Privacy Impact Assessment

for the

DHS Data Analysis Tools

DHS Reference No. DHS/ALL/PIA-055(a)

June 13, 2023

Homeland Security

Abstract

The Department of Homeland Security (DHS) Office of Intelligence & Analysis (I&A) is developing, deploying, and using Data Analysis Tools (DAT) to perform enhanced analysis of DHS datasets and other data sources available to DHS in support of its homeland security mission.[1] This Privacy Impact Assessment (PIA) examines the privacy implications of DATs, as they will analyze data sources that contain personally identifiable information (PII), and describes the types of tools the Department may develop, how the tools will use data, what sources of information the tools will use, how information will be protected when it is used in DATs, and the oversight process for DAT deployment. In addition, this Privacy Impact Assessment describes testing environments on the unclassified and classified DHS networks that I&A will use to facilitate the development and testing of DATs.

DHS is updating and replacing the original Privacy Impact Assessment[2] to capture technological and programmatic evolutions since it was first published in 2016. In that time, DHS has consistently advanced its development and analytical capabilities to reduce cost and improve effectiveness. Since publishing the original Privacy Impact Assessment, the Department has decommissioned and replaced two capabilities: the Cerberus IT system and the Data Framework search capability. Cerberus was decommissioned by the Department and replaced by a more efficient and more effective cloud-based solution referred to as the Data Management Hub.[3] Similarly, the Data Framework has been retired and replaced by several new cloud-based environments: including a classified "environment" for development and deployment of tools and analytics on the classified fabric and the Rapid Requirements Development and Data Environment (R2D2E) (an unclassified "environment") for unclassified development and deployment of tools and analytics.

An "environment" is a cloud-based location wherein I&A develops, tests, operates, and maintains DATs. Each of these "environments" may host a variety of "tenant environments," which are generally subordinate virtual environments within the higher-level environments, wherein individual organizations may develop, test, and deploy DATs, in addition to many other technical tasks. In the world of cloud environments, there are many lower-level environments

[1] The five Homeland Security missions include: Prevent terrorism and enhance security; secure and manage our borders; Enforce and administer our immigration laws; Safeguard and secure cyberspace and; Ensure resilience to disasters. For more information, see the "Mission" description on the DHS website, available at http://www.dhs.gov/mission.

[2] The original version of this Privacy Impact Assessment, published August 2016, is available on the DHS Privacy website. *See* DEPARTMENT OF HOMELAND SECURITY, PRIVACY IMPACT ASSESSMENT FOR DHS DATA ANALYSIS TOOLS, DHS/ALL/PIA-055, *available at* https://www.dhs.gov/privacy-documents-department-wide-programs.

[3] *See* DEPARTMENT OF HOMELAND SECURITY, PRIVACY IMPACT ASSESSMENT FOR THE DATA MANAGEMENT HUB, DHS/ALL/PIA-076, *available at* https://www.dhs.gov/privacy-documents-department-wide-programs.

that inherit many of the security and access controls and tools maintained at the highest-level environment. The Data Framework search capability has been replaced by search-style DATs residing in these and other DHS environments.

Further revisions to this Privacy Impact Assessment include additional information about the oversight framework for DATS, the security authorization process, and training requirements to access DATs. Further, in the original DAT Privacy Impact Assessment, the DHS Privacy Office noted its intention to conduct a Privacy Compliance Review (PCR) of DATs within two years of publication, with the goal of addressing the DAT program's compliance with the privacy safeguards. As part of its due diligence to launch the Privacy Compliance Review, the Privacy Office determined that there was no significant use of DATs warranting review at that time. In fact, the DAT program as described in the original Privacy Impact Assessment was overtaken by alternative analysis processes and there was no DAT activity to review at the time of review. Accordingly, a Privacy Compliance Review of DATs has not been completed. The DHS Privacy Office, however, may conduct a Privacy Compliance Review in accordance with the Chief Privacy Officer's statutory authority[4] and DHS Directive 047-01, "Privacy Policy and Compliance"[5] at any time.

This Privacy Impact Assessment outlines DHS's current framework and development of DATs. I&A will update this Privacy Impact Assessment should further changes occur.

Overview

Introduction

Among its many responsibilities, DHS I&A, in conjunction with the Chief Information Officer of the Department, is charged with developing and using DATs to access, analyze, and disseminate information, including to provide information and support to other parts of DHS.[6] I&A continues to improve its ability to equip the Department with the intelligence and information it needs to keep the homeland safe, secure, and resilient.

To fulfill its homeland security missions, DHS relies on the analyses and reporting of its operational and intelligence analysts. To enhance DHS's analytical capabilities, I&A will continue to develop DATs to perform enhanced analysis of DHS datasets and other data sources that are available to DHS to support its homeland security mission. DATs are Department-wide capabilities that may be used by DHS personnel. These DATs will improve the analysis and reporting of DHS data by personnel across the Department, which in turn will help the

[4] 6 U.S.C. § 552.
[5] *See* DHS Directive 047-01, "Privacy Policy and Compliance," January 19, 2017, *available at* https://www.dhs.gov/privacy-policy-guidance.
[6] 6 U.S.C. § 121(d)(13) and (17).

Department fulfill its mission.

DHS is issuing this Privacy Impact Assessment to provide greater transparency into its intelligence and data analysis activities. These DATs permit the use of DHS data only in support of the user's work supporting authorized Component or Office missions, and the DATs do not provide DHS personnel with access to DHS data to which they are not authorized access. Instead, the DATs are technical tools that help DHS personnel use more effectively the data to which they already have access. DHS personnel use DATs in support of missions consistent with the purpose for which the data was originally obtained by DHS and therefore only use DATs for purposes that have already been approved and documented in applicable privacy compliance documents. Consequently, the users and uses of DHS data described in the existing System of Records Notice (SORN) or Privacy Impact Assessment for a particular program or dataset remain unchanged. DAT oversight procedures, including privacy compliance, are discussed below.

Data Analysis Tools

I&A will deploy DATs that enable the Department to perform enhanced analysis—particularly using standard statistical or quantitative methods—of DHS datasets and other data sources available to DHS in support of its homeland security mission. Specifically, DATs may access: (1) DHS-collected datasets; (2) datasets made available to DHS by U.S. Government or foreign partners, some of which may be classified; and (3) commercial or publicly available datasets. These tools will typically be deployed into one of DHS's classified or unclassified accredited environments but may also be deployed into a classified or unclassified accredited environment hosted by other government agencies, or state, local, tribal, territorial (SLTT), or private sector partners in support of collaborative efforts that advance DHS's homeland security mission[7]. Any such deployment must be consistent with the terms and conditions established for such future potential collaborative efforts, which will minimally establish parameters to ensure use comports with DHS and partner missions, authorities, and policies. Dependent on the DAT's current stage of development (i.e., testing or operational), DATs will most commonly be deployed into DHS's unclassified environment, the Rapid Requirements Development and Data Environment, or DHS's classified environment. DATs assist homeland security efforts by helping DHS personnel to use the data they already have more efficiently and effectively. They

[7] One hypothetical example is a DAT that extracts "selectors" (e.g., e-mail addresses, phone numbers, credit card numbers) from a body of text. Such a tool would be built on logic that recognizes these types of selectors. Providing the DAT code to a partner who has a need to perform the same function does not provide any DHS data or information. The tool could be used consistent with the partner's own authorities. In a collaborative scenario, one DHS component may partner with other federal agencies on an activity, and an element of DHS could develop a DAT that could be leveraged by all parties in the collaborative scenario, consistent with a joint Concept of Operations (CONOPS) to establish roles, responsibilities, and authorities, ensuring all users had the proper authorities to leverage the DAT, including any data it might access. Again, these hypothetical examples are theoretical and provided for illustration purposes only.

also assist DHS in performing data-driven analysis.

DATs are applications or analytics used in the unclassified or classified network environments that help analysts discover patterns and insights in data, perform searches of data (e.g., querying one or more databases simultaneously), or better understand search results (e.g., by using data visualization capabilities or performing a quantitative or statistical analysis). DATs allow authorized DHS users to query one or more datasets simultaneously and structure the resulting information in a way that provides context to the data or allows analysts to more helpfully interpret the data.

There are three types of DATs, which are described below.

