Privacy Impact Assessment Update
for the

DHS Data Framework–

Retention

DHS/ALL/PIA-046(d)

March 15, 2017

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

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

Abstract

The Department of Homeland Security's (DHS or Department) Data Framework is a scalable information technology (IT) program with built-in capabilities to support advanced data architecture and governance processes. The DHS Data Framework is the Department's "big data" solution to build in privacy protections while enabling a more controlled, effective, and efficient use of existing homeland security-related information. Beginning in January 2017, DHS is modifying how the Department manages source IT system retention requirements for the DHS Data Framework. DHS is updating the Data Framework Privacy Impact Assessment (PIA) to account for a modification in the way the Data Framework manages and complies with retention periods and the capabilities to support internal retention management and enforcement within the Data Framework of the source system retention schedules, when internal management is required.

Overview

In a Privacy Impact Assessment (PIA) published on November 6, 2013, and subsequent PIA updates, the Department of Homeland Security (DHS or Department) previously described the Department's development of the DHS Data Framework.[1] The DHS Data Framework creates a systematic repeatable process for providing controlled access to DHS data across the Department. The DHS Data Framework is DHS's "big data" solution that builds in privacy protections while enabling a 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 DHS Data Framework includes the Neptune[2] and Cerberus[3] systems and the Common Entity Index.[4]

Reason for the PIA Update

Initially, the DHS Data Framework planned to enforce existing source IT system data requirements, including retention, in the DHS Data Framework by relying on the source IT systems to notify the DHS Data Framework of changes, deletions, or corrections to data. As contemplated, DHS would not delete any data until the DHS Data Framework received a deletion notification from the source IT system. However, DHS discovered that some source IT systems are not always able to accommodate this request to send such delete notifications due to a number of constraints

---

[1] *See* DHS/ALL/PIA-046 DHS Data Framework, November 6, 2013, DHS/ALL/PIA-046(a) DHS Data Framework, August 29, 2014, DHS/ALL-PIA-046(b) DHS Data Framework, February 27, 2015, DHS/ALL/PIA-046(c) DHS Data Framework, March 30, 2016, *available at* www.dhs.gov/privacy.
[2] *See* DHS/ALL/PIA-046-1 Neptune Pilot, September 25, 2013, DHS/ALL/PIA-046-1(a) Neptune Pilot, August 29, 2014, DHS/ALL/PIA-046-1(b) Neptune, February 27, 2015, *available at* www.dhs.gov/privacy.
[3] *See* DHS/ALL/PIA-046-3 Cerberus Pilot, November 22, 2013, DHS/ALL/PIA-046-3(a) Cerberus Pilot, August 29, 2014, DHS/ALL/PIA-046-3(b) Cerberus, February 27, 2015, *available at* www.dhs.gov/privacy.
[4] *See* DHS/ALL/PIA-046-2 Common Entity Index Prototype, September 26, 2013, *available at* www.dhs.gov/privacy.

(*e.g.*, resources, legacy systems, disruptions to operational support).

While the Department continues to prefer that the source IT systems provide deletion notifications to the DHS Data Framework and, as described in previous DHS Data Framework PIAs, it is clear that this solution is not always possible or practicable. To address this issue, DHS developed and deployed a data management capability within the DHS Data Framework that manages the source IT system's data retention rules to ensure and enforce compliance. Specifically, the DHS Data Framework employed a process to capture and validate the source IT system retention rules and develop the software to enable the Data Framework to replicate the source IT system rules to be compliant with the rules that govern the source IT system. This process ensures that the DHS Data Framework continues to follow and comply with the retention periods provided for each source IT system.

When the Department determines that it must manage retention of a particular data set internally, based on a lack of notification from a source system, the DHS Data Framework Program Office presents an internal retention management request to the DHS Data Framework governance structure for formal approval.[5] Before presenting the internal retention management request to the governance structure, the Data Framework Program Office, in collaboration with associated stakeholders, prepared and documented an analysis of the compliance impacts associated with the internal retention management. This documentation includes an assessment of potential risks and proposed mitigations, as well as consistency with DHS Data Framework compliance protections for approval by the Oversight Offices and ultimately to the DHS Data Framework governance structure.

The Department continues to maintain a strong preference for retention management by source IT systems for the data in the DHS Data Framework and will seek to implement these notification capabilities in the development process of new systems. By ensuring that this analysis and oversight occurs when alternative retention management changes are needed, however, the Department will continue to ensure that risks are understood, documented, and mitigated. Oversight and governance awareness and approval will ensure continued transparency and compliance.

The DHS Data Framework continues to rely on the accuracy of source systems and the data validation during ingest and data quality processing to ensure accurate tagging and index. The DHS Data Framework Program Office is establishing timelier refresh of information from the source systems with near real-time being the refresh goal. 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 DHS Data Framework.

[5] The Department governs the DHS Data Framework through the Data Framework Working Group (DFWG). The DFWG consists of Component Mission Operators, System Owners, Data Stewards, Component and Headquarters Oversight Offices and other Stakeholders as needed. The Oversight Offices include the Privacy Office, the Office for Civil Rights and Civil Liberties, and the Office of General Counsel.

