Privacy Impact Assessment

for the

Data Analytics Technology Center (DA-TC)

DHS Reference No. DHS/S&T/PIA-040

August 13, 2020

Homeland Security

Abstract

The U.S. Department of Homeland Security (DHS) Science and Technology Directorate (S&T) operates the Data Analytics Technology Center (DA-TC) to research the performance of, and identify opportunities and risks associated with emerging and next-generation advanced data analytics and computation capabilities. The DA-TC research activities involve data from government, commercial, and publicly available data, to refine technical problems, capability gaps, data sets, technology, operational assessments, and mission performance that meet DHS and the Homeland Security Enterprise (HSE) critical mission and operational needs. S&T is conducting this Privacy Impact Assessment (PIA) to identify and mitigate privacy risks associated with DA-TC collecting, maintaining, and sharing personally identifiable information (PII) during research, development, testing, and evaluation (RDT&E) of advanced analytic and computational capabilities. This PIA also assesses the privacy risks associated with the two distinct computing environments: a physical on-premise environment and off-premise cloud environments.

Overview

DHS S&T conducts RDT&E of advanced computational and analytics technologies to support critical DHS missions throughout the Components and fulfill responsibilities under 6 U.S.C. § 182, *Responsibilities and Authorities of the Under Secretary for Science and Technology*. These DHS operational Component critical missions and authorities relate to counterterrorism; border security; countering transnational criminal networks; cybersecurity; critical infrastructure protection; international trade security; national transportation security; maritime security; and national preparedness, response, and resilience.

The DA-TC works directly with DHS S&T programs, DHS components, the HSE,[1] academia, industry, and non-profit researchers to facilitate technical problem-solving for homeland security analytic missions, such as counterterrorism, border security, and national preparedness and resilience. DA-TC's mission is to conduct RDT&E activities to deliver effective and innovative insight, methods, and solutions to address DHS and HSE critical needs.

DA-TC has a state-of-the-art data analytics laboratory (referred to as the "Lab" hereafter) in a DHS-controlled information technology (IT) security environment that is designed to support mission-relevant evaluations of emerging technologies, rapid experimentation, and strategic RDT&E efforts. The Lab has two computational environments: (1) an on-premise non-networked physical computing capability, and (2) a cloud-based computing environment. This PIA provides an overview of the nature of DA-TC's activities, as described in the following subsections. The

[1] U.S. DEPARTMENT OF HOMELAND SECURITY, DHS INSTRUCTION NUMBER 034-06-001, *DEPARTMENT REPORTING REQUIREMENTS*, defines the Homeland Security Enterprise as "[t]he collective efforts and shared responsibilities of federal, state, local, tribal, territorial, non-governmental, private-sector, and international partners—as well as individuals, families, and communities—to maintain critical Homeland Security capabilities."

specific types of tools, applications, data, and analysis to be used for any given DA-TC project will depend on the mission use case(s) supporting the project. This PIA will be updated with appendices to enable the assessment of potential privacy risks and document mitigation strategies for individual projects, and the data sets in use.

The Lab serves as a resource where authorized users develop and test hypotheses and methods in a secure environment. DA-TC's customers and researchers include DHS Components, other federal entities, and state and local partners to include university, industry, and other non-profit partnerships. The Lab enables DA-TC researchers to understand mission relevance of analytics solutions using customer-provided, open source, research data sets, and proprietary and commercially available data. The Lab is used to accelerate the definition of next-generation analytics technical, mission, and operations requirements. Further, the Lab enables rapid technology assessments that inform next-generation enterprise data analytics acquisitions, architectures, and solutions for DHS. The Lab controls and monitors access to specific applications or data sets based on a user's need to know and sensitivity of the data. All Lab users sign rules of behavior to advise them of their roles and responsibilities when accessing and using DHS IT systems.

DA-TC RDT&E Activities

DA-TC is tasked with defining and scoping DHS critical mission problems and identifying technical solutions in support of the missions of the Department, DHS Components, S&T programs, and the HSE. DA-TC leadership determines the technical efficacy and value propositions regarding the initiation, execution, and termination of research activities to support these needs. DA-TC selects actionable research activities based on optimizing the return from limited research and development resources.

As part of each research activity, DA-TC works with the S&T Privacy Office and the S&T Office of General Counsel, as well as other compliance functions, to establish guidelines for the collection, analysis, and disposition of data. When DA-TC conducts RDT&E without an operational partner those activities do not result in any operational action; rather DA-TC uses technical results from experiments and pilots to inform the development of next-generation mission and operations technologies. DA-TC also invokes operational Components' authorities when conducting RDT&E. DA-TC may report information to appropriate federal, state, or local authorities, should exigent circumstances arise (such as significant threat to life and property).

DA-TC's RDT&E activities include:

- Data Audits: are used to examine government, commercial, and open source data sets to assess their content, completeness, reliability, and characteristics that can be used to advance the DHS mission. Data audits are essential to understanding of how data elements contribute to analytic processes, methodologies, and decision making. The results of DA-

TC's data audits help the rest of DHS to determine the value of data sets. DA-TC's data audits also help identify what data is needed to support DHS missions.

- Technology and Tool Assessments: are used to examine the application of emerging data sets, technologies, system components, algorithms, and tools. DA-TC's assessment of a technology or tool seeks to measure its performance of technical functions, mission and operations functions, human performance, operational costs, privacy implications, and security mechanisms. DA-TC assesses data, tools, and technologies for general or specific use cases. The assessments identify potential value or risk to DHS. These efforts are often executed as research experiments and result in the development of technical performance baselines that can be used to gauge the success of RDT&E activities. They may also result in analyses that can be useful to support DHS S&T, other DHS components, or other government acquisition formulations and decisions.
- Prototyping: is used to help DHS identify technical approaches that can be used to solve data analytics and related computational problems by exploring custom software or hardware in the Lab. These rare efforts are used to assemble and create a variety of technologies to achieve a proof-of-concept functional capability. DA-TC will then use that capability to answer technical and mission questions as well as gain initial indications of mission performance related to algorithms, systems, and architectures.
- Pilots: are used to apply and integrate advanced analytics and computational technologies in a training exercise or operational environment to improve a specific critical DHS mission or operation. DA-TC conducts pilots in conjunction with DHS Component authorities. In these efforts, technologies are directly exposed to the mission environment, analytic staff, and decision-making processes to more fully assess technical, human, and mission performance.

