ADVISE Report

DHS Privacy Office Review of the Analysis, Dissemination, Visualization, Insight and Semantic Enhancement (ADVISE) Program

July 11, 2007

Homeland Security

DHS Privacy Office Review of the Analysis, Dissemination, Visualization, Insight and Semantic Enhancement (ADVISE) Program

Privacy Office
U.S. Department of Homeland Security
Washington, DC

July 11, 2007

TABLE OF CONTENTS

I.	**EXECUTIVE SUMMARY**	**1**
II.	**OVERVIEW**	**1**
A.	Clarifying the term" ADVISE"	2
B.	Focus of the Report	3
C.	Summary of Review	3
D.	Summary Recommendation	3
III.	**PRIVACY COMPLIANCE REQUIREMENTS**	**4**
A.	Privacy Impact Assessment	4
B.	System of Records Notice	4
IV.	**THE ADVISE TECHNOLOGY FRAMEWORK**	**5**
V.	**ANALYSIS**	**7**
A.	Interagency Center for Applied Homeland Security Technology (ICAHST)	7
B.	All-Weapons of Mass Effect (All-WME)	8
C.	Biodefense Knowledge Management System	10
1.	BKMS Pathogen Demonstration	10
2.	BKMS Bio Encyclopedia	10
3.	BKMS Next Generation	10
D.	Remote Threat Alerting System (RTAS)	10
E.	ICE Demonstration (ICE Demo)	12
F.	Threat Vulnerability Integration System (TVIS)	16
VI.	**SUMMARY OF THE PRIVACY OFFICE REVIEW**	**19**
VII.	**RECOMMENDATIONS**	**20**
A.	Short-Term Responsive Action	20
B.	Long Term Responsive Action	21
VIII.	**GOING FORWARD**	**22**

ADVISE Report
DHS Privacy Office
July 11, 2007

I. Executive Summary

The U.S. Department of Homeland Security Privacy Office (Privacy Office) reviewed the DHS Science and Technology Directorate's (S&T) operation of the Analysis, Dissemination, Visualization, Insight and Semantic Enhancement (ADVISE) program. In its review, the Privacy Office determined: (1) some deployments of ADVISE used personally identifiable information (PII) without first conducting Privacy Impact Assessments (PIA) as required; and (2) all ADVISE deployments used or generated data that were contained in existing systems of records maintained under duly published System of Records Notices (SORN), as required by the Privacy Act.

In response to the use of PII in these ADVISE deployments, the Privacy Office recommends a set of short- and long-term responsive actions.

- Short-term recommendations focus on ensuring full compliance with privacy protection requirements before continuing with ADVISE deployments and making better use of non-PII data during research and development efforts.
- Long-term recommendations focus on integrating privacy compliance requirements into S&T's overall project development processes and developing additional privacy guidance for future S&T programs.

The full set of recommendations and a current status update are listed at the conclusion of this report.

II. Overview

The Privacy Office conducted a review of the ADVISE program, which is a program funded and managed by S&T. The Privacy Office conducted this review pursuant to Section 222 of the Homeland Security Act of 2002, which designates the Chief Privacy Officer as the DHS senior official responsible for ensuring that PII is used in full compliance with the fair information practices of the Privacy Act of 1974 and for reporting on complaints of privacy violations.[1] The Privacy Office's review responded to discussions with S&T during which it appeared that ADVISE deployments were using PII without first meeting privacy compliance requirements. Prior to this report, S&T discontinued all ADVISE-related efforts pending resolution of privacy issues and completion of required privacy documentation.

This report provides the results of the Privacy Office's review and concludes with short- and long-term recommendations to better integrate privacy protection into future research

[1] Homeland Security Act of 2002, Section 222, 6 U.S.C. § 142.

and development efforts and provide general guidance to ensure privacy protections are integrated into the overall project development process.

A. Clarifying the term" ADVISE"

The term "ADVISE" has been used interchangeably for two different stages of research and development:

- The first refers to a toolset or development kit - a set of generic tools to gather, link, and present information.
- The second refers to a collection of deployed systems to test the effectiveness of the toolset in specific settings.

Since each of these references to "ADVISE" raises a different set of privacy protection risks, it is important to distinguish between the risks presented by a development kit and the risks presented by a deployed system. This report uses the following separate terms:

- The terms "ADVISE technology framework" or "Framework" refer to the first stage of research and development: the toolset/development kit.
- The term "ADVISE deployments" refers to the second stage of research and development: implementations of the ADVISE technology framework.

The capability of the ADVISE technology framework to use large volumes of PII[2] in complex ways raises privacy protection issues. As discussed below, the Privacy Office determined that the most effective way to address these issues would be to build a combination of technical privacy safeguards and a new privacy assessment mechanism directly into the Framework. The new assessment mechanism, the Privacy Technology Implementation Guide (PTIG), will present recommendations for incorporating privacy protections into the early stages of project development.[3] The results of these early stage privacy protections will be reflected in the required privacy compliance documentation, including the Privacy Threshold Analysis (PTA) and, as appropriate, the PIA and SORN. These privacy compliance documents will remain a requirement after the PTIG is issued, for each specific deployment of the Framework.

[2] The E-Government Act of 2002 uses the terms "personal information" and the term "information in identifiable form." The Homeland Security Act of 2002 uses the term "personal information." The Privacy Office uses the term "personally identifiable information" to address all similar terms and establish the threshold for information that triggers privacy compliance requirements. The Privacy Office defines PII as "Any information that permits the identity of an individual to be directly or indirectly inferred, including any other information which is linked or linkable to that individual regardless of whether the individual is a U.S. Citizen, Legal Permanent Resident, or a visitor to the U.S. This definition includes information about DHS employees and contractors."

