Privacy Impact Assessment
for the

Repository for Analytics in a Virtualized Environment (RAVEn)

DHS/ICE/PIA-055

May 13, 2020

Contact Point
Alysa D. Erichs
Acting Executive Associate Director
Homeland Security Investigations
U.S. Immigration & Customs Enforcement
(202) 732-5100

Reviewing Official
Dena Kozanas
Chief Privacy Officer
Department of Homeland Security
(202) 343-1717

Abstract

The United States Immigration and Customs Enforcement (ICE) Homeland Security Investigations (HSI) Innovation Lab is developing an analytical platform called the Repository for Analytics in a Virtualized Environment (RAVEn). RAVEn will facilitate large, complex analytical projects to support ICE's mission to enforce and investigate violations of U.S. criminal, civil, and administrative laws. RAVEn also enables users to develop new tools to analyze trends and isolate criminal patterns as HSI mission needs arise. ICE is publishing this Privacy Impact Assessment (PIA) and its associated appendices because the analytical tools that reside on RAVEn access and store personally identifiable information (PII) retrieved from data systems owned by DHS, other governmental agencies, and commercial databases. ICE will regularly update the appendices to this PIA to reflect any new system connection or tool developed on the RAVEn platform.

Overview

HSI is a directorate of ICE that investigates, disrupts, and dismantles transnational criminal threats facing the United States. As the largest investigative unit in DHS, HSI uses its unique immigration and customs legal authorities to protect the United States from illegal activity with a border nexus. This activity includes immigration crime; human rights violations; human smuggling; smuggling of narcotics, weapons, and other types of contraband; child exploitation; financial crimes; cybercrime; and export enforcement issues. HSI is composed of eight separate but interconnected divisions organized around functional and subject matter areas. Overall those divisions host 185 domestic field offices, 62 international attaché offices, and a number of intelligence fusion centers, joint task forces, and other specialized units. Historically, these offices and units separately developed and purchased analytical tools to assist their specific investigative efforts. HSI Innovation Lab was created to centralize HSI's development and use of analytical tools in order to reduce duplicative efforts and increase HSI's investigative efficiency.

HSI Innovation Lab's primary mission is developing products and tools for use by HSI special agents and analysts in the field by turning data that has been collected by ICE into valuable insights. HSI Innovation Lab is not an investigative unit;rather, it is dedicated to identifying types of analytical tasks required by HSI offices and matching them with the tool or combination of tools best suited to accomplish each task. HSI Innovation Lab focuses on merging the latest in open source data management technologies with a flexible development philosophy that can quickly be adapted to the ever-changing landscape that is combating complex criminal organizations.

At a high level, the HSI Innovation Lab:

- Identifies a mission need, likely affecting multiple programmatic areas, that is best addressed using analytics;[1]
- Selects and tailors an analytical tool, which may be repurposed from another ICE analytical application outside of RAVEn, to address the mission need;
- Iterates and tests the tool in the non-production environment of the RAVEn analytical platform for accuracy and effectiveness (i.e., the tool's analysis yields sufficiently predictive results); and
- Deploys the tool, or the results generated by the tool, in a manner that best meets the mission need, on the RAVEn platform for use by appropriate HSI analysts and special agents.

HSI Innovation Lab is dedicated to building reusable analytical tools that are designed to be modular and focused on accomplishing specific functions. The objective is to reduce overall cost and time by avoiding duplication of efforts across programmatic areas. HSI Innovation Lab capabilities are powered by RAVEn, an advanced analytical platform.

RAVEn

RAVEn is a cloud-based platform that enables HSI users, who are law enforcement officers or support law enforcement, to perform analytics across raw or unevaluated datasets using a suite of search, analytical, and reporting tools. It is specifically designed to combine and maximize the efficiency and capabilities of open-source tools.[2] RAVEn leverages capabilities from tools purchased by an individual HSI program/division so that they may be reused for multiple tasks, reducing duplication of efforts across HSI.

RAVEn will not replace ICE's traditional criminal investigatory case management systems. Rather, RAVEn will primarily perform large, complex analytical projects at HSI. RAVEn will curate and chain together seemingly disparate raw datasets by performing advanced analytics across multiple datasets, thus enabling users to accomplish tasks currently considered too large or complex for existing systems. HSI Innovation Labs will use Artificial Intelligence (AI), or machine learning, in many RAVEn tools to better recognize patterns in data and enhance the tool's effectiveness.

[1] Analytics is a computation of data and statistics for the purposes of evaluation, analysis, or prediction.
[2] Open Source can be defined as "software for which the human-readable source code is available for use, study, reuse, modification, enhancement, and redistribution by the users of that software." *See* HOST - Open Source in Government Challenges and Opportunities, *available at* https://www.dhs.gov/sites/default/files/publications/Open%20Source%20Software%20in%20Government%20%E2%80%93%20Challenges%20and%20Opportunities_Final.pdf.

For example, using the RAVEn analytical tools framework, HSI Innovation Lab is training a tool using AI to analyze HSI Reports of Investigation (ROI) stored in the HSI Data Warehouse[3] and extract relevant information and relationships in the data. The tool enables users to search free text narratives and some structured metadata in ROI data to extract entities (e.g., individuals, businesses, vehicles, phone numbers, or addresses) and identify interconnections (i.e., relationships and patterns) among those data fields. This type of analysis can identify relationships in the data that might not otherwise have been found without machine learning due to the number of records and the complexity of their differences.

To train the tool, HSI Innovation Lab personnel annotate records to depict the information and relationships of interest. The tool will use the annotated records to create a model (a computation or a formula formed as a result of an algorithm[4]) that is then used to extract relevant information and relationships. A subset of ROI records, which were initially set aside, are separately analyzed and used to evaluate the efficacy of the model (i.e., ensure the relevant information and relationships the tool is uncovering are the same as those found in the manually analyzed data). This subset of records is commonly referred to as the testing dataset. As needed, HSI Innovation Lab personnel further train the tool (e.g., adjust parameters or run the training data through the model additional times) until the results of the tool match the results of the testing dataset. The model is not considered suitable for deployment until it reliably recognizes and extracts the selectors (i.e., phone numbers, addresses, names) and associations (i.e., the person that uses the phone number that was extracted) that were manually annotated in the testing dataset.

Once the tool has been determined to yield accurate results, it can then be deployed to HSI operational units to continuously and automatically analyze large ROI datasets for connections specified by HSI special agents or analysts. HSI Innovation Lab can then determine whether the tool is applicable on other datasets, such as arrest records found in the ICE Enforcement Integrated Database (EID),[5] to meet future HSI investigative needs.

The RAVEn-developed tools will standardize and organize data; conduct analyses to isolate criminal patterns; conduct trend analyses; and identify weaknesses in criminal organizations that can be exploited by investigators. The incorporation of a vendor tool or tool developed outside RAVEn into the RAVEn environment will necessitate updates to the tool as business rules change. To address this issue, HSI Innovation Lab personnel will train the tool to ensure it continues to produce relevant and accurate results within the RAVEn platform. The type or format of a tool's analytical work product will vary from tool to tool. Each analytical tool HSI

[3] For more information on ROIs and the HSI Data Warehouse, see DHS/ICE/PIA-045 Investigative Case Management System (ICM) *available at* www.dhs.gov/privacy.
[4] An algorithm is a process or set of rules to be followed in calculations or other problem-solving operations.
[5] *See* DHS/ICE/PIA-015 Enforcement Integrated Database, *available at* www.dhs.gov/privacy.

Innovation Lab develops for the RAVEn platform is examined in greater detail in the appendices of this PIA.[6]

RAVEn will ingest either datasets from other systems or manual uploads of investigative records from HSI agents. RAVEn tools can also search and query other systems through application programming interfaces (APIs).[7] RAVEn has incorporated all datasets from its predecessor, the ICE Big Data Environment (ICE-BDE).[8] RAVEn will ingest datasets and maintain the data in an ICE-owned and controlled cloud computing environment. While all ingested data will reside within the cloud system, every tool will be segregated into separate operating environments through user and system access requirements. Data is aggregated in the RAVEn environment and user access is controlled using a centrally managed Attribute Based Access Control (ABAC). Every record brought into the system is assigned one or more attribute(s), referred to as a "Security Bucket ID" within RAVEn. Users are then granted permissions to view and add records to that Security Bucket based on their need-to-know and job duties. Regardless of which tool a user is viewing, the user's ability to see a certain type of record is uniform and consistent because it is managed by the RAVEn system as a whole. Similarly, users are granted access to tools based on their need-to-know and job duties. RAVEn's central data store eliminates the need for a separate data store for each application, which would result in multiple copies of many datasets containing PII. RAVEn tools are created with a specific mission need in mind and user roles will vary by analytical tool.

Whether data is ingested into RAVEn or merely queried from the original databases will be determined by the use case as each new tool is developed. At the time HSI Innovation Lab makes a connection between RAVEn and another database, it will execute an interconnection agreement (ICA) with the owner of that database that will document system auditing, logging, oversight, and permissible uses of datasets. HSI personnel may only access the data associated with their use of RAVEn and for which they have a designated need-to-know. This access decision will be made on a tool-by-tool basis.

As stated above, RAVEn is an analytical platform and *not* a case management system. RAVEn also does not alter original source system data. Information and analyses generated within RAVEn and pertaining to ongoing investigations will be manually added to the relevant case file/case management system by the reviewing analyst or agent. For example, if RAVEn ran ROI

[6] The appendices for this PIA detail: all source systems that provide information to RAVEn, all analytical tools developed by the HSI Innovation Lab to support RAVEn, and the data elements each RAVEn tool uses.
[7] An API allows two separate computer systems or software applications to communicate with one another.
[8] ICE-BDE provided users with the ability to perform analytics across disparate ICE datasets in order to identify anomalous behavior. The project fell under the ICE Analytics Program, which seeks to provide users with a tool suite of analytical products from which they can perform search, analytics and reporting. ICE-BDE generated investigative leads and conducted trend analysis used to identify entities of interest to investigators and analysts. BDE was decommissioned in 2018 and its data and capabilities were migrated to RAVEn.

data in the tool described in the example above, any relevant trends or criminal patterns the tool identified would be reviewed and verified by HSI special agents and analysts. The special agent or analyst would then be required to manually input RAVEn work products into a new ROI for investigative follow up. Investigative tips or leads that result from RAVEn products are subject to further investigation by HSI special agents prior to any concrete law enforcement action.

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?

Pursuant to the Homeland Security Act of 2002,[9] the Secretary of Homeland Security has the authority to enforce numerous federal criminal and civil laws. These include laws residing in Titles 8, 18, 19, 21, 22, 31, and 50 of the U.S. Code. The Secretary delegated this authority to ICE in DHS Delegation Number 7030.2, Delegation of Authority to the Assistant Secretary for the Bureau of Immigration and Customs Enforcement and the Reorganization Plan Modification for the Department of Homeland Security (January 30, 2003). ICE has been authorized to collect information under 5 U.S.C. § 301; 8 U.S.C. §1103 and 1105; 8 U.S.C. § 1225(d)(3); 8 U.S.C. § 1324(b)(3); 8 U.S.C. § 1357(a); 8 U.S.C. § 1360(b); 19 U.S.C. § 1; and 19 U.S.C. § 1509.

1.2 What Privacy Act System of Records Notice(s) (SORN(s)) apply to the information?

The data ingested and maintained by RAVEn from other DHS or Federal agencies is controlled and covered by source system Systems of Records Notices (SORN). The specific SORNs for each relevant dataset will be listed in the appendices of this PIA.

Manual uploads of records by HSI onto the RAVEn platform are covered by the DHS/ICE-009 External Investigations SORN[10] as they will be collected through investigative processes and used for investigations of violations of the law within ICE's jurisdiction.

RAVEn analytical products that are used for lead generation are governed by the system of record in which the product will ultimately be stored. For example, if ICE stored these analytical products in its Investigative Case Management system (ICM),[11] then the DHS/ICE-009 External Investigations SORN would provide coverage.

---

[9] Pub. L. 107-296, Nov. 25, 2002.
[10] DHS/ICE-009 External Investigations, 75 FR 404 (Jan. 5, 2010). Note: this SORN is currently in the process of being updated.
[11] *See* DHS/ICE/PIA-045 Investigative Case Managements System (ICM) *available at* www.dhs.gov/privacy.

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

ICE has created a System Security Plan (SSP) to support RAVEn's Authority to Operate (ATO).

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

Retention schedules for data within RAVEn will be determined by the source systems from which they originate. All data ingests are also tagged with the source system retention schedule. Data in RAVEn will be refreshed from the source systems at a regular rate, and therefore will adhere to the source system schedules. As source system information refreshes, it will delete any data within RAVEn designated for destruction. Ad-hoc data uploads will be retained in the same manner as their associated case file. Case files are routinely retained for 20 years after the case is closed in accordance with legacy customs schedule N1-36-86-1-161.3 (inv 7B).[12] An ICE-wide updated schedule for investigative records is being developed and will be submitted to NARA for approval.

The RAVEn platform ties any visualizations (i.e., maps, graphs, charts of data points) or analytical products it creates to the underlying records that a RAVEn tool analyzed. When RAVEn updates source records on the platform the analytical product derived from those records are also updated on the RAVEn platform. The appendices of this PIA contain citations to all published privacy documentation of ingested datasets, which contain the relevant retention schedules for ingested data. Analytical products that do not generate a lead or are unused are considered intermediary records and will be deleted when no longer needed as specified by General Records Schedule 5.2 item 020. If analytical records are marked by an analyst or agent as connected to an ongoing investigation or case, then the record will be retained for the same length of time as the associated case file, 20 years after the case is closed.

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.

RAVEn does not collect information directly from individuals. All information accessed and analyzed by RAVEn is provided by government agencies and commercial providers. Some

[12] Records retention is made in accordance with legacy customs schedule N1-36-86-1-161.3 (inv 7B), *available at* https://www.archives.gov/files/records-mgmt/rcs/schedules/departments/department-of-the-treasury/rg-0036/n1-036-86-001_sf115.pdf.

source systems may be subject to the PRA and will state the OMB control number in their respective PIAs. The PIA for each source system can be found in the appendices of this PIA.

Section 2.0 Characterization of the Information

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

RAVEn operates as a platform for a variety of analytical tools that operate across disparate datasets. RAVEn employs user access restrictions at the data element level and robust user auditing controls to compartmentalize data based on a user's need-to-know. The information contained within the RAVEn platform is sourced from systems as described in this PIA's appendix and from information collected by HSI personnel during their investigations. The information contained in the RAVEn platform will continuously change as analytical tools are developed.

On an ad hoc basis, an HSI agent may manually upload records directly into RAVEn for use by a specific tool for a pre-designated purpose. Since RAVEn is not a case management system, the agent must also include the information in appropriate recordkeeping systems, such as ICM. The following are examples of the types of information that are obtained through HSI investigative processes and could be uploaded to RAVEn:

- Photographs, video, and/or documents obtained during surveillance of subjects of an investigation.
- Audio or video recordings of interviews conducted by HSI special agents.
- Documents or other information obtained pursuant to search warrants, subpoenas, or court orders.
- Information shared by foreign partners with HSI special agents pursuant to treaties or other legal frameworks.

RAVEn also accesses and stores law enforcement, immigration, border inspection, criminal, visa, and publicly available information from U.S. government and commercial databases. RAVEn obtains information from other systems either via bulk data transfer or through queries of the system. ICE uses RAVEn to isolate patterns of activity which are indicative of criminal activity and provide investigators access to the information needed to successfully disrupt and dismantle criminal networks. Pattern isolation is most successful if a tool has all relevant information and large datasets, thus the more information ingested by the tool will dramatically decreases the risk of introducing error or bias into RAVEn machine learning models.

Certain U.S. government data systems accessed by RAVEn are governed by information sharing rules which prohibit bulk data transfer. For these systems, RAVEn allows users to conduct queries based on specific information. The following examples are representative of the differences between information obtained in bulk versus information obtained via query:

- The RAVEn platform will ingest and store U.S. and foreign trade records from other governmental systems.[13] Tools being developed within the RAVEn platform will identify organizations that are involved in exporting sensitive material in ways that are dissimilar to other organizations in their geographic area.
- The RAVEn platform performs queries in the National Crime Information Center (NCIC)[14] using specific identifiers (e.g., name, date of birth, sex). These queries are conducted once an individual has been identified as a subject of interest. The queries return information related to a subject's criminal history and other relevant information, which is then added to a RAVEn analytical product.

