US Department of Health and Human Services

Privacy Impact Assessment

Date Signed:

04/29/2024

OPDIV:

NIH

Name:

National Covid Cohort Collaborative (N3C)

PIA Unique Identifier:

P-5515149-785938

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?

No

Identify the operator.

Contractor

Is this a new or existing system?

Existing

Does the system have Security Authorization (SA)?

Yes

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

PIA Validation

Describe in further detail any changes to the system that have occurred since the last PIA.

Updated the Privacy Impact Assessment (PIA) with specifics on whose (Personally Identifiable Information (PII) is hosted in National Covid Cohort Collaboration (N3C).

The website is no longer available to the public, but only to registered users.

Describe the purpose of the system.

The National Coronavirus (CoVid 19) Cohort Collaborative (N3C) system serves as a cloud-based hosting environment for
multiple data models that are transformed into a common
analytic model for performing cohort analysis of CoVid-19
datasets and ongoing research involving public health actions,
clinical care, and treatment.

N3C is a collaborative partnership between the National Center for Advancing Translational Sciences (NCATS), the National Patient-Centered Clinical Research Network (PCORnet), Observational Health Data Sciences and Informatics (OHDSI), Accrual to Clinical Trials Informatics for Integrating Biology and the Bedside (ACT/i2b2), TriNetX, and several HHS agencies, including Food and Drug Administration (FDA) and Center for Disease Control (CDC).

TriNetX is not an acronym.

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

The N3C resources CoVid-related clinical data from the following:

Limited data sets (LDS) which may include dates of service (such as admission, discharge), date of birth (DOB), date of death (DOD); city, state, zip code, and age.

De-identified data of medical records (visits & procedure occurrences, diagnosis, diseases, prescription & drugs, medical devices and supplies, medical conditions, date of death and health care provider locations).

Synthetic data sets that include an artificial, statistically comparable, computational derivative of the original data.
These data sets are authorized through Data Transfer Agreements (DTA) and Data Sharing & User Agreements (DUA) titled the NIH CoVid-19 Data Warehouse.

N3C users and employees log in using one of the following:
NIH Identity, Credential, and Access Management (IAM) Services, which maintains its own unique privacy impact assessment (PIA) on record, including all legal authoritiesdocumented.

In-Common, through NIH IAM Services, provides single sign-on (SSO) access to cloud and local services, and seamless global collaboration for students, faculty, staff, and researchers.

HHS Personal Identity Verification (PIV), a US Federal government wide credential used to access Federally controlled facilities and information systems at the appropriate security level.

Login.gov -a publicly available secure online access to participating government programs.

N3C system is hosted on the Palantir Foundry Cloud Service (PFC) platform, a Federal Risk and Authorization Management Program (FedRAMP) authorized cloud service.

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

The N3C resources CoVid-related clinical data from the following:

LDS' which may include dates (such as admission, discharge, service), year of birth, DOD; city, state, zip code, and age.

De-identified data of medical records (visits & procedure occurrences, diagnosis, diseases, prescription & drugs, medical devices and supplies, medical conditions, date of death and health care provider locations).

Synthetic data sets that include an artificial, statistically comparable, computational derivative of the original data.
These data sets are authorized through DTAs and DUAs titled the NIH CoVid-19 Data Warehouse.

N3C users log in using one of the following:
NIH IAM, which maintains its own unique PIA on record, including all legal authorities documented.

In-Common, through NIH IAM Services, provides SSO access to cloud and local services, and seamless global collaboration for students, faculty, staff, and researchers.

HHS PIV, a US Federal government wide credential used to access Federally controlled facilities and information systems at the appropriate security level.

Login.gov -a publicly available secure online access to participating government programs.

N3C is a collaborative partnership between NCATS, PCORnet, OHDSI, ACT/i2b2, TriNetX, and several HHS agencies including FDA and CDC, TriNetX (not an acronym), biotech, and other private research organizations.

N3C system is hosted on the Palantir Foundry Cloud Service (PFC) platform, a Federal Risk and Authorization Management Program (FedRAMP) authorized cloud service.

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

Yes

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

Date of Birth
Name
E-Mail Address
De-identified datasets for research
Business/ Research Partner & Collaborators - names, email addresses title & organization
Limited datasets for research - admission/discharge dates, date of death; city, state, zip code, and age.

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

Employees

Business Partner/Contacts (Federal/state/local agencies)

How many individuals' PII is in the system?

