US Department of Health and Human Services

Privacy Impact Assessment

Date Signed:

07/31/2020

OPDIV:

CDC

Name:

CSELS Data Hub (CDH)

PIA Unique Identifier:

P-8520604-354462

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.

Agency

Is this a new or existing system?

New

Does the system have Security Authorization (SA)?

No

Describe the purpose of the system.

CSELS Data Hub (CDH) receives public health data feeds from Electronic Health Records (EHR), Centers for Disease Control and Prevention (CDC) surveillance systems, CDC funded studies, state, local, tribal jurisdictions and various partners to support the COVID-19 data needs and build a repository of datasets relevant for responding to public health events now and in the future.

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

The data in CDH is provided by private and public organizations including healthcare agencies, universities, labs and federal agencies. It is important to note that the CDH platform is not considered to “collect” the data directly from users but receives large datasets collected from the agencies.

CDH data types collected from facilities, labs, pharmacies, National Emergency Medical Services (EMS) Information System (NEMSIS), and patients consist of organizational structure, services, beds, utilization, staffing, expenses, physician arrangements, system affiliation, geographic indicators, accreditations, approval codes, hospital readiness, adoption level of computerized systems, decision support, specialty, National Provider Information (NPI) code, specimen data, enrollment data, inpatient and outpatient medical claims, outpatient pharmacy claims, event code, procedures done by Emergency Medical Technicians (EMT), medications administered by EMTs, Reason for call (if reported to 9-1-1), non-identifiable unique patient ID.

Patient's PII data consists of demographics, name, mailing address, email addresses, phone number, occupation, medical notes, medical history, contact tracing, symptoms, test results, hospitalization dates, date of birth, sex, race, ethnicity, county, marital status, census tract, healthcare facilities where the patient appointments occur, standardized test order codes, Logical Observation Identifiers Names and Codes (LOINC), standardized result codes, and interpreted results as well as numeric result values and reference ranges, results include both positive and negative test results.

External non-CDC users from participating health agencies accessing the system are identified and authenticated via Amazon Web Services (AWS) Cognito service which is a Federal Risk and Authorization Management Program (FedRamp) approved service or AWS Identity Access Management (IAM) tool. User profiles are collected and stored in Cognito and IAM and consist of username, email address, phone number, full name and password. Internal CDC users accessing the system are identified and authenticated via CDC's Active Directory (AD); AD is a separate system with its own PIA.

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

The CSELS Data Hub Platform works in collaboration with participating commercial labs, vendors having public health datasets, EHR vendors, public organizations, federal agencies and universities that have agreed to share data with proper data use agreement (DUA) with CDC. The system does not directly collect the data from individuals. The data will be used by the CDC system Data Collation and Integration for Public Health Event Response (DCIPHER), Health and Human Services (HHS) Protect (Cloud Palantir) system to enrich the emergency department data and other datasets being provided to platform by multiple federal agencies to create a national picture. The data will also be used by CDC to do the needed research for their mission, by Emergency Operations Center (EOC) to derive useful information needed for timely decision making, by CDC studies to answer public health questions, and perform exploratory analysis and visualization. The data will be stored in the FedRAMP Amazon Web Services (AWS) environment. The data will be shared securely using AWS secure file transfer protocol service, access to Simple Scalable Storage (S3) service and Application Programming Interface (API) by enforcing strict role based access control (RBAC) and only providing read only access to data. The data will be maintained as per the duration agreed upon in the DUAs. It will be purged after the agreed upon duration or moved to archival (Amazon Glacier) when the latest data arrive, and older data is no longer needed.

External non-CDC users from participating health agencies accessing the system are identified and authenticated via Amazon Web Services (AWS) Cognito service which is a Federal Risk and Authorization Management Program (FedRamp) approved service or AWS Identity Access Management (IAM) tool. User profiles are collected and stored in Cognito and IAM and consist of username, email address, phone number, full name and password. Internal CDC users accessing the system are identified and authenticated via CDC's Active Directory (AD); AD is a separate system with its own PIA.

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

Mailing Address

Phone Numbers

Medical Records Number

Medical Notes

Race/Sex/Ethnicity

Occupation

User credentials and password

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

Public Citizens

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

Patients

How many individuals' PII is in the system?

1,000,000 or more

For what primary purpose is the PII used?

The PII data is strictly used for research purposes only, disease surveillance and reporting.

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

Stored user credentials and passwords are used to authenticate external users.

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

Public Health Service Act, Section 301, "Research and Investigation," (42 U.S.C. 241); and Sections 304, 306 and 308(d) which discuss authority to maintain data and provide assurances of confidentiality for health research and related activities (42 U.S.C. 242 b, k, and m(d)).

