CAOE | CENTER FOR ACCELERATING OPERATIONAL EFFICIENCY

A DEPARTMENT OF HOMELAND SECURITY CENTER OF EXCELLENCE

PETS4HSE Report

Ross Maciejewski
Professor, Arizona State University
Director, Center for Accelerating Operational Efficiency - a Department of Homeland Security Center of Excellence.

January 31, 2023

ASU USC Northwestern UTEP UCI UBC UNIVERSITY AT ALBANY State University of New York PennState ILLINOIS

University of Nebraska Omaha UNLV THE UNIVERSITY OF ALABAMA MICHIGAN STATE UNIVERSITY Creighton UNIVERSITY UB Western Carolina UNIVERSITY Monmouth COLLEGE JOHN JAY COLLEGE OF CRIMINAL JUSTICE

Privacy Enhancing Technologies for the Homeland Security Enterprise (PETS4HSE)

- Workshop held June 21, 2022
- Sponsors:
- Arizona State University
  Center for Accelerating Operational Efficiency, A Department of Homeland Security Center of Excellence
- DHS Chief Privacy Officer
- https://pets4hse.org/
  - “Privacy-enhancing technologies (PETs) under development promise the ability to control the sharing and use of sensitive information while minimizing the risk of unauthorized use.
  - “These technologies have been under development by researchers for nearly four decades but have been slow to migrate from the research lab into operational use.
  - “This workshop will help to speed the pace of change by engaging researchers and practitioners in a joint endeavor to solve the “hard problems” that will enable us to solve the practical problems so we can put new technologies into practice while still keeping risk levels low.”

Acknowledgment

- The PETS4HSE Workshop could not have happened without the immense support of two individuals:

Simson Garfinkel

Lynn Parker Dupree
Chief Privacy Office
Department of Homeland Security

What are PETS?

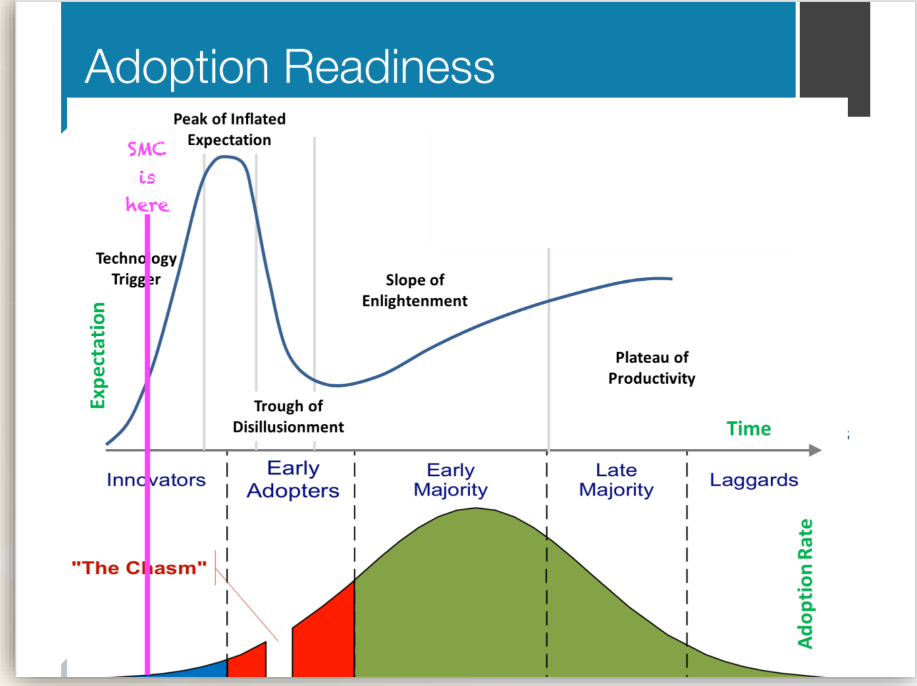
- “PETS” – Term emerged in 2003
    - “The Status of Privacy Enhancing Technologies (PET) online and offline,” John Borking, Associate Commissioner of the Dutch Data Protection Authority, The Netherlands. https://link.springer.com/content/pdf/10.1007/978-0-387-35696-9_15.pdf
    - http://dud.inf.tu-dresden.de/Anon_Terminology.shtml
- Originally covers a broad range of functional requirements:
    - Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management.
    - Examples: PGP (Pretty Good Privacy), TLS (Transport Layer Security), TOR (The Onion Router)
- Increasingly applied to technology for Privacy Preserving Data Publishing and Privacy Preserving Data Analysis:
    - **Differential Privacy** – Publish aggregate statistical data while limiting privacy loss to Individuals
        - Example: 2020 Census
    - **Secure Multiparty Computation** – Learn the results of a computation without revealing inputs to each other.
        - Example: Find the average age of everyone in a room without revealing anyone’s age to anyone else.
    - **Homomorphic Encryption** – Performing operations on encrypted data.
    - **Private Set Intersection** – Find common entries between two datasets without revealing either.
    - **Private Search** – Perform a search on a database without revealing search terms or retrieved records.

