Privacy Impact Assessment
for the

Common Entity Index Prototype (CEI Prototype)

DHS/ALL/PIA-046

September 26, 2013

Contact Point
Paul Reynolds
Program Manager
DHS/MGMT/OCIO
(202) 617-5068

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

Abstract

The Department of Homeland Security (DHS) Office of the Chief Information Officer (OCIO) is developing a new system called the Common Entity Index Prototype (CEI Prototype). The CEI Prototype will enable DHS to correlate and consolidate a limited set of identity data from select component-level systems and organize key identifiers collected about individual members of the public. The purpose of this prototype is to determine the feasibility of establishing and effectively controlling access to a centralized index of select biographic information, enabling DHS to provide correlated and consolidated identities. This PIA is being conducted because it will use datasets provided by select DHS components containing personally identifiable information (PII) for testing and evaluation purposes.

As the CEI Prototype is a new use of PII collected from members of the public using information technology, DHS is publishing this PIA pursuant to Section 208 of the E-Government Act of 2002. If the system meets the operational criteria during the testing and evaluation stage and DHS transitions the system to operational use, 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 complimentary 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. Presently, data is isolated across DHS components, which limits access and increases retrieval effort and lead time. Even with the requisite authorization, clearance, and safeguarding, DHS personnel currently lack the technical capability to simultaneously query across all DHS databases available to them. Instead, DHS operators must log into multiple databases to conduct routine searches on a particular individual. This manual retrieval process is time intensive, complex, and susceptible to error.

DHS has published a PIA on the overall framework, which is known as the DHS Data Framework.[2] 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 piloting 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

[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.
[2] See DHS/ALL/PIA-046 Data Framework PIA at www.dhs.gov/privacy.

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 that 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.

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.

Prior to deploying an operational system, DHS OCIO is developing a prototype of CEI to evaluate in a test environment using datasets provided by DHS components. The CEI Prototype will have advanced mechanisms that control access to its information based on the attributes of the requestor, the type of data, and the policies governing the information's use and purpose. This correlation and consolidation of identity information, together with appropriate access controls, will enhance DHS's ability to carry out several of its missions, such as quickly identifying individuals who may be a person of interest across DHS components and data sets. DHS is building the CEI Prototype with an initial set of data from U.S. Customs and Border Protection's (CBP) Electronic System for Travel Authorization (ESTA), U.S. Immigration and Customs Enforcement's (ICE) Student and Exchange Visitor Information System (SEVIS), and U.S. Transportation Security Administration's (TSA) Alien Flight Student Program (AFSP). If determined feasible, the CEI Prototype will help determine whether a larger similar system will facilitate and improve DHS's ability to carry out its national and departmental missions.

CEI Prototype

The purpose of the CEI Prototype is to determine the feasibility of establishing a system that correlates and consolidates limited identity information from select DHS source system data sets, resolves differences in the data, and consolidates the data as a more comprehensive identity record about an individual, while including references to the relevant source system records. The CEI Prototype is being tested and evaluated by DHS to determine whether it can deliver a more authoritative biographic reference point on individuals held by DHS using limited number of biographic data elements from the source systems. The resulting correlation and consolidation will be maintained in the CEI Prototype system of records[3] for the period of time the records from the source system are maintained.

In testing the CEI Prototype, DHS seeks to demonstrate two of the elements of the DHS Data Framework: that it can effectively correlate individual records across Departmental databases and that it can enforce effective, dynamic access controls to the data. The CEI Prototype will correlate biographic data such as name, date of birth, country of birth, government-issued document number(s), phone number, physical address, and email address when available in the source systems. This information will be organized into an updated, common record pertaining to a specific individual. The CEI Prototype will also display the mathematical probability of the correlation for each identity that is returned. The mathematical probability of the correlation will be used as part of the testing for effectiveness.

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

As part of the CEI Prototype, the DHS Privacy Office (PRIV), Office for Civil Rights and Civil Liberties (CRCL), Office of Policy (PLCY), and Office of the General Counsel (OGC), in coordination with DHS components, will provide policy recommendations and/or oversight of the correlation process and evaluate the effectiveness of the Prototype. In considering the effectiveness of the CEI Prototype, oversight offices will review how the system generates a match, what minimum elements are required for two records to be matched, how easily a match can be "unmatched," whether CEI can help provide redress, and how authorization for access to the records in CEI will be granted.

