# US Department of Health and Human Services

## Privacy Impact Assessment

### Date Signed:

04/21/2026

### OPDIV:

NIH

### Name:

NCI Center for Technical and Operational Support Cancer Research Ecosystem

**PIA Unique Identifier:**

P-9865729-871788

**The subject of this PIA is which of the following?**

Minor Application (child)

**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.**

Agency

**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.**

The Center for Technical and Operational Support (CTOS) Cancer Research Ecosystem (CCRE) supports cancer research communities through an established environment, facilitating and streamlining research data submissions, access to cloud-based tools, analyzing harmonized data, and interactive innovation and collaboration with the National Cancer Institute (NCI). The CCRE is comprised of sub-systems, data repositories, and controlled-access data repositories (CADRs).

The CCRE systems support the mission and objectives of the NCI Cancer Research Data Commons (CRDC) and Childhood Cancer Data Initiative (CCDI) programs. The subsystems in the CCRE system boundary offer web-based interfaces for exploring meta-data that describe data files that are stored in public data repositories and CADRs. Users of the subsystems generate file manifests that itemize data files of interest which can be accessed from within NCI supported high-performance computing (HPC) environments, such as the Seven Bridges Cancer Genomics Cloud (SBCGC). The meta-data that is provided to support manifest generation and the content of the manifests exported from CCRE subsystems do not contain sensitive or controlled data.

For select CCRE subsystems, controlled-access data is made available through the sub-system user interface for authorized end users. These subsystems are classified as components of controlled-access data repositories (CADR).

The CADRs within the CCRE system boundary are identified as follows:

General Commons Data Repository (GC): A dedicated storage environment for controlled-access genomic, imaging, and proteomic data submitted from NIH-funded cancer research studies.

Kids First Data Repository (KF): A dedicated storage environment for controlled-access genomic and clinical data submitted from the Gabriella Miller Kids First Pediatric Research Program.

Clinical and Translational Data Commons (CTDC): A web-based subsystem for searching and analyzing clinical data and reports, as well as metadata that describes controlled-access molecular and genomic data files stored in the General Commons Data Repository. While the data stored in the GC is not exposed through the web interface, the controlled-access clinical data is available for authorized users in the CTDC user interface.

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

The CCRE maintains personally identifiable information (PII) within its subsystems, data repositories, and CADRs for the purpose of supporting global cancer research objectives. The PII within CCRE are clinical data and include demographics (sex, age, ethnicity), medical notes, medical records number (MRN), genomic, proteomic, and imaging.

There are only two pathways for end users to access controlled access data stored in the subsystems, data repositories, and CADRs. The controlled-access data stored within the CCRE boundary is protected by authentication and authorization solutions. The pathways respective authentication and authorization mechanisms are described as follows:

Sub-system web interface: Users are required to authenticate using the NIH Login service, an NIH-managed authentication service under the NIH Identity, Credential, and Access Management (IAM) Services. IAM maintains its own Privacy Impact Assessment (PIA).

Authorized access through one of the NCI-supported high-performance computing environments (HPC), such as the Seven Bridges Cancer Genomics Cloud (SB-CGC). Authentication and authorization is implemented by way of the NCI Data Commons Framework (DCF) services, which leverages the NIH Database of Genotypes and Phenotypes (dbGaP) authorizations to determine user permissions for controlled access data.

Changes to authentication and authorization for both pathways is anticipated to change in 2026. All CCRE subsystems, data repositories, and controlled-access data repositories (CADR) are in the process of adopting the NIH Researcher Auth Service (NIH RAS), a separate NIH Identity, Credential, and Access Management (IAM) Services solution.

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

The Center for Technical and Operational Support (CTOS) Cancer Research Ecosystem (CCRE) supports cancer research communities through an established environment, facilitating and streamlining research data submissions, access to cloud-based tools, analyzing harmonized data, and interactive  innovation and collaboration with the National Cancer Institute (NCI). The CCRE is comprised of sub-systems, data repositories, and controlled-access data repositories (CADRs).

The CCRE systems support the mission and objectives of the NCI Cancer Research Data Commons (CRDC) and Childhood Cancer Data Initiative (CCDI) programs. The subsystems in the CCRE system boundary offer web-based interfaces for exploring meta-data that describe data files that are stored in public data repositories and CADRs. Users of the subsystems generate file manifests that itemize data files of interest which can be accessed from within NCI supported high-performance computing (HPC) environments, such as the Seven Bridges Cancer Genomics Cloud (SBCGC). The meta-data that is provided to support manifest generation and the content of the manifests exported from CCRE subsystems do not contain sensitive or controlled data.

For select CCRE subsystems, controlled-access data is made available through the sub-system user interface for authorized end users. These subsystems are classified as components of controlled-access data repositories (CADR).

The CADRs within the CCRE system boundary are identified as follows:

General Commons Data Repository (GC): A dedicated storage environment for controlled-access genomic, imaging, and proteomic data submitted from NIH-funded cancer research studies.

Kids First Data Repository (KF): A dedicated storage environment for controlled-access genomic and clinical data submitted from the Gabriella Miller Kids First Pediatric Research Program.

Clinical and Translational Data Commons (CTDC): A web-based subsystem for searching and analyzing clinical data and reports, as well as metadata that describes controlled-access molecular and genomic data files stored in the General Commons Data Repository. While the data stored in the GC is not exposed through the web interface, the controlled-access clinical data is available for authorized users in the CTDC user interface.

The CCRE maintains personally identifiable information (PII) within its subsystems, data repositories, and CADRs for the purpose of supporting global cancer research objectives. The PII within CCRE are clinical data and include demographics (sex, age, ethnicity), medical notes, medical records number (MRN), genomic, proteomic, and imaging.

