platform allows for selective collaboration via secure multi-party computations in an audited and transparent fashion.

Auditability

Inherent in any quality control system is the need to document the observation of policies and procedures. The DNAnexus Platform incorporates several automatic features that provide audit trails necessary to document compliance. These include the following:

- **Logging.** Access and changes to data are logged to a dedicated server, and logs are maintained for at least 6 years. All user uploads are logged and “hashed” to verify integrity. All data analyses are stamped with the date and time processed, along with the tool (including version) used to process them.
- **Automated Generation of Audit Evidence.** By implementing the processes from DNAnexus’ policies into automation on the platform, DNAnexus can provide real-time evidence of compliance.

- **Transparency and traceability of authorization.** Of particular importance for demonstrating governance In collaborative environments, the Platform supports consistent, informed and ongoing authorization decisions, reciprocity, and the transparency and traceability of security and privacy information with regard to data, algorithms, ownership and relevant system attributes. Examples of transparency include the withdraw of consent, as required by the "right to be forgotten" as specified by GDPR.
- **Human Readable Audit Trail.** As discussed elsewhere in this document, the DNAnexus Platform supports compliance with 21 CFR § 11, and as such, provides a human readable audit trail. All licensed users can turn on this audit trail. The audit trail is generated into the user's Project folder at the end of the day for all activity that transpired within the previous 24 hours in that org.
- **Records retention.** Customers have the ability to delete data and reports when no longer needed or when patient or donor consent is revoked. Customer data are stored until deleted by the customer, providing complete control over record retention and destruction. Project administrators can lock projects to prevent accidental deletion of any files by anyone other than the project administrators.

Availability

DNAnexus has also taken steps to provide users with confidence in the availability of their data:

- **Secure Facilities**. All user data are stored and processed in high-security data centers with backup power. Facilities have strict physical access controls.
- **Backups**. All user data are redundantly stored on multiple devices across multiple facilities of the DNAnexus cloud infrastructure to provide 99.99999% durability and 99.99% availability of objects over a given year, and designed to sustain the concurrent loss of data in two facilities. Project files cannot be modified and can only be deleted if permitted by the project administrator. *Note:* While DNAnexus takes responsibility for the backup and restoration of the system, public apps, user metadata, and logging information – the user is responsible for the backup and restoration of their own data.
- **Disaster Recovery and Incident Response Plans**. Consistent with ISO 27002 standards, disaster recovery and incident response plans are in place to ensure that if a disaster occurs, DNAnexus takes appropriate recovery steps and notifies stakeholders in a prompt and compliant manner.

Intellectual Property Protection

Theft or misuse of proprietary applications in genomic analysis has become an increasing threat. With the high degree of operational security of the Platform and all the security of all the building blocks during their Software Development Lifecycle (SDLC), we have extended the security scanning, digital signing, and audited change controls to the production of highly secure and qualified (GxP)

applications. Where users are concerned about theft of intellectual property, the DNAnexus Platform can be used to tightly control the invocation, execution, and output of an application running on any DNAnexus Platform instance anywhere in the world.

FedRAMP

The Federal Risk and Authorization Management Program (FedRAMP) is a U.S. government-wide program to provide a standardized approach to security assessment, authorization and continuous monitoring of cloud products and services. It is a frequent requirement of U.S. government sponsored work. The DNAnexus Platform operates in accordance with FedRAMP. The precisionFDA community platform for NGS assay evaluation and regulatory science exploration (https://precision.fda.gov/) is a portal that uses the DNAnexus Platform as the back end. FedRAMP is constructed around the applicable United States Codes (USC) covering security, personal data privacy, internal controls, and the interchange of information to upstream and downstream systems. FedRAMP is also framed around many of the NIST 800 family of standards and guidance. While FedRAMP is a mandatory requirement for government sponsored cloud service providers, it also provides a known standard to non-U.S. government agencies and non-government customers to evaluate the security level of the DNAnexus Platform.