Data Sources and Use

To execute the above RDT&E activities, DA-TC uses data from different sources that are defined below. DA-TC receives, sends, and uses data only within the scope of established data handling guidance and in compliance with DHS policies.[2]

- Government Data: In some instances, government data owners associated with different missions and Components will make full or limited data sets available for specific RDT&E activities. DHS S&T, DHS Components, other federal agencies, foreign governments, and state, local, or tribal organizations have established authorities to

[2] DHS 4300 is a series of information security policies, which are the official documents that create and publish Departmental security standards in accordance with DHS Management Directive 140-01, *Information Technology System Security*. *See* U.S. DEPARTMENT OF HOMELAND SECURITY, DHS 4300A SENSITIVE SYSTEMS HANDBOOK, *available at* https://www.dhs.gov/publication/dhs-4300a-sensitive-systems-handbook.

collect and use various data sets related to homeland security missions. For critical missions of DHS, and in accordance with DHS authorities, DA-TC may collaborate with government data owners in data analytics RDT&E activities. In those cases, legal determinations, appropriate collaborative agreements, and privacy oversight will be used to establish guidelines for specific RDT&E activities. Appropriate data handling guides will be developed for each data source that is part of RDT&E activity.

- Commercial Data: DA-TC may work with commercial data providers in its RDT&E activities. This work may examine the use of commercial data, or if there is mutual benefit, may lead to the development of new commercial data services. DA-TC also uses relevant third-party data collected by non-profit organizations and made available for a fee. Use of these sources in RDT&E activities allows DA-TC to better understand the potential value of such data sets, such as saving analytic resources or increasing the accuracy of decision making. DA-TC may also use commercial data in its RDT&E activities to enrich data it receives from other sources.

- Research Data: DA-TC uses research data from government, industry, and academia. DA-TC will use appropriate data from these sources, where data has been developed and curated for the furtherance of research and development. For example, the Linguistic Data Consortium, established by the Defense Advanced Research Projects Agency, provides data that supports long-term research in language and linguistics technology.

- Open Source Data: DA-TC uses data sets developed through the collection of open source information, that includes publicly available social media. DA-TC leverages open source data to characterize effective use of analytics technology within DHS operations. The RDT&E may evaluate the use of new types of data found in open sources or the use of existing open source data to evaluate new analytic tools. For example, DA-TC may use open source data in its development of an automated situational awareness capability for the DHS Components to use in operations. A more detailed discussion of DA-TC's and DHS S&T's use of open source data, such as publicly available social media, can be found in a forthcoming PIA.[3] Project-specific legal and privacy reviews will be conducted before open source data is collected to provide DA-TC with guidance that ensures adherence with all legal, privacy, and DHS policies.

Once the proper legal and privacy reviews are conducted, the above data sets may potentially be comingled with DHS Component or other government data in furtherance of research and development. For example, DA-TC researchers may use government trade data and open source

[3] Once finalized, U.S. DEPARTMENT OF HOMELAND SECURITY, SCIENCE AND TECHNOLOGY DIRECTORATE, PRIVACY IMPACT ASSESSMENT FOR THE S&T PUBLICLY AVAILABLE INFORMATION IN RESEARCH, DEVELOPMENT, TESTING, AND EVALUATION will be *available at* https://www.dhs.gov/privacy-documents-st.

and commercial data sets, in coordination with human review to assess an analytics technology's ability to correctly disambiguate company names. Such an assessment would produce technical and mission metrics that indicate the value of the analytic technology. Results of DA-TC's research would then be used by DHS for developing mission strategy, understanding technical risk, justifying acquisitions, characterizing capability gaps, or driving technology-based performance improvements. No data sets containing PII that are collected for a specific purpose and evaluated as an RDT&E activity can be re-purposed for other or future initiatives. Findings of one project are not shared with other partners unless specific agreements are put in place to address data sharing requirements.

DA-TC may conduct testing and evaluation in an operational setting, using Component authorities, where the data collected or received includes PII. Prior to using that data, DA-TC coordinates with the DHS Component or partner to ensure that DA-TC's role is limited to research, development, testing, and evaluation. This is documented in the development of a Memorandum of Understanding (MOU). Each authorized data set that DA-TC uses is subject to the data owners' specified controls and handling guidance, privacy oversight, and security controls that are documented in a data handling guide. The guide includes instructions for sharing, destruction, and certification of removal of data from DA-TC or other DHS S&T systems at the end of the activity. Once the RDT&E effort is concluded, DA-TC may delete or return data sets to the data owner for future use. Per the DHS/S&T-001 Research, Development, Test and Evaluation Records System of Records Notice (SORN),[4] PII collected during the project is retained for the duration of the project. At the conclusion of the project, PII is destroyed in accordance with applicable federal record schedules. Researchers may retain aggregated research data (without PII) in accordance with applicable federal record schedules, as it may help inform future RDT&E efforts.

As a result of legal and privacy reviews, the types of data to be used in each activity are rationalized and minimized to the data sets required to enable appropriate research results. Data types that may be used in DA-TC research activities include the following information:

- Full Name;
- Home Street Address;
- Home Phone Number;
- Business Street Address;
- Business Phone Number;

[4] *See* DHS/S&T-001 Research, Development, Test, and Evaluation Records, 78 Fed. Reg. 3019 (January 15, 2013), *available at* https://www.dhs.gov/system-records-notices-sorns.

- Email Address;
- Employer Identification Number (EIN);
- Social Security number (SSN);
- Tax Payer Identification Number (TIN);
- IP addresses;
- Facial images;
- Video; and
- Audio.

Information collected from publicly available social media may include information such as:

- Social media handles;
- Account names;
- Email addresses;
- Pictures, video speech (live stream);
- IP addresses;
- Latitude and longitude of geo-tagged content;
- Phone numbers; and
- Posted content.

