Privacy Impact Assessment
for the

Neptune Pilot

DHS/ALL/PIA-046-1

September 25, 2013

Contact Point
Tom Bush
Common Vetting Task Force
Department of Homeland Security
202-447-4750

Reviewing Official
Jonathan R. Cantor
Acting Chief Privacy Officer
Department of Homeland Security
202-343-1717

Abstract

The Department of Homeland Security (DHS) Neptune Pilot v0.1 (Neptune Pilot) is a part of the overall DHS Data Framework, which directly supports DHS's need to use and share homeland security-related information across systems and components with privacy and civil liberties protections built directly into the data. The Neptune system is designed to ingest authoritative unclassified data sets of person-centric identity information that DHS gathers during its routine interactions with the public for the purpose of tagging the data, and providing the tagged data to authorized systems. Neptune tags the data by assigning both tag names and values to all of the ingested information. During the first sixty to ninety days, the Neptune Pilot will tag and share data with two different systems: the Common Entity Index Prototype (CEI Prototype) and the Cerberus Pilot. Both the CEI Prototype and Cerberus Pilot are described in separate Privacy Impact Assessments (PIA). The Neptune Pilot is a DHS-wide pilot administered by the Common Vetting Task Force (CVTF) with the support of the Office of Intelligence and Analysis (I&A) acting in coordination with the Office of the Chief Information Officer (OCIO). DHS is publishing this PIA pursuant to Section 208 of the E-Government Act of 2002 since the Neptune system handles personally identifiable information (PII).

If the system passes the testing and evaluation stage and DHS moves to an operational system, a new PIA will be published.

Overview

DHS Data Framework

Section 101 of the Homeland Security Act of 2002, Pub. Law No. 107-296 (Nov. 25, 2002), as amended, established the Department of Homeland Security (Department or DHS) as an executive department of the United States. The primary mission of the Department is to, among other things, prevent terrorist attacks within the United States, reduce the vulnerability of the United States to terrorism, minimize the damage and assist in the recovery from terrorist attacks that do occur within the United States, support the missions of its legacy components, monitor connections between illegal drug trafficking and terrorism, coordinate efforts to sever such connections, and otherwise contribute to efforts to interdict illegal drug trafficking. At the same time, the Department also has the primary responsibility to ensure that the privacy and civil rights and civil liberties of persons are not diminished by efforts, activities, and programs aimed at securing the homeland. To enable the Department to carry out these complementary missions, the Homeland Security Act eliminated information firewalls between government agencies by consolidating multiple agencies under the Department of Homeland Security.

DHS is changing the way it structures its information architecture and data governance to further this consolidation of information in a manner that fully protects individuals' privacy and civil rights and civil liberties. Since 2007, DHS has operated under the "One DHS" policy, which was implemented to afford DHS personnel timely access to the relevant and necessary homeland-security information they need to successfully perform their duties. The existing architecture of DHS databases, however, is not

conducive to effective implementation of the "One DHS" policy. Because this information is collected under different authorities and for various purposes, and is concurrently subject to privacy, civil rights and civil liberties, and other legal protections, DHS personnel requesting such information must (1) have an authorized purpose, mission, and need to know before accessing the information in the performance of their duties; (2) possess the requisite security clearance; and (3) assure adequate safeguarding and protection of the information. In the past, this access was technically cumbersome, time-intensive, and required personnel to log on and query separate databases in order to determine the extent of DHS holdings pertaining to a particular individual.

The Secretary and Deputy Secretary through the Common Vetting Task Force (CVTF)[1] and a collaboration between the Office of the Chief Information Officer, Office of Policy, Office of Intelligence and Analysis, the Privacy Office, Office for Civil Rights and Civil Liberties, Office of General Counsel, and operational components developed the DHS Data Framework. The goal of the DHS Data Framework is to enable a single user to search datasets extracted from multiple DHS systems for a specific purpose and view the authorized information in a clear and accessible format. The DHS Data Framework will create a systematic repeatable process for providing controlled access to DHS data across the enterprise. The DHS Data Framework will enable efficient and cost-effective searches across DHS databases in both classified and unclassified domains. The searches will identify key DHS data associated with an individual or identifier. The DHS Data Framework will ensure access to the most authoritative, timely, and accurate data available in DHS to support critical decision making and mission functions. Finally, the DHS Data Framework will enable controlled information sharing in both classified and unclassified domains in a manner that manages search parameters and access to the underlying data while maintaining the authoritative source of data at the source system.