These oversight offices will also provide guidance regarding the tagging of source data that is ingested into the CEI Prototype. The Prototype will only ingest source data that has been tagged and stored for access control purposes in the Neptune Pilot.[4] The CEI Prototype will test how well the tags enable access controls to the data. For example, the Neptune Pilot tagged each data element as Core,[5] Extended,[6] or Encounter.[7] The CEI Prototype will only receive core and extended data elements for correlation.

Initially, DHS will use certain biographic data elements and necessary metadata from the following source data sets to populate the CEI Prototype:

(1) CBP's ESTA, covered by the DHS/CBP-009 - Electronic System for Travel Authorization (ESTA) System of Records Notice (SORN) (July 30, 2012, 77 FR 44642);[8]

(2) ICE's SEVIS, covered by the DHS/ICE-001 - Student and Exchange Visitor Information System SORN (January 5, 2010, 75 FR 412)[9] ; and

(3) TSA's AFSP, covered by the DHS/TSA-002 - Transportation Security Threat Assessment System SORN (May 19, 2010, 75 FR 28046).[10]

---

[4] See DHS/ALL/PIA-045 Neptune Pilot for additional information: www.dhs.gov/privacy. Neptune is a sensitive but unclassified (SBU) repository that ingests and tags authoritative mission data.
[5] "Core biographic" data consists of name; date of birth; gender; country of citizenship; and country of birth.
[6] "Extended biographic" data is additional biographic information about an individual who is the subject of a DHS screening, vetting, law enforcement or immigration-related encounter. 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.
[7] "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 (e.g., an application for a DHS administered benefit), or 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 program points of contact.
[8] DHS/CBP-009 - Electronic System for Travel Authorization (ESTA) July 30, 2012, 77 FR 44642.
[9] DHS/ICE 001 - Student and Exchange Visitor Information System January 5, 2010, 75 FR 412.
[10] DHS/TSA 002 - Transportation Security Threat Assessment System May 19, 2010, 75 FR 28046.

These three data sets were identified for the prototype in order to demonstrate how data sets from different components can be correlated while maintaining appropriate access controls. If additional data sets are added to the CEI Prototype, this PIA will be updated.

As part of the tagging design process, the component data/business owners (CBP for ESTA, TSA for AFSP, and ICE for SEVIS), PRIV, CRCL, PLCY, and OGC have provided guidance regarding the tags applied to the data (core, extended, encounter) and the appropriate users and uses of the data.

The CEI Prototype uses technical access control to provide results to a user's query that are based on the user's need to know. These controls take into account user attributes, data tags, and the policies governing the data to enforce appropriate privacy and policy safeguards for the query result. Essentially, the CEI Prototype policy-based access control (PBAC) will assess the attributes of the user (e.g., agency, role) requesting the data, the source of the data (ESTA, AFSP, SEVIS), data tag values (core, extended, encounter), and the purpose of the request (e.g., national security) to determine the type and amount of information to which the user will be provided access. This approach ensures that appropriate legal, privacy, civil rights and civil liberties, policy, and safeguarding requirements for the information are satisfied.

Correlation results obtained during the testing of the CEI prototype will not be used to support operational decisions impacting any individuals. No information will be shared outside of DHS. DHS will publish a new or updated PIA for the CEI Prototype or its operational successor, if the Department decides to deploy the system beyond the Prototype. DHS will also provide further safeguards by developing a process for keeping CEI records synchronized with source systems to provide CEI users with the most accurate and current correlated records, a governance process for adding new data sources, and a redress process for handling incorrectly correlated data—identified either by the individual or by DHS users. If the CEI Prototype does not meet the needs of the Department, all information will be deleted from the CEI Prototype.

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 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.

OCIO operates pursuant to:

- Homeland Security Act, 6 U.S.C. § 343;

- Clinger-Cohen Act of 1996, 40 U.S.C. § 11101, et seq.

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

The CEI Prototype is covered by DHS/ALL-035 Common Entity Index Prototype (CEI Prototype) SORN, published on August 23, 2013.

Source systems are covered by:

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

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?

Each of the source systems from which data elements are ingested is subject to an approved records retention schedule. The ingested data elements that make up correlated identities in CEI are subject to the records retention schedules of the source systems from which they came. The correlated identities themselves are not subject to a separate records schedule, but instead are dynamic records that necessarily change—and are potentially deleted—through adherence to the retention schedules of the source systems. If the system moves from a prototype to an operational system, DHS will work with NARA on a retention schedule that is specific for CEI but continues to recognize that DHS does not retain information in CEI beyond the retention period for the source system.

1.5 If the information is covered by the Paperwork Reduction Act (PRA), provide the OMB Control number and the agency number

[11] DHS/TSA 002 - Transportation Security Threat Assessment System May 19, 2010, 75 FR 28046.