Information collected from commercial vendors may include information such as:

- First name;
- Middle name(s);
- Last name;
- Date of birth;
- Gender (Male, Female, Other);
- Street;
- City;
- State code;
- Zip code;

- Phone;
- Email;
- IP addresses;
- Latitude/Longitude; and
- ID numbers assigned by the commercial sources (such as secure hashes or LexisNexis ID).

Other information collected and stored by DA-TC includes:

- Protected critical infrastructure information;
- Law Enforcement Sensitive data provided by a Component to DA-TC that is historical in nature and not related to an active criminal case; and
- Video recording data.

The Account Request form for a DA-TC Lab system or cloud environment user collects the individual's:

- First name;
- Middle initial;
- Job title;
- DHS Personnel Identity Verification (PIV) Card number;
- Office phone;
- Office address;
- Office location;
- City and state;
- Organization; and
- Government or contractor status.

Lab Users and Partners

The U.S. Government, Federally Funded Research and Development Centers, DHS Components, HSE, industry, and academic researchers are potential users of the Lab resources. The HSE users include critical infrastructure sector representatives and federal, state, local, or tribal law enforcement, as appropriate. Users of the system are assigned specific roles, as system administrators, researchers, analysts, or users, as appropriate and in accordance with the legal and

policy guidelines developed for each research activity. These assignments enable DA-TC to control user access to appropriate data sets and applications required for each research activity based on a user's role (i.e., the user's research area and need to know). Data collections associated with each research activity are managed as separate storage volumes that require credentialed access to enable data use for each of the user roles defined—at a minimum access is granted, logged, and reviewed at the file level to administrator, project group, and individual levels.

Partner participation in or use of the Lab will be governed by an MOU that outlines the mission rationale for use of the Lab and establishes the basis for partnership with the Department. Access to Lab facilities and resources is granted through an Account Request form that is processed and approved by the DA-TC Director and the appointed Information System Security Officer (ISSO). Accounts and access are granted in accordance with DHS policies regarding access to information technology systems, requiring users to be cleared by personnel security and agreements to be executed including rules of behavior for usage of government systems. Further, each user is trained annually using general annual security and privacy training, and is provided specific training on handling different types of data being used, when required, prior to gaining access to any DA-TC system by DHS S&T or its research partners.

DA-TC Lab Systems: Physical On-Premise and Cloud Environment Systems

The Lab uses both a physical environment and cloud computing environment. Each of these environments is delineated by the security boundary accredited by S&T, which includes equipment, IT security tools, servers, and applications to allow for DA-TC staff to perform RDT&E activities.

The physical environment is neither physically nor virtually connected to the cloud environment. When data is to be transferred from a cloud environment to the physical environment for testing, the data is downloaded from the cloud environment onto an encrypted hard drive.[5] The data is then transferred into the physical environment in accordance with established data-handling guidance and in compliance with DHS policies.

DA-TC may also store data sets and perform RDT&E activities, such as data audits, technology assessments, and research experiments, in cloud computing environments. These cloud environments may include DHS, DHS S&T, DA-TC, or DHS component cloud commercial (Amazon Web Services, Microsoft Azure, and Google Cloud Platform), or the federal government's computing infrastructure. DA-TC uses the cloud computational environment as appropriate, as approved by the data owner, and in accordance with appropriate

[5] S&T ensures that the module is compliant with the Federal Information Processing Standard 140-2 (FIPS 140-2). FIPS 140-2 is a U.S. Government computer security standard used to approve cryptographic modules. *See* NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, CRYPTOGRAPHIC MODULE VALIDATION PROGRAM, *available at* https://csrc.nist.gov/Projects/Cryptographic-Module-Validation-Program/Standards (last visited August 10, 2020).

oversight and compliance guidance.

DA-TC RDT&E Data Lifecycle

DA-TC receives, collects, and uses data, including PII, to conduct RDT&E. When working with another DHS component or other operational partner, DA-TC codifies the use, sharing, and retention of the data in an MOU and data handling guide. The results of DA-TC's RDT&E may produce reports or knowledge products to inform DHS about technical capabilities. These reports focus on technical performance, mission impact, costs, and other characteristics of specific technologies. DA-TC redacts PII from such reports, unless PII is absolutely necessary for providing context to explain the results. DA-TC will mark documents with PII as sensitive to control dissemination. In exigent circumstances, such as indicators of a potential terrorist threat or activities potentially involving human trafficking, DA-TC will share any relevant data and information that it has collected with the proper DHS Component authorities with the appropriate federal, state, or local authorities for response. This sharing of PII is limited to what is necessary to respond and provide assistance to the individual. For example, DA-TC may share an individual's name; social media user name, handle, or alias; address or approximate location; phone number, email address, or other contact information that is made publicly available on social media; and possibly details of the individual's relevant circumstances. If any sharing must occur under exigent circumstances, DA-TC will confer with the S&T Privacy Office and the S&T Office of General Counsel.

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?

DA-TC, as part of DHS S&T, conducts RDT&E of advanced computational and analytics technologies that have potential to benefit critical DHS missions throughout the Components and fulfill its mission responsibilities under 6 U.S.C. § 182, *Responsibilities and Authorities of the Under Secretary for Science and Technology*. Prior to initiating a new project, S&T authorities under 6 U.S.C § 182 are viewed in conjunction with the requirements set forth in the Privacy Act of 1974 along with DHS policy to ensure all data sets are properly collected.

Data is collected and used in accordance with authorities that enable RDT&E activities within DHS S&T or in partnership with the data owner. This PIA will be updated or include appendices to discuss such authorities for collection and use. All adequate legal and privacy reviews will be conducted before DA-TC acquires or procures additional data sets. Commercial data is made available to DA-TC under terms and agreements that enable appropriate research and development. Research consortia data is made available to DA-TC under general terms and agreements that enable homeland security research and development to use this data in a manner

that is consistent with its intended use. Open source data will be collected and used in accordance with the scope of the RDT&E activity. Once the proper legal and privacy reviews are conducted, the above data sets may potentially be comingled with DHS Component data in furtherance of research and development.