PETS4HSE Origin and Implementation

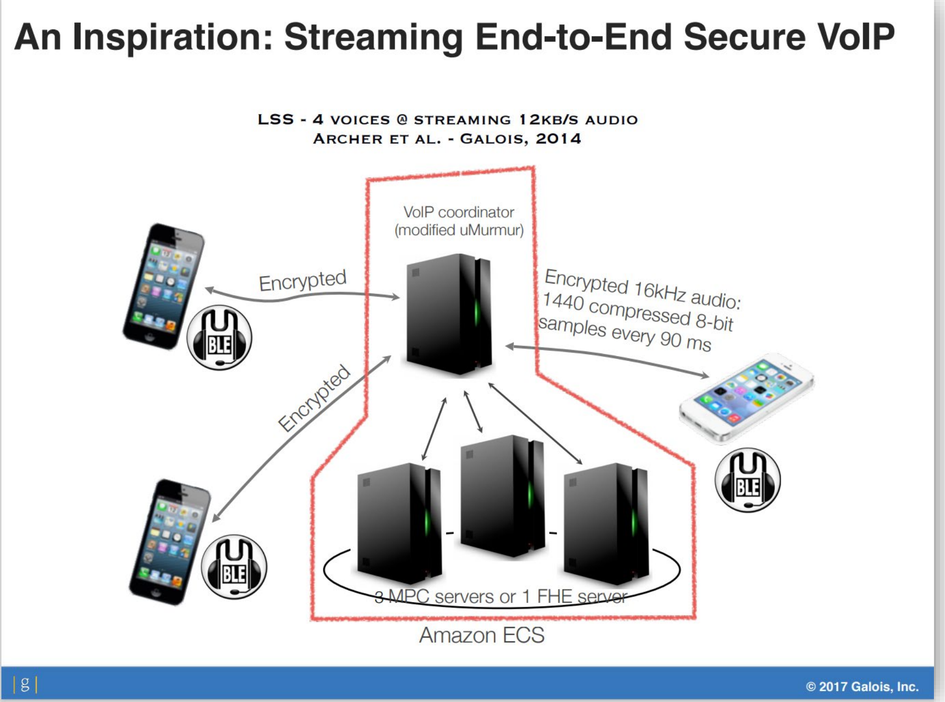
- November 2017 – Workshop on Multi-party Computing at US Census Bureau
  - Census Bureau knew that it wanted to use the technology but wasn’t sure how.
  - A day of presentations about what was available from academia and industry.
  - Main result: November 2020 JASON report, “Secure Computation for Business Data” (JSR-20-2E)

PETS4HSE Origin and Implementation

- Two Slides from Galois Presentation



Adoption Readiness



An Inspiration: Streaming End-to-End Secure VoIP

LSS - 4 voices @ streaming 12kb/s audio
Archer et al. - Galois, 2014
VoIP coordinator (modified uMurmur)
Encrypted
Encrypted 16kHz audio: 1440 compressed 8-bit samples every 90 ms
Encrypted
BLE
BLE
BLE
3 MPC servers or 1 FHE server
Amazon ECS
g

