NEW AMERICA

March 2025

How to Protect Government Data with Privacy-Enhancing Technology

Sydney Saubestre

Open Technology Institute

Last edited on March 25, 2025 at 1:00pm

Acknowledgments

This report benefited greatly from the expertise and support of several individuals. I am most grateful to Joel Yong for his thought partnership, outstanding research, and drafting contributions. I also appreciate the background research conducted by Emma Seeman and the insightful feedback and edits from Prem Trivedi and Nat Meysenburg. I would like to thank the Gates Foundation for their generous support of this work.

Editorial disclosure: The views expressed in this report are solely those of the author and do not reflect the views of New America, its staff, fellows, funders, or board of directors.

About the Author

Sydney Saubestre is a senior policy analyst at New America's Open Technology Institute focusing on data usage and privacy issues.

About New America

We are dedicated to renewing the promise of America by continuing the quest to realize our nation's highest ideals, honestly confronting the challenges caused by rapid technological and social change, and seizing the opportunities those changes create.

About Open Technology Institute

OTI works at the intersection of technology and policy to ensure that every community has equitable access to digital technology and its benefits. We promote universal access to communications technologies that are both open and secure, using a multidisciplinary approach that brings together advocates, researchers, organizers, and innovators.

Contents

Purpose of the Report

Data is vital to the functioning of government, helping agencies make informed decisions, improve public services, and drive policy innovation. However, data sharing comes with significant privacy risks, including unauthorized access, data breaches, re-identification of individuals from data associated with them, and potential misuse of sensitive information.

Privacy-enhancing technologies (PETs) offer solutions to address many of these risks by enabling secure and privacy-preserving data use. These technologies allow organizations—whether they are government agencies, private corporations, or nonprofit service providers—to share and analyze data while ensuring compliance with privacy regulations and maintaining public trust. However, PET adoption presents challenges, including technical complexity and resource constraints.

While PETs have a broad range of utility, this report specifically outlines the role of PETs in government operations. It provides an overview of available PET options and serves as a practical guide for data professionals in local, state, and federal government to determine the most suitable PETs for their specific needs. As PETs continue to evolve, governments must invest in capacity-building and policy frameworks to maximize their benefits for public sector innovation and responsible data use.

Key Goals

This report aims to:

- Explain the significance of PETs in protecting sensitive data while enabling responsible data sharing;
- Provide an accessible introduction to different types of PETs and their potential applications; and
- Offer practical guidance to help government data professionals, policymakers, and privacy engineers select PETs based on specific data-sharing requirements.

Introduction to Privacy-Enhancing Technologies (PETs)

Privacy Matters

Governments should and do rely on data to allocate resources, assess policy impact, and improve public services. From tracking public health trends to optimizing educational attainment, data-driven decision-making enables more efficient and informed governance. However, the growing scale of data collection and sharing also heightens privacy risks—particularly as more personal information is aggregated and stored across public and private systems[1]. Consumer data, including financial records, location history, and online activity, is increasingly intertwined with government-held information, creating broader exposure to breaches, misuse, and re-identification[2].

Without proper safeguards, sensitive personal data can be exploited, leading to real-world harm. For example, the 2015 U.S. Office of Personnel Management (OPM) breach compromised the records of millions of federal employees, exposing Social Security numbers, personnel records, and even extensive information about employees' friends and relatives provided as part of applications for security clearance.[3] Similar risks exist in the private sector, where high-profile breaches have exposed data from credit card details to genetic information.[4] Such incidents erode public trust and illustrate the dangers of concentrating vast amounts of sensitive information in centralized, highly accessible systems.[5]

> **"When data is consolidated across agencies and sectors, insider threats, unauthorized access, and political misuse become greater risks."**

Large, integrated databases offer efficiency and convenience, but they also create single points of failure.[6] When data is consolidated across agencies and sectors, insider threats, unauthorized access, and political misuse become greater risks.[7] Recent high-profile cases have underscored how individuals with privileged access—whether through government positions or corporate control—can exploit these databases in ways that put people and systems at risk.[8]

Privacy-enhancing technologies (PETs) can provide technical solutions to mitigate these risks at every stage of the data lifecycle—collection, processing, use, storage, and sharing.[9] By decentralizing access, limiting exposure of sensitive data, and enabling secure analysis, PETs help balance data utility with privacy protection. As governments and companies continue to modernize their digital infrastructure, PETs must be a core component of responsible data governance, ensuring that data utility does not come at the expense of individual privacy and security.[10]

What Are PETs?

Privacy-enhancing technologies (PETs) refer to a diverse set of tools and methodologies designed to ensure that data can be used, analyzed, and shared without compromising the privacy of the people whose data has been collected.[11] These technologies mitigate risks by applying cryptographic techniques, anonymization methods, and secure computation processes.[12] PETs are particularly critical for safeguarding the privacy of individuals whose data is held by governments, businesses, and research institutions—whether it's citizens interacting with public services, consumers generating digital footprints, or patients contributing to medical research. By reducing the risk of exposure, PETs allow organizations to extract insights from data while maintaining trust and compliance with privacy protections.[13]

Why Do PETs Matter?

Government data breaches can have far-reaching consequences, exposing sensitive personal information and undermining confidence in public institutions.[14] With vast amounts of data—including Social Security numbers, health records, and immigration details—at risk, breaches can lead to identity theft, financial fraud, or personal safety risks.[15] Additionally, compromised government databases can be exploited for political or foreign interference, weakening national security.[16] Public distrust in data security can reduce participation in critical government programs, limiting the effectiveness of services and policy initiatives.[17] By integrating PETs, governments can minimize these risks, enhancing both the security of public data systems and the trust of the people they serve.

As data use grows across sectors, the need for robust privacy safeguards becomes more urgent. PETs contribute to safeguards by:

- **Minimizing trust requirements**: Using technical measures to enforce restrictions on data access and processing, instead of relying solely on policies and contracts to protect privacy.
- **Enabling secure data collaboration**: Facilitating secure data sharing across organizations without exposing raw datasets, preserving confidentiality and privacy.
- **Supporting ethical data use and legal compliance**: Helping governments and organizations adhere to laws that mandate de-identification, access restrictions, and limitations on purpose and use, safeguarding ethical data-handling practices.[18]

One of the key advantages of PETs is their ability to facilitate public-interest data sharing. Governments often need to share information across agencies, with research institutions, or with private-sector partners to drive innovation and improve public services.[19] PETs can enable this type of collaboration while limiting disclosure and without compromising individuals' privacy.[20]

Government agencies handle vast amounts of personal, financial, and health-related data. Ensuring that this data is shared and processed securely is critical for:

- **Cross-agency collaboration**: Enabling government entities to securely analyze and link data across departments while maintaining privacy protections in compliance with legal and ethical standards.
- **Public trust and transparency**: Encouraging citizen support for data-driven initiatives by ensuring their privacy is safeguarded through secure data practices.
- **Privacy policy and regulatory compliance**: Helping agencies meet legal and regulatory requirements by ensuring personal data is processed in accordance with privacy laws, including data minimization, consent management, and data retention policies.[21]

In practice, PETs can support a wide range of use cases, such as:

- **Health data research**: Enabling secure data-sharing frameworks for medical research while protecting patient confidentiality.
- **Fraud detection and prevention**: Analyzing financial transactions securely without exposing personal financial details.
- **Census and demographic analysis**: Aggregating census data to analyze trends without accessing personally identifiable information.

By embedding PETs into government data strategies, agencies can unlock the value of data while protecting individuals' rights and upholding ethical standards.

Types of PETs and Plain-Language Explanations: A Glossary

Privacy-enhancing technologies (PETs) are highly technical. This makes it challenging for non-experts to understand their functions, in turn making it harder to fully understand their benefits. To bridge this gap, this report includes a glossary that provides clear, non-technical explanations of key PETs.

Additionally, after the glossary, there is a table to help practitioners navigate which PET may be best suited for different needs (see Table 1). The goal is to ensure that policymakers, data managers, and engineers can make informed decisions about which PETs to use in different scenarios.

Methodology for PET Selection

The PETs covered in this report were chosen based on their relevance to government use cases, their ability to protect privacy while maintaining data utility, and their compliance with regulatory requirements. The selection process considered:

- **Strength of privacy protections**: The extent to which each PET minimizes exposure of sensitive data.
- **Suitability for government applications**: Whether the PET can be effectively integrated into government workflows and data-sharing initiatives.
- **Regulatory and ethical considerations**: How well each PET aligns with privacy laws and ethical data-use principles.

By understanding and implementing PETs, government agencies can ensure that data-driven initiatives remain both effective and privacy-compliant. The following sections provide an in-depth look at different PETs, their applications, and how they can be leveraged for secure data sharing.