[3] As of this report, the Privacy Technology Implementation Guide is still in draft form.

B. Focus of the Report

This report summarizes the Privacy Office's privacy compliance requirements and reviews the ADVISE technology framework against those requirements. The report then identifies privacy issues with the Framework and suggests privacy protection mechanisms to address those issues. Finally, the report assesses the ADVISE deployments with respect to the privacy compliance requirements and recommends a series of responsive actions to improve the integration of privacy protections into the research stage of technology development.

C. Summary of Review

The ADVISE technology framework: The Privacy Office determined that since the ADVISE technology framework does not in itself collect or maintain PII, potential privacy issues regarding the Framework are best addressed through the PTIG. When completed, the PTIG can be used to articulate privacy policy and compliance requirements so they may be incorporated into the architecture of the ADVISE technology framework and other toolsets. As discussed below, the Privacy Office believes that the traditional PIA is better suited for operational systems, including deployments.

The ADVISE deployments: The Privacy Office's review of the ADVISE deployments focused on two issues:

1. Whether the ADVISE deployments complied fully with the existing privacy compliance requirement to conduct a PIA prior to using PII.

   No. The Privacy Office determined that some of the previously conducted ADVISE deployments did not conduct PIAs. In contrast, ongoing ADVISE deployments are drafting PIAs, as necessary, prior to using PII.

2. Whether the ADVISE deployments used or generated data that were contained in existing systems of records maintained under duly published SORNs as required by the Privacy Act..

   Yes. The Privacy Office determined that the ADVISE deployments used or generated data that were contained in existing systems of records maintained under duly published SORNs as required by the Privacy Act.

D. Summary Recommendation

As S&T continues to research and development efforts related to the ADVISE technology framework and associated deployments, the Privacy Office recommends S&T further integrate privacy protections including compliance with privacy policy, statutes, regulations, and recommended best practices. This report concludes with short- and long-term recommendations for implementing this and other responsive actions..

III. Privacy Compliance Requirements

This section describes the Privacy Office's key privacy compliance requirements. The first is to conduct a PIA prior to loading or using PII. The second is to publish a SORN as required by the Privacy Act of 1974 when PII is collected and maintained in a system of records.

A. Privacy Impact Assessment

Section 222 of the Homeland Security Act of 2002, as amended, states that the Chief Privacy Officer is responsible for "assuring that the use of technologies sustain, and do not erode, privacy protections relating to the use, collection, and disclosure of personal information."[4]

Section 208 of the E-Government Act of 2002 obligates federal agencies to "conduct a Privacy Impact Assessment ... before developing or procuring information technology that collects, maintains, or disseminates information that is in an identifiable form; or [before] initiating a new collection of information that will be collected, maintained, or disseminated using information technology."[5]

The Privacy Office implements these statutory authorities through a privacy compliance process formalized in the PIA. In an effort to streamline the initial vetting process, the Privacy Office created a short form, the Privacy Threshold Analysis (PTA) which is used to determine if a program or system meets the initial threshold of the PIA including the anticipated use of PII.

The Privacy Office developed the PIA form to help programs analyze how they collect, use, disseminate and maintain PII. A completed PIA provides transparency into the use of PII and the mitigation of any associated privacy impacts. Upon approval by the DHS Chief Privacy Officer, the Privacy Office publishes PIAs on the Privacy Office website, www.dhs.gov/privacy.[6]

B. System of Records Notice

The Privacy Act of 1974 defines the organizational norms for federal agencies in dealing with an individual's PII. For information under the control of an agency and retrieved by personal identifier, the Privacy Act shapes the collection, use, maintenance, and release

[4] Homeland Security Act of 2002, Section 222(1), 6 U.S.C. § 142(1).

[5] § 208(b)(A-B) of the E-Government Act of 2002.

[6] An exception to publication on the Privacy Office website exists for national security programs. Although these programs are required to complete the PIA, in order to protect the classified nature of the underlying program, the PIA may not be publicly available. Nonetheless, the Privacy Office works with the programs to identify options to discuss these programs publicly.

of covered PII and grants individuals certain rights under specific circumstances to access and correct records about themselves. The Privacy Act also requires a SORN[7] to be published in the Federal Register when PII is maintained by a federal agency in a system of records.[8]

The SORN identifies the purpose for the system of records, what categories of individuals are covered by information in the system of records, what categories of information are maintained about the individuals, and how the information is shared by the agency (also known as "routine uses").[9] The SORN also provides public notice regarding the available mechanisms to exercise the rights granted through the Privacy Act to access and correct the PII that an agency maintains.[10]

IV. The ADVISE Technology Framework

This section describes the ADVISE technology architecture and the applicable privacy protection mechanisms recommended by the Privacy Office.[11]

The ADVISE technology framework is not itself an operational system, rather it is a set of basic building blocks that can be used to build individual systems. The ADVISE technology framework consists of three layers of technology. The first layer is called the "Information Layer," which facilitates bringing data into the ADVISE technology framework. The second layer is called the "Knowledge Layer," which processes the data, enabling relationships to be built between particular pieces of information. The third layer is the "Application Layer" which provides the user with an interface to the entire ADVISE technology framework. These three functional layers are supported and enforced by a security infrastructure including privacy protective role- and rule-based access.[12]

---

[7] 5 U.S.C. § 552a(e)(4).

[8] 5 U.S.C. § 552a(5) "[T]he term 'system of records' means a group of any records under the control of any agency from which information is retrieved by the name of the individual or by some identifying number, symbol, or other identifying particular assigned to the individual."