In order to achieve this objective, DHS is creating two central repositories for DHS data, Neptune and Cerberus. Through these new systems, DHS will apply appropriate safeguards for access and use of DHS data and deliver new search and analytic capabilities such as entity resolution through correlation. New technology and the subsequent lower cost of aggregating large volumes of data collected by DHS have made this initiative possible. These technological developments enable more advanced, efficient analytics, while simultaneously offering stronger safeguards.

DHS has developed a comprehensive approach to automating and improving the access to and use of its mission data. The DHS Data Framework will implement four primary elements for controlling data: (1) user attributes (characteristics about the individual requesting access such as organization, clearance, and training); (2) data tags identifying the type of data involved, where the data originated, and when it was ingested for authoritative mission data; (3) context, which will combine purpose for which data can be used with the function which will identify what type of search and analysis can be conducted; and (4) dynamic access control policies that evaluate user attributes, data tags, and authorized purpose to grant or deny access to DHS data in the repository based on legal authorities and appropriate policies of the Department. DHS will log all activities to aid in oversight.

[1] The CVTF is a Department-wide task force comprised of representatives from support and operational components dedicated to improving the efficiency of the Department's screening and vetting activities.

Initially, DHS Data Framework will test three different capabilities needed to implement the full vision. The user attribute hub is being developed through a separate effort and will be incorporated into the DHS Framework. The following capabilities will test the other three elements of the framework:

- ***Neptune Pilot:*** The Neptune Pilot, residing in the Sensitive but Unclassified (SBU) domain, will ingest and tag data in the Neptune repository. Data in the Neptune Pilot will be shared with the CEI Prototype and the Cerberus Pilot, but will ***not*** be accessible for other purposes. This pilot will test the second element of the DHS Data Framework ("data tags").

- ***CEI Prototype:*** The CEI Prototype, residing on the SBU domain, will receive a subset of the tagged data from the Neptune Pilot and correlate data across component data sets. The CEI Prototype will test the utility of the Neptune-tagged data—specifically, the ability to ensure that only users with certain attributes are able to access data based on defined purposes using the dynamic access control process, which is described below in greater detail. This pilot will test the third and fourth elements of the DHS Data Framework ("authorized purpose/use" and "dynamic access control," respectively).

- ***Cerberus Pilot:*** The Cerberus Pilot, residing in the Top Secret/Sensitive Compartmented Information (TS/SCI) domain, will receive all of the tagged data from the Neptune Pilot and test the ability to ensure that only users with certain attributes are able to access data based on defined purposes using the dynamic access control process, which is described below in greater detail. This pilot will also test the "authorized purpose/use" and "dynamic access control" elements of DHS Data Framework.

The Neptune Pilot

The primary purpose of the Neptune Pilot is to test and demonstrate the basic technical capabilities of data tagging in support of the DHS Data Framework. Data tagging will preserve the lineage of the source data and the purpose and intent for its collection even when it is not resident in its source system. Ultimately, the tagging of data will be one part of the equation that helps ensure that information is delivered only to authorized users in conformance with law, policy, and best-practice standards.

Phase One of the Neptune Pilot will be limited to the staging of unclassified data currently contained in three DHS component databases: the U.S. Customs and Border Protection (CBP) Electronic System for Travel Authorization (ESTA)[2], the U.S. Immigration and Customs Enforcement (ICE) Student and Exchange Visitor Information System (SEVIS)[3], and the Transportation Security Administration (TSA) Alien Flight Student Program (AFSP)[4]. Data will be used for (1) tagging to test the technical control capabilities that will ultimately support controls necessary for protection of privacy, civil rights, and civil liberties, and (2) transfer to the CEI Prototype and Cerberus Pilot.

[2] DHS/CBP-009 - Electronic System for Travel Authorization (ESTA) July 30, 2012, 77 FR 44642.
[3] DHS/ICE 001 - Student and Exchange Visitor Information System January 5, 2010, 75 FR 412.
[4] DHS/TSA 002 - Transportation Security Threat Assessment System May 19, 2010, 75 FR 28046.

The tagged data will be stored on the DHS-owned Neptune data storage system and accessed by a limited number of DHS staff in order to assess whether Neptune tagging is successfully working. The initial tags necessary to support privacy, civil rights and civil liberties, and ultimately policies for access, use, and sharing of information remain subject to additional development through the efforts of the CVTF and future governance processes. Data inconsistencies and errors discovered while processing will be rejected and placed in a holding area for incompatible records to be manually reviewed. The ingest process will also generate auditable information as to the number of records processed and rejected. This overall approach to tagging will also allow the Department to apply a consistent taxonomy across the data sets.

