Privacy Impact Assessment
for the

Cerberus Pilot

DHS/ALL/PIA-046-3

November 22, 2013

Contact Point
Clark Smith
Intelligence and Analysis/Chief Information Officer
Department of Homeland Security
202-282-8973

Reviewing Official
Karen L. Neuman
Chief Privacy Officer
Department of Homeland Security
202-343-1717

Abstract

The Department of Homeland Security (DHS) Cerberus Pilot (Cerberus Pilot) is a part of the overall DHS Data Framework.[1] The goal of Framework is to help alleviate mission limitations associated with stove piped Information Technology (IT) systems that are currently deployed across multiple operational components in DHS. More specifically, Cerberus Pilot is intended to test the feasibility of a more controlled, effective, efficient use and sharing of available homeland security information across the DHS Enterprise while protecting privacy and safeguarding personal data. During the Cerberus Pilot the system ingests certain DHS component unclassified data about individuals and maintains the data in a DHS owned and controlled cloud computing environment on a Top Secret/Sensitive Compartmented Information (TS/SCI) network, where it is available for classified searches and evaluation using various analytical tools. The goal of the Cerberus Pilot is to 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[2]. The Cerberus Pilot will be populated with data imported from the Neptune Pilot[3] data store via bulk load. The DHS Data Framework and Neptune Pilot are described in separate Privacy Impact Assessments (PIA).

The Cerberus Pilot is a DHS-wide pilot administered by the Office of Intelligence and Analysis (I&A) on behalf of the Department in coordination with the Office of the Chief Information Officer (OCIO) and overseen by the Common Vetting Task Force (CVTF). DHS is publishing this PIA pursuant to Section 208 of the E-Government Act of 2002 because the Cerberus Pilot system handles personally identifiable information (PII).

Overview

DHS Data Framework

The primary mission of DHS is, among other things, to 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 has the primary responsibility to ensure that the privacy, civil rights, and civil liberties of individuals 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 of 2002

[1] See DHS/All/PIA-046 DHS Data Framework PIA, published November 11, 2013.

[2] Dynamic Access Control are the policy rules that evaluate data tags, user attributes, and the context of a request to determine who gets access to what data and who can use what types of tools.

[3] See DHS/All/PIA-046-2 Neptune Pilot PIA, published November 11, 2013.

eliminated information stovepipes between government agencies by consolidating multiple agencies under DHS.

DHS is changing the way it structures its information architecture and data governance to further this consolidation of information in a manner that protects individual privacy, civil rights, and civil liberties.

Since February 1, 2007, DHS has operated under the DHS Policy for Internal Information Exchange and Sharing ("One DHS") policy, which was implemented to afford DHS personnel timely access to relevant and necessary homeland-security information they need to successfully perform their duties. Because DHS 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 performance of their duties; (2) possess the requisite background or security clearance; and (3) assure adequate safeguarding and protection of the information. This access is cumbersome, time-intensive, and requires personnel to log onto and search separate databases in order to determine what information separate DHS datasets contain about a particular individual.

The Secretary and Deputy Secretary developed the DHS Data Framework through the Common Vetting Task Force (CVTF)[4] and collaboration among the Offices of the Chief Information Officer (OCIO), Policy (PLCY), Intelligence and Analysis (I&A), operational components, and the Oversight Offices, Privacy (PRIV), Office for Civil Rights and Civil Liberties (CRCL), and Office of the General Counsel (OGC). The goal of the DHS Data Framework is to provide a user the ability to search different datasets extracted from multiple DHS systems for a specific purpose, retrieve accurate and responsive information, and view the information in a clear and accessible format.

To further this goal, the DHS Data Framework creates an efficient, systematic, repeatable process -- that is also cost-effective -- for providing highly controlled access to DHS data and searches across the enterprise in both classified and unclassified domains. The searches allow a user to identify key DHS data associated with an individual or identifier. Instead of relying on multiple copies of certain data sets distributed in multiple DHS systems, the DHS Data Framework will ensure authorized DHS personnel have access to the most authoritative, timely, and accurate data available in DHS databases to support critical decision making and mission functions. Thus, the DHS Data Framework will enable controlled information sharing in both the classified and unclassified domains in a manner that manages search parameters and access

[4] The CVTF is a Department-wide task force comprised of representatives from support and operational components and PRIV.

to the underlying data while maintaining the authoritative source of data at the authoritative system.