PETS4HSE Origin and Implementation

- January 2022 – Initial idea to have a workshop "in which people from DHS talk about their algorithmic needs for privacy, and people from the privacy community talk about their bag of solutions."
  - Core questions: funding, venue, format
  - Explored partnership with other federal agencies
  - Meeting between DHS CPO and OSTP
  - DHS Chief Privacy Officer agrees to fund

- Key goals of the workshop:
  - Teach DHS leaders and program managers about the opportunities offered by PETs.
    - This is a fundamental new way of thinking about data analysis within the government.
    - PETs do not fit well with current policy, IT systems, training, skills, etc.
  - Expose PETs researchers to the needs of government.
  - Explain to PETs researchers the "messiness" of government data practices.
    - Most PETs can't be used in a government setting without a lot of adoption.

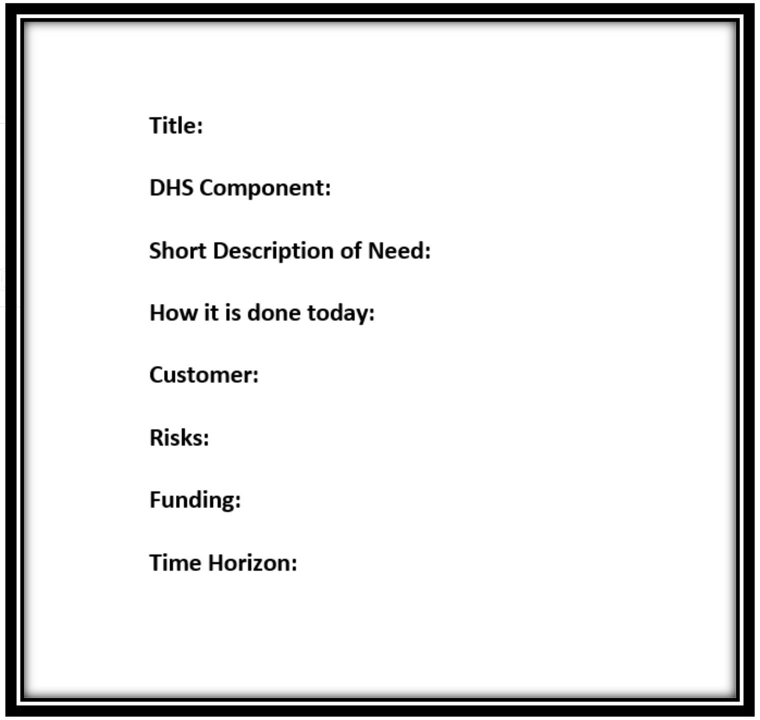
PETS4HSE Origin and Implementation

- February 2022 - DHS develops a tasking order for DHS Center of Excellence at Arizona State University
  - Administrative support
  - Venue
  - Food
  - Travel support for invited speakers

- We decided to go with a crazy break-next deadline of a workshop in June
  - July is dead for the federal government; academics would be too busy in August and Sept.
- Two calls for participation!
  - Call #1 – Use cases
  - Call #2 – PETs researchers

Soliciting Use Cases

- March 2022 – DHS Privacy Officer solicits “use cases” / “problem statements” from DHS components
  - Operational Components: USCIS, Coast Guard, CBP, CISA, FEMA, FLETC, ICE, Secret Service, TSA
  - Support Components: MGMT, S&T, CWMD, I&A, Operations
- We said that the use cases would be public.
- We develop a template for use cases.
  - They didn’t know what to submit!
- We collected and edited the use cases.
- We then *verified that we could publicly release them!*
  - Good thing! We lost several of the use cases. ☹
  - 11 use cases from OCHIO, CISA, ICE, OBIM and USCIS
  - https://pets4hse.org/PETS4HSEUseCases.pdf



Title:

DHS Component:

Short Description of Need:

How it is done today:

Customer:

Risks:

Funding:

Time Horizon:

Use Cases: Lessons Learned

- Mistakes we made:
  - “Use case” template should have included instructions and examples.
  - We should have used consistent terminology (not “use cases” and “problem statements.”