[9] 5 U.S.C. § 552a(7) "[T]he term 'routine use' means, with respect to the disclosure of a record, the use of such record for a purpose which is compatible with the purpose for which it was collected."

[10] 5 U.S.C. § 552a(e)(4).

[11] This section is based upon information provided by S&T.

[12] According to S&T, privacy and security protections were designed and deployed in the earliest development stages of the ADVISE technology framework: "Primary security for the ADVISE system is ensured through the NEBRASKA Security Infrastructure (NSI). The NSI is intended to prevent unauthorized access to services and data, ensure privacy and integrity of data transmitted over the network, and provide a secure audit log of all service requests. It authenticates the server proxy and client proxy, ensures privacy and integrity of all data sent between the client proxy and server proxy, and

The underlying goal of the ADVISE technology framework is to provide an improved mechanism for analyzing previously collected information in greater detail across larger volumes of information. All data must be loaded into the ADVISE technology framework by a human operator. DHS analysts currently review incoming information manually and use standard desktop software (email, spreadsheet, web browser, word processor) to search for keywords in order to find and record relevant information and the relationships between individual pieces of information. The ADVISE technology framework is designed to provide a better way to visualize each piece of information and the relationships between all pieces of information.

The ADVISE technology framework can accommodate large volumes of data of highly diverse data types. More importantly, the ADVISE technology framework continues to operate in real time even as the data volume and diversity increases. The Framework enables searches based on established relationships (which must be identified by human operators), which means it can assist in identifying links between particular pieces of information across large and varied data collections that could otherwise go unnoticed. The Framework also allows analysts to share the results of each other's efforts and to retain links to the original source of each piece of information which further enhances the efficiency and reliability of the Framework. This capability enables the ADVISE technology framework to support hypothesis testing and decision support.

Ontologies (sets of nouns used to categorize information) define the scope of data types and relationships that can be loaded into an implementation of the ADVISE technology framework. Once an analyst maps data to an ontology and loads that data into an implementation of the ADVISE technology framework, all data are fused together using exact matches.[13] From a privacy compliance perspective, the use of this new information (that the two separate pieces of information refer to the same item) must also be reviewed to ensure that privacy protections are sustained.

The Privacy Office determined that a PIA is not the appropriate vehicle for assessing privacy protection issues related to the ADVISE technology framework because the

---

provides role-based and rule-based access control to services. It provides a technically non-repudiable audit of service requests as well as failed attempts to access the system, and provides user credentials to internal Nebraska components to facilitate the development of additional security functions such as result set filtering."

[13] S&T explained further that the ADVISE technology framework identifies a common node for the two occurrences. This common point links back to the source documents where this data originated. If the source documents are database records, then these two records are now linked inside the ADVISE technology framework through the fusion process. The ADVISE technology framework does not create conclusions. Conclusions are drawn by professional analysts looking at the link diagram that displays the node.

Framework itself does not contain any data and is not focused on any specific implementation. This determination is further supported by the inability of a single PIA to address the variety of data and uses of that data that could be used by multiple implementations of the Framework. Each deployment of the Framework will use definable information for a particular purpose and can therefore be assessed to determine if a PIA would be required. As needed, a full PIA would be conducted for each ADVISE deployment prior to loading or using PII.

The Privacy Office determined that the most effective and efficient method of integrating privacy protections into the ADVISE technology framework is to apply a new type of privacy guidance - one that can be adapted to match the architecture of the Framework itself. This new privacy guidance document, the PTIG, will provide step-by-step guidance to integrate privacy compliance requirements into the development process for individual deployments of the ADVISE technology framework.

The Privacy Office and S&T also plan to collaborate on the development of a PTIG specifically tailored to the research and development process. The Privacy Office and S&T plan to use the results of these two PTIGs to build a PTIG specific to the ADVISE technology framework.

V. Analysis

This section presents the Privacy Office review of the ADVISE deployments and evaluates the extent to which each deployment complied with DHS privacy protection requirements.

The Privacy Office conducted interviews with personnel from S&T, Lawrence Livermore National Laboratory (LLNL), and the Los Alamos National Laboratory (LANL) and reviewed documentation regarding the ADVISE technology framework and the ADVISE deployments. The Privacy Office also engaged in additional discussions with DHS Components associated with the ADVISE deployments. The following is a review of each ADVISE deployment and its associated privacy analysis.

A. Interagency Center for Applied Homeland Security Technology (ICAHST)

S&T manages ICAHST, a research facility located at the Johns Hopkins University Applied Physics Laboratory. Using only synthetic data,[14] ICAHST evaluates promising homeland security technologies for DHS and other government stakeholders within the

[14] Synthetic data is artificially created data designed to mimic the characteristics of personally identifying information without using information that refers to actual individuals. One value of synthetic data is the ability it provides to test system functionality with minimal impacts on privacy.

homeland security technology community. ICAHST's evaluation of the ADVISE technology framework is currently on hold, but ICAHST plans to continue evaluating the Framework to address issues such as overall performance, hardware and software installation, and user interface.

According to S&T, ICAHST completed evaluations of the ADVISE technology framework including setup, configuration, and basic functionality. Tests for capacity, detailed functionality, usability, and utility are at various stages of completion. Testing of basic functionality indicates that data from multiple sources can be collected and categorized in a common framework and data can be queried and visualized. Finally, preliminary tests show that the ADVISE technology framework correctly and accurately performs data fusion (e.g., merging multiple, heterogeneous data sources into a single environment).

S&T completed a PTA for the ICAHST evaluation indicating that no PII was being used and thus a PIA was not required.

B. All-Weapons of Mass Effect (All-WME)

The Department of Energy started the All-WME effort at the LLNL in October 2002, prior to the formation of DHS. In 2003, S&T supplied funding for All-WME. Initially, All-WME used existing data management and analysis tools developed by LLNL and LANL scientists.