In order to achieve this objective, DHS is creating two central repositories for DHS data: Neptune and Cerberus Pilot. Through these new systems, DHS will apply appropriate safeguards for the access and use of DHS data and deliver new search and analytic capabilities such as entity resolution[5] through the correlation of disparate DHS data sets. A third system, the Common Entity Index (CEI) Prototype will test the usefulness of data in the Neptune system. All three of these systems are described more fully below. 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.

The DHS Data Framework implements four elements for controlling data:

(1) **User attributes** identify characteristics about the user requesting access such as organization, clearance, and training;

(2) **Data tags** label the data with the type of data involved, where the data originated, and when it was ingested;

(3) **Context** combines what type of search and analysis can be conducted (i.e., **authorized function**), with the purpose for which data can be used (i.e., **authorized purpose**); and

(4) **Dynamic access control policies** evaluate user attributes, data tags, and context to grant or deny access to DHS data in the repository based on legal authorities and appropriate policies of the Department.

DHS will log user activities to aid audit and oversight functions.

Initially, the DHS Data Framework will test the data tags, context, and dynamic access. Representatives from the CVTF, operational components, OCIO, and Oversight offices will participate the testing of the capabilities. To support the first element of the DHS Data Framework (user attributes), DHS is developing the User Attribute Hub through a separate effort and will subsequently incorporate it into the DHS Data Framework. The User Attribute Hub is where DHS will maintain a listing of a system user's attributes (e.g., component in which the individual works, location, job series, security clearance) for determining access control. The following capabilities will test the other elements of the framework:

- ***Neptune Pilot:***[6] The Neptune Pilot, residing in the Sensitive but Unclassified (SBU) domain, will ingest and tag data in the Neptune repository. This pilot will test the data tagging element of the DHS Data Framework by applying appropriate data tags, to data

[5] Entity resolution is the process by which a system or person identifies two records that represent the same entity and reconcile and merge the records or link the records.

[6] See DHS/ALL/PIA-046-2 Neptune Pilot PIA.

from multiple component datasets. Data in the Neptune Pilot will be shared with the Common Entity Index (CEI) Prototype and the Cerberus Pilot, but will ***not*** be accessible for other purposes.

- ***CEI Prototype:***[7] The CEI Prototype, also residing on the SBU domain, will receive a subset of the tagged data from the Neptune Pilot and will correlate this data, which is derived from multiple component datasets. 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 prototype will use data tags to test the third and fourth elements of the DHS Data Framework (authorized purpose/function 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 in a separate data repository known as Cerberus 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 leverage data tags to test the authorized purpose/function and dynamic access control elements of DHS Data Framework.

The Cerberus Pilot

The objective of the Cerberus Pilot is to enhance DHS's analytical capabilities by providing a more complete picture of DHS-collected data and significantly enhance the speed with which DHS is able to identify relevant information in a classified environment. The Cerberus Pilot's architecture will allow DHS to separate the data from the authoritative system and thus move to a "data-as-a-service."[8] This environment will enable mission users and intelligence analysts to work with data with the necessary controls in place to protect privacy, civil rights, and civil liberties. The architecture allows DHS to maintain the data in one location with the necessary controls without having to copy the data from one system to the next. The Cerberus Pilot will test and is intended to demonstrate the ability to import tagged data from the Neptune Pilot to the TS/SCI level, and test access control based on user attributes, context, and data tags. The Pilot receives its authoritative data from the Neptune Pilot. Data stored in Neptune Pilot is tagged as core biographic data, extended biographic data, and detailed encounter data and is then imported into the Cerberus Pilot via bulk load.[9] Once the data is ingested into the Cerberus Pilot, the dynamic access control will determine what information a user may access.

[7] See DHS/ALL/PIA-046-1 CEI Prototype PIA and DHS/All/-034 CEI Prototype SORN.

[8] Data-as-a-service offers data on demand based appropriate use and purposes but not tied to a particular software program.

[9] The tagging process is described in a PIA for the Neptune Pilot.

