Privacy Impact Assessment Update
for the

DHS Data Framework

DHS/ALL/PIA-046(a)

August 29, 2014

Contact Point
Paul Reynolds
Data Framework Program Management Office
Department of Homeland Security
(202) 447-3000

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

Abstract

The DHS Data Framework ("Framework") is a scalable information technology program with built-in capabilities to support advanced data architecture and governance processes. The Framework is DHS's "big data" solution to build in privacy protections while enabling more controlled, effective, and efficient use of existing homeland security-related information across the DHS enterprise and with other U.S. Government partners, as appropriate. Currently, the Framework includes the Neptune and Cerberus systems, and the Common Entity Index. Between November 2013 and August 2014, DHS deployed a pilot/prototype to test different capabilities needed to implement the Framework. After the successful completion of the pilot/prototype phase, DHS now intends to mature the Framework by entering into the next phase—limited production capability. DHS is updating the original Framework Privacy Impact Assessment to reflect this transition to limited production capability.

Introduction

In a Privacy Impact Assessment (PIA) published on November 6, 2013, the Department of Homeland Security (Department or DHS) previously described the Department's development of the Framework. The *DHS Data Framework Overview* in that PIA is summarized here for ease of reference, followed by a description of the *DHS Data Framework Pilot/Prototype Phase*, and an explanation of the next phase in the maturation of the project.

DHS Data Framework Overview

1. Background

The Department's primary mission 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 the Department's 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 individuals' privacy rights, civil rights and civil liberties 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 sought to eliminate information firewalls between government agencies by consolidating multiple agencies under DHS.

Since 2007, DHS has operated under the "One DHS" policy,[1] which was implemented to afford DHS personnel timely access to the relevant and necessary homeland security information they need to successfully perform their duties. DHS personnel requesting this 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) ensure adequate safeguarding and protection of the information. The Department's existing architecture for its IT systems and databases, however, is not conducive to effective implementation of the One DHS policy because information exists in multiple separate databases. The result is a technically cumbersome, time-intensive process to determine what information DHS has about a particular individual.

The Secretary of Homeland Security and the DHS Deputy Secretary directed the development of the Framework to automate execution of the One DHS policy through a collaborative effort among the Department's Common Vetting Task Force (CVTF),[2] the Office of the Chief Information Officer, the Office of Policy, the Office of Intelligence and Analysis, the "oversight offices," including the Privacy Office, the Office for Civil Rights and Civil Liberties, the Office of the General Counsel, and DHS's operational components.

2. Objective

The Framework will create a systematic repeatable process for providing controlled access to DHS data across the Department. The Framework will enable the implementation of efficient and cost-effective search and analysis across DHS databases in both classified and unclassified domains. The searches will identify key DHS data associated with an individual or identifier. Adhering to the 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 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 the Framework's goal, DHS created two central repositories for DHS data: Neptune and Cerberus. Neptune serves as the repository in the unclassified domain. Cerberus resides in the Top Secret/Sensitive Compartmented Information domain. Through

[1] *DHS Policy for Internal Information Exchange and Sharing*, February 1, 2007.
[2] The CVTF is a Department-wide task force comprised of representatives from support and operational components dedicated to improving the efficiency of DHS's screening and vetting activities.

these systems, DHS applies appropriate safeguards for access and use of DHS data and delivers search and analytic capabilities.[3]

The Framework defines 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 based on the type of data involved, the authoritative system from which the data originated, and when it was ingested into the Framework;
(3) **Context** combines what type of search and analysis can be conducted (function), with the purpose for which data can be used (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 and/or Components.

The Framework uses the dynamic access control policies to enable the automated enforcement of access requirements so that a user sees only the information that he or she would otherwise be entitled to view as a matter of law and policy. The Framework includes these elements and related processes to ensure: (1) accurate data tagging; (2) data integrity as data is copied and transferred from its original location; and (3) enforced access control policies. The Framework also enables the Department to log user activities to aid audit and oversight functions.