The Neptune Pilot contemplates the ingestion and tagging of data elements from three unclassified databases collected by DHS components in the course of their mission execution. These data elements will be extracted from the source systems by the data owners as a one-time snapshot and placed on pre-determined secure media (e.g., DVD-ROMs, external hard drives) as data files. During the ingest process, the data will be tagged to align with the pilot's internal metadata structure. This will enable testing of proposed access controls and information management policies based on data owner and privacy, civil rights and civil liberties, and safeguarding considerations. An audit log of loaded data will be created.

For purposes of the Neptune Pilot, the tagged data stored in the Neptune repository will not be updated if there are changes to information in the component authoritative data systems. During the Neptune Pilot, this "one-time snapshot" data from the component databases will be used only for the testing and demonstration of data tagging/data protection capability and the provision of correct subsets of data to the CEI Prototype and the Cerberus Pilot. If the Neptune Pilot were determined to be successful at tagging data and appropriate for operational use, DHS would conduct a new PIA to cover the operational use of the system. At that time, data updates from the source system would be implemented and new data sets would be considered for addition to the operational Neptune system and considerations would be made for updating information in Neptune as the source system data changes.

The original source discs for information supplied to the Neptune system will be placed in an appropriate safe, with only authorized users given access to them. The discs are needed as backup to maintain a record of what data is transferred into the system. The source discs will be destroyed at the conclusion of the pilot when (1) ingest of the original source data is no longer required and (2) the pilot is complete as determined by the CVTF. This is consistent with the applicable retention schedules governed by the National Archives and Records Administration (NARA), which stipulate that tests and pilots do not need to be retained beyond their use to the agency.

Use of PII in the Neptune Pilot

The Neptune system will download PII originally collected in the ESTA, SEVIS, and AFSP data sets in order to demonstrate the ability to combine and tag data to produce a single secure repository of consistently tagged data that supports the CEI Prototype and the Cerberus Pilot. Tagged data will ultimately increase DHS capabilities for data safeguarding based on privacy and civil liberties protections, and data discovery using ontology mapping, which allows DHS to tag data consistently

from different databases. This will provide a complete integrated knowledge management solution for all data sets from structured to unstructured, and on all domains from Unclassified to Top Secret. The capability can be extended to all data types (e.g., text, images, video imagery, etc.) and allow for various types of analytics.

The Neptune Pilot will not support direct electronic searches, queries, or analysis and will not be used to identify predictive patterns or anomalies. Testing of the tags as one aspect of determining who gains access to particular data will be conducted in the Cerberus Pilot and the CEI Prototype. As part of Phase One of the Neptune Pilot, DHS will identify how to transfer data from the source systems to Neptune for tagging in order to maintain an accurate depiction of the data. Based on the pilots, DHS will determine next steps and provide updates to this PIA, as appropriate.

Section 1.0 Authorities and Other Requirements

1.1 What specific legal authorities and/or agreements permit and define the collection of information by the project in question?

Pursuant to 6 U.S.C. § 112, the Secretary of Homeland Security is directly charged by Congress to take reasonable steps to ensure that information systems and databases of the Department are compatible with each other and with appropriate databases of other departments and agencies. In fulfilling those responsibilities, the Secretary exercises direction, control, and authority over the entire Department, and all functions of all departmental officials are vested in the Secretary.

The DHS CIO operates pursuant to Homeland Security Act, 6 U.S.C. § 343; and Clinger-Cohen Act of 1996, 40 U.S.C. § 11101, et seq. The Under Secretary for I&A is responsible for performing the functions specified in Title II of the Homeland Security Act, 6 U.S.C. § 121. Those responsibilities include (but are not limited to):

- Integrating information gathered by the Department, and, as appropriate, making that information available within the Department and to other appropriate departments and agencies;
- Providing intelligence and information support to other elements of the Department;
- Establishing and using a secure communications and information technology infrastructure, including data-mining and other advanced analytic tools, to access, receive, and analyze data and information, and to disseminate information acquired and analyzed by the Department, as appropriate; and
- Ensuring that any information databases and analytical tools developed or used by the Department are compatible with one another and with relevant information databases of other agencies of the federal government, and that such information is treated in a manner that complies with applicable privacy requirements.