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 CEI Prototype. 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 is 1651-0111, and for AFSP is 1652-0021.

Section 2.0 Characterization of the Information

The following questions are intended to define the scope of the information requested and/or collected, as well as reasons for its collection.

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

The CEI Prototype will correlate core and extended biographic information from three DHS systems. The information consists of:

- Full Name;
- Alias(es);
- Gender;
- Date of Birth;
- Country of Birth;
- Country of Citizenship;
- Phone Number;
- Physical Address;
- Email Address;
- Fingerprint Identification Number; and
- Document Type, Number, Date of Issue, and Location of Issuance for the following types of government issued documents or numbers:
    - Passport;
    - Driver's License;
    - ESTA;
    - SEVIS;
    - Alien Registration; and
    - Visa.

- DHS will use the tags created in the Neptune Pilot and its purpose and use rules to determine access. Those tags include metadata related to the following:
  - The source system name;
  - the system identification number, which ties the biographic information back to the source system record; and
  - the date the record was ingested into the CEI Prototype.

The CEI Prototype will also display the mathematical probability of the correlation for each identity that is returned.

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

The information being ingested by the CEI Prototype was initially collected directly from the individuals at the point of their interaction with DHS or from the school or sponsor based on a request from the individual. The CEI Prototype will receive data from Neptune Pilot, a SBU repository that ingests and tags authoritative mission data. DHS will use data from the following source systems to populate the CEI Prototype:

(1) DHS/CBP-009 - Electronic System for Travel Authorization, July 30, 2012, 77 FR 44642;

(2) DHS/ICE-001 - Student and Exchange Visitor Information System, January 5, 2010, 75 FR 412; and

(3) DHS/TSA-002 - Transportation Security Threat Assessment System, May 19, 2010, 75 FR 28046.

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 CEI Prototype uses neither information from commercial sources nor publicly available data.

2.4 Discuss how accuracy of the data is ensured.

The CEI Prototype does not check for accuracy of the ingested data elements, but instead relies on the source systems (ESTA, SEVIS, and AFSP) to ensure that the information in their systems is accurate. The CEI Prototype displays the mathematical probability of the correlation for each identity that is returned. This is an automated process that will be reviewed by DHS oversight offices and mission operators and fine-tuned by the system administrators to ensure the

system is as effective as possible. As part of the CEI Prototype, the source system data owners, as well as mission subject matter experts, will review the results of sample correlations to determine the accuracy of the correlated identity. If the system has incorrectly correlated two records, there is a process by which that information can be unlinked.

For the duration of the CEI Prototype, data from source systems will be ingested once and not updated. Upon completion of the CEI Prototype for the three data sets, DHS will either delete all information (records and correlated identities) or update this PIA with the description of the means by which the data will remain accurate over time given the planned use of the system.

2.5 Privacy Impact Analysis: Related to Characterization of the Information

Privacy Risk: Correlated identities may provide additional information that a user should not see in the CEI Prototype, thus degrading proper access control and safeguarding of the data.

Mitigation: The CEI Prototype provides results to the query that are based on a user's attributes and associated PBAC in the data sets to foster appropriate legal, privacy, policy, and safeguarding capabilities for the correlated records. The data elements that compose a new record will retain the tags that identify the source systems (ESTA, AFSP, SEVIS) and type of data (core, extended) and can be restricted or displayed based on the access control policies being enforced. PRIV, CRCL, PLCY, and OGC in coordination with the component data providers, will provide policy recommendations and/or oversight of the correlation process.

Privacy Risk: There is a risk that the incorrect tag is applied to a particular data element, which would give users more access than they should receive.

Mitigation: The CEI Prototype and Neptune Pilot are in place to test and ensure that the data elements are properly tagged. PRIV, CRCL, PLCY, and OGC in coordination with the component data providers, will provide policy recommendations and/or oversight of the access and correlation process on an ongoing basis.

Privacy Risk: There is a privacy risk that CEI correlation will be inaccurate and bring two or more records of different individuals together.

Mitigation: As part of the CEI Prototype, DHS will develop standard operating procedures for identifying and de-conflicting, or "un-linking," two or more records. Similarly, DHS will have a governance process for identifying when the correlation may be creating too many inaccurate records. The governance process includes both DHS oversight offices and the relevant operational components. Additionally, results from the CEI Prototype will not be used to support any operational decisions and will have no impact on the individual. Because this is a

pilot, any inaccurate matches will be informative and help DHS make improvements on the system.

