Privacy Impact Assessment (PIA): OS - HPDSP - QTR1 - 2022 - OS1132807

Signed Date: 2/9/2022

Approved

Acronyms
ATO - Authorization to Operate
CAC - Common Access Card
FISMA - Federal Information Security Management Act
ISA - Information Sharing Agreement
HHS - Department of Health and Human Services
MOU - Memorandum of Understanding
NARA - National Archives and Record Administration
OMB - Office of Management and Budget
PIA - Privacy Impact Assessment
PII - Personally Identifiable Information
POC - Point of Contact
PTA - Privacy Threshold Assessment
SORN - System of Records Notice
SSN - Social Security Number
URL - Uniform Resource Locator

General Information

		PIA ID:	1419448
PIA Name:	OS - HPDSP - QTR1 - 2022 - OS1132807	**Title:**	OS - HHS Protect Data Sharing Platform
OPDIV:	OS	**PIA Queue:**	

PTA

PTA - 1A:	Identify the Enterprise Performance Lifecycle Phase of the system	Operations and Maintenance
PTA - 1B:	Is this a FISMA-Reportable system?	Yes
PTA - 2:	Does the system include a website or online application?	Yes
PTA - 2A:	Are any of the URLs listed accessible by the general public (to include publicly accessible log in and internet websites/online applications)?	

URL Details

Type of URL	List Of URL
Publicly accessible website with log in	https://protect.hhs.gov
Publicly accessible website with log in	https://protect.hhs.gov
Publicly accessible website with log in	https://protect.hhs.gov

PTA - 3:	Is the system or electronic collection, agency or contractor operated?	Contractor
PTA - 3A:	Is the data contained in the system owned by the agency or contractor?	Agency
PTA - 5:	Does the system have or is it covered by a Security Authorization to Operate (ATO)?	Yes
PTA - 5A:	If yes, Date of Authorization	12/21/2020
PTA - 5B:	If no, Planned Date of ATO	
PTA - 8:	Please give a brief overview and purpose of the system by describing what the functions of the system are and how the system carries out those functions?	The Department of Health and Human Services (HHS) Office of the Chief Information Officer (OCIO) urgently requires a data sharing platform

		Protect cloud services to support extending the HHS Protect Data Sharing Platform (HPDSP) to create a modern, scalable, cloud-based, data infrastructure serving the needs of HHS. The HPDSP team assists the HHS and our federal partners working with state, local, tribal, and territorial governments, and with the private sector to execute a holistic government response to fight the Coronavirus (COVID-19) pandemic and protect the public's health.
PTA - 9:	List and/or describe all the types of information that are collected (into), maintained, and/or shared in the system regardless of whether that information is PII and how long that information is stored.	All data sources integrated into HPDSP are intended to proportionally enable this whole of government response effort to the Coronavirus (COVID-19) pandemic.

In accordance with the afore mentioned necessity and proportionality principles for fair and appropriate use of public health data, HPDSP takes a cautious, granular, and graduated approach to data integration, management, and usage. HPDSP currently integrates limited Personally Identifiable Information (PII) data intended to support future critical public health workflows. Immediate workflows derivative of these data sources, however, do not require access to PII or Protected Health Information (PHI) sensitive fields which are therefore rigorously restricted by access controls to provide de-identified views of select data to specific users. The restricted sensitive PII/PHI fields are accessible on a limited, need-to-know basis to select HHS administrative personnel and contractors working at the direction of HHS personnel on critical Platform management related tasks (e.g., data integrations and pipeline administration).

Data is stored in accordance with the contract terms in HHS Data Use Agreements. Historical versions of datasets are retained for 30 days.