Phase I: Framework Pilot/Prototype

Earlier this year, the Department successfully completed testing the initial Framework capabilities through the Neptune Pilot, Cerberus Pilot, and Common Entity Index (CEI) Prototype.[4] The Department used three data sets in the pilot/prototype phase: the U.S. Customs and Border Protection's (CBP) Electronic System for Travel Authorization (ESTA), the U.S. Immigration and Customs Enforcement's (ICE) Student and Exchange Visitor Information System (SEVIS), and the Transportation Security Administration's (TSA) Alien Flight Student Program (AFSP). The data sets were copied from the relevant component IT system, transferred into the Neptune platform and tagged, and then the tagged data elements were pushed to the CEI and Cerberus platforms. The pilot/prototype phase successfully demonstrated important foundational elements of the Framework, including, but not limited to those capabilities

[3] During limited production capability, the search and analytic capabilities will be limited to the three basic search functions deployed in the pilot/prototype phase: person search, characteristic search, and trend search.
[4] The PIAs for these pilots and prototype are published at http://www.dhs.gov/privacy-documents-department-wide-programs.

described below. More information on each of these capabilities is described in the Neptune and Cerberus PIAs.[5]

- **Data Management and Transfer –** The pilot/prototype phase demonstrated that Neptune could ingest the data from the three data sets, apply access control tags and relevant metadata, and transfer the tagged data to the Common Entity Index and Cerberus. The data tags identified the type of data involved, when the data originated, when it was ingested as authoritative mission data, and whether the data elements are designated as core, extended, or encounter.[6]

- **User Authentication and Attributed-Based Access –** The pilot/prototype phase demonstrated users could be authenticated with appropriate certificates and that their attributes were properly set with predetermined functions and purposes. Upon login, a user's attributes were retrieved from an attribute authority.[7] Where a user had more than one function and purpose (i.e., the user needed access to data while acting in different capacities), the user was able to select the appropriate functions and purposes for accessing data. Once a user was positively authenticated, the user would have access to the Cerberus system and could request data.

- **Policy-Based Access Control –** The pilot/prototype phase demonstrated that DHS could apply policy-based access controls to determine the type of basic search tools[8] the user could use and what data the user could access. Given a particular user's attributes, an Access Control Server asked what function and purpose the user performs to then determine what privileges the user had. The user's function

[5] *See* DHS/ALL/PIA-046-1(a) Neptune PIA Update and DHS/ALL/PIA-046-3(a) Cerberus PIA Update, published concurrently with this PIA Update.

[6] Core biographic data is basic biographic information, to include name, date of birth, gender, country of citizenship, and country of birth. Extended biographic data is additional biographic information about an individual that is not considered core biographic information, such as address, phone number, email address, passport number, and/or visa number. Encounter data is information that derives from a DHS screening, vetting, law enforcement, or immigration-related event/process and is collected in accordance with DHS authorities and regulations. For more information on these concepts, *see* DHS/ALL/PIA-046-1(a) Neptune PIA Update and DHS/ALL/PIA-046-3(a) Cerberus PIA Update, published concurrently with this PIA Update.

[7] During the pilot/prototype phase, attributes were self-asserted based on pre-defined choices. In a related effort, the Department is developing an authoritative user attribute hub which will include DHS and Intelligence Community user attributes. Once developed, the Framework will employ this authoritative attribute hub, called the Trusted Identity Exchange. The Trusted Identity Exchange will eliminate the need for users to self-assert their attributes. For the pilot/prototype phase, the Department used a Lightweight Directory Access Protocol server as a stand-in for the DHS Trusted Identity Exchange.

[8] Query tools included (1) [Specific] Person or Entity-Based Search; (2) Characteristic-Based Search; and (3) Pattern-Based Search. These queries are described in greater detail in the Fair Information Practice Principles analysis later in the document.

controlled the basic search tools (i.e., the type of query that could be performed) that the user could use. The user's purpose determined which data sets and which type of data (i.e., core, extended, or encounter) the user could access. The Department tested a variety of purpose and function combinations to test whether the Access Control Server gave the user access to the correct tools and data. In each instance, the demonstrations showed that the policy-based controls were appropriately applied and that users only had access to the search tools, data sets, and types of data that they were permitted to access under DHS policy.

- **Audit Logging –** The pilot/prototype phase also demonstrated that DHS could log the application of policy-based controls as it was occurring. The policy decision log showed the policy enforcement when a user requested access and evaluating the policy rules to determine the user's privileges to data or tools. The audit log reader also captured the queries a user made and statistics regarding query results, aiding in audit and oversight, including verification of compliant data usage.