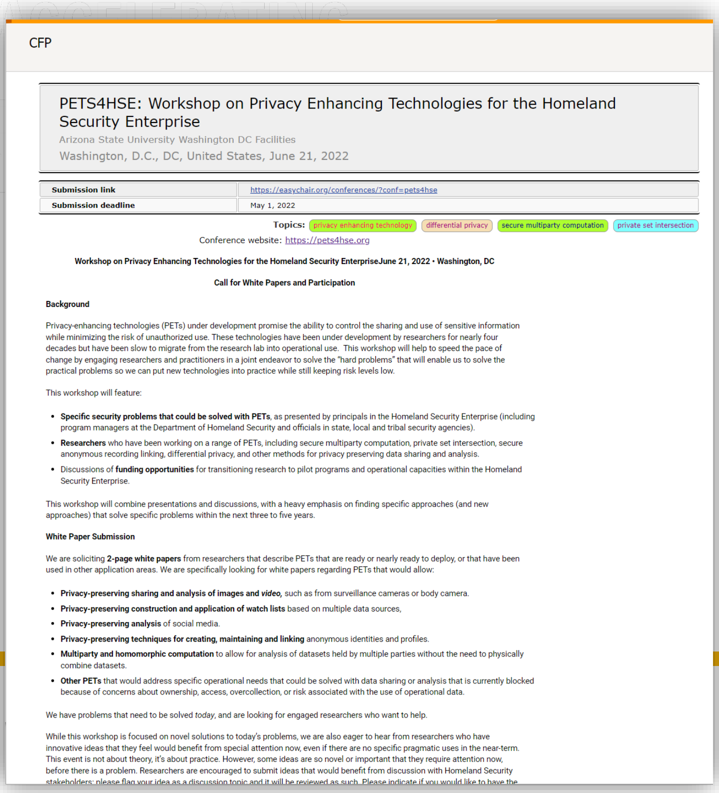
- Challenges encountered from Call #1
  - Components had difficulties coming up with use cases
  - Some components did not want use cases public even though they were deemed unclassified

- We would have received more use cases with more effort:
  - One-on-one meetings to explain PETs in detail.
  - Developing the use uses in collaboration with the component.
  - More focus on making the use case releasable.

- Moving forward:
  - Use cases inspire both developers & other potential users.
  - It would be useful to have a catalog of potential uses of PETs within the government.

Engaging PETS Researchers

- Call developed between Chief Data Officer Directorate, Privacy, and ASU
- Call posted to EasyChair at https://easychair.org/cfp/PETS-4-HSE
  - EasyChair set up as submission site
- Call promoted by email and Twitter.
- Lessons Learned:
  - Use a consistent acronym
    - (We used PETS-4-HSE and PETS4HSE)
  - Have the conference site ready to go
  - Build the program committee early.



CFP

PETS4HSE: Workshop on Privacy Enhancing Technologies for the Homeland Security Enterprise

Arizona State University Washington DC Facilities

Washington, D.C., DC, United States, June 21, 2022

Submission link	https://easychair.org/conferences/?conf=pets4hse
Submission deadline	May 1, 2022

Topics: privacy enhancing technology | differential privacy | secure multiparty computation | private set intersection

Conference website: https://pets4hse.org

Workshop on Privacy Enhancing Technologies for the Homeland Security EnterpriseJune 21, 2022 • Washington, DC

Call for White Papers and Participation

Background

Privacy-enhancing technologies (PETs) under development promise the ability to control the sharing and use of sensitive information while minimizing the risk of unauthorized use. These technologies have been under development by researchers for nearly four decades but have been slow to migrate from the research lab into operational use. This workshop will help to speed the pace of change by engaging researchers and practitioners in a joint endeavor to solve the "hard problems" that will enable us to solve the practical problems so we can put new technologies into practice while still keeping risk levels low.

This workshop will feature:

- **Specific security problems that could be solved with PETs**, as presented by principals in the Homeland Security Enterprise (including program managers at the Department of Homeland Security and officials in state, local and tribal security agencies).
- **Researchers** who have been working on a range of PETs, including secure multiparty computation, private set intersection, secure anonymous recording linking, differential privacy, and other methods for privacy preserving data sharing and analysis.
- Discussions of **funding opportunities** for transitioning research to pilot programs and operational capacities within the Homeland Security Enterprise.