- **Search or Process Automation Tools:** A search tool provides a user with the ability to perform searches on the datasets that the user is permitted to access. An example of a search DAT is one that allows an analyst to quickly compare two datasets to which they already have access, such as comparing lost and stolen passports against passports utilized by known or suspected terrorists Process automation tools automate standard day-to-day authorized work processes of analysts and aim to accelerate that work through machine assistance. An example of a DAT that conducts process automation is one that automates the process of manually copying, pasting, and formatting data elements from many disparate reports into a standardized format suitable for analysis.
- **Exploratory Analysis Tools:** Exploratory analysis tools assist a user in understanding more about the data, so that the user can determine the next steps they should take in the analysis. For example, exploratory analysis tools may provide the user with descriptive statistics (e.g., average, maximum, minimum, count, and odds ratio) or graphs (e.g., box plot, dot plot, and histogram). The goal of these tools is not to make conclusions, but to describe the data in a meaningful way that allows the analyst to interpret the data more easily. An example of an exploratory analysis tool is one that identifies the number of lost and stolen passports by country of issuance or the number of passports that were lost or stolen by month.
- **Advanced Analysis Tools:** Advanced analysis tools assist a user interpreting the data to answer intelligence or operational questions. These tools could include inferential statistics (e.g., confidence intervals, hypothesis testing, classification, regression), entity resolution algorithms, natural language processing (NLP) or understanding (NLU) algorithms, time series-based techniques, or other data modeling capabilities (e.g., network analysis, trend analysis, geospatial analysis) to include other Artificial Intelligence approaches, such as Machine Learning (AI/ML) algorithms. An example of an advanced analysis DAT is one that plots lost and stolen

passports by geographic location and facilitator, which helps an analyst to visualize the geographic locations where false passports are most often identified or encountered and the underlying networks purveying or using these passports. See Section 3.2 for additional information on advanced analysis tools.

Analysts may use one or more DATs when attempting to answer an intelligence question. Using the scenarios described above, an analyst could first use a search DAT to identify the lost or stolen passports used by known or suspected terrorists. Before the creation of the search DAT, the analyst would have compiled the data and performed the comparison manually. The use of the search DAT allows the analyst to perform the comparison more quickly and reduces human error in the comparison process. Next, the analyst could use an exploratory analysis DAT to determine whether lost or stolen passports were more common from particular countries or used more frequently in certain months. Finally, the analyst could use an advanced analysis DAT to plot the networks of lost or stolen passports on the map with shifts over time. Seeing these locations on a map may help an analyst better understand where the activity is occurring and identify potential trends and facilitation networks. Before the creation of this advanced analysis DAT, the analyst would have performed the network analysis and mapping visualization in separate tools and, in some cases, would have relied on I&A's Media Services team to produce the map.

With the use of the DATs, the resulting product could be an intelligence report or paper that shows that known or suspected terrorists are using stolen passports most frequently from country X to travel along route Y, with peak travel occurring in months A and B. The analyst could have prepared a similar report without the assistance of DATs, but the use of DATs speeds up the data analysis process by making it less manual and improves the analyst's ability to make data-driven conclusions rather than highlighting a few observations that may or may not constitute a trend. Consequently, DATs may also strengthen the objectivity and timeliness of DHS's analysis. The analyst will use the lost and stolen passport data for the same purpose as before the creation of the DATs —to identify known or suspected terrorists or the tools and tactics used by known or suspected terrorists. The DATs do not change the originally intended use; rather, the DATs strengthen the analyst's ability to use the data for the same purpose.

This Privacy Impact Assessment describes DATs at a high level to protect DHS operations and analytical sources and methods. DATs described in this Privacy Impact Assessment do not include the use of basic office productivity software (e.g., Microsoft Excel, Microsoft Access) by analysts in support of the homeland security mission. DATs are not part of the underlying databases or source information technology (IT) systems. DATs do not change any data in a source system or dataset, nor cause data to be retained permanently. (See Section 5.1 for more information on retention). Rather, DATs are external applications used to search or interpret existing datasets. They may be used in either the classified or unclassified DHS

networks, where both unclassified and classified information can be stored, searched, and analyzed.

DAT User Base

DATs may be used with data that resides on either the DHS classified or unclassified networks, or partner networks, consistent with uses of the data as approved in applicable information sharing and access agreements (ISAAs).

DAT Testing Environments

When a user comes to I&A's technical staff with a mission requirement for a tool, the staff first create the DAT on either an unclassified or classified standalone development environment. This development environment allows developers to build and test the functionality of the tool in a separate space where they will not disrupt day-to-day production operations on the network. The development environment connects to a secure, controlled enclave, known as the "testing environment," in which the developer and requestor of the DAT can test the tool against live or real data[8,9] to streamline the development process and to test whether a DAT may help answer an intelligence or operational question.

Currently, developers create and test prototypes using synthetic data when it is available and determined to be the most appropriate and useful data to support development activities. However, live or real data is often formatted differently. To make sure the tool is compatible with live or real data, developers often must re-write the code multiple times, which increases development timelines. The ability to test tools with live or real data before moving them to a production environment will reduce development timelines and provide capabilities to end users more quickly. The use of live or real data for testing and development also allows the mission requestor to better understand whether a DAT can help answer an intelligence question, what gaps may remain, and whether there is value pursuing the tool. Some tools may not be pursued

[8] In accordance with Report 2013-01, Data Privacy and Integrity Advisory Committee (DPIAC) on Privacy Recommendations on the Use of Live Data in Research, Testing, or Training, DHS components take a risk-based approach regarding the use of live data. This analysis begins with a rebuttable presumption that the use of live or real data is not approved. As it relates to DHS I&A's development and use of DATs, the DHS Privacy Office conducted a rigorous privacy risk analysis, as recommended by Data Privacy and Integrity Advisory Committee Report 2013-01, and determined that the use of live or real data is justified. As outlined in this Privacy Impact Assessment, DHS I&A specifically justified its need to use live or real data to develop DATs by specifying its intended use, explaining why synthetic data would not suffice for its intended purpose, and describing the security and technical controls in place to mitigate risk. *See* https://www.dhs.gov/publication/dpiac-recommendations-paper-2013-01.

[9] Per Report 2013-01, Data Privacy and Integrity Advisory Committee (DPIAC) on Privacy Recommendations on the Use of Live Data in Research, Testing, or Training, live data is defined as information containing personally identifiable information that comes from a production system, vendor, or public records, or any other dataset that otherwise contains operational data. Personally identifiable information that has been extracted from production systems for research, testing, or training is commonly referred to as real data. Real data is considered a subset of live data.

after initial development and testing.

Live or real data in the testing environment is only used for development and testing purposes and may not be used for operational purposes. If the testing and development process reveals information about an exigent threat, the mission requestor and I&A technical team will follow a streamlined approval process to move the tool to the production environment (i.e., where it may be used by the mission requestor) or use the results within the testing environment. The details of this streamlined approval process will be documented in an I&A Standard Operating Procedure, developed in coordination with the DHS Office of the General Counsel, Privacy Office, and Office for Civil Rights and Civil Liberties.

For mission users to apply the results from DAT testing to operations, the DAT must be transferred to the production environment on the unclassified or classified DHS network. As part of the transfer process, I&A IT Security staff will conduct a vulnerability and compliance assessment of the DAT.

Oversight of DATs

The DHS Privacy Office, Office for Civil Rights and Civil Liberties, Office of the General Counsel, and I&A's Intelligence Oversight team provide oversight for DATs—as outlined below—through a process that allows DHS oversight offices to focus on DATs that may be higher risk, ensures DATs are reviewed by appropriate oversight groups, and meets DHS's operational timelines. As part of this process, all DATs enter the oversight process by submission of a Privacy Threshold Analysis (PTA) to the DHS Privacy Office, unless they are already covered under an existing Privacy Threshold Analysis. Then, tools are routed through one of two paths, depending on the tool's function, its user base, and the data it is accessing, i.e., "high risk' DATs versus "lower risk" DATs. These paths are described below.

- **Data Access Review Council (DARC) Review ("high risk"):** DATs that 1) implicate the bulk collection of U.S. Persons Information (USPI), and/or are 2) created for the DHS Intelligence Enterprise (DHS IE) for use with Intelligence Community (IC) data are reviewed by the DHS Privacy Office, Office for Civil Rights and Civil Liberties, Office of the General Counsel, and I&A's Intelligence Oversight team through the Data Access Review Council.