Different Types of PETs

De-Identification

What it is: De-identification is the process of removing or altering personally identifiable information from datasets so that individuals cannot be easily identified. This is essential when sharing data for analysis, ensuring that personal details such as names, addresses, or contact information are not exposed. Redaction is a specific form of de-identification, where sensitive information in documents (e.g., health records or contracts) is blacked out or removed. This helps prevent privacy breaches while still allowing the relevant data to be used for research or policymaking.

How it works: De-identification can involve several techniques:

- **Pseudonymization**: Replacing direct identifiers, like a name or Social Security number, with a pseudonym or code.
- **Suppression**: Entirely removing sensitive identifiers that could lead to re-identification.
- **Generalization**: Replacing exact data points with broader categories. For example, replacing a specific age with an age range like "30–40."
- **Redaction**: Systematically identifying and blacking out sensitive content, such as health information, financial details, or personal identifiers, so that unauthorized users cannot view or misuse the data.

These techniques preserve the utility of the data for analysis while ensuring individuals' privacy.

Plain metaphor example: Imagine a basketball game where players wear jerseys with numbers but no names. You can analyze their performance, but you don't know who they are unless you have the key to the roster.

Use case: De-identification of student data is essential for compliance with the U.S. Family Education Rights and Privacy Act (FERPA), as it removes or obscures personally identifiable information to minimize the risk of unintended disclosure.[22] When properly de-identified, a subset of data can be shared without obtaining further consent, allowing for educational research while maintaining confidentiality and ensuring privacy protections.

Differential Privacy

What it is: Differential privacy is a mathematical technique used to protect individual privacy while sharing data for analysis.[23] It ensures that the inclusion or exclusion of any single person's data cannot be determined based on the results of a statistical query or analysis.[24] This is achieved by adding controlled random noise to the data or query responses, making it difficult to identify any specific individual's information while still preserving overall patterns and trends.[25] In simpler terms, differential privacy protects individuals by blending their data into the crowd, allowing researchers and analysts to learn about the group without revealing personal details about anyone.

How it works: Differential privacy works by adding random noise to datasets, making it difficult to link individual records back to a person.[26] It is worth noting that with differential privacy, there is an inverse relationship between privacy and precision (the more noise is added for individual privacy, the less precise the data is).[27]

Plain metaphor example: Think of differential privacy like altering a picture to protect someone's identity. Imagine you have a group photo where everyone's faces are clear. If you want to share that photo without revealing anyone's identity, you blur the faces just enough that no one can recognize individuals but can still tell it's a group photo.

Differential privacy adds this "blur" by introducing small, random changes (noise) to the data. This ensures that no one can tell if a particular person's is present in the photo, but overall trends, like average height or hair color, remain visible and useful. It protects individual privacy while still allowing meaningful patterns to emerge.

Use case: In the 2020 Census, the U.S. Census Bureau implemented differential privacy to enhance the protection of individual respondents' data.[28] This approach involved adding controlled noise to the data before it was released, making it difficult to identify specific individuals while still allowing for meaningful analysis of population trends and demographics.[29]

Encryption

What it is: Encryption is the process of converting data into a secure format that can only be accessed or decrypted by authorized parties. It is a critical component of data security, ensuring that information remains confidential and protected from unauthorized access during storage, transmission, and processing.

Think of it like locking valuables in a safe—only those with the key can access them. Whether the items are being transported (in transit) or stored in the safe (at rest), the lock ensures they remain protected from anyone who shouldn't have access.

Encryption: In Transit

What it is: Encrypting data in transit ensures that information transmitted over a network remains confidential and protected from interception. Transport Layer Security (TLS) is the protocol used to secure communications, commonly seen in HTTPS connections, which prevents attackers from accessing data during transmission.[30]

How it works: When a client (such as your web browser) connects to a server, they negotiate a secure connection by agreeing on encryption methods and exchanging keys. Data is then encrypted before transmission, ensuring that only the intended recipient can decrypt and read it, even if the data is intercepted.[31]

Plain metaphor example: Imagine two spies using a secret language to communicate over a public radio channel. Even if someone listens in, they won't understand the message without knowing the code.

Use case: Online banking websites and e-commerce platforms use TLS encryption to protect data transmitted between users and servers.[32] When you log into your bank account or make a purchase online, TLS ensures that sensitive information—like login credentials or payment details—is encrypted during transmission. Even if an attacker intercepts the data, they won't be able to read it because the data is encrypted before it leaves your device and decrypted only by the intended recipient server. This encryption process is facilitated by certificate authorities, which validate the authenticity of the encryption keys to ensure secure communication.[33]

Encryption: At Rest

What it is: Encrypting data at rest protects data stored on devices, servers, or databases. Before data is saved to disk, it is encrypted using algorithms like the Advanced Encryption Standard (AES). When the data is needed, it is decrypted by authorized systems or users. This method ensures that even if someone gains access to the physical storage, they cannot read the data without the decryption key.[34]

How it works: When data is saved to a storage medium, it is encrypted using a cryptographic algorithm and an encryption key. The data remains encrypted until it is needed, at which point it is decrypted by authorized users or systems with the correct decryption key.

Plain metaphor example: Storing encrypted data is like keeping valuables in a high-security bank vault. Even if a thief gets inside the bank, they can't open the vault without the right combination.

Use case: Services like Dropbox and Google Drive, used across government settings, employ at-rest encryption to protect files stored on their servers.[35] This ensures that if a hacker gains access to the storage, they cannot read the contents of the files without the decryption keys.

Homomorphic Encryption

What it is: Homomorphic encryption allows computations to be performed on encrypted data without needing to decrypt it first. This technique ensures that sensitive data remains encrypted throughout the entire process, even during analysis, making it a powerful tool for privacy-preserving computations.

How it works: With homomorphic encryption, data is encrypted using a special algorithm that allows mathematical operations, like addition or multiplication, to be performed directly on the encrypted data.[36] The result of these operations is still encrypted, and only after the computations are complete can the encrypted result be decrypted to reveal the final output. This means that sensitive data, such as personal information or financial details, can be processed and analyzed by third parties without ever exposing the original data. It is especially useful in cloud computing, where data privacy is crucial, as it allows for secure data sharing and processing without compromising security.

Plain metaphor example: Homomorphic encryption is like sending a locked box of ingredients to a chef to cook in a private kitchen. The chef uses those ingredients to prepare a meal and then locks the finished product inside the box. Only the person with the key can unlock the box and enjoy the final result, ensuring no one else can steal the ingredients or alter the meal.

Use case: IBM has used homomorphic encryption for cloud computing solutions, enabling clients to analyze sensitive data, such as medical or financial information, while keeping it encrypted throughout the process.[37] Similarly, governments can use homomorphic encryption to securely outsource data processing and storage to cloud providers while maintaining control over the data and ensuring its confidentiality, which is critical for large-scale processing of sensitive data.

Federated Data Science

Federated data science is a collaborative approach to data analysis where multiple parties work together to analyze decentralized data without transferring or sharing sensitive information. By using techniques like federated learning and federated analytics, organizations can derive insights from distributed datasets while ensuring privacy and compliance with data protection regulations.

Federated Learning

What it is: Federated learning is a machine learning technique that allows multiple devices or parties to collaboratively train a model without sharing their raw data.[38] Instead of sending data to a central server, each participant trains the model locally on their own device and only shares the model updates.

How it works: In federated learning, a global machine learning model is built collaboratively by many devices or entities (e.g., smartphones, hospitals, or organizations). Each participant trains the model on their local data, then sends only the model parameters (such as weights or gradients) to a central server, rather than the raw data itself. The server aggregates the updates from all participants to improve the model. This process is repeated across multiple rounds, allowing the model to learn from a diverse set of data sources without any party needing to expose their private data.

Plain metaphor example: Federated learning is like a group of chefs each perfecting their own recipe in separate kitchens. Each chef works with their own set of ingredients, refining their dish based on what they've learned. After a round of cooking, the chefs share what improvements they've made to their recipes, but none of them reveal their specific ingredients or the exact methods they used. The central restaurant combines these improvements to create the best dish possible, without needing to know what's inside each chef's kitchen.

Use case: As governments make strides to train their own artificial intelligence (AI) models, federated learning provides an opportunity to train on decentralized data rather than with general data that may put personally identifiable information (PII) at risk.[39] The Centers for Disease Control and Prevention and National Institutes of Health could use federated learning to train AI models on COVID-19 patient data from multiple hospitals, without hospitals sharing raw patient data.[40]

Federated Analytics

What it is: Federated analytics is similar to federated learning, but instead of training machine learning models, it focuses on performing data analysis collaboratively while keeping the data decentralized and private.[41]

How it works: In federated analytics, data remains on the local devices or servers, and only aggregated insights or analysis results are shared. For example, rather than sending raw data to a central server, each participant can perform calculations on their own data and then share only the aggregated results, such as averages or statistical summaries. This ensures that the original, detailed data is never exposed or transmitted, but collective insights can still be derived from all participants' datasets.

