Privacy Impact Assessment Update
for the

DHS Data Framework

DHS/ALL/PIA-046(b)

February 27, 2015

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. Beginning in April 2015, DHS intends to mature the Framework during an Initial Operational Capability phase, which will include new DHS data sets, additional DHS users, and new technical capabilities (e.g., data refresh) for use within a controlled operational context. DHS is updating the Framework Privacy Impact Assessment (PIA) to reflect the transition to this Initial Operational Capability phase.

Introduction

In a Privacy Impact Assessment (PIA) published on November 6, 2013, and a PIA update published August 29, 2014, the Department of Homeland Security (Department or DHS) previously described the Department's development of the Framework.[1] The Framework will create a systematic repeatable process for providing controlled access to DHS data across the Department. 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.

Neptune is the unclassified "data lake," which DHS currently uses to receive, store, and tag the data from unclassified DHS information technology systems. Once tagged, unclassified DHS data sets from Neptune are transferred to Cerberus, which is the classified data lake that DHS currently uses to perform classified searches of unclassified DHS data sets. The Common Entity Index is an unclassified correlation engine that will allow DHS to connect disparate DHS data sets to view all available information about an identified individual.

To ensure appropriate technical and policy governance of the program—including the incorporation of robust privacy, civil rights, and civil liberties protections—DHS is deploying the Framework in an iterative fashion. Below is a summary of the Data Framework phases to date:

[1] DHS/ALL/PIA-046 DHS Data Framework, November 6, 2013, and DHS/ALL/PIA-046(a) DHS Data Framework, August 29, 2014.

- ***Pilot Phase:*** Between November 2013 and August 2014, DHS deployed a Framework Pilot phase to test the mission utility, technical feasibility, and policy protections of the Framework in a non-operational context.[2]
- ***Limited Production Capability Phase:*** Between August 2014 and April 2015, DHS deployed a Limited Production Capability to further test the Framework's capabilities within a controlled operational context.[3]
- ***Initial Operational Capability Phase:*** Beginning in April 2015, DHS intends to mature the Framework during an Initial Operational Capability phase. Initially, the Framework's uses, users, and capabilities (i.e., the basic search functions) will remain the same as during the Limited Production Capability phase. However, during the Initial Operational Capability phase, the Framework will include new DHS data sets, as described in Appendix A, and may add new types of DHS users and new technical capabilities (e.g., increased data refresh capabilities) for use within a controlled operational context. The search and analytic capabilities continue to be limited to the three basic search functions deployed in the pilot/prototype and Limited Production Capability phase: person search, characteristic search, and trend search. DHS is updating the Framework PIA to reflect the transition to this Initial Operational Capability phase.

In order to achieve the Framework's ultimate goals, 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 these systems, DHS applies appropriate safeguards for access and use of DHS data and delivers search and analytic capabilities.

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;[4]

[2] For more information about the Pilot phase, please see the following privacy impact assessments: DHS/ALL/PIA-046 DHS Data Framework, November 6, 2013; DHS/ALL/PIA-046-1 Neptune Pilot, September 25, 2013; DHS/ALL/PIA-046-2 Common Entity Index Prototype, September 26, 2013; and DHS/ALL/PIA-046-3 Cerberus Pilot, November 22, 2013.

[3] For more information about the Limited Production Capability phase, please see the following privacy impact assessments: DHS/ALL/PIA-046(a) DHS Data Framework, August 29, 2014; DHS/ALL/PIA-046-1(a) Neptune Pilot, August 29, 2014; and DHS/ALL/PIA-046-3(a) Cerberus Pilot, August 29, 2014.

[4] Neptune currently ingests data as collected by unclassified DHS systems and transfers that data to Cerberus within an approved tagging scheme that includes Core and Extended Biographic information and Encounter information related to individuals. Core Biographic data is basic biographic information, to include name, date of birth, gender,

(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.

Framework Pilot and Limited Production Capability Phases

Between November 2013 and August 2014, the Department successfully completed testing initial Framework capabilities through the Neptune Pilot, Cerberus Pilot, and Common Entity Index (CEI) Prototype.[5] In August 2014, the Framework began a Limited Production Capability phase, which DHS will conclude in April 2015.[6] During the Limited Production Capability phase, DHS established the ability to manually refresh data from the original DHS IT system to the Framework. DHS has publicly declared its intention to develop these capabilities prior to the operational use of data in the Framework, and the Limited Production Capability provided the next step in implementing refresh.