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

DA-TC only collects, maintains, and uses data for RDT&E purposes consistent with the purpose and routine uses of the DHS/S&T-001 Research, Development, Test and Evaluation Records SORN[6] and respective component SORNs used by data owners for the original data collection.

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

The physical computing environment has a System Security Plan that was last updated on October 23, 2019. The system security documentation for the cloud environments was last updated on October 18, 2019. The Federal Information Security Modernization Act of 2014 (FISMA) requirements treat the physical DA-TC Lab and cloud environments as distinct FISMA systems that will have separate system security plans to be approved as part of the Certification and Accreditation process. These systems are undergoing review for authority to operate (ATO), pending publication of this PIA.

As DA-TC implements new tools and applications, S&T will assess the use of privacy sensitive technologies and how they collect, use, and maintain data for each specific project under the accredited security boundary for the DA-TC Lab. These privacy analyses will inform updates to this PIA, require updates to other PIAs, or be documented in a new PIA.

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

DA-TC abides by General Records Schedule 3.1, General Technology Management Records, item 11, which covers test files, data, and evaluation. Test and Evaluation files are considered temporary and cut off at the end of the calendar year after completion or cancellation of a project. All records are then destroyed or deleted five years after cutoff or one year after responsible office determines it is no longer needed for legal, audit, administrative, or business purposes.

Data owners will follow their agency's previously established records schedules for source

[6] The DHS/S&T-001 Research, Development, Test and Evaluation Records SORN is being updated to provide more sufficient transparency of S&T's collection of information related to RDT&E.

system records. In addition, DA-TC follows NARA and DHS applicable record schedules throughout the course of each project.

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.

The PRA requirements are accounted on a project-by-project basis. S&T will update this PIA accordingly should PRA requirements arise.

Section 2.0 Characterization of the Information

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

Information collected from government data sources may include information such as:

- Full Name;
- Home Street Address;
- Home Phone Number;
- Business Street Address;
- Business Phone Number;
- Email Address;
- EIN;
- SSN;
- TIN;
- IP addresses;
- Facial images;
- Video; and
- Audio.

Information collected from publicly available social media may include information such as:

- Social media handles;
- Account names;

- Email addresses;
- Pictures, video speech (live stream);
- IP addresses;
- Latitude and longitude of geo-tagged content;
- Phone numbers; and
- Posted content.

Information collected from commercial vendors and research consortia may include information such as:

- First name;
- Middle name(s);
- Last name;
- Date of birth;
- Gender (Male, Female, Other);
- Street;
- City;
- State code;
- Zip code;
- Phone;
- Email;
- IP addresses;
- Latitude, Longitude; and
- ID numbers assigned by the commercial sources (for example, secure hashes or LEX ID).

Other information collected and stored by DA-TC includes:

- Protected critical infrastructure information;
- Law Enforcement Sensitive data provided by a Component to DA-TC that is historical in nature and not related to an active criminal case; and
- Video data.

The Account Request form for a DA-TC Lab system or cloud environment user collects the individual's:

- First name;
- Middle initial;
- Job title;
- DHS PIV card number;
- Office phone;
- Office address;
- Office location;
- City and state;
- Organization; and
- Government or contractor status.

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

The data used in DA-TC RDT&E activities originates from S&T programs, other DHS Components, government agencies, commercial data providers, open source, and research consortia. Once the proper project-specific legal and privacy reviews are conducted, DA-TC researchers may independently collect open source data, or receive open source data provided to DA-TC by a data-sharing partner which is properly documented and governed through an information sharing access agreements (ISAA) or MOUs.

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. DA-TC's RDT&E activities use data from commercial sources to better understand the potential benefits of DHS using such data sets, such as saving analytic resources or increasing the accuracy of decision making. Project-specific legal and privacy reviews will be conducted before DA-TC collects open source data to ensure proper adherence to all legal, privacy, and DHS policies. Open source data, such as information from publicly available social media, may be used for the development or evaluation of a tool, such as an open source media aggregator, that may be recommended for deployment by a DHS component. Such data may also be used to enrich or provide context to other data provided to DA-TC by a data owner.

2.4 Discuss how accuracy of the data is ensured.

The data owner is responsible for the accuracy, completeness, and quality of the data provided to DA-TC. In most instances, the data quality is managed by the data owner or source since DA-TC is not the original source of the data. However, data sets may contain inaccurate information. As a research organization, DA-TC will try to improve DHS's strategic capability to increase data accuracy and understand the impact on analytic outcomes and decision making. However, DA-TC does not attempt to make, nor does it have authority to make, operational decisions. If a Component decides to use DA-TC's data for an operational purpose, the Component is responsible for verifying the accuracy of S&T data.

2.5 Privacy Impact Analysis: Related to Characterization of the Information

Privacy Risk: There is a risk that DA-TC is collecting more PII than necessary for DA-TC to perform its RDT&E activities.

Mitigation: This risk is partially mitigated. The information required is not always identified at the initiation of a project. At that time, researchers meet with the S&T Privacy Office to discuss how to minimize the PII used while performing the required research and then work with the program to complete a Privacy Threshold Analysis (PTA) for each project. For every activity, data obtained or collected is limited to the purpose of that specific activity. Data collected for one purpose or activity is not shared or used for another project or activity. All PII is deleted or returned to the data owner once the project is completed. However, DA-TC may retain aggregated and anonymized research data indefinitely, as it may help inform future RDT&E efforts.

Privacy Risk: There is a risk that information about individuals is not accurate, timely, and complete.

Mitigation: This risk is partially mitigated. In most instances, the owner of the data (e.g., another DHS component) manages data quality. DA-TC will inform the data owner if it identifies data quality or integrity issues with the data it receives. The data owner is responsible for the accuracy and quality of their data. Regardless, DA-TC uses information solely for RDT&E, not to make an operational decision about an individual. One of the goals of DA-TC's RDT&E is to evaluate the accuracy, timeliness, and completeness of the data DHS uses.