Plain metaphor example: Federated analytics is like a line of cashiers closing out their registers at a supermarket for the night. Each cashier counts their bills and reports their respective final amounts to the supermarket's log, but nobody knows how many 1s, 5s, 10s, 20s, 50s, or 100s there were in registers besides their own. Each cash register remains private, but the supermarket can still create a complete picture by combining their totals.

Use case: Google reported in 2020 that it used federated analytics to support the Now Playing feature on Google's Pixel phones. This approach enhances privacy by ensuring that song recognition happens locally on the device, without transmitting raw or processed audio data. Because each phone received the same database, the feature ensured privacy while maintaining functionality.[42]

Generalization

What it is: Generalization is a privacy technique where specific, granular data is replaced with broader categories or ranges to protect individuals' identities while maintaining useful data for analysis. For example, rather than recording an individual's exact salary, the data could be generalized into salary ranges such as "$40,000–$60,000." Generalization helps ensure that specific individuals cannot be identified based on the data, even when combined with other publicly available information.

How it works: Generalization involves transforming detailed data into higher-level categories or ranges. For example, instead of recording an exact age, an age range (such as "30s") might be used to avoid identifying an individual. Similarly, a person's exact geographic location might be replaced with a broader region or area. This approach reduces the risk of re-identification by ensuring that data points are not unique to a person. The challenge is balancing the level of detail retained for analysis with the level of privacy provided to individuals.

Plain metaphor example: Instead of displaying an exact street location, generalization zooms out so you only see the general area, like a heatmap that highlights trends without revealing individual lots.

Use case: Generalization helps protect patient privacy in health care by replacing specific data points—such as exact ages or test results—with broader categories, enabling medical research while reducing the risk of re-identification. Transforming patient data into fixed intervals and replacing values with carefully calculated averages, for example, allows researchers to analyze trends without exposing individual identities.[43]

Hashing

What it is: Hashing is a process that converts input data, such as passwords or files, into a fixed-length string of characters, known as a hash. This hash is a unique identifier that represents the original data but cannot be reversed back to the original data—unlike a token, which can be reversed.[44] Hashing is widely used for ensuring data integrity, verifying authenticity, and securely storing sensitive information like passwords.

How it works: Hashing algorithms take an input (such as a password) and apply a mathematical function that produces a fixed-length hash value. The important property of hashing is that it is a one-way function, meaning that once the data is hashed, it cannot be converted back to its original form. For example, when storing passwords, instead of saving the password itself, systems store the hash of the password. During login, the entered password is hashed and compared to the stored hash. If the hashes match, the password is correct. This approach ensures that even if the hash is exposed, the original password cannot be easily recovered.

Plain metaphor example: Hashing is like making a smoothie from a mixture of fruit. Once the fruits are blended into the smoothie, you can't separate the individual pieces of fruit back out, but you can still tell the smoothie was made from those fruits based on its flavor.

Use case: In password security, including government passwords, when a user logs in, the system hashes the entered password and compares it to the stored hash.[45]

K-Anonymity

What it is: K-anonymity is a privacy principle that ensures any individual's data is indistinguishable from at least k-1 other individuals in a dataset. This reduces the risk of identifying a specific person when analyzing or sharing data. K-anonymity works by ensuring that for any individual, their attributes (such as age, zip code, or gender) are shared by at least k-1 other people in the dataset. For example, when k=2, an individual's data is indistinguishable from at least one other person's data. The higher the value of k, the stronger the privacy protection.

How it works: K-anonymity is typically achieved by modifying data through generalization or suppression. For example, if an individual's exact birth date is included in a dataset, it could be replaced with a broader date range, such as “January 1, 1980–December 31, 1989,” so that at least k-1 other individuals share the same birth date range. Similarly, exact geographic locations could be generalized into broader regions to ensure multiple individuals share the same location. K-anonymity ensures that data analysis is still possible without compromising individual privacy, even if other data sources are available for cross-referencing.

Plain metaphor example: K-anonymity is like trying to spot your friends in a crowded stadium full of people wearing the home team’s colors. If you’re surrounded by 10 people in the same jersey, it’s hard to identify who’s who. The more people in the crowd with the same jersey, the harder it becomes to single out anyone, keeping everyone’s identity protected in the sea of fans.

Use case: K-anonymization protects privacy in local-level-area datasets by grouping data and suppressing low-value cells, reducing the risk of identification. Agencies like the U.S. Census Bureau use this approach to prevent attacks on the data, such as reverse geocoding and differencing, while preserving data utility.[46]

Private Set Intersection

What it is: Private set intersection (PSI) is a type of multi-party computation that allows two parties to compare their datasets to identify common elements, while keeping the rest of the data private.[47] PSI ensures that no party learns anything about the other party’s data, except for the items that are present in both sets.

How it works: In PSI, each of the two parties has a private set of data, and they engage in a protocol that allows them to securely compute the intersection (i.e., the common elements) between the two sets. During this process, neither party reveals any information about their private data outside of the common elements. This is accomplished using cryptographic techniques such as encryption or secure hashing. The result of the protocol is the list of common items, without any leakage of additional information about the sets. PSI is useful in scenarios where both parties need to know shared data, but neither is willing to share their entire dataset.

Plain metaphor example: PSI is like two people comparing their contact lists to see which friends they share, without showing each other their entire list of contacts. They can identify and only reveal the common names, keeping their individual contact lists private.

Use case: The strengths of PSI lies in enabling the comparison of datasets without revealing unnecessary information. The FBI and local police departments have the potential to use PSI to check if a suspect appears on watchlists without revealing entire law enforcement records.[48]

Secure Multi-Party Computation

What it is: Secure multi-party computation (SMPC) allows multiple parties to collaboratively compute a function over their private inputs while keeping those inputs confidential.[49] The key feature of SMPC is that no participant learns anything about the others' private data during the computation, ensuring privacy and security.

How it works: SMPC involves dividing the computation process into smaller pieces and distributing them across multiple parties, each of which performs a computation using its own private data. At no point does any participant receive access to the other participants' raw data; they only receive partial results, which are combined at the end to produce the final output. This allows for joint computations, like analyzing shared data, comparing results, or making collective decisions, without revealing any individual's private information. The security comes from the fact that the computations are designed in such a way that no single participant has enough information to infer anything about others' inputs.

Plain metaphor example: SMPC is like a group of people each solving different pieces of a puzzle. Each person only sees their own piece of the puzzle, and at the end, they combine their pieces to see the complete image, but no one ever learns what the other pieces look like before the final combination.

Use case: In 2020, a group of European health care organizations leveraged SMPC to securely analyze patient data across multiple hospitals without exposing sensitive personal data.[50] Using SMPC, they could compute joint results of disease spread and treatment efficacy on encrypted data without any hospital revealing its internal data or jeopardizing patient privacy.

Synthetic Data

What it is: Synthetic data is artificially, computer-generated information that mirrors the structure, patterns, and statistical properties of real-world data but does not contain any actual personal or sensitive information.[51] It is created using algorithms and statistical models, replicating the patterns and relationships found in the real data. By design, synthetic data behaves similarly to real data in analysis, but because it doesn't trace back to any individual's real information, it can be shared more freely.

How it works: Synthetic data is created by models that learn from real data to generate new, similar data. One common method, generative adversarial networks (GANs), uses two parts: a generator that creates new data and a discriminator that checks if the data looks real. Other techniques focus on relationships between data points, like how age might be related to income. Instead of just changing ages randomly, the model keeps the overall pattern intact while altering the data, making it hard to reverse-engineer but still useful for analysis.[52]

Plain metaphor example: An artist creates a painting based on a photograph. The artist doesn't replicate the photo exactly but captures its key elements—such as color, shape, proportions, and scale—to create a unique painting that still holds a strong resemblance to the source photograph.

Use case: During the U.K.-U.S. PETs Prize Challenge, innovators were tasked with developing federated learning solutions to improve pandemic forecasting while maintaining privacy. Participants used a synthetic dataset that was created by the University of Virginia's Biocomplexity Institute as a digital twin of a real population, preserving statistical and behavioral properties without exposing actual personal data. This approach demonstrated how synthetic data can support critical public health responses by enabling secure data sharing and analysis without compromising individual privacy.[53]

Tokenization

What it is: Tokenization is the process of replacing sensitive data elements, such as credit card numbers or personal identifiers, with unique identifiers, or "tokens." Unlike hashing, tokenization is reversible, making it suitable for situations where the original data may need to be retrieved. This makes tokenization ideal for scenarios where re-association may be needed, and it should not be used if the goal is to prevent re-identification by the primary data processor. The tokens do not have any meaningful value outside the context of the system that issued them, ensuring that they cannot be used to access the original data.

How it works: Tokenization works by generating a random string of text, based on a piece of sensitive data, to be used as the token. The actual sensitive data is never transmitted or stored with the token. When the data needs to be accessed or processed, the system can use the token instead of the original data, ensuring that even if the token is intercepted, it has no useful value. This prevents sensitive information from being exposed during transactions or data storage.