- **Intelligence Oversight and Office of General Counsel Review ("lower risk"):** DATs that do not meet the criteria of the Data Access Review Council also require coverage by a Privacy Threshold Analysis and are reviewed by the DHS Office of the General Counsel and I&A's Intelligence Oversight Office. The DHS Privacy Office and the Office for Civil Rights and Civil Liberties may "spot check" these DATs to ensure that they do not raise any privacy civil rights, or civil liberties concerns. Any DATs that raise privacy, civil rights, or civil liberties concerns will be referred for Data Access Review Council review. Tools that

only automate analyst's existing manually performed activities qualify as a Search and Process Automation, DAT; however, they may be referred for Data Access Review Council review if they raise privacy or civil rights and civil liberties concerns. For example, a tool that extracts passport numbers from unstructured text or normalizes phone numbers (e.g., adding or removing a country code, fixing formatting issues) automates functions performed manually by analysts and would not be referred for Data Access Review Council review.

To facilitate Data Access Review Council review of a DAT, where required, DHS I&A drafts a written concept of operations (CONOPS) and other supporting documentation. Additionally, the DHS oversight offices receive a written summary of the DAT that includes detailed information, such as:

- A description of the tool's functionality;
- Its anticipated categorization (i.e., search and process automation, exploratory analysis, advanced analysis, or multi-capability), and a description of how it will be used once deployed;
- The intelligence or operational question or process challenge the tool is designed to address (i.e., the authorized mission need for the data);
- The intended outcome (i.e., mission impact) for the mission requestor;
- The data that will be used or accessed;
- A list of the anticipated DHS organizations that would use the tool;
- A description of the information security controls for safeguarding against risks such as loss, unauthorized access or use, or inappropriate disclosure of the data;
- Whether and what types of personally identifiable information the DAT will access;
- Whether the DAT will generate or create data;[10]
- Where or how the data accessed, generated, or created will be stored;
- How long the data will be retained and the process for correcting inaccurate data or destroying data that is no longer needed,
- Whether the DAT accesses commercial or publicly available data;
- Whether the DAT is envisioned to perform data mining;
- Whether the DAT relies on or attempts to identify individual characteristics that are

[10] For example, AI algorithms could use one or more datasets that produce a subsequent new dataset, which represents the relevant information to be used in analysis. In such a case, the new data generated would be assessed by the oversight offices in the processes described above.

protected (e.g., nationality, gender), accesses categories of information with additional protections (e.g., asylum records), or implicates other individual rights; and,

- What measures (e.g., metrics, audits) are in place to evaluate the tool's effectiveness.

With this detailed information, the DHS oversight offices will conduct an informed review of the proposed DAT to ensure compliance with legal, privacy, and civil rights and civil liberties requirements. I&A initiates the review process by providing the written summary to the oversight offices and requesting review of the new tool.

To meet the operational timelines, the oversight offices have ten business days from the date the request to review the tool is received by the oversight offices. The DHS Privacy Office, Office for Civil Rights and Civil Liberties, Office of the General Counsel, or I&A's Intelligence Oversight may jointly or independently identify objections to the tool, confirm they have no objections, request more information (e.g., a briefing) on a tool, or request changes to a tool. If an oversight office requests additional information, that oversight office and I&A will jointly establish a deadline for provision of that information, notify the other oversight offices of the new deadline to examine those questioned aspects of the tool, and invite the other oversight offices to receive and review the new information.

If the tool review is not completed within ten business days of receipt by the oversight offices (or, in the event of a request for additional information, by the deadline jointly established by I&A and the oversight office(s) requesting the information), then the Under Secretary for Intelligence and Analysis (USIA), the Principal Deputy Under Secretary for Intelligence and Analysis (PDUSIA), or the relevant I&A Deputy Under Secretaries may approve operational use for 30 days, or until the review is complete (whichever comes first). The oversight offices may raise an objection to the provisional use of the tool and elevate any outstanding issue to senior leadership. If an oversight office does not complete the review within 30 days, and is responsible for the delay, elevation to senior leadership is required before additional provisional operational approval may be granted by I&A leadership.

Under exigent circumstances, the Under Secretary for Intelligence and Analysis, the Principal Deputy Under Secretary for Intelligence and Analysis, or the relevant I&A Deputy Under Secretaries may determine that a tool must be deployed immediately to address an exigent crisis or situation. In these circumstances, the approving official will notify in writing the DHS General Counsel, Chief Privacy Officer, and Officer for Civil Rights and Civil Liberties of the exigent crisis or situation and I&A's development and deployment of the tool. The I&A approving official will provide this notification as soon as practicable, but no later than ten business days following deployment. When I&A deploys a DAT under exigent circumstances, the oversight offices will receive the normal written summary of the DAT at the same time the

Deputy Under Secretaries for Intelligence and Analysis notify them of its deployment, and those offices will review the tool e, consistent with the procedures outlined above.

As noted earlier, DATs do not permit access to previously inaccessible DHS data or permit that data to be used in ways incompatible with DHS's original purpose for collecting the data, and the users and uses of that data are described in the applicable privacy documentation. To ensure compliance with applicable privacy documentation, the DHS Privacy Office identifies the applicable System of Records Notice or Privacy Impact Assessment that apply to the users and use of data. I&A's development of DATs for DHS is an enterprise-wide capability. The DHS Privacy Office and Office of the General Counsel will consult with Component privacy and legal offices, as appropriate.

Section 1.0 Authorities and Other Requirements

1.1 What specific legal authorities and/or agreements permit and define the collection of information by the project in question?

DATs do not collect any new information from the public. DHS's authority to collect the information is documented in the System of Records Notice of the DHS source IT system to which the DAT applies. Similarly, DATs do not change I&A's or the DHS Components' authorities to access DHS datasets. DHS provides its employees with access to DHS datasets if those employees have a need for the data in the performance of their official duties. The applicable System of Records Notice is identified as part of the DAT oversight process.

Users must also comply with any other applicable laws or policies governing their use of data. For example, I&A users must comply with Executive Order No. 12,333, United States Intelligence Activities.[11] Among other requirements, Executive Order 12,333 places limitations on the collection, retention, and use of United States Person information.

I&A is authorized to develop DATs on behalf of the DHS Intelligence Enterprise pursuant to 6 U.S.C. § 121(d)(13) and (17), which charges the Under Secretary for Intelligence and Analysis with establishing and utilizing, in conjunction with the DHS Chief Information Officer, a secure communications and information technology infrastructure, including advanced analytical tools, to access, receive, and analyze data and information that provides intelligence, information analysis, and support to other elements of the Department. These Department-wide DATs are consistent with and promote I&A's carrying out these responsibilities.

1.2 What Privacy Act System of Records Notice(s) (SORN(s))

[11] See Executive Order 12,333, United States Intelligence Activities, as amended, July 30, 2008. *Available at* http://www.archives.gov/federal-register/codification/executive-order/12333.html.

apply to the information?

There are two types of data discussed in this Privacy Impact Assessment that are subject to a System of Records Notice. The first type of data is the raw or source data—the data accessed or used through a DAT. The second type of data is the analytical results—what the analyst has learned. For example, a list of lost and stolen passports and a list of passports used by known or suspected terrorists can both constitute raw data, while the analyst's report about the stolen passports used by a particular terrorist constitutes analytical results.

Raw data accessed or used through a DAT remains covered by the source IT system's System of Records Notice. DATs do not alter the data in the source systems. Generally, DHS will use DATs to support the Department's mission to prevent terrorism and other threats to homeland security. However, the authorized use of a particular dataset is described in the System of Records Notice and Privacy Impact Assessment for that data or program, and DATs may only use data consistent with the System of Records Notice and Privacy Impact Assessment for a particular dataset. DATs do not provide a user with any access to DHS data that is not described in the applicable System of Records Notice and Privacy Impact Assessment. Furthermore, all policy and legal controls that apply to a particular dataset are maintained when the data is used with DATs. For example, if a DAT uses Passenger Name Record (PNR) data, then the DAT must apply the requirements of the Automated Targeting System (ATS) System of Records Notice[12] and its associated Privacy Impact Assessment,[13] as well as the U.S. European Union Passenger Name Record Agreement.[14]

DHS work products must also be covered by an appropriate System of Records Notice if records are retrieved by personal identifier. This System of Records Notice may or may not be the same System of Records Notice that covers the data from a source IT system. For example, if I&A and U.S. Customs and Border Protection (CBP) analysts were both analyzing Passenger Name Record data stored in the Data Management Hub, the Automated Targeting System System of Records Notice would cover the Passenger Name Record data used by both analysts because Automated Targeting System is the source system for Passenger Name Record data. While the CBP analyst's work product would be covered under the Automated Targeting System System of Records Notice, the I&A analyst's work product would be covered under

[12] *See* DHS/CBP-006 Automated Targeting System, 77 FR 30297 (May 22, 2012), *available at* https://www.dhs.gov/system-records-notices-sorns.
[13] *See* DEPARTMENT OF HOMELAND SECURITY, U.S. CUSTOMS AND BORDER PROTECTION, PRIVACY IMPACT ASSESSMENT FOR THE ATOMATED TARGETING SYSTEM, DHS/CBP/PIA-006, *available at* https://www.dhs.gov/privacy-documents-us-customs-and-border-protection.
[14] *See* "Agreement Between the United States of America and the European Union on the Use and Transfer of Passenger Name Records to the United States Department of Homeland Security," dated December 14, 2011, *available at* https://www.dhs.gov/sites/default/files/publications/dhsprivacy_PNR%20Agreement_12_14_2011.pdf.

the Enterprise Records System (ERS) System of Records Notice.[15]

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

There is a full system security plan for the creation of the tools. As presented in the plan, the tools will undergo a security review for compliance to relevant policies and processes.