Section 3.0 Uses of the Information

3.1 Describe how and why the project uses the information.

DA-TC's collection and use of government, commercial, research consortia, and open source data is scoped at the beginning of each research activity to accomplish an RDT&E

objective. This scoping occurs as a collaboration between the DHS S&T Privacy Office, the DHS S&T Office of General Counsel, and DA-TC researchers. For any research effort, DA-TC identifies proposed data sets, proposed data Components, and a technical approach that will be needed to complete the activity. DA-TC identifies potential partners such as federal, state, and local government entities to participate in research activities. S&T enters into ISAAs, such as an MOU, with the partners or data providers that set the terms and conditions that govern data access, use, sharing, and deletion. Final agreements must be reviewed and approved by the respective legal and privacy oversight functions of the parties, prior to agreement execution.

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

Yes. DA-TC will produce information regarding the functional performance of electronic search, query**,** and analysis technologies. The results of RDT&E will inform partners (e.g.**,** DHS operational Components) regarding the accuracy, reliability**,** and other operational characteristics of such technologies for decision making.

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

The research activity manager determines assigned individuals on the project and the system owner reviews and approves access. Only DHS-badged individuals, including those from other DHS Components, may access the Lab, which consists of both a physical on-premise computing environment and a cloud computing environment. All users must have a DHS Entrance on Duty (EOD) clearance and valid Personal Identity Verification (PIV) card and must complete and sign a DA-TC User Account Request and Rules of Behavior form. Users of the DA-TC Lab only have access to the data set(s) and application(s) required based on that user's need to know, which is determined by the DA-TC Lab Project Managers and the System Owner.

3.4 Privacy Impact Analysis: Related to the Uses of Information

Privacy Risk: There is a risk that users may use information stored in the Lab for unauthorized purposes.

Mitigation: This risk is partially mitigated. All users and computing resources are audited on a regular and ad-hoc basis. These audits record lab usage and individuals that are accessing data sets. Storage and file access are part of the audit. Patterns of usage should be consistent with the research goals and interim results that an individual's work products indicate. For example, if a user is regularly reviewing algorithms for international global trends in a data set, and his or her

audit reveals a repeated review of singular locations, an inquiry would be opened and an audit leading to an authorized rationale or a violation, resulting in an immediate removal of access would be recommended. Discovery of unauthorized usage of these systems will be reported immediately to management, compliance, and legal authorities for remedial or punitive action. Intentional unauthorized use will result in a prohibition of access to the facility permanently.

Privacy Risk: Users may use information stored in the Lab beyond the purpose of an RDT&E activity.

Mitigation: This risk is partially mitigated. Lab users and administrators are held accountable for the protection of PII. All Lab users receive training for their system user role and the acceptable use of the rights/privileges associated with their user role. All users complete, sign, and agree to the DA-TC User Account Request and Rules of Behavior. All DHS personnel are required to complete privacy training. This includes Lab users, data owners, and systems administrators. Users are required to complete data-specific training when information requires special protection. The Lab system audits access to PII in compliance with privacy principles and all applicable privacy protection requirements. Data obtained for one project is only collected for the purpose of the specific project. Data collected for one purpose is not shared or used for another project or initiative. Data is purged or returned to data owner at the end of RDT&E activity. All Lab system users must have an EOD and valid PIV to access the Lab.

Privacy Risk: There is a risk that information collected or generated under RDT&E authorities may be used for operational purposes.

Mitigation: This risk is partially mitigated. DA-TC does not use Lab data for operational purposes, except in exigent circumstances when DA-TC provides the information to the appropriate authorities.

In the event DA-TC partners with an operational Component and the operational Component seeks to use the RDT&E information for operational purposes, the Component must determine that the proposed use falls within the Component's operational authorities. The Component also must determine that the data use is consistent with the Component's applicable SORN(s). In addition, Components may be required to complete additional privacy compliance requirements (e.g., PTA, PIA) for the operational use.

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.

This PIA functions as the primary form of notice to individuals that DA-TC conducts

research on government, commercial, research consortia, and open source data sets. Adequate legal and privacy reviews will be conducted before data is collected to ensure all legal, privacy, and DHS policies are properly adhered to appropriately. Government data will be collected in accordance with authorities that enable research and development activities within DA-TC or in partnership with the data owner.

For DHS Component source system data, Components provide individuals with notice at the initial collection, as appropriate. The source systems have undergone appropriate privacy analysis and documentation. In addition, the data's use by DA-TC has been determined to be compatible with the purpose for which the data was originally collected. S&T will conduct similar analysis when using other government data.

Commercial data will be made available to DA-TC under terms and agreements that enable appropriate research and development. Consortia data will be made available to DA-TC under general terms and agreements that enable homeland security research and development to use this data in a manner that is consistent with its intended use.

Open source data will be collected and used in accordance with the scope of the RDT&E activity. Notice of DA-TC's use of publicly available social media data will be provided in a forthcoming PIA on that topic.

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

In most cases, DA-TC ingests copies of data provided by data owners, such as other government agencies or commercial data vendors. For government agency data, consent for the collection of the data, and the opportunity to decline or opt out, occurred when the government agency initially collected the data, as appropriate. The use of government agency data by DA-TC must be determined to be compatible with the purpose of the original collection. In the instance of commercial data, individuals consent to uses, decline to provide information or opt out based on the privacy policy of the owner of the source that is providing this data to DA-TC.

DA-TC does not provide notice to individual social media users when collecting open source data. Notice of this collection is provided by this PIA and a forthcoming social media PIA. Nevertheless, DA-TC only collects publicly available social media information that users have posted to social media platforms available to anyone with an internet connection.

4.3 Privacy Impact Analysis: Related to Notice

Privacy Risk: There is a risk that individuals will not receive notice that their information may be used in DA-TC RDT&E activities.