This workshop will combine presentations and discussions, with a heavy emphasis on finding specific approaches (and new approaches) that solve specific problems within the next three to five years.

White Paper Submission

We are soliciting **2-page white papers** from researchers that describe PETs that are ready or nearly ready to deploy, or that have been used in other application areas. We are specifically looking for white papers regarding PETs that would allow:

- **Privacy-preserving sharing and analysis of images and *video*,** such as from surveillance cameras or body camera.
- **Privacy-preserving construction and application of watch lists** based on multiple data sources,
- **Privacy-preserving analysis** of social media.
- **Privacy-preserving techniques for creating, maintaining and linking** anonymous identities and profiles.
- **Multiparty and homomorphic computation** to allow for analysis of datasets held by multiple parties without the need to physically combine datasets.
- **Other PETs** that would address specific operational needs that could be solved with data sharing or analysis that is currently blocked because of concerns about ownership, access, overcollection, or risk associated with the use of operational data.

We have problems that need to be solved *today*, and are looking for engaged researchers who want to help.

While this workshop is focused on novel solutions to today's problems, we are also eager to hear from researchers who have innovative ideas that they feel would benefit from special attention now, even if there are no specific pragmatic uses in the near-term. This event is not about theory, it's about practice. However, some ideas are so novel or important that they require attention now, before there is a problem. Researchers are encouraged to submit ideas that would benefit from discussion with Homeland Security stakeholders; please flag your idea as a discussion topic and it will be reviewed as such. Please indicate if you would like to have the

Accepted Submissions

Authors	Title
Murat Kantarcioglu	Privacy-preserving, error resilient record linkage
Mark Blunk, Paul Bunn, Samuel Dittmer, Steve Lu and Rafail Ostrovsky	Linking Without Leaking: Private Set Intersection
Vishesh Tanwar, Sanjay Madria and Sajal Das	Privacy-Preserving Video Surveillance System over Cloud
Dongqi Fu, Jingrui He, Hanghang Tong and Ross Maciejewski	Privacy-preserving Graph Analytics: Secure Generation and Federated Learning
Rakibul Hasan	Privacy-preserved capturing and processing of images and videos
Jose-Luis Ambite, Srivatsan Ravi and Greg Ver Steeg	Secure Federated Learning
Emily Shen, R. Nicholas Cunningham, J. Parker Diamond, Noah Luther, David A. Wilson and David Bigelow	Rapid Prototyping of Secure Multi-Party Computation Applications
David Archer	Privacy Enhancing Technologies Ready for the Homeland Security Enterprise
Kurt Rohloff	Homomorphic Encryption for Privacy-Protected Linking and Querying of Watchlists
Mohamed Hussein	Twin Finder: Discovering and Assessing the Vulnerability to AI-Generated Twin Identities
Xiao Wang and Jennie Rogers	VaultDB: Facilitating Secure Analytics over Multiple Private Data Sources
Chen Chen, Mubarak Shah and Ishan Dave	Self-supervised Deep Learning for Privacy-preserving Video Analytics

https://pets4hse.org/accepted-white-papers.html

Reflections on PETS4HSE

- There is an extraordinarily broad range of PPDSA technologies.
  - Researchers focused on secure multiparty computation, homomorphic encryption, federated learning.
  - These might all be useful for the DHS components, but components could benefit from simple tech too.

- This is one of the *first efforts* to bring together researchers and potential users of the technology.
  - What are the structural barriers to researchers seeking out potential users?
  - Mechanisms exist for academia to partner with government. Why aren't they being used?

Discussion Topics

- How to bring privacy into the forefront of system design?

- How to bring privacy into the forefront of procurement?
  - Can we specify functionality without a specific approach or vendor?
  - How do we validate solutions?
  - How do we avoid vendor lock-in?

- Each company has its own technology and as agencies procure different technologies, how do we be more forward leaning with respect to privacy?

- There is a general lack of awareness about PETs!
  - How can awareness be improved within a Department? USG? Industry?
  - Is there a role for Centers of Excellence or public-private partnerships?