RAVEn may contain PII relating to individuals who are non-immigrants, immigrants, U.S. citizens, or lawful permanent residents. The categories of information collected, used, disseminated, and maintained in the RAVEn source systems include:

- *Biographic* – Includes an individual's name, spouse's name, children's names, aliases, gender, date of birth, birth certificate, place/country of birth, address (current and former), phone number, country of citizenship, country of residence, Alien Registration Number (A-Number), Social Security number (if available), passport number, email address, usernames for social media, etc.
- *Biometric* – Includes fingerprint images, fingerprint identification numbers, and photographs.
- *Travel* – Includes visa information (e.g., number, country of issuance, expiration date), passport number, border crossing card number, and arrival and departure information.
- *Location-Related* – Includes address information, geotags from metadata, or geolocation information from surveillance activities, witness accounts, or commercially available data. Source systems might also include data derived from third party license plate reader cameras.[15]

[13] *See* Appendix A of this PIA for more information.
[14] See National Crime Information Center (NCIC) Privacy Impact Assessment, *available at* https://www.fbi.gov/file-repository/pia-ncic.pdf/view.
[15] For more information on License Plate Reader technology, *see* DHS/ICE/PIA-039 License Plate Reader Data from a Commercial Service *available at* www.dhs.gov/privacy.

- *Immigration-Related* – Includes class of admission (e.g., visa type), immigration status, immigration benefit application information (e.g., adjustment of status), immigration history, and employment history.
- *Criminal History* – Includes outstanding warrants, criminal charges and arrests, arrest dispositions, NCIC codes for crimes charged and convicted, FBI number, and sentencing data.
- *Financial Data* – Includes data on suspicious financial activity, currency transaction reports, and currency or monetary instrument reports.
- *Telecommunications Data* - Includes telecommunication device identifiers (e.g., Internet Protocol Addresses, Electronic Serial Number), telecommunications usage data (e.g., date/time of call, dialed number), and biographic information on targets of investigations, potential targets, associates of targets, or any individuals or entities that receive calls from these individuals. This does not include the content of phone calls.
- *Case-Related* – Includes case number, digital copies of evidence, court records, incident reports, arrest reports, seizure reports, electronic surveillance reports (ELSURs), and contents of Reports of Investigation (ROIs). ROIs are narratives documenting investigative activities. ROIs may describe case details and statuses, summaries of events (e.g., target encounters, witness or victim interviews, surveillance activities), agent observations, descriptions of evidence, and any other information relevant to a case. Case-related data may also include publicly available open source information, such as social media posts or information found in public databases (e.g., county and court records) as well as information obtained from the darknet.[16]

ICE will update its privacy compliance documentation to account for any changes to RAVEn that would impact the collection, use, or maintenance of PII. This could include RAVEn ingesting new datasets or developing new analytical tools. When a new data set is being added to RAVEn that is not owned by ICE, a Privacy Threshold Analysis and an updated appendix to this PIA will be coordinated with the owning agency. The types of data ingested and used by each RAVEn tool is reflected in the relevant appendices of this PIA.

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

All raw datasets analyzed by RAVEn are provided by other government and commercial databases. Some datasets may be ingested in bulk, while other datasets are only queried within their source systems by a RAVEn API (see sec 2.1, above). The data sources and the collection

[16] Information that is only accessible through anonymizing cryptographic software (Tor).

methods for Federal source systems are explained in detail in the source system PIAs and SORNs, which are referenced in the appendices of this PIA. Non-Federal source systems are discussed in the appendices of this PIA.

RAVEn does not collect information directly from individuals. On an ad hoc basis, HSI agents may manually upload records directly into RAVEn for use by a specific tool for a pre-designated purpose. These records stem from investigative requests, such as warrants or subpoenas, or other investigative activity.

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

Yes. Commercially available data and open source (public) data may be ingested or accessed by RAVEn. ICE routinely uses publicly available commercial data to verify or update information about an individual, such as the person's current address/geolocation, civil litigation records, criminal history, or incorporation records. This data is targeted and is used to cross-check, confirm, and broaden the scope of investigations and intelligence gathering efforts. Information RAVEn accesses from commercial data sources is further detailed in the appendices.

In addition, HSI users may also upload an individual's or business's publicly available information on an ad hoc basis. This information will only be collected pursuant to active and ongoing investigations. Investigators will manually capture and upload information to the RAVEn platform including:

- Social media history;
- Content from public web sites;
- Advertisement/marketing posting from both the open internet and dark net; and
- Business registrations.

RAVEn will also provide tools to HSI users that will simplify the process of collecting information from open source systems by using automated scripts that mimic the actions of investigators. RAVEn tools that use commercial data or assist in its collection will be detailed in the appendix section of this PIA after they are developed.

2.4 Discuss how accuracy of the data is ensured.

All raw datasets analyzed by RAVEn are provided by other government and commercial databases. RAVEn relies on the accuracy and integrity of source system data. The accuracy of data from manual uploads will be dependent on the collection methods used by the HSI agent. The accuracy of DHS-owned data, other government agency data, and commercial and public source

data depends on the original source. HSI Innovation Lab endeavors to ensure data from ingested sources is routinely refreshed in RAVEn as close to real time as possible. This way RAVEn data can be corrected when the data in source systems is updated. RAVEn cannot alter data in source systems. If RAVEn users notice inaccurate data in RAVEn, the RAVEn system owner will update or notify the source system administrators accordingly.

If an HSI agent manually uploads investigative evidence into RAVEn, it is the responsibility of the agent and his or her supervisor to ensure the data is accurate. In the event uploaded data is later identified as inaccurate, that agent is required to modify his or her own uploads to correct the data. If the user who uploaded the data no longer has access privileges for RAVEn, it is the responsibility of a supervisor or systems administrator to make the appropriate changes to the incorrect data.

In addition, data quality is strengthened by the policy requirement that all RAVEn users attach a case number to the uploaded data, when one is available. Attaching a case number links the data to a particular investigation or analysis project, thereby helping to ensure the inclusion of the data is appropriate for investigative or analytical purposes.

Moreover, RAVEn analytical products that result in investigative tips or leads are subject to further investigation by HSI special agents prior to any concrete law enforcement action. HSI agents and analysts receive training on the importance of verifying information from RAVEn before including it in any analytical report or using it as the basis for any formal law enforcement action, such as opening an investigation or conducting an enforcement activity. Information is always handled with concern for its ultimate potential use as evidence in court; as such, HSI personnel are very careful to ensure the quality and integrity of the information to avoid damaging an investigation. HSI personnel are also responsible for ensuring that the information is relevant to an investigation and if an analytical product is found to be irrelevant or incorrect ICE will not retain the information.

2.5 Privacy Impact Analysis: Related to Characterization of the Information

Privacy Risk: There is a risk that information will be included in the system that is not necessary or relevant to accomplish the system's purpose.

Mitigation: The risk is mitigated. Initial information collections are governed by source system business rules and HSI requirements to comply with law and DHS policy. At the time of each database connection, an ICA will be created between the source system owner and RAVEn personnel that will determine whether a dataset is necessary for the platform's mission purpose. As each tool is developed or any new data set is added, HSI Innovation Lab personnel will complete a Privacy Threshold Analysis in consultation with the ICE Privacy Division and the owning agency to ensure that the ingestion or querying of datasets is relevant and necessary for

the tool's purpose. The governance process is overseen by HSI Innovation Lab senior managers. The existence of this governance process will help to ensure new data sources are appropriately vetted, as well as compliance with the DHS Fair Information Practice Principles.[17] RAVEn imposes user access restrictions and permissions at the record level so that manually uploaded data is only viewable to a user if it is tagged as associated with his or her profiles. Access controls follow the data and attach to an analytical work product. This prevents manually uploaded data from being accessed, used, or transferred in contravention with the security requirements of that data.

Privacy Risk: There is a risk when ingesting bulk data that changes or corrections made to PII in the underlying source systems will not be reflected in RAVEn, thus leading to inaccurate or out-of-date information being stored, shared, or used for mission purposes.

Mitigation: This risk is partially mitigated. At the time of each database connection HSI Innovation Lab will create, in consultation with source system owners, an ICA that will determine system refresh rates, auditing, logging, oversight, and data transfer rates between the systems. Under the agreement, system performance will be monitored to ensure the rate of data flows. HSI Innovation Lab endeavors to ensure routine ingests occur as often as possible. If HSI users determine that information is inaccurate in a source system, the RAVEn system administrator will notify the administrators of that source system. The data will then be refreshed in RAVEn according to the refresh schedule. All analysis in RAVEn is conducted via real time interfaces, meaning that a tool's work product would change as soon as the underlying data is changed.

This risk is further mitigated through the ICE analytical vetting process. ICE places great importance on ensuring analytical products are thoroughly vetted by trained analysts before being sent to the field for further investigation. HSI analysts using a RAVEn tool (not Innovation Lab personnel) will check source systems to corroborate and confirm the accuracy of an analytical match. At that time, any inconsistencies between source system data and RAVEn data can be reconciled.

Privacy Risk: There is a risk that data may become corrupted during transmission from RAVEn source systems to RAVEn tools.

Mitigation: The risk is mitigated. The HSI Innovation Lab has created technical measures to ensure that data transmittals between source systems and RAVEn do not affect the integrity of the datasets. Although data may be accessed by many RAVEn tools, the source datasets are ingested only once into RAVEn. Files ingested by the RAVEn platform have a

[17] For more information on the Fair Information Practice Principles *see* Privacy Policy Guidance Memorandum 2008-01/Privacy Policy Directive 140-06, The Fair Information Practice Principles: Framework for Privacy Policy at the Department of Homeland Security. *available at* www.dhs.gov/privacy.

cryptographic hash created at the time of ingestion. When a file or record is analyzed by a RAVEn tool, it is checked against the original cryptographic hash to ensure it has not been modified in any way.

Section 3.0 Uses of the Information

3.1 Describe how and why the project uses the information.

RAVEn user analysis of data will vary depending on the tool accessing the data on the platform. An examination of the use of data for each analytical tool is available in the appendices of this PIA. RAVEn will be a critical platform for the development of tools to analyze information and identify connections within disparate datasets that had proved to be previously too difficult to synthesize. The automated nature of RAVEn analysis greatly increases the efficiency and effectiveness of certain aspects of HSI's otherwise manual and labor-intensive work. In so doing, RAVEn facilitates more efficient investigations of immigration crime; human rights violations and human smuggling; smuggling of narcotics, weapons and other types of contraband; child exploitation; transnational gangs; financial crimes; cybercrime; and export enforcement issues.

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.

RAVEn will have the ability to conduct queries of datasets accessed from multiple databases to discover predictive patterns and connections between entities and events. RAVEn tools will assist users in recognizing relationships between disparate or previously un-synthesizable data. HSI users will do this to develop timely, actionable leads needed to accomplish law enforcement and criminal intelligence missions. All analytical products created by an agent or analyst are reviewed and refined by at least one other analyst or agent before a RAVEn product is entered into a case management system as a lead. RAVEn analytical products that result in investigative tips or leads are subject to further investigation by HSI special agents prior to any concrete law enforcement action.

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

Law enforcement personnel from other DHS components may be granted user access to certain tools residing on the RAVEn platform if they are designated as Task Force Officers (TFOs)

with HSI,[18] have the required security clearance, and have a demonstrated need-to-know. At the time of each new tool's development or system connection, business rules for the system and privacy compliance documentation will be updated. The business rules and privacy compliance documentation will determine user access, system auditing, permissible uses of the tool or dataset, and oversight of the tool.

3.4 Privacy Impact Analysis: Related to the Uses of Information

Privacy Risk: There is a risk of unauthorized access to or inappropriate use or disclosure of information contained in RAVEn.

Mitigation: This risk is mitigated by training, controls on user access, and oversight by HSI management. ICE requires that all personnel take annual Information Assurance Awareness Training, which stresses the importance of appropriate and authorized use of personal data in government information systems. As tools are developed and system connections are made, HSI Innovation Lab reviews and updates all relevant business rules and compliance documentation for the system, which includes user access controls.

While the ability to access RAVEn may be widespread throughout HSI, HSI Innovation Lab will verify and consult with HSI program managers regarding user access to tools and datasets. As a default, a user's access privileges to a tool or dataset on RAVEn is limited to access he or she has to a source system. System administrators or HSI program managers verify personnel's need-to-know before granting access to RAVEn tools. A RAVEn administrator could grant access to general use tools, but a HSI project manager would need to confirm the requestor's need for access to more specialized tools or data. Additionally, when users perform manual ad hoc uploads of data to RAVEn, the HSI Innovation Lab restricts access to the uploaded data to only those users with a need-to-know through record level data tagging. These access layers also apply to tools and their corresponding algorithms. No tool or user can view or access data until it has been determined that access is relevant and necessary. As described in Section 8.3, security and access controls are in place to mitigate the risk of unauthorized individuals gaining access to RAVEn. Regular auditing, logging, and oversight by HSI Innovation Lab personnel ensure that unauthorized access to the systems does not occur.

Privacy Risk: There is a risk that information about individuals unassociated with illicit activities will be included in analysis conducted by RAVEn's tools.

Mitigation: This risk is partially mitigated. ICE's policies and procedures are targeted toward limiting the amount of information that is held by ICE to that which is relevant and necessary to execute its law enforcement mission. The majority of raw datasets analyzed by

[18] Task Force Officers are personnel from other law enforcement agencies that work in concert with HSI for a specific purpose (i.e., human trafficking or counterterrorism) to share information and deconflict efforts on investigations.

RAVEn are accessed from other DHS components and partners that collect under specific law enforcement authority. RAVEn confirms through the data sharing agreement process that systems performing the original collection provide accurate data that is relevant to the administration of the law or other law enforcement purposes.

Furthermore, RAVEn analytical tools are created to determine specific patterns of illegality or criminal threats. Therefore, RAVEn tools' narrow focus should filter out an irrelevant individual's information from the final analytical product. Finally, all analytical products created by an agent or analyst are reviewed and refined by at least one other analyst or agent before a RAVEn product is entered into a case management system. The analyst can remove any data deemed irrelevant to the tool's stated law enforcement purpose.

Privacy Risk: There is a risk RAVEn tools will use data from source systems for purposes beyond the purpose of its original collection.

Mitigation: This risk is mitigated through the development process of the analytical tool and/or system connection. All system connections to RAVEn will be accompanied by a data sharing agreement, which will be reviewed by both the HSI Innovation Lab and the source system owners to ensure the transferred data is used only for purposes consistent with the original collection of the data. As tools develop, the HSI Innovation Lab will update compliance documentation (such as privacy documentation and business rules) before a tool will be allowed to operate. During the review process, the HSI Innovation Lab and ICE Privacy will confirm that the tool's use of information aligns with the stated purposes of its collection as noted in the relevant SORNs and PIAs of the source system.

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.

RAVEn does not directly collect information from individuals, and therefore is unable to provide direct notification that information is being collected. Due to the nature of an analytical tool's use in law enforcement, it is also not feasible to notify an individual prior to the use of their information due to the possibility of harming ongoing law enforcement activities and investigations. Further, providing notice could alert the target of an actual or potential criminal, civil, or regulatory investigation or reveal ICE's investigative interest in a subject. However, general notice of the existence, contents, and uses of this system, and the systems from which it routinely derives its data, are provided by the publication of this PIA and the associated SORNs. When information is obtained from Federal Government forms, notices on such forms state that information may be shared with law enforcement entities.

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

As RAVEn does not directly collect information from individuals, there is no opportunity for individuals to consent, decline, or opt out of providing information to the system. The agency or program that collected the information from individuals is best positioned to provide them with the opportunity to consent, decline to provide information, or opt out. These programs, however, may not be able to provide an individual with the opportunity to consent or decline to the use of their information, as their systems are maintained for a law enforcement purpose.

4.3 Privacy Impact Analysis: Related to Notice

Privacy Risk: There is a risk that individuals are not aware that their information is contained within RAVEn and may not understand how ICE uses the information collected about them.

Mitigation: This risk is partially mitigated by the publication of this PIA, which serves as public notice of the existence of RAVEn and the data its tools access and store. Also, public notice is provided by Federal agencies through source system SORNs that information contained therein could be used for law enforcement purposes.