In October of 2018, DNAnexus was awarded the “Moderate” Authority to Operate (ATO) by U.S. Health and Human Services. DNAnexus joins an elite group of approximately 125 companies in the world with this security rating and is the only industry or Federal cloud-based platform for biomedical informatics and data management. FedRAMP is based on internationally recognized standards such as NIST 800-53r4, FIPS 199, FIPS 200, OMB A-130 and other NIST Guidance documents, raising the bar beyond current standards such as annual SOC2 audits. FedRAMP certification provides a common language for security controls and processes. Given the continuous nature of these processes, DNAnexus has automated most of these processes to allow real time evidence for audits. The foundation provided by FedRAMP sets a foundation for current and future security, privacy and quality controls – “future proofing” our users’ investment in pursing secure science on DNAnexus.

Privacy

A person's DNA provides many of the instructions that define an individual. Protecting the privacy of individual's whole genome sequence data is a critical function of the DNAnexus Platform.

Privacy and security are interrelated, but they are not the same thing. Without security, you cannot have privacy. However, you *can* have security without privacy. Privacy controls are more granular, controlling who, what, and when information can be accessed. Privacy focuses on the following concepts:

- What data should be collected?
- What are the permissible uses of the data?
- With whom can the data be shared?
- How long should the data be retained?
- Can the subject of the data get a copy, request a correction, or have their data be erased?

As sequencing technology produces higher resolution reads on ever-increasing long reads of DNA (and RNA), an additional challenge of re-identifying the genomic donor via long variant matching becomes an issue, even for sequences stripped of the 18 PHI identifiers. Thus, DNAnexus helps users explore different schemes of using synthetic data to achieve anonymity as well as differential privacy for sharing results.

DNAnexus' Platform is configured to address the regulations of the jurisdictions where it operates. The range of privacy regulations spans from Regulation (EU) 2016/679, commonly known as GDPR , and GB/T 35273-2017, also known as the "PI National Standards" of the Republic of China, to industry-specific regulations focused on specific types of transactions (e.g. HIPAA, PIPEDA, PCI). Many of these regulations include data localization restrictions (e.g. data of citizens must stay in their country), a determination of "important" or "sensitive" information.

DNAnexus as your Data Processor

DNAnexus takes the position of being a "data processor" to our customers, who are data controllers. As a "data processor," DNAnexus fulfills the necessary regulatory requirements while facilitating the data controllers to achieve their compliance. As a data processor, DNAnexus shall:

1. Follow the instructions of the customer in the management of their data. DNAnexus shall not opportunistically mine or use personal data that it is entrusted, aside from those mentioned elsewhere in this document or under the written instructions of the customer.
2. Obtain written permission from the customer before engaging a subprocessor and assume liability for failures of the subprocessors to meet the requirements of GDPR.
3. Upon request, delete or return all personal data to the customer at the end of the service contract.
4. Enable and contribute to compliance audits conducted by the customer's controller or a competent representative of the controller.
5. Take reasonable steps to secure data, such as encryption, stability and uptime, backup and disaster recovery, and regular security testing.
6. Notify the customer without undue delay upon learning of data breaches.
7. Make every effort to restrict personal data transfer to a third country only if legal safeguards are obtained.
8. Make the Data Protection Officer available to address concerns or questions upon request.

More detail on DNAnexus' Data Protection Addendum (DPA) may be obtained here. The Data Protection Officer of DNAnexus is a senior executive available to work with customers to help maintain their data protection.

Consent

Under DNAnexus Privacy Policy, users are responsible for ensuring that the patients or donors of samples from which genomic information is generated have provided informed consent in accordance with how information from the samples will be used.