Mitigation: This risk is partially mitigated. This PIA provides notice that DA-TC may

collect or store information about individuals for research purposes in accordance with legal, privacy, and DHS policies. Much of the data transferred to DA-TC is owned by DHS S&T's data sharing partners. Individuals consent to uses, decline to provide information, or opt out during the initial information collection, if applicable, by the data sharing partner. Components, whose data is used by DA-TC, would have provided notice for that data through PIAs, SORNs, and Privacy Act Statements/Notices, as appropriate. Research consortia often use willing and paid or volunteer participants in studies that result in research data sets. The DHS S&T Office of General Counsel and S&T Privacy Office will review the authorities and terms associated with the use of each data set and ensure they are acceptable for DA-TC RDT&E activities. Only data that complies with legal, privacy, and DHS policies for research can be used by DA-TC.

While DA-TC does not provide individual users notice prior to viewing or collecting open source data, DA-TC only collects information that users have contributed to publicly available websites. Users that post content to publicly available social media platforms make that information available to all members of the public. DA-TC provides notice of its collection of open source data through this PIA and more specifically under a forthcoming PIA for S&T's use of publicly available information in RDT&E.

Section 5.0 Data Retention by the project

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

DA-TC will not retain research data beyond the end of a specific research activity. Once the RDT&E effort is concluded DA-TC must delete or return data sets to the data owner, depending on the data source. Per the DHS/S&T-001 Research, Development, Test and Evaluation Records SORN, PII collected during the project is retained for the duration of the project; at the conclusion of the project, PII is returned to the providing Component or destroyed. Researchers may retain aggregated research data (without PII) indefinitely, as it may help inform future RDT&E efforts.

5.2 Privacy Impact Analysis: Related to Retention

Privacy Risk: There is a risk that DA-TC may retain data longer than necessary.

Mitigation: This risk is mitigated. DA-TC fully mitigates this risk via routine reviews and audits. At the end of each effort, the ISSO will ensure that data stored in the physical and/or cloud computation environments is properly deleted, including providing certificates of deletion to the data owners and the S&T Privacy Office.

Privacy Risk: There is a risk that data may be replicated in different cloud computing environments (for example, an Amazon Web Services cloud environment and a Microsoft Azure cloud environment) during transition within the cloud environments.

Mitigation: This privacy risk is partially mitigated. Through established business

processes, the planning for experiments using multiple cloud environments will address the need for any additional controls necessary to protect data in each cloud environment as part of the specific compliance documentation required for that experiment. Planning includes the development of effective data handling guidance and approved PTAs. DA-TC is conducting research that will improve and strengthen security and protection of data for future multi-cloud architectures.

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.

For any research effort, DA-TC identifies proposed data sets, proposed data elements, and a technical approach that will be needed to complete the activity. DA-TC also identifies any potential partners, including other federal, state, and local government entities, whose security posture will be improved or informed via the research activity. DA-TC establishes agreements with any partners or data providers, such as interagency acquisition (IAA) agreements, ISAAs, MOUs, or Cooperative Research and Development Agreements (CRADA), that set the terms and conditions that govern data access, use, sharing, and deletion. Final agreements must be reviewed prior to execution by the respective legal and privacy oversight functions of the parties. As a result, DA-TC may only share data if such sharing is permitted by the data owner as described in the original agreement.

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

DA-TC's sharing of information it collected is consistent with Routine F of the DHS/S&T-001 Research, Development, Test and Evaluation SORN, which permits sharing with contractors and their agents, grantees, experts, consultants, and others performing or working on a contract, service, grant, cooperative agreement, or other assignment for DHS, when necessary to accomplish an agency function related to this system of records. Individuals provided information under this routine use are subject to the same Privacy Act requirements and limitations on disclosure as are applicable to DHS officers and employees.

DA-TC's information sharing is also consistent with Routine Use G of the DHS/S&T-001 Research, Development, Test and Evaluation Records SORN, which permits sharing with an appropriate federal, state, tribal, local, international, or foreign law enforcement agency or other appropriate authority charged with investigating or prosecuting a violation or enforcing or implementing a law, rule, regulation, or order, where a record, either on its face or in conjunction with other information, indicates a violation or potential violation of law, which includes criminal,

civil, or regulatory violations and such disclosure is proper and consistent with the official duties of the person making the disclosure.

Any external sharing of Component source system data would require analysis of the specific SORN involved to ensure that the sharing is compatible with the purpose for which the information was initially collected.

6.3 Does the project place limitations on re-dissemination?

Any external information sharing would require an ISAA. The ISAA oversight and approval process will require explicit consideration of appropriate limitations and the inclusion of terms and conditions on the re-dissemination of the data covered by the agreement.

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

Data owners control disclosures by incorporating disclosure terms and conditions in the ISAA and provide any data handling guides developed for the use, disclosure, and disposition of data sets related to each specific research and development activity.

6.5 Privacy Impact Analysis: Related to Information Sharing

Privacy Risk: There is a risk that DA-TC may unintentionally share PII with a third party or when providing equipment to research partners for RDT&E activities.

Mitigation: This risk is mitigated. The terms of data sharing are reviewed, documented, and agreed upon by DA-TC and partner organizations prior to initiating a project. Information can only be shared in accordance with the authorities for data sharing. If, as part of an RDT&E activity, DA-TC provides equipment to research partners, DA-TC will sanitize this equipment as appropriate to ensure no PII or other data is inadvertently provided to the research partner. In final reports, DA-TC will remove all PII, unless it provides necessary context. In these cases, DA-TC will mark documents to communicate that the information requires special handling.

In the event DA-TC unintentionally shares PII with a third party, or in the case of any other suspected or confirmed privacy incident, DA-TC will respond to and mitigate the incident in accordance with DHS's Privacy Incident Handling Guidance.[7]

[7] *See* U.S. DEPARTMENT OF HOMELAND SECURITY, PRIVACY INCIDENT HANDLING GUIDANCE, *available at* https://www.dhs.gov/publication/privacy-incident-handling-guidance-0.

Section 7.0 Redress

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

For information that DA-TC receives from government data owners, access by individuals to their information is governed by the respective data owner's Freedom of Information Act (FOIA) and/or Privacy Act (PA) request policy.