1.2 What Privacy Act System of Records Notice(s) (SORN(s)) apply to the information?

The data used in the Neptune Pilot are covered by the following source SORNs:

- Electronic System for Travel Authorization (ESTA) SORN, 77 Fed. Reg. 44642 (July 30, 2012)[5].
- Student & Exchange Visitor Information System (SEVIS) SORN, 75 Fed. Reg. 412 (Jan. 5, 2010)[6].
- Transportation Security Threat Assessment SORN, 75 Fed. Reg. 28046 (May 19, 2010)[7].

The Neptune Pilot will not result in the creation of a new system of records because it is merely a copy of data maintained in a source system of records and the individual records are tagged to permit the component data providers to maintain control over the contents of the records. The addition of data tags allows DHS to identify the source system, but not information about the individual. The Neptune Pilot cannot be queried or retrieved by personal identifiers.

1.3 Has a system security plan been completed for the information system(s) supporting the project?

The CEI Prototype team has completed Section 1 of the draft System Security Plan, per the request of the Data Center 1 (DC1) Information System Security Manager (ISSM). The current Federal Information Processing Standards (FIPS) is High-High-Moderate. The anticipated date of Security Authorization completion is January 2014.

1.4 Does a records retention schedule approved by the National Archives and Records Administration (NARA) exist?

A NARA-approved record schedule is available for each of the three data sets in the Neptune Pilot. The NARA General Records Schedule for Electronic Data, GRS-20, covers other electronic records created by the Neptune Pilot, while GRS-24, Information Technology Operations and Management Records, covers audit logs and other compliance-related documentation in the Neptune Pilot.

If the Neptune Pilot is not successful or the Neptune system does not become operational, the data received from the three systems will be deleted and destroyed from the system. If the Neptune system were to move from a pilot to operational status, DHS would identify how best to maintain accurate records in Neptune.

---

[5] http://www.gpo.gov/fdsys/pkg/FR-2012-07-30/html/2012-18552.htm

[6] http://edocket.access.gpo.gov/2010/E9-31268.htm

[7] http://edocket.access.gpo.gov/2010/2010-11919.htm

1.5 If the information is covered by the Paperwork Reduction Act (PRA), provide the OMB Control number and the agency number for the collection. If there are multiple forms, include a list in an appendix.

The provisions of the Paperwork Reduction Act of 1980, 44 U.S.C. §§ 3501-21, are not applicable to the Neptune Pilot. The information maintained in the underlying data sets is subject to the Paperwork Reduction Act. The OMB control number for SEVIS is 1653-0034, for ESTA 1651-0111, and for AFSP 1652-0021.

Section 2.0 Characterization of the Information

2.1 Identify the information the project collects, uses, disseminates, or maintains.

The Neptune Pilot effort will initially ingest, tag, retain and transfer information originally collected by the ESTA, SEVIS, and AFSP component database systems. The data elements that will be transferred to Neptune are listed in Appendix A.

The Neptune Pilot will create new metadata during ingestion of the source data. Data will be transformed into a standard form that supports storage and query operations. During that stage, data quality evaluation occurs. This will include detecting and rejecting corrupt records.

The data will be tagged with the following metatags:

- The source system;
- The date the information was replicated in the Neptune Pilot;
- A source system identifier, which will allow the specific data element to be traced back to the source system; and
- The contact information for the component data provider.

The data transferred to the Neptune Pilot system will also be grouped into core biographic data, extended biographic data, or detailed encounter data, with a tag indicating the grouping as follows:

- *Core biographic* data is basic biographic information to include name; date of birth; gender; country of citizenship; and country of birth.
- *Extended biographic* data is additional biographic information about an individual, such as address, phone number, email address, passport number, and/or visa number that is not considered Core Biographic information. Extended biographic data pertains to the subject of the encounter rather than associated third parties. Extended biographic data does not include DHS derogatory data or detailed information about DHS encounter(s) or transactions with an individual or associated third parties.
- *Detailed encounter* (Encounter) data is information that derives from a DHS screening, vetting, law enforcement, or immigration-related event/process and that is collected in accordance with DHS authorities and regulations. The term "encounter" is used to

describe a face-to-face meeting, an electronic or paper-based transaction (such as an application for a DHS-administered benefit), or as the result of information provided to the United States by a foreign government, aircraft operator, or other private entity. Detailed encounter data may contain DHS derogatory information, screening/vetting results, or information pertaining to third parties, such as sponsors or school program points of contact.

Neptune Pilot will transfer some Core and Extended biographic information to the CEI Prototype but no Encounter information; and Core, Extended, and Encounter information to Cerberus for the Cerberus Pilot.