Section 5.0 Data Retention by the project

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

Raw datasets accessed or ingested by RAVEn will be governed by different SORNs and different NARA approved retention periods. The retention of raw data within RAVEn is determined by the source system connection. As an analytical platform, RAVEn will not hold raw data longer than the source system. As source system information refreshes, it will delete any data within RAVEn designated for destruction. All data ingests are also tagged with the source system retention schedule, thus if a source system is decommissioned (such as ICE-BDE), records will be retained for the relevant retention schedule in RAVEn. Ad-hoc uploads and data from source systems with no retention schedule will be tagged with its associated case file. That data will be retained for 20 years after the case is closed. The retention period for each dataset is outlined in the published privacy documentation cited in the appendix of this PIA.

All visualizations and analytics products created by RAVEn contain data tags that point to the underlying records in the RAVEn database. Analytical products are considered intermediary records which are destroyed upon verification of successful creation of the final document or file (such as a generated lead), or when no longer needed for a business use, whichever is later. When underlying records are deleted through system refreshes, the analytical product will also be deleted

automatically unless marked for an investigation. If analytical products are marked by an analyst or agent as connected to an ongoing investigation or case, then the record will be retained for 20 years after a case is closed.[19] These products will be transferred to the relevant case management system, which has its own processes for ensuring proper data retention and destruction. An ICE-wide updated schedule for investigative records is being developed and will be submitted to NARA for approval.

5.2 Privacy Impact Analysis: Related to Retention

Privacy Risk: There is a risk that information will be retained in the RAVEn environment for longer than is necessary to accomplish the purpose for which the information was originally collected.

Mitigation: The risk is mitigated. RAVEn datasets are refreshed and updated regularly from source systems. Data will also be tagged with the source system retention schedule when it is ingested. Moreover, information sharing agreements or system inter-connection agreements will be finalized by both HSI Innovation Lab and source system data owners. These agreements will specify the applicable records retention policies and procedures for RAVEn to access or ingest source system data. As discussed in section 8 of this PIA, HSI Innovation Lab will regularly audit RAVEn to ensure that data is not inadvertently retained longer than what is reflected in the ICA.

Handling and retention requirements will remain consistent for data as it is accessed by different tools on the platform. Products of RAVEn's analytical tools will be connected to the underlying records in RAVEn. As the underlying records are deleted in accordance with source system retention periods, the analytical products will be updated to ensure that records do not remain in a RAVEn analytical work product past the source system retention schedule.

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.

Access to RAVEn tools and datasets will be determined on a case-by-case basis and is explained in further detail specific to each tool in the appendices of this PIA. ICE may share final analytical products of RAVEn with law enforcement or intelligence agencies that demonstrate a need to know the information in the performance of their missions and in furtherance of HSI's

[19] Records retention is made in accordance with legacy customs schedule N1-36-86-1-161.3 (inv 7B), *available at* https://www.archives.gov/files/records-mgmt/rcs/schedules/departments/department-of-the-treasury/rg-0036/n1-036-86-001_sf115.pdf.

own law enforcement analyses or investigations. These agencies can include federal, state, tribal, local, and foreign law enforcement agencies, as well as relevant fusion centers, FBI Joint Terrorism Task Forces, and international organizations such as INTERPOL.

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

The sharing of PII with law enforcement agencies outside of the Department is compatible with the original purpose for RAVEn source system collections, namely to conduct criminal and civil law enforcement investigations and activities, to administrate the Immigration and Nationality Act,[20] and to ensure public safety. All external sharing of source system information will be determined in the ICA process to ensure that sharing falls within the scope of applicable law, including the published routine uses in the associated SORNs, as listed in the appendices of this PIA.

For example, RAVEn ingests data from the Enforcement Integrated Database (EID).[21] EID is governed by the Criminal Arrest Records and Immigration Enforcement Records (CARIER) System of Records SORN.[22] Like all DHS SORNs, CARIER has routine use G that allows for the sharing of data with Federal, state, local, tribal, and foreign law enforcement agencies if the record, in conjunction with other information, indicates a violation of the law under that agency's jurisdiction. HSI Innovation lab will, through the tool development process and ICA process, ensure that onward sharing of that tool's analytical product is for a purpose in line with CARIER's routine use G. Similarly, if the RAVEn tool compiled EID data with data from the Department of Treasury's Financial Crime Enforcement Network (FinCEN)System,[23] HSI Innovation Lab will ensure that the purpose of the tool, and any authorized onward sharing, also aligns with routine use 3 of the FinCEN SORN.[24] That routine use allows information or records from FinCEN to be shared with authorized domestic governmental agencies charged with administering the law.

The sharing of information that is manually uploaded into RAVEn with law enforcement is compatible with the original purpose for collection, namely to conduct criminal law enforcement investigations and other enforcement activities, to uphold and enforce the law, and to ensure public safety. All external sharing falls within the scope of applicable law, including the published routine uses in the DHS/ICE-009 External Investigations SORN.

---

[20] 8 U.S.C. 1101 *et al*.
[21] *See* DHS/ICE/PIA-015 Enforcement Integrated Database (EID) *available at* www.dhs.gov/privacy.
[22] DHS/ICE-011 Criminal Arrest Records and Immigration Enforcement Records (CARIER) System of Records, 81 FR 72080 (Oct. 19, 2016).
[23] For more information about FinCEN, *see* FinCEN PIA-Data Collection, Storage, and Dissemination *available at* https://www.fincen.gov/sites/default/files/shared/FinCEN_DCSD_PIA.pdf.
[24] FinCEN .003 - Bank Secrecy Act Reports System - 79 FR 20969 (Apr. 14, 2014).

6.3 Does the project place limitations on re-dissemination?

Re-dissemination of RAVEn information by an agency external to DHS is prohibited unless the third agency receives ICE's express authorization. Every ICA that will be created for RAVEn will have a section for unique limits on re-dissemination.

RAVEn users will also follow the Third Agency Rule, which mandates that prior to sharing information or data to a third agency (not contemplated in the original sharing agreement), the agency that intends to share will acquire consent from the agency that provided the data or information. Only individuals with a need to know will be able to gain access to RAVEn analytical products.

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

Each tool on the RAVEn platform that provides access to individuals outside of DHS will be developed with the ability to maintain logs of information shared between agencies. Any disclosure of information derived from an analytical work product created by a RAVEn tool will be noted in the case management system in which the information is documented. HSI users are required to complete and retain DHS Form 191, Privacy Act Disclosure Record, when making any disclosures outside of DHS.

6.5 Privacy Impact Analysis: Related to Information Sharing

Privacy Risk: There is a risk RAVEn data will be disclosed to external partners without a need-to-know.

Mitigation: This risk is mitigated. RAVEn users are required by law and policy to share information with only those external partners who have a demonstrated law enforcement, intelligence, or national security need-to-know. Parameters for re-dissemination of a particular tool's analytical work product will be determined during the development process and noted in the appropriate compliance documentation. All external sharing will be documented in audit logs that will be regularly reviewed by HSI Innovation Lab administrators. RAVEn users will be trained on the appropriate sharing of PII for each RAVEn tool to which they are granted access. If users are unsure whether PII can be shared with certain partners, they will be instructed to contact the ICE Privacy Division for guidance.

Privacy Risk: There is a risk that information for individuals designated as members of a Special Protected Class (SPC)[25] will be shared without authorization.

Mitigation: This risk is partially mitigated. Some datasets ingested or queried by RAVEn may contain information about SPCs. The special sharing and handling requirements required by

[25] *See* 8 U.S.C. § 1367 Penalties for unauthorized disclosure of information of special protected classes.

law and DHS policy[26] will be implemented as part of the RAVEn tool development process and interconnection agreement. RAVEn will properly identify and tag SPC data within the system. As different source system capabilities and ingest methods differ, the method of tagging will vary, but will be a requirement of the tool development process. HSI personnel will be required to take training related to the special restrictions on handling, use, and disclosure of SPC data.

Section 7.0 Redress

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

Individuals seeking notification of and access to any of the records covered by this PIA may submit a request in writing to the ICE Freedom of Information Act (FOIA) Officer by mail or facsimile:

U.S. Immigration and Customs Enforcement
Freedom of Information Act Office
500 12th Street SW, Stop 5009
Washington, D.C. 20536-5009
(202) 732-0660
http://www.ice.gov/foia/

All or some of the requested information may be exempt from access pursuant to the Privacy Act or the Freedom of Information Act (for those individuals who are not U.S. citizens or lawful permanent residents and whose records are not covered by the Judicial Redress Act) in order to prevent harm to law enforcement investigations or interests. Providing individual access to these records could inform the target of an actual or potential criminal, civil, or regulatory violation investigation or reveal investigative interest on the part of DHS or another agency. Access to the records could also permit the individual who is the subject of a record to impede the investigation, to tamper with witnesses or evidence, and to avoid detection or apprehension.

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

Individuals seeking to correct records contained in RAVEn, or seeking to contest its content, may submit a request in writing to the ICE Office of Information Governance and Privacy by mail:

U.S. Immigration and Customs Enforcement

[26] *See* DHS Directive 002-02-01 Implementation Of Section 1367 Information Provisions.

Office of Information Governance and Privacy
Attn: Privacy Division
500 12th Street SW, Stop 5004
Washington, D.C. 20536-5004
(202) 732-3300
http://www.ice.gov/management-administration/privacy

All or some of the requested information may be exempt from correction pursuant to the Privacy Act or the Judicial Redress Act in order to prevent harm to law enforcement investigations or interests.

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

ICE provides general notice via this PIA, source system SORNs, and on ICE's public-facing website about the procedures for submitting Freedom of Information Act and Privacy Act requests.[27] No direct notification to individuals about procedures for correcting RAVEn records is currently provided, since the information in RAVEn is not collected directly from individuals and records in RAVEn contain material compiled for law enforcement purposes. RAVEn contains copies of datasets owned by ICE, DHS components, and the offices of other agencies. Therefore, individuals may also have the option to seek access to and correction of their data directly from those source system owners. Depending on the system in which the data resides, the corresponding SORN might be exempt from certain Privacy Act requirements, such as access and amendment. If so, then individuals' right to be notified about these procedures is limited.

7.4 Privacy Impact Analysis: Related to Redress

Privacy Risk: There is a risk that individuals will be unable to participate meaningfully in the use of their data as maintained in this system or determine whether the system maintains records about them.

Mitigation: This risk is not mitigated. Because the data in RAVEn originates from other systems of record with a law enforcement purpose, individuals' rights to be notified of the existence or non-existence of data about them, and to direct how that data may be used by ICE, are limited. Notification to affected individuals could compromise the existence of ongoing law enforcement activities and alert individuals to previously unknown investigations of criminal or otherwise illegal activity. This could cause individuals to alter their behavior in such a way that certain investigative tools, such as wiretaps or surveillance, will no longer be useful. Permitting

[27] More information is *available at* https://www.ice.gov/foia/request.

individuals to direct the agency's use of their information will similarly interfere with the intended law enforcement use of the system.

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?

RAVEn leverages various technological and policy-based controls, described in greater detail below, to ensure information is used in accordance with the stated practices in this PIA.

Robust User Access Controls: ICE policy requires that RAVEn limits a user's access based on the user's need-to-know and job responsibilities. Access to each tool within RAVEn is controlled by access control lists created at the system and user level in RAVEn. For datasets routinely ingested into RAVEn from another source, the access control lists are based on the user's original access privileges in the source system. This safeguard prevents users from being able to access data in RAVEn that they are unable to access in the source system.

Robust and Accessible User Auditing: RAVEn also implements extensive auditing of user actions in the system. User actions are recorded and stored in audit logs accessible only to authorized personnel. User auditing captures the following activities: logon and logoff, search query strings, records viewed by the user, changes in access permissions, records/reports extracted from the system, and records/reports printed by the system. The system also keeps a complete record of all additions, modifications, and deletions of information in the system and the date, time, and user who performed the action. This information is readily accessible by supervisors and ICE IT security personnel.

General Supervisory Oversight and Monitoring: ICE policy requires that users grant their supervisors access rights to all work they are performing within RAVEn. This enables supervisors to view how their staff are using the system, including the specific data they are working with, and the types of investigations and/or analyses they are conducting. This policy helps to deter and identify individuals who are using the system or its data for unauthorized purposes, and to identify unauthorized use of the system.

Tagging, Supervisory Monitoring, and System Auditing of Ad Hoc Data Uploads: When investigative data is manually imported into RAVEn, HSI agents are required by policy to electronically share this data with their supervisors for review. The HSI supervisor will review the data while it remains in RAVEn's upload queue. Data will only be uploaded from the queue if it is approved. HSI supervisors are responsible for identifying any data imported in contravention of ICE policy. Supervisors may request that the system administrator delete any improperly uploaded data from the system. HSI agents are also required to enter information describing the data being

uploaded, such as source name/category and date retrieved, which helps the supervisor evaluate whether the upload complies with ICE policy and helps other users better understand and evaluate the data. RAVEn will segregate all manually uploaded data through tagging and access permissions. An HSI agent can only upload information for use by a tool that he or she has permission to use. Finally, RAVEn keeps an audit log of all manual uploads by recording user name and date/time of upload.

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

All personnel who have access to the ICE network are required to take annual privacy and security training, which emphasizes the importance of appropriate and authorized use of personal data in government information systems. Users of any tool that contain SPC information will have additional training dedicated to the appropriate handling of such information. In addition, RAVEn users must complete system-specific training that includes rules of behavior, appropriate uses of system data, uploading and tagging records, disclosure and dissemination of records, and system security before they gain access to a system.

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

HSI Innovation Lab system administrators establish user accounts and update user role-based permissions, as needed. Access roles are assigned by an HSI supervisor based on the user's need to know and implemented by an HSI Innovation Lab system administrator. Administrators will limit users to the least amount of privileges within the system whenever possible. Users will have to justify access on a tool by tool basis. Administrators and HSI Supervisors review user access roles regularly to ensure that users have the appropriate level of access. Individuals who no longer require access are removed from the access list.

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?

Any new uses or sharing of information for RAVEn will be approved by HSI Innovation Lab managers. The existence of this governance process will help to ensure that new data sources are appropriately vetted. Any new sharing of information will require an interconnection agreement with RAVEn and, as appropriate, an update to RAVEn's privacy compliance documentation.

Responsible Officials

Jordan Holz
Privacy Officer
U.S. Immigration & Customs Enforcement
Department of Homeland Security

Approval Signature

[Original, signed copy on file with the DHS Privacy Office]

_____

Dena Kozanas
Chief Privacy Officer
Department of Homeland Security

APPENDIX A
Source Systems[28]

Investigative Case Management System (ICM)

Datasets: ICM is the primary case management tool used by HSI for law enforcement investigations into violations of criminal, customs, and immigration laws. ICM stores case information (reports of investigation), subject and associate records, evidence and descriptions of evidence, law enforcement intelligence, reports of suspicious activities, investigative tips and leads, and warrant returns from third parties, including telecommunications data.

Associated Compliance Documentation

- PIA: DHS/ICE/PIA-045 ICE Investigative Case Management (ICM)[29]
- SORN: DHS/ICE-009 External Investigations[30]

Ingest/Refresh Schedule: Daily

Enforcement Integrated Database (EID)

Datasets: EID is an ICE database repository that captures investigation, arrest, booking, detention, and removal information generated through ICE and CBP operations throughout a subject's interaction with the DHS enforcement processes. Records stored in EID can include records documenting arrests, booking, detention, removal, and any information pertaining to an encounter between a subject and a CBP or ICE law enforcement officer.

Associated Compliance Documentation:

- PIA: DHS/ICE/PIA-015 Enforcement Integrated Database (EID)[31]
- SORN: DHS/ICE-011 Criminal Arrest Records and Immigration Enforcement Records (CARIER)[32]

Ingest/Refresh Schedule: Daily

---

[28] This Appendix will be updated when new source systems are queried or their data is ingested into the RAVEn environment.
[29] *See* DHS/ICE/PIA-045 Investigative Case Management (ICM) *available at* www.dhs.gov/privacy.
[30] DHS/ICE-009 External Investigations, 75 FR 404 (Jan. 5, 2010).
[31] *See* DHS/ICE/PIA-015 Enforcement Integrated Database, *available at* www.dhs.gov/privacy.
[32] DHS/ICE-011 Criminal Arrest Records and Immigration Enforcement Records (CARIER) System of Records, 81 FR 72080 (Oct. 19, 2016).