During the Cerberus Pilot, DHS personnel with authorized access will conduct searches for test and evaluation purposes based upon mission requirements, their status (role), data access rules, and DHS policy constraints. The Pilot will specifically 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. Testing and evaluation will be conducted on the DHS TS/SCI network. The Pilot will also leverage the "data tags" to test the "authorized purpose/use" and "dynamic access control" elements of the broader DHS Data Framework, including the ability to perform simple and complex searches across different component datasets using different analytical tools.

CVTF, operational component staff, OCIO staff, and oversight staff will closely review the testing..

Data involved in the Cerberus Pilot will be stored in a DHS-owned cloud computing environment on a physically secure TS/SCI network in order to facilitate searches using classified search terms. Initially, the Cerberus Pilot will not contain any classified data.

DHS will not perform any additional tagging during the import of data from the Neptune Pilot to Cerberus during the Pilot. The initial tags for supporting privacy, civil rights, and civil liberties protections during the pilot, which will ultimately be incorporated into policies for the access, use, and sharing of information. These tags will remain subject to additional development through the efforts of the CVTF and future governance processes.

The Cerberus Pilot will include the following elements:

- Creation of a secure cloud computing environment on a TS/SCI network to enable the processing, analysis, and use of DHS data from the Neptune Pilot.
- Initial importation, through the Neptune Pilot, of three DHS component data sets: the U.S. Customs and Border Protection (CBP) Electronic System for Travel Authorization (ESTA),[10] the U.S. Immigration and Customs Enforcement (ICE) Student and Exchange Visitor Information System (SEVIS),[11] and the Transportation Security Administration (TSA) Alien Flight Student Program (AFSP).[12]
- The testing of the proposed access controls and information management policies is based on data owner, privacy, civil rights and civil liberties, and information safeguarding considerations.

The Cerberus Pilot provides a controlled data service that will allow DHS to audit data searches submitted to Cerberus Pilot to ensure they comply with authorized purpose and functions.

Use of PII in Cerberus Pilot

DHS will use the PII maintained in Cerberus for counterterrorism purposes demonstrate whether and how Cerberus works. DHS test users will be able to conduct specific person based

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

searches, which is a query to retrieve information about an identified individual of interest using specific, detailed biographic information as the only search criteria. DHS test users will also be able to conduct characteristic search, which is a query to retrieve information about a partially identified individual(s)s of interest using limited specific biographic information in combination with general attributes known or reasonably believed to apply to the subject of the search. The policy and legal controls of the authoritative systems are maintained while the data is stored in the Cerberus Pilot. For example, only a user with appropriate access to data in the authoritative system will be able to conduct a search in Cerberus to discover or analyze the information that is returned. The returned data will be internally cached and purged daily, but the underlying source data that was queried will remain in the Cerberus Pilot in accordance with the authoritative system's retention policies, as defined by law and policy, including Executive Order 12333.

All three systems of records used in the pilot have a mission to screen individuals against national security and law enforcement databases to identify individuals with possible links to terrorism. The use of Cerberus to conduct counterterrorism searches is consistent with the purposes for which DHS collected the information. If a search identifies individuals across all three data sets that relates to counterterrorism as part of this pilot, the records would be covered by the DHS/I&A system of records notice, I&A's Enterprise Records SORN (73 FR 28128, May 15, 2008). DHS will not be using any of the results of these searches for operational purposes.

Access Controls.

The Cerberus Pilot employs dynamic access control to enable the automated enforcement of access requirements. In other words, a user sees only that information that he or she would otherwise be entitled to view in the authoritative system as a matter of law and policy. These access controls will provide data security and limit the dissemination of the information only to those who have a need-to-know – currently: personnel from Customs and Border Protection (CBP), Immigration and Customs Enforcement (ICE), Transportation Security Administration (TSA), Operations Coordination (OPS), and I&A During the Cerberus Pilot, information will not be shared outside of DHS.

The unclassified data maintained in Cerberus Pilot remains unclassified. If a specific piece of unclassified information is connected with classified material, then that association or connection will be classified. The fact that the data will be maintained on the TS/SCI environment does not change the nature of the data. The classification of the data will follow the authoritative system of records. PRIV will work with component FOIA offices and Cerberus Pilot project team to ensure appropriate access to the unclassified information maintained in Cerberus Pilot is available to respond to Privacy Act requests.