1.4 Does a records retention schedule approved by the National Archives and Records Administration (NARA) exist?

Each of the DHS datasets used by DATs has an approved National Archives and Records Administration (NARA) record schedule or an interim retention schedule described in the applicable Privacy Impact Assessment. These retention schedules are implemented when a DAT uses the data. Retention schedules are implemented through (1) data refreshes or (2) manual deletions.

For datasets on the classified network with refreshes from the underlying source data, the data is updated or deleted with each refresh of data, so that the data on the classified network mirrors the underlying source system and adheres to the source system retention schedule. Retention periods for data in the Data Management Hub (located on a classified network) are enforced through refreshes of data, meaning that retention periods apply to the data based on the time at which the data is refreshed. Refresh rates for each dataset located in the Data Management Hub are established with the data originator prior to initial data transfers, and technology and procedures are in place to notify the recipient of any unexpected delays in refresh transmissions to enable intervention 24x7. If Data Management Hub data is not refreshed in a near real-time or on a recurring basis, then users must validate the information in the underlying DHS source system before taking any action (e.g., writing a report).

For unclassified data maintained outside of the Data Management Hub, data refreshes are outside the control of I&A. Accordingly, manual deletions at the end of an applicable retention period may be necessary because a DAT could have accessed data from a one-time extract or another data source that is not refreshed in a timely matter. In these situations, manual deletions will be performed in accordance with the applicable records retention schedule or other guidelines that govern appropriate retention periods, whichever is more restrictive. As an added protection, users of these types of datasets must also validate the information in the underlying DHS source system before taking any action.

Analytic results are retained according to the applicable System of Records Notice for analysts' work products. Please see Section 5.1 for more information on the retention of analytic

[15] *See* DHS/IA-001 Enterprise Records System (ERS), 73 FR 28128 (May 15, 2008), *available at* https://www.dhs.gov/system-records-notices-sorns.

results.

The NARA General Records Schedules[16] 3.1, "General Technology Management Records"; 3.2, "Information Systems Security Records"; 4.1, "Records Management Records"; and 4.3, "Input Records, Output Records, and Electronic Copies," cover other electronic records created by DATs, including searches and results, audit logs, and other compliance-related documentation.

1.5 If the information is covered by the Paperwork Reduction Act (PRA), provide the OMB Control number and the agency number for the collection. If there are multiple forms, include a list in an appendix.

DATs do not collect information from individuals. Information is analyzed only from the source datasets; therefore, the provisions of the Paperwork Reduction Act of 1980, 44 U.S.C. §§ 3501-21, are not applicable. For all information maintained in the underlying datasets that is subject to the Paperwork Reduction Act, the OMB Control Numbers and the agency numbers can be found within their respective Privacy Impact Assessments.

Section 2.0 Characterization of the Information

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

DATs enhance DHS's ability to analyze and understand the data the Department already collects pursuant to its homeland security missions. Consequently, DATs do not collect information directly from individuals. Instead, DATs use information obtained from various data sources available to DHS, many of which collected information directly from individuals. DATs access and analyze, but do not permanently retain, the information.

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

As noted previously, DATs may access: (1) DHS-collected datasets; (2) datasets made available to DHS by U.S. Government or foreign partners, some of which may be classified; and (3) commercial or publicly available datasets. DATs do not collect information directly from individuals. These datasets may be accessed consistent with DHS's authorities and applicable privacy documentation. DATs do not expand or alter DHS's ability to collect data or access particular datasets.

[16] *See* https://www.archives.gov/records-mgmt/grs/.

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

DATs may access commercial or publicly available data in two forms. First, DATs may access commercial or publicly available data that has been incorporated into a DHS system of records. This commercial or publicly available data may contain personally identifiable information, and its use is documented in the applicable System of Records Notice and Privacy Impact Assessment. For example, the Automated Targeting System uses commercial data to research individuals and cargo requiring additional screening, which is documented in the Automated Targeting System System of Records Notice and Privacy Impact Assessment. Similarly, a DHS source system may maintain publicly available social media information, and the use of the data in a DAT would be covered by the applicable System of Records Notice.

Second, DATs may access commercial or publicly available data not already in a DHS system of records for reference purposes. This reference data does not include personally identifiable information and is therefore not covered by a DHS System of Records Notice. DHS uses this data to aid in DHS's ability to analyze or interpret a DHS-collected dataset or a dataset provided to DHS by a U.S. Government or foreign partner. For example, DHS may use the Official Airline Guide (OAG), which contains information about real-time and historical flight data, to provide context to DHS's own travel data sources.

2.4 Discuss how accuracy of the data is ensured.

To ensure the accuracy and integrity of data to which DATs are applied, DHS data will be obtained from the source systems or the DHS system (e.g., the Data Management Hub), which obtains its data from the source systems. DATs will not alter or transform data in the source systems. If DHS data is not refreshed on a near real-time basis, users must check the underlying source systems to confirm the data is still accurate before taking any action (e.g., writing a report).

Data provided to the Department by U.S. Government or Intelligence Community (IC) partners is considered accurate. If there are any questions regarding the accuracy or timeliness of the data, then DHS analysts will work with the originating agency to confirm the information. The accuracy of commercial or publicly available data ingested into a DHS system of records will be ensured through the mechanisms documented in the Privacy Impact Assessment for the applicable source dataset. Commercial or publicly available data used for reference purposes will not contain personally identifiable information and will only be ingested from sources that are considered reliable by industry standards.

2.5 Privacy Impact Analysis: Related to Characterization of

the Information

Privacy Risk: In some instances, the age of the data used by the DATs may reduce the effectiveness of findings or render them obsolete. For example, if the underlying source system refreshes its data hourly or daily, but the same data residing on the classified network is only refreshed monthly, then a user may miss relevant information that may have been added to the record in that time (e.g., a benefit decision, recent travel in or out of the United States). Therefore, there is a risk that an individual may be included in intelligence reporting based on DAT use on inaccurate or obsolete data.

Mitigation: This risk is partially mitigated. The ability to react and respond to threat information is dependent on the timely receipt of accurate and reliable data. For example, the Department cannot take the appropriate action or put in place mitigation tactics if an illicit activity has already occurred by the time DHS receives notice. DHS therefore has an operational imperative to ensure the information used in its reports and operational activities is as accurate, timely, and relevant as possible. In support of this operational imperative, if the data is not updated in near real-time, users must verify/corroborate the analysis produced by a DAT in the source system—to the extent the analyst has access to the source system—before including the findings in reports or as the basis for operational activities.

DHS considers data provided by other government agencies or certain private sector partners to the Department for analytical and operational purposes to be authoritative. If there are any questions regarding the accuracy of other agency-provided data, recipients will work with the originating agency to confirm the information. Finally, when developing intelligence reports, papers, or other work products, DHS intelligence analysts will follow good tradecraft practices, which include documenting the source of data and independently assessing its timeliness and reliability. If DHS becomes aware of a change to the accuracy of previously disseminated information or intelligence in intelligence reports, papers, or other work products, DHS follows standard Intelligence Community procedures to administratively or substantively revise or recall these products in the appropriate dissemination systems, which publish those updates to all consumers to whom the information was originally made available.[17]

Privacy Risk: There is a risk that the use of commercial data may result in inaccurate information being used by DHS or stored in a DHS system as analytical results.