Detainee Telephone Services system (DTS)

Datasets: Enforcement and Removal Operations (ERO) contracts with a provider for management of telephone services at their detention facilities. Under the Detainee Telephone Services (DTS) contract, the Contractor provides detainees with telephone access and provides ICE with investigative reports of those calls. The Contractor reports minimal information on call activity in the reports, including the detainee name and A-Number, the number called, date/time and duration of the call, and the detention facility information.

Associated Compliance Documentation:

PIA: DHS/ICE/PIA-015(b) EID ENFORCE Alien Removal Module (EARM 3.0)[33]

SORN: DHS/ICE-011 Criminal Arrest Records and Immigration Enforcement Records (CARIER)[34]

Ingest/Refresh Schedule: Daily

Pen-Link

Datasets: Pen-Link, and its associated application the Telecommunication Linking System, is HSI's national repository of case-related telecommunications information derived from any type of investigative law enforcement case or event. This data is usually obtained via a subpoena to a telecommunications company (e.g., phone company) and contains transactional details about telecommunications activities. It does not contain the contents of any communications.

Associated Compliance Documentation

- PIA: DHS/ICE/PIA-045 ICE Investigative Case Management (ICM)[35]
- SORN: DHS/ICE-009 External Investigations[36]

Ingest/Refresh Schedule: Daily

Bond Management Information System (BMIS)

Datasets: BMIS is an immigration bond financial management database used to track the issuance, maintenance, cancellation, and revocation of bonds. It contains subject records and records on

[33] *See* DHS/ICE/PIA-015(b) EID ENFORCE Alien Removal Module (EARM 3.0), *available at* www.dhs.gov/privacy.
[34] DHS/ICE-011 Criminal Arrest Records and Immigration Enforcement Records (CARIER) System of Records, 81 FR 72080 (Oct. 19, 2016).
[35] *See* DHS/ICE/PIA-045 Investigative Case Management (ICM) *available at* www.dhs.gov/privacy.
[36] DHS/ICE-009 External Investigations, 75 FR 404 (Jan. 5, 2010).

obligors, including bond management companies. The records contain biographic information, tax information, and financial information.

Associated Compliance Documentation:

PIA: DHS/ICE/PIA-005 Bond Management Information System[37]

SORN: DHS/ICE-004 Bond Management System (BMIS)[38]

Ingest/Refresh Schedule: Daily

Exodus Accountability Referral System (EARS)

Datasets: EARS is a tool that supports efforts by ICE and CBP to enforce U.S. federal export control laws. ICE and CBP must consult with relevant regulatory agencies to investigate whether an export of a particular commodity or service is controlled. ICE and CBP request and track information from licensing agencies using EARS. EARS collects and maintains PII about individuals who are the Principal Party in Interest into possible criminal violations of U.S. federal export control laws. The PII includes business contact information, country of import, license type and number, and type of Principal Party in Interest (exporter, manufacturer, or subject of investigation).

Associated Compliance Documentation:

PIA: DHS/ICE/PIA-021 Exodus Accountability Referral System[39]

SORN: DHS/ICE-009 External Investigations[40]

Ingest/Refresh Schedule: Daily

Significant Event Notification System (SEN)

Datasets: SEN is a reporting and law enforcement intelligence transmissions tool developed by ICE. SEN allows the manual entry, query, and modification of various reports to provide timely information to ICE managers on notable incidents, events, or activities that involve or impact ICE agents and staff in the field. SEN is also used to track news stories regarding ICE and its work. SEN includes data on individuals who are the subject of past or anticipated encounters by ICE personnel (such as witnesses, victims, suspects, and detainees) and individuals from other law

[37] See DHS/ICE/PIA-005 Bond Management Information System, *available at* www.dhs.gov/privacy.
[38] DHS/ICE-004 Bond Management Information System (BMIS), 76 FR 8761 (February 15, 2011). Note: this SORN is in the process of being updated.
[39] *See* DHS/ICE/PIA-021 Exodus Accountability Referral System, a*vailable at* www.dhs.gov/privacy.
[40] DHS/ICE-009 External Investigations, 75 FR 404 (Jan. 5, 2010).

enforcement agencies who contact ICE requesting assistance. SEN also includes data on individuals who are of interest to ICE but are not necessarily part of a past or anticipated encounter in support of its law enforcement intelligence function.

Associated Compliance Documentation:

PIA: DHS/ICE/PIA-023 Significant Event Notification System[41]

SORN:

- DHS/ICE-006 Intelligence Records System[42]
- DHS/ICE-009 External Investigations[43]

Ingest/Refresh Schedule: Daily

ICE Subpoena System (ISS)

Datasets: ICE uses ISS to automate the process of generating, logging, and tracking subpoenas, notices, and summonses that ICE issues in furtherance of its investigations into violations of customs and immigration laws. ISS retains ICE employee and case data for the individual who created the document and the case it pertains to; data regarding the recipient of the subpoena or summons, data regarding the target of the subpoena or summons, and commercial information regarding telephone service providers to identify owners of particular telephone numbers.

Associated Compliance Documentation:

PIA: DHS/ICE/PIA-027 ICE Subpoena System[44]

SORN: DHS/ICE-009 External Investigations[45]

Ingest/Refresh Schedule: Daily

Southwest Border Transaction Record Analysis Center (TRAC)

Datasets: TRAC is a centralized searchable database that contains information about the financial transactions made by global money services businesses (MSBs). The term "money services business" includes any person doing business in one or more of the following capacities: currency dealer or exchanger; check casher; Issuer/Seller/Redeemer of traveler's checks or money orders;

[41] *See* DHS/ICE/PIA-023 Significant Event Notification System, *available at* www.dhs.gov/privacy.
[42] DHS/ICE-006 Intelligence Records System (IIRS), 75 FR 9233 (March 1, 2010).
[43] DHS/ICE-009 External Investigations, 75 FR 404 (Jan. 5, 2010).
[44] *See* DHS/ICE/PIA-027 ICE Subpoena System, *available at* www.dhs.gov/privacy.
[45] DHS/ICE-009 External Investigations, 75 FR 404 (Jan. 5, 2010).

or transmitter for money wires.[46] TRAC provides data and analysis to over 200 law enforcement and regulatory agencies with jurisdiction over money laundering or criminal activity near the southwest border.

Associated Compliance Documentation:

PIA: None. This is a database run by a state government.

SORN: None. This is a database run by a state government.

Ingest/Refresh Schedule: RAVEn will facilitate on-demand queries of the TRAC and will retain the results while the information is relevant to the HSI case for which it was queried. This retention is governed by an MOU.

Privacy Risk: There is a risk that TRAC may misuse or may not properly safeguard ICE data, as it is not governed by the Privacy Act and has no applicable privacy compliance documentation.

Mitigation: This risk is mitigated. RAVEn only ingests data from TRAC; ICE does not share data with TRAC. All queries will be created using non-identifying information, such as date ranges or transaction types. There is no mechanism to share or push data to TRAC within the RAVEn platform.

Privacy Risk: There is a risk that an individual may not know their information collected by TRAC will be used for a law enforcement purpose.

Mitigation: This risk is not mitigated. TRAC records are collected from commercial entities in the MSB. Notice to customers about cooperation with law enforcement is dependent upon the MSB.

Department of the Treasury Financial Crimes Enforcement Network (FinCEN)

Datasets: FinCEN captures information from forms filed by financial institutions in the United States on transactions that exceed certain dollar threshold amounts or may appear on the face of the transaction to be furthering illicit activities. The information is collected via Bank Secrecy Act forms, which include subject biographic information, bank account information, occupation, passport number, country, amount and type of transactions, and other relevant information regarding the financial transaction that an individual has conducted.

[46] A full definition of "Money services business" is provided in 31 CFR 1010.100(ff).

Associated Compliance Documentation:

PIA: FinCEN PIA-Data Collection, Storage, and Dissemination[47]

SORN: Treasury/FinCEN .003 – Bank Secrecy Act Reports System[48]

Ingest/Refresh Schedule: Daily

Federal Bureau of Investigation (FBI) National Crime Information Center (NCIC)

Datasets: The NCIC is a national database maintained by the FBI accessible for use by every criminal justice agency and law enforcement agency nationwide. The NCIC is a computer index of criminal justice information as reported to the FBI by law enforcement agencies throughout the United States and internationally. The NCIC contains records on both property and persons, including warrants, protection orders, stolen property, wanted persons, missing persons, victims of identity theft, violent gangs, terrorists, and other persons of interest to law enforcement. The Interstate Identification Index, which contains automated criminal history record information, is accessible through the NCIC.[49]

Associated Compliance Documentation:

PIA: Privacy Impact Assessment National Crime Information Center (NCIC)[50]

SORN: JUSTICE/FBI-001 National Crime Information Center (NCIC)[51]

Ingest/Refresh Schedule: RAVEn will facilitate on demand queries of the NCIC and will retain this information while the information is relevant to the HSI case for which it was queried.

Office of Biometric Identity Management (OBIM) Automated Biometric System (IDENT) and Homeland Advanced Recognition Technology (HART)

Datasets: OBIM's authoritative biometric database, the Automated Biometric Identification System (IDENT), is the central DHS-wide system for the storage and processing of biometric data. IDENT stores and processes biometric data—digital fingerprints, facial images (photographs), and links biometrics with biographic information, immigration information, and criminal history information associated with identities in the system. OBIM is currently modernizing the biometric

[47] *See* Privacy Impact Assessment Data Collection, Storage, and Dissemination, *available at* https://www.fincen.gov/sites/default/files/shared/FinCEN_DCSD_PIA.pdf.
[48] FinCEN .003 - Bank Secrecy Act Reports System - 79 FR 20969 (Apr. 14, 2014).
[49] For more information on the Interstate Identity Index *see* https://www.bjs.gov/content/pub/pdf/iiince.pdf.
[50] *See* Federal Bureau of Investigation Privacy Impact Assessment of the National Crime Information Center *available at* https://www.fbi.gov/file-repository/pia-ncic.pdf/view.
[51] JUSTICE/FBI-001 National Crime Information Center (NCIC), 84 FR 47533 (Sept. 10, 2019).

database, which will be redeployed as the Homeland Advanced Recognition Technology (HART) system.

Associated Compliance Documentation:

PIA: DHS/OBIM/PIA-001 Automated Biometric Identification System (IDENT)[52]

SORN: DHS/ALL-041 External Biometric Records (EBR)[53]

Ingest/Refresh Schedule: RAVEn will facilitate on demand queries of IDENT/HART and will retain this information while the information is relevant to the HSI case for which it was queried.

U.S. Department of Agriculture (USDA) National Finance Center (NFC) Payroll System

Datasets: NFC is a Government-wide payroll system owned by the USDA that is used to set up federal employee payroll profiles as well as manage payment of salary and benefits. Datasets loaded into RAVEn contains ICE employee personnel information.

Associated Compliance Documentation:

PIA: DHS/ALL/PIA-053 DHS Financial Management Systems[54]

SORN: DHS/ALL-019 Department of Homeland Security Payroll, Personnel and Time and Attendance Records[55]

Ingest/Refresh Schedule: HR data is used for the purposes of personnel visualization in the I-GIS portal and will be updated on an ad hoc basis by the group owner.

Social Security Administration (SSA) Earnings Recording and Self-Employment Income System from the Master Earnings File (MEF)

Datasets: The Master Earnings File (MEF) contains individual earnings histories for each of the 350+ million Social Security numbers (SSN) that have been assigned to workers. These earnings histories are the basis for determinations of eligibility for retirement, survivor, disability and health insurance benefits programs of the Social Security Act and for computations of benefit amounts payable under both of those programs. SSA will transmit to ICE earnings data reported on SSN assigned to non-immigrants who are not authorized to be employed, as required by section

[52] *See* DHS/OBIM/PIA-001 Automated Biometric Information System (IDENT), *available at* www.dhs.gov/privacy. DHS is retiring IDENT and replacing it with the Homeland Advanced Recognition Technology System (HART), which will be discussed in a forthcoming PIA.
[53] DHS/ALL-041 External Biometric Records (EBR), 83 FR 17829 (April 24, 2018).
[54] *See* DHS/ALL/PIA-053 DHS Financial Management Systems *available at* www.dhs.gov/privacy.
[55] DHS/ALL-019 Department of Homeland Security Payroll, Personnel and Time and Attendance Records, 80 FR 58283 (September 28, 2015).

290(c)(2) of the Immigration and Nationality Act.[56] HSI will use this data to determine who may be violating the terms of their admission to the United States. SSA will share the following data elements: SSN, full name, address, employer's name, employer's address, amount of earnings, and tax year.

Associated Compliance Documentation:

PIA: 016-00-SSA/DCS-M-004 Earnings Record Maintenance System[57]

SORN: Earnings Recording and Self-Employment Income System (MEF) SORN, 60-0059[58]

Ingest/Refresh Schedule: Data will be transmitted by SSA annually. Records will be purged after they are processed by RAVEn unless certain data is marked as pertinent to an ICE investigation.

Small Business Administration (SBA) Paycheck Protection Program and Economic Injury Disaster Loan Data

Datasets: The Council of the Inspectors General on Integrity and Efficiency (CIGIE) Pandemic Response Accountability Committee (PRAC) will furnish a copy of the Small Business Administration (SBA)'s Paycheck Protection Program Loan-Level Data (PPP Data) and Economic Injury Disaster Loan-Level Data (EIDL Data) to HSI. PRAC will contribute data related to companies who have applied for and/or received economic relief from the SBA under the PPP and/or EIDL programs. This data will include biographical data, financial data, corporate workforce data (number of employees and duration of employment), and nationality data where that information is known. The purpose of HSI conducting checks against its data holdings is to identify companies who pose a high risk of defrauding the U.S. Government.

Associated Compliance Documentation:

PIA: Office of Capital Access, Capital Access Financial System Privacy Impact Assessment.[59]

SORN: SBA 35-Non-Employment Related Background Checks[60]

Ingest/Refresh Schedule: Data will be transmitted by PRAC on an ad hoc basis as the dataset is updated. Records will be retained in accordance with HSI's records retention schedule.

---

[56] (INA; 8 U.S.C. § 1360(c)(2)).
[57] 016-00-SSA/DCS-M-004 Earnings Record Maintenance System Privacy Impact Assessment, *available at* https://www.ssa.gov/privacy/pia/Earnings%20Record%20Maintenance%20System.updtd%20Sept%2028.htm
[58] 71 Fed. Reg. 1819 (January 11, 2006).
[59] *See* Office of Capital Access, Capital Access Financial System Privacy Impact Assessment, *available at* https://www.sba.gov/sites/default/files/2021-06/SBA%20CAFS%20PIA%20FY21.pdf.
[60] 74 FR 14889 (April 1, 2009).

APPENDIX B
Analytical Tools[61]

Name: Lead Tracking Tool (LTT)

Purpose and Use:

The RAVEn Lead Tracking Tool (LTT) is designed to address the operational needs of HSI field offices by providing a centralized and formalized way to capture lead referrals and outcomes. HSI uses the LTT for initiating, adjudicating, tracking, and collecting and reporting statistics on investigative leads. In the past, HSI managed this process via repeated emails and phone calls and documented the results locally at a field office on a spreadsheet. LTT automates this process and provides greater efficiency in lead management and accuracy in reporting.

The LTT is a role-based application, meaning that a user's system access is defined by their rights and permissions. The application hosts two roles: lead creators (users responsible for developing and referring leads) and lead recipients (users responsible for accepting or rejecting a lead). Lead creators can create and modify all information; lead recipients can add additional information such as people and businesses, attachments and notes, but cannot modify the lead information itself.

The LTT captures, saves, and shares lead generation information manually input by HSI field offices and then captures the outcomes of the leads provided to other HSI field offices. The LTT provides a workflow that guides the lead creator and recipient through a series of steps:

1. An HSI field office discovers a lead that requires follow up by another HSI field office. The initial field office creates a lead within the LTT and routes it to a lead recipient in the other field office.
2. The recipient field office determines whether follow-up action is required. It can either accept the lead for follow up or reject the lead.
3. The recipient field office then provides feedback in the LTT to the creator in the form of reasons for rejection or outcomes of the follow up.