Plain metaphor example: Tokenization is like getting a new membership card at a gym, where your real name and contact details are replaced with a unique number. Even though the number is on your card, no one can learn your real identity from just seeing the number—it's stored securely by the gym.

Use case: Government programs such as the Supplemental Nutrition Assistance Program (SNAP), Medicare, and unemployment insurance limit data sharing with other agencies to safeguard citizens' sensitive data. While this is a measure to minimize risk, keeping the data separate can hurt agencies' ability to extract insights and understand their populations. Tokenization offers an opportunity for governments to share data across agencies without compromising individual privacy.[54]

Trusted Execution Environment

What it is: A trusted execution environment (TEE) is a secure area within a processor that runs code in isolation from the rest of the system, ensuring that sensitive data is processed in a trusted and confidential manner.[55] TEEs are designed to protect data and computations from being accessed or tampered with, even by the operating system or malicious software.

How it works: A TEE creates a secure enclave within a processor, where both code and data are isolated from the rest of the system. When a program runs inside a TEE, it is protected from external interference or observation, ensuring that sensitive operations can occur in a trusted environment. TEEs are typically used to process sensitive data, such as encryption keys or financial information, ensuring that this data remains private and secure even when the system itself may be compromised.

Plain metaphor example: A TEE is like having a locked safe inside your house that only you can access. Even if someone else enters your house, they can't open the safe and see what's inside because it's securely isolated.

Use case: TEEs can be leveraged in any context in which a government entity might deal with sensitive data. For example, the Social Security Administration and the U.S. Department of Labor could use TEEs to detect fraudulent disability and unemployment claims without exposing the full databases of all individuals receiving government welfare.

Zero-Knowledge Proof

What it is: A zero-knowledge proof (ZKP) is a cryptographic method that allows one party to prove to another party that they know a piece of information (e.g., a password or secret) without revealing the information itself.[56]

How it works: In a ZKP, the prover (who knows the secret) and the verifier (who wants to be convinced) engage in a protocol where the prover demonstrates knowledge of the secret without ever revealing it. The protocol typically involves the prover presenting evidence that they can correctly solve a problem or answer a question based on the secret, without actually disclosing the secret. ZKPs are used in many PETs to allow for secure authentication or transactions without revealing private data.

Plain metaphor example: A ZKP is like a magician proving they know how a trick works, without ever revealing the secret behind the trick. They show you the result, but not the method used to achieve it.

Use case: Zero-knowledge proofs could enable users to prove they meet age requirements without revealing their exact age or identity. A 2022 demonstration developed by the innovation laboratory at France's National Commission on Information and Liberty (CNIL) showcases a privacy-preserving age-verification system where a trusted third party certifies a user's eligibility without disclosing personal data. This approach strengthens online privacy while ensuring compliance with age restrictions, offering a scalable solution for secure digital identity verification.[57]

Key Considerations for Decision-Making

When deciding which privacy-enhancing technology (PET) to use, there are several key factors to consider. These factors include the type of data, the sensitivity level of the data, the data-sharing needs, the required privacy guarantee level, and scalability requirements. Each of these considerations plays a critical role in determining which PET is best suited for a particular use. Table 1 breaks this down in a more visual way, making it easier to compare and understand how each factor influences the choice of the most appropriate PET for a specific situation. For a more detailed explanation of the questions used to evaluate each PET, please refer to **Appendix 2**.

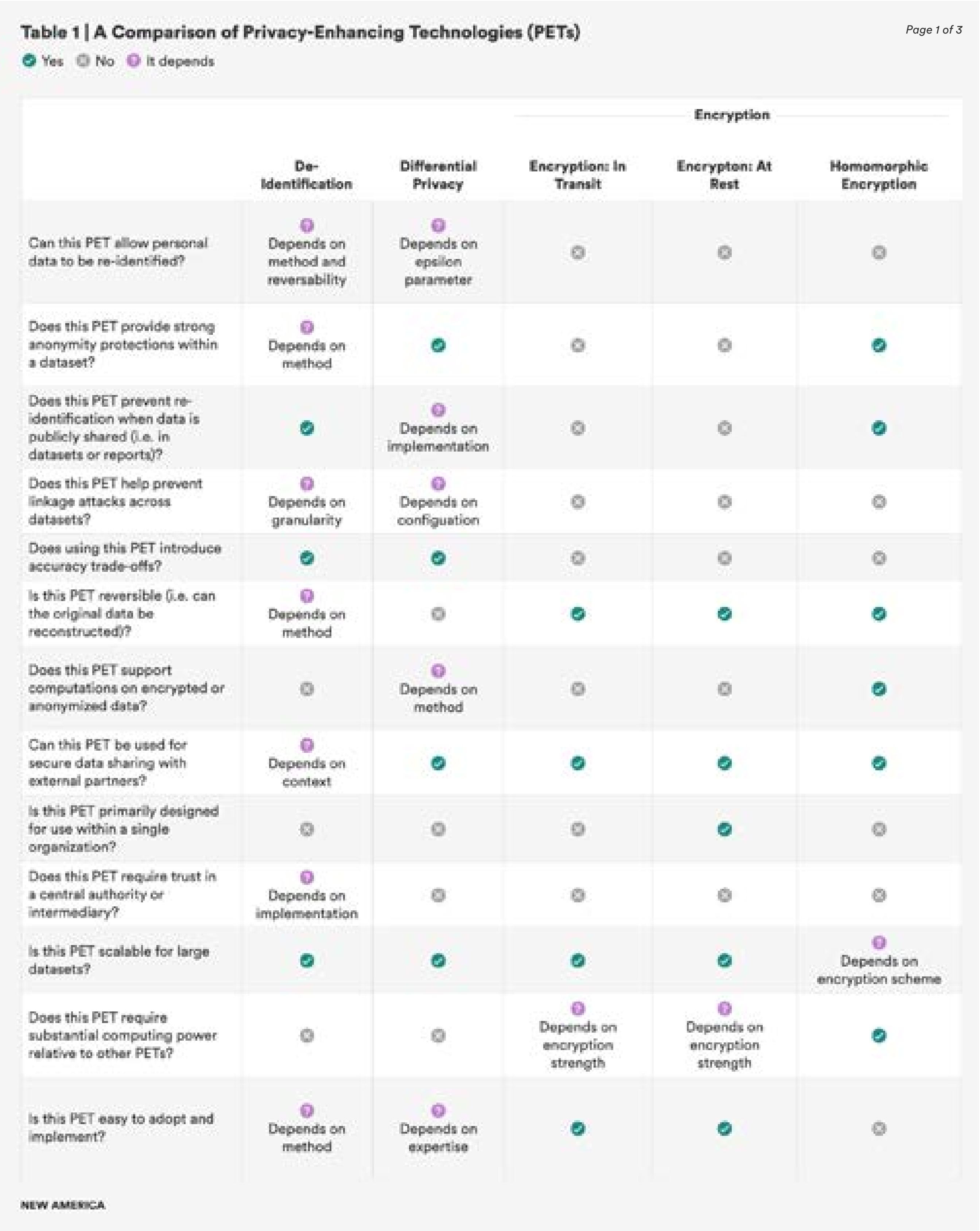


Table 1 | A Comparison of Privacy-Enhancing Technologies (PETs)

Page 1 of 3

✓ Yes ✗ No ? It depends

			Encryption		
	De-Identification	Differential Privacy	Encryption: In Transit	Encrypton: At Rest	Homomorphic Encryption
Can this PET allow personal data to be re-identified?	? Depends on method and reversability	? Depends on epsilon parameter	✗	✗	✗
Does this PET provide strong anonymity protections within a dataset?	? Depends on method	✓	✗	✗	✓
Does this PET prevent re-identification when data is publicly shared (i.e. in datasets or reports)?	✓	? Depends on implementation	✗	✗	✓
Does this PET help prevent linkage attacks across datasets?	? Depends on granularity	? Depends on configuation	✗	✗	✗
Does using this PET introduce accuracy trade-offs?	✓	✓	✗	✗	✗
Is this PET reversible (i.e. can the original data be reconstructed)?	? Depends on method	✗	✓	✓	✓
Does this PET support computations on encrypted or anonymized data?	✗	? Depends on method	✗	✗	✓
Can this PET be used for secure data sharing with external partners?	? Depends on context	✓	✓	✓	✓
Is this PET primarily designed for use within a single organization?	✗	✗	✗	✓	✗
Does this PET require trust in a central authority or intermediary?	? Depends on implementation	✗	✗	✗	✗
Is this PET scalable for large datasets?	✓	✓	✓	✓	? Depends on encryption scheme
Does this PET require substantial computing power relative to other PETs?	✗	✗	? Depends on encryption strength	? Depends on encryption strength	✓
Is this PET easy to adopt and implement?	? Depends on method	? Depends on expertise	✓	✓	✗