2.2 What are the sources of the information and how is the information collected for the project?

The Neptune Pilot will ingest, tag, retain, and transfer information originally collected by the ESTA, SEVIS, and AFSP component database systems. The data will be divided into three categories: Core Biographic, Extended Biographic, and Encounter data. The specific information collected in those systems is set forth in each program's respective SORN and PIA.

Electronic System for Travel Authorization[8]: ESTA is an automated system administered by CBP that determines the eligibility of visitors (nonimmigrant aliens entering for business or pleasure for 90 days or less) to travel to the United States under the Visa Waiver Program (VWP). The ESTA application collects biographic information and answers to VWP eligibility questions and uses the information to vet applicants against various security and law enforcement databases in order to identify high-risk applicants. Authorization via ESTA does not determine admissibility to the United States. Rather, CBP officers determine admissibility upon a traveler's arrival.

Student and Exchange Visitor Information System[9]: SEVIS is an ICE program that monitors information about exchange visitors and international students and scholars (those with F-, M-, or J-visa status) while in the United States. The system, which is web-accessible, requires schools and programs approved to host students and scholars on these visas to report biographic and other information.

Alien Flight Student Program[10]: The AFSP is a TSA program to screen prospective flight student candidates who are not citizens of the United States before they are allowed to undergo pilot training. The mission of the program is to ensure that foreign students seeking training at flight schools regulated by the Federal Aviation Administration do not pose a threat to aviation or national security. Candidates log on to the AFSP Candidate website to submit their background information and flight training request(s). Once the application process is completed, TSA conducts a security threat assessment to determine whether the candidate poses a threat to aviation or national security.

A computer readable extract from each system will be added to the Neptune Pilot. The pilot will create new metadata during ingestion. Data will be transformed into a standard form that supports

[8] DHS/CBP/PIA-007 Electronic System for Travel Authorization PIA.
[9] DHS/ICE/PIA-001 Student and Exchange Visitor Information System PIA.
[10] DHS/TSA/PIA-026 Alien Flight Student Program PIA.

storage and query operations and tagged for core, extended, encounter, and other data noted above. This information in conjunction with individual user attributes and policy-based controls will allow DHS to provide dynamic access control to its data. An audit log for actions taken in the Neptune Pilot system will also be created that captures who requested access, when access was requested, what data was requested and for what purpose, and what actions were taken as a result of the request.

2.3 Does the project use information from commercial sources or publicly available data? If so, explain why and how this information is used.

No. The data ingested into the Neptune Pilot (ESTA, SEVIS, and AFSP) does not include information from public or commercial sources.

2.4 Discuss how accuracy of the data is ensured.

The Neptune Pilot relies on the accuracy of source systems. The original extraction of the data will be logged, stored, and available to verify consistency. The data will be evaluated to determine whether it conforms to defined formats and schemas (e.g., "date" format of "MM/DD/YYYY"). Data inconsistencies/errors discovered while mapping the source data elements to the Neptune Pilot data schema will be rejected and placed in a holding area for incompatible records to be manually reviewed. Manual Review will be conducted by the development team to determine whether the data was rejected due to problems in the ingest process or because of an actual problem with the source data. The ingest process will also generate auditable information as to the number of records processed and rejected.

2.5 Privacy Impact Analysis: Related to Characterization of the Information

Privacy Risks: There is a risk when reviewing the data for tagging that the nature of the data is impacted by the quality measures taken to create a consistent group of data across three different data sets.

Mitigation: All data ingested by Neptune has been obtained for purposes that are consistent with DHS authorities. Notwithstanding, the privacy rights of individuals about whom DHS collects, maintains, and uses PII depend on the quality of the information - its completeness, accuracy, timeliness, and relevance. The information contributed initially from the three underlying data sets is information that was collected directly from the subjects of the information or at the request of the individual by the school or sponsor, a factor that enhances accuracy. In order to mitigate this risk, the pilot will include a review of these quality assurance measures so that the data owners, CVTF, and DHS oversight offices (PRIV, CRCL, OGC) understand whether changes need to be made at the ingestion process.

Privacy Risk: There is a risk that the data will be inaccurate because DHS is taking a snapshot in time rather than routinely updating the data.

Mitigation: This risk is mitigated by the fact that DHS will not use the data in the Neptune Pilot for any purpose beyond testing the efficacy of the tagging process. This risk will be need to be addressed if the system becomes operational.