When creating the leads, the lead creator can indicate specifics such as lead type, lead source, seizure date, and lead priority from drop-down menus in LTT data fields. Once a lead is created, the LTT automatically generates an email to the lead recipient. A lead creator may attach program codes to leads, which are used to track investigative and enforcement efforts agency-wide. The codes help to facilitate statistical reporting that may otherwise have gone untracked. The lead creator can also add general identifying information on relevant persons or businesses.

[61] This Appendix will be updated when new tools are added to the RAVEn environment.

When an HSI field office sends a lead to a recipient using the LTT, that recipient will receive an email notification that includes a hyperlink to the LTT. From the LTT, the recipient can view all the information entered into the system by the lead creator. A recipient may accept or reject a lead, as well as add additional recipients (including recipients from additional field offices) to that lead. If a lead is rejected, the recipient must select the reason for the rejection from a provided list. Current selections for rejection are "Insufficient Resources available," "Declined by Assistant United States Attorney," and "Other."

When a lead is accepted and worked by the recipient, he or she can then provide feedback in the form of outcomes to the lead creator. The lead outcome workflow in the LTT allows the recipient to provide the lead creator with information such as if an enforcement action took place, if arrests or seizures were made, and/or additional persons were identified in the course of the investigation. These metrics have typically not been reported back to a lead creator under the current HSI lead dissemination process. The LTT not only provides closure to lead creators, but also enables detailed statistical reporting that was previously unavailable.

Information Collected, Retained, and Disseminated:

The LTT will capture, save, and disseminate lead generation information input by HSI field offices and then capture the outcomes of the leads provided to other HSI field offices.

Information collected, retained, and disseminated could incorporate any information entered by a lead creator that he or she deems relevant to the lead. This information can be manually derived from any of the datasets ingested by RAVEn for which the user has permission to access.[62] Lead information can include: biographical information (name, date of birth, address, etc.), immigration and travel history, citizenship, known family and associates, criminal history, passport or national identification information, organizational information, financial information, employment data, vehicle information, educational history, and case information derived from an investigation (which may include telecommunications data, location information, or information derived from publicly available social media).

Individuals Impacted:

Individuals who are subject to ICE investigations and are material to investigative leads (i.e., subjects of investigation, family members of subjects, known associates). Individuals who are material to leads may be U.S. citizens or lawful permanent residents.

[62] For a complete list of ingested system and their associated datasets, please see appendix A of this PIA.

Additional Privacy Risks and Mitigations:

Privacy Risk: Because the LTT contains information from multiple data sources, there is a risk of creating a lead with inaccurate data or more information than is necessary for the purpose of the lead referral.

Mitigation: This risk is partially mitigated. Lead creation is a manual process and all lead creators undergo user training which stresses the importance of confirming the accuracy of information entered into a lead. Personnel who use the LTT are HSI agents or analysts who are trained on effectively analyzing the information they collect to determine whether it is helpful in developing a lead. The data fields within the LTT are restricted to those that fulfill the purpose of creating and referring a lead.

Privacy Risk: The LTT may refer or disclose information to personnel who are not authorized to access the information

Mitigation: This risk is mitigated. Access to leads is determined by RAVEn's user-based access controls. Lead users may only view leads within the same office as either the creator or the recipient. HSI Innovation Lab limits lead generation capability to select HSI Offices and agents working on certain cases**.** Lead creators may only include data in a lead from a RAVEn dataset to which they have access. Recipients will only have access to data contained within the lead that was referred, and further restrictions on the information can be added depending on the data linked to the lead. All LTT users are trained to disseminate information to only those individuals with a need-to-know to accomplish their respective missions. Lead recipients will only be able to view the information entered into the lead, as receipt of a lead does not grant any permissions or access to any other RAVEn tool or data. Further, users agree to accept the RAVEn system's terms of use, including protecting against unauthorized or improper use or access. HSI Innovation Lab or an LTT user's supervisor can revoke a user's access in the case of misuse.

Name: Natural Language Processing (NLP)

Purpose and Use:

The HSI Innovation Lab uses Natural Language Processing (NLP) on the narrative sections of raw datasets ingested by RAVEn. Typically, datasets such as reports to have structured fields (e.g., dates, times, or locations) and narrative portions entered into unstructured data fields. NLP is a type of Artificial Intelligence whereby a system is trained to recognize, understand, and analyze human language as it is naturally written. Since every individual writes differently (e.g., unique syntax, grammar, terminology, and use of slang or abbreviations), it is difficult for standard analytical systems to process the narrative data. NLP allows a system not only to extract meaning of text from stand-alone words or terms, but also from whole sentences. The technology has to be trained to search the narrative for specific information, so HSI Innovation Lab must identify, in consultation with the end-users, which words, phrases, or concepts should be targeted for analysis. HSI Innovation Lab must then train the NLP until it recognizes text and patterns. NLP cannot be deployed for a program or project until it successfully recognizes and extracts information and associations as confirmed by HSI Innovation Lab in a testing dataset. The process requires intense labor and time for development before the technology can be used to search for specific data. NLP, therefore, can only be used for narrowly defined and pre-determined investigative purposes within RAVEn.

NLP analyzes the investigative free text narratives and some structured data fields to extract entities (e.g., individuals, businesses, vehicles, phone numbers, or addresses) and identify interconnections (i.e., relationships and patterns). NLP analysis can identify relationships across datasets that might not otherwise have been found. For example, a narrative section of an investigative report may include a subject's family ties or points of contact. The NLP tool can analyze the written text in the narrative to determine that an individual listed as a father in one report is also a point of contact in a second report, and a business owner in a third report. This previously unknown connection allows HSI to further analyze potential connections between the individual and persons in seemingly disparate HSI law enforcement actions.

Information Collected, Retained, and Disseminated:

NLP analyzes narrative sections of DHS law enforcement investigatory records.[63] A narrative entry could include any information collected or observed by a DHS agent or officer (including Border Patrol Agents, Customs and Border Protection Officers, ICE Officers, and HSI Special Agents) that he or she deems relevant to the law enforcement encounter or activity. The unstructured data field may include: biographical information (name, date of birth, address, etc.),

[63] Currently the Natural Language Processing tool is used for analyzing narrative sections from DHS Form I-213, *Record of Deportable/Inadmissible Alien*, which is stored in EID, and ICE Reports of Investigation. As more source systems are entered into the NLP tool, Appendix A of this PIA will be updated.

immigration and travel history, citizenship, familial ties, criminal history, humanitarian claims made by an individual, and passport or national identification information. Narrative sections could also contain information about individuals associated with the subject who are deemed relevant to the law enforcement activity, action, or investigation, for example a subject's emergency point of contact information.

Individuals Impacted:

The NLP tool may analyze the PII of individuals who are encountered or investigated by DHS agents or officers, including information about individuals related to or associates of a subject encountered during a law enforcement investigation.

Additional Privacy Risks and Mitigations:

Privacy Risk: There is a risk that more information will be ingested into the NLP tool than is needed to conduct the analysis.

Mitigation: This risk is partially mitigated. As noted above, the data is ingested for analysis by the NLP tool from DHS investigative records systems. The DHS officers who collected or recorded the information are trained to note only information relevant to a law enforcement activity. By design, ingested data may contain more information than is needed, but the NLP tool is trained to recognize information relevant to an investigation. NLP only extracts entities and relationships that it was trained to find. When the NLP tool creates an analytical product or report, only information deemed by an analyst or agent to be of investigative value or that identifies illicit activity will be referred as part of an investigative lead.

Privacy Risk: There is a risk that using the NLP tool on unstructured datasets will result in data errors and the use of incorrect data in investigative analyses.

Mitigation: This risk is partially mitigated. NLP use across these datasets is governed by query or search parameters established by HSI Innovation Lab personnel and believed to be likely to produce relevant results. All analytical products are reviewed and refined by at least one analyst or agent before a lead is referred to the field; this human interaction with information allows for the identification of a need to refine NLP parameters to reduce data errors.

Privacy Risk: There is a risk that individuals will not be given adequate notice that their information will be subject to NLP analysis by ICE.

Mitigation: This risk is partially mitigated. The publication of this PIA helps to mitigate the lack of direct notice to the individual whose information is analyzed by RAVEn. This PIA provides a description of the types of records that will be placed into RAVEn on a routine or ad hoc basis, the purposes for which the information will be used by ICE and the tools, including NLP, that will be used to analyze the data. All records analyzed by NLP are either a DHS Officer's recorded interactions with the public or findings in an investigation. Direct notice for RAVEn's

use of investigative data cannot be mitigated due to concerns of alerting individuals of an ongoing investigation.

Name: ICE Geographic Information Systems (GIS) Portal

Purpose and Use:

The ICE GIS Portal is a mapping platform for visualization and analytics of geospatial and geolocation data. GIS will house data, some of which may contain PII, paired with location information. The ICE GIS Portal allows ICE enterprise users to securely share geospatial information on critical infrastructure, law enforcement data, imagery, and other user-defined geospatial data; as well as perform spatial analyses such as querying,[64] geocoding, routing functions,[65] and zonal statistics.[66]

The ICE GIS Portal acts as a service and platform provider and does not create or own the data managed and shared within. The ICE GIS Portal allows for collaboration and sharing of datasets, maps, applications, and other geographic information between groups and among other system users. The GIS Portal is unique to other RAVEn tools in that is not restricted to HSI use and is intended to be an ICE-wide tool. The ICE GIS Portal also supports other RAVEn tools. RAVEn tools requiring mapping or visualization of location data can leverage the GIS Portal mapping tools natively within their own application without sharing any information with the ICE GIS Portal. The ICE GIS Portal provides the following capabilities:

- Secure access to enterprise geospatial information (i.e., imagery, base maps);
- Mapping and visualization services;
- Geospatial tools and analytic services;
- Geocoding; and
- Web application templates.

The GIS Portal has a default internal geocoder. Geocoding is the process of assigning map coordinates to data like a physical address. Geocoding can be used in reverse as well, giving a physical address to coordinates selected on a map. Some geocoding information cannot be verified by HSI Innovation Lab, such as when an address is in a foreign country with restrictions on sharing geolocation data. HSI Innovation Lab will then use an external geocoder supplied by a vendor. When using an external geocoder, HSI only sends addresses to the vendor, and does not include any of the paired data or additional information. The vendor will then match the physical address with a map coordinate and return the data back to the GIS Portal.

[64] Querying is a filtering and search function for the data paired with location information (i.e., find all ICE Offices in a geographic location).
[65] Routing Functions show pathways between geolocations (i.e., quickest route, shortest route).
[66] Zonal Statistics are values calculated by data that resides within determined geographical zones.

GIS Portal users are limited to individuals with access and credentials on the ICE network. All ICE employees have access to the ICE GIS Portal after signing a rules of behavior (ROB) which specifies the appropriate uses of the GIS Portal. Access to the platform only provides the user the ability to visualize geospatial data. Data within the system or created by other users is restricted. A user's access to maps or information is dependent upon their role and/or group assignment. All initial accounts have a **Viewer** role in the system and must request to access a **Group** or have roles elevated to **Power User**, **Group Owner**, or **Administrator**.

- **Groups** - a collection of users and datasets, often related to a specific region, subject, or project, that are created and managed by the group owner. Members of the group can share and edit information only within the group. **Group Owners** and **Administrators** can add members to the group.
- **Viewer** - can view content shared with all portal users and ask to join groups or add data to a map. A viewer cannot share any data (either with a group or other member on the system).
- **Power User –** is able to see a customized view of the site, use the group's maps, applications, layers, and analytic tools. Power Users can also create maps and applications, edit features, add items to the portal, and view and share content with the Groups of which they are members.
- **Group Owner –** is like a Power User but can also approve requests and add members to Groups. Group Owners are the only non-Administrator users that can share data with all ICE GIS Portal users. To mitigate incorrect dissemination of data, the number of Group Owners is limited. Group Owners are trained on requirements for granting access and disseminating information within GIS.
- **Administrator -** can see all data and groups within the site. Administrators can create new groups and change the ownership of groups and data. Administrators can directly bulk load users to the ICE GIS system and assign them to groups. Administrators are the only users that can change an ICE GIS Portal user role. This user role will be limited in number.
- **Senior Leader -** can see all data within the system but will not have the ability to create or change data in the system. The Senior Leader role is limited to Associate Director level and above and is only granted when requested by Group Owners.

Information Collected, Retained, and Disseminated:

The ICE GIS Portal contains data that is consistent with the ICE mission and authorities. All data remains under the ownership of the users that provide it (i.e., individual ICE GIS Portal users or ICE GIS Portal Data providers). There are no restrictions on the amount or type of data elements that may be used by the GIS Portal. The GIS Portal uses a combination of data created by users and data pulled from both ICE datasets and external datasets. External datasets provide

baseline geospatial points to enrich a map and can be chosen by Power Users or Group Owners to incorporate into their maps. They include publicly available data from other federal agencies (National Oceanic and Atmospheric Administration,[67] Federal Emergency Management Agency,[68] Department of Energy,[69] and the Department of Commerce[70]) and state and local governments (road, imagery, or infrastructure data). Any GIS Power User or higher can also add data from a system from which they have permission to share. This information could include biographical data, data derived from investigations or arrests, or financial transactions.[71]

Additional Privacy Risks and Mitigations:

Privacy Risk: There is a risk data will be shared with individuals who are not authorized to view the information or do not have a need-to-know.

Mitigation: This risk is mitigated. GIS Portal data cannot be shared outside the system. All Group Owners are trained and sign Rules of Behavior (ROB) stipulating that they will verify the need-to-know of any users requesting access to a group with data that contains PII. Only two user roles, Group Owners and Administrators, can share data layers with other groups on the GIS Portal. This limitation reduces errors in data sharing. All data shared within the ICE GIS platform is monitored by GIS Portal administrators in a data usage dashboard to insure no data is improperly disseminated.

Privacy Risk: There is a risk that a user may ingest, upload, or share inaccurate information on the GIS.

Mitigation: This risk is not mitigated. Source systems are responsible for the accuracy of information within their databases that is ingested by the GIS Portal. In addition, users of the GIS Portal maintain full control of the data that they upload (e.g., information taken from spreadsheets) to the GIS Portal. There is a screen that presents to all users when they log onto the GIS Portal that states that users are responsible for managing their own data. Other users of a GIS group may inform the Group Owner of any information on the GIS platform that appears to be incorrect, but it is the responsibility of the data owner to update the information. All Group Owners are trained and sign a ROB stating they are responsible for the quality of the data they share.

Privacy Risk: There is a risk that information within the GIS Portal may be used in a manner that is inconsistent with the original purpose of its collection.

---

[67] For more information *see* https://www.weather.gov/gis/.
[68] For more information *see* https://gis.fema.gov/.
[69] For more information *see* https://edx.netl.doe.gov/group/doe-gis-group.
[70] For more information *see* https://www.census.gov/programs-surveys/geography.html.
[71] The GIS Portal currently holds a dataset generated from addresses extracted from officer arrest and encounter reports derived from EID and ICE personnel data provided by ICE Human Resources. As more systems are added to GIS and RAVEn, Appendix A of this PIA will be updated.

Mitigation: All users of the system are bound by the code of conduct under the ROB they sign to receive access to the system**.** A user may only upload information into GIS from systems that he or she has been granted the authority to collect and permission to share. Administrators of the GIS Portal routinely monitor data creation on the system and will revoke a user's access if information is used for a purpose in violation of the ROB. Further, any system-to-system connection made with the GIS Portal is subject to a Memorandum of Understanding (MOU) or other information sharing agreement in which the connection must be explained and justified.

Privacy Risk: There is a risk that GIS may retain information longer than permitted by the data's retention schedule.

Mitigation: This risk is not mitigated. All data is controlled by the data owner and the GIS Portal does not have a mechanism to track the retention periods of user data. All Group Owners are trained and sign a ROB stating they are responsible for proper maintenance of the data they share.

Name: Optical Character Recognition (OCR) Tool

Purpose and Use:

Many large scale HSI investigations involve the examination and processing of paper documents retrieved by agents. Optical Character Recognition, or OCR, is a technology that enables a system to convert different types of documents, such as scanned paper documents, PDF files, or images captured by a digital camera, into machine-readable text data. The OCR Tool is designed to ingest large quantities of hard-copy records into the RAVEn environment for auditors, analysts, or agents to ensure individuals and businesses are in compliance with labor, customs, and immigration laws. The OCR tool application provides greater efficiency and accuracy in HSI investigations that require entry of information from hard-copy documents into electronic systems for further analysis.