NEW AMERICA

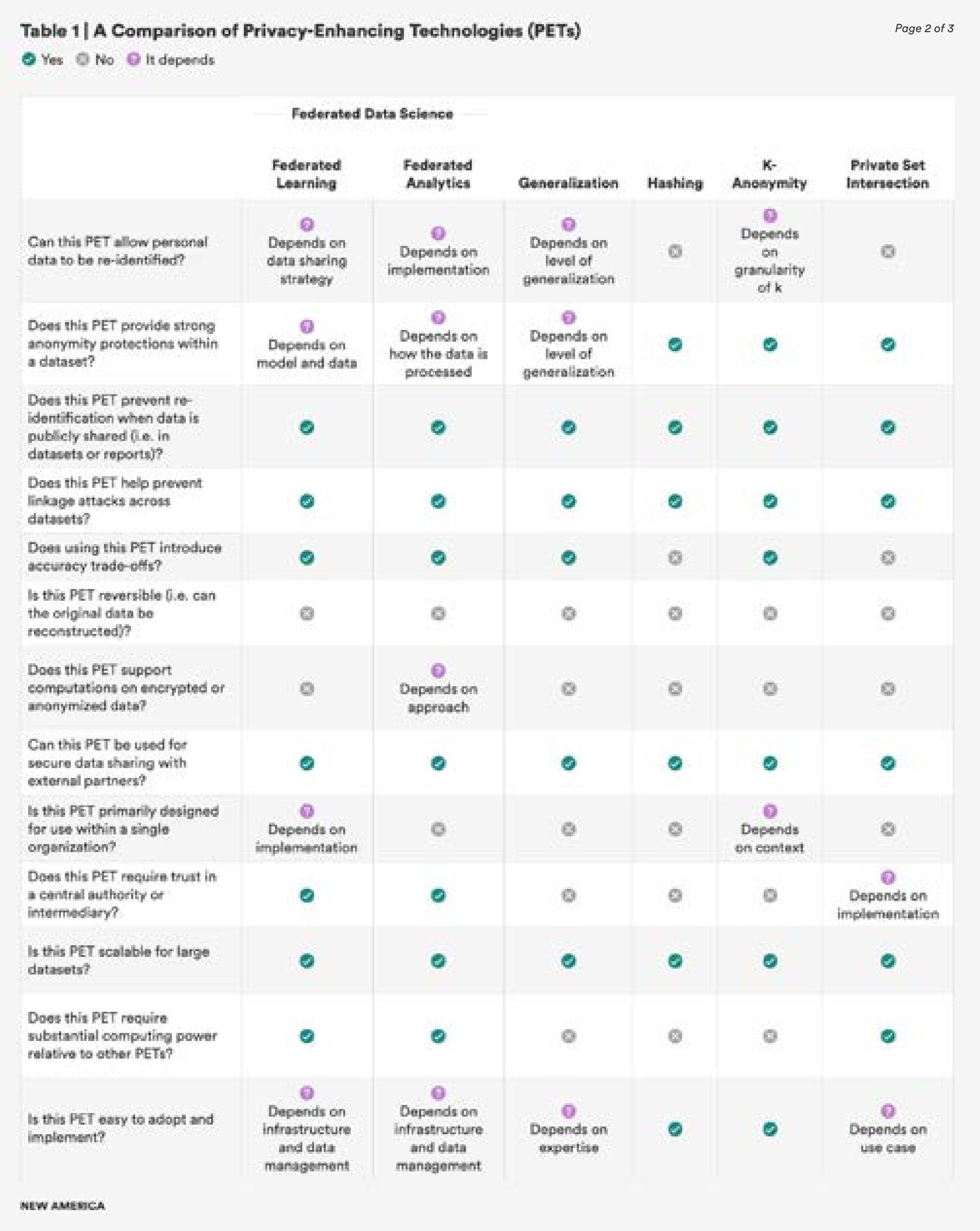


Table 1 | A Comparison of Privacy-Enhancing Technologies (PETs)

Page 2 of 3

✓ Yes ✗ No ? It depends

	Federated Data Science					
	Federated Learning	Federated Analytics	Generalization	Hashing	K-Anonymity	Private Set Intersection
Can this PET allow personal data to be re-identified?	? Depends on data sharing strategy	? Depends on implementation	? Depends on level of generalization	✗	? Depends on granularity of k	✗
Does this PET provide strong anonymity protections within a dataset?	? Depends on model and data	? Depends on how the data is processed	? Depends on level of generalization	✓	✓	✓
Does this PET prevent re-identification when data is publicly shared (i.e. in datasets or reports)?	✓	✓	✓	✓	✓	✓
Does this PET help prevent linkage attacks across datasets?	✓	✓	✓	✓	✓	✓
Does using this PET introduce accuracy trade-offs?	✓	✓	✓	✗	✓	✗
Is this PET reversible (i.e. can the original data be reconstructed)?	✗	✗	✗	✗	✗	✗
Does this PET support computations on encrypted or anonymized data?	✗	? Depends on approach	✗	✗	✗	✗
Can this PET be used for secure data sharing with external partners?	✓	✓	✓	✓	✓	✓
Is this PET primarily designed for use within a single organization?	? Depends on implementation	✗	✗	✗	? Depends on context	✗
Does this PET require trust in a central authority or intermediary?	✓	✓	✗	✗	✗	? Depends on implementation
Is this PET scalable for large datasets?	✓	✓	✓	✓	✓	✓
Does this PET require substantial computing power relative to other PETs?	✓	✓	✗	✗	✗	✓
Is this PET easy to adopt and implement?	? Depends on infrastructure and data management	? Depends on infrastructure and data management	? Depends on expertise	✓	✓	? Depends on use case

NEW AMERICA

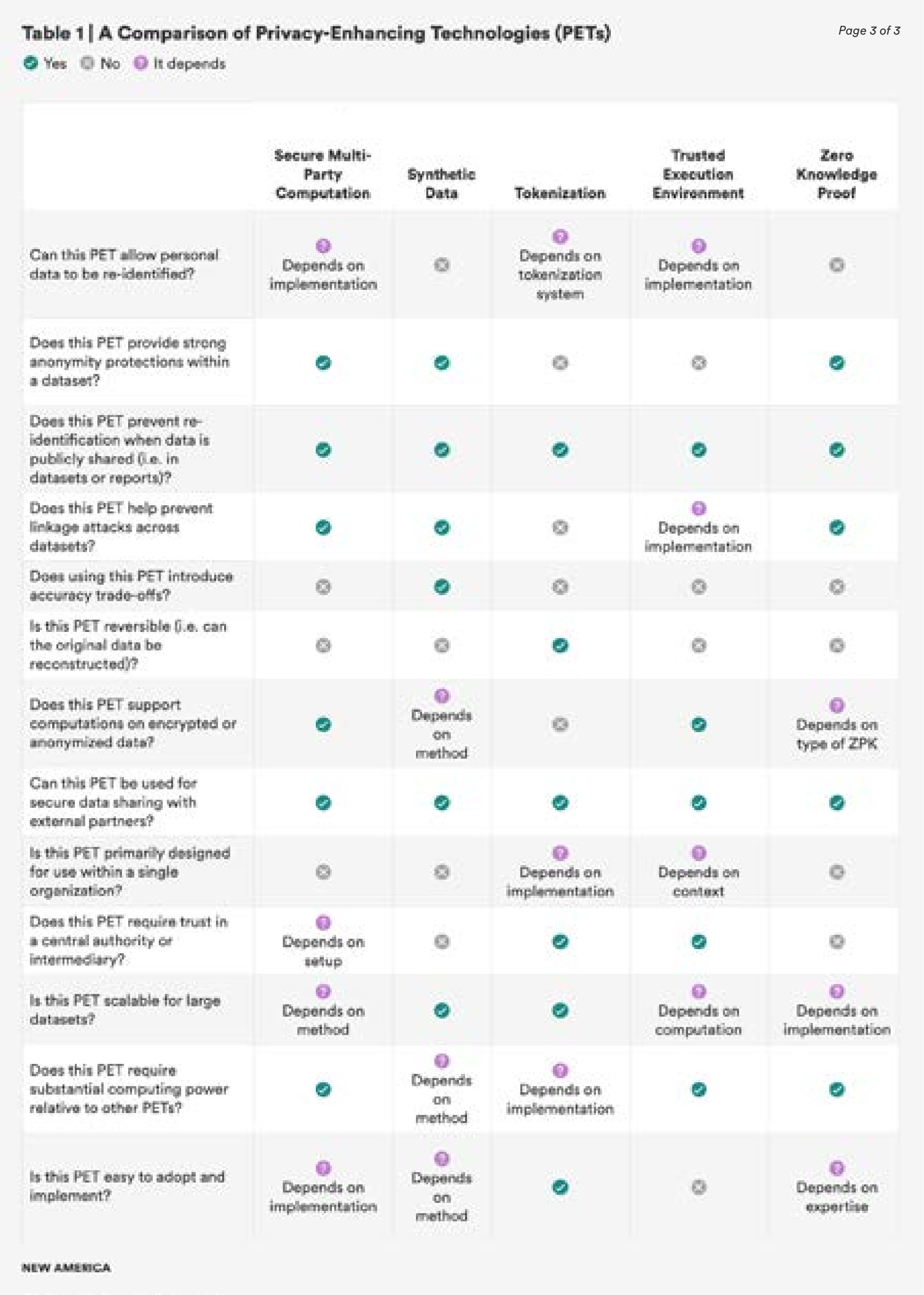


Table 1 | A Comparison of Privacy-Enhancing Technologies (PETs)