Updated Governance and Oversight Process

As noted in the last PIA, DHS recognized the need for a robust governance structure. To address this need, the Department established a more formal governance structure to develop and implement necessary processes, procedures, and decision-making for the Framework systems,

---

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. The three attributes categories were defined as part of the Department's enterprise data attribute and tagging definition process. During the Initial Operational Capability phase, all integrated data sets will be mapped to the approved tag categories (e.g., Core, Extended, and Encounter) to ensure the developed policy controls are enforced.

[5] Please see the following privacy impact assessments: DHS/ALL/PIA-046 DHS Data Framework, November 6, 2013; DHS/ALL/PIA-046-1 Neptune Pilot, September 25, 2013; DHS/ALL/PIA-046-2 Common Entity Index Prototype, September 26, 2013; and DHS/ALL/PIA-046-3 Cerberus Pilot, November 22, 2013.

[6] Please see the following privacy impact assessments: DHS/ALL/PIA-046(a) DHS Data Framework, August 29, 2014; DHS/ALL/PIA-046-1(a) Neptune Pilot, August 29, 2014; and DHS/ALL/PIA-046-3(a) Cerberus Pilot, August 29, 2014. The Limited Production Capability phase did not include the Common Entity Index.

platforms, and uses.[7]

The Department is establishing the DHS Data Framework Steering Group (DFSG), an executive steering committee, with a charter approved by the Secretary. The charter defines the mission, authority, membership, responsibilities, and operating principles for the DFSG. The mission of the DFSG is to provide effective governance, oversight, coordination, and direction to the Framework and all related projects and initiatives and to ensure its successful and timely delivery in compliance with all policy- and user-based requirements. In addition to DHS mission and technical representatives, the DFSG's membership includes "oversight offices," i.e., the Privacy Office, the Office for Civil Rights and Civil Liberties, and the Office of the General Counsel.

The Under Secretary for Intelligence and Analysis chairs the DFSG. The Department's Chief Information Officer serves as the Vice Chair with a rotating Co-Vice Chair from one of the Department's operational components chosen in a predetermined order of rotation for a two-year term. Because of the number of U.S. Customs and Border Protection (CBP) data sets incorporated into the Framework, an official from CBP serves as the initial Co-Vice Chair.

Active and continued enhancement of the governance structure is among the most significant changes that DHS is making to the Framework since the publication of the PIA for the Limited Production Capability phase. By design, the DFSG brings together expertise in a number of areas including mission operations, information technology, and oversight to ensure the Framework delivers mission capabilities while ensuring all legal, policy, technical, security, privacy, civil rights, and civil liberties protections and requirements are met.

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 its intention to develop these capabilities before implementing the operational use of data in the Framework,[8] and the Limited Production Capability provided the next step in implementing refresh by demonstrating DHS's ability to manually update data from the underlying IT systems. To support the development of this long-term capability, the Program Management Office and data providers identified the timelines for refreshing each data set; tested DHS's ability to manually refresh data sets; and began planning for implementation of data set refresh automation.

As new data is brought into the Framework during the Initial Operational Capability phase, manual data refresh timelines will be identified as part of the onboarding process. These

[7] *Please see* DHS/ALL/PIA-046(a) DHS Data Framework, August 29, 2014.
[8] *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.

timelines will be based on operational need, available resources, and technical capabilities. The goal is to have regular manual refreshes (e.g., monthly) of data according to the refresh timelines established for each data set. During the Initial Operational Capability, it is unlikely that DHS will have near real-time data refresh. Consequently, DHS has developed user training to mitigate the impact of the ongoing data latency (due to limited refresh capabilities). Users are trained to verify information at the source system before completing any final analysis or using the information operationally. To facilitate human review and verification at the source IT system before operational use, the Department included source system contact information in the data tagging.

Initial Operational Capability Phase

Summarized below are the expanded capabilities for the Initial Operational Capability. No new capabilities will be deployed until they have been formally approved by the Framework governance structure, including the oversight offices. If approved, new capabilities will be described in updates to this PIA.

Data Sources

During the Initial Operational Capability phase, the Framework will continue to ingest new data sets from Department IT systems. In the long-term, the Department plans to ingest three to five data sets each year, with a total of 20 to 24 data sets added over the next several years. The timeline for approving and ingesting new data sets will depend on the technical, legal, and policy complexities of each data set. To provide transparency to the public during this aggressive data integration timeline, DHS has developed Appendices to this PIA to fully detail the data sources, users, and uses of data within the Framework. These Appendices will be updated to memorialize the DFSG approvals and provide transparency to the public prior to any changes to the Framework.[9] Please see Appendix A for a list of the data sources that have been approved for inclusion in the Framework.