After PETS4HSE

- CAOE worked with DHS to develop a request for proposals under its center funding mechanism
- Call was developed in August of 2022 and 8 proposals were submitted
- Proposals undergo a two-phase review process
  - External scientific merit review
  - DHS components then review proposals that pass merit review for relevancy to the mission area

Results from the RFP

- Four proposals were selected, and awards announced on January 27, 2023
  - Cerberus: Guarding Sensitive Data with Trigeneous Secure Computations
  - Privacy-Preserving Analytics for Non-IID Data
  - A Federated Query Optimizer for Privacy-Preserving Analytics and Machine Learning

CAOE Seminar Series

- Challenges and Opportunities for Privacy Enhancing Technologies in the Homeland Security Enterprise
  - Nov. 10, 2022 - Dr. Rafail Ostrovsky (UCLA)
  - Dec. 8, 2022 – Dr. L. Jean Camp (University of Indiana)
  - Jan. 12, 2023 – Dr. Steve Lu (Stealth Software Technologies, Inc.)
  - Feb. 9, 2023 – Dr. Alisa Frik (UC Berkeley)
  - Mar. 9, 2023 – Dr. Laura Brandimarte (University of Arizona)
  - Apr. 13, 2023 – TBD
  - May 18, 2023 – Dr. Lorrie Faith Cranor (Carnegie Melon)

Grant Acknowledgment

- The material in this presentation is based upon work supported by the U.S. Department of Homeland Security under Grant Award Number, 17STQAC00001-03-03."

- Disclaimer. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

Acknowledgment

- The PETS4HSE Workshop could not have happened without the immense support of two individuals:

Simson Garfinkel

Lynn Parker Dupree
Chief Privacy Office
Department of Homeland Security

Backup Slides: The Use Cases

- All of the following material can be found at pets4hse.org

Component: Office of the Under Secretary for Management, Office of the Chief Human Capital Officer (OCHCO)

- **Use Case: Identifiable Human Resources (HR) Reports**
- **Description of Need:** OCHCO frequently needs to transmit identifiable Human Resources (HR) reports to DHS Components and other customers for use in HR processing and other HR-related matters. These reports are based on existing data sets housed in OCHCO's HR data warehouse. Currently, OCHCO needs to manually remove certain sensitive data elements such as social security numbers (SSN) from these existing data sets.
- **Proposed Solution:** OCHCO is interested in any technology that could automatically identify and mask these sensitive data elements for this type of reporting.

Component: Office of the Under Secretary for Management, Office of the Chief Human Capital Officer (OCHCO)

- **Use Case: Human Resources Analytics Anonymization**
- **Description of Need:** OCHCO regularly sends reports to DHS leadership and other DHS customers to support HR analytics. The analytics reporting should not be linkable to individual personnel.
- **Proposed Solution:** OCHCO is interested in any technology that could automatically anonymize this reporting used to generate HR analytics.

- **Use Case: Social Security Number Anonymization**
- **Description of Need:** DHS policy requires that SSN only be used when there are technical, legal, or managerial barriers to using an alternate unique Identifier (ID). The Office of Personnel Management, National Finance Center, and certain other agencies OCHCO needs to share data within order to facilitate HR processing still rely on SSN as a unique ID; and therefore, OCHCO must continue to transmit SSN to them on a regular basis.
- **Proposed Solution:** OCHCO is interested in any technology that could better protect SSN when communicating with these agencies.

Component: Cybersecurity & Infrastructure Security Agency (CISA)

- **Use Case:** Privacy Enhanced Information Sharing Through Synthetic Data
- **Description of Need:** CISA needs to share data with vendors and external data scientists for analysis of threat and intelligence gaps. CISA needs to coordinate with industry professionals and the academic research community regarding emerging threats and dynamics in tradecraft. The sharing of data is limited due to privacy, statutory (e.g., Personally Identifiable Information (PII), Protected Critical Infrastructure Information (PCII)), and classification considerations.
- The result would be to create blended data using Generative Artificial Intelligence (GAI). This would entail using synthetic data with real training data (e.g., endpoint data collected within departments and agencies) to develop a sharable dataset for use across and outside CISA. This real training data is modeled to assist in the generation of the synthetic data producing information that is privacy enhanced and sharable. Additionally, this sharable data could be further protected through the use of homomorphic encryption (HE).
- **How it is done today:** CISA is unable to share data with external vendors and scientists due to privacy, statutory, and classification restrictions.