Data currently integrated into HPDSP includes the following categories and corresponding sources:

- **COVID-19 Case & Death Counts** - Multiple COVID-19 case and death counts sources to ensure comprehensive visibility. This includes Center for Disease Control and Prevention (CDC) forms filled out in data collation and integration for public health event response and public sources of case and death counts, as well as data from USAFacts
- **COVID-19 Testing and Labs Data** – State-level information on cases and testing for the 50 US states; Diagnostic and serology testing data provided by United States Government (USG) agencies, public health labs, commercial providers, states, and in-house labs provided data; wastewater COVID-19 surveillance tests.
- **Hospital Capacity and Utilization Data** - Hospital capacity, utilization, inventory, workforce, and supply provided by states and territories, Electronic Health Records (EHRs), and hospitals.
- **Emergency Department Data** - COVID and influenza-like ED data from public and EHRs along with symptoms provided by the CDC
- **Supply Chain** - Supply chain data from government and

industry

- **Community-based testing sites** – data related to testing locations, volumes, and sites managed by CDC & HHS ASPR, as well as FEMA.
- **Demographics** - Census population, mortality statistics, demographic statistics, behavioral risk factors, equity indicators, at-risk Medicare beneficiary group data, diabetes, and heart disease data for identification of vulnerable populations from publicly available data
- **Local Policy Actions** – state and local policy actions (e.g., stay at home orders, reopening, mask orders) from publicly available data
- **Infrastructure and Geospatial Data** – public infrastructure information (e.g., schools, facilities, etc.), geographic information system (GIS) data, and more from publicly available data
- **Drug and Vaccine Development Data** - available information about the development and distributions of tests, vaccines, and therapeutics related to COVID-19 or COVID-19-like research from USG agencies
- **Mobility Data** - data to measure broader trends in the population such as leveraging deidentified mobility data. This data provides information on the type of facilities and number of visits to establishments.
- **Deidentified EHR patient trajectory data** - de-identified data from EHRs on patient trajectories.
- **Models & Projections** - public and private models and projections
- **School Reopening** – local school district reopening's status trackers

PTA -9A:	Are user credentials used to access the system?	Yes
PTA - 9B:	Please identify the type of user credentials used to access the system.	HHS User Credentials HHS Password HHS Username Non-HHS User Credentials CAC Card Email address Password Username
PTA - 10:	Describe why all types of information is collected (into), maintained, and/or shared with another system. This description should specify what information is collected about each category of individual	HPDSP is an analysis and sharing platform to integrate, manage, and analyze critical data assets in order to enable authorized users to develop insights into how COVID-19 is spreading and how the federal government can best apply resources to mitigate and prevent spread. Data sources are gathered from across the federal government; state, local, tribal, and territorial governments; healthcare facilities; colleges and universities; and in partnership with select private commercial entities. Through HPDSP, HHS is streamlining data collection and integration with the goal of eliminating the need for insecure and unwieldy methods of transference, including faxes and email attachments to report this data. As one critical component of this streamlining process, HHS assigns a unique identifier for each specific entity within HHS Protect, which increases clarity by allowing for easier deduplication and data management. This system of using unique identifiers also increases the security of the data. Information is shared within the Coronavirus Task Force which includes members of FEMA, Veterans Affairs (VA), and Department of Defense (DoD), along with state and local governments.
PTA - 10A:	Are records in the system retrieved by one or more PII data elements?	No
PTA - 10B:	Please specify which PII data elements are used.	
PTA - 11:	Does the system collect, maintain, use, or share PII?	Yes

PIA

PIA - 1:	Indicate the type of PII that the system will collect or maintain	Name E-Mail Address Date of Birth User Credentials Patient ID Number Others - Date of death, user phone number, patient dates of service/diagnoses, demographics, health measurements, physician data, hospital and state contact data, zip codes

PIA - 2:	Indicate the categories of individuals about whom PII is collected, maintained or shared	Employees/ HHS Direct Contractors Patients Other
PIA - 3:	Indicate the approximate number of individuals whose PII is maintained in the system	Above 2000
PIA - 4:	For what primary purpose is the PII used?	Limited PII on HHS Protect Data Sharing Platform

(HPDSP) includes:

Patient Data – Patient Dates of Service (dates on which care was sought and provided at a health facility), Diagnoses, Demographics (age, race, and sex), ID numbers, health measurements (e.g., blood pressure, body mass index), birth and death dates, and zip code of residence or where care was received are integrated onto the system for purposes of facilitating analyses of healthcare (e.g., hospital, Emergency Department) utilization and capacity, along with analyses of population trends. Specifically, patient dates of service are used to understand utilization, COVID testing, and COVID cases time series; diagnoses to understand patient dispositions; demographics to analyze population trends; IDs for unique counts; zip codes for geographical clustering. This restricted sensitive PII/PHI fields are accessible on a limited, need-to-know basis to a limited set of HHS administrative personnel and contractors working at the direction of HHS personnel on critical Platform management related tasks (e.g., data integrations and pipeline administration). In addition, lab testing and reported COVID cases may contain information on health measurements (e.g., blood pressure, body mass index), birth dates, and death dates from their source systems. This data is immediately removed and stripped from the dataset prior to broader user access.