The RAVEn OCR tool is only used for typed forms that contain delineated and structured data fields, such as name blocks. The OCR tool must be customized for each type of form it reads to analyze specific data fields on that form before it can be used. In order to do this, HSI Innovation Lab first divides an image into lines, words, and then characters. Once the characters have been isolated, the OCR tool will compare each character to a set of training images annotated by HSI Innovation Lab. The OCR tool then makes a hypothesis about what alphabetical or numerical character is present in the image. HSI Innovation Lab personnel then confirm whether the hypothesis was correct. By training an OCR tool against a large number of training data, these programs can make highly accurate hypotheses about characters in bounded data fields.

This effort is intended to get maximum value from appropriate implementation and tuning of an OCR tool. As part of this project, additional research and development related to improving the accuracy of OCR, particularly related to handwriting, is required. In the future, HSI Innovation Lab plans to build an in-house handwriting recognition solution to process handwritten forms. This appendix entry will be updated accordingly.

The OCR tool will only be used on scanned forms that were acquired through a formal law enforcement request for information (warrants, subpoenas, and notices of inspection). The OCR tool will automate the process of transferring data from a scanned form into RAVEn, thus increasing efficiency and reducing human error. The OCR tool is designed with an interface for a user to compare the data fields that were auto-filled by the OCR tool against a scanned copy of the form. The user must manually verify that all entries by the OCR are correct before the work product is saved to RAVEn. The OCR tool does not analyze the information that it transfers to RAVEn and is not paired with any technology to identify trends or patterns within the OCR tool's output. There is always an ICE auditor, analyst, or agent monitoring and verifying the work product produced by OCR prior to any further action or manipulation of the data.

Information Collected, Retained, and Disseminated:

The OCR tool will be used to analyze structured fields in standardized hard-copy (paper) documents.[72] These records are scanned directly into the RAVEn platform. On an ad-hoc basis HSI agents and analysts will upload scanned forms into RAVEn. Scanned documents are treated as manual uploads and will be retained after upload for 20 years after the case has closed. The scanned form, like any other data manually uploaded to RAVEn, will only be accessible to HSI agents and analysts designated by the user who uploaded the file. The structured information can include: biographical information (name, date of birth, address, etc.), employer information, translator information, and passport or national identification information.

Individuals Impacted:

Individuals and businesses subject to a law enforcement request for information (subpoena, warrant, notice of investigation, request for information) by HSI.

Additional Privacy Risks and Mitigations:

Privacy Risk: There is a risk that more information than is needed will be ingested into the OCR tool.

Mitigation: This risk is mitigated. The OCR tool is designed to only analyze specifically bounded fields in pre-determined documents. Thus, the only documents uploaded into the RAVEn environment for use by the OCR tool are those the HSI Innovation Lab has trained the tool to analyze. The data that is then extracted during the OCR process is specified through pre-set bounding fields developed to identify the exact data field locations to be processed. Data fields outside the bounding areas are not captured by the tool.

Privacy Risk: There is a risk that using the OCR tool will result in data errors and the use of incorrect data in investigative analyses.

Mitigation: This risk is partially mitigated. The OCR tool will only auto-fill characters it is able to identify. The system will prompt the user to manually review any fields that the OCR tool cannot decipher. If a document is too damaged to read, scanned poorly, or uses an unrecognized font, the OCR tool will not conjecture what the character could be. There are also instances where the OCR tool may make the wrong hypothesis as to the character presented in the image. For example, it may mistake the lowercase letter l for the number 1. In all cases, however, an HSI analyst, agent, or auditor, must manually verify the OCR work product against the scanned copy before it is saved in RAVEn.

[72] Currently the OCR tool is analyzing biographical sections from DHS Form I-9, *Employment Eligibility Verification*. As more datasets are entered into the OCR tool, Appendix A of this PIA will be updated.

Name: RAVEn Collaborate, Organize, Research, Explore (CORE)

Purpose and Use:

In an increasingly digital world, HSI must expand its ability to extract and analyze the vast volumes of data it encounters over the course of a wide range of investigative categories. The purpose of RAVEn is centered upon providing more effective tools to address this challenge. Collaborate, Organize, Research, Explore (CORE) is the main search and analytic tool for all RAVEn data and any additional future data holdings that HSI will acquire. CORE enhances HSI's ability to conduct data analytics, which will be used to identify criminal trends and links among investigative targets that HSI had not previously identified.

CORE provides a simple, easy-to-use search bar, similar to widely used publicly available search engines, from which users search RAVEn datasets. CORE allows users to sort, filter, group, and compare search results from these datasets. These functionalities are available at the data element level and can also be run on the metadata and data tags of the records ingested into RAVEn. This includes data in free text fields, as well as structured data fields. CORE can also assist users in discovering relationships between disparate data by linking and organizing different data elements (e.g., temporally or geographically). Through these methods, users can identify evidence relevant to criminal investigations that are difficult to extract or locate in large datasets.

CORE users can customize their investigative or analytical workflow by displaying search results from multiple data sources in one workspace, called a "canvas." When a user conducts a search within CORE, the results are displayed in a search pane. From that search pane, users can select or drag and drop the relevant results onto the canvas. Users may also import and add external data, such as subpoena return results and publicly available information, to the canvas. The canvas owner can share the canvas and the data contained therein with others by creating a collaborative team associated with that particular canvas and adding users to the team. Within the canvas, users can use the returned datasets to build out and analyze criminal networks and associations, as well as represent linkages of data by creating maps, charts, timelines, and graphs. CORE allows for different methods to visualize data within RAVEn, but it does not alter original source system data. The canvas can be exported in multiple file formats, as well as shared within CORE across a collaborative team. All changes to the canvas are automatically saved within the RAVEn platform. The analytic results of the data, however, are only maintained so long as the data remains in RAVEn. If the source system that feeds RAVEn deletes a record, the data in that record will no longer be available to the user. Users can also share the analytical data and results with limited numbers of other users who have the necessary permissions to view both the data and the related shared canvas.

CORE aggregates data from various sources and allows users to analyze and share the data in analytically useful ways. RAVEn's data ingest and indexing process from external sources into

RAVEn's operational data store is key to making these operations more efficient. HSI Innovation Lab created and optimized custom Extract, Transform, Load (ETL) pipelines for various data sources and data types. Through this ETL process, ingested data is normalized and deduplicated, and entity resolution is performed. To reduce any discrepancies between source systems and CORE, data ingests occur frequently to refresh and update the data. While the CORE tool will have access to nearly all datasets available within the RAVEn platform, HSI personnel may only use CORE to access data associated with their individual RAVEn user permissions and for which they have a designated need-to-know.

The output from CORE may be an organized report or network chart that captures the most pertinent evidence discovered via the application. Examples of these outputs are connections between previously unlinked HSI investigations, new investigative leads, or charts of relationships between targets of an investigation. Information and analyses generated within CORE which pertain to an ongoing investigation will be added to the relevant case file/case management system manually by the reviewing analyst or agent. CORE does not share information externally from RAVEn. The format of the reports will enable use in other HSI law enforcement systems, principally HSI's Investigative Case Management system (ICM).[73]

CORE also has an integrated chat functionality that facilitates collaboration across investigative teams, as well as between HSI offices. This collaboration is essential for identifying targets being investigated by multiple offices, resolving any conflicts, and coordinating linked investigations between HSI offices.

As an application within RAVEn, CORE will possess the same technical, administrative, and physical safeguards as other tools within the RAVEn environment. Data retention and user access is controlled at the record level. Ad hoc uploads of data into RAVEn for use in CORE is assigned the same case number that is associated with an ongoing investigation. HSI supervisors approve any ad hoc data loaded into RAVEn for use in CORE. RAVEn complies with all auditing and logging requirements and tracks all user activities within CORE, including executed searches and results.

Information Collected, Retained, and Disseminated:

Since CORE is available for use on all datasets searchable within RAVEn, potentially any dataset listed within Appendix A of this PIA can be searched by a user if he or she has the appropriate access.[74] Record types and data elements searchable by RAVEn CORE include:

[73] *Supra* note 3
[74] For more information on RAVEn potential pool of data, *see* DHS/ICE-018 Analytical Records, 86 FR 15246 (March 22, 2021), *available at* https://www.dhs.gov/system-records-notices-sorns.

- Biographic information and identifying numbers that reside in source systems;
- Financial data, including data reported pursuant to the Bank Secrecy Act (e.g., certain transactions over $10,000) and other financial data obtained via official investigations, legal processes, or legal settlements;
- Licensing information related to applications by individuals or businesses to hold or retain a customs broker's license, operate a customs-bonded warehouse, or be a bonded carrier or bonded cartman;
- Trade analysis data, including trade identifier numbers (e.g., for manufacturers importers, exporters, and customs brokers) and bill of lading data (e.g., consignee names and addresses, shipper names and addresses, container numbers, carriers); other financial data related to trade required for the detection and analysis of financial irregularities and crimes;
- Location-related data, including address; geotags from metadata associated with other record categories collected; and geolocation information derived from authorized law enforcement activities, ICE-owned devices, witness statements, or commercially available data;
- Law enforcement records, including criminal history, subject records, and investigative records; visa security information; and trade-based and financial sanction screening lists;
- Open source information, including news articles or other data available to the public on the internet or in public records, including content from the dark net and publicly available information from social media;
- Commercially available data, including public and proprietary records available for a subscription;
- Cargo and border crossing data, including inbound/outbound shipment records and border crossing information;
- Information or evidence seized or otherwise lawfully obtained during the course of an ICE investigation, including business records, third-agency records, public records (e.g., courts), transcripts of interviews/depositions, or records and materials seized or obtained via subpoena or other lawful process; and
- Tips concerning illegal or suspicious activity from the public and other law enforcement agencies.

Individuals Impacted:

RAVEn CORE can be utilized on any dataset ingested into the RAVEn platform and can

be used for any HSI statutory mission. Therefore, any individual whose record is in RAVEn may be impacted by the tool. These individuals include:

- Individuals identified in law enforcement, intelligence, crime, and incident reports (including financial reports under the Bank Secrecy Act and law enforcement bulletins) produced by DHS and other government agencies;
- Individuals identified in U.S. passport, visa, border, immigration, and naturalization benefit data, including arrival and departure data;
- Individuals identified in DHS law enforcement, licensing, and immigration records;
- Individuals who participate in the import or export of goods to or from the United States or to or from nations with which the United States has entered an agreement to share trade information;
- Individuals (e.g., subjects, witnesses, associates, assigned government personnel) associated with customs enforcement, immigration enforcement, administrative actions, detainer requests, or law enforcement investigations/activities conducted by DHS; and
- ICE personnel or personnel from partner law enforcement agencies who are mentioned in reports that concern law enforcement operations, injuries to law enforcement personnel, or other significant incidents reported within ICE.

Additional Privacy Risks and Mitigations:

Privacy Risk: There is a risk that users may be impaired in their ability to assess data quality because the data is in aggregated form.

Mitigation: This risk is mitigated. RAVEn CORE identifies the source of each data element in the search results and analytical results. As trained investigators and analysts, system users are already familiar with the quality of the various datasets they are accessing through CORE and can use this information to analyze and weigh the results generated by the system. HSI Innovation Lab endeavors to ensure data from ingested sources is routinely refreshed in RAVEn as close to real time as possible.[75] This way data that is corrected in source systems is updated in RAVEn. RAVEn's data tagging functionality also enables users to identify data sources that might be contributing incorrect records or otherwise poor quality data, so a user may alert HSI Innovation Lab to follow up with the data owner for correction or other appropriate remediation. All data available within RAVEn and displayed in CORE is linked back to the original data source, enabling users to weigh the reliability or quality of a data element and/or independently verify the information at the original source.

[75] The schedules of source system data refreshes are discussed in Appendix A of this PIA.

Privacy Risk: There is a risk that when disparate datasets are aggregated, the use of the aggregated data will be inconsistent with the purpose for which the data was originally collected.

Mitigation: This risk is mitigated. ICE mitigates this risk by limiting use of CORE to purposes that are related to the HSI mission, namely, to identify, disrupt, and dismantle transnational criminal organizations. This is in line with the overall purpose of the RAVEn platform. As such, HSI Innovation Lab ensures that all data can be used for that purpose prior to its initial ingestion into the platform. HSI Innovation Lab personnel do so by entering into information sharing agreements or similar contractual arrangements with the data holder. In that process, HSI Innovation Lab confirms with the data holder that RAVEn's use of the data is compatible with the original purpose of its collection. Similarly, ad hoc data uploads would possibly be aggregated, as well, but the original source of the data would be noted as legally obtained through a law enforcement mechanism, such as a subpoena. CORE has logging and auditing capabilities that track a user's access to data and its analytical output. Thus, supervisors and HSI Innovation Lab personnel have oversight of the CORE tool and can ensure it is only used for law enforcement purposes. HSI Innovation Lab will also limit CORE users to HSI special agents, criminal analysts, and full-time HSI Task Force Officers. As such, the job duties of the personnel who access CORE will be aligned with the purpose of the tool.

Privacy Risk: There is a risk that because CORE aggregates and searches data from multiple systems, it is possible that its users may be able access records using CORE that they otherwise could not view in the source system and are inappropriate for them to access.

Mitigation: This risk is mitigated. For data sets routinely ingested into RAVEn, HSI has established technical rules to ensure that the user privileges of the source system carry forward and apply to that user in RAVEn, no matter what tool the user accesses. Permissions to ad hoc data uploads must be manually assigned on each upload. As a result, a user's access privileges to the data stored in RAVEn are identical to their access privileges to that same data in the source system. This prevents CORE from being used, intentionally or unintentionally, to undermine or defeat the role-based access controls established by the source system.

Name:

HORUS Application Programming Interface (API)/ JANUS Facial Recognition Algorithm

Purpose and Use:

In the course of investigations, Homeland Security Investigations (HSI) routinely encounters digital images of potential victims or individuals suspected of crimes but cannot connect those images to identifiable information through existing investigative means and methods. HSI, therefore, may submit those images to government agencies and commercial vendors, hereinafter called Facial Recognition Services (FRS), to compare against the FRSs' digital image galleries via their own facial recognition processes. The agencies and vendors query their databases for potential matches and return lists of potential candidate matches that HSI can use to produce investigative leads.[76]

HSI, through the RAVEn system, is developing a connection with approved third party FRSs for HSI agents to submit probe photos.[77] This will allow HSI to format probe photo submissions to the American National Standards Institute (ANSI)/NIST Type 10 record format for data exchange using a biometric image messenger on RAVEn,[78] as well as to log and track all submissions by HSI and all returns by FRSs to ensure adequate security of the data and oversight of the use of FRSs. Other ICE offices, such as Enforcement and Removal Operations (ERO) do not use this tool.