500-4,999

For what primary purpose is the PII used?

Information is required to grant researchers, partners & collaborators access to the dataset for Covid-related research.

Employees' and direct contractors' PII are required to manage, operate, and maintain the N3C information technology platform.

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

NA - There is no secondary use.

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

5 USC 301, Departmental regulations.

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

No

Identify the sources of PII in the system.

Directly from an individual about whom the information pertains

Online

Government Sources

Other HHS OPDIV

Other Federal Entities

Non-Governmental Sources

Private Sector

Other

Identify the OMB information collection approval number and expiration date

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

Is the PII shared with other organizations?

No

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

Prior notice is not required because researchers, business partners/collaborators voluntarily sign data agreements when requesting access to N3C.

Employees and direct contractors voluntarily provide their personal information during the NIH staff onboarding process.

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 not process for individual to opt-out of providing their PII. Under the data use agreements, researchers, partners and collaborators voluntarily provide their PII when requesting access to N3C datasets.

Employees and direct contractors' PII are acquired based on their roles and responsibilities to develop, operate, and maintain the system.

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

The process to notify and obtain consent is published in our privacy statement. Changes to our policy are communicated by email and also posted on our N3C website.

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.

There is not a process in place to respond to individual concerns if they believe the PII had been inappropriately obtained, used, disclosed or inaccurate. The DTA identified the provider as the legal authority to collect and share the data with the recipient.

Researchers have signed DTA, DUA, and Data Use Request (DUR) on file.

Employees and direct contractors information are processed during the NIH staff onboarding and have signed non-disclosure statements.

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

Reviews are conducted for researchers, partners, and collaborators during an annual renewal or conclusion of their data use agreements (DTA, DUA, DUR).

Employee PII reviews are conducted during NIH internal processes including annual security awareness training and the termination processes.

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.

Access to Personally Identifiable Information (PII) is assigned to personnel based upon current job responsibilities. The system uses NIH IAM services to assign permissions/user roles which is considered PII.

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.

The NIH Security Awareness Training course is used to satisfy this requirement. According to NIH policy, all personnel who use NIH applications must attend security awareness training every year. There are five categories of mandatory information technology (IT) training (Information Security, Counterintelligence, Privacy Awareness, Records Management and Emergency Preparedness).

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

Users are provided in-platform walk-through training that covers core functionality and appropriate and authorized workflows in the N3C system. All data used and displayed in training is hypothetical.

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.

Records are retained and disposed of under the authority of the NIH Records Retention Schedule.

Item 03-005 - Patient Medical Records
These records document admissions and medical treatment for a patient accepted in a research project. These records are the primary source of evaluation and analysis for either clinical care or clinical research study.
Disposition: 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

Item 03-007 - Pathology Test Records
Pathology test records including media preparation case files, indices and formulas as required by 42 CFR 493. The records contain information related to requisitions for laboratory media and cells, including a description of the method of preparation and the ingredients used.
Disposition: Cut off annually after the date of reporting. Destroy 10 years after cutoff. DAA-0443-2012-0007-0012

Item 03-002 - Radiology and Imaging records
These records are comprised of electromagnetic radiation (Xrays) and other roentgenographic images produced by devices and procedures, such as body/head scans created by computerized transaxial tomography (CT). Files may include physician interpretations of images/scans.
DispositionCut off in 5 year intervals by fiscal year after file becomes inactive or when no longer needed for clinical reference, whichever is longer. Destroy 60 years after cutoff. DAA-0443-2012-0007-0007

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

Administrative

Users log in using secure government portals. Platform and data access are determined by the appropriate administrative authorities as described in the N3C Institutional Review Board (IRB) protocol. Administrative controls are reinforced by the technical and physical controls laid out below.

Technical

N3C hosted in Palantir Foundry Cloud Service (PFC), a Fed-ramp authorized cloud service and contractor's management service of N3C, ensures that data is secured in the system via several technical means. Across the platform, data is encrypted in transit and at rest. Palantir provides highly configurable access controls. Palantir Foundry platform supports comprehensive auditing of all data processing and access. It captures meta-data about the source of all data and maintains records of data imports, reads, writes, searches, exports, and deletions.

Physical

PFC is a cloud service platform hosted in Amazon Web Services (AWS). PFC system architecture is aligned with a number of framework and classified as a moderate security category with a FedRAMP Authorization to Operate and host a sensitive data.

Note: web address is a hyperlink.