Compliance and Assessment

- **Internal Review**. DNAnexus follows a rigorous quarterly internal review of security controls. In addition, a formal annual review is conducted on the entire security management system, security policies, and security and privacy risks.
- **Third-Party Assessments.** DNAnexus uses third-party objective expert services for regular security vulnerability scanning and for full network and application penetration tests. These assessments approach the DNAnexus Platform as an attacker would and attempt to find any security vulnerabilities that could be exploited. DNAnexus promptly addresses any issues uncovered in these assessments.
- **External Audits**. DNAnexus has achieved ISO 27001 certification by independent third parties and maintains this compliance with annual on-site audits. DNAnexus also engaged an independent third-party to review compliance with Regulation (EU) 2016/679 (GDPR) and maintains this compliance with annual audits. It is DNAnexus' position that these audits provide reasonable confidence that DNAnexus complies with GB/T 35273-2017 (PI National Standards) of the People's Republic of China as well as many other privacy regulations.
- **EU-US and Swiss-US Privacy Shield.** DNAnexus participates in the Privacy Shield Frameworks as designed by the U.S. Department of Commerce, the European Commission and the Swiss Administration to provide a mechanism to comply with data protection requirements when transferring personal data from the EU and Switzerland. Should the UK split from the EU, DNAnexus will also acquire the UK-US Privacy Shield. See: https://www.privacyshield.gov/

HIPAA Security and Privacy Rules

Below is an overview of the Security and Privacy Rules issued by the U.S. Department of Health and Human Resources ("HHS") under the Health Insurance Portability and Accountability Act of 1996 ("HIPAA" or 45 CFR § 160) and the Health Information Technology for Economic and Clinical Health Act of 2009 ("HITECH").

Who is subject to the HIPAA Privacy Rule?

The Privacy Rule applies to health plans, healthcare clearinghouses, and any health care provider that transmits health information in electronic form in connection with transactions regulated by HHS ("Covered Entities"). 45 CFR §160.102 and 160.103.

Under HITECH, the Privacy Rule obligations extend to "Business Associates", which generally refers to contractors to whom a covered entity delegates some or all of its Privacy Rule obligations.

What Information is Subject to the Privacy Rule?

The Privacy Rule protects "individually identifiable health information," which it calls "Protected Health Information" or "PHI." 45 CFR § 160.103.

The Privacy Rule defines "PHI" as information relating to:

- An individual's past, present, or future physical or mental health condition,
- The provision of health care to an individual, or
- The past, present or future payment for the provision of health care to the individual, **if any such information identifies the individual or if there is a reasonable basis to believe that the information can be used to identify the individual.** 45 CFR § 160.103.

As a corollary, there are no restrictions on the use or disclosure of de-identified health information. 45 CFR §§ 164.502(d)(2), 164.514(a) and (b). The Privacy Rule provides a "safe harbor" method of de-identification, which requires removal of 18 specified identifiers, such as name, address, and dates relating directly to an individual (e.g., birth date), social security number, and the like. 45 CFR § 164.514(b).

NOTE: A researcher who has no clinical relationship with a tissue donor and who only has access to de-identified tissue samples or genomic information is not subject to the Privacy Rule.

Privacy of De-Identifying Genomic Data

At this time, it is unclear whether an individual can be identified solely on the basis of their whole genome sequence (WGS). In 2008, David Craig, a computational biologist at the Translational Genomics Research Institute of Phoenix AZ was able to take someone's DNA sample and compare it with the summary statistics of an NIH study, sorting through the data until he found a match. In the following decade, other researchers such as Yaniv Erlich[2,3] have raised serious questions about the

[2] See: https://www.nature.com/news/privacy-protections-the-genome-hacker-1.12940

[3] See: https://www.nature.com/news/privacy-loophole-found-in-genetic-databases-1.12237#ref-link-1

vulnerabilities of public databases containing genetic data. It is DNAnexus' position that this information needs to be secured with strong privacy controls – as provided in the DNAnexus Platform.

Quality

As customers integrate the DNAnexus Platform into their regulated processes for clinical trials of pharmaceutical Active Pharmaceutical Ingredients (API), medical diagnostic devices and ongoing manufacturing of pharmaceutical APIs (e.g. microbiome-oriented APIs), DNAnexus' Quality Management Systems (QMS) must integrate with the QMS of our customers. DNAnexus maintains an accredited QMS to comply with the U.S. 21 CFR family of regulations and quality guidelines of the International Council of Harmonization (ICH).