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.

Any risk that the data will not accurately reflect ongoing data changes at the source system because of refresh latency is mitigated by not relying on the system data for any operational actions or decisions. The systems of record are the source data systems so users retrieving information via Data Framework will be trained to understand the risk associated with the latency of data and trained to verify that information at the source system.

Privacy Impact Analysis

Authorities and Other Requirements

The data used by the DHS Data Framework continues to be covered by the source system System of Records Notices (SORN). All source datasets are listed in Appendix A of the DHS Data Framework PIA,[6] including the applicable source system SORNs. Appendix A will be updated as new systems are added to the DHS Data Framework.

There is no change to the Paperwork Reduction Act requirements, which remain non-applicable to this effort.

Characterization of the Information

There is no change to the information the project collects, uses, disseminates, or maintains. The sources of the information and how the information is collected are described in previous DHS Data Framework PIAs and in Appendix A.

This update does not have a privacy impact related to characterization of the information.

Uses of the Information

There is no change to how and why the Framework uses the information, which is described in previous DHS Data Framework PIAs. These previous PIAs also describe how the DHS Data Framework uses technology to conduct electronic searches, queries, or analyses in an electronic database to discover or locate a predictive pattern or an anomaly.

This update does not have a privacy impact related to the uses of information.

[6] *See* DHS/ALL/PIA-046 DHS Data Framework PIA Appendix A, *available at* www.dhs.gov/privacy.

Notice

This PIA update provides additional notice to describe the internal retention management capability the DHS Data Framework will employ in certain instances.

Data Retention by the project

To ensure appropriate technical and policy governance of the program, including the incorporation of robust privacy, civil rights, and civil liberties protections, the DHS Data Framework institutes a process for managing the lifecycle of a source IT system's data. Data is to be maintained according to the retention, use, and handling provisions of the respective SORN for the source IT system and approved National Archives and Records Administration source IT system records schedules.

As part of this process, the DHS Data Framework originally planned and incorporated into public-facing privacy documentation the requirement for the original source IT systems to notify the DHS Data Framework of changes, deletions, or corrections to data with each refresh of data.

However, coordination of data delivery with source dataset owners has revealed that the source IT systems are not always able to accommodate this request to send such a notification due to a number of constraints (*e.g.*, resources, legacy systems, disruptions to operational support).

Therefore, the DHS Data Framework is developing data management capabilities to internally manage and comply with the source IT system's data retention rules. While the implementation of data management capabilities will provide significant enhancements to the DHS Data Framework's core offerings, it is still the preference of the DHS Data Framework Program Office to receive automatic notifications of changes, deletions, or corrections to business rules from the source IT system whenever possible.

Privacy Risk: There is a risk that the DHS Data Framework will not capture the retention rules accurately and maintain information longer than the NARA-approved retention period.

Mitigation: To mitigate this risk, the Department employed a formal governance process to address DHS Data Framework management of source IT system retention rules to ensure consistency and compliance with the retention period for the source system. As part of the governance process, the DHS Data Framework Program Office, in collaboration with associated stakeholders, prepared a request and documented the analysis of the retention and other business rules and compliance impacts associated with the internal retention management, including an assessment of potential risks and proposed mitigations, as well as consistency with DHS Data Framework compliance protections. The Program Office submitted these materials to the Oversight Offices and the DHS Data Framework governance structure for approval.

Information Sharing

This update does not impact internal or external sharing and disclosure, which is described in previous DHS Data Framework PIAs.

Redress

This update does not impact how access, redress, and correction may be sought.

Nor does this update impact data refresh rates and the possibility of data latency. As the Department brings data into the DHS Data Framework, the Department identifies refresh rates as part of the onboarding process. These refresh timelines are based on operational need, available resources, and technical capabilities. The goal is to have regular refreshes of data according to the refresh timelines established for each data set. Currently, DHS has not achieved near real-time data refresh for all of the data sets. Consequently, DHS continues to mitigate this risk by developing 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.

Auditing and Accountability

This update does not impact how the Framework ensures that information is used properly. However, to provide transparency and traceability, the DHS Data Framework Program Office will provide a report regularly to each respective Component source dataset owner, and the Component Privacy Office. DHS Data Framework governance structure, in order to validate that data in the source IT system and DHS Data Framework are in sync.

This update does not impact the privacy training provided to users. Users will continue to be required to take general privacy training as well as training specific to the DHS Data Framework. 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.

This update does not impact the DHS Data Framework procedures in place to determine which users may access the information and how the DHS Data Framework determines who has access. These procedures continue to be overseen and governed by the Oversight Offices and the DHS Data Framework governance structure.

Responsible Official

Donna Roy
Office of Chief Information Officer
Department of Homeland Security

David Bottom
Office of Intelligence and Analysis
Department of Homeland Security

Approval Signature

Original signed copy on file with the DHS Privacy Office.

_____

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