Lessons Learned

The Department intends to mature the Framework in an incremental manner and learned a number of important lessons as a result of the Framework's pilot/prototype phase, including those listed below:

1. Developing a scalable big data architecture means DHS needs to **establish a governance process** to evaluate the integration of new data, new missions, new users, and new analytical tools.

2. **Incremental development** of the Framework allows the Department to deploy new capabilities and then verify that those capabilities comply with legal and policy requirements. This approach allows DHS to ensure that it delivers new capabilities that support DHS's operational mission while protecting privacy, civil rights, and civil liberties. By incorporating these protections from the beginning, the Department is building a sustainable and scalable Department-wide big data architecture.

3. Establishing long-term operational utility and protecting privacy, civil rights, and civil liberties depends on DHS's ability to **refresh and update data** and to **incorporate appropriate redress** mechanisms into the Framework.

4. Conducting **more stakeholder engagement**—with mission operators, system administrators, and data stewards—will facilitate widespread adoption this Department-wide big data solution.

5. **Promoting transparency** will help the public understand how DHS is using its data and support a robust public dialogue on the appropriate use of big data solutions within the U.S. Government. Throughout the pilot/prototype phase, DHS published multiple privacy impact assessments, gave public briefings at the DHS Data Privacy and Integrity Advisory Committee (DPIAC) meetings, and asked the DPIAC for recommendations to further promote transparency.

Applying Lesson Learned 1: Governance and Oversight Process

To address the lessons learned, described above, the Department is establishing a more formal governance structure to develop necessary processes, procedures, and decision-making for the Framework systems, platforms, and uses. Governance includes establishing an executive steering committee under the authority of the Department's Information Sharing and Safeguarding Governance Board.

Part of this governance process includes overseeing the data set selection process. DHS owns numerous data sets, and it became evident during the pilot/prototype phase that choosing the most relevant and appropriate data sets for Framework repositories can be a complicated issue. For this reason, this section will provide an overview of the data set selection process and describe the structure for making selection decisions.

Some of the key goals of a Department-wide big data solution are to reduce the number of times DHS data is replicated throughout the Department; to maintain a high level of data quality and integrity by ingesting data from the original IT system to eventually eliminate the multiple aggregated data sets that exist throughout the Department; and to improve DHS's ability to comply with existing law and policy by identifying individuals who are subject to additional legal or policy requirements (e.g., U.S. Persons, certain special protected classes of aliens). When evaluating which data sets should be included, the Department must keep these goals in mind.

In addition to the governance body, a new Program Management Office will manage the staff level day-to-day implementation of the Framework, including Department-wide enterprise services, standards, processes, and procedures, in close coordination with the DHS oversight and compliance offices based on the mission priorities established by an executive steering committee.

An executive steering committee will act as the primary body for making decisions on what data is ingested and validates the access control rules applicable to each data set. An executive steering committee will also prioritize data sets for ingestion based on mission needs and requirements. Once a data source is identified, an executive steering committee will

convene to examine the risks and benefits of adding the data source. The system owners, in coordination with the Program Management Office, are responsible for presenting all the technical, policy, and security considerations for the ingestion, handling, and deletion of their data.

For each data set, the originating system remains the authoritative source for the data. This means that all rules, policies, and guidelines that are applied to the data source will be applied when the data is ingested. Updates to any rules, policies, and guidelines for that data set will be communicated to an executive steering committee and Program Management Office by the system owner or respective component.

The enhanced governance structure, along with the resulting data selection process, is the most significant change that DHS is making to the Framework since the publication of the Privacy Impact Assessment for the pilot/prototype phase. Changes to the systems to which the Framework will be applied – Neptune and Cerberus – will be described in revised assessments for those systems. The revisions take into account the experience and lessons learned from the pilots for each of these systems.

Applying Lesson Learned 2: Incremental Development

The long-term success of a large-scale, Department-wide initiative of new technology is undoubtedly a complex endeavor. Incremental development affords DHS three key benefits:

- The opportunity to verify that Framework privacy protections are being effectively implemented;
- The time to secure necessary resources and build and deploy the technical capabilities; and
- The ability to receive and incorporate feedback from key operational and public stakeholders as the Framework progresses.