Because privacy is integral to maintaining public trust in DHS programs and initiatives, the Cerberus Pilot has been designed to embed privacy protections, including access control, at the data and level. Such data level privacy protections ensure that even though multiple data sets are aggregated to provide a more complete picture of an individual identity, access to that information is strictly limited to users whose permissions and roles allow it. Moreover, the Pilot supports immutable audit logs that are reviewed periodically to ensure that user permissions and roles are being adequately enforced by the system.

PRIV will review the tagging efforts and the context (i.e., authorized function along with authorized purpose) to develop effective dynamic access controls. If DHS determines that the Cerberus Pilot is successful at controlling access to data, analytical tools usage, and is appropriate for operational use, DHS will conduct a new PIA to cover the operational use of the system. At that time, DHS anticipates that it will automate data updates from Neptune system.

Oversight

In order to ensure that a robust oversight process is in place, PRIV, in addition to other DHS oversight offices, will review audit processes and logs by examining the actual use of the system to confirm that the access controls are functioning properly. PRIV will work with CVTF and others to codify the robust governance process that is being used to develop the tags and context and ensure that this governance process continues to include both DHS components and oversight offices.

Conclusion

There are both privacy risks and privacy benefits to a centralized location for DHS data that has been properly tagged. This PIA focuses on the Cerberus Pilot and its mission to provide data-as-a service to the DHS Enterprise. In order to move Cerberus from Pilot to operational, the following privacy risks must be appropriately mitigated and documented in published privacy compliance documents.

- A clearly defined leadership and governance structure for the ongoing development and maintenance of Cerberus and Neptune incorporates both operational and oversight components. The governance structure must include clear criteria for approving additional data sets, additional uses, and new analytic tools.
- Demonstration that the dynamic access controls work so that only those who should have access to the data do have for the specified limited purpose.
- Demonstration that data tagging is occurring properly in Neptune.
- Improved transparency to the public at the point of collection, in the applicable SORNs, and PIAs for data being stored in Cerberus.
- Defined process for providing access and redress, as appropriate to the data maintained in Cerberus and Neptune.
- Defined process for auditing the immutable logs and development plan for using technology to identify unusual or anomalous behavior in the system.

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 charged with taking reasonable steps to ensure that the Department's information systems and databases are compatible with each other and with appropriate databases of other departments and agencies. In fulfilling these 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 Cerberus Pilot is consistent with and promotes carrying out these responsibilities. In addition, the Cerberus Pilot is supported by the authorities described in the SORNs for the underlying data sets.

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

The data that will be imported initially into the Cerberus Pilot are covered by the following SORNs:

- Electronic System for Travel Authorization (ESTA) SORN, 77 FR 44642 (July 30, 2012), http://www.gpo.gov/fdsys/pkg/FR-2012-07-30/html/2012-18552.htm.
- Student & Exchange Visitor Information System (SEVIS) SORN, 75 FR 412 (Jan. 5, 2010), http://edocket.access.gpo.gov/2010/E9-31268.htm.
- Transportation Security Threat Assessment SORN, 75 FR 28046 (May 19, 2010) http://edocket.access.gpo.gov/2010/2010-11919.htm.
- In addition, the pilot will relate to counterterrorism mission. To the extent a record from all three of the above systems of records relates to counterterrorism, the SORN for I&A, the Office of Intelligence & Analysis Enterprise Records System (ERS) SORN, 73 FR 28128 (May 15, 2008), http://edocket.access.gpo.gov/2008/E8-10888.htm, applies to the use and maintenance, if any, of analytical or metadata output generated by the integration and analysis of the datasets described above within the Cerberus Pilot.

The Cerberus Pilot will not result in the creation of a new system of records. It is merely a copy of data maintained in an authoritative system of records. As part of the pilot PRIV will consider whether additional SORNs are required in order to improve transparency of its activities or because the nature of the data has changed based on pilot results.

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

The Cerberus Pilot 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 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?

Each of the initial three data sets that will be used in the Cerberus Pilot has an approved NARA record schedule.[13] In addition, the I&A ERS SORN has its own approved retention schedule for any analytical output generated by the use of Cerberus Pilot. The NARA General Records Schedule for Electronic Data, GRS-20, covers other electronic records created by Cerberus Pilot, including searches and results, while GRS-24, Information Technology Operations and Management Records, covers audit logs and other compliance-related documentation in Cerberus Pilot.