In 2005, S&T funded two specific efforts related to the use of the ADVISE technology framework. The first effort focused on whether the ADVISE technology framework could improve upon existing analysis. The second effort tested the technical performance of the ADVISE technology framework itself. Both tests were conducted in 2006.

The specific testing goals of the ADVISE technology framework included:

- Identifying information and knowledge from high-value Field Intelligence Element source documents and manipulating that information in ways that were previously not possible;
- Capturing and sharing analysis amongst analysts; and
- Combining information from multiple data sources.

No operational decisions were made from this performance test and evaluation; and currently there are no plans to continue this deployment in the future.

The All-WME deployment used classified message traffic collected by the National Labs' Field Intelligence Elements (FIEs). The FIE message traffic included information related to foreign groups and organizations involved in WME material flows and illicit trafficking. In a few cases, the characterization of capabilities included information

related to individuals with knowledge, skills and technologies used to advance WME programs. This message traffic contained PII.

The PII in the FIE message traffic was collected from Joint Worldwide Intelligence Communications System (JWICS) intelligence analysis products. These products were vetted, marked, and released by U.S. intelligence agencies as appropiate for use in intelligence analysis. Part of this initial review process included identifying and excluding data that related to U.S. persons. As a result, none of the performance testing and evaluation utilized data related to U.S. persons.

As of this report, there are no present or future plans to continue this ADVISE deployment. If S&T restarts this ADVISE deployment, all privacy compliance requirements must be met prior to loading or using PII.

Privacy Analysis

Once S&T started funding the All-WME deployment, DHS privacy compliance requirements applied. In reviewing the All-WME deployment, the Privacy Office determined that the All-WME deployment did not comply with the privacy compliance requirement to conduct a PIA prior to using PII. The FIE message traffic included information related to individuals which constituted PII and therefore triggered the PIA requirement.

The Privacy Office determined that this ADVISE deployment used or generated data that did not meet the threshold definition of a "system of records" and thus did not trigger the Privacy Act requirements. Only when information relates to an "individual" do the obligations of the Privacy Act apply. The Privacy Act defines an "individual" as "a citizen of the United States or an alien lawfully admitted for permanent residence."[15] The PII used by this ADVISE deployment after DHS started funding the project, did not pertain to U.S. persons. Since the All-WME deployment did not use information pertaining to persons fitting the definition of "individual" under the Privacy Act, the requirements of the Privacy Act did not apply.

Leading up to the cancellation of future All-WME deployments, S&T drafted and the Privacy Office reviewed a PTA and PIA for the ongoing portion of this deployment. If S&T reinititiates this deployment, all draft privacy compliance documentation must be updated and finalized before PII is loaded or used in connection with this deployment.

[15] 5 U.S.C. § 552a(A)(2).

C. Biodefense Knowledge Management System

S&T's Biodefense Knowledge Center (BKC) is pursuing a series of three separate deployment initiatives with the overall goal of identifying better methods to assist DHS analysts in identifying and characterizing biological threats posed by terrorists.[16] BKC's research will begin with the first deployment, "BKMS Pathogen Demonstration" followed by "BKMS Bio Encyclopedia" and "BKMS Next Generation." Of these three initiatives, only "BKMS Next Generation" will use the ADVISE technology framework.

The following is a summary of the series of BKC deployments.

1. BKMS Pathogen Demonstration

This is the first of BKC's three deployments and will serve the DHS Office of Intelligence and Analysis (I&A) and the DHS Office of Health Affairs (OHA). According to S&T, this deployment will not use the ADVISE technology framework and will not use PII

S&T completed a PTA for this deployment which was validated by the Privacy Office indicating that since PII will not be used, a PIA is not required.

2. BKMS Bio Encyclopedia

This is the second of BKC's three deployments and will also serve I&A and OHA. According to S&T, this deployment will not use the ADVISE technology framework, but will use PII.

S&T completed a PTA which was validated by the Privacy Office, indicating that a PIA will be required prior to loading or using PII.

3. BKMS Next Generation

This deployment will utilize the ADVISE technology framework. S&T confirms that this deployment will comply with privacy compliance requirements and will not load or use PII until all applicable privacy compliance requirements are met. Plans for this deployment are currently on hold.

D. Remote Threat Alerting System (RTAS)

This deployment, which was initiated in May 2005, sought to determine if the ADVISE technology framework could assist DHS Customs and Border Protection (CBP) in identifying anomalous shipments based on the cargo type and originating country. The

[16] While only the third deployment contemplates using the ADVISE technology framework, the first two are briefly summarized to provide additional context to the ultimate use of the ADVISE technology framework.

development effort involved collaboration between S&T, the Pacific Northwest National Laboratory (PNNL), and LLNL.

In December 2005, CBP supplied S&T with open-source data that was ultimately used in a demonstration of RTAS in June 2006. All RTAS activities officially ended in September 2006, at which point RTAS was decommissioned. The data was removed, the hardware components were disassembled, and the hard drives were erased.

RTAS was not used in operations; no operational decisions were made based on the deployment, and no further activity is occurring related to this deployment.

According to S&T, CBP supplied S&T with publicly available data related to trade and shipment patterns for the period of April through September 2005 (a total of 3,636,901 records). This data came from the Port Import Export Reporting Service (PIERS), a commercial data provider. PIERS data fields ("Name," "Fname," and "NTF_Name" and other fields related to "shipper" and "consignee") indicate the use of PII, and while there is no indication of whether the actual data used in this deployment included PII, the potential exists and thus the Privacy Office's analysis addressed the potential privacy impacts. It is possible that the actual data only included names of businesses and did not include names of individuals. S&T removed the data at the termination of RTAS and, as a result, there is no record of whether PII was actually used in this deployment.