Applying Lesson Learned 3: Redress and Data Refresh

The ability to continuously update data from the original DHS IT system to the Framework is a key capability that must be developed before the Framework initiatives can be fully operational. DHS has publicly declared[9] its intention to develop these capabilities prior to the operational use of data in the Framework, and the limited production capability provides the next step in implementing refresh. To support the development of this long-term capability, the limited production capability will: identify the timelines for refreshing each data set; test DHS's

[9] *See* public briefing on the Framework presented during the DPIAC meeting on September 12, 2013 and January 30, 2014. Available on the DHS Privacy website at: http://www.dhs.gov/dhs-data-privacy-and-integrity-advisory-committee-meeting-information.

ability to refresh each data set; and begin implementation of limited data set refreshes. These refresh timelines will be based on operational need, available resources, and technical capabilities. Limited production capability will start with an initial bulk data ingest of the source systems (i.e., a 'snapshot' in time) into Neptune, followed by increased data updates as the limited production capability progresses and DHS tests its ability to refresh each data set. The goal is to have regular refreshes of data by the end of the calendar year, according to the refresh timelines established for each data set.

Applying Lesson Learned 4: Stakeholder Engagement

As mentioned above, increased stakeholder engagement is an integral part of ensuring the Framework meets operators' needs and is widely adopted throughout the Department. The limited production capability will allow select DHS operators to continue their evaluation of the Framework systems and provide valuable feedback to support the long-term success of the Framework.

Applying Lesson Learned 5: Transparency

A robust transparency initiative is critical to the success of the Framework. Accordingly, DHS has undertaken a variety of activities to promote transparency regarding the Framework outside of the traditional privacy compliance documentation process. DHS gave a presentation about the Framework as part of the White House Big Data Review.[10] DHS has provided two public briefings on the Framework during meetings of its Federal Advisory Committee, the DHS DPIAC.[11] DHS plans to continue its public briefings at DPIAC meetings as the Framework progresses. Finally, DHS has tasked the DPIAC with developing recommendations regarding how DHS can further provide transparency into the Framework.

Phase II: Limited Production Capability

Based on the Framework's success to date, the Department is moving from the pilot/prototype phase to a limited production capability for both the Neptune and Cerberus systems.[12] During limited production capability, DHS will test the ability to refresh data from

[10] *See* the White House 90-Day Review for Big Data website for more information. Available at: http://www.whitehouse.gov/issues/technology/big-data-review.
[11] *See* the DHS Privacy website for archived meeting materials. Available at: http://www.dhs.gov/dhs-data-privacy-and-integrity-advisory-committee-meeting-information.
[12] The Department currently is not entering into a new phase of development with regard to Common Entity Index.

the original DHS IT system to the Framework. DHS has publicly declared[13] its intention to develop these capabilities prior to the operational use of data in the Framework, and the limited production capability provides the next step in implementing refresh.

The limited production capability shares many of the pilot/prototype phase conditions, except that limited production capability provides for limited evaluation in the operational environment. For example, the data elements the source systems transferred to Neptune for ingestion remain the same as in the Neptune Pilot. The data tags remain the same except that Neptune will be adding metadata tags to support future access rules related to the sensitivity of or ability to release the data (i.e., data about persons who receive additional protections, such as U.S. Person minimization pursuant to Executive Order 12333 or the non-disclosure provisions of 8 U.S.C. § 1367 for certain special protected classes of aliens. In addition, limited production capability will introduce data quality processing that will validate ingest tagging, "common schema mapping" or common information fields used across the data sets, (e.g., a last name is mapped to a first name), and will generate data quality metrics for performance and compliance reporting.

Because the Framework controls are still maturing, policy-based access to the integrated data and tools will remain limited to the combination of user functions and purposes defined for the pilot/prototype phase. The data sets and basic search tools will also remain the same. Specifically, the search tools used during limited production capability are the three basic search functions deployed in the pilot/prototype phase: person search, characteristic search, and trend search. Similarly, limited production capability will involve a pre-approved number of users from selected components, no new information sharing roles from those previously established, and limited production capability respects current access controls embedded in operational systems. There will be no users external to DHS.