Privacy Risk: There is a risk that the CEI Prototype will identify inaccurate or inconsistent information about an individual across two different systems.

Mitigation: For the CEI Prototype, DHS will not be updating records with new or corrected information because it is using data as a snap shot in time. Before the CEI Prototype can become operational, DHS will develop standard operating procedures for identifying possibly inaccurate information across data sets and having the information updated at the source system.

Privacy Risk: The CEI Prototype uses a snapshot of data from one point in time, and this data will be out of date as soon as it is loaded into the CEI Prototype.

Mitigation: The CEI Prototype will not be used for any operational purposes. Visual cues/banners to this effect will be clearly displayed on the CEI Prototype to remind testers. Prior to becoming an operational system, the Department would need to address how it keeps the system accurate and up-to date.

Section 3.0 Uses of the Information

The following questions require a clear description of the project's use of information.

3.1 Describe how and why the project uses the information.

The purpose of the CEI Prototype is to determine the feasibility of establishing a centralized index of select biographic information that will enable DHS to create and control access to consolidated and correlated identity records and provide those records to authorized individuals for facilitating DHS's national security, law enforcement, and benefits missions. The CEI Prototype will use information solely for testing and evaluation purposes at this time.

In considering the effectiveness of the CEI Prototype, oversight offices will review the correlation performance, the data elements required to improve correlation, the process for de-conflicting or "un-linking" a correlated identity, the redress benefits derived from CEI, and the effectiveness of the CEI Prototype data access controls. Similarly, the components will review the value of the CEI Prototype.

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.

No, the CEI Prototype will not perform predictive analysis.

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

ICE, TSA, CBP, the DHS Office of Intelligence and Analysis (I&A), and the DHS Office of Operations Coordination and Planning (OPS) will participate in user testing to validate that the prototype works as designed, has value for operational use, and will help mission operators.

3.4 Privacy Impact Analysis: Related to the Uses of Information

Section 4.0 Notice

The following questions seek information about the project's notice to the individual about the information collected, the right to consent to uses of said information, and the right to decline to provide information.

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

DHS provided notice in the DHS/ALL-035 Common Entity Index Prototype SORN, 78 FR 52553, (published on August 23, 2013), in the Federal Register and in this PIA. For the data in the CEI Prototype, individuals receive notice from the source systems when they provide their information. The CEI Prototype will retain the rules and policies for authorized use of the data from any source system.

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 CEI Prototype.

4.3 Privacy Impact Analysis: Related to Notice

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

Mitigation: DHS provides notice to individuals through this PIA and the CEI Prototype SORN, which serve as public notice of the existence of the CEI Prototype, the data collected and maintained, and the routine uses associated with the information collected. The information is

used only for the purposes described in the public notice of this PIA. Additionally, as part of the prototype and piloting process, DHS is assessing whether additional notice should be provided in the source systems of records. This decision will be made prior to any system becoming operational.

Section 5.0 Data Retention by the project

The following questions are intended to outline how long the project retains the information after the initial collection.

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

DHS is testing and evaluating the CEI Prototype. As part of the prototype, DHS is evaluating the system's effectiveness. If the CEI Prototype is not successful, then all records will be deleted at the conclusion of the test.

If the CEI Prototype is successful, an assessment will need to be made with the system owners as to the disposition of the data. If CEI were to become an operational system, the system is designed to retain the data elements based on retention guidelines of the source system, and an appropriate retention schedule would be developed with NARA. Based on these retention guidelines, the CEI Prototype will take the 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 AFSP data is retained as follows: (1) 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 the individual is an actual 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.

5.2 Privacy Impact Analysis: Related to Retention

Privacy Risk: The CEI Prototype might retain data for longer than it would be retained in the source system or for longer than is necessary for this project.

Mitigation: Data elements have been tagged based on the retention schedules for the source system. Based on these schedules, the CEI Prototype will take the appropriate action to handle the data by either archiving or retaining it.

Section 6.0 Information Sharing

The following questions are intended to describe the scope of the project information sharing external to the Department. External sharing encompasses sharing with other federal, state and local government, and private sector entities.

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 CEI Prototype does not share information outside of DHS as part of the normal agency operations, and the CEI Prototype has no mission/operational/production use at this time.

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

The CEI Prototype does not share information outside of DHS as part of the normal agency operations and has no mission/operational/production use at this time.

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

The CEI Prototype is a prototype that has no mission/operational/production use at this time. If a disclosure needs to be made outside of the Department pursuant to the limited routine uses set forth in the CEI Prototype SORN, DHS OCIO will keep a record of any disclosures.

