US Department of Health and Human Services

Privacy Impact Assessment

Date Signed:

06/30/2025

OPDIV:

NIH

Name:

NCI Genomic Data Commons

PIA Unique Identifier:

P-5098014-753392

The subject of this PIA is which of the following?

Minor Application (stand-alone)

Identify the Enterprise Performance Lifecycle Phase of the system.

Operations and Maintenance

Is this a FISMA-Reportable system?

No

Does the system include a Website or online application available to and for the use of the general public?

Yes

Identify the operator.

Contractor

Is this a new or existing system?

New

Does the system have Security Authorization (SA)?

Yes

Indicate the following reason(s) for updating this PIA.

Describe the purpose of the system.

National Cancer Institute’s (NCI’s) Genomic Data Commons (GDC) is a next generation cancer knowledge network that supports the hosting and standardization of genomic and clinical data from cancer research programs, the harmonization of raw sequence data, and the application of state-of-the art methods for generating high level data (e.g. mutation calls, structural variants, etc.). The NCI Office of Cancer Genomics (OCG) of the National Institutes of Health (NIH) established the GDC to provide the cancer research community with a data service supporting the receipt, quality control, integration, storage, and redistribution of standardized cancer genomic data sets derived from cancer studies. The purpose of the GDC is to provide the cancer research community with a unified data repository that enables data sharing (submission of data and access to data) across cancer genomic studies in support of precision medicine.

Describe the type of information the system will collect, maintain (store), or share.

The GDC system provides aggregated cancer genomics data in a uniform and co-localized database and to serve as a foundation for future expanded data access, computational capabilities and bio-

informatics cloud research. IThe GDC stores eRA (not an acronym) Commons ID for logged in users and Internet Protocol addresses for every user to track usage and prevent denial of service (DoS) attacks. eRA Commons IDs are imported from eRA. eRA maintains its own Privacy Impact Assessment (PIA), including all legal authorities documented.

Users requiring access to this system log in using the NIH Identity, Credential, and Access Management (IAM) Services, with validation with eRA (eRA Commons ID). Both of which maintain their own unique PIAs on record, with all legal authorities documented. The purpose of IAM Services is to authenticate all users and computers in a Windows domain type network assigning and enforcing information security policies for all computers and installing or updating software. The IAM Services collects unique user credentials from eRA and stores them in an encrypted format. The IAM Service is an essential service which facilitates and governs network access to various resources. The GDC uses the eRA Commons ID to then authorize user access to studies with controlled access requirements with the database of genotypes and phenotypes (dbGaP). dbGaP also maintains its own PIA on record, with all legal authorities documented.

Provide an overview of the system and describe the information it will collect, maintain (store), or share, either permanently or temporarily.

National Cancer Institute's (NCI's) Genomic Data Commons (GDC) is a next generation cancer knowledge network that supports the hosting and standardization of genomic and clinical data from cancer research programs, the harmonization of raw sequence data, and the application of state-of-the art methods for generating high level data (e.g. mutation calls, structural variants, etc.). The NCI OCG of the National Institutes of Health (NIH) established the GDC to provide the cancer research community with a data service supporting the receipt, quality control, integration, storage, and redistribution of standardized cancer genomic data sets derived from cancer studies. The purpose of the GDC is to provide the cancer research community with a unified data repository that enables data sharing across cancer genomic studies in support of precision medicine.

The GDC system provides aggregated cancer genomics data in a uniform and co-localized database and to serve as a foundation for future expanded data access, computational capabilities and bio-informatics cloud research. IThe GDC stores eRA (not an acronym) Commons ID for logged in users and Internet Protocol addresses for every user to track usage and prevent denial of service (DoS) attacks. eRA Commons IDs are imported from eRA. eRA maintains its own Privacy Impact Assessment (PIA), including all legal authorities documented.

Users requiring access to this system log in using the NIH Identity, Credential, and Access Management (IAM) Services, with validation with eRA (eRA Commons ID). Both of which maintain their own unique PIAs on record, with all legal authorities documented. The purpose of IAM Services is to authenticate all users and computers in a Windows domain type network assigning and enforcing information security policies for all computers and installing or updating software. The IAM Services collects unique user credentials from eRA and stores them in an encrypted format. The IAM Service is an essential service which facilitates and governs network access to various resources. The GDC uses the eRA Commons ID to then authorize user access to studies with controlled access requirements with the database of genotypes and phenotypes (dbGaP). dbGaP also maintains its own PIA on record, with all legal authorities documented.

Does the system collect, maintain, use or share PII?

Yes

Indicate the type of PII that the system will collect or maintain.

eRA Commons ID
IP Addresses

Indicate the categories of individuals about whom PII is collected, maintained or shared.

Employees
Public Citizens
Business Partner/Contacts (Federal/state/local agencies)
Vendor/Suppliers/Contractors

How many individuals' PII is in the system?

100,000-999,999

For what primary purpose is the PII used?

eRA Commons ID and IP Address are used to protect the system from a DoS attack (malicious or non-malicious).

Describe the secondary uses for which the PII will be used.

Detect any security or system related issues and monitor performance/user metrics.

Identify legal authorities governing information use and disclosure specific to the system and program.

42 U.S. Code § 241
42 U.S. Code § 282
42 U.S. Code § 284
42 U.S. Code § 285a

Are records on the system retrieved by one or more PII data elements?

Yes

Identify the number and title of the Privacy Act System of Records Notice (SORN) that is being use to cover the system or identify if a SORN is being developed.

09-90-1401, Records About Requesters of Restricted Datasets

09-90-0777, Facility and Resource Access Control Records

Identify the sources of PII in the system.