Physician Data – Physician name and ID number are maintained on the system. This data provides information on healthcare facility resources, utilization, and capacity data. This restricted sensitive PII/PHI fields are accessible on a limited, need-to-know basis to select HHS administrative personnel and contractors working at the direction of HHS personnel on critical Platform management related tasks (e.g., data integrations and pipeline administration).

Hospital and State Contact Data – State liaison and hospital contact information is maintained on the system for operations of communicating with hospital and state leads.

HPDSP User Data – Name, Email address, User credential and phone number are maintained on the platform. This is required data for users to access the system, where users include employees, contractors, and individuals from partner organizations such as states and hospitals. This data is provided by users as part of the account request, approval, and provisioning process managed by HHS Office of the Chief Information Officer (OCIO). For in-platform collaboration purposes, users can see the names and email addresses of other users with the same permissions.

Deployed Personnel – Employee-specific information for Office of the Assistant Secretary for Preparedness and Response (ASPR) deployed personnel was submitted to the HPDSP through May 2021 and discontinued thereafter.

Sensitive fields submitted include, name (full), flight number, departure time / origin airport / destination airport, team roster number, deployment location, hotel information.

ASPR and Associated-Employee Data — In response to state requests for support from the federal government, ASPR collects employee-specific information from requesting groups such as hospitals and states, government coordinators, and points of contact such as names, emails, phone numbers, email contents, and other notes.

PIA - 5:	Describe any secondary uses for which the PII will be used (e.g., testing, training or research)	Beyond limited and restricted PII that is integrated into the platform, HPDSP also tracks limited end-user information. For platform users, first and last name, email, are collected to ensure the stability of the website, to monitor information security, and to the extent the information has been de-identified, to improve the portal to be more useful to more visitors. User audit logs are also preserved in access restricted form for supervisory use to ensure accountability and oversight over Platform use and access to data.
PIA - 7:	Identify legal authorities, governing information use and disclosure specific to the system and program	HHS is authorized by Sections 301 and 319D of the Public Health Service Act [42 U.S.C. § 241 and 247d-4], as amended, to maintain active surveillance of diseases through epidemiological and laboratory investigations and data collection, analysis, and distribution.
PIA - 8:	Provide the number, title, and URL of the Privacy Act System of Records Notice (SORN) that is being used to cover the system or indicate whether a new or revised SORN is in development.	N/A: A System of Record Notice (SORN) isn't required because the system isn't a system of record and therefore the Privacy Act doesn't apply.
PIA - 9:	Identify the sources of PII in the system	Government Sources Within the OPDIV Other HHS OPDIV State/Local/Tribal Other Federal Entities Non-Government Sources Other
PIA - 9A:	Identify the OMB information collection approval number or explain why it is not applicable.	We are working with Paperwork Reduction Act Officer to determine if an Office of Management and Budget (OMB) Information collection approval number and the expiration date are required
PIA - 9B:	Identify the OMB information collection expiration date.	
PIA - 10:	Is the PII shared with other organizations outside the system's Operating Division?	No
PIA - 11:	Describe the process in place to notify individuals that their personal information will be collected. If no prior notice is given, explain the reason	A prior notice is not applicable as HPDSP is not involved in the collection process. HPDSP receives data through Data Use Agreements with its providers. Regarding HPDSP User PII, Direct contractors are informed of purposes and types of information collected via the in-platform privacy notice.
PIA - 12:	Is the submission of PII by individuals voluntary or mandatory?	Voluntary
PIA - 13:	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	HPDSP receives data through Data Use Agreements with its provider; HPDSP is not involved in the collection of the data. Therefore,