If the Cerberus Pilot is not successful or if the Cerberus Pilot system is not made operational, the data received from the three authoritative systems will be deleted and destroyed from the Cerberus Pilot system. If the Cerberus Pilot system is moved from a pilot to operational status, the records will remain in Cerberus and DHS will identify how best to maintain these records.

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 Cerberus Pilot project. 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 objective of Cerberus Pilot is to enhance the ability of DHS personnel to conduct the analysis necessary to protect homeland security. The initial Cerberus Pilot receives information originally collected by the ESTA, SEVIS, and AFSP database systems via an import from Neptune, which will have tagged and applied appropriate context to the data in order to support

[13] See individual system PIAs for more information on retention schedules.

dynamic access control. See the Neptune PIA, Appendix A, for a list of the data elements that have been ingested and tagged in Neptune and then transferred to Cerberus Pilot.

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

Cerberus Pilot will receive information originally collected through DHS mission execution via a computer-readable export from the Neptune data store. Neptune will provide tagged information originally collected by the ESTA, SEVIS, and AFSP database systems. The specific information collected in those systems is set forth in each program's respective PIA and SORN. A brief description of each system follows:

Electronic System for Travel Authorization:[14] 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 applicants who pose a law enforcement or security risk to travel. 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:[15] 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:[16] The AFSP is a TSA program that enables the screening of prospective flight student candidates who are not citizens of the United States before they are allowed to receive 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.

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

[14] DHS/CBP/PIA-007 Electronic System for Travel Authorization PIA.

[15] DHS/ICE/PIA-001 Student and Exchange Visitor Information System PIA.

[16] DHS/TSA/PIA-026 Alien Flight Student Program PIA.

No. The data to be imported into Cerberus Pilot (ESTA, SEVIS, and AFSP) does not include information from public or commercial sources.

2.4 Discuss how accuracy of the data is ensured.

It is the responsibility of the authoritative database owners (i.e., CBP, TSA, and ICE) to ensure the accuracy of the information. Once the information is in the databases, it is assumed to be correct. Neptune inherits the accuracy of the underlying authoritative databases, and Cerberus in turn inherits the accuracy of the data from Neptune. Neptune's process for data ingestion, auditing, and quality review is described in the Neptune Pilot PIA.

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.

2.5 Privacy Impact Analysis: Related to Characterization of the Information

Privacy Risks: There is a risk that data tagged in Neptune and subsequently Cerberus will be inaccurate.

Mitigation: In order to mitigate this risk, DHS will review quality assurance measures so that the data owners, CVTF, and DHS oversight offices (PRIV, CRCL, OGC) understand whether changes need to be made to 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 Cerberus Pilot for any purpose beyond testing the efficacy of the safeguarding and analytical use of the tagged data. This risk will be further addressed if the system becomes operational.

Section 3.0 Uses of the Information

3.1 Describe how and why the project uses the information.

In making data available as a service, Cerberus Pilot will enable DHS personnel to use tools to evaluate the analytical potential of the aggregated data store. The Cerberus Pilot uses information from ESTA, SEVIS, and AFSP to test and evaluate the ability to identify individuals and to investigate threats to the homeland in a manner that protects with privacy, civil rights, and civil liberties. The Cerberus Pilot is being deployed for demonstration and evaluation within DHS only. During the pilot, there will be no sharing of information outside of DHS.

For the Cerberus pilot, DHS will test counterterrorism searches to identify how the system functions.

The Cerberus Pilot will enable the following analytical searches:

- **Specific Person- or Entity-Based Search**: A search to retrieve information about an identified individual or entity of interest using specific, detailed biographic information as the only search criteria. A person-based search is the narrowest type of search and therefore the type of search that will generate the highest percentage of responses that are relevant to the underlying search (e.g., the name "Joseph A. Smith" or a date of birth of 3/29/70).
- **Characteristic-Based Search**: A search to retrieve information about an identified individual or entity of interest using limited and specific biographic information in combination with general attributes known or reasonably believed to apply to the subject of the search. A characteristic-based search is broader in scope than a person-based search; it is still designed to provide additional information concerning an individual based upon the limited identifying information available to the user at the time of the search. (e.g., male, Citizen of X Country, arriving on an international flight at John F. Kennedy International Airport on xx/xx/2013).
- **Pattern-Based Search**: A general search to identify one or more individuals or entities of interest using specific credible criteria based upon available information that is reasonably indicative of a threat to homeland security. Unlike person-based or characteristic-based searches, which are designed to identify additional information pertaining to an individual or entity already identified as a threat to homeland security, a pattern-based search is designed to identify previously unknown individuals who pose threats to homeland security. (e.g., Male, 18-35 years old, traveling from X City to Y City, between xx/xx/2013-xx/xx/2013).