The PIERS data was combined with historical trade data from the U.S. Census. The Census data covered U.S. maritime imports during fiscal year 2005, and contained data fields for "country of origin," "district port," "HTS [Harmonized Tariff Schedule] code," "value," and "entry count."

According to S&T, RTAS used various scenarios to demonstrate the value of the ADVISE technology framework. These scenarios used data related to shipments from countries of concern, shipments of products of concern, shipments of potential avian flu-carrying products, and shipments of products related to anti-dumping duty orders. A separate scenario examined shipments of items in violation of intellectual property rights (e.g., unlicensed DVD hardware). Part of this last scenario included identifying information related to a given shipment including the shipper and other parties. As mentioned above, the "shipper" and "parties" data fields could include data related to either a business or an individual. There is no indication whether data actually used in the deployment included the names of individuals. S&T states that the intellectual property search was conducted, but asserts that S&T did not use PII as part of the search query.

Privacy Analysis

Based on the lack of specific information regarding the actual data used in the RTAS deployment, the Privacy Office could not reach a determination regarding the privacy compliance requirement to conduct a PIA for this deployment.

The Privacy Office further determined that RTAS used or generated data that were contained in an existing System of Records maintained under a duly published SORN as required by the Privacy Act. DHS's use of PIERS data was covered by the existing DHS SORN entitled, "Automated Commercial System SORN, Treasury/CS .278." As described in the "Categories of Records" section, this SORN covers information related to "commodity and merchandise processing information relating to Customs administration of trade laws." PIERS data fits within that definition.

Based on the operation of the ADVISE technology framework, RTAS fused data so as to create a new collection of information. In reviewing this new collection of information, the Privacy Office determined that the new information was not retrieved by personal identifier thereby placing it outside the definition of a "system of records" and outside the applicability of the Privacy Act of 1974.

S&T completed a PTA after the conclusion of the RTAS deployment which the Privacy Office validated. The Privacy Office's analysis in the PTA indicated that since the deployment was already decommissioned a PIA was not required but that further analysis at the time would have focused on the actual existence of PII in the data set used. The Privacy Office and S&T are now collaborating to ensure an effective privacy compliance assessment and documentation process for ongoing and anticipated ADVISE deployments.

E. ICE Demonstration (ICE Demo)

This deployment was operated by S&T and LLNL to determine whether the ADVISE technology framework could assist DHS Immigration and Customs Enforcement (ICE) in better utilizing existing ICE data.

ICE identified the data to be used in the deployment and delivered that data to S&T on June 7, 2005. This data was not actually used until two weeks prior to the demonstration of the deployment which occurred on July 28, 2005. As planned, the ICE Demo was dismantled shortly after the demonstration.

The ICE Demo deployment was not used in operations, no operational decisions were made based on the deployment, and no further activity is occurring related to this deployment.

S&T reports that a small sample of data from up to eight different data sources were provided by ICE to S&T. S&T and ICE cannot confirm if data from all of the data sources were used, nor can either component provide information on how much data were loaded from the data sources used. According to S&T, the information loaded into ICE Demo was used in a single manner for a single purpose. During the one-time demonstration of ICE Demo, the fused data was used to display a selection of foreign students associated with multiple driver's licenses.

The following is the Privacy Office's analysis of the ICE Demo deployment, including a summary of the data provided by ICE and used during the ICE Demo deployment.

Privacy Analysis

The Privacy Office determined that the ICE Demo deployment used PII without first conducting a PIA. Further analysis at the time would have focused on the requirement to use PII in the pilot. If PII was required for the pilot, a PIA would be required. This level of review is now part of the ongoing collaboration between the Privacy Office and S&T to document privacy compliance for ongoing and anticipated deployments.

The Privacy Office determined that ICE Demo used or generated data that were contained in existing systems of records maintained under duly published SORNs as required by the Privacy Act. According to S&T, the fused data was used to display a selection of foreign students associated with multiple driver's licenses which did not include retrieval by personal identifier. The Privacy Office determined that since PII was not retrieved by personal identifier, the use information did not meet the definition of a "system of records" and thus did not trigger the requirements of the Privacy Act.

The remaining Privacy Act issue is whether each use of each data set contained in an existing system of records was maintained under a duly published SORN as required by the Privacy Act. The following is a privacy analysis of each data source.

- The Student and Exchange Visitor Information System (SEVIS) is maintained by ICE and contains information related to foreign exchange students in the United States studying for an advanced degree. This data source contains data fields for PII including name, date of birth, address – this is confirmed in the SEVIS SORN, DHS/ICE-001. Given that ICE operates SEVIS, the privacy compliance issue is whether PII from SEVIS used in ICE Demo was covered by an applicable SORN. The Privacy Office determined that the PII in SEVIS was covered by the DHS/ICE-001 SORN and thus as it relates to SEVIS data, ICE Demo used data contained in an existing system of records maintained under a duly published SORN as required by the Privacy Act.
- The Law Enforcement Support Center (LESC) is maintained by ICE. As described in the "Categories of Individuals" section of the SORN JUSTICE/INS-023, the information contained in LESC relates to "immigrants

who have the status of legal permanent residents and/or United State citizen and who are either the subject of an investigation, or have been arrested, charged with and/or convicted of criminal or civil offenses." The data included name and date of birth – data considered to be PII. This conclusion is reinforced by the "Categories of Records" section of the SORN that lists biographic identifiers including date, place of birth, and Social Security Numbers. LESC is also an ICE program and just as with SEVIS, the privacy compliance issue is whether PII from LESC used in ICE Demo was covered by an applicable SORN. The Privacy Office determined that the LESC data was covered by the JUSTICE/INS-023 SORN and thus, as it relates to LESC data, ICE Demo used data contained in an existing system of records maintained under a duly published SORN as required by the Privacy Act.[17]