Quality Principles

The DNAnexus QMS applies to the development and operation of the DNAnexus products, services, infrastructure and assets that support the development and operations of the DNAnexus Platform. The DNAnexus QMS follows the seven principles:

- *QMP1 – Customer focus*: meet customer requirements and strive to exceed customer expectations.
- *QMP2 – Leadership*: leaders at all levels establish unity of purpose, direction and create conditions in which people are engaged in achieving the organization's objectives.
- *QMP3 – Engagement of people*: Competent, empowered and engaged people at all levels throughout the organization are essential to enhance its capacity to create and deliver value.
- *QMP4 – Process approach*: Consistent and predictable results are achieved more effectively and efficiently when activities are understood and managed as interrelated processes that function as a coherent system.
- *QMP5 – Improvement*: Successful organizations have an ongoing focus on improvement.
- *QMP6 – Evidence-based decision making*: Decisions based on the analysis and evaluation of data and information are more likely to produce the desired results.
- *QMP7 – Relationship management:* For sustained success, an organization manages its relationship with interested parties, such as its suppliers and customers.

Quality Details

In addition, the DNAnexus Management Team is committed to providing customers in the pharmaceutical, government, academic, laboratory, and clinical research sectors a secure, high-quality software platform for storing and analyzing genome sequence data by:

1. Designing and developing a secure, state-of-the-art platform that responds to evolving customer needs,
2. Testing all changes in a staging environment before they are released for production,

3. Striving for error-free platform updates, and
4. Tracking and resolving any needed corrections through a company-wide ethos of continual improvement.

This commitment to developing, testing, and continually improving the DNAnexus Platform is reflected in the company's four quality objectives, described in more detail below. Specific quality objectives related to the quality policy include:

- **Pipeline results are consistent and reproducible**. Pipelines can be preconfigured to preset parameters, ensuring consistent analysis of samples. All data, tools and workflows are tracked and version-controlled to ensure auditability and reproducibility. Because the DNAnexus Platform automatically logs the tool version used to process data, the consistency of results is never compromised by inadvertent use of differing versions of an analysis tool. Runtime consistency enables all users to work within a common operating environment; job runs are consistent regardless of platform updates.
  Metrics: All failed regression tests are documented and approved before production release.

- **Customer production data are secure and available.** DNAnexus restricts user data to high-security facilities. All data are encrypted during transit and while in storage. Controls restrict users and organizations with whom data can be shared or copied. Unauthorized access is prevented through two-factor authentication, password complexity and change requirements, and session timeout features. Disaster recovery and incident response plans are in place.
  Metrics: If an interruption to the DNAnexus customer-facing services occurs, the Recovery Time Objective (RTO) is 8 hours. The Recovery Point Objective (RPO) for customer data is 4 hours. The availability objective for the DNAnexus Platform is 99.9% scheduled uptime, with less than 1% unscheduled downtime.

- **Customer production data integrity is maintained at every step**. All uploads are logged and "hashed" to verify integrity; audit logs are maintained for at least six (6) years. DNAnexus users have complete control over data retention (and destruction); project administrators can lock projects to prevent accidental deletion of any files. All data uploaded to DNAnexus are validated using a checksum to ensure the uploaded data are consistent with the source data; customers can validate the integrity of data downloaded from the platform in a similar manner.
  Metrics: No reported integrity failures, as recorded via customer support tickets.

- **A continual improvement process is in place to detect issues and control changes.** The QMS uses feedback and tests to detect issues or problems, identify and understand root causes, and develop solutions, which are applied throughout the organization and monitored for their success. This process ensures that any changes are controlled, documented, and support the overall quality and performance objectives.
  Metrics: Qualitative feedback indicates that customers are satisfied with how the DNAnexus Platform performs and how the DNAnexus team responds when customers require changes.