The Department developed a repeatable onboarding plan during the Pilot and Limited Production Capability phases. During the Initial Operational Capability phase, the Department will codify the process for engaging, prioritizing, documenting, vetting, and then loading new data sets. This onboarding process, governed by the DFSG, will include the completion of Privacy Threshold Analyses, which are the Department's internal documents to assess whether

[9] Appendix A provides detailed information regarding each data set currently integrated in the Framework and known pending data sets that will be added during the Initial Operational Capability phase. Appendix B provides detailed information regarding the Framework's current approved uses. Appendix C provides detailed information regarding the current Data Framework users. As DHS approves additional Framework uses and users, Appendix B and C will be updated to memorialize the DFSG approvals and provide transparency to the public prior to any changes to the Framework.

changes to PIAs or System of Records Notices (SORN) are required.

Data Updates, Corrections, Deletions, and Refresh

During the Initial Operational Capability phase, the Framework will continue to rely on the source IT systems to notify the Framework of updates and corrections to the data sets. Updates and corrections to the data sets will be incorporated into the Framework after each manual data refresh. Because of the lack of automated, near real-time refresh, there will be a delay between when updates or corrections are made in the source IT system and when those updates or corrections are incorporated into the Framework. Any corrections or changes to the data will happen at the source IT system, and will be incorporated into the Data Framework by the Program Management Office during the subsequent data refresh.

If a source IT system changes any of the rules, policies, or guidelines for a data set, then the source IT system owner or DHS Component will be responsible for communicating those changes to the Program Management Office so that the Framework access control rules can be updated accordingly. The source IT system owner or DHS Component will also be responsible for communicating those changes to the DFSG for its awareness.

Until the Department establishes a near real-time data refresh capability, DHS personnel will not use data from the Data Framework without verifying the data in the underlying source IT system. This extra step is a privacy protection that ensures data quality and an operational protection to ensure that DHS personnel are using accurate information in DHS operations.

Data will be maintained according to the retention, use, and handling provisions of the respective SORNs for those mission systems of records. Because DHS is relying on the source IT systems to notify the Framework of changes, deletions, or corrections to data, DHS will not delete data until it receives a deletion notification from the source IT system. (Note: "Deletions" will be applied as defined by the source IT system. This may mean that data is overwritten, masked, fully removed, marked as "inactive," or archived, etc.). As part of the metadata tagging process, DHS tags each data set with a retention period, and therefore DHS can remind the underlying source IT system of the upcoming retention expiration date if the Framework has not already received a deletion notification.

Users and Uses

During the Initial Operational Capability, DHS will expand the users of the Data Framework to the DHS Intelligence Enterprise. The uses will remain limited to counterterrorism, border security, and immigration. The DHS Intelligence Enterprise[10] consists of the intelligence offices of the following DHS Components:

[10] More information about the DHS Intelligence Enterprise is *available at* http://www.dhs.gov/more-about-office-intelligence-and-analysis-mission.

- Customs and Border Protection;
- Immigration and Customs Enforcement;
- U.S. Citizenship and Immigration Services;
- U.S. Coast Guard;
- Transportation Security Administration;
- U.S. Secret Service; and
- Federal Emergency Management Agency.

As the Framework matures, DHS will evaluate whether to expand the authorized users and uses. New uses and users must be approved by the DFSG, including oversight offices. New DHS users and uses will be included in Appendix B of this PIA.

The Framework is designed first and foremost to support DHS users' mission needs while mitigating privacy risk. However, DHS recognizes that the Framework provides a controlled process for meeting some of its interagency information sharing needs. If DHS opens the Framework to include non-DHS users or uses, then DHS will update the body of this PIA to reflect that expansion and analyze the privacy risks and mitigations associated with that expansion. DHS will list the new non-DHS users and uses in Appendix B of this PIA.

Data Processing and Delivery

During Limited Production Capability, the Department developed "cross domain guards" to automate the secure transfer of Neptune data from the unclassified domain to Cerberus in the classified domain. During the Initial Operational Capability phase, the Department also will work to automate transfer from source systems to Neptune as each of the source systems can support automated transfer.

Program Management

As directed by the DFSG, the Program Management Office will continue its stakeholder outreach efforts during the Initial Operational Capability phase. In addition, the DFSG will develop a data source and mission prioritization process to ensure that the expansion of Framework uses appropriately meet Departmental and operational priorities. As directed by the DFSG, the Program Management Office will implement the data source and mission prioritization process. As directed by the DFSG, the Program Management Office will also develop and implement Framework performance monitoring to ensure the Framework achieves set goals and objectives.