Limited production capability will include an initial bulk data ingest of the source systems (i.e., a 'snapshot' in time) into Neptune, followed by increased instances of data refreshes as the limited production capability progresses. The Framework Program Management Office and the source system owners are actively exploring ways in which to establish continuous update mechanisms to ensure that Framework systems and users have accurate, relevant, timely, and complete data, but there will be some data latency – i.e., the period between data refreshes – in the limited production capability phase. The goal is to have regular updates by the end of 2014. Cerberus users will be trained to understand the risk associated with data latency and to verify information at the source system. Until continuous data updates can be

[13] *See* public briefing on the Framework presented during the DPIAC meeting on September 12, 2013 and January 30, 2014. Available on the DHS Privacy website at: http://www.dhs.gov/dhs-data-privacy-and-integrity-advisory-committee-meeting-information.

accomplished, no operational use of the data will occur without a human review and verification of the information within the source system. To facilitate this human review, data is tagged with the source system and contact information.

During limited production capability, the Department will validate controls in the operational environment and transfer data from the unclassified to the classified domain. As mentioned, the Program Management Office is working with component data stewards to establish source system data delivery to Neptune for initial limited production capability as well as data refreshes agreements, processes, and mechanisms to support continuous updates. The Program Management Office is also working with the components to develop Framework mission use cases to evaluate the Framework's mission use potential and to broaden component review and input for the evolving Framework control elements.

Fair Information Practice Principles (FIPPs)

The Department applies the following Fair Information Practice Principles, developed from the Privacy Act's underlying concepts, to account for the nature and purpose of the information being collected in relation to the Department's missions. While some of the principles analysis remains unchanged from the initial Framework pilot/prototype phase, the privacy impacts resulting from limited production capability require additional analysis.

As described above, the creation of a robust governance structure is a principal means through which the Department intends to enhance the Framework's adherence to the Fair Information Practice Principles and further ensure the proper privacy and civil rights and civil liberties protections are in place for the Framework.

1. Principle of Transparency

Principle: DHS should be transparent and provide notice to the individual regarding its collection, use, dissemination, and maintenance of PII. Technologies or systems using PII must be described in a SORN and PIA, as appropriate. There should be no system the existence of which is a secret.

Privacy Risk: There is a risk that individuals may not be aware their PII is being compared against other DHS information in this DHS-wide big data project.

Mitigation: The existing System of Records Notices for ESTA, SEVIS, and AFSP provide notice to the public that the information may be compared against other data sets and be subject to analysis for DHS's counterterrorism and immigration missions. The ESTA System of Records[14] notes that DHS's purpose of collecting the information includes "vetting [individuals']

[14] *See* DHS/CBP-009 – Electronic System for Travel Authorization (ESTA), July 30, 2012, 77 FR 44642. Available at: http://www.gpo.gov/fdsys/pkg/FR-2012-07-30/html/2012-18552.htm.

information against various security and law enforcement databases and identifying high-risk applicants." The SEVIS System of Records Notice[15] notes one of its purposes is to support "the analysis of information in the system for law enforcement, reporting, management, and other mission-related purposes." The AFSP System of Records Notice[16] lists one of its purposes as "To permit the retrieval of the results of security threat assessments, employment investigations, and evaluations performed for security purposes; including criminal history records checks and searches in other governmental, commercial, and private data systems, performed on the individuals covered by this system." Additionally, DHS is updating this PIA and the PIAs for Neptune and Cerberus to reflect deployment of the limited production capability.

In addition, while the existing privacy documentation may permit the use of individuals' PII in this context, DHS is pursuing ways to provide transparency outside of the traditional privacy documentation process because of the privacy sensitivities surrounding big data technology and use. DHS promoted the Framework as part of the White House Big Data Review,[17] and the Framework is described in the White House's final big data report.[18] DHS has provided two public briefings on the Framework during meetings of its Federal Advisory Committee, the DHS DPIAC.[19] DHS plans to continue its public briefings at DPIAC meetings as the Framework progresses. Finally, DHS has tasked the DPIAC with developing recommendations regarding how DHS can further provide transparency into the Framework.

2. Principle of Individual Participation

Principle: DHS should involve the individual in the process of using PII. DHS should, to the extent practical, seek individual consent for the collection, use, dissemination, and maintenance of PII and should provide mechanisms for appropriate access, correction, and redress regarding DHS's use of PII.