Consistency of Results

To ensure the consistency of results, DNAnexus has implemented a number of features in the DNAnexus Platform to support these requirements:

- **Preconfigured Pipelines**. The DNAnexus Platform allows lab bioinformatics specialists to configure pipelines which chain together a set of analysis tools and datasets, and also allows for the use of preset parameters, thereby ensuring consistent analysis of patient samples. These pipelines can be packaged as separate apps for use by more basic users who can "point and click" to run their analyses and generate reports.
- **Version Control**. The DNAnexus Platform automatically logs the tool version used to process data, allowing labs to ensure that the consistency of results is not compromised by inadvertent use of differing versions of an analysis tool.
- **Runtime Consistency**. The DNAnexus Platform provides a consistent runtime environment and provides users with the ability to incorporate additional runtime resources into their applications. The applications consistently deploy the specified runtime environment when run. Tools and data can be shared with other users without encountering runtime environment inconsistencies.

Good Clinical and Laboratory Practices (cGCP, cGLP, and cGMP)

DNAnexus enables compliance with the requirements of current Good Clinical Practices ("cGCP"), current Good Laboratory Practices ("cGLP"), current Good Manufacturing Practices ("cGMP") and since the DNAnexus Platform uses electronic records, 21 CFR § 11 by those who use and submit genomic data to the U.S. Food and Drug Administration ("FDA") and comparable regulatory organizations outside the United States.

What are cGCP, cGLP, cGMP and 21 CFR § 11? Who is Subject to them?

cGCP, cGLP, cGMP and 21 CFR § 11 all apply to data submitted to the FDA.

- cGCPs are regulations and guidelines that are intended to ensure data quality and protect human subjects. cGCPs set minimum standards for the conduct of clinical trials involving human subjects to test the safety and efficacy of drugs, diagnostics and medical devices. They consist of an international set of principles, adherence to which is "universally recognized as a critical requirement to the conduct of research involving human subjects."[4]

[4] See http://www.fda.gov/ScienceResearch/SpecialTopics/RunningClinicalTrials/default.htm

- cGLPs are practices prescribed by FDA regulations and apply to nonclinical laboratory studies in support of applications for research or marketing for FDA-regulated products.[5] Such nonclinical laboratory studies are performed in laboratory conditions in order to determine the safety of test articles and do not include clinical trials utilizing human subjects or animal field trials.[6]
- The requirements of 21 CCFR § 11 set forth the criteria by which the FDA determines the equivalence of records in electronic form to paper records and the FDA's acceptance of electronic records in lieu of paper records.[7] Anyone who submits data processed or stored electronically to the FDA must comply with these regulations, including, without limitation:
    - Clinical trial sponsors,
    - Clinical research organizations ("CROs") conducting trials on sponsors' behalf, and
    - Laboratories hired by sponsors to perform pre-clinical studies under GLP for submission to the FDA.

Computer System Validation - Who Does What?

21 CFR § 11 requires the "system" undergo and be maintained in a validated state. For purposes of validation, DNAnexus treats its offering as two separate elements:

- The base Platform and the applications
- Pipelines that run on the Platform

DNAnexus develops and maintains the Computer System Validation of the Platform and provides this validation documentation to customers upon request. As the Platform is updated on a weekly basis (i.e. ~50 releases per year), DNAnexus releases GxP Release Notes and Pre-Release Notes on a weekly basis. This allows our customer's Quality Assurance and Computer System Validation (CSV) teams to understand the evolution of the Platform in a timely fashion. Our customers treat the underlying Platform as a Commercial Off The Shelf (COTS) product.

For applications and pipelines supported by DNAnexus, the company provides qualification documents to the customer's CSV team to complete their validation activities (e.g. performance qualification tests). For customers who develop their own products, DNAnexus can offer guidance on how to achieve sustainable validation of their applications and pipelines, but these activities are owned by the customer.

---

[5] See the cGLP regulations at 21 CFR § 58. FDA questions and answers regarding cGLPs are available at http://www.fda.gov/ICECI/EnforcementActions/BioresearchMonitoring/NonclinicalLaboratoriesInspectedunderGoodLaboratoryPractices/ucm072738.htm.

[6] 21 CFR § 58.3(d).

[7] 21 C.F.R. § 11 generally applies to records in electronic form that are submitted to the FDA as required by agency regulations or under the Federal Food, Drug and Cosmetic Act or the Public Health Service Act, but not including paper records submitted by electronic methods. (21 C.F.R. § 11.1(b)).