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 phases, the privacy impacts resulting from deploying operational capabilities 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, 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 personally identifiable information (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 SORNs for data sets incorporated into the Framework during the Pilot and Limited Production Capability phases 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. As DHS adds new data sets to the Framework, the Privacy Office will review the applicable SORNs to see if they provide appropriate notice. For example, to support the ingestion of some data sets in the Initial Operational Capability phase, DHS is publishing SORN updates to notify the public that the unclassified data will also be replicated to the classified network. Additionally, DHS is updating this PIA and the PIAs for Neptune and Cerberus to reflect the transition to the Framework's Initial Operational Capability phase and the incorporation of new data sets. DHS will continue to review its PIAs when it adds new data sets or capabilities to the Data Framework.

Despite the existing mitigations, DHS recognizes that the long-term success of the Framework depends on robust transparency. Accordingly, 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,[11] and the Framework is described in the White House's final big data report.[12] DHS has provided two public briefings on the Framework during meetings of

[11] *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.
[12] *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.

its Federal Advisory Committee, the DHS DPIAC.[13] DHS plans to continue its public briefings at DPIAC meetings as the Framework progresses. Finally, DHS will analyze the DPIAC's recommendations regarding how DHS can further provide transparency into the Framework and implement them as appropriate.

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: This risk remains during the Initial Operational Capability and will remain until: (1) DHS has near real-time refresh and (2) DHS a process to provide an individual with the same access and redress opportunities in the Framework that he or she would have in the original DHS IT system.[14] To partially mitigate this risk, users of the Framework must verify the accuracy of the data in the source IT system before using the data operationally. Also, during the Initial Operational Capability phase, the Program Management Office will draft a Data Quality Plan to begin discussions within the Department about a more robust redress process.

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: This risk remains during the Initial Operational Capability and will remain until DHS has near real-time refresh. Refresh during the Initial Operational Capability continues to be a manual process. Refresh timelines outlined in Appendix A are established by operational need, available resources, and technical capabilities. To help mitigate the risk due to these manual refreshes, 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 issuing any raw intelligence (e.g., intelligence information report) or 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 product or before the information is used operationally.

Also, during the Initial Operational Capability phase, the Program Management Office will draft a Data Quality Plan to begin discussions within the Department about a more robust

[13] *See* the DHS Privacy website for archived meeting materials, *available at* http://www.dhs.gov/dhs-data-privacy-and-integrity-advisory-committee-meeting-information.
[14] The Framework does not impact an individual's ability opportunity to receive appropriate access, correction, and redress in the original IT system.

redress process. The Program Management Office will also continue its technical development and work with data providers to increase the timeliness and automation of data refresh.

3. Principle of Purpose Specification

Principle: DHS should specifically articulate the authority that 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: Until the DFSG approves new uses based on the onboarding process and mission use case methodology, described above, DHS users will only use the data for immigration, border security, and counterterrorism purposes. The SORNs for the existing data sets within the Framework specify that DHS collected the information for these purposes. See Appendix A for information about current and pending data sets to be added to the Framework during the Initial Operational Capability phase. DHS is publishing SORN updates to account for the classified storage of the data, as needed. Any changes to the data sets, users, and uses will trigger a review to determine whether the purpose remains compatible and whether this risk is impacted by the addition of new data sets, uses, or users.

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 that are necessary to fulfill the purposes authorized under the Framework.

Mitigation: To minimize this risk, DHS will continue to carefully evaluate each data set to determine whether its use is directly relevant and necessary to accomplish the purposes authorized under the Framework. The Department estimates the Framework will ingest at least three-five data sets per year with a total of 20-24 data sets added over the next several years. Appendix A will be updated continually to document information about current and pending data sets to be added to the Framework. DHS is publishing SORN updates to account for the classified storage of the data, as needed.

For each new data set added during the Initial Operational Capability, the Department will employ the onboarding process for engaging, prioritizing, documenting, vetting, and then loading new data sets. This onboarding process, governed by the DFSG, will be comprehensive, to include an assessment of data minimization under DHS privacy policy and the development of Privacy Threshold Analyses to ensure formal consideration of PIA and SORN update

requirements of the source systems, as well as access control rules or user types that need to be reviewed, modified, or created.

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

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. The number of system data delivery methods will decrease as the Framework can provide a more controlled release and transfer of information from the Department. 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.