		the sources systems are responsible for providing methods for individuals to opt-out of the collection or use of their PII. Regarding PII of HPDSP users (Name, Email/User Credentials, Phone Number if necessary for authentication), the PII collected are required for the user’s account to be configured properly and are collected by HHS OCIO as part of account approval and provisioning processes. Opting out of providing this information for account provision would result in no account access being provided.
PIA - 14:	Describe the process to notify and obtain consent from the individuals whose PII is in the system when major changes occur to the system (e.g., disclosure and/or data uses have changed since the notice at the time of original collection). Alternatively, describe why they cannot be notified or have their consent obtained	There is not a process in place to respond to individual concerns. HPDSP is not involved in the collection of the data; as such, concerns would be directed to the source systems of record.
PIA - 15:	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 no process exists, explain why not	HPDSP is not involved in the collection of the data; HPDSP receives data through Data Use Agreements with its provider. Therefore, the sources systems are responsible for providing methods for individuals to raise and resolve concerns about the handling of their PII and any concerns about PII contained in HPDSP would be directed to those source systems. HPDSP captures changes in the underlying source systems of record and updated in the HPDSP platform through set pipeline schedules.
PIA - 16:	Describe the process in place for periodic reviews of PII contained in the system to ensure the data's integrity, availability, accuracy, and relevancy. Please address each element in your response. If no processes are in place, explain why not	There is not a process in place for periodic reviews of PII contained in HPDSP. The data provided from underlying source systems is considered to be “As Is.” However, once data is cleaned and stripped of PII in HPDSP, validation checks are run on scheduled intervals to affirm the integrity of the resulting dataset as compared to the raw data from the underlying source system.
PIA - 17:	Identify who will have access to the PII in the system and the reason why they require access	Administrators Contractors
PIA - 17A:	Provide the reason of access for each of the groups identified in PIA -17 Administrators and direct contractors - provide support to the system, integrate and clean the data of PII.	
PIA - 17B:	Select the type of contractor	HHS/OpDiv Direct Contractor
PIA - 18:	Describe the administrative procedures in place to determine which system users (administrators, developers, contractors, etc.) may access PII	Administrative controls regarding PII access are made by both HHS as the determiner of user access and the data source owner (e.g., Centers

		for Disease Control and Prevention (CDC), HHS). Determinations are made based on role-based access controls and a need-to-know basis, allowing users access to the minimum amount of data to perform their job. HPDSP enables administrators to ensure data is appropriately access controlled and only available to authorized users, as carried out by the technical means described in the following question.
PIA - 19:	Describe the technical methods in place to allow those with access to PII to only access the minimum amount of information necessary to perform their job	To carry out these organizational access determinations described above, HPDSP provides highly configurable access controls that enable administrators to implement flexible, granular permissions for entire projects, datasets, or even specific rows or columns within a dataset. The implementation of self-propagating authorizations enables administrators to ensure downstream compliance. Data can be further secured proportionate to its sensitivity by applying granular, project-based access controls to restrict or grant dataset sharing capabilities. These technical methods ensure that datasets containing PII are locked down to a highly restrictive user set. Once the data is cleaned of PII, access controls are applied to mirror administrative policies such that a broader, but still purpose-based and limited group, may see the resulting data. In addition, any analysis of user interactions with HHS Protect will be carried out for the purposes of improving the portal to be more useful to visitors and will be limited to de-identified information.
PIA - 20:	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 users sign the HHS Rules of Behavior prior to gaining access to HPDSP and complete any requisite training at the direction of HHS.
PIA - 21:	Describe training system users receive (above and beyond general security and privacy awareness training).	Users are provided in-platform walk-through trainings covering core functionality and appropriate and authorized workflows in the HHS Protect. All data used and displayed in trainings is data accessible by users with the most minimal level of access in the platform.
PIA - 23:	Describe the process and guidelines in place with regard to the retention and destruction of PII. Cite specific NARA records retention schedule(s) and include the retention period(s)	We are working with the Records Management Office to determine if a Records Retention Schedule is required. We will maintain records