Mitigation: This risk is partially mitigated. As noted before, this Privacy Impact Assessment does not describe a new collection of personally identifiable information by DHS because DATs do not collect new information. DATs will not use public or commercial

[17] *See* Intelligence Community Standards and Procedures for Revised or Recalled Intelligence Products (2020), *available at* https://www.dni.gov/index.php/what-we-do/ic-related-menus/ic-related-links/intelligence-community-policy-memorandums.

information containing personally identifiable information unless that information has already been incorporated into a DHS or other government system of records. In those instances, the pertinent System of Records Notice and Privacy Impact Assessment apply, and the users of DATs rely on the methods the source systems use to ensure data accuracy for public or commercial information containing personally identifiable information.

Additionally, no personally identifiable information is permanently retained or stored in a DAT, so public or commercial information containing personally identifiable information will not be permanently retained or stored outside of the source IT system unless it is part of the analytical results that are separately retained by a DAT user at the conclusion of their use. Please see Section 5.1 for more information on the retention and storage of data. DHS has a mission imperative to ensure its operational and intelligence analysis and reporting is timely, accurate, and relevant, and intelligence analysts employ tradecraft measures (e.g., cross-checking sources, noting when information is derived from commercial sources or when accuracy is uncertain) to ensure the accuracy of their analysis. If a DAT user retains public or commercial information as part of their results, then the results would have undergone a human review to ensure the information is as timely, accurate, and relevant as possible.

In some instances, DATs may access commercial or public information that does not contain personally identifiable information for reference purposes only. In these situations, DHS uses information from public or commercial sources that are considered reliable by industry standards. Intelligence analysts must still employ good tradecraft measures (e.g., noting the information is from a commercial source).

Section 3.0 Uses of the Information

3.1 Describe how and why the project uses the information.

As noted in the "Overview" section, DATs are technical tools that help DHS personnel use more effectively the data to which they already have access. These DATs do not permit uses of DHS data that are incompatible with the purpose for which DHS originally collected that data, and the DATs do not provide DHS personnel with new access to DHS data. Instead, the DHS personnel only use DATs for purposes that have already been approved for their authorized mission uses, consistent with their respective Component or Office authorities, as approved pursuant to the "Oversight of DATs" section of the Privacy Impact Assessment. Generally, DHS analysts will use the data to identify individuals, associations, relationships, or patterns in support of homeland security missions. The users and uses of DHS data described in the applicable System of Records Notice or Privacy Impact Assessment for a particular program or dataset remain unchanged.

For example, a CBP analyst may use Passenger Name Record data from the Automated

Targeting System in a DAT to analyze the travel patterns of known and suspected terrorists. The analyst may use a search tool to identify the travel records of known or suspected terrorists and then use an advanced analysis tool to visualize the travel patterns on a map. In this scenario, the CBP analyst is using the DATs to support existing CBP uses of data to prevent, detect, investigate, and prosecute terrorist offenses and related crimes and to identify individuals who would be subject to additional questioning upon arrival or departure from the United States. These uses of travel data are already articulated in the Automated Targeting System Privacy Impact Assessment and System of Records Notice, as is the comparison and correlation of data from various datasets. DATs are simply technical tools that enhance DHS's ability to use its existing data collections for purposes that have already been conveyed in System of Records Notices and Privacy Impact Assessments.

Once an analyst has finished their analysis, information obtained from DATs may be included in intelligence cables, such as an Intelligence Information Report (IIR), finished intelligence products, such as a bulletin or report, or other operational work products (e.g., a list of individuals selected for additional screening). As with the results of any operational analysis or intelligence product—whether developed using a DAT or through other means (e.g., using traditional office productivity software or non-technical means)—DHS may use this analysis to inform its operational activities. For example, network analysis of DHS travel data and classified holdings of derogatory information may reveal an individual with ties to a terrorist organization who has traveled to the United States. DHS now knows that this individual has ties to terrorism and access to the United States, and DHS can use this information to inform its operations, such as referring the individual to secondary inspection for additional screening, writing a report to inform other Government or Intelligence Community agencies of the connection, denying an immigration benefit, or determining that the individual is inadmissible to the United States when they try to enter the country.

3.2 Does the project use technology to conduct electronic searches, queries, or analyses in an electronic database to discover or locate a predictive pattern or an anomaly? If so, state how DHS plans to use such results.

Yes. DHS analysts could use data in exploratory analysis or advanced analysis tools to discover or locate a predictive pattern or anomaly.

Analysts could use exploratory analysis tools such as graphs or descriptive statistics (e.g., a statistically significant standard deviation) to identify anomalies in the data that require further analysis. An outlier (or infrequent observation) could be observed in a graph. Because outliers can have a profound influence on results, they should be removed if they are not relevant to the analysis (i.e., an outlier that results from a typographic error would not be relevant; an

outlier that identifies aberrant behavior as an indicator of a terrorist may be relevant).

Analysts can use advanced analysis tools to discover predictive patterns or anomalies. Three types of advanced analysis tools and their potential uses are described below.

- **Classification models** group data based on similar attributes and could be used to identify individuals who have similar suspicious behavior patterns to known or suspected terrorists or terrorist groups. For example, DHS may use classification models to identify individuals traveling through routes known to be exploited by illicit individuals. Individuals traveling along these routes may warrant additional scrutiny when they travel to or from the United States.
- **Network analysis** is a standard data modeling technique that can produce representation of a particular network. Modeling the network may reveal how interconnected the network is or predict how information might flow throughout the network. For example, DHS may visually map the interactions of various members of a human trafficking group. The visual representation may help DHS identify the trafficking group's leader or important lieutenants and how information is likely to flow throughout the network.
- **Regression analysis** is a standard statistical practice that can describe the relationship between one or more predictor variables and a dependent variable. For example, when DHS analyzes the age range of individuals who are known to have traveled for illicit purposes, such as seeking to join a terrorist group, a regression analysis may reveal that there is a strong correlation between individuals traveling to join a terrorist group and individuals who are in a particular age range (e.g., an age range of 20-25 might be a better predictor than other age ranges for illicit travel, whereas an illicit traveler aged 50 might be an anomaly). Although the regression analysis would not tell DHS why there is a link between a particular age range and traveling to join a terrorist group, DHS could use this information to help better inform its analysis and operations.

The information obtained using these DATs assists users in either refining their analysis or formulating new searches to collect additional information.

To the extent that these activities qualify as data mining under the Federal Agency Data Mining Reporting Act of 2007, 42 U.S.C. § 2000ee-3, they will be included in DHS's annual data mining report to Congress.

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

Yes. I&A develops or adopts the DATs. DAT capabilities will be shared across DHS.

DHS personnel will have access to DATs that aid in functions consistent with their job duties. For DATs used in the DHS unclassified or classified networks, access determinations will be made based on a combination of rule-based access controls and attribute-based access controls (ABAC), which rely on user attributes such as organization or job series. If necessary, DATs can be restricted to specific user groups that have been approved by the DHS oversight offices. Please see the "DAT User Base" sub-section of the Overview for more information on DAT users.

3.4 Privacy Impact Analysis: Related to the Uses of Information

Privacy Risk: There is a risk that DHS may use DATs on data for purposes other than those for which the data was collected.

Mitigation: This risk is mitigated. DATs do not give a user the ability to access DHS datasets which they are not authorized to access. Consequently, to access a dataset through a DAT, a user must be able to demonstrate that they (1) already have authority to access a particular dataset (e.g., through a memorandum of agreement) for the purpose in question or (2) are participating in a project with appropriate project documentation (e.g., concept of operations, letter of intent) that is approved by the DHS Data Access Review Council.

In the first scenario, DHS has existing processes in place that verify need-to-know and privacy compliance for particular datasets. For example, a user who has access to Passenger Name Record data has followed both DHS and CBP processes for receiving access to that dataset, which includes a verification of the user's need-to-know, compliance with binding international agreements, etc.

In the second scenario, a user may request access to a dataset as part of a Department-level project overseen by the DHS Data Access Review Council. The Data Access Review Council, which includes DHS oversight offices and invited Components, will verify that the use is compatible with the purpose for which the data was collected. For example, if I&A, CBP, and U.S. Immigration and Customs Enforcement (ICE) analysts are working on a special project to detect and disrupt human trafficking, an ICE analyst may need access to a CBP dataset and vice versa. Through the Data Access Review Council, the DHS Privacy Office would verify compliance with applicable privacy documentation, including coordination with CBP, ICE, and I&A privacy offices, as needed. In no circumstance is a DAT used on data for purposes other than for which the data was originally collected.