There are only two pathways for end users to access controlled access data stored in the subsystems, data repositories, and CADRs. The controlled-access data stored within the CCRE boundary is protected by authentication and authorization solutions. The pathways respective authentication and authorization mechanisms are described as follows:

Sub-system web interface: Users are required to authenticate using the NIH Login service, an NIH-managed authentication service under the NIH Identity, Credential, and Access Management (IAM) Services. IAM maintains its own Privacy Impact Assessment (PIA).

Authorized access through one of the NCI-supported high-performance computing environments (HPC), such as the Seven Bridges Cancer Genomics Cloud (SB-CGC). Authentication and authorization is implemented by way of the NCI Data Commons Framework (DCF) services, which leverages the NIH Database of Genotypes and Phenotypes (dbGaP) authorizations to determine user permissions for controlled access data.

Changes to authentication and authorization for both pathways is anticipated to change in 2026. All CCRE subsystems, data repositories, and controlled-access data repositories (CADR) are in the process of adopting the NIH Researcher Auth Service (NIH RAS), a separate NIH

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

Yes

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

Medical Records Number
Medical Notes
Demographics (sex, age, ethnicity)
genomic, proteomic, and imaging

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

Public Citizens
Vendor/Suppliers/Contractors
Patients

**How many individuals' PII is in the system?**

5,000-9,999

**For what primary purpose is the PII used?**

To provide relevant data, statistics, and insights related to cancer research activity, including NCI-sponsored cancer clinical trials, extramural cancer research entities, and public cancer data registries to drive scientific discovery among the greater cancer research communities.

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

NA

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

Public Health Service Act. (42 U.S.C. 241, 242, 248, 281, 282, 284, 285a-285q, 287, 287b, 287c, 289a, 289c, and 44 U.S.C. 3101)

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

No

**Identify the sources of PII in the system.**

Government Sources
Within OpDiv
Other HHS OpDiv
Other Federal Sources
Public

**Identify the OMB information collection approval number and expiration date**

N/A. Public Law 114-255, Section 2035, exempts research conducted by NIH from Paperwork Reduction Act (PRA) requirements.

**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.**

Users fill out a data use/data access agreement before they are allowed to have access to the CCRE CADRs and raw PII.

**Describe the procedures for accounting for disclosures.**

NA. The system and its subsystems are not Privacy Act System of Records.

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

CCRE is not the source system. Processes to notify individuals that their information will be collected is done at the point of data intake by partnered registries, organizations, principal investigators (PIs), and during the clinical trial recruitment.

**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.**

There is no opt-out option. CCRE is not the source system. Processes for individuals to opt-out of the collection of their PII is done at the point of data intake by partnered registries, organizations, PIs, and during the clinical trial recruitment.

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

CCRE is not the source system. Processes to obtain consent from individuals is done at the point of data intake by partnered registries, organizations, PIs, and during the clinical trial recruitment.

**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.**

CCRE is not the source system. Processes to resolve concerns from individuals is done at the point of data intake by partnered registries, organizations, PIs, and during the clinical trial recruitment.

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

CCRE is not the source system. Processes for periodic review of data is done at the point of data intake by partnered registries, organizations, PIs, and during the clinical trial recruitment.

**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.**

Processes have been established to ensure least privilege and separation of duty to prevent unintended and unauthorized access to sensitive data, such as PII. Developers, administrators, and contractors do not have access to the underlying data in a production environment. Production environments that contain sensitive data are managed by the NCI Cloud One operations team.

**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.**

Processes have been established to ensure least privilege and separation of duty to prevent unintended and unauthorized access to sensitive data, such as PII. Developers, administrators, and contractors do not have access to the underlying data in a production environment. Production

environments that contain sensitive data are managed by the NCI Cloud One operations team. Changes to the system are propagated from non-production environments to production environments using automated delivery systems, ensuring support teams have the minimum amount of access to information required to fulfill project responsibilities.

**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).**

NA

**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.**

03-005, Patient Medical Records. Cut off patient case file annually after 5 years of inactivity. Destroy when case file
is no longer needed for scientific reference (DAA-0443-2012-0007-0010).

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

Administrative Controls:
For system users, access to sensitive data is secured using administrative controls adopted by the National Cancer Institute's Cloud One operating model, which ensures that administrative practices are in place to maintain separation of duty and least privilege. Cloud One operations staff leverage policies to prevent unauthorized individuals to access data and systems in production environments, including system development teams. Users with access to sensitive data in production environments are reviewed and audited on a routine basis as specified in the NCI Cloud One GSS authorization. For end users, access to controlled access data requires an approved Data Access Request (DAR) that is reviewed by the NCI Data Access Committee (DAC).

Technical Controls:
System user access to the system is controlled by NIH/NCI approved systems which authenticates the system user prior to granting access. Access level and permissions are associated with authenticated users that assume a role with greater permissions than unauthenticated users. 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. End users accessing controlled-access data are required to authenticate and the data access authorization mechanisms enforce access only to data that has been approved by the NCI DAC.

Physical Controls:
The system is entirely hosted on NCI Amazon Web Services Cloud One. All system infrastructure components are maintained in Federal Risk and Authorization Management Program certified environments, and physical security controls are inherited under the AWS shared responsibility model.

**Identify the publicly-available URL:**

https://datacatalog.ccdi.cancer.gov/
https://studycatalog.cancer.gov/

https://moleculartargets.ccdi.cancer.gov/
https://ccdi.cancer.gov/
https://caninecommons.cancer.gov/
https://general.datacommons.cancer.gov/
https://clinicalcommons.ccdi.cancer.gov/
https://clinical.datacommons.cancer.gov/
https://hub.datacommons.cancer.gov/
https://cbioportal.ccdi.cancer.gov/
https://populationsciences.datacommons.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?**

No

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

Yes