What is the Difference between Validation and Qualification?

Validation

Validation is the process of "establishing documented evidence" that provides a high degree of assurance that a specific process will consistently produce a product or system that meets predetermined specifications (URS, FS, Design Spec) and quality attributes. Simply stated, it is the act, process or instance to substantiate something on a reliable basis. It is the process to prove the system matches the requirements.

Qualification

Qualification is the process to establish confidence that a process or system will work correctly and consistently operate *within established limits and tolerances*. Qualifications tend to be smaller in scope, more static in nature, and are usually subsets for a greater validation initiative. Qualification allows for tests to be performed on one element or component of the process to be validated against a specified outcome. Simply stated, it is the act or process to assure that something complies with an expected outcome, standard or set of specific requirements.

How do Qualification and Validation Work Together?

Validation incorporates the concept of qualification. Qualification is used to assess something dynamic that is likely to change (e.g., you qualify a computer network -- how the packets are routed will change, but the process of communication happens). Validation is larger in scope and can incorporate various qualifications to achieve the end result. In both qualification and verification, you still must establish evidence against a predetermined criterion. Validation criteria tend to be more exacting than qualification, where the latter tolerates variation within the *range* of acceptable criteria.

Clinical Laboratory Improvement Amendments of 1988 (CLIA)

Who is subject to CLIA?

Congress passed CLIA in 1988 to establish quality standards for all laboratory testing to ensure the validity and reliability and timeliness of laboratory examinations and procedures, handling of specimens, and reporting of results.[8] For purposes of CLIA, a "laboratory" is any facility that performs laboratory testing on specimens derived from humans for the diagnosis, prevention, or treatment of disease or impairment or assessment of health in humans.[9]

How do CLIA standards apply to clinical labs' management and analysis of next-generation genome sequencing ("NGS") data?

CLIA requires the "consistent performance" by laboratories of "valid and reliable laboratory examinations and other procedures."[10] CLIA requirements include, without limitation: maintenance of quality assurance and quality control programs to ensure the validity and reliability of the lab's examination and procedures and the proper handling of specimens and reporting of results; maintenance of records, equipment, and facilities necessary for the proper and effective operation of the laboratory; qualification under a proficiency testing program meeting applicable standards; and assurance of the adequacy and competency of staff.[11]

With regard to patient information and reports, CLIA regulations provide the following overall standard:

> The laboratory must have an adequate manual or electronic system(s) in place to ensure test results and other patient-specific data are accurately and reliably sent from the point of data entry (whether interfaced or entered manually) to final report destination, in a timely manner.[12]

CLIA regulations also restrict the disclosure of patient data:

> The laboratory must ensure confidentiality of patient information throughout all phases of the total testing processes that are under the laboratory's control.[13]

---

8 42 U.S.C. §263a (f).
9 42 U.S.C. §263a (a); 42 CFR §493.2.
10 42 U.S.C. §263a (f).
11 Id.
12 42 C.F.R. § 493.1291(a).
13 42 C.F.R. §§ 493.1231. See also 42 C.F.R. § 493.1291(f) ("Test results must be released only to authorized persons and, if applicable, the individual responsible for using the test results and the laboratory that initially requested the test.")

The CLIA standard requiring labs to consistently perform valid and reliable testing has many implications for laboratory information systems that track data resulting from analysis of patient samples, which are summarized below.

The general standards for the management and analysis of data from patient samples can be found in the applicable regulations.[14] Additional detail is available in guidelines and checklists published by private organizations, notably the College of American Pathologists ("CAP"), which is a CLIA accrediting body approved the Centers for Medicare & Medicaid Service.[15] In addition, the US Centers for Disease Control has convened a working group on Next-generation Sequencing: Standardization of Clinical Testing, which has developed guidelines, some of which address validation of informatics pipelines used by clinical labs to analyze genomic information.[16] Many of these requirements are directed at ensuring the integrity of data generated, the consistency of analytical methods used, and their auditability and availability.