***Privacy Risk*:** There is a risk that an individual will not be able to receive appropriate access, correction, and redress regarding DHS's use of PII.

***Mitigation*:** To mitigate this risk in the long-term, DHS will develop (1) a process to provide an individual with the same access and redress opportunities in the Framework that he or

[15] *See* DHS/ICE-001 – Student and Exchange Visitor Information System, January 5, 2010, 75 FR 412. Available at: http://www.gpo.gov/fdsys/pkg/FR-2010-01-05/html/E9-31268.htm.
[16] *See* DHS/TSA-002 – Transportation Security Threat Assessment System, May 19, 2010, 70 FR 33383. Available at: http://www.gpo.gov/fdsys/pkg/FR-2010-05-19/html/2010-11919.htm.
[17] *See* the White House 90-Day Review for Big Data website for more information. Available at: http://www.whitehouse.gov/issues/technology/big-data-review.
[18] *See* the White House report "Big Data: Seizing Opportunities, Preserving Values," May 2014. Available at: http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf.
[19] *See* the DHS Privacy website for archived meeting materials. Available at: http://www.dhs.gov/dhs-data-privacy-and-integrity-advisory-committee-meeting-information.

she would have in the original DHS IT system[20] and (2) the ability to refresh the data that is ingested into the Framework.

With respect to access and redress, the Program Management Office will employ the formal Framework governance structure to create this permanent process moving forward. The absence of an access and redress process that extends from the original DHS IT system to the Framework is one of the reasons that DHS chose to deploy a limited production capability instead of pursuing full operational use of the Framework.

***Privacy Risk*:** There is a risk that changes made to PII in the underlying DHS IT system as a result of correction and redress will not be replicated into the Framework.

***Mitigation*:** With respect to data correction, DHS must create a process to refresh the data provided from the original DHS IT system to the Framework. One of the main goals of the limited production capability is to identify the timelines for refreshing each data set, test DHS's ability to refresh each data set, and begin implementation of limited data set refreshes. These refresh timelines will be based on operational need, available resources, and technical capabilities. Limited production capability will start with an initial bulk data ingest of the source systems (i.e., a 'snapshot' in time) into Neptune, followed by increased data updates as the limited production capability progresses and DHS tests its ability to refresh each data set. The goal is to have regular refreshes of data by the end of the calendar year, according to the refresh timelines established for each data set.

To help mitigate this risk during the limited production capability, DHS requires users of the Framework to go back to the original DHS IT system and verify that an individual's information has not been updated pursuant to redress or correction before completing any final analysis or using the information operationally. Requiring users to verify information in the original DHS IT system will ensure that any updates pursuant to redress or correction will be incorporated into any final analytical product or before the information is used operationally.

3. Principle of Purpose Specification

Principle: DHS should specifically articulate the authority which permits the collection of PII and specifically articulate the purpose or purposes for which the PII is intended to be used.

Privacy Risk: There is a risk that DHS will include data in the Framework for a purpose other than the purpose for which is was collected in the original DHS IT system.

Mitigation: During limited production capability, DHS users will only use the data for immigration, border security, and counterterrorism purposes. The ESTA, SEVIS, and AFSP

[20] The Framework does not impact an individual's ability opportunity to receive appropriate access, correction, and redress in the original IT system.

System of Records Notices specify that DHS collected the information for these purposes. For example, the ESTA System of Records[21] states that "[t]he purpose of this system is to collect and maintain a record of nonimmigrant aliens who want to travel to the United States under the [Visa Waiver Program (VWP)], and to determine whether applicants are eligible to travel to the United States under the VWP by vetting their information against various security and law enforcement databases and identifying high-risk applicants." The SEVIS System of Records Notice[22] notes that SEVIS allows DHS "to monitor the progress and status of lawfully admitted F/M/J nonimmigrants residing in the United States, to ensure they comply with the obligations of their U.S. admittance…." and that the information may be used "to support other homeland security and immigration activities…." The AFSP System of Records Notice[23] states that the purpose of the system includes the "[p]erformance of security threat assessments, employment investigations, and evaluations performed for security purposes that Federal statutes…" and "the retrieval of information from other terrorist-related, law enforcement, immigration and intelligence databases on the individuals covered by this system."