6.4 Privacy Impact Analysis: Related to Information Sharing

Privacy Risk: There is a risk that combining data will lead to more access to unauthorized users.

Mitigation: The CEI Prototype does not share information outside of the Department except in limited circumstances outlined in the published SORN. The CEI Prototype has no mission/operational/production use at this time. Before CEI would move in to a production phase, the CEI Prototype would implement dynamic access controls and audit logging to prevent information from being shared with users who do not have the proper authorization or a need to know.

Privacy Risk: There is a risk that the disclosures of records will not be accurately maintained.

Mitigation: The CEI Prototype is a prototype that has no mission/operational/production use at this time. If a disclosure needs to be made outside of the Department pursuant to the limited routine uses set forth in the CEI Prototype SORN, DHS OCIO will keep a record of any disclosures. If the system goes operational, DHS will identify the most effective means for logging such disclosures.

Section 7.0 Redress

The following questions seek information about processes in place for individuals to seek redress which may include access to records about themselves, ensuring the accuracy of the information collected about them, and/or filing complaints.

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

Based on the CEI Prototype SORN, individuals seeking notification of and access to any record contained in this system, or seeking to contest its content, may submit a request in writing to the DHS Headquarters Freedom of Information Act (FOIA) Officer. Instructions for filing a FOIA or Privacy Act request are available at http://www.dhs.gov/foia.

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

In order to correct inaccurate or erroneous information received from the authoritative source system by CEI Prototype, the individual requester should contact the DHS Headquarters FOIA Officer with the proposed corrections. The CEI Prototype team will review the request to determine whether the inaccuracy is in the source data or in the correlation data and will coordinate with the source data owners to determine how to address the proposed correction consistent with the Privacy Act.

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

Individuals are notified concerning procedures for correcting their information through this PIA and the CEI Prototype SORN. If an individual would like to request a change to a record, he or she should contact DHS Headquarters FOIA and DHS will decide how to address the proposed correction consistent with the Privacy Act.

7.4 Privacy Impact Analysis: Related to Redress

Privacy Risk: Individuals may not be able to address redress requests in the CEI Prototype since this is a prototype.

Mitigation: Based on the CEI Prototype SORN, individuals seeking notification of and access to any record contained in this system, or seeking to contest its content, may submit a request in writing to the Headquarters FOIA Officer.

As part of the CEI Prototype, DHS will identify how to handle redress requests effectively and efficiently and create standard operating procedures prior to CEI becoming an operational system.

Section 8.0 Auditing and Accountability

The following questions are intended to describe technical and policy based safeguards and security measures.

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

The CEI Prototype provides results to a query based on a user's attributes and associated PBAC for the search. The access controls apply and execute appropriate legal, privacy, policy, and safeguarding capabilities for the new record. PRIV, CRCL, PLCY, and OGC, in coordination with the component data providers, will provide policy recommendations and oversight of the correlation process.

The CEI Prototype is also implementing audit logging so that user requests and the results returned to those requests will be logged and will include date and timestamps of these transactions. The CEI Prototype will provide read-only access so that source data cannot be changed in the CEI Prototype.

The data logs will provide information on what is being accessed for oversight purposes. Additionally, the dynamic access controls will limit the data that is viewed and the users who are permitted to view it.

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

The CEI Prototype project team members working on the Prototype, including members from OPS, I&A, CBP, TSA, and ICE, will be provided specific privacy training on the CEI Prototype. If CEI were to be made operational, personnel would be required to take source system (AFSP, ESTA, SEVIS) training, as well as DHS privacy and security training. Before user access is granted or any information is provided, the CEI Prototype will also test the effectiveness of rules intended to enforce the proper training (e.g., privacy, security) requirements based on a user's role and the data that he or she is trying to access.

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

A final user list will be created for the CEI Prototype and will include roles based on the access control requirements collected from the different data owners (ESTA, SEVIS, and AFSP). Individuals participating in the Prototype will be allowed to see certain data elements based on the authorized purpose of their search, the type of search, the value of the assigned data tags, and the policies defined in the CEI Prototype.

The CEI Prototype follows appropriate security measures to permit only appropriate personnel access to data residing in its database.

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?

The CEI Prototype is for testing and evaluation purposes only, and does not have mission/operational/production use at this time. Data sets are being acquired from ESTA, SEVIS and AFSP for the purposes of testing and evaluation.

Responsible Officials

Donna Roy

Executive Director

Information Sharing Executive Office

Department of Homeland Security

Approval Signature

Original signed on file at the DHS Privacy Office.

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