If the Cerberus Pilot is successful, the technology will support advanced analytical tools. These analytical tools could be used to support link and trend analysis, entity disambiguation, data filtering, the creation of alerts, and other uses permitted by the DHS mission authority. The information obtained using such analytical tools assists analysts in either refining their analysis or formulating searches to obtain additional information upon which to base decisions or actions regarding individuals. If the Cerberus Pilot is successful, then the creation and use of such analytical tools would be approved under a governance framework and described in a subsequent PIA.

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 Cerberus Pilot data service provides electronic search capability available through a controlled interface. DHS will evaluate these results to determine the potential of the integrated data for DHS analysis. The Cerberus Pilot does not provide predictive analytics.

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

Yes. Data available in Cerberus Pilot will be shared, with DHS personnel from CBP, ICE, TSA, Operations Coordination (OPS), and I&A, consistent with law, memoranda of agreements, and policy, including, Executive Order (EO) 12333 and I&A's Interim Intelligence Oversight Procedures. There will be no sharing of information from derived from Cerberus Pilot outside of DHS.

3.4 Privacy Impact Analysis: Related to the Uses of Information

The following discussion identifies (1) the risks presented by the capabilities of the Cerberus Pilot, (2) the management of those risks, and (3) specific risk mitigation and management capabilities of the Cerberus Pilot.

Privacy Risks: There is a risk that more information than is necessary will be available to users.

Mitigation: A key aspect of privacy mitigation for the Cerberus Pilot is maintaining appropriate control over access to the data. Access includes both *who* can search the database and for what *purposes* such searches are permissible. DHS will change access control from the existing Role Based Access Control (RBAC) approach to one that includes a dynamic, granular access control mechanism and provides enhanced protection of privacy, civil rights, and civil liberties through definition and enforcement of who (User Attributes) is allowed access to individual data elements (Data Tags) for particular purposes (Context equals Purpose of the data added to the Function of the data).

With its attribute- and policy-based access control, the Cerberus Pilot provides new ways to ensure that privacy and civil liberties are well protected, through definition and enforcement of who is allowed access to specific data elements for particular purposes. Users requesting access to information are described in a standardized way (through user identity attributes) and data sources are defined in a standardized way (through data tagging). With these two requirements in place, rules can be created to automatically evaluate access requests and make policy-based decisions to permit or deny access to information. Any authorized user can then request information from any tagged system (without the need for separate logins to each system), with assurance that the information can be obtained as long as the user has the necessary permissions and access is consistent with defined policy. Ultimately, tagged data will enhance the capability to ensure that data from the original authoritative sources will be used only by authorized personnel for authorized purposes and functions.

During the Cerberus Pilot users/evaluators will have specified roles for authorized purposes and will have access only to the data necessary for pilot requirements gathering and evaluation of Cerberus Pilot's capabilities. All pilot users will be vetted and required to have a security clearance that is approved for access up to the highest classification of data on the system. Users with administrator privileges will be able to log directly on to the Cerberus Pilot system for maintenance and monitoring. All administrator actions will be logged and auditable. All other pilot users will be authenticated through an official certification authority service that will validate the user's public key infrastructure (PKI) certificate. Once authenticated, the user will only be able to access the system functions and data authorized for his or her assigned role. Determinations of what operations the user may or may not undertake will be made by comparing his/her authenticated identification and current role to an access control database of file permissions, program permissions, and data rights.

The Cerberus Pilot leverages the data tagging and access policies defined for the Neptune Pilot data cloud. For that effort, the analysis of the source data systems was performed according the DHS Data Framework and with the participation of the system data owners, DHS OCIO, PRIV, CRCL, and OGC to ensure that the rules appropriately safeguard the data in accordance with the requirements.