Page 3 of 3

✓ Yes ✗ No ? It depends

	Secure Multi-Party Computation	Synthetic Data	Tokenization	Trusted Execution Environment	Zero Knowledge Proof
Can this PET allow personal data to be re-identified?	? Depends on implementation	✗	? Depends on tokenization system	? Depends on implementation	✗
Does this PET provide strong anonymity protections within a dataset?	✓	✓	✗	✗	✓
Does this PET prevent re-identification when data is publicly shared (i.e. in datasets or reports)?	✓	✓	✓	✓	✓
Does this PET help prevent linkage attacks across datasets?	✓	✓	✗	? Depends on implementation	✓
Does using this PET introduce accuracy trade-offs?	✗	✓	✗	✗	✗
Is this PET reversible (i.e. can the original data be reconstructed)?	✗	✗	✓	✗	✗
Does this PET support computations on encrypted or anonymized data?	✓	? Depends on method	✗	✓	? Depends on type of ZPK
Can this PET be used for secure data sharing with external partners?	✓	✓	✓	✓	✓
Is this PET primarily designed for use within a single organization?	✗	✗	? Depends on implementation	? Depends on context	✗
Does this PET require trust in a central authority or intermediary?	? Depends on setup	✗	✓	✓	✗
Is this PET scalable for large datasets?	? Depends on method	✓	✓	? Depends on computation	? Depends on implementation
Does this PET require substantial computing power relative to other PETs?	✓	? Depends on method	? Depends on implementation	✓	✓
Is this PET easy to adopt and implement?	? Depends on implementation	? Depends on method	✓	✗	? Depends on expertise

NEW AMERICA

The **type of data** is one of the most important considerations when selecting a PET. For example, if the data contains personally identifiable information (PII), such as names, Social Security numbers, or addresses, the technology chosen must be capable of protecting individuals' identities. PETs like encryption, tokenization, and differential privacy are specifically designed for handling sensitive PII and ensuring that individual records are protected.[58] On the other hand, if the data is aggregated or anonymized—such as trends or summaries without any direct identifiers—the requirements for privacy protection may be less stringent. In these cases, techniques like generalization or federated analytics may be sufficient, as they preserve privacy while allowing for broader data analysis without exposing specific individual information.

The **sensitivity level** of the data also determines the privacy measures needed. Highly sensitive data, such as medical records or financial information, demands more robust security measures to prevent breaches or misuse. For such sensitive datasets, PETs like homomorphic encryption or secure multi-party computation (SMPC) can provide strong privacy guarantees by ensuring that the data remains encrypted even during processing and analysis, thus minimizing the risk of exposure.[59] In contrast, data with medium or low sensitivity, such as business analytics or public records, may not require as strict of protections. For these types of data, techniques like de-identification or k-anonymity may offer a sufficient balance between privacy and usability, as they can anonymize or group data without significantly diminishing its usefulness for analysis.[60]

Another critical factor to consider is **data-sharing needs**. The level of privacy protection required can vary greatly depending on whether the data will be shared within a trusted environment or with external partners. For example, if data is being shared within a government agency or between departments with strong internal security controls, the privacy requirements may be less strict, and basic encryption techniques or tokenization might be adequate.[61] However, if data is being shared with external entities, such as third-party vendors, contractors, or researchers, stronger privacy protections are necessary to ensure that sensitive information remains secure and that unauthorized access is prevented. In such cases, PETs like federated learning, private set intersection, or differential privacy allow for secure data sharing and analysis while ensuring that individual-level data is not exposed to external parties.[62]

The **required privacy-guarantee level** plays a significant role in selecting a PET. Some applications prioritize maximum privacy, even if this means sacrificing data precision. For example, when dealing with extremely sensitive data, high-privacy guarantees may be necessary, even if this reduces the precision of the data. Techniques like differential privacy introduce random noise to datasets, ensuring that the privacy of individuals is protected while allowing for aggregate analysis.[63] On the other hand, some use cases require a more balanced approach, where a moderate level of privacy is acceptable but maintaining data precision is still important. In these scenarios, PETs like k-anonymity or tokenization may offer a good compromise, as they protect personal information while still allowing for detailed, actionable insights from the data.[64]

Finally, **scalability requirements** are a crucial consideration when selecting a PET. As datasets grow in size, the scalability of the privacy-enhancing technology becomes more important. For smaller datasets, more resource-intensive PETs like homomorphic encryption or SMPC might be feasible because the computational overhead is manageable.[65] However, for large datasets, it's essential to choose a PET that can scale without compromising performance. Techniques such as federated learning or federated analytics are particularly useful in large-scale environments because they allow data to be processed across multiple devices or servers, ensuring that privacy is maintained without the need to centralize sensitive information.[66] Additionally, distributed encryption and generalization techniques can scale well to handle large volumes of data while still preserving privacy.[67]

In summary, selecting the appropriate PET depends on the specific data characteristics and the intended use case. By carefully evaluating the type of data, its sensitivity level, sharing needs, privacy guarantees, and scalability requirements, organizations can choose the right PET to meet their privacy and security goals while still enabling effective data analysis.

Combining PETs to Maximize Utility and Privacy

The combination of privacy-enhancing technologies (PETs) depends on the data flow and lifecycle—which include how data is collected, processed, shared, and analyzed. Aligning PET selection with the data cycle and intended utility helps organizations maximize both privacy and the value of their data.

Some PETs require a sequential approach because certain protections must be in place before the next stage of processing. For example, a government agency managing benefit enrollment might first apply encryption in transit to secure personal and financial information as it moves between systems and then use homomorphic encryption to analyze eligibility criteria without exposing individuals' raw data. Similarly, k-anonymity might be applied to de-identify enrollment data before generating synthetic datasets that enable policy analysis without revealing specific applicants.

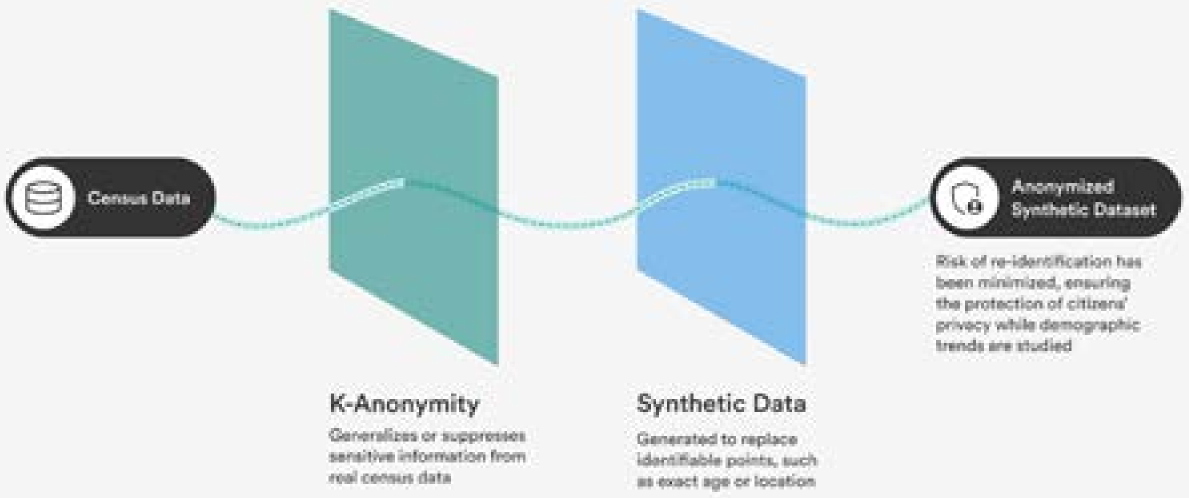
In other cases, PETs can function independently or in parallel, depending on the privacy risks and data-use requirements. For example, differential privacy can be applied directly to reports on program participation to prevent re-identification, while federated learning allows agencies to collaborate on improving benefit delivery without sharing individual records. Since these techniques address different risks, they do not need to be applied in a strict sequence.

Choosing the right PET approach requires understanding the full data lifecycle. This means assessing the sensitivity of the data, the risks at each stage, and the level of privacy protection required while ensuring the data remains useful. If data will be shared across agencies or undergo multiple transformations, a sequential approach may be necessary to maintain privacy at every step. If privacy risks are distinct and manageable separately, PETs can be applied independently.

