CAOE | Center for Accelerating Operational Efficiency

A Department of Homeland Security Center of Excellence

Workshop on Privacy Enhancing Technologies for the Homeland Security Enterprise

June 21, 2022

Arizona State University

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

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 document collects 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).

Disclaimer

This event is hosted by the Center for Accelerating Operational Efficiency, a Department of Homeland Security Center of Excellence in collaboration with the Department of Homeland Security Privacy Office.

This material is 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 should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security

All use cases provided in this document are approved for public release. These are being distributed as examples of what the Department of Homeland Security would like to do with Privacy Enhancing Technology; however, these examples do not represent an official notice of any Department of Homeland Security plans or any programs that are underway, and they do not constitute a statement of need by the Department of Homeland Security.

For more information, contact CAOE at CAOE@ASU.edu or visit https://caoe.asu.edu/.

Table of Contents

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.

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.

Customers:
- CISA/Cybersecurity Division (CSD)
- CISA/National Risk Management Center (NRMC)

The American people would benefit from this sharing of information and increase ability to collaborate by enhancing:
- The security posture of the Federal Government
- State, Local, Tribal, and Territorial Government
- Predictive Analytics regarding Cyber Threats
- Security of Critical Infrastructure

Risks: The dynamics of detection and analysis of threats will continue to lag behind the adversarial tradecraft contributing to persistent attacks and breaches.
If this capability does not perform reliably, CISA's ability to collaborate with external vendors and data scientists will remain limited.

Time Horizon: This has been long standing need at CISA. This capability is necessary immediately and would be an ongoing need for CISA's collaboration internally and externally.

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: Big data analytics and AI can assist law enforcement agencies in investigations and criminal intelligence by curating and chaining together seemingly disparate raw datasets and performing advanced analytics across multiple datasets. These tools enable personnel to accomplish tasks too large or complex for traditional systems and more effectively prepare information for manual analysis and decision-making.

AI allows users to better recognize patterns in data and enhances a tool's effectiveness over time. In order for AI to be optimally effective, it must be trained and tested on large quantities of data. Law enforcement investigations involve an incredible volume and scope of data that is often unique in nature from one case to the next. This can present a challenge in training AI to, for example, identify connections between criminal acts and enterprises that are not readily apparent without using real data. The use of real data in this way may conflict with established privacy principles involving use limitation and data minimization and may increase costs as a result of requirements for securing personally identifiable and other protected information. It also could have the unintended consequence of training bias into an AI tool if not managed effectively. 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.

Customer: ICE Privacy enables the ICE mission by overseeing and advising on the agency's compliance with federal privacy laws, regulations, policies, and principles.

Risks: As AI becomes a mainstay in government databases, the need to train those algorithms will become commonplace. Without a sufficient answer as to how to produce adequate synthetic datasets for use in training, it is foreseeable that real data will be used, which can present risks to individual privacy, civil liberties, and may strain government resources.

Time Horizon: ICE Privacy is interested in considering use cases for this functionality in the near-term.

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

Customer: Advanced security of data during transmission and the matching process would safeguard the biometric and biographic information collected on foreign and domestic individuals encountered by DHS Components. Implementation would increase public perception and trust of the handling and protection of biometric and biographic information by DHS. This assurance would increase the acceptance of use of biometrics for streamlined processing during Component's mission execution.

Risks: The decryption of biometric data for matching creates vulnerabilities to cybersecurity attacks, an incident of this kind would decrease public perception and trust of the handling and protection of biometric and biographic information by DHS.

Time Horizon: This capability would be necessary in the near (1-3 year) and would be an enduring need.

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.

Customer: The Information Marketplace for Policy and Analysis of Cyber-risk & Trust (IMPACT) program was developed to support the global cyber risk research community by coordinating and developing real world data and information sharing capabilities – tools, models, and methodologies. This framework also implements modern security controls over data sharing.

The IMPACT data sharing model could be adopted and adapted for sharing of operational biometric and identity data to improve research, test, and evaluation. IMPACT, developed by DHS S&T for the sharing of cybersecurity data, could be adapted for a prototype to share biometric and identity data with National Science Foundation (NSF) / University Cooperation Research Center (UARC) Center for Identification Technology Research (CITeR) research universities.

Risks: The design of a prototype using the IMPACT data sharing framework would only use Privacy approved data, thereby reducing the risk. Specifically, the OBIM curated MEDS III data set of deceased identities would provide a usable data set. (Note: MEDS III is still in the curation phase).

Time Horizon: This capability would be necessary in the near term (18 months) and would be an enduring need.

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.

Customer: Advanced security of data transmissions would safeguard the biometric and biographic information collected on foreign and domestic individuals encountered by DHS Components. Implementation would increase public perception and trust of the handling and protection of biometric and biographic information by DHS. This assurance would increase the acceptance of the use of biometrics for streamlined processing during Component's mission execution.

Risks: Quantum computers have the potential to provide the means to decrypt current RSA encryption protocols through faster factorization of large integers than traditional computing technologies. While it may take decades for quantum computers to reach an operational level to be able to implement a cybersecurity attack capable of toppling current cybersecurity protections, we need to consider a counter to a potential future where these encryption-cracking capabilities are available. If data transmission security does not keep pace with emerging cyber threats public confidence in the handling of biometric and biographic information by DHS would rapidly decrease.

Time Horizon: This capability would be necessary in the near to mid-term (1-5 year) and would be an enduring need.

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.