4. Principle of Data Minimization

Principle: DHS should only collect PII that is directly relevant and necessary to accomplish the specified purpose(s) and only retain PII for as long as is necessary to fulfill the specified purpose(s). PII should be disposed of in accordance with DHS records disposition schedules as approved by the National Archives and Records Administration (NARA).

Privacy Risk: There is a risk that DHS will include more data sets in the Framework than those which is necessary to fulfill the purposes authorized under the Framework.

Mitigation: To minimize this risk, DHS has carefully evaluated each data set to determine whether its use is directly relevant and necessary to accomplish the purposes authorized under the Framework. The pilot/prototype phase demonstrated that these three data sets were effectively used together to support DHS's immigration, border security, and counterterrorism missions. During the limited production capability, DHS will not be including any new data sets in the Framework.

Privacy Risk: There is a risk that the Framework will encourage DHS to replicate data sets across the Department, proliferating data across the Department.

---

[21] *See* DHS/CBP-009 – Electronic System for Travel Authorization (ESTA), July 30, 2012, 77 FR 44642. Available at: http://www.gpo.gov/fdsys/pkg/FR-2012-07-30/html/2012-18552.htm.
[22] *See* DHS/ICE-001 – Student and Exchange Visitor Information System, January 5, 2010, 75 FR 412. Available at: http://www.gpo.gov/fdsys/pkg/FR-2010-01-05/html/E9-31268.htm.
[23] *See* DHS/TSA-002 – Transportation Security Threat Assessment System, May 19, 2010, 70 FR 33383. Available at: http://www.gpo.gov/fdsys/pkg/FR-2010-05-19/html/2010-11919.htm.

Mitigation: An important goal of the Framework is to reduce the number of copies of data sets across the Department. By creating a Department-wide big data solution, DHS will actually reduce the number of copies of data sets across the Department in the long-term. Eventually, some data aggregation systems may be decommissioned as their capabilities are replicated and centralized within the Framework. To implement this mitigation, however, DHS must successfully replicate the capabilities of other systems and build operator support. The limited production capability is the next step in an iterative process toward these goals.

Privacy Risk: There is a risk that data will be retained in the Framework for longer than is allowed in the original DHS IT system.

Mitigation: DHS has determined that the retention period for the original DHS IT system will also apply when that information is ingested into the Framework. To comply with the established retention periods, DHS has tagged data with the time that it was ingested into the original IT systems so that the information can be deleted when the retention period ends.

5. Principle of Use Limitation

Principle: DHS should use PII solely for the purpose(s) specified in the notice. Sharing PII outside the Department should be for a purpose compatible with the purpose for which the PII was collected.

Privacy Risk: There is a risk that Framework users will access more PII than is necessary to accomplish their specified purpose.

Mitigation: One of the hallmarks of the Framework is the ability to restrict access to PII within a particular data set based on the user's specified purpose. To accomplish this, DHS has tagged elements from each data set as belonging to one of three categories—core biographic, extended biographic, and encounter information—and users are only able to access the categories that are necessary to perform their function. This use of data tags allows DHS to minimize data access according to specified purpose, which is an improvement in the implementation of data minimization within the Department.

Privacy Risk: There is a risk that DHS users will use the data for purposes other than those authorized during the limited production capability.

Mitigation: During limited production capability, DHS users will only use the data for immigration, border security, and counterterrorism purposes. As described earlier in the PIA, access to data is determined by a user's purpose and function. The Framework's policy-based controls will ensure that a user is only able to access information that is permitted for a particular purpose and function.

Privacy Risk: There is a risk that the elements of data access and control are insufficiently developed or incorrectly implemented and will fail to limit the use of the data to the purposes authorized for the limited production capability.

Mitigation: The pilot/prototype phase tested the user attributes, tags, and context to verify that the controls performed correctly. During limited production capability, DHS will continue to evaluate the application of these controls. Additionally, DHS provided demonstrations of these controls to subcommittees of the DPIAC and requested recommendations from the DPIAC on what auditing and oversight capabilities DHS could develop to ensure that these controls are not circumvented.

Privacy Risk: There is a risk that DHS will share PII outside of the Department for a purpose that is not compatible with the purpose for which the PII was collected.