For information DA-TC receives or collects from commercial data providers, research consortia, and open sources, any individual who may desire to access whatever information DA-TC may have collected under this initiative may submit a FOIA or PA request to the DHS FOIA Office:

Chief Privacy Officer/Chief Freedom of Information Act Officer
U.S. Department of Homeland Security
2707 Martin Luther King Jr. Avenue, SE
Washington, D.C. 20528

The DHS/S&T-001 Research, Development, Test, and Evaluation Records SORN contains instructions for accessing information under the "Notification Procedure" section.

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

For information provided to DA-TC originating from another government data-sharing partner, individuals desiring to correct inaccurate or erroneous information can seek the support of the originating system's public records or Component FOIA/Privacy Act Officer.

DA-TC collects information from publicly available social media and uses in its RDT&E activities. This content was made publicly available by users whom have published it voluntarily. As a result, users have the opportunity not to provide information prior to posting. Individuals may also choose to adjust the privacy settings on their various social media platforms or other websites to restrict access to this information as they see fit. Lastly, individuals also retain the right and ability to remove previously posted information from their respective online accounts via the policies and practices of the respective websites.

Those persons from whom PII may be collected and who are seeking access to any record collected under this initiative may submit a FOIA or PA request. DHS/S&T-001 Research, Development, Test, and Evaluation Records SORN contains instructions for accessing information under the "Notification Procedure" section.

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

DA-TC notifies individuals of the redress procedures for this initiative through this PIA and through the DHS/S&T-001 Research, Development, Test, and Evaluation Records SORN. The respective owners of data sets shared with the DA-TC for RDT&E, such as another government agency or a commercial data vendor, also have established procedures for correcting information. Individuals seeking to correct their information contained in data sets provided to DA-TC from a data-sharing partner should follow the procedures of the data-sharing partner.

7.4 Privacy Impact Analysis: Related to Redress

Privacy Risk: There is a risk that an individual will not know how to correct inaccurate information about themselves that is used in DA-TC RDT&E activities.

Mitigation: This risk is partially mitigated. DA-TC is not using PII for operational purposes and thus, does not depend on the accuracy of the data regarding an individual. However, much of the data used by DA-TC is a copy of data provided by systems owned by others (not including publicly available information). Therefore, redress of inaccurate or erroneous information is obtained through the originating owner of the data. DA-TC does not provide procedures to correct publicly available information. The individual about whom the PII is collected, in many instances, made the information publicly available.

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?

All data within the Lab is audited while it is maintained within the system, regardless of the environment. Actions against data, including events, such as data loading, editing, and access by applications or users, is logged and timestamped. All system logs are maintained on a dedicated host within the DA-TC system. Deletion documentation is provided to the end user once the pilot/test is completed to ensure the data has been purged appropriately. Users are required to follow the DA-TC Rules of Behavior and their DHS's information use policies.

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

All DHS personnel, including DA-TC users and system administrators, are required to complete privacy training. In addition, the DA-TC team conducts an annual training for DA-TC staff to review security policies and procedures. Data-specific training will be required by DHS Components as necessary (e.g., Cybersecurity and Infrastructure Security Agency (CISA)

Protected Critical Infrastructure (PCII) training will be provided to any user accessing that data).

Orientation/user guidance, including privacy guidance and training, as needed, is a continual part of the use of any DA-TC resources and is specific to the research activity. All users sign DHS user behavior agreements that spell out the general guidelines of all systems involved. Depending on the project involved, the DA-TC team will receive additional training and guidance prior to and while engaged with the project.

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

To access the on-premise physical computing environment, a user must be physically present at DA-TC's facilities. An individual seeking access to the physical environment must have an EOD clearance and a valid PIV card, and must complete and sign a DA-TC User Account Request and Rules of Behavior form.

Individuals who seek access to the physical environment must submit the DA-TC Account Request Form, which is signed by the requestor and by the individual's Federal Sponsor (or Security Office). A user's Account Request must then be approved by (1) his or her Supervisor; (2) Security; and (3) System Owner. The Account Request is completed by the DA-TC Lab System Administrator, once the necessary approvals are in place.

A Lab user's access to data and applications must be approved by the Lab's System Owner. A Lab user's role, is determined through the User's Account Request. DA-TC Lab Project Managers and the System Owner provided project specific access requirements once a user is approved for access. The physical environment provides end users with access to the data set(s) and application(s) required based on that end user's need to know.

This segmentation from the data and work products of other users is implemented through various operational and technical system controls. The physical environment employs role-based access to provide an initial set of rights to end users, and this assignment enables protection from users logging into systems and applications to which they are not authorized. System protections prevent users from accessing servers that may house data, work products and Administrator functions. Application-level access controls prevent end users from using tools that may be able to read, write, or delete data in the physical environment through data obfuscation and limited functionality.

Like a user of the physical on-premise computing environment, an individual seeking access to the cloud computing environment must have a DHS EOD clearance and a valid PIV card and must complete and sign a DA-TC User Account Request and Rules of Behavior form. The only difference in the access approval process between the physical on-premise system and the

cloud environment is that cloud environment users need not be physically present at the DA-TC's facilities. Additionally, only badged DHS employees or contractors will have access to the cloud environment.

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?

DA-TC establishes documented agreements with any partners or data providers, such as IAAs, ISAAs, MOUs, or CRADAs, that set the terms and conditions that govern data access, use, sharing, and deletion. Final agreements must be reviewed and approved by the respective legal and privacy oversight functions of the parties, prior to execution.

DA-TC privacy risk analysis also addresses proposed information sharing for specific initiatives. A DHS adjudicated PTA is required before any data is transferred, stored, and available for access by DA-TC personnel. The PTA will discuss the PII and PII sources, how DA-TC will use the PII, how the DA-TC shares and disposes of the PII, and other PII-related matters. Once the project PTA is approved, PII may be uploaded and stored within the Lab in connection with the specific DA-TC RDT&E activity.

Privacy Risk: There is a risk that the "One DHS" policy may be misinterpreted to permit unauthorized users to access the physical and cloud instances without the proper authorization or agreements in place.