Section 3.0 Uses of the Information

3.1 Describe how and why the project uses the information.

The Neptune Pilot will demonstrate the feasibility of delivering tagged data to CEI and Cerberus so that users of these other systems can access the data based on controls that incorporate roles and rules that protect privacy, civil rights, and civil liberties, and other legal protections. The data will not be used for operational purposes in the Neptune Pilot.

3.2 Does the project use technology to conduct electronic searches, queries, or analyses in an electronic database to discover or locate a predictive pattern or an anomaly? If so, state how DHS plans to use such results.

The Neptune Pilot will not support direct electronic searches, queries, or analysis and will not be used to identify predictive patterns or anomalies. Any analytical tools will be described in the DHS Data Framework PIA, the Cerberus Pilot PIA, and the CEI Prototype PIA.

3.3 Are there other components with assigned roles and responsibilities within the system?

Only I&A Information Technology and OCIO staff will have direct access to the Neptune Pilot to tag the data. PRIV, CRCL, PLCY, and OGC will review the results of the tagging, but will not have direct access to the Neptune Pilot.

3.4 Privacy Impact Analysis: Related to the Uses of Information

Privacy Risks: There is a risk that data tags are insufficiently developed or incorrectly implemented.

Mitigation: This risk is mitigated by the fact that DHS will test the tagging of data through the Neptune Pilot before considering deploying an operational system. If the system were to achieve operational status, the data tags would have been developed and tested on the basis of input from data providers in conjunction with a specially constituted working group that includes the DHS OCIO, the component data provider, PRIV, CRCL, PLCY, and OGC. Further mitigation is provided by the fact that the tags would be modified based on the pilot results and thereafter as needed.

Privacy Risks: There is a risk of errors in the transfer of data from the source system.

Mitigation: As part of the pilot, DHS is working to minimize this risk through close oversight of the data transfers. The original extraction of the data will be logged and stored and available to verify consistency. Data inconsistencies and errors discovered while processing according to the mapping of the source data elements to the standard Neptune Pilot data schema will be rejected and placed in a holding area for incompatible records to be manually reviewed. After successful transfer of the data files from the transfer media, the transfer media will be placed in an appropriate safe, with only authorized users given access to it. The processed/tagged data will be stored on a DHS-owned storage system located at DHS Data Center 1. The storage media used in the transfer process will remain

available only for the purpose of re-ingestion of the data as necessary to adjust row key definitions and re-indexing to optimize the data. The ingest process will also generate auditable information as to the number of records processed and rejected.

Section 4.0 Notice

4.1 How does the project provide individuals notice prior to the collection of information? If notice is not provided, explain why not.

DHS is providing notice of this program through this PIA. In addition, DHS provides transparency through its SORNs, PIAs available on the DHS Privacy Office website, http://www.dhs.gov/privacy, and civil rights and civil liberties policies and procedures available on the DHS Office for Civil Rights and Civil Liberties website, http://www.dhs.gov/civilliberties.

ESTA, SEVIS, and AFSP records are covered by PIAs and SORNs as described above, which are available on the DHS Privacy Office website, and each program provides a Privacy Act statement that provides specific notice about the collection and use of the relevant information to the individual.

4.2 What opportunities are available for individuals to consent to uses, decline to provide information, or opt out of the project?

Individuals do not have the opportunity to consent to the use of their data in the Neptune Pilot.

4.3 Privacy Impact Analysis: Related to Notice

Privacy Risk: Individuals may not be aware that their information is being used in the Neptune Pilot.

Mitigation: DHS provides notice to individuals through this PIA, which serves as public notice of the existence of the Neptune Pilot, the data collected and maintained, and the routine uses associated with the information collected. The information is used only for the purposes mentioned in the public notice of this PIA. Additionally, as part of the prototype and piloting process, DHS will assess whether additional notice should be provided in the source system of records. This decision will be made prior to any operationalization of the Neptune system.

Section 5.0 Data Retention by the Project

5.1 Explain how long and for what reason the information is retained.

DHS is testing and evaluating the Neptune Pilot. As part of the pilot, DHS is evaluating the effectiveness of the system. If the Neptune Pilot is not approved for further development, all records will be deleted at the conclusion of the test and the component data providers notified of that deletion.

If the Neptune system were to become operational, DHS would retain the data elements based on the retention guidelines of the source system. Based on these retention guidelines, DHS would work on an appropriate retention schedule for the Neptune system with NARA.