How do CLIA standards apply to DNAnexus?

The DNAnexus Platform lies substantially outside of the boundaries of CLIA regulation. DNAnexus does not receive patient care reports, does not directly generate patient care reports, does not interpret data received from partner/client healthcare providers, nor does it provide direct-to-consumer testing nor reporting.

DNAnexus does manipulate raw data, in that it provides a platform for the analysis of genomic data. As such DNAnexus is required to demonstrate the integrity of data at the interfaces to the DNAnexus Platform. The DNAnexus Upload Agent performs a checksum comparison of the uploaded data to ensure integrity with the source data. Data egressing from DNAnexus is also evaluated by checksum analysis. The data consumer is responsible for verifying that a locally calculated checksum of the downloaded data matches the checksum provided by DNAnexus.

All data uploaded to DNAnexus is immutable, analysis tools are version controlled, and the platform maintains detailed logs describing every analysis performed. These features provide the ability to demonstrate reproducibility and to track the provenance of analysis results, which simplifies the process of adhering to CLIA standards on the part of DNAnexus customers.

---

[14] See 42 C.F.R. § 493.1230 et seq.

[15] See, for example, "Laboratory General Checklist" College of American Pathologists (Jan. 4, 2012) ("CAP Lab General Checklist").

[16] See Gargis et al., "Assuring the quality of next-generation sequencing in clinical laboratory practice" 11 Nature Biotechnology 11 at p. 1033 (November 2012).

NCBI Database of Genotypes and Phenotypes (dbGaP) Security Best-Practices

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, are potentially subject to "controlled access."

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."[17] dbGaP datasets are organized into two tiers: Open Access and Controlled Access data.[18]

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[19]

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.[20]

Consistent with this approach, the application for access to controlled data requires that investigators agree to adhere to specified security best practices.[21] To obtain access to TCGA, an investigator must similarly agree to a Data Use Certification Agreement, which includes a provision requiring compliance with dbGaP security best practices.[22]

For a detailed analysis of the official dbGaP best practices requirements, and how the DNAnexus Platform supports compliance of each item, see "*Appendix A: dbGaP Security Best-practices Checklist*."

---

[17] http://www.ncbi.nlm.nih.gov/gap
[18] http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/about.html
[19] http://www.ncbi.nlm.nih.gove/projects/gap/cgi-bin/study.cgi?study_id=phs000178.v7.p6
[20] https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login
[21] Id.
[22] https://tcga-data.nci.nih.gov/tcga/tcgaAccessTiers.jsp

ISO/IEC 27001:2013 Certification

DNAnexus maintains independent certification of its compliance with the ISO 27001 management standard applicable to "Information Security Management Systems" for the DNAnexus Platform. This certification demonstrates the depth of DNAnexus' commitment to protecting the security of DNAnexus users' genomic information processed and stored in the Platform.

What are ISO 27001 and 27002?

ISO 27001 was established by the International Organization for Standards and the International Electrotechnical Commission. It is an internationally recognized management standard that describes best-practices for an "Information Security Management System (ISMS)."[23]

The ISO 27001 standard provides a model for "establishing, implementing, operating, monitoring, reviewing, maintaining and improving" an ISMS.[24] To obtain this certification, DNAnexus has had to demonstrate to an independent Accredited Registrar that it has:

- Systematically evaluated the risks to the security of its information systems, including their impacts on DNAnexus and its users;
- Designed and implemented a comprehensive set of security controls to address those risks; and
- Adopted management processes for planning, implementing, monitoring and improving those controls.

ISO 27001 is a management standard that is supported by specific control objectives and definitions defined in ISO 27002. Compliance with ISO 27001 requires a holistic approach to security that begins with management's identification of information security as a key strategic imperative, a thoughtful assessment of the risks associated with inadequate security, and a disciplined approach to the development, implementation and evaluation of controls designed to provide the desired protections. Certification of our compliance with this standard represents a milestone in the commitment of DNAnexus to the security of its users' data.

What does ISO 27001 certification mean for the security of information that DNAnexus stores and processes for its users?