Mitigation: This risk is mitigated through current technical controls and business processes. For example, to access the on-premise physical environment, a DA-TC guest must be escorted at all times. DA-TC maintains logs of guests and the purpose of their visit.

The escort must be an approved DA-TC user who meets the following criteria: (1) DHS employee or contractor with (2) approved unescorted access to DHS information after undergoing a successful suitability investigation and determination; (2) a completed user account request form; (3) a signed Rules of Behavior agreement; and (4) completed DA-TC and project specific training. These are the same requirements for the cloud instance. All users are granted access to select application(s) and data sets based on need to know as determined by ISSO and System Owner. Access is limited to the timespan associated with the specific project. Only DHS personnel are granted access to any of the DA-TC environments.

Privacy Risk: There is a risk that unauthorized individuals may access DA-TC's research data stored in the Lab.

Mitigation: This risk is mitigated. DA-TC requires encryption of the data when stored in the physical and cloud computation environments. Access to DA-TC's research data is controlled by requiring a user to physically insert a PIV card into a reader device and provide a PIN. Access

is further controlled by limiting access only to DA-TC-designated personnel with a need to know. Similarly, DA-TC requires encryption when data is transferred between the physical and cloud environments.

Privacy Risk: There is a risk that improper configuration of the cloud computing environment would expose data to all cloud environment users with access as opposed to DA-TC-authorized users only.

Mitigation: This risk is mitigated. DA-TC executes system assessments, independent control assessments, and continuous monitoring in the same manner as all DHS IT systems. Compliance and vulnerability scans of the system are conducted at least on a weekly basis.

Contact Official

Stephen Dennis
Data Analytics Technology Center
Science and Technology
(202) 254-5788

Responsible Officials

James Johnson
Director, Office of Science and Engineering Technology Centers
Science and Technology Directorate
U.S. Department of Homeland Security

Maria Petrakis
Privacy Officer
Science and Technology Directorate
U.S. Department of Homeland Security

Approval Signature

Original, signed copy on file at the DHS Privacy Office.

_____

Dena Kozanas
Chief Privacy Officer
U.S. Department of Homeland Security
(202) 343-1717

Appendix A

Data Audits

At the publication of this PIA, there are no Data Audits to document. S&T will update Appendix A when new Data Audits are added.

Appendix B

Technology and Tool Assessments

Background

The DHS S&T Data Analytic Technology Center (DA-TC) is conducting technical assessments and providing technical support to the development and use of commercial-grade speech processing and analysis technology. Specifically, DA-TC is facilitating technology assessments that enable a government determination of technical performance to support mission applications.

Technology assessments of the capabilities will consist of word spotting on limited data sets. These assessments will also be used to make determinations about the robustness of transcription for a non-English language. The goals of this U.S. Customs and Border Protection's (CBP)[8] research, development, testing, and evaluation (RDT&E) activity, for which DA-TC is providing support, are to provide the Department with information and technical guidance to: (1) support critical homeland security missions, and (2) understand the state of research and performance of emerging technology related to future DHS applications.

DA-TC's Role

During this technology assessment, DA-TC will coordinate and provide technical support through subject-matter experts (SME). Technical SMEs will provide advice regarding operation of the tools, the recording of technical performance metrics, the determination of appropriate metrics for mission impact, and the translation of technical performance to potential mission outcomes. DA-TC staff will collaborate in the formulation, editing, and production of a technical report that helps end users document the findings from the assessment for their internal use.

DA-TC will also provide CBP with equipment with the speech processing and analysis tools installed.

Secure Multi-Party Computation

The DHS S&T Data Analytic Technology Center (DA-TC) is assessing the potential homeland security applications of secure multi-party computation (SMPC) tools available in the commercial marketplace. SMPC, a subfield of cryptography, is a class of technologies focused on creating methods for multiple parties to conduct computation across data sets while preserving the privacy and security policies of the data owners. SMPC tools could permit DHS to independently and simultaneously analyze data held by multiple Components without having to transfer data between systems for analysis. Thus, the many benefits include eliminating the need for sharing

[8] *See* U.S. DEPARTMENT OF HOMELAND SECURITY, U.S. CUSTOMS AND BORDER PROTECTION, PRIVACY IMPACT ASSESSMENT FOR BORDER SURVEILLANCE SYSTEMS, DHS/CBP/PIA-022 (2014 and subsequent updates), *available at* https://www.dhs.gov/privacy-documents-us-customs-and-border-protection.

and storing identical data in multiple places.

DA-TC's effort will encompass the characterization of the technical merits and initial performance of these tools. DA-TC and its contractor staff will test and evaluate SMPC tools in the Lab to validate the capabilities of each respective tool. The results will ensure that S&T and the rest of DHS have a foundation upon which to make potential future procurement and architecture decisions. In the future, partnering DHS Components may deploy these tools in their own approved environments to best reflect realistic scenarios for testing and evaluation.

Background

DHS S&T is part of an interagency working group with the DHS Countering Weapons of Mass Destruction Office (CWMD) and the Department of Defense's Joint Program Executive Office for Chemical, Biological, Radiological, and Nuclear Defense (JPEO-CBRND), known as the Countering Weapons of Mass of Destruction Alliance, whose mission is to identify, plan, and execute collaborative activities that enable the development and implementation of joint chemical, biological, radiological, and nuclear (CBRN) capabilities to achieve common, compatible, and complementary capabilities. The Alliance focuses on development of capabilities for CBRN detection and initial response capability for U.S. government mission needs.

DA-TC's Role

The DHS S&T DA-TC supports the joint missions of Alliance partners. DA-TC will develop and evaluate methods and tools for analyzing data collected by CBRN detection sensors. Alliance partners will share their data with DA-TC for use in research and evaluation. Alliance partners will leverage the results of these activities for future decision and analytic applications.

Appendix C

Prototyping

At the publication of this PIA, there are no Prototyping projects to document. S&T will update Appendix C when new Prototyping projects are added.

Appendix D

Pilots

At the publication of this PIA, there are no Pilot projects to document. S&T will update Appendix D when new Pilot projects are added.