		indefinitely until the appropriate schedule (if required) has been determined. Data is stored in accordance with the contract terms in HHS Data Use Agreements. Historical versions of datasets are retained for 30 days. In HPDSP, retention policies can be set on historical versions of datasets for configurable time periods to run automatically. These policies, and any changes to them, are tracked within the platform. For data ingested directly from source systems, HPDSP can be configured to mirror the retention rules of those systems, such that deletions will propagate through the platform to the user front end. HPDSP user PII - when any Palantir Foundry platform is decommissioned, the instance is shut down such that administrative access by users is blocked; data in the system is cordoned off and scheduled for full deletion of the cluster. Depending on security configuration, this could result in a waiting period of up to 180 days, though the data will be rendered unreadable or, at the very least, made inaccessible, prior to its removal.
PIA - 24:	Describe how the PII will be secured in the system using administrative, technical, and physical controls. Please address each element in your response	<u>Administrative -</u> Data governance decisions regarding access to HPDSP is determined by HHS. Access to

		datasets is determined by the appropriate data source owner (e.g., CDC, HHS). Administrative controls are reinforced by the technical and physical controls laid out below. Technical - Palantir Foundry 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 that enable administrators to implement flexible, granular permissions for entire projects, datasets, or even specific rows or columns within a dataset. These access controls rely on validation based on a system of tokens, which gets passed through different services only if a user is authorized. The resulting permissions flow through the system into derived data, analytics, or reports from the underlying secure fields to ensure all data is secure at all times. Additionally, Palantir Foundry platform supports comprehensive auditing of all data processing and access. It captures metadata about the source of all data. It also maintains records of data imports, reads, writes, searches, exports, and deletions. This metadata can be used to track revision history and to manage compliance with data auditing and oversight requirements. Palantir software includes several data and information protection functionalities to comply with regulations and industry requirements, such as California Consumer Privacy Act, Criminal Justice Information Services (CJIS), Department of Defense IL-4, Federal Information Security Management Act (FISAMA), General Data Protection Regulation (EU GDPR), and Health Insurance Portability and Accountability Act (HIPAA). Physical - Palantir systems architecture is aligned with the following certifications, frameworks, and attestations: SSAE18 (Not an acronym) System and Organization Controls (SOC) 2 Type II; International Standard on Assurance Engagements (ISAE) 3000 SOC 2 Type II; FedRAMP Moderate; and Trusted Information Security Assessment Exchange (TISAX) (in process).
PIA - 25:	Describe the purpose of the web site, who has access to it, and how users access the web site (via public URL, log in, etc.). Please address each element in your response	HHS and the CDC, within HHS, are working to obtain the data needed to address this public health emergency. Timely and accurate data are

		the critical resource in responding to COVID-19 to protect the public and mitigating the spread and impact of this virus. With the increasing number of cases in the United States and across the globe, the United States Government (USG) must access and assimilate all available resources to support efforts in the response using HHS Protect. User Accounts to HPDSP are provisioned via the HHS Identity and Access Management (HIAM) system, CDC Secure Access Management Service Identity Management System, or the Palantir Identity Management System (an Azure Active Directory authentication source). Data Governance decisions, including who has access to the HPDSP platform, are determined by HHS. Access to data sources is determined by the appropriate data owner (i.e., CDC, HHS, etc.). Users are able to access the system by hitting the public URL, then logging in via the account information they have been provisioned as described above.
PIA - 26:	Does the website have a posted privacy notice?	Yes
PIA - 27:	Does the website use web measurement and customization technology?	Yes
PIA - 27A:	Select the type of website measurement and customization technologies is in use and if it is used to collect PII	Session Cookies - Does Not Collect PII
PIA - 28:	Does the website have any information or pages directed at children under the age of thirteen?	
PIA - 28B:	Is there a unique privacy policy for the website and does the unique privacy policy address the process for obtaining parental consent if any information is collected?	No
PIA - 29:	Does the website contain links to non-federal government websites external to HHS?	Yes
PIA - 29A:	Is a disclaimer notice provided to users that follow external links to websites not owned or operated by HHS?	No
PIA - 29B:	Is a TPWA needed for this system?	No