Sequential: K-Anonymity and Synthetic Data

Governments could apply k-anonymity to real census data, ensuring that individuals' identities are protected by generalizing or suppressing sensitive information. Then, they could generate synthetic datasets to replace specific, identifiable data points, such as exact age or location, enabling continued research and policy development without compromising citizens' privacy. This approach ensures that demographic trends are studied while minimizing the risk of re-identification.

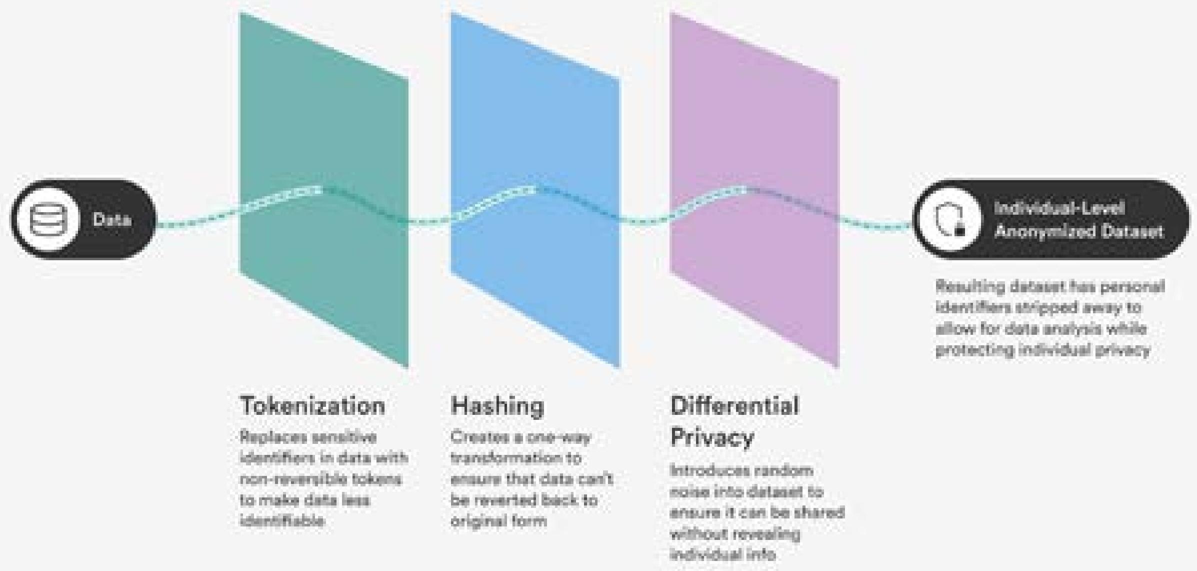
Figure 1 | Layering PETs: K-Anonymity + Synthetic Data


Census Data
Anonymized Synthetic Dataset
Risk of re-identification has been minimized, ensuring the protection of citizens' privacy while demographic trends are studied
K-Anonymity
Generalizes or suppresses sensitive information from real census data
Synthetic Data
Generated to replace identifiable points, such as exact age or location

Sequential: Tokenization, Hashing, and Differential Privacy

The combination of tokenization, hashing, and differential privacy in a sequential process works by progressively securing data at each step. First, tokenization replaces sensitive identifiers with non-reversible tokens, making the data less identifiable. Then, hashing further secures the tokenized data by creating a one-way transformation, ensuring that even if data is exposed, it can't be reverted back to its original form. Finally, differential privacy ensures that the data can be shared or analyzed without revealing individual-specific information by introducing random noise into the dataset. This sequential approach is critical in creating an individual-level anonymized dataset, as seen in the census, where personal identifiers are stripped away to allow for data analysis while protecting privacy.

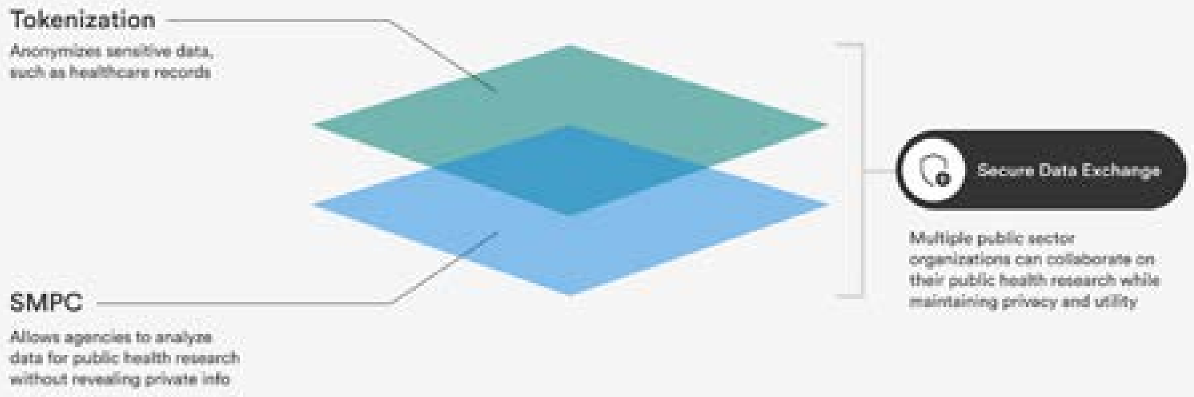
Figure 2 | Layering PETs: Tokenization + Hashing + Differential Privacy


Data
Individual-Level Anonymized Dataset
Resulting dataset has personal identifiers stripped away to allow for data analysis while protecting individual privacy
Tokenization
Replaces sensitive identifiers in data with non-reversible tokens to make data less identifiable
Hashing
Creates a one-way transformation to ensure that data can't be reverted back to original form
Differential Privacy
Introduces random noise into dataset to ensure it can be shared without revealing individual info

Non-Sequential: Tokenization and Secure Multi-Party Computation

Tokenization and secure multi-party computation (SMPC) can be used by public sector organizations to enhance privacy while enabling collaboration. Tokenization anonymizes sensitive data, such as health care records, before sharing it between agencies, ensuring personal information remains protected. SMPC allows agencies to analyze the data for public health research without revealing private information. Since tokenization and SMPC operate independently, this approach is non-sequential while maintaining privacy and utility.

Figure 3 | Layering PETs: Tokenization + Secure Multi-Party Computation (SMPC)


Tokenization
Anonymizes sensitive data, such as healthcare records
SMPC
Allows agencies to analyze data for public health research without revealing private info
Secure Data Exchange
Multiple public sector organizations can collaborate on their public health research while maintaining privacy and utility

Practical Considerations and Barriers to PET Adoption

Adopting privacy-enhancing technologies (PETs) presents several challenges, including technical complexity, usability issues, and regulatory gaps. These barriers can slow the widespread implementation of PETs and complicate their integration into organizations' data privacy practices.

Technical complexity is a key barrier to PET adoption. Many PETs, especially those involving advanced cryptographic techniques such as homomorphic encryption or secure multi-party computation, require specialized knowledge and are computationally intensive.[68] For example, differential privacy—used by the U.S. Census Bureau in the 2020 Census—adds noise to data to protect individual privacy.[69] However, this approach can reduce accuracy, particularly in small communities with limited data.[70] The challenge lies in balancing privacy protection with data utility, as introducing too much noise can lead to inaccurate or unrepresentative insights.

Usability issues in PETs are another obstacle. Many PETs are complex to configure and manage, making them less accessible to non-technical users.[71] Poorly designed interfaces can hinder adoption, as organizations may struggle to implement these technologies correctly.[72] To overcome this, user-friendly interfaces that abstract the technical complexity of PETs could help organizations deploy these tools more effectively, making privacy protections easier to use without sacrificing security.

Regulatory and standardization gaps also hinder PET adoption. For example, the European Union's General Data Protection Regulation (GDPR) requires "appropriate technical and organisational measures" to protect personal data but does not specify which PETs ensure compliance.[73] This lack of clarity can create confusion about which technologies are legally acceptable, especially when jurisdictions interpret privacy laws differently. Developing standardized guidelines for PET usage would provide clearer compliance paths, helping organizations confidently choose the right technologies for their needs while adhering to legal requirements.[74]

Cost constraints are a major hurdle in PET adoption. High implementation costs, ongoing maintenance expenses, and integration complexities make it difficult for organizations to justify investment, particularly when privacy protections do not directly contribute to revenue generation.[75] Many PETs require specialized infrastructure or technical expertise. Advanced techniques such as homomorphic encryption or secure multi-party computation demand substantial computing resources, driving up costs.[76] If the return on investment isn't clear, organizations may deprioritize PET adoption and the data sharing allowed in favor of other pressing technological needs. Having a clear understanding of the value of PET-enabled data sharing can help, as can implementing them in phases.