Current retention schedules for the three data sets are as follows:

(1) CBP ESTA data is retained for no more than three years;

(2) ICE SEVIS data is retained for 75 years; and

(3) TSA AFSP data is retained as follows: For individuals who are not identified as possible security threats, records are destroyed one year after DHS/TSA is notified that access based on security threat assessment is no longer valid; (2) when an individual is identified as a possible security threat and subsequently cleared, records are destroyed seven years after completion of the security threat assessment or one year after being notified that access based on the security threat assessment is no longer valid, whichever is longer; and (3) when an individual is an actual match to a watchlist, that individual's record(s) are destroyed 99 years after the security threat assessment or seven years after DHS/TSA is notified that the individual is deceased, whichever is shorter.

5.2 Privacy Impact Analysis: Related to Retention

Privacy Risks: There is a risk that PII will be retained longer than is necessary and relevant to the purposes specified.

Mitigation: The requirements for retention of data in the Neptune Pilot are predicated on Federal law, DHS policy, and NARA-approved retention schedules for component systems. Data replicated in the Neptune Pilot will be subject to the retention policies governing the component databases from which the data is replicated. Information in the audit log is covered by General Records Schedules, GRS 20, approved by NARA.

Section 6.0 Information Sharing

6.1 Is information shared outside of DHS as part of the normal agency operations? If so, identify the organization(s) and how the information is accessed and how it is to be used.

The Neptune Pilot is being deployed to support the CEI Prototype and the Cerberus Pilot internally for DHS. There will be no information sharing outside DHS.

6.2 Describe how the external sharing noted in 6.1 is compatible with the SORN noted in 1.2.

Not applicable (no external sharing).

6.3 Does the project place limitations on re-dissemination?

Not applicable (no external sharing).

6.4 Describe how the project maintains a record of any disclosures outside of the Department.

Not applicable (no external sharing).

6.5 Privacy Impact Analysis: Related to Information Sharing

Not applicable (no external sharing).

Section 7.0 Redress

7.1 What are the procedures that allow individuals to access their information?

For the Neptune Pilot, individuals will not be able to access their data because it is not retrieved by personal identifier. Individuals interested in their records should either follow the directions outlined in the source system SORN or the Common Entity Index Prototype SORN[11].

7.2 What procedures are in place to allow the subject individual to correct inaccurate or erroneous information?

The component systems selected for initial ingestion in many cases allow users to access, self-correct, and update their information. The source system procedures for individuals to address possibly inaccurate or erroneous information are described in the respective SORNs for the source systems.

7.3 How does the project notify individuals about the procedures for correcting their information?

Because the data replicated in Neptune is the same as the data in the underlying systems, notification to individuals of the procedures for correcting the data ingested into the Neptune Pilot is the same as that of the source systems. Those procedures are set forth in the SORNs for the source systems.

7.4 Privacy Impact Analysis: Related to Redress

Privacy Risks: The Neptune Pilot contains PII that is not accurate.

Mitigation: The Neptune Pilot will rely on the accuracy of the underlying component systems that supply the information. To the extent those systems collect information directly from the individual involved, the opportunity is provided (as detailed above) for the individual to ensure the accuracy of the data submitted. An additional opportunity exists for individuals to request access to and/or correction of their record(s) in the underlying component systems, as permitted by law, DHS policy, and described in the applicable SORNs.

Section 8.0 Auditing and Accountability

8.1 How does the project ensure that the information is used in accordance with stated practices in this PIA?

The Neptune Pilot has no operational users. It will not be used for operational purposes or support operational decisions. The data store itself will be secured against accidental or deliberate

[11] DHS/ALL-035 Common Entity Index Prototype System of Records Notice, published August 23, 2013, 78 FR 52553.

unauthorized access, use, alteration, or destruction of information. The Neptune Pilot ensures that the information contained in Neptune is used in accordance with the practices stated in this PIA through specific auditing, accountability, and oversight measures. The specific auditing measures for the Neptune Pilot will include the following:

- A tamper-resistant audit log for robust continuous monitoring of data access, and
- Metrics for assessing the performance of the program and its compliance with stated practices.

8.2 Describe what privacy training is provided to users either generally or specifically relevant to the project.

DHS provides mandatory privacy training to all employees and contractors who have access to or use PII, and all users are required to complete mandated information security training that addresses privacy as well as the proper and secure use of DHS applications.

In addition, the DHS Privacy Office offers role-based training for agency employees involved with information sharing. The Office for Civil Rights and Civil Liberties offers several training products through its Civil Liberties Institute, accessible at http://www.dhs.gov/civillibertiesinstitute.