To further this effort RAVEn has incorporated the HORUS API to assist agents in detecting faces in images, isolating faces for formatting, and clustering faces found in multiple images so agents or analysts can choose the best image available for submission to an FRS. HORUS is a Graphical User Interface (GUI) that will allow HSI users to submit images and visualize the results of the JANUS Facial Recognition Algorithms in one tool. The Intelligence Advanced Research Projects Activity (IARPA) JANUS program face recognition algorithm is the same tool used by the Department of Defense in its Automated Biometric Identity System (ABIS).[79] While many early face recognition algorithms were developed and trained on "constrained" images with similar poses of the subject and environmental conditions (lighting, camera distance), JANUS was developed specifically to match images that were unconstrained and variable. JANUS moves

[76] For more information on HSI's facial recognition processes see DHS/ICE/PIA-054 ICE Use of Facial Recognition Services, *available at* www.dhs.gov/privacy.
[77] Probe photos are facial images that are lawfully obtained pursuant to an authorized criminal investigation and submitted for facial recognition matching.
[78] ANSI/NIST ITL 1-2011, Update 2015, Data Format for the Interchange of Fingerprint, Facial, and Other Biometric Information, https://www.nist.gov/programs-projects/ansinist-itl-standard.
[79] *See* U.S. Department of Defense PIA for Defense Automated Biometrics Identification System (DOD ABIS) *and* Defense Biometric Identification Records System A0025-2 Provost Marshal General Defense Forensics and Biometrics Agency DoD 80 Fed. Reg. 8292 (February 17, 2015) *available at* https://www.federalregister.gov/documents/2015/02/17/2015-03123/privacy-act-of-1974-system-of-records.

beyond two-dimensional image matching into "model-based matching" that fuses all views from whatever video and stills are available to create a composite face template of an individual. Instead of relying on a "single best frame approach," JANUS addresses common challenges that beset face recognition algorithms (e.g., pose, illumination, and expression) by exploiting all available imagery uploaded to the gallery.[80] JANUS has also been vetted for accuracy and bias by the National Institute of Standards and Technology (NIST).[81]

HSI personnel will use HORUS to detect faces in media they have uploaded to RAVEn. The tool will detect faces within an image or video and will isolate the images of the face for review. The tool will also cluster similar faces within an upload (e.g., clustering multiple frames of a video or multiple photos from a digital photo album). Users will review the cluster of isolated face images and confirm all isolated images are the suspect or victim the user wishes to identify. During this review process, HSI will select isolated images from the clusters that are best suited for facial recognition processes. The HSI user will ensure he or she isolates a facial image that has the highest image quality possible, contains the least concealment of the individual's face, and is most similar to a constrained image, such as a mugshot or passport photograph. This is because HSI endeavors to isolate images as similar as possible to the galleries of images held by an FRS, which are usually constrained. HORUS will then pass the image to a biometric image messenger which formats the selected face image into the proper standard for submission to an FRS that is connected to RAVEn. Those connections are noted in Appendix A of this PIA.

HSI programs may work with the HSI Innovation Lab to build a repository of images processed by HORUS, called a gallery. These galleries are program specific and are only accessible to users designated by the program who uploaded the data. This allows a program, prior to submitting an image to a third-party FRS, to use HORUS to run a query against that program's specific gallery of images. This would allow the HSI program to link new probe photos to cases already under investigation by that HSI program. For example, linking similar photographs that were submitted in different benefit applications may help generate a lead on a suspect of identity fraud. Likewise, linking two videos of human rights abuses that were generated years apart by a common perpetrator may bolster a human rights investigation. HORUS is not capable of querying any images outside the program's gallery, to include other data in RAVEn and other programs' existing HORUS galleries.

An HSI user will compare information in each case linked by HORUS to other information available to HSI from various sources to vet the potential match outside the HORUS application. Additional evidence leading to validation or elimination of a candidate as a possible match could

[80] For more information *see* https://www.iarpa.gov/index.php/research-programs/janus.
[81] The JANUS algorithm was tested by NIST in their Face Recognition Vendor Test. For more information on the testing process, *see* NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, "Ongoing Face Recognition Vendor Test (FRVT) Part 2: Identification" (November 2018) *available at* https://nvlpubs.nist.gov/nistpubs/ir/2018/NIST.IR.8238.pdf.

include biographic information, current and previous addresses, telephone numbers, vehicles, criminal history, immigration history, and information derived from publicly available social media. HSI agents will not attempt to act as biometric face examiners and will instead compare HORUS returns through non-biometric investigative processes.

The HSI user will then run the probe photo against an FRS connected to RAVEn to determine if the new image will return a list of candidates for lead generation purposes. All images uploaded to a programmatic gallery must be submitted to an FRS, regardless of any case linkages or information found from a HORUS link. The HSI user will then vet any candidate returns from the FRS in accordance with HSI policy described in ICE's FRS PIA.[82] If an FRS-generated lead ultimately results in a positive identification of an individual, all relevant data pertaining to the individual's case will be retained in the relevant case file or ICM.[83]

HORUS is not an authoritative biometric database and cannot be used alone to determine the identity of an individual. There is no PII directly viewable with the isolated images searched by HORUS. RAVEn does maintain original, uncropped, source data which can be viewed by a user if they need to see the context of the file. That original image or video may contain incidental PII (such as a scanned fraudulent identity document with name and address, or a video with an individual shouting a name). HORUS has a splash screen disclaimer that must be accepted by users prior to accessing HORUS. The disclaimer instructs the user that any data in addition to the image is not verified, should be considered inaccurate, and should not be used to establish an individual's identity. The HORUS workflow requires all images to be ultimately run against an external FRS to generate a lead to identify an individual. HORUS is only used for processing images prior to submission and potentially linking cases within an HSI program's gallery.

Image uploads to the HORUS environment will be cataloged and organized into a user repository, called a gallery, by the HSI program. Any HSI programs that establish an ongoing connection with HORUS will submit a Privacy Threshold Analysis (PTA) to ICE Privacy for review of potential risks to individual privacy. HSI users are required by policy to electronically share data uploads with their supervisors for review. The HSI supervisor will review the data while it remains in RAVEn's upload queue. Data will only be uploaded from the queue if it is approved by the supervisor.

HSI users are also required to enter information describing the data being uploaded, such as source name/category, date retrieved, and corresponding case number. RAVEn keeps an audit log of all manual uploads by recording username and date/time of upload. When a user selects an image to be sent to an FRS through RAVEn, the name of the FRS, the date/time of submission, and the existence of any candidate matches are also noted within RAVEn. Images within HORUS

[82] *Supra* note 1.
[83] *See* DHS/ICE/PIA-045 Investigative Case Management (ICM), *available at* https://www.dhs.gov/privacy.

are considered temporary, and HORUS will be routinely purged of images not associated with an ongoing programmatic gallery. All HORUS ingested media and any FRS returns collected by an HSI user will be retained in HSI case files or HSI electronic case management systems, such as ICM. Images collected by HSI and determined to have no relation to a case are deleted prior to being ingested into HORUS.

Information Collected, Retained, and Disseminated:

HORUS will process and analyze photographs, videos, or other visual media determined to be relevant to an HSI investigation. It will create and match isolated digital face images derived from that media. HSI will collect this information within HSI policy detailed in the PIA for ICE's use of FRS.[84] Media within HORUS is either associated with an ongoing investigation or with an established HSI program whose use of HORUS has been reviewed by ICE Privacy to ensure compliance with ICE and DHS policies and applicable laws and regulations. According to HSI program mission needs, metadata and researcher notes may be associated with the image such as category/type of image, source of collection, date/time of collection, case number, name of HSI user, or other data needed for cataloguing or for evidentiary purposes. The images in the gallery will be assigned an identifier associated with the case or collection. There will not be any additional PII associated with the media or face image within HORUS or the biometric image messenger.

Individuals Impacted:

HORUS may analyze the photographs or videos of individuals who are encountered or investigated by HSI agents or analysts, including potential victims of a crime within HSI's statutory mission. Individuals captured within the photograph that may not be a target of an investigation will not be isolated, matched, or retained in ongoing program galleries.

Additional Privacy Risks and Mitigations:

The privacy risks and mitigations below relate directly to HSI using the HORUS API for the purposes described above. For a further discussion of the privacy risks, and corresponding mitigations, related to HSI's use of FRS generally, see ICE's FRS PIA.[85]

Privacy Risk: There is a risk that individuals will not have sufficient notice that HSI used their image as a probe photo in HORUS or that their information was obtained by HSI through an FRS.

Mitigation: This risk is partially mitigated. Suspects in probe photos or identified via FRS data may not be advised they are being investigated. Notice to these individuals could inform them that they are the target of an actual or potential criminal investigation or reveal investigative interest on the part of DHS or another agency. Access to the records might also permit the

[84] *Id.*
[85] *Id.*

individual who is the subject of a record to impede the investigation, to tamper with witnesses or evidence, harm victims, or to avoid detection or apprehension.

All individuals present in the United States, however, have constitutional protections in criminal proceedings entitling them to discovery production.[86] The discovery obligations of federal criminal prosecutors are generally established by the Federal Rules of Criminal Procedure 16 and 26.2, 18 U.S.C. § 3500 (the Jencks Act), *Brady v. Maryland*,[87] and *Giglio v. United States.*[88] In immigration proceedings, each party is responsible for producing evidence upon which it seeks to rely in the litigation. Therefore, if ICE seeks to use information derived from an FRS to sustain any charge or otherwise as evidence, it will produce that information.

This PIA and the ICE FRS PIA do provide notice to that public that ICE uses these technologies.

Privacy Risk: There is a privacy risk that information will be retained within HORUS for longer than is needed to accomplish the purpose for which the information was originally collected.

Mitigation: The risk is mitigated. By default, images within HORUS are considered temporary records and are deleted after they have exhausted their business purpose. Any probe photos, isolated imagery, or candidate returns used by HSI are subsequently saved in the relevant hard case file for investigation or ICM.

The 20-year retention period for ICM and other case file records is consistent with the retention schedules for other investigative records within DHS. By ensuring that information pertaining to individuals who are encountered repeatedly over a span of time can be linked, this retention period supports HSI's effective enforcement of U.S. civil immigration authorities, customs authorities, and federal criminal authorities. Closed cases can contain information that may be relevant to a new or existing case and need to be readily searchable and accessible for at least a period of time. The addition of probe photos and candidate returns to a case file will not affect the existing retention processes in ICE systems. Probe photos and candidate returns will be destroyed when the case file is destroyed.

In instances where an HSI program has created a gallery of unknown images within HORUS, that program will submit a PTA to ICE Privacy before ingestion into RAVEn. ICE Privacy will work with ICE Records and Data Management Unit to ensure each gallery has an appropriate retention schedule that aligns with requirements established by the National Archives and Records Administration (NARA). The standard, currently, is that HORUS will retain unidentified images within the program's gallery for 75 years after original collection, in

[86] Discovery is the general process of a criminal defendant obtaining information possessed by a prosecutor regarding the government's criminal case against the defendant.
[87] 373 U.S. 83 (1963).
[88] 405 U.S. 150 (1972).

accordance with a DHS-wide biometrics retention schedule, DAA-0563-2013-0001-0006.[89] Every program gallery, however, will be individually assessed during the PTA process.

Privacy Risk: There is a risk HSI will submit low quality images or probe photos that would otherwise increase the likelihood of false case linkages from HORUS.

Mitigation: The risk is partially mitigated. ICE Privacy and HSI have developed training, that all HSI agents and analysts are required to take prior to use of HORUS or any FRS, that helps maximize image quality during their collections. The training discusses common failures within facial recognition technologies and how to choose images most likely to improve accuracy within an algorithm. Moreover, the primary purpose of HORUS is to assist HSI agents and analysts to find and isolate the highest quality face images from standard media collections. HORUS' functionality allows HSI users to quickly compare and identify isolated images of suspects and victims from unprocessed media. This step will occur prior to an HSI user running a check against a programmatic gallery. HSI users are also trained to vet all returns from an FRS, as per the policy discussed in ICE's FRS PIA.[90] This vetting will extend to HORUS linkages. An HSI user will compare existing evidence between the two cases to ensure that any image linkages are corroborated by other information or evidence (e.g., same methods of criminality, same geographical area of operation.)

Privacy Risk: There is a risk that HSI will use biographic information viewable in source data to identify an individual.

Mitigation: This risk is partially mitigated. Original images or videos uploaded to HORUS may contain incidental PII. That PII is cropped out of the viewable image when HORUS isolates a face image for searching. The PII is only viewable in RAVEn if a user must access the original file for context. HORUS has a splash screen disclaimer that must be accepted by users prior to accessing HORUS. The disclaimer instructs the user that any data in addition to the image is not verified, should be considered inaccurate, and should not be used to establish an individual's identity. Users are admonished to search the original case file of any image linked by HORUS and are directed to run all images through an FRS prior to generating a lead.

Privacy Risk: There is a risk that HSI will use biographic or derogatory information received from a linked case that is inaccurate.

Mitigation: This risk is partially mitigated. Data contained within ICE case files are gathered for law enforcement and/or national security purposes. Law enforcement and national

[89] *See* NATIONAL ARCHIVES AND RECORDS ADMINISTRATION, REQUEST FOR RECORDS DISPOSITION AUTHORITY, RECORDS SCHEDULE NUMBER DAA-0563-2013-0001, U.S. DEPARTMENT OF HOMELAND SECURITY, BIOMETRIC WITH LIMITED BIOGRAPHICAL DATA (2013), *available at* https://www.archives.gov/files/records-mgmt/rcs/schedules/departments/department-of-homeland-security/rg-0563/daa-0563-2013-0001_sf115.pdf.
[90] *Supra* note 1.

security personnel are trained to review all information they collect for accuracy, as errors may detrimentally affect prosecutions and investigations. Similarly, if ICE personnel are alerted that data within their case files or systems are inaccurate, they will work to correct the data as soon as practicable. This increases the likelihood that the information within a case file or ICM has been previously vetted for accuracy. Additionally, the HSI user will conduct his or her own research and investigation to determine if the information returned by a linked case is accurate before generating a lead.

Privacy Risk: There is a risk that media ingested into HORUS will be unverifiable.

Mitigation: The risk is partially mitigated. HSI Innovation Lab will not verify the provenance of images prior to ingestion into RAVEn. HSI generally does not have the technical capacity to analyze all media collected to determine whether it has been digitally manipulated (i.e., "deep fakes"). However, all ingestions must be approved by HSI supervisors prior to upload into HORUS. HSI will use traditional investigative methods to vet the veracity of any media prior to collection. The media collected by HSI will not be used to determine benefit eligibility, nor will it be used for establishing probable cause in an investigation. Media collected by HSI will be considered investigative leads until additional evidence either validates or refutes what was collected.

Privacy Risk: There is a risk HSI users may inappropriately access images or use information within HORUS for purposes outside the scope of this Appendix.

Mitigation: The risk is mitigated. User roles and access controls are incorporated into RAVEn at the record level, so that only users with a need to know and pre-approval can access the tool or programmatic galleries. Due to technical constraints in the HORUS algorithm, an image or video in one gallery cannot be shared with another gallery internally. Any image would have to be submitted individually to each program's gallery. At each submission, the HSI supervisor of that program would have to verify that the data upload was for a legitimate mission need. Any HSI personnel involved in the collection, processing, vetting, or sharing of face images are required to undergo training on the appropriate uses of data collected by HSI and received from partners. Personnel will receive access only after satisfactory completion of the training and with approval from an HSI supervisor. Anyone who is found to have used HORUS in an unauthorized manner will be disciplined in accordance with ICE policy and/or federal law.

Name: RAVEN GO

Purpose and Use:

RAVEn GO is a mobile application which serves as a portable version of the RAVEn system. RAVEn GO is available for download by authorized users (i.e., Homeland Security Investigations (HSI) Special Agents, Criminal Analysts, and a limited number of other HSI employees) on ICE phones through a comprehensive mobility solution that allows ICE users to securely access ICE information via their ICE mobile device.

RAVEn GO provides search functions, data feeds, and real-time secured message and data exchange. At the most basic level, RAVEn GO operates as a remote access point to RAVEn. HSI agents and analysts can:

- Perform simple, quick searches or more advanced, filter-based searches on RAVEn datasets for which they have been granted access;
- Set up automatic alerts that occur when new data matching criteria enters the RAVEn system;
- Send secure text messages and photos to other members in the field for real-time analysis (he application stamps the transmissions with date, time, and geocode/geolocation);
- Record and save relevant details, add photos, and auto-scan (capture textual data from images) driver's licenses, passports, and visas with reports and Encounter Cards (EC); and
- Submit reports and ECs to the RAVEn platform for further analysis.

RAVEn GO allows HSI agents to directly input data taken from the field in the form of an EC. In the context of HSI criminal investigations, HSI agents use Field Interview Cards (FIC) (ICE Form 73-011) to document field interviews. The FIC is a form on which agents may record information about investigative subjects, witnesses, and other third parties. The EC is RAVEn's version of the FIC that also allows agents to note interactions and observations that are not tied to interviews (e.g., encountered vehicles, phones, addresses).

RAVEn GO provides a fillable form (the EC) on the ICE phone and securely transmits the information collected during field interviews into the RAVEn platform. This electronic collection provides a more secure means to collect and transmit Sensitive PII and allows for the direct electronic ingest of data into the RAVEn platform. This minimizes the opportunity for data entry errors and inconsistencies between the data collected on the original instrument (the EC) and the data that ultimately resides in RAVEn.

The RAVEn GO EC function expedites the availability of this information to other HSI agents with appropriate access permissions who may encounter the same individuals and therefore have a need to know the details of a previous HSI interview. Using a mobile application does not change

the collection of information by HSI agents; it simply changes the way the information is transmitted to RAVEn for analysis and use by other HSI personnel.