To obtain ISO 27001 certification, DNAnexus has implemented a comprehensive set of security controls to protect its users' information. These controls are modeled after the best-practices list of controls found in ISO 27002, which is the most common standard implemented by organizations adopting the

[23] See www.itgovernance.co.uk.iso27001.aspx
[24] See www.27000.org/iso-27001.htm

ISO 27001 ISMS approach to security management. ISO 27002 details best practices for information security.[25]

Like most security systems, the ISO 27002 controls are informed by the key goals of any best-practices information security system:

- ***Confidentiality*** (ensuring that information is accessible only to those who need to use it);
- ***Integrity*** (safeguarding the accuracy of information and the methods used to process it); and
- ***Availability*** (ensuring that authorized users have prompt access to the information when they need it).

Given the numerous regulatory requirements applicable to genomic information depending on the context (including HIPAA, CLIA, GCP, GLP, and European Data Privacy regulations), the security control framework provided by ISO 27002, managed by the ISMS in accordance with ISO 27001, provides a comprehensive framework by which DNAnexus has developed its Platform, enabling DNAnexus users to confidently entrust their information to the Platform.

[25] See http://www.standardsconsultants.com/iso-27001-v-iso-27002

Appendix A: dbGaP Security Best Practices Checklist

The following is a detailed analysis of the official dbGaP Best practices Requirements, available from http://www.ncbi.nlm.nih.gov/, and how the DNAnexus Platform supports compliance of each item.

"Think Electronic Security"

1 Requirement: "Download data to a secure computer or server and not to unsecured network drives or servers"

Compliance with DNAnexus: DNAnexus provides secure servers for downloaded and stored data.

2 Requirement: "Make sure these files are never exposed to the Internet"

Compliance with DNAnexus: Data stored with DNAnexus is not exposed to the internet by default. It can be shared selectively and securely using DNAnexus sharing controls, rather than posting, sharing via FTP, or emailing insecurely.

3 Requirement: "Have a strong password for file access and never share it."

Compliance with DNAnexus: DNAnexus enforces strong passwords, including length and character variety.

4 Requirement: "If you leave your office, close out of data files or lock your computer."

Compliance with DNAnexus: DNAnexus automatically locks sessions after 15 minutes of inactivity.

5 Requirement: "Data stored on laptops must be encrypted."

Compliance with DNAnexus: Data stored with DNAnexus does not require downloading to laptops for processing. Downloading of data can be prohibited while still allowing results to be generated and viewed.

"Think Physical Security"

1 Requirement: If the data are in hard copy or reside on portable media, treat it as though it were cash.

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 data unattended or in an unlocked room.

Compliance with DNAnexus: DNAnexus data are stored in highly secure data centers.

3 Requirement: Consider locking data up.

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

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.

Compliance with DNAnexus: All data uploaded to DNAnexus is by default inaccessible from the Internet. 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.

Compliance with DNAnexus: DNAnexus applies all relevant security patches within 30 days

3 Requirement: dbGaP data on the systems must be secured from other and if exported via file sharing, ensure limited access to remote systems.

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.

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.

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 are 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 are cached on local systems, directory protection must be kept, and data must be removed when processing is complete.

Compliance with DNAnexus: After 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.

Compliance with DNAnexus: Retain your uploaded data on DNAnexus and consider 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.

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.

Compliance with DNAnexus: Storing data in DNAnexus makes hard copies unnecessary.

2 Requirement: Delete electronic files securely.

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.

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.

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

  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)

  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.

  Compliance with DNAnexus: DNAnexus servers that store protected data 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.

  Compliance with DNAnexus: Portable media and laptops are not needed for data stored on DNAnexus.

- Requirement: Use of updated anti-virus/anti-spyware software.

  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.

  Compliance with DNAnexus: DNAnexus performs regular scans, audits, and intrusion detection on its systems.

- Requirement: Use of strong password policies for file access.

  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.

Compliance with DNAnexus: Users are able to delete their DNAnexus projects and/or files when any of the above occurs.