Directly from an individual about whom the information pertains
Other
Government Sources
Within the DOI
Non-Government Sources
Public
Private Sector

Identify the OMB information collection approval number and expiration date

0925-0752, expiration date: 4.30.2026

Is the PII shared with other organizations?

Yes

Identify with whom the PII is shared or disclosed and for what purpose.

Describe any agreements in place that authorizes the information sharing or disclosure.

The GDC maintains contracts with Leidos Biomedical Research, Inc., and the University of Chicago as third-party vendors that maintain the system.

Describe the procedures for accounting for disclosures.

For data that is disclosed as part of a request under a routine use, GDC administrators maintain a accounting of those disclosures.

Describe the process in place to notify individuals that their personal information will be collected. If no prior notice is given, explain the reason.

The GDC Data Portal provide a popup that a user must acknowledge when first accessing the portal. The GDC Web Site also provides links to the Privacy Policy and the Privacy Act notification statement.

Is the submission of PII by individuals voluntary or mandatory?

Voluntary

Describe the method for individuals to opt-out of the collection or use of their PII. If there is no option to object to the information collection, provide a reason.

For access to the GDC for controlled data, there is no opt-out option. Users may decline to log into the system if they wish. However, in doing so, they will not be able to submit data or access the system's controlled information. Users will only be able to access the publicly available information.

Process to notify and obtain consent from individuals whose PII is in the system when major changes occur to the system.

When major system changes occur that involve PII, users are notified via a News Note on the GDC Web Site or announcement to the GDC Listserv and are provided with links to Release Notes, when applicable. If privacy related changes are made, the GDC would update the GDC Data Portal message that users must acknowledge and provide additional details in the GDC Web Site privacy statement.

Describe the process in place to resolve an individual's concerns when they believe their PII has been inappropriately obtained, used, or disclosed, or that the PII is inaccurate.

If a user believes their PII is inappropriately obtained, used, or disclosed, the user can contact the GDC Help Desk for further action.

Describe the process in place for periodic reviews of PII contained in the system to ensure the data's integrity, availability, accuracy and relevancy.

There are no processes in place for periodic reviews of PII within the system. the eRA Commons information is maintained by eRA. eRA has their own PIA, including documented processes for periodic reviews. For IP addresses, they are only reviewed to resolve system issues, such as DoS attacks.

Identify who will have access to the PII in the system and the reason why they require access.

Describe the procedures in place to determine which system users (administrators, developers, contractors, etc.) may access PII.

Determinations are made based on role-based access controls and least privilege. User rights are

provisioned based on controls within the system, allowing users only access to the minimum amount of PII necessary to perform their job.

Describe the methods in place to allow those with access to PII to only access the minimum amount of information necessary to perform their job.

Determinations are made based on role-based access controls and least privilege. User rights are provisioned based on controls within the system, allowing users only access to the minimum amount of PII necessary to perform their job.

Identify training and awareness provided to personnel (system owners, managers, operators, contractors and/or program managers) using the system to make them aware of their responsibilities for protecting the information being collected and maintained.

According to NIH policy, all personnel who manage or operate NIH applications must successfully complete annual security awareness training. Training is completed on the http://irtsectraining.nih.gov site with valid NIH credentials.

Describe training system users receive (above and beyond general security and privacy awareness training).

Users must review and agree to the HHS Rules Of Behavior before accessing the GDC.

Do contracts include Federal Acquisition Regulation and other appropriate clauses ensuring adherence to privacy provisions and practices?

Yes

Describe the process and guidelines in place with regard to the retention and destruction of PII.

07-203, System access records. Systems not requiring special accountability for access. These records are created as part of the user identification and authorization process to gain access to systems. Records are used to monitor inappropriate systems access by users. Destroy when business use ceases (DAA-GRS-2013-0006-0003).

001-003, Records of All Other Intramural Research Projects. Intramural research records related to planning, development, oversight and execution of biomedical research projects and programs performed by NIH research staff, contractors or under collaborative research and development agreements (CRADAs). These records span the project life-cycle. Cut off annually at termination of project/program or when no longer needed for scientific reference, whichever is longer. Destroy 7 years after cutoff (DAA-0443-2012-0007-0003).

Describe, briefly but with specificity, how the PII will be secured in the system using administrative, technical, and physical controls.

Administrative Controls: All GDC personnel are required to sign and acknowledge a user access agreement and complete annual user training before accessing the GDC. Administrative Controls: Access is limited by job role with approval by system owner or designated representative. Temporary access granted to troubleshoot specific issue such as a data integrity concerns. All personnel have taken mandatory security and privacy training classes and annual refreshers.

Technical Controls: Access to the system is controlled by NIH log in which authenticates the user prior to granting access. Access level and permissions are controlled by the system and based on user, role, organizational unit, and status of the report. All servers have been configured to remove all unused applications and system files and all local account access except when necessary to manage the system and maintain integrity of data.

Physical Controls: The servers reside in the University of Chicago Center data center and Amazon Web Services hosting locations where policies and procedures are in place to restrict access to the machines. This includes guards at the front door and entrance to the data center.

Identify the publicly-available URL:

https://portal.gdc.cancer.gov
https://api.gdc.cancer.gov
https://docs.gdc.cancer.gov
https://login.gdc.cancer.gov
uat-portal.gdc.cancer.gov
uat-api.gdc.cancer.gov
Note: web address is a hyperlink.

Does the website have a posted privacy notice?

Yes

Is the privacy policy available in a machine-readable format?

Yes

Does the website use web measurement and customization technology?

Yes

Select the type of website measurement and customization technologies is in use and if it is used to collect PII.

Does the website have any information or pages directed at children uner the age of thirteen?

No

Does the website contain links to non- federal government websites external to HHS?

Yes

Is a disclaimer notice provided to users that follow external links to websites not owned or operated by HHS?

Yes