- No Fly List and Selectee List. The No Fly List is produced by the by the Terrorist Screening Center and contains information related to persons identified and prevented from boarding commercial airplanes. The data fields for the No Fly List include first, middle, and last name; date and place of birth; citizenship; and passport/ID number – data considered to be PII. The Selectee List is also produced by the Terrorist Screening Center and contains information related to persons identified as requiring additional screening before being permitted to board commercial airplanes. The data fields for the Selectee List include first, middle, and last name; date and place of birth; citizenship; and passport/ID number – data considered to be PII. Both lists are covered by the TSC SORN entitled "Terrorist Screening Records System (TSRS)," JUSTICE/FBI-019. The privacy compliance issue is whether ICE Demo used PII from the No Fly and Selectee Lists within the scope of the JUSTICE/FBI-019 SORN.

  In order to share information with another agency, the Privacy Act requires an articulation of the sharing and the purpose of the sharing in a routine use. The JUSTICE/FBI-019 SORN identifies multiple routine uses for disclosing information including: "[F]or the purpose of the development, testing, or modification of information technology systems used or intended to be used during or in support of the screening process."

  The Privacy Office determined that S&T used PII from the No Fly List and the Selectee List in order to develop and test the ADVISE technology framework's capability to improve the screening process. Therefore, this use of PII from the No Fly and Selectee lists fits within the scope of routine uses articulated within the JUSTICE/FBI-019 SORN and pursuant to the "routine use" provision of §552a(b)(3) of the Privacy Act of 1974, the Privacy Office

[17] JUSTIC/FBI-019 is a "legacy" SORN which means it was created and is still associated with the agency that was responsible for the information prior to the creation of DHS. The Privacy Office is reviewing all legacy SORNs and reissuing those SORNs as DHS SORNs.

determined that the ICE Demo deployment used No Fly and Selectee List data contained in an existing system of records maintained under a duly published SORN as required by the Privacy Act.

- ICE's Pattern Analysis and Information Collection (ICEPIC) system. This data source is operated by ICE. Data retrieved from this data source included data fields labeled "person id" and "address" – data considered to be PII. While some additional data may be stored directly in ICEPIC, the majority of the PII used by ICEPIC actually comes from other systems.

  The first privacy compliance issue is whether the PII used as part of ICE Demo was specific to ICEPIC or whether the PII actually originated with one of the underlying systems that are part of ICEPIC. Based on information supplied by ICE, the Privacy Office determined that the two PII data fields S&T identified as used in ICE DEMO ("person id" and "address") are not the types of information that would be added directly into ICEPIC and thus the PII originated from underlying systems.

  The second privacy compliance issue is whether the PII in the originating systems was covered by a SORN. The Privacy Office determined that each of the underlying systems was covered by a SORN. Thus, as it relates to PII contained in the ICEPIC data set, ICE Demo used data contained in an existing system of records maintained under a duly published SORN as required by the Privacy Act.

- National Security Entry Exit Registration System (NSEERS). According to ICE, NSEERS is operated by ICE and contained data related to individuals from special interest countries who must register with DHS when entering or leaving the United States. These data fields include subject name, VISA number, passport number, and I94 number – data considered to be PII. The privacy compliance issue is whether PII from NSEERS used in ICE Demo was covered by an applicable SORN. The Privacy Office determined that PII from NSEERS used in ICE Demo was covered by the SORN for the Treasury Enforcement Communications System (TECS), Treasury/CS .244.[18] Thus, as it relates to PII contained in NSEERS data, ICE Demo used data contained in an existing system of records maintained under a duly published SORN as required by the Privacy Act.

- Unconfirmed Overstays. As reported by ICE, this data collection is a subset of the US-VISIT registration data set and relates to individuals who were not confirmed leaving the United States and may be "out of status" regarding immigration. ICE states that this subset of data is not used by any single system and that ICE received this data set from US-VISIT. This data source includes data fields for name, date of birth, and document number – data

[18] See footnote 17regarding the Privacy Office's legacy SORN reissuing initiative.

considered to be PII. As reported by ICE, the Unconfirmed Overstay data comes from DHS's US-VISIT program. The privacy compliance issue is whether PII from Unconfirmed Overstays used in ICE Demo was covered by an applicable SORN. The Privacy Office determined that PII from the Unconfirmed Overstays system used in ICE Demo was covered by the Arrival and Departure Information System (ADIS) SORN, DHS/ICE CBP-001-03. Thus, as it relates to PII contained in Unconfirmed Overstays data, ICE Demo used data contained in an existing system of records maintained under a duly published SORN as required by the Privacy Act.

- Special Interest Tracking System DATA (SITSDATA). According to ICE, SITSDATA was used by INS National Security Unit to track INS Joint Terrorism Task Force (JTTF) cases and incidents. This system is no longer active and no other system currently hosts the data. The privacy compliance issue is whether PII from SITSDATA used in ICE Demo was covered by an applicable SORN. The Privacy Office determined that PII from SITSDATA used in ICE Demo was covered by the SORN for the Treasury Enforcement Communications System (TECS), Treasury/CS .244.[19] Thus, as it relates to PII contained in the SITSDATA data set, ICE Demo used data contained in an existing system of records maintained under a duly published SORN as required by the Privacy Act.