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 apply when that information is ingested into the Framework. Because DHS is relying on the source IT systems to notify the Framework of changes, deletions, or corrections to data, DHS will not delete data until it receives a deletion notification from the source IT system. (Note: "Deletions" will be applied as defined by the source IT system. This may mean that data is overwritten, masked, fully removed, marked as "inactive," or archived, etc.). As part of the metadata tagging process DHS tags each data set with a retention period, and therefore DHS can remind the underlying source IT system of the upcoming retention expiration date if the Framework has not already received a deletion notification.

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: As noted above, the Framework is designed first and foremost to support DHS users' mission needs while mitigating privacy risk. One of the hallmarks of the Framework is the ability to restrict access to types of data, including PII, within the Framework 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. During the Initial Operational Capability phase, the Department plans to expand the Framework's approved tagging scheme elements (i.e., common information fields used across the data sets) and refine data controls/tagging as the Department operationalizes the Framework in accordance with the governance onboarding process, described above, and as approved by the DFSG.

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

Mitigation: Only the DFSG can approve new users and uses, with input from oversight offices. Once a user or use is approved by the DFSG, technical controls ensure the user is only able to access data for the use or uses for which he or she has been approved. 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.

Mitigation: The Department successfully tested the user attributes, tags, and context to verify that the controls performed correctly. At the Department's request, the DPIAC provided recommendations on what auditing and oversight capabilities DHS could develop to ensure that these controls are not circumvented. The Department will analyze, document, and report on its implementation of the DPIAC's recommendations using the DFSG governance structure.

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: DHS will not use the Framework share information outside of the Department during the Initial Operational Capability phase without updating this PIA and reassessing this risk. In the long-term, DHS does plan to use the Framework to share information externally, including to reduce and replace the number of external bulk transfers of DHS data. However, such an expansion will require validated mission use cases and the approval of the DFSG (including oversight offices).

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.

Privacy 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 partially mitigate this risk, DHS will develop a Data Quality Plan during the Initial Operational Capability. In accordance with the Data Quality Plan, the Program Management Office is establishing a feedback mechanism to report data quality issues to source systems. In addition, the Limited Production Capability phase introduced data quality processing that assures all data is received unaltered and correctly processed; that data anomalies are identified, logged, and reported to the source data owners for potential correction; and that the Framework is generating data quality metrics for performance and compliance reporting.

This risk will not be fully mitigated until DHS develops a near real-time refresh capability. The Program Management Office also identified the timelines for manually refreshing each existing data set, and began implementation of limited manual data set refreshes. For each new data set, DHS will use the onboarding process, described above, to identify and implement manual data refresh timelines. To provide additional mitigation during the Initial Operational Capability phase, Framework users will continue to 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 issuing any raw intelligence, (e.g., intelligence information report), 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.

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

Mitigation: The Department follows the requirements for information assurance and security and the development of sensitive systems[15] and handling of sensitive information[16] for the Framework systems. Both Framework systems, Cerberus and Neptune, have system security plans and the Chief Information Security Officer's approval for Authority to Operate. Security and policy based controls enforced with these systems include:

- The use of data encryption of any media and during any transmission of data to prevent PII exposure.
- Only users with administrator privileges will be able to directly access the delivered data within the Neptune or Cerberus systems to initiate ingest and data quality processing. These users will be vetted and approved for access up to the highest level of data on the system.

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

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.

Privacy 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: DHS has demonstrated that the Framework's audit capabilities were adequate to support an audit of whether PII 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. 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, at DHS's request, the DPIAC provided recommendations on what auditing and oversight capabilities DHS could develop to ensure that these controls are not circumvented. The Department will analyze, document, and report on its implementation of the DPIAC's recommendations using the DFSG governance structure.

Privacy 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 Initial Operational Capability phase, the Program Management Office will pull a random selection of queries from Framework systems and manually review them to determine compliance with the Framework policies. The Program Management Office will present its findings to the DFSG, which includes the Privacy Office, the Office of Civil Rights and Civil Liberties, and the Office of the General Counsel.

Additionally, to mitigate this risk, DHS requested the DPIAC develop recommendations for how DHS can use audit logs in a meaningful way to ensure robust oversight. The Department will analyze, document, and report on its implementation of the DPIAC's recommendations using the DFSG governance structure.

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, helping to demonstrate the ability to meaningfully use technology including dynamic access controls to mitigate risk. 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
Department of Homeland Security

Clark Smith
Office of Intelligence and Analysis
Department of Homeland Security

Approval Signature

Original signed copy on file with DHS Privacy Office.

_____

Karen L. Neuman
Chief Privacy Officer
Department of Homeland Security