Further, during DAT use, there may be log management and analysis tools that can monitor and assess audit data population and network processing to identify issues related to erroneous data, false inclusion/exclusion of access and/or information, and to prove that the audit capability is immutable. I&A can conduct periodic reviews for compliance within the program and uses of the tools (e.g., Artificial Intelligence and its sub-disciplines such as

Machine Learning), to ensure that the information is used in accordance with the uses documented in the appropriate Memoranda of Understanding/Agreement, System of Records Notices, information sharing and access agreements, or other technical and business documentation.

Privacy Risk: There is a risk that the technology used by a DAT may pose unique privacy risks.

Mitigation: This risk is partially mitigated. DHS recognizes that some types and uses of technology either inherently carry or have the potential to create privacy risks. For example, the same tool that allows DHS to map lost and stolen passports could be used with another dataset to engage in prohibited or inappropriate behavior (e.g., mapping individuals based on religion) based on the bias of users or designers. Consequently, DATs are developed in coordination with DHS's Office of the General Counsel to ensure they operate consistent with applicable law and policy and by the DHS Privacy Office and Office for Civil Rights and Civil Liberties[18] to ensure that they use information in a manner that appropriately protects individuals' privacy, civil rights, and civil liberties. These DHS oversight offices may, for example, stipulate that DATs not be able to perform certain functions. Additionally, if appropriate, these oversight offices may limit the use of a particular DAT to a select group of users with special training. Any privacy risks and mitigations will be addressed and memorialized in writing.

Privacy Risk: There is a risk that DHS will take information out of context or rely on pattern and anomaly analysis or modeling to take actions or make decisions that would affect a particular group or categories of individuals in a disproportionately negative manner.

Mitigation: This risk is partially mitigated. Mitigation of the potential for bias that might affect a particular group begins with appropriate data selection prior to performing analytics. Data selection is a core element of review in the initial Privacy Threshold Analysis process for any new DATs, affording oversight office engagement in discussions related to this potential risk prior to design or use. Further, actions or decisions are not based on a single report or opinion. They are informed by thorough research and assessments that consider multiple sources of information. Additionally, all DATs that detect patterns, anomalies, or produce models are based on proven statistical methods and must be explainable. Further, analyses based in whole or in part on this information must also characterize uncertainty and limitations of statistical methods. The availability of additional data sources through authorized access to DHS unclassified or classified repositories, information sharing agreements, or Intelligence Community repositories enables users to conduct more robust and comprehensive research. Users may also receive assistance from data scientists, who may reduce the risk of users

[18] More information about DHS Office of Civil Rights and Civil Liberties' oversight of screening, vetting, and intelligence activities is available at: http://www.dhs.gov/security-intelligence-and-information-policy-section.

misinterpreting the data. Further, pre-existing processes within the DHS Intelligence Enterprise, such as tradecraft best practices and the review of intelligence products through manager review chains, are in place to ensure that recommendations or assessments—to include those informed by research and analysis using DATs—are sound and defensible. Finally, the use of DATs will increase DHS's ability to produce data-driven reports, which helps increase the objectivity of DHS's analysis.

Privacy Risk: Users with access to DATs may abuse their privileges by accessing and analyzing data that is outside the scope of their assignment, such as performing searches on themselves, friends, relatives, or neighbors.

Mitigation: This risk is partially mitigated. As previously noted, DATs do not permit an analyst to access information that they do not otherwise have authority to access. Most DATs are tailor-designed to address a specific mission need, such that the DAT can only perform a specific task or function that accesses a particular pre-approved dataset that is known to be relevant to a user's authorized mission needs. For example, a user may simply click an "update" button and the DAT runs the pre-programmed query, thus mitigating the potential for the user to access data outside the scope of their assignment when using that particular DAT.

Further, users and administrators of Rapid Requirements Development and Data Environment- or classified-managed systems are held accountable for the protection of personally identifiable information by I&A's use of access controls, audit logs, and training. All user activity in DHS-managed systems is logged. User activity will be reviewed on regular and ad-hoc bases. Patterns of usage should be consistent with the analysis goals and results that an individual's work products indicate. Discovery of unauthorized use of systems, datasets, or DATs will be reported immediately to management, compliance, and legal authorities for remedial or punitive action. Intentional unauthorized use will result in a permanent prohibition of access, as outlined in the Rules of Behavior and Terms of Service. For data residing on the classified network access is restricted to a limited set of trusted users. Similar Rules of Behavior and Terms of Service apply, including auditing and enforcement.

Section 4.0 Notice

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

When DHS collects information from individuals, DHS provides notice through Privacy Act Statements[19] and its publication of System of Records Notices and Privacy Impact

[19] Pursuant to 5 U.S.C. § 552a(e)(3), agencies are required to provide what is commonly referred to as a Privacy Act

Assessments.[20] DATs do not collect information from the public, and the use of DATs does not change the circumstances of or purpose for DHS's collection of information. DATs provide new technical capabilities to support DHS's existing use of its information and do not change the purpose for which DHS uses the information. Consequently, DHS does not provide a separate notice of its use of DATs at the point at which DHS collects information from individuals.

4.2 What opportunities are available for individuals to consent to uses, decline to provide information, or opt out of the project?

DATs use data from DHS's existing datasets and do not collect information from individuals. Consequently, the source system dataset defines an individual's opportunity to consent to uses or decline to provide information; individuals do not have a separate opportunity to consent to the use of their data in DATs.

4.3 Privacy Impact Analysis: Related to Notice

Each of the source data systems has its own notice requirements and mechanisms for providing public notice. As explained above, no separate notice is provided to the public regarding the use of DATs. Because the DATs do not change DHS's collection or use of the information it receives from the public, the privacy impacts on the public stem from DHS's original collection and use of the data, and there are not any new privacy impacts related to notice that arise from the use of DATs. For example, the Privacy Act Statement for the CBP Electronic System for Travel Authorization (ESTA)[21] informs applicants that DHS solicits information:

> "to determine the eligibility of, and whether there exists a law enforcement or security risk in permitting, the alien to travel to the United States. Upon review of such biographical information, the Secretary of Homeland Security shall determine whether the alien is eligible to travel to the United States under the program."[22]

The Electronic System for Travel Authorization System of Records Notice also notes that "[t]he information provided through ESTA is also vetted—along with other information that the Secretary of Homeland Security determines is necessary, including information about other persons included on the ESTA application—against various security and law enforcement

Statement to all persons asked to provide personal information about themselves if that information will go into a system of records (i.e., the information will be stored and retrieved using the individual's name or other personal identifier such as a Social Security number).

[20] All DHS System of Records Notices and unclassified Privacy Impact Assessments are available on the DHS Privacy Office website at www.dhs.gov/privacy.

[21] *See* DHS/CBP-009 Electronic System for Travel Authorization (ESTA), 81 FR 39680 (June 17, 2016), *available at* https://www.dhs.gov/system-records-notices-sorns.

[22] *See* "ESTA Privacy Act Statement." Available at: https://esta.cbp.dhs.gov/esta/application.html?execution=e1s1.

databases to identify those applicants who pose a security risk to the United States." DHS provides notice, both at the point of collection and through the Electronic System for Travel Authorization System of Records Notice and Privacy Impact Assessment,[23] that the U.S. Government is vetting the individual's application to determine whether they pose a security risk to the United States. Whether this vetting involves the manual plotting of terrorist networks on network analysis charts or the use of an advanced analysis tool, the authorized purpose (i.e., determining eligibility to travel to the United States) remains the same.

Section 5.0 Data Retention by the Project

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

DATs do not permanently retain raw data accessed through a source system. Data used in the DATs will be internally "cached" (i.e., temporarily stored) for the information to be analyzed or processed by the tool. Generally, all cached data will be purged when the user closes the DAT. However, some tools may run continuously in the background and therefore the data remains temporarily stored. For example, an analyst may receive alerts that new information is available about a person of interest. The implementation of retention schedules through (1) data refreshes or (2) manual deletions is described in Section 1.4.