S&T completed a PTA after the conclusion of the ICE Demo deployment which the Privacy Office validated. Further analysis at the time would have focused on the requirement to use PII in the pilot. If PII was required for the pilot, a PIA would be required.

F. Threat Vulnerability Integration System (TVIS)

TVIS used a series of data sets to identify opportunities to test the capability of the ADVISE technology framework to help analysts in I&A. According to S&T, TVIS received Authority To Operate (ATO) and version 1.0 of the ADVISE technology framework was installed on May 18, 2004. In March of 2005, S&T installed version 1.1. In March 2006, S&T installed version 2.0, and in February 2007, S&T installed version 2.1.

S&T and I&A reviewed the existing data sets available to I&A and selected those data sets that appeared to be most useful for the TVIS deployment. As described above and based on the mechanics of the ADVISE technology framework, all data loaded into TVIS was fused together. DHS S&T reports that different combinations of the above data sets were loaded and fused at different times.

[19] Id.

According to DHS S&T, the previously operated TVIS pilot was dismantled in early 2007 and all data was removed. S&T anticipates beginning another TVIS deployment, operated in coordination with I&A. The anticipated deployment of TVIS will be located within I&A's classified computing environment and will be operated by S&T for a period of six months after the data is loaded. The only authorized users of TVIS will be S&T staff. I&A analysts will only be allowed to view demonstrations of the deployment and will not have direct access to any portion of the deployment.

The goal of this next phase of the TVIS deployment is to determine whether the ADVISE technology framework is more effective than the tools currently used by I&A, and to assess which types of documents and data are most compatible with the technical requirements of the Framework's visualization tools, and which areas of research and analysis are most compatible with the ADVISE technology framework.

At the completion of the deployment, S&T will determine whether the effort to load the data and use the ADVISE technology framework's visualization tools for that area of research is effective based upon feedback from I&A. If TVIS proves successful and I&A so desires, all data will be removed from the deployment and TVIS will be transitioned into an operational system. At that time, a separate PIA will be conducted to cover the specifics associated with the operational version of TVIS.

The following is the Privacy Office's analysis of the initial TVIS deployment, including a summary of the data used in the deployment.[20]

Privacy Analysis

The Privacy Office determined the TVIS deployment did not comply with the requirement to conduct a PIA prior to using PII. Given the range of potential PII from the different data sources discussed below, a PIA would have articulated the actual data to be used and described the focus and scope of the testing scenarios and the requirement for PII.

The Privacy Office determined that the TVIS deployment used or generated data that were contained in an existing system of records maintained under a duly published SORN as required by the Privacy Act.

Based on S&T reporting, the Privacy Office determined that any new information that might have been created through the data fusion process were contained in an existing system of records maintained under a duly published SORN as required by the Privacy

[20] I&A provided additional data to S&T for the TVIS deployment that either did not include PII or was not actually used for technical reasons including software conflicts and complexities regarding mapping unstructured data.

Act. The Privacy Office determined that if the new fused data collection was used in a way that included retrieval by personal identifier, that new data would constitute a system of records and be covered by the SORN for the Homeland Security Operations Center Database (HSOC), DHS2005-0028. Thus, as it relates to the new collection of information, the TVIS deployment used data contained in an existing system of records maintained under a duly published SORN as required by the Privacy Act.

The remaining Privacy Act issue is whether each use of each data set contained in an existing system of records was maintained under a duly published SORN as required by the Privacy Act. The following is a privacy analysis of each data source.

- No Fly & Selectee Lists and TSC Daily Summaries. The No Fly and Selectee Lists are the same data sources used in the ICE Demo deployment and described above. The TSC Daily Summaries contain information related to incidents reported by the Terrorist Screening Center and included names and dates of birth of individuals attempting to cross the U.S./Canadian border through one of the ports of entry, as well as the location where the encounter with the individual occurred – data considered to be PII. The No Fly List, Selectee List, and TSC Daily Summaries are produced by the Terrorist Screening Center from the Terrorist Screening Database and covered by the Terrorist Screening Records System (TSRS), JUSTICE/FBI-019. The privacy compliance issue is whether TVIS used PII from these data sources within the scope of the JUSTICE/FBI-019 SORN. As stated in "routine use" section of this SORN, information can be disclosed "[F]or the purpose of the development, testing, or modification of information technology systems used or intended to be used during or in support of the screening process." The Privacy Office determined that PII from these sources was used by S&T in the development of technology to support screening and intelligence purposes. Thus, pursuant to the "routine use" provisions of §552a(b)(3) of the Privacy Act of 1974, TVIS used data contained in an existing system of records maintained under a duly published SORN, as required by the Privacy Act.
- Student and Exchange Visitor Information System (SEVIS). SEVIS data described the names, birthdates, residence location, citizenship, nationality, school attending, and major sought for individuals who traveled to the U.S. seeking a graduate degree – data considered to be PII. The privacy compliance issue is whether PII from SEVIS used in TVIS was covered by an applicable SORN. The Privacy Office determined that the DHS/ICE-001 SORN covered the data within the SEVIS. Thus, as it relates to PII from SEVIS, TVIS used data contained in an existing system of records maintained under a duly published SORN as required by the Privacy Act.
- I&A Finished Reporting. This data source included narrative text describing a topic of interest usually based upon a request for information from a member of the Intelligence Community, DHS or national leadership, law enforcement,

> or infrastructure protection agencies. This data may have included PII. The privacy compliance issue is whether the data used in TVIS was covered by an applicable SORN. The Privacy Office determined that the Homeland Security Operations Center Database (HSOC) SORN, DHS2005-0028 covered the data within the I&A Finished Reporting and thus as it relates to PII from I&A Finished Reporting, TVIS used data contained in an existing system of records maintained under a duly published SORN as required by the Privacy Act.