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 the Cerberus Pilot through this PIA. 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. Each program provides a Privacy Act statement that provides specific notice about the collection and use of the relevant information to the requesting individual. I&A’s ERS SORN is published on the DHS Privacy website and provides notice that data collected by DHS is used by I&A for analysis relating to counterterrorism.

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 Cerberus Pilot. DHS does not provide an opportunity to consent for data used by DHS for counterterrorism analytical purposes because this is the purpose for which the authoritative data was collected by the DHS component. Cerberus is a copy of the data from the authoritative system and follows the existing policies for use of the data. If as part of the pilot, DHS identifies records in all three data sets that relate to counterterrorism, the DHS I&A ERS SORN will apply to these records.

4.3 Privacy Impact Analysis: Related to Notice

Privacy Risks: There is a risk that individuals may not know what information is maintained within the Cerberus Pilot and how their information is used by DHS personnel.

Mitigation: This risk is mitigated by providing notice to individuals at the time of collection of information that is maintained in the authoritative systems (ESTA, SEVIS, and AFSP) and individuals can elect not to provide the information if they so choose. The SORNs for each of the systems are being updated to provide notice that information will be imported into the Cerberus Pilot. 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 requesting individual.

Additionally I&A's ERS SORN provides notice that I&A assesses and analyzes all terrorism, homeland security, and related law enforcement and intelligence information received by DHS, under Title II of the Homeland Security Act (6 U.S.C. § 121, et seq.), in support of the overall DHS mission.

PRIV will also assess after the pilot whether additional transparency at the point of information collection as well as SORNs or SORN updates are required to provide additional transparency.

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 Cerberus Pilot and the effectiveness of the Cerberus system. If the Cerberus Pilot is not successful, then all records in the Cerberus system will be deleted at the conclusion of the test.

If the Cerberus Pilot is successful and becomes an operational system, DHS will retain the data elements based on the retention guidelines of the authoritative system. Pursuant to these retention guidelines, the Cerberus Pilot will take appropriate action to handle the data by either archiving or retaining it.

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 AFS data is retained as follows:

a) for individuals who were not identified as possible security threat, records will be destroyed one year after DHS/TSA is notified that access based on security threat assessment is no longer valid;
b) when an individual was identified as a possible security threat and subsequently cleared, records will be 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
c) when the individual is identified as a security threat due to a positive match to a watchlist, records will be destroyed 99 years after the security threat assessment or seven years after DHS/TSA is notified the individual is deceased, whichever is shorter.

(5) I&A ERS SORN data is retained as follows:

a) Records schedule N1-563-07-16-3 applies to program records held by the analytic divisions, and includes raw reporting files, which are temporary records to be destroyed or deleted 30 years after cut-off.

b) Records schedule N1-563-09-07-1 applies to records concerning U.S. Persons, which are not permanently retained. Such records are destroyed within 180 days from the date the information is collected or when there is no longer a mission need to retain the information.

The returned data is internally cached and will be purged daily. The underlying source data that was queried remains in Cerberus in accordance with the authoritative system’s retention policies as required by law and policy.

5.2 Privacy Impact Analysis: Related to Retention

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

Mitigation: The requirements for retention of data are predicated on federal law and policy. DHS approved the retention schedules for component systems and Cerberus Pilot rules. Data imported into the Cerberus Pilot will be subject to the retention policies of the component databases (i.e., CBP ESTA, ICE SEVIS, TSA AFSP). New data composed of searches and results in the audit log are covered by General Records Schedules and the I&A ERS SORN.

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 Cerberus Pilot is being deployed for demonstration and evaluation within DHS. During the pilot, there will be no sharing of information derived from Cerberus Pilot outside of 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?

Individuals may seek access to their records based on the directions outlined in the authoritative system SORNs. Component FOIA offices will work with Cerberus Pilot staff to access the unclassified information on Cerberus Pilot in order to respond, as appropriate to Privacy Act Requests.

Individuals may seek access to their records based on the directions outlined in the authoritative system SORNs. Data maintained in Cerberus remains tagged separately and is part of the authoritative system of records although stored in a different location. Component FOIA offices will work with Cerberus Pilot staff to access the unclassified information on Cerberus Pilot in order to respond, as appropriate to Privacy Act Requests.

Because the authoritative systems that are ingested into the Cerberus Pilot may contain classified and sensitive but unclassified information related to intelligence, counterterrorism, homeland security, and law enforcement programs, certain records will be exempted from notification, access, and amendment to the extent permitted by subsection (j) and (k) of the Privacy Act and as described in the Code of Federal Regulations.