Access to RAVEN RAVEn GO is restricted to authorized HSI special agents, criminal analysts, and a limited number of other HSI employees/full-time task force officers. All RAVEn GO users agree to a Rules of Behavior (ROB) statement at every login. Audit logs are maintained for all actions taken within the RAVEn GO application. Any user found to have violated the ROB may have his or her access revoked and is subject to disciplinary action.

Information Collected, Retained, and Disseminated:

RAVEn GO allows users to access any datasets available to them through RAVEn. This includes all datasets discussed in Section 2.1 of this PIA, as well as datasets discussed in Appendix A. RAVEn GO does not connect to any systems other than the RAVEn platform and does not share data beyond what is available within RAVEn.

The following data may be entered on the EC, when available and pertinent:

- Agent/Officer name and identifier;
- Criteria for interview and organizational affiliation indications;
- Date and time;
- Field Interview Location;
- Citizenship;
- Family information, including citizenship;
- Contact information;
- Vehicle information;
- Miscellaneous information and additional notes;
- Photo identification number;
- Photographs;
- Descriptions of scars, marks, and tattoos;
- Social Security number;
- A-Number; and
- TECS Case Number.

Individuals Impacted:

RAVEn GO can perform simple searches on any dataset ingested into the RAVEn platform and can be used for any HSI statutory mission. Therefore, any individual whose record is in RAVEn may be impacted by the tool. Most commonly, RAVEn GO will be used when HSI agents or analysts encounter individuals during field work for their investigations.

Additional Privacy Risks and Mitigations:

Privacy Risk: There is a risk data will inadvertently be retained on the ICE mobile device.

Mitigation: This risk is mitigated. The deployment of RAVEn GO is managed by HSI's instance of a Mobile Device Manager (MDM). The MDM provides HSI with the ability to control the operations undertaken by RAVEn GO at a granular level, including RAVEn GO's ability to write data to the ICE mobile device. This prevents RAVEn GO data from being captured or stored on the ICE mobile device unless it is an explicit functionality created by the HSI Innovation Lab. The MDM additionally provides a local encrypted store so any data that is stored locally, intentionally or otherwise, will not exist in plain text. The MDM provides HSI the ability to manage how RAVEn GO interacts with ICE mobile devices and allows HSI to mitigate the risk of RAVEn GO data being inadvertently retained on the ICE mobile device.

Privacy Risk: There is a risk that the mobile application will allow unauthorized users more opportunities to access the RAVEn platform via ICE phones.

Mitigation: This risk is mitigated. All ICE phones are encrypted and require either face, fingerprint, or passcode authentication to access the contents of the phone. ICE utilizes certificate technology to grant authorized users secure access to restricted information. All methods and endpoints for interacting with RAVEn GO are protected by the same authentication and authorization mechanisms as RAVEn. As a result, a user's access privileges via RAVEn GO are the same as the platform. Access privileges to data stored in RAVEn are identical to a user's access privileges to that same data in a RAVEn source system. Since RAVEn GO is protected by the same authentication and authorization mechanisms as RAVEn, ICE mobile devices, or any other device, do not present additional risk of unauthorized access.

Privacy Risk: There is a risk RAVEn GO may collect information, such as metadata or geolocation information, while not in active use.

Mitigation: This risk is mitigated. RAVEn GO is not able to run any background processes which would collect information when RAVEn GO is not running on the ICE mobile device. Any risk to this type of information collection by RAVEn GO is mitigated by HSI's deployment of MDM. RAVEn GO's MDM allows HSI to control how RAVEn GO can run, including what information it can collect, even after the deployment of RAVEn GO onto a device. The RAVEn

GO application is partitioned off from any other mobile functionalities or applications on the phone through the MDM.

Privacy Risk: There is a risk information entered into the EC will not be vetted for accuracy prior to its use in RAVEn searches and reports.

Mitigation: This risk is partially mitigated. All information entered into the EC is based on information provided by the subject of the interview, documents provided during the interview, and/or observations made by trained HSI agents. The ability to capture and auto-scan documents minimizes the risk of data entry errors that are present in traditional forms of data capture associated with the interview process. The entering agent's name and contact information is linked to each EC so any discrepancies that may be later identified can be addressed with the agent and corrected.

Privacy Risk: There is a risk that RAVEn GO users will be able to access information for which they do not have permission.

Mitigation: This risk is mitigated. For data sets routinely ingested into RAVEn, the HSI Innovation Lab has established technical rules to ensure that the user privileges of the source system carry forward and apply to that user in RAVEn, no matter what tool the user accesses. Permissions to ad hoc data uploads must be manually assigned on each upload. As a result, a user's access privileges to the data stored in RAVEn are identical to their access privileges to that same data in the source system. This prevents RAVEn GO from being used, intentionally or unintentionally, to undermine or defeat the role-based access controls established by the source system.

Name:

Mobile Device Analytics (MDA) and Email Analytics (EA)

Purpose and Use:

MDA is an analytical tool within the RAVEn environment that can search and analyze data extracted from mobile devices. Email Analytics is a similar analytical tool that is used on data seized from email accounts and servers. HSI has identified cellular telephones and email communications to be an important means to support criminal activity. HSI determined there is a high likelihood that evidence exists on cellular telephones, other mobile devices (hereinafter collectively referred to as "mobile device(s),") and email accounts identified in connection with criminal investigations.

During operations, HSI will either gain consent from an individual or obtain subpoenas, search warrants, or court orders to extract data from mobile devices or email accounts linked to targets of criminal investigations. For MDA, through separate and additional processes, HSI will make mobile device data accessible to agents and analysts for analysis.[91] For EA, Each email extraction will be manually loaded by HSI personnel after appropriate authority has been granted (e.g., search warrant). When investigative data is manually imported into RAVEn. An HSI supervisor will review the data to ensure the data's upload is appropriate and adheres to ICE policy. MDA and EA are simply analytical tools and cannot open, access, or "hack" devices or email accounts. For security purposes, the mobile device will not be connected to the internet.

The MDA and EA applications will have the capabilities to sort, filter, correlate, and discover relationships and pertinent evidence lawfully obtained from mobile devices and/or email accounts. The interface will enable members of investigative teams to organize and intuitively visualize lawfully collected information and facilitate collaboration and communication among investigators. The implemented advanced analytics strategies will rapidly extract patterns and illuminate connections within the collected information.

As an application within RAVEn, MDA and EA will possess the same technical, administrative, and physical safeguards as other tools within the RAVEn environment. Data retention and user access is controlled at the record level, and as an ad hoc upload, all data ingested by RAVEn for MDA and EA is assigned a relevant case number that is associated with an ongoing investigation. HSI supervisors approve the data loaded into RAVEn for MDA and EA. RAVEn has auditing and logging functionalities that track the use of the MDA and EA tools and associated data.

The output from MDA and EA will be an organized report that captures the most pertinent

[91] For more information on ICE processes of access and extraction of mobile devices *see* DHS/ICE/PIA-042 Forensic Analysis of Electronic Media, *available at* www.dhs.gov/privacy.

evidence discovered via the application. The format of the reports will enable use in other HSI law enforcement systems, principally the Investigative Case Management system (ICM).[92] Investigators and analysts can also use MDA or EA to isolate and consolidate information relevant to an investigation for later extraction and use within the RAVEn environment. For example, links of email addresses identified by MDA/EA could be migrated to the RAVEn CORE[93] tool to create a visualization or to be checked against ICE holdings in ICM.

Information Collected, Retained, and Disseminated:

The MDA and EA applications will store data found on mobile devices or within email accounts obtained via court orders, search warrants, subpoenas, or consensual searches from individuals, which includes:

- Biographic information (e.g., names);
- Contact information (e.g., phone numbers, email addresses, physical addresses);
- Social media information (e.g., accounts, usernames, public posts);
- Text/chat messages;
- Financial accounts and transactions;
- Browser history;
- Emails, including email addresses and email content;
- Information on calendars and other note applications on a device or linked to an email account;
- IP addresses;
- Geolocation information (such as addresses and/or geotags associated with data in emails or on the mobile device);
- Images; and
- Telecommunications information (e.g., call logs, cell site information).

Individuals Impacted:

The MDA and EA tools analyze the PII of individuals whose information may be contained within the mobile devices or email accounts seized by HSI personnel.

[92] *See* DHS/ICE/PIA-045 Investigative Case Management System (ICM), *available at* www.dhs.gov/privacy.
[93] For more information on RAVEn CORE *See* Appendix B of this PIA.

Additional Privacy Risks and Mitigations:

Privacy Risk: There is a risk that MDA/EA collects information that is irrelevant to an investigation. This includes a risk that the PII of individuals unassociated with an investigation will be retained by ICE.

Mitigation: This risk is partially mitigated. There is a risk of overcollection due to the fact that the entire contents of a mobile device will be uploaded into MDA and the entire contents of an email account will be uploaded to EA. The sheer volume of data within a mobile device or email account, however, makes it impractical to view the data outside of analytical processes. MDA and EA only conduct focused searches using predetermined queries. Only data that is directly tied to the search is then displayed. After the query, the agent or analyst may determine which information is relevant and may remove information deemed out of scope. Any data collected outside the scope of the MDA or EA application will not be used or introduced as evidence in an investigation.

Privacy Risk: There is a risk that data uploaded to MDA or EA may have been acquired by means other than via legal processes, such as a subpoena or search warrant.

Mitigation: This risk is mitigated. MDA and EA are purely analytical tools and exist as only a part of a forensic or investigative analysis process that has rigorous protections to ensure the integrity of data and chain of custody for evidence. The process of opening or accessing the contents of a mobile device occurs earlier in the chain of custody of evidence. At that time in the process, HSI verifies that the mobile device was obtained through consent or legal processes. Accessing the contents of the phone is accomplished outside of HSI Innovation Lab by digital forensic examiners through a separate extraction device. HSI agents or analysts seeking data from a mobile device must gain supervisor approval prior to forensic examiners granting access. Supervisors will ensure that the chain of custody was maintained and HSI's possession of the mobile device is lawful. Moreover, MDA logs all devices that are connected to the application, and tags the records extracted from the mobile device with a HashID. That ID must be associated with a relevant case or investigation, thereby precluding any improper uploads to MDA.

Each email extraction will be manually loaded by HSI personnel after appropriate authority has been granted (e.g., search warrant). When investigative data is manually imported into RAVEn, HSI agents are required by policy to electronically share this data with their supervisors for review. The HSI supervisor will review the data to ensure that it was acquired via appropriate processes while it remains in RAVEn's upload queue. Data will only be uploaded from the queue if it is approved. These uploads are logged within the RAVEn system. The logs of devices connected to MDA and data uploaded to EA are routinely audited by HSI Innovation Lab personnel. If HSI finds that an agent or analyst has uploaded data inappropriately to MDA or EA, that information will be deleted.

Privacy Risk: There is a risk that the MDA or EA tool will be used for purposes beyond what is outlined in this PIA. This risk is heightened due to the large amounts of data present in mobile devices and email accounts.

Mitigation: The risk is partially mitigated. HSI's use of MDA and EA are governed by requirements specified in the ICE Computer Forensics Handbook and by federal laws, regulations, and policies that govern the acquisition, handling, and preservation of electronic evidence, including personally identifiable information. MDA and EA have been developed to log all user access and use of data. MDA and EA create a HashID of ingests at the record level that notes what case the data is associated with. HSI Innovation Lab is developing security measures to ensure that only individuals associated with a relevant case and with a need to know the information can view specific records within MDA or EA. Until that time, MDA and EA user access is restricted to those users with a need to use the tool.

Further, all user search terms and queries are automatically logged by RAVEn within the MDA and EA applications. HSI supervisors will conduct routine audits on the system to ensure proper use. Any user that is found to be using MDA or EA inappropriately will have his or her access revoked and may face disciplinary action. Finally, if data derived from MDA or EA is used in a law enforcement action, such as an arrest or search, then that data must be presented to a court, where ICE must explain how the information was obtained. If a court determines that evidence was acquired inappropriately, it may deny HSI's request for a warrant or subpoena. Therefore, all data analyzed by MDA and EA will only be used for criminal investigations and appropriate law enforcement purposes.

Privacy Risk: There is a risk that data will be retained within the MDA or EA tool longer than is necessary for an investigation.

Mitigation: The risk is mitigated. Data that HSI extracts for use in MDA or EA will reside in the same data repository as all data within RAVEn. It will be similarly tagged at the record level with the relevant case number and retention schedule. MDA data extracts will be both automatically and manually audited with the rest of the RAVEn data environment to ensure data is deleted in accordance with ICE policies and retention schedules.

Name: Airline Reporting Corporation (ARC) Travel Intelligence Program (TIP) (ARC TIP)/ Homeland Security Investigations (HSI) Office of Intelligence's (Intel) use of Airlines Reporting Corporation (ARC) Travel Intelligence Program (TIP) program.

Purpose and Use:

ARC TIP is a commercial airline intelligence database that provides information needed for HSI Intel investigations supporting federal criminal investigations, such as child exploitation, human trafficking, and financial crime. HSI Intel analysts use passenger name, airline, or credit card number to search ARC TIP. HSI Intel analysts may incorporate information relevant to investigations found through ARC TIP queries manually into Reports of Analysis (ROA) with RAVEn integration maintained by the HSI Innovation Lab. ARC TIP maintains derived HSI Intel queries for 45 days and data is purged automatically.

HSI Intel has a broad and complex public safety mission which is furthered through the collection and sharing of timely and accurate intelligence on illicit trade, travel, and financial activity with a nexus to the United States. Through ARC TIP, data is received daily, containing the previous day's ticket sales, encompassing actual ticket information showing true intent to travel. The TIP application collects difficult and, in most instances, unable-to-access "ticket face" information to include the full flight itinerary, passenger name record, fare detail, and form of payment. The proprietary TIP database unlocks over one billion records containing thirty-nine months of past and future travel, previously unavailable due to no other system containing the vast collection of data.

Information collected through ARC TIP is available to ICE HSI Intel (Intelligence Analysis Division, Intelligence Collection Division, Intelligence Integration & Emergency Management Division, and Intelligence Enterprise Services Division). ICE users query the ARC TIP tool by passenger name and by credit card number. Further, the travel records database contains information for tickets issued by ARC Accredited Travel Agencies in the United States, both traditional agencies and online agencies (i.e. Expedia and Orbitz), both domestic and international. However, if the passenger buys a ticket directly from the airline, then the search done by ICE will not show up in an ARC report. ICE only receives the ticketing data that matches the criteria used for the search. For example, if the search criteria is Person A, the system will not search for any other name but Person A. If a search includes a common name, it would not be unusual for the portal to return thousands of lines of data in the report that is sent back to the portal. Thus, ICE minimizes its search parameters to ensure the most accurate results.

Information Collected, Retained, and Disseminated:

The ARC program enables HSI Intel to collect ticketing information through ARC TIP enabling authorized users to search ARC's air ticketing database for law enforcement and national security purposes only.

Authorized users can search using the following criteria:

- Passenger name;
- Credit card number;
- Airport code (anticipated departure and/or arrival city); and
- Estimated dates when travel was ticketed ARC currently offers three TIP services: TIP Online, TIP Monitoring and TIP Search.

Individuals Impacted:

Individuals impacted include DHS employees and contractors and members of the public, including U.S. Persons and Non-U.S. Persons.

Additional Privacy Risks and Mitigations:

Privacy Risk: There is a risk of unauthorized access to sensitive information, such as personal data of passengers and investigations details, by external adversaries or internal personnel without the proper clearance.

Mitigation: This risk is mitigated. Robust access control measures are implemented for ARC TIP with regular audits of access logs conducted. Access is granted on a least-privilege basis, with access only granted to the information necessary for the analyst's role.

Privacy Risk: There is a risk sensitive information could be inadvertently exposed through misconfigurations, weak data handling policies, or during the transfer of data between systems.

Mitigation: This risk is mitigated. Strict data handling and classification policies are developed and implemented that dictate how data is stored, transmitted, and destroyed. Encryption is utilized with secure protocols for data transmission security assessments and vulnerability scans are conducted regularly to identify and remediate potential weaknesses. Additionally, data loss prevention (DLP) tools are implemented to monitor and prevent unauthorized data transfer.