Operational challenges further complicate PET implementation. Many organizations or government agencies lack the in-house expertise required to deploy and maintain these technologies effectively.[77] Government agencies may struggle with the technical demands of PETs, from configuring privacy-preserving algorithms to ensuring ongoing compliance with evolving regulations and best practices. Without dedicated resources, PETs may become underutilized or misconfigured, reducing their effectiveness. Streamlining PET integration through managed services, a dedicated support staff or team, and a phased approach can help address these operational challenges.

Awareness and expertise gaps—a lack of awareness and understanding of PETs among key stakeholders—present another significant barrier. Many decision makers are unfamiliar with PETs' functionality, potential integration with existing systems and data lifecycles, and benefits to improving data use and sharing efforts, leading to hesitation around investing.[78] Moreover, PETs are often perceived as more complex or resource-intensive than they are, which makes it difficult to appropriately build a strong case for their implementation. In areas where privacy risks or data-sharing efforts are not immediately pressing, organizations and agencies may prioritize more familiar measures, or just default to not sharing the data generally.[79] Adopting PETs also requires cross-functional collaboration, yet expertise is typically confined to specific departments or data-sharing initiatives, such as the census.

This can result in fragmented efforts that reduce the overall utility of PETs. Overcoming these gaps with targeted training and knowledge sharing can help integrate PETs more effectively into data-governance strategies.

Advancing the Use of Privacy-Enhancing Technologies

To incentivize the implementation of privacy-enhancing technologies (PETs), governments, research institutes, philanthropic organizations, and private partners should focus on creating an ecosystem that fosters both development and widespread adoption. Given the growing importance of privacy-preserving data sharing, these stakeholders can take key steps to drive this transformation in public sector data practices.

First, governments should align procurement policies with criteria that prioritize affordability, scalability, and effectiveness in real-world, public-sector applications. For instance, the Massive Data Institute's Privacy-Enhancing Technologies initiative provides guidance to school districts and education agencies on how to use PETs to securely use student data to improve service delivery and learner outcomes.[80] By establishing clear benchmarks—particularly for solutions that can be adapted across various use cases—that support both privacy and data utility, governments can encourage vendors to competitively bid on their ability to deliver strong, secure, and efficient technologies.[81] Long-term contracts or public-private partnerships can further encourage innovation by providing a predictable revenue stream, motivating vendors to invest in scalable solutions that meet evolving needs.

Secondly, governments should also issue clear guidelines for determining when PETs are sufficient for meeting privacy and security requirements. Defining specific use cases where specific PETs—such as encryption or anonymization—are effective in mitigating risks while maintaining data utility can ensure that solutions are both secure and practical. These guidelines and frameworks also empower government users to confidently adopt a range of PET solutions, knowing they are in compliance with regulatory and privacy requirements.

Thirdly, the market alone may not fully address the need for effective PETs, especially given that the case for data sharing to improve government programs and services may not always be prioritized. To fill this gap, governments and other stakeholders should provide incentives for innovation through mechanisms like targeted grants or collaborative research initiatives. The Future of Privacy Forum's Privacy-Enhancing Technologies Research Collaboration Network, supported by the U.S. National Science Foundation and Department of Energy, seeks to build a repository of existing use cases; this effort is an example of how government-supported collaboration can foster development and use-case clarity.[82]

International partnerships like the one between the U.K. Information Commissioner's Office and the U.S. National Institute of Standards and Technology demonstrate the value of cross-border efforts. A key initiative supported by these entities was a global competition in which experts from academic institutions and private companies competed for a combined $1.6 million prize pool.[83] These efforts show that when countries align, they can lower costs and drive the development of emerging technologies.

Conclusion

As data becomes increasingly essential for effective governance, protecting privacy remains critical. Privacy-enhancing technologies (PETs) provide a means for governments to unlock the value of data while safeguarding sensitive information. By promoting innovation, fostering collaboration, and offering clear guidelines, governments can create an environment where data serves the public good without compromising privacy. This guide aims to lower barriers to understanding the benefits and applications of various PETs. There is an urgent need for public sector stakeholders to prioritize securely using data to drive positive outcomes while upholding the privacy and rights of individuals.

Appendix 1. Key Term Definitions

- **Algorithm**: a step-by-step set of instructions a computer follows to solve a problem or perform a task.
- **Cryptographic method**: a technique used to secure data by transforming it into a format that is unreadable to unauthorized users, often used to protect sensitive information.
- **Data aggregation**: the process of collecting and summarizing data from multiple sources to form a comprehensive set or report, often used for analysis and reporting.
- **Data de-identification**: the process of removing or modifying personally identifiable information from datasets, so that individuals cannot be readily identified from the data, while maintaining its utility for analysis.
- **Data integrity**: the accuracy and reliability of data.
- **Data lifecycle**: the stages through which data passes from creation or collection, through processing and analysis, to sharing, storage, and eventual deletion or archival.
- **Data processor**: a person or company that handles data on behalf of another organization.
- **Data sharing**: the practice of making data available for access, use, or collaboration with other parties.
- **Input data**: the information that goes into a system (i.e., a search term you type into Google).
- **Machine learning model**: a type of computer program that learns patterns from data and makes predictions or decisions without being explicitly programmed.
- **Noise**: random data or alterations deliberately introduced into a dataset to protect individual privacy by preventing re-identification, commonly used in techniques like differential privacy.

- **Operating system**: the software that runs on a computer or phone and manages all its basic functions, like running apps, storing files, and connecting to the internet.
- **Output data**: the result that comes back from a search (i.e., the list of results you see).
- **Risk of exposure**: the likelihood or potential for sensitive data to be accessed, disclosed, or misused by unauthorized individuals, systems, or entities.
- **Security measures**: the technical, administrative, and physical actions taken to safeguard data against unauthorized access, alteration, destruction, or theft.
- **Sensitive information**: any data in need of extra protection due to its confidential nature, such as health records, financial details, or Social Security numbers that carry more harm if exposed.
- **Server**: a powerful computer that stores and processes data and makes it available to other devices. When visiting a website, for example, you're getting information from a server.
- **Statistical properties**: the characteristics of data, such as averages, trends, and distributions, that can be used to understand patterns without revealing specific personal details.
- **Transaction**: the process of transferring or exchanging data between parties, such as in financial exchanges or when querying databases.
- **Unintended disclosure**: the accidental or inadvertent disclosure of sensitive data, potentially due to technical flaws or human error.

Appendix 2. Key Evaluation Questions for Privacy-Enhancing Technologies (PETs)

This appendix outlines the series of key questions used to evaluate privacy-enhancing technologies (PETs) in **Table 1**. Each question is designed to assess specific aspects of a PET, such as its ability to maintain data privacy, prevent re-identification, and ensure scalability, among other factors. The answers to these questions help determine a PET's strengths, limitations, and suitability for different use cases, supporting informed decisions in the adoption of these technologies for privacy protection.

1. **Does this PET allow personal data to be re-identified?** This checks whether the method or process could enable unauthorized parties to re-identify personal data, either in transit or after protection is applied.

2. **Does this PET provide strong anonymity protections?** This checks whether the method ensures that individual data points are indistinguishable within a dataset.

3. **Does this PET prevent re-identification in publicly shared data?** This asks whether the technology ensures that personal data cannot be re-identified when shared publicly, such as in a dataset or report.

4. **Does this PET help prevent linkage attacks across datasets?** This checks whether the PET helps prevent adversaries from combining or correlating datasets to re-identify individuals.

5. **Does this PET introduce accuracy trade-offs?** Some PETs may reduce the quality or accuracy of data to protect privacy. This question asks whether accuracy is sacrificed.

6. **Is this PET reversible (i.e., can the original data be reconstructed)?** This assesses whether the protected data can be reversed or mapped back to the original data, such as through decryption or re-identification techniques.

7. **Does this PET support computations on encrypted or anonymized data?** This checks whether analysis or operations can be performed on encrypted or anonymized data without first decrypting or de-anonymizing it.

8. **Does this PET support secure data sharing with external partners?** This determines if the technology enables data to be securely shared with outside entities while maintaining privacy protections.

9. **Is this PET best suited for use within a single organization?** This checks whether the PET is mainly intended for internal use rather than multi-party collaboration.

10. **Does this PET require trust in a central authority or intermediary?** Some PETs rely on a central entity to manage access, keys, or data processing. This checks whether trust in such an intermediary is required.

11. **Is this PET scalable for large datasets?** Scalability refers to whether the PET can handle large volumes of data efficiently without performance degradation, which is essential for big data applications.

12. **Does this PET require substantial computing power?** Some PETs, especially advanced cryptographic techniques, can be computationally intensive. This checks if the PET is resource-heavy relative to other PETs.

13. **Is this PET easy to implement?** This evaluates how practical and straightforward it is to adopt and integrate the PET into existing systems, including infrastructure and technical support.

Notes

[1] Organization for Economic Cooperation and Development (OECD), *Enhancing Access to and Sharing of Data* (OECD Publishing, 2019