S&T completed a PTA for this anticipated next stage of the TVIS deployment which was validated by the Privacy Office, indicating that a PIA is required. A draft PIA is currently under review. The future TVIS deployment will adhere to the privacy compliance requirement that no PII be loaded or used until after all privacy protection requirements are met.

VI. Summary of the Privacy Office Review

The Privacy Office determined that the decision to use real data to test general functionality of new technology without first conducting a PIA created unnecessary privacy risks. DHS privacy compliance requirements state that programs (including those focused on research and testing) identify when PII is used, specify the considerations and protections accorded PII during collection, use, maintenance, and dissemination, and explain the specific privacy protections built into the system to minimize any potential privacy risks.

All DHS uses of technology, including research and testing, are required to address potential privacy protection concerns before PII is loaded or used. In conducting the above described ADVISE deployments, S&T did not conduct the required PIAs before loading or using PII. All of the ADVISE deployments, however, used data contained in existing systems of records maintained under duly published SORNs as required by the Privacy Act.

VII. Recommendations

In response to the previous ADVISE deployments and the overall approach to testing new technologies, the Privacy Office recommends a two-tiered response to S&T's testing of the ADVISE technology framework, one short term and the other long term.

A. Short-Term Responsive Action

The Privacy Office recommends the following short-term actions in response to the privacy issues raised by S&T's use of the ADVISE technology framework.

1. Research and development efforts determined to not include the use of PII do not raise privacy protection concerns and can, from a privacy compliance perspective, be reinitiated and continued - these include the ICHAST work, as well as the BKMS Pathogen Demonstration and BKMS Bio Encyclopedia deployments.
2. Other research and development efforts of the ADVISE technology framework and ADVISE deployments which use PII can, from a privacy compliance perspective, be reinitiated and continued provided S&T complete all required privacy compliance documentation and incorporate technical and procedural privacy protections demonstrating compliance with privacy protection requirements before PII is loaded or used.
3. Testing of the ADVISE technology framework should initially use either synthetic data or non-PII in order to demonstrate the value of the Framework. Only after that value is established and only after available data sources are inventoried to identify data well-suited to the use of the ADVISE technology framework should PII be used during future ADVISE deployments. In general, it is inappropriate to use PII to test the generic functionality of technology.
4. All applicable privacy compliance documentation must be approved by the DHS Chief Privacy Officer and published (as appropriate) prior to loading or using PII in connection to the ADVISE technology framework. This documentation must demonstrate a thorough analysis of the potential privacy protection issues raised by the ADVISE technology framework itself, the testing of the Framework (including scope and measures of success), and whether there is a possibility that a successful test using PII (if justified) could lead to an operational decision.

B. Long Term Responsive Action

The Privacy Office recommends the following long-term actions in response to the privacy issues raised by S&T's use of the ADVISE technology framework.

1. The Privacy Office and S&T should collaborate to ensure that privacy compliance requirements are integrated into S&T's integrated process team management approach. This will ensure that all future research efforts comply with the privacy protection requirements as early as possible in the research life cycle.
2. The Privacy Office and S&T should collaborate to review S&T's inventory of research and development projects to ensure that all current S&T projects using PII comply with privacy protection requirements.
3. The Privacy Office and S&T should collaborate to develop a collection of synthetic privacy compliant test data so that S&T testing can benefit from the unique characteristics of PII without creating unnecessary privacy protection risks. Data sets containing PII may only be used in testing pursuant to the approval of both the Privacy Office and S&T.
4. The Privacy Office and S&T should collaborate to develop privacy technology guidance for all S&T research and development efforts so that future research will benefit from an integration of privacy awareness and protections. This guidance would provide recommendations for the integration of privacy protections into the early stages of research and development efforts.
5. The Privacy Office and S&T should collaborate to develop privacy technology guidance specific to the ADVISE technology framework so that all uses of the ADVISE technology framework can benefit from awareness and integration of the specific aspects of privacy protection requirements that are raised by the nature and use of the ADVISE technology framework.

VIII. Going Forward

As of this report, there are three anticipated ongoing deployments of the ADVISE technology framework: ICAHST, TVIS, and BKMS Next Generation. S&T is current with all privacy compliance requirements related to these three deployments.

- ICAHST: The PTA indicated no PII and therefore no further privacy compliance documentation is required at this time.
- TVIS: The PTA indicated that PII will be used and therefore a PIA is required. A draft PIA is currently under review.
- BKMS Next Generation: A PTA will be completed which will indicate whether PII will be used and thus whether further privacy compliance documentation will be required.

S&T and the Privacy Office are currently working through the privacy compliance documentation process for these deployments and PII will not be loaded or used until the privacy compliance documentation is approved by the Chief Privacy Officer, and, as appropriate, published to the public.

The Privacy Office and S&T are collaborating to identify the appropriate administrative mechanism to integrate privacy compliance requirements into both the management and life cycle of S&T research portfolios. In addition, S&T recently announced the appointment of a full-time Privacy Act Officer.

The Privacy Office is working on a PTIG for operational technologies to further embed privacy protection into Departmental uses of operational technology. The Privacy Office also recommended that privacy protection requirements be added into the DHS System Development Life Cycle. These efforts will be used to build privacy technology guidance to support the substantive aspects of research programs such as the ADVISE technology framework and the procedural mechanisms governing all S&T research efforts.

When research initiatives use PII, privacy compliance requirements apply and become part of the requirements that must be factored into the nature of the research, development, and ultimately the effect of the technology. The efforts involved in this review and the collaboration between the Privacy Office and S&T as they work through the above recommendations will hopefully guide future research efforts that trigger privacy protection concerns.