Any privacy training specific for the CEI Prototype and Cerberus Pilot users will be detailed in the PIAs for those systems.

8.3 What procedures are in place to determine which users may access the information and how does the project determine who has access?

The Neptune Pilot is only authorized for pre-operational testing by DHS information technology support personnel.

8.4 How does the project review and approve information sharing agreements, MOUs, new uses of the information, new access to the system by organizations within DHS and outside?

Only I&A and DHS OCIO information technology support staff participating in the development or maintenance of the system will be able to access the data for the purposes of system administration or performing data tagging.

Responsible Officials

Thomas Bush
Common Vetting Task Force

Approval Signature

Original signed on file at the DHS Privacy Office.

Jonathan R. Cantor
Acting Chief Privacy Officer
Department of Homeland Security

Appendix A

The following data elements will be ingested into Neptune:

ESTA

- Full Name (First, Middle, and Last);
- Date of birth;
- Gender;
- Email address;
- Phone number;
- Travel document type (e.g., passport), number, issuance date, expiration date and issuing country;
- Country of Citizenship;
- IP address;
- ESTA application number;
- Department of Treasury Pay.gov Payment Tracking Number (i.e., confirmation of payment; absence of payment confirmation will result in a "not cleared" determination);
- Country of Birth;
- Date of Anticipated Crossing;
- Airline and Flight Number;
- City of Embarkation;
- Address while visiting the United States (Number, Street, City, State);
- Whether the individual has a communicable disease, physical or mental disorder, or is a drug abuser or addict;
- Whether the individual has been arrested or convicted for a moral turpitude crime, drug possession or use, or has been sentenced for a period longer than five years;
- Whether the individual has engaged in espionage, sabotage, terrorism. or Nazi activity between 1933 and 1945;
- Whether the individual is seeking work in the U.S.;
- Whether the individual has been excluded or deported, or attempted to obtain a visa or enter U.S. by fraud or misrepresentation;
- Whether the individual has ever detained, retained, or withheld custody of a child from a U.S. citizen granted custody of the child;
- Whether the individual has ever been denied a U.S. visa or entry into the U.S., or had a visa cancelled, and, if so, the location and date of that denial or cancellation;
- Whether the individual has ever asserted immunity from prosecution; and
- Any change of address while in the U.S.

SEVIS

Biographical information for F/M/J nonimmigrants:

- names;
- U.S. domestic address;
- foreign address (F/M/J nonimmigrants only);
- date of birth;

- birth country and city;
- country of citizenship;
- country of legal permanent residence;
- username;
- email addresses;
- the DHS-assigned Immigrant Identification Number (IIN);
- Alien Registration Number (A-Number);
- National Identity Number (for F/M/J nonimmigrants only); and
- passport information (number, issuing country, expiration date).

F-1, M-1, or J-1 nonimmigrant educational and financial information:

- program of study;
- school registration information;
- program completion or termination information;
- transfer information;
- leave of absence information and study abroad;
- extensions;
- change of education level;
- student ID number;
- I-901 fee payment information; and
- financial information (for F/M nonimmigrants, financial information includes data on source of funds--personal or school, and average annual cost--tuition, books, fees, and living expenses; for J nonimmigrants financial information includes total estimated financial support, financial organization name, and support amount).

F/M/J nonimmigrant status and benefit information:

- The DHS-assigned Fingerprint Identification Number (for individuals 14 years of age and older);
- U.S. visa number, issuing country, expiration date;
- class of admission;
- immigrant benefit application information (primarily reinstatement, employment authorization, 212e waiver, etc.); and
- arrival and departure information (port of entry, date of entry/exit).

AFSP

- Full name (and any other names used previously);
- Any unique student identification number issued previously to the candidate by the Department of Justice or TSA (such as for other flight training);
- Passport and visa (if any) number;
- The candidate's current airman certificate, issuing country, certificate number, and type rating;
- The type of training for which the candidate is applying, the date of the candidate's prior recurrent training (if any);
- A copy of the training form documenting that recurrent training; the dates and location of the candidate's requested training;

- Candidate’s date of birth; gender;
- Birth country;
- Nationality;
- Height, weight, eye color, and hair color;
- Country of citizenship;
- Type of citizenship (current, dual, or historical);
- Whether citizenship is acquired through birth or naturalization;
- Dates of citizenship;
- Passport information (issue and expiration date, status, city of issuance); and
- His or her address, dates at the address, phone number, and email address.