How it is Done Today: Currently, the review of this information is performed manually based on scans of hardcopy evidence having a volume of 500- 1000+ pages. Neither the scanning nor the review of such hardcopy evidence is ideal. It may be easier to help rapidly identify and redact PII if all USCIS forms related to EB-5 adjudications were submitted electronically instead of via paper.

Customer: Faster identification of PII would expedite USCIS responds to Congressional inquiries, FOIA requests, and media inquiries in addition to achieving greater transparency of routine FOIA requested information that would benefit the U.S. public. Internal USCIS stakeholders, the Office of Chief Counsel, and the Immigrant Investor Program Office, would also benefit in the sharing of electronic documents related to eligibility concerns, which are not easily transported between the two physical offices, and require additional scanning of hardcopy evidence to advance the eligibility discussion.

Risks: USCIS form types for EB5 immigration benefits require a substantial amount of PII to establish eligibility. Their hardcopy filings make them particularly attractive targets to any number of monetary schemes that could maliciously use their PII and financial information.

Time Horizon: USCIS receives new EB5 FOIAs weekly that contain multiple parts and require hardcopy review and/or scanning to respond to the requestor. This manual labor is not sustainable, and we anticipate an increase in FOIA requests since the new legislation passed on March 15, 2022, will establish new processes and policies that the private attorneys will seek to benefit their USCIS EB-5 form submissions.

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: For 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.

Customer: The entire Homeland Security Enterprise would benefit. We (DHS) are required by law and/or policy to protect data, especially when it comes to protected status individuals (e.g., 8 USC 1367).

Risks: Improper safeguarding, retention, and use of personal data. DHS would be in violation of law and/or policy and liable and, in some circumstances, civil penalties could be levied on federal personnel.

Time Horizon: Now – permanent sustainment.

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

<u>Use Case</u>: 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-petitioners[1], 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.

Customer: Victims of domestic and intimate partner violence, human trafficking, and other crimes with pending and approved immigration benefits protected by 8 U.S.C. 1367. This new capability will enable protected individuals to get updates on their case, to request disability accommodation and expedite requests, and generally improve USCIS's customer service to vulnerable populations. USCIS components will benefit by reducing processing delays, moving away from solely a paper-based system, and modernizing customer service channels. As a result,

[1] USCIS also extends the provisions of 8 U.S.C. 1367 to abused spouses of certain persons applying for employment authorization under INA 106.

[2] USCIS Policy Manual Guidance instructs that USCIS cannot release any information relating to a protected person until the identity of the requestor of information is verified and that person's authorization to know or receive the protected information is verified. Such identity and eligibility verification must be done before responding to any inquiry, expedite request, referral, or other correspondence. Upon identity verification, USCIS can provide protected information directly to the protected person or his or her representative authorized to receive 1367-protected information. At this time, the Contact Center does not have method to incorporate enhanced identity verification methods necessary for Section 1367-protected individuals.

this would increase operational efficiencies, support allocating resources to other top priorities, and advance DHS's victim-centered approach.

Risks: The risks of not developing this capability will mean that USCIS customer service channels for protected individuals will continue to be extremely limited and increase barriers for vulnerable applicants. However, any new capability must ensure that multiple options for identity verification are available. DHS's adoption of a victim-centered approach means that a protected individual must be able to choose which identify verification option they are comfortable using. For example, survivors of trauma may be reluctant to use certain forms of technology like facial recognition.[3]

What are the risks of not developing this capability?
In February 2021, the White House issued an Executive Order on Restoring Faith in Our Legal Immigration Systems and Strengthening Integration and Inclusion Efforts for New Americans, which instructs DHS to identify barriers that impede access to immigration benefits and to make recommendations on how to remove these barriers. One significant barrier that stakeholders identified is the lack of access to USCIS customer service tools for 8 USC § 1367-protected individuals. If this capability is not developed, survivors will continue to have limited access to customer service tools and will continue to encounter hurdles when seeking case updates. Furthermore, 1367 protected individuals may need to waive their protections solely to receive information about their case(s).[4]

In addition, DHS issued an agency-wide policy in October 2021 to adopt victim-centered approaches into all policies, programs, and activities governing DHS interactions with victims of crime. Continuing with the current restrictive customer service avenues for protected individuals does not comport with this directive.

What are the risks if the capability does not perform reliably?
An unauthorized disclosure of information which relates to a protected person can have significant consequences. The safety of victims of domestic violence and intimate partner violence, victims of trafficking, and victims of crime can be put at risk, as can their family members, if information is provided to a person who is not authorized.

What are the risks to the program if the "privacy problem" is not solved?
The risks of not developing this capability will also continue to place the agency at risk of actual or potential violations of the confidentiality provisions of 8 U.S.C. 1367. Anyone who willfully uses, publishes, or permits any information pertaining to such victims to be disclosed in violation of the above-referenced confidentiality provisions may face disciplinary action and be subject to a civil penalty of up to $5,000 for each violation.

Internally, by not enhancing identity verification capabilities for protected individuals, USCIS will continue to use antiquated communication methods by requiring them to contact us by U.S.

---

[3]There is also the concern about the disparities in the accuracy of facial recognition systems based on gender and racial biases.
[4] 8 U.S.C. 1367(b)(4).

mail. This paper-based system increases security vulnerabilities, causes unmet mission needs, creates staffing issues, and increases overall costs.

Time Horizon: We require this capability as soon as it is operationally feasible, and it would be needed indefinitely.