Mitigation: Non-DHS users will not have access to the Framework during the limited production capability. Moreover, DHS is not sharing information outside of the Department during the limited production capability.

6. Principle of Data Quality and Integrity

Principle: DHS should, to the extent practical, ensure that PII is accurate, relevant, timely, and complete within the context of each use of the PII.

Risk: There is a risk that PII transferred outside of the original IT system and into the Framework will not be accurate, relevant, timely, or complete.

Mitigation: To mitigate this risk in the long-term, DHS must create a process to refresh the data provided from the original DHS IT system to the Framework, so that updates or corrections are replicated from the original DHS IT system into the Framework. As noted in the "Principle of Individual Participation," one of the main goals of the limited production capability is to identify the timelines for refreshing each data set, test DHS's ability to refresh each data set, and begin implementation of limited data set refreshes. More information on data refresh is provided in the "Principle of Individual Participation" and throughout the PIA.

To mitigate this risk during the limited production capability, Framework users will be trained to understand the risk associated with data latency (due to limited refresh capabilities). Users will also be required to verify information at the source system before completing any final analysis or using the information operationally.

7. Principle of Security

Principle: DHS should protect PII (in all forms) through appropriate security safeguards against risks such as loss, unauthorized access or use, destruction, modification, or unintended or inappropriate disclosure.

Risk: There is a risk the Framework systems will not have appropriate security safeguards.

Mitigation: The Department followed the requirements for information assurance and security and the development of sensitive systems[24] and handling of sensitive information[25] for the Framework systems. Both Framework systems involved in limited production capability, i.e., Cerberus and Neptune, have system security plans and the Chief Information Security Officer's approval for Authority to Operate.

Information will be encrypted and safeguarded during transport and storage, and limited production capability will involve only a limited number of pre-approved users whose access to data, data sets and query tools will be determined based on their authenticated attributes and their predetermined functions and purposes. No new information sharing roles have been established.

8. Principle of Accountability and Auditing

Principle: DHS should be accountable for complying with these principles, providing training to all employees and contractors who use PII, and should audit the actual use of PII to demonstrate compliance with these principles and all applicable privacy protection requirements.

Risk: There is a risk that the use of PII will not be auditable to demonstrate compliance with these principles and all applicable privacy protection requirements.

Mitigation: As part of the pilot/prototype phase, DHS determined that the Framework's audit capabilities were adequate to support an audit of whether personally identifiable information was accessed properly and that the dynamic access controls could sufficiently limit the data that is viewed to the users who are permitted to view it. During limited production capability, the Framework will continue to employ tamper-resistant audit logs, which will also provide metrics for assessing the capture of all successful and unsuccessful attempts to log in, to access information, and other meaningful user and system actions. The audit logs will contain the user name and the query performed, but not the responses provided back.

Additionally, DHS provided demonstrations of the audit log capabilities to subcommittees of the DPIAC and requested recommendations from the DPIAC on what auditing and oversight capabilities DHS could develop to ensure that these controls are not circumvented.

Risk: There is a risk that DHS will not perform reviews of the audit logs to determine compliance with the Framework policies.

Mitigation: During the limited production capability, the PMO will pull a random selection of queries from Framework systems and manually review them to determine

[24] *See* DHS 4300A Sensitive Systems Handbook.
[25] *See* DHS Handbook for Safeguarding Sensitive Personally Identifiable Information.

compliance with the Framework policies. The PMO will present its finding to an executive steering committee, which includes PRIV, CRCL and OGC.

Additionally, to mitigate this risk in the long-term, DHS has tasked the DPIAC with developing recommendations for how DHS can use audit logs in a meaningful way to ensure robust oversight.

Conclusion

DHS developed the Framework specifically to ensure that it is consistently using DHS data for the purposes for which it was collected. There are several privacy risks to the overall Framework that have been mitigated during the pilot phase. This helped demonstrate the ability to effectuate meaningful dynamic access controls. As the Framework continues to mature, this Privacy Impact Assessment will be updated periodically to account for any major changes to the information architecture and data governance.

Responsible Officials

Donna Roy
Office of Chief Information Officer

Clark Smith
Office of Intelligence and Analysis

Approval Signature

Original signed copy on file with the DHS Privacy Office.

_____

Karen L. Neuman
Chief Privacy Officer
Department of Homeland Security