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

State/Local/Tribal

Other Federal Entities

Non-Governmental Sources

Private Sector

Other

Identify the OMB information collection approval number and expiration date

N/A

Is the PII shared with other organizations?

Yes

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

Within HHS

The data is used for disease surveillance and reporting. The system provides data for analysis to associate disease trends among groups like people within a certain age bracket, gender, geographic location, nationality, or race in order to derive useful information for decision making by public health officials, influence policies and communicate critical information.

Other Federal Agencies

The data is used for disease surveillance and reporting. The system provides data for analysis to associate disease trends among groups like people within a certain age bracket, gender, geographic location, nationality, or race in order to derive useful information for decision making by public health officials, influence policies and communicate critical information.

State or Local Agencies

The data is used for disease surveillance and reporting. The system provides data for analysis to associate disease trends among groups like people within a certain age bracket gender, geographic location, nationality, or race in order to derive useful information for decision making by public health officials, influence policies and communicate critical information.

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

CSELS Data Hub platform requires data use agreements (DUAs) between all entities that connect to the platform that governs the use of all the data including PII in compliance with the Federal Information Security Management Act (FISMA).

Describe the procedures for accounting for disclosures.

CDH information is not disclosed to other entities; agencies participating in the CDH platform are able to access the CDH, but they can only access their own data or data from other entities with whom they have data sharing agreements. Hence there are no specific accounting procedures for information 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.

No prior notice is given by CDC because CDH does not collect information directly from any individuals. The actual collection of the data is done by participating data providers. As the original collectors of data, obtaining consent from individuals and notifying individuals about data collection and use are the responsibility of those participating agencies.

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.

The CDH platform is a "downstream" recipient of data that has already been collected by data providers through the established agreements, procedures and arrangement they traditionally havefollowed for data collection. Individuals requesting to opt-out must do so according to the policies and procedures, systems, and options provided by data providers to individuals.

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

The CDH platform does not have a process to obtain consent from or notify individuals about data collection and use. CDH is a data repository of data that has already been collected by data providers through their agreements with healthcare facilities, participating organizations and individuals; obtaining consent from and notification of individuals about data use is the responsibility of the data providers that collect it and when changes occur to the system.

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.

The CDH platform does not have a process in place to work with individuals regarding concerns about their PII stored in the system because the records in the system are not subject to the Privacy Act. Further, consent, notification, and such interactions are conducted between individuals and the data provider collecting the PII and is out of scope of the CDH platform.

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

It is the primary responsibility of the data provider to maintain integrity, availability, accuracy and relevancy of the data in the system of origin used for providing data to CDC. In CDH original data received from data provider is maintained in its raw form to meet the requirement, while copy of data is created for data processing, visualization, reporting and analytic need. The data provider may periodically provide composite data that includes all the data previously provided and new data. The version process in the system maintains the old version while making the latest data available to user, thus maintains integrity and accuracy of data while making multiple versions available and option to revert to any previous version for relevancy.

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

Users:

By using shared data from multiple jurisdictions (shared per fully executed data-use agreements), agencies can perform disease analysis.

Administrators:

Administrators are required to have access to the database to maintain the system.

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

CDH program management review, on a case-by-case basis, which system users may access PII. The decision is based on the DUA and users' job requirements consistent with Role Based Access.

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.

Users are given access according to their DUA only and after proofing and approval. The data steward oversees the approval process and determines who gets access to the information for which the entity is responsible. The Least Privilege model is used for all grants of access and enforced with row-level security, column level security and filtered down view in logical/physical database. The system also implements scope down policy that enforce access to different state of data to a user based on DUA and the job they need to perform.

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.

All CDC personnel are required to take annual Privacy and Security Awareness Training (SAT).

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

None

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.

CDH data is kept by the CDC as a historical public health record, per CDC's "Scientific and Research Project Records Control Schedule", section 1a ("Authorized Disposition: PERMANENT"). Records Schedule N1-442-09-1. The data will be securely purged from the system if the agreed upon DUA requires purging the data after the agreed upon duration.

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

Physical controls:
CDH data is protected by physical controls including physical barriers and locked doors protecting restricted areas, security guards, video cameras, climate control (cooling), redundant power systems, and fire prevention and control systems.

Technical controls:
CDH technical controls include Role-Based Access Control for all users (system administrators, developers, and users), encryption, multiple firewalls, and system redundancy. The system also undergoes continuous monitoring using automated monitoring systems.

Administrative controls:
CDH administrative controls include restricted access to the system; each individual user is vetted by CDH program management. Security Awareness Training is required and must be updated annually. Background checks are required for all users. OMB (Office of Management and Budget), HHS and CDC security and privacy policies and standards are followed by all users of the system.