A user may choose to retain the results of their analysis. If a DAT user retains results, they will do so consistent with the applicable System of Records Notice for their work product. For example, the System of Records Notice for the Enterprise Records System applies to any I&A user's work product or user output from the DATs. During its review of tools, the DHS Privacy Office, in coordination with Component privacy offices as appropriate, will identify the applicable System of Records Notice that covers the retention of any results and document the System of Records Notice in the written tool summary. Typically, users will maintain the output of the tools (such as electronic results or written analysis) in a shared space (e.g., access-controlled SharePoint sites) in which users may collaborate with other DHS users or other U.S. Government partners. This storage of results must also be consistent with the System of Records Notice that covers the user's analytical results.

5.2 Privacy Impact Analysis: Related to Retention

Privacy Risk: The data used with DATs includes unclassified or classified datasets that have been copied outside of their source systems (e.g., as agreed to in Information Sharing and Access Agreements). Consequently, the integrity of the data is dependent on regular refreshes of data, and there is a risk that data may be maintained on the classified network longer than the

[23] *See* DEPARTMENT OF HOMELAND SECURITY, U.S. CUSTOMS AND BORDER PROTECTION, PRIVACY IMPACT ASSESSMENT FOR THE ELECTRONIC SYSTEM FOR TRAVEL AUTHORIZATION, DHS/CBP/PIA-006, *available at* https://www.dhs.gov/privacy-documents-us-customs-and-border-protection.

source system retention period if data accessed by DATs is not refreshed in near real-time.

Mitigation: This risk is partially mitigated. This risk is not specific to the use of DATs; rather, this risk applies to the general activity of copying data outside of the source IT system or dataset. The full mitigation of this risk is near real-time refreshes of data through processes or projects (e.g., application programming interface connections established via Interconnection Security Agreements, other projects on DHS's classified network) that are not described in this Privacy Impact Assessment. However, if a dataset is not refreshed in near real-time, DAT users mitigate this risk by confirming the accuracy or integrity of the results they produce with the source system or data provider before taking appropriate action. Furthermore, with access to data provided from multiple systems, there is a greater likelihood inconsistent data will be recognized than if users accessed individual systems.

Privacy Risk: There is a risk that analysts will retain analytical results on local computers, shared drives, collaboration spaces, or in other locations and that they will not delete those results in accordance with the applicable retention period.

Mitigation: This risk is partially mitigated. This risk is not specific to DAT use and applies to the use of data by DHS personnel in general. This risk is partially mitigated by the annual training conducted by the I&A Intelligence Oversight Officer regarding Executive Order No. 12,333, United States Intelligence Activities, as amended July 30, 2008, which includes safeguarding information concerning U.S. Persons and reviewing constitutionally protected activities. Intelligence Oversight training is mandatory for all DHS I&A personnel (including employees, detailees, and contractors). All DHS employees are also required to complete annual privacy training.

Additionally, I&A analytical products—including information accessed by DATs—are subject to review for Analytic Standards.[24] The Standards promote the protection of privacy and civil liberties by ensuring the objectivity, timeliness, relevance, and accuracy of personally identifiable information used in analytic products. If reviewers or managers question the source, timeliness, or accuracy of the data, DAT users will be asked to validate and confirm the data in question.

Section 6.0 Information Sharing

6.1 Is information shared outside of DHS as part of the normal agency operations? If so, identify the organization(s) and how the information is accessed and how it is to be used.

[24] See Intelligence Community Directive (ICD) 203, Analytic Standards, accessible at: http://www.dni.gov/files/documents/ICD/ICD%20203%20Analytic%20Standards.pdf, dated 2 January 2015.

Only DHS personnel authorized to access the underlying data will be permitted to use DATs that have DHS data containing personally identifiable information, and DATs do not share DHS information outside of DHS. DHS may provide DATs that do not contain personally identifiable information or do not contain DHS data to external partners.

Analytic work products (e.g., intelligence assessments) that may be informed or supported by DATs will be shared with external partners consistent with Components' pre-existing information sharing and dissemination guidelines approved by DHS oversight offices. These existing processes provide that, among other things, personally identifiable information included in an analytic work product will be disseminated consistent with the user's authorities, policies, and procedures, including an applicable routine use outlined in the System of Records Notice for any system of records in which the results are maintained, as well as the laws and policies governing the dissemination of the underlying information provided by the source system.

To ensure that results are disseminated in a manner consistent with the System of Records Notice in which the results are maintained, the written summary of the DAT will identify the system of records for the data user in which the results of their use apply. For example, I&A personnel will share analytic work products consistent with the Enterprise Records System's System of Records Notice, and CBP personnel will share analytic work products consistent with the Automated Targeting System's System of Records Notice. DATs will include instructions for users regarding information sharing.

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

No raw data from source systems is shared. Only work products may be shared, and the DAT must identify the System of Records Notice which applies to the user's work products. For example, I&A personnel's work products are covered by the Enterprise Records System's System of Records Notice, which includes routine uses such as Routine Use D, which permits the disclosure of information "[t]o a Federal, State, local, tribal, or territorial government or agency lawfully engaged in the collection of intelligence (including national intelligence, foreign intelligence, and counterintelligence), counterterrorism, homeland security, law enforcement or law enforcement intelligence, and other information, where disclosure is undertaken for intelligence, counterterrorism, homeland security, or related law enforcement purposes, as authorized by U.S. Law or Executive Order, and in accordance with applicable disclosure policies." If I&A prepares an intelligence product on terrorist networks using lost or stolen passports to travel to the United States that is informed or supported by a DAT and shares this product with another federal intelligence agency for counterterrorism purposes, then Routine Use D would cover this sharing. This sharing is compatible with the purposes

articulated in the Enterprise Records System's System of Records Notice because it advances I&A's mission to "identify and assess the nature and scope of terrorist threats to the homeland"[25] and "detect and identify threats of terrorism against the United States,"[26] and would fall within its responsibility "[t]o disseminate, as appropriate, information analyzed by the Department within the Department, to other agencies of the Federal Government with responsibilities relating to homeland security."[27]

Similarly, for example, if an analyst at CBP's National Targeting Center uses a DAT to analyze data in the Automated Targeting System to prepare a list of terrorists using lost or stolen passports, the analyst's work product is covered by the Automated Targeting System's System of Records Notice, which includes routine uses such as Routine Use H, which permits disclosure of information "[t]o federal and foreign government intelligence or counterterrorism agencies or components where DHS becomes aware of an indication of a threat or potential threat to national or international security, or to assist in anti-terrorism efforts." If CBP writes an intelligence product based on the product created by the DAT informing another federal counterterrorism agency that a known or suspected terrorist on the list has traveled to the United States using a lost or stolen passport, then Routine Use H would cover this sharing. This sharing is compatible with the Automated Targeting System's System of Records Notice because it supports CBP's efforts to "perform targeting of individuals who may pose a risk to border security or public safety, may be a terrorist or suspected terrorist, or may otherwise be engaged in activity in violation of U.S. law" and supports the "enforcement of the laws enforced or administered by DHS, including those related to counterterrorism."

6.3 Does the project place limitations on re-dissemination?

Because only work products may be disseminated, only work products may be re-disseminated. Restrictions on re-dissemination of work products will be imposed consistent with existing processes. These processes address a variety of policy and legal requirements, such as the proper handling of information protected by statute or regulation, such as information about asylum records[28] or victims of certain qualifying crimes.[29]

Additionally, products may be subject to dissemination controls related to classified national security information, which may require that an individual has a certain security clearance to access a classified product. In all cases, individuals must establish a need-to-know to access classified information.

[25] 6 U.S.C. § 121(d)(1)(A).
[26] 6 U.S.C. § 121(d)(1)(B).
[27] 6 U.S.C. § 121(d)(8).
[28] 8 CFR § 208.6.
[29] 8 U.S.C. § 1367.

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

DAT users will follow their existing Component processes to maintain a record of any disclosures outside of the Department. For Department-level initiatives, users will follow processes approved by the DHS oversight offices and outlined in appropriate project documentation.

6.5 Privacy Impact Analysis: Related to Information Sharing

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

Mitigation: This risk is mitigated. Results or outputs informed or supported by DATs will be shared with external partners consistent with Components' pre-existing information sharing and dissemination guidelines. For Department-level initiatives, DAT users will follow procedures approved by the Office of the General Counsel, the Privacy Office, and the Office for Civil Rights and Civil Liberties. These business rules provide that, among other things, personally identifiable information included in an analytic work product will be disseminated consistent with the user's authorities, policies, and procedures, including an applicable routine use outlined in the System of Records Notice for any system of records in which the results are maintained, as well as the laws and policies governing the dissemination of the underlying information provided by the source system.