A request for access to non-exempt records in this system may be made by writing to the FOIA Officer, Office of Intelligence and Analysis, Department of Homeland Security, Washington, DC 20528, in conformance with 6 C.F.R. Part 5, Subpart B, which provides the rules for requesting access to Privacy Act records maintained by DHS.

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

The component systems selected for initial importation in many cases allow users to access, self-correct, and update their information. The authoritative system procedures for individuals to address possibly inaccurate or erroneous information are described in the respective SORNs for the systems. During the Cerberus Pilot, no updates will be performed to Cerberus Pilot records.

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

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

7.4 Privacy Impact Analysis: Related to Redress

Privacy Risks: There is a risk that Cerberus Pilot contains PII that is not accurate and individuals are adversely affected.

Mitigation: The Cerberus Pilot relies 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 before it is ingested into Neptune and Cerberus Pilot. An additional opportunity exists for individuals to request access to and/or correction of their record(s), as permitted by law and DHS policy and described in the applicable SORNs.

While an individual's record will not be updated in Cerberus Pilot, the risk to the individual is minimized by DHS not taking operational action based on the records in Cerberus Pilot.

As part of the Pilot, PRIV will assess how best to give access to the component FOIA offices so that they may respond to any applicable Privacy Act requests.

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 Cerberus Pilot has specific auditing, accountability, and oversight measures to ensure that the information it contains is used in accordance with the practices stated in this PIAt. The Cerberus data store itself will be secured against accidental or deliberate unauthorized access, use, alteration, or destruction of information. The specific auditing measures for the Cerberus Pilot will include the following:

- A tamper-resistant record that allows for robust continuous monitoring of data access, and
- Metrics for assessing the performance of the program and its compliance with policy-based access controls.

During the Cerberus Pilot evaluation, log management and analysis tools will monitor and assess audit data population and network processing to identify issues related to erroneous data, false inclusion/exclusion of access and/or information, and to prove that the audit capability is immutable.

- The security controls for access and use will be specific and granular to both limit access to appropriate users and regulate appropriate uses of information.
- The Cerberus Pilot will have structures in place to secure against accidental or deliberate unauthorized access, use, alteration, or destruction of information.
- Access to data in the Cerberus Pilot will be restricted to persons with a legitimate need to know and protected by appropriate access controls, taking into account the sensitivity of the data.

Legal and policy controls on the use and protection of information will be implemented through the policy process and integrated into the technology. This allows data protections to be built-in to Cerberus technology, and monitored enterprise-wide.

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 in the development and use of information sharing. The Office for Civil Rights and Civil Liberties also offers several training products through its Civil Liberties Institute, accessible at http://www.dhs.gov/civil-rights-and-civil-liberties-institute.

All Cerberus Pilot users will be provided with appropriate training before being afforded access to the system, and assistance for users will be available as needed.

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 Cerberus Pilot is only authorized for pre-operational testing. Therefore, users will be limited initially to a small number of I&A analysts that will serve as evaluators of the Cerberus Pilot. All users will be operating at the TS/SCI classified level and must have security clearances and access approvals commensurate with that level. In addition, I&A analytic personnel will access the information stored in the Cerberus Pilot in accordance with I&A's Interim Intelligence Oversight Procedures. The use of access controls, user PKI certificates, and the DHS user credentialing data store will ensure identity. If noncompliance is discovered, appropriate disciplinary and corrective actions will be taken.

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?

To ensure that the Cerberus Pilot is working consistently within the confines of law and policy, DHS will set up a governance structure that includes the oversight offices, components providing data, users of the system, and the CIO. The governance process will work through both the CVTF and the Information Sharing and Safeguarding Governance Board. During the Cerberus Pilot, the CVTF will serve as the acting governance body. As such, CVTF will review data service functionality and user authorities during the pilot review/approval process and will oversee the access control definition process. The review and approval of information sharing and access agreements and any uses or access to the system will be reviewed by the CVTF.

Responsible Officials

David Hong
Program Manager
Department of Homeland Security

Approval Signature

Original signed and on file with the DHS Privacy Office
Karen L. Neuman
Chief Privacy Officer
Department of Homeland Security