Component: U.S. Immigration and Customs Enforcement (ICE) Office of Information Governance & Privacy

- **Use Case:** Artificial Intelligence (AI) compatible synthetic data production for training law enforcement tools
- **Description of Need:** ICE Privacy is interested in learning about technologies that can create sophisticated synthetic datasets from large quantities of disparate data that could be used to effectively train AI tools while protecting individual privacy. As an added functionality, the technology could report demographic deviations across a dataset to ensure any dataset used for training a tool is appropriately representative and minimizes potential bias. Agencies with law enforcement missions could benefit from robust, timely, and reliable analytics using AI that is optimally trained with representative datasets. System developers could benefit from reduced development and compliance costs. There could also be benefits to members of the public from minimizing the use of individual data and minimizing the potential for bias.

Component: Office of Biometric Identity Management (OBIM)

- **Use Case:** Enhancing privacy of biometric matching
- **Description of Need:** A fully homomorphic encryption systems to allow for biometric matching within an encrypted domain.
- **How it is done today:** OBIM is responsible for the storage, matching, and sharing of biometric information collected during Component missions. The data is transmitted through encrypted pathways, decrypted for matching and then re-encrypted for transmission back to the Components. During decryption the data become more vulnerable to attacks; if this stage could be bypassed and matching performed while the data are encrypted, the risk of attack would be minimized

Component: Office of Biometric Identity Management (OBIM)

- **Use Case:** Enhancing cybersecurity of biometric data transmissions to support biometric and identity research, using trusted data environments.
- **Description of Need:** Biometric and identity research, testing, and evaluation could be dramatically improved if representative data was available to researchers. Currently, biometric and identity research is conducted using small, incomplete, and unlabeled data sets that do not accurately represent the DHS Component real world scenarios. This results in a potential gap in performance of those solutions being evaluated for operational use.
- **How it is done today:** DHS biometric and identity data is **not** currently shared with research organizations.

Component: Office of Biometric Identity Management (OBIM)

- **Use Case:** Enhancing cybersecurity of biometric data transmissions
- **Description of Need:** Secure data transmission between Component collection of biometric and biographic information and the OBIM biometric repository of record to protect against advanced cyber threats. With the advancement of quantum computers, current cybersecurity protocols will very likely be at risk.
- **How it is done today:** Biometric data is currently submitted in an RSA-encrypted form.

Component: U.S Citizenship and Immigration Services (USCIS), Office of Citizenship and Applicant Information Services (CAIS)

- **Use Case:** Two-factor authentication for the online account.
- **Description of Need:** When someone signs up for or access to their myUSCIS account, they must use two-factor authentication. Two-factor authentication (2FA) works by adding an additional layer of security to the online account. It requires an additional login credential beyond the username and password. This typically involves the system emailing or texting a one-time passcode to the user's registered email or cell phone number, which must be entered to gain access to the online account.
- **Deleting draft data in the online account.** myUSCIS deletes draft data after 30 days of no activity. This was a request from the USCIS Office of Privacy to increase privacy and security since the data is no longer available.

Component: U.S Citizenship and Immigration Services (USCIS), Office of Legislative Affairs

- **Use Case:** Constituent identity verification
- **Description of Need:** USCIS would like to have a reliable way to verify the identity of individuals seeking assistance from Congress in obtaining information or resolving problems with their cases. Current practice is to require a signed privacy release (with a hand-written signature comparable to other signatures on file with USCIS).
- **Use Case:** Waiving statutory nondisclosure restrictions
- **Description of Need:** CIS needs a way for applicants and petitioners to waive statutory nondisclosure restrictions (such as in asylum, refugee, Temporary Protected Status, or abuse cases) without signaling that the application or petition is subject to those restrictions.

Component: U.S Citizenship and Immigration Services (USCIS), Field Operations Directorate

- **Use Case:** Employment Based Fifth Preference (EB-5) Immigration Benefit Adjudications
- **Description of Need:** The eligibility requirements for an employment-based fifth preference visa requires a substantial amount of information to determine eligibility. The submitted evidence is sensitive information that is far beyond the personally identifiable information (PII) found in USCIS adjudications. For example, initial evidence for Form I-526, Immigrant Petition by Alien Entrepreneur, filings routinely include sensitive financial information pertaining to U.S. businesses, foreign nationals, and U.S. citizens. Generally, a Form I-526 submission package includes several hundred pages consisting of personal and business-related bank statements, tax returns, and organizational documents and business plans related to a petitioner's capital investment. Similarly, this is again the circumstance when the Form I-526 petitioner submits a Form I-829, Petition by Investor to Remove Conditions on Permanent Resident Status, within a two-year period seeking eligibility to remove his or her conditions.
- Operationally, this makes identifying and exercising exemptions to the Freedom of Information Act very difficult. USCIS is obliged to identify and redact sensitive information, including hundreds and hundreds of pages of PII. It would be extremely helpful if there were an automated way to rapidly identify and redact PII so we can more quickly respond to FOIA and other inquiries.

Component: U.S Citizenship and Immigration Services (USCIS), Office of Performance and Quality

- **Use Case:** Consistent Marking and Handling of Data Across Sharing Mediums
- **Description of Need: F**or a majority of data sharing activities, internal and external to DHS, there is a need to move sensitive/confidential data across mediums in a secure way. This is a challenge for legacy systems that do not have built-in privacy-centric requirements designed to parse through data types with the required level of granularity and mark it (e.g., protected status) appropriately for handling. The other part of the equation, even if the data were marked and passed securely, is the need for the partner system to receive the marked data (i.e., standardized with corresponding technical schema) and have the markings move with the data downstream to authorized users. Additionally, if the data were to change (e.g., legal, policy, privacy reasons), many receiving systems do not have the ability to refresh data in near-real time and have it effectively cascade to other authorized systems/users.
- **How it is done today:** The marking of sensitive/confidential data is not done effectively and consistently. It takes a tremendous amount of time, resources, and human labor. Again, even if the sending partner were to do the required marking and technical implementation, the receiving partner would still have to be in technical alignment and interpret what it was sent for operationalization. Data sharing agreements (DSAs) and other supporting business documentation used today call out the types of data shared, along with the appropriate safeguarding provisions. These provisions are memorialized in procedural guidance (at the receiving end) to ensure data are handled appropriately. DHS relies on its data partners to abide by the terms and conditions of the agreements. Audits are performed on occasion to ensure compliance.

Component: U.S Citizenship and Immigration Services (USCIS), Office of Policy & Strategy

- **Use Case:** Enhancing Customer Service Options for Victims of Domestic Violence and Intimate Partner Violence, Human Trafficking and other Crimes.
- **Short Description of Need:** USCIS seeks to enhance ways in which individuals protected under 8 U.S.C. § 1367 can safely and efficiently interact with existing customer service channels. The statutory protections under 8 U.S.C. § 1367 apply to VAWA self-petitioners1, petitioners for U Nonimmigrant Status, and applicants for T Nonimmigrant Status and their family members ("protected individuals.")
- Background on 8 U.S.C. § 1367 Protections
- Applicants and recipients of immigration benefits covered by 8 U.S.C. § 1367 are entitled to special protections with regard to privacy and confidentiality. This statute prohibits the unauthorized disclosure of information about petitioners and applicants for, and beneficiaries of VAWA, T, and U-related benefit requests to anyone other than an officer or employee of DHS, the Department of Justice (DOJ), or the Department of State (DOS) for a legitimate agency purpose.
- **How it is done today:** At this time, protected individuals do not have access to most USCIS customer service options, such as the USCIS Contact Center, due to current policy regarding identity verification.2 As these protections generally continue through the protected individual's entire immigration journey, the current process for obtaining information about their cases is burdensome for both the protected individual and the agency. The process is especially burdensome for unpresented individuals. If protected individuals are not represented by an attorney or accredited representative, their only option for customer service is either going in person to a USCIS District Office or sending written correspondence to USCIS via U.S. mail