All DATs must identify the System of Records Notice that covers its output. Section 6.2 provides two examples of how a routine use could be used to share externally information that is supported by or derived from a DAT. Further, DHS provides mandatory privacy training to all employees and contractors who have access to or use personally identifiable information, and all users are required to complete mandated information security training that addresses privacy as well as the proper and secure use of DHS applications.

Section 7.0 Redress

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

The DATs' work product may contain classified and sensitive but unclassified information related to intelligence, counterterrorism, homeland security, and law enforcement programs. Such records are exempted from notification, access, and amendment to the extent permitted by subsection (j) and (k) of the Privacy Act, as described in the Code of Federal Regulations, and as delineated in the applicable System of Records Notice. However, the System

of Records Notices for the DHS datasets explain the procedures by which data subjects may request amendment of their information. If information is corrected or removed in the underlying source systems, its accuracy is reflected through refreshes of records in the repository or data extract from the source system, subject to the search capability and DATs. The procedures for individuals to address possibly inaccurate or erroneous information are described in underlying System of Records Notices.

Individuals may seek access to their records from the underlying source system and from the system that covers that users' analytical output by following the directions set forth in the appropriate System of Records Notices.

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

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

As noted above, DAT work product may contain classified and sensitive but unclassified information related to intelligence, counterterrorism, homeland security, and law enforcement programs. Such records are exempted from notification, access, and amendment to the extent permitted by subsection (j) and (k) of the Privacy Act, as described in the Code of Federal Regulations, and as delineated in the applicable System of Records Notice. However, the System of Records Notices for the DHS datasets explain the procedures by which data subjects may request amendment of their information. If information is corrected or removed in the underlying source systems, its accuracy is reflected through refreshes of records in the repository or data extract from the source system, subject to the search capability and DATs. The procedures for individuals to address possibly inaccurate or erroneous information are described in underlying System of Records Notices for the source datasets.

Individuals may seek to amend their records in the underlying source system and in the system that covers that DHS users' analytical output by following the directions set forth in the appropriate System of Records Notices.

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

DATs do not collect any new information from the public. The authority to collect the source information and procedures for correcting source information are documented in the source IT systems' System of Records Notices. DHS provides general notice to the public on

filing a Privacy Act request on its website at: http://www.dhs.gov/file-privacy-act-request.

7.4 Privacy Impact Analysis: Related to Redress

DATs do not collect or permanently retain any new information from the public. The System of Records Notices for the underlying source data explain the procedures by which data subjects may request amendment of their information. If information is corrected or removed in the underlying source systems, its accuracy is reflected in results returned by the search capability. See Sections 2.4 and 2.5 above for further discussion on accuracy of data.

Privacy Risk: There is a risk that individuals may not have sufficient redress for exempted records.

Mitigation: This risk is partially mitigated. Even if individuals are not able to amend exempted records, if they believe these records may have caused or are causing difficulties during travel screenings, they may submit general complaints about treatment or requests for redress to the DHS Traveler Redress Inquiry Program (TRIP), 601 South 12th Street, TSA 901, Arlington, VA 22202-4220 or online at www.dhs.gov/trip. In addition, individuals seeking access to and notification of any record contained in a system of records, or seeking to contest its content, may submit a request in writing to the DHS Chief Privacy Officer, I&A Freedom of Information Act Officer, or the appropriate Component from where the data originates, all whose contact information can be found at http://www.dhs.gov/foia. If an individual believes more than one component maintains Privacy Act records concerning them, the individual may submit the request to the Chief Privacy Officer and Chief Freedom of Information Act Officer, Department of Homeland Security, Washington, D.C. 20528-0655. Even if the Privacy Act does not provide a right of access, certain records about the individual may be available under the Freedom of Information Act.

Section 8.0 Auditing and Accountability

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

Specific auditing, accountability, and oversight measures include:

- A written summary of the DAT (described in the Overview);
- Regular review of DATs by the DHS Office for Civil Rights and Civil Liberties, DHS Privacy Office, DHS Office of the General Counsel, and I&A's Intelligence Oversight team; and
- Additional security measures described in Section 3.4.

During DAT use, log management and analysis tools will monitor and assess audit data

population and network processing to identify issues related to erroneous data, false inclusion/exclusion of access and/or information, and to prove that the audit capability is immutable.

Legal and policy controls on the use and protection of information will be implemented through the policy process and integrated into the technology via access control rules enforced when a DAT directly accesses data. This allows data protections to be built-in to the initial search and monitored enterprise-wide. Access controls are described above in Section 3.3. DHS I&A periodically conducts reviews for compliance within the program and uses of the tools (e.g., Artificial Intelligence and its sub-disciplines such as Machine Learning), to ensure that the information is used in accordance with the uses documented in the appropriate Memoranda of Understanding/Agreement, System of Records Notices, information sharing and access agreements, and other technical and business documentation.

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

DHS provides mandatory privacy training to all employees and contractors who have access to or use personally identifiable information, and all users are required to complete mandated information security training that addresses privacy as well as the proper and secure use of DHS applications.

Persons employed by or detailed to I&A are required to attend annual training conducted by the I&A Intelligence Oversight Officer regarding Executive Order No. 12,333, United States Intelligence Activities, as amended July 30, 2008, which includes safeguarding information concerning U.S. persons and reviewing constitutionally protected activities.

Additionally, users are provided training as part of the process to gain access to certain data repositories or tools (e.g., training agreed to in Information Sharing and Access Agreements). For example, data-specific training may be required by DHS Components as necessary and training will be provided to any user accessing that data. Depending on the project involved, additional training and guidance may be provided prior to, and while engaged with, the project. Software developers, data scientists, and support staff provide DAT training to new users and are available for ad hoc questions and requests.

All DHS DAT users prior to gaining access and on an annual basis must complete:

(i) mandatory DHS privacy and security training, which includes but is not limited to, instruction on the appropriate access, handling, and use of personally identifiable information, or special protected class information (e.g., asylum records);

(ii) information security training that covers the proper and secure use of DHS applications; and

(iii) training conducted by the I&A Intelligence Oversight Officer regarding Executive Order No. 12,333, United States Intelligence Activities, as amended on July 30, 2008, which includes safeguarding information concerning U.S. persons and reviewing constitutionally protected activities, regardless of whether their DATs access such information.

Offices that create DATs will also provide tailored training as part of the process for authorized users to gain access to a DAT. This training will cover issues related to the appropriate use of various datasets accessed through the DATs, as well as demonstrations and hands-on training related to the functionality (i.e., running) of the analytics, interacting with the results, how to report perceived inaccuracies, how to handle sensitive information (e.g., U.S. Person personally identifiable information), and other best practices, including requirements to verify and corroborate findings. Software developers, data scientists, and support staff will also provide ad-hoc training to users and will be available for impromptu questions and requests.

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

To access a dataset through a DAT, a user must be able to demonstrate that they (1) already have access to a particular dataset (e.g., through an Information Sharing and Access Agreement) for the purpose in question or (2) are participating in a project with appropriate project documentation (e.g., concept of operations, letter of intent) that is approved by the DHS Data Access Review Council.

All users accessing classified data will be operating at the Top Secret/Sensitive Compartmented Information classified level and must have security clearances and access approvals commensurate with that level. In addition, I&A analytic personnel will access DATs in accordance with I&A's Intelligence Oversight Procedures. The use of access controls, user PKI certificates, and the DHS user credentialing data store will ensure identity of DAT users as appropriate. If noncompliance is discovered through periodic audit reviews, appropriate disciplinary and corrective actions will be taken.

8.4 How does the project review and approve information sharing agreements, MOUs, new uses of the information, new access to the system by organizations within DHS and outside?

DATs do not share DHS personally identifiable information externally and are therefore not subject to any information sharing agreements or Memoranda of Understanding for sharing personally identifiable information externally. The DAT itself may be subject to a sharing agreement as well as the underlying source data used by DATs or analytic products produced

by the DATs.

Responsible Officials

Bryan Pendleton
Chief Data Officer
Office of Intelligence and Analysis
Department of Homeland Security
(202) 282-8183

Approval Signature

Original, signed version on file with the DHS Privacy Office.

Mason C. Clutter
Chief Privacy Officer
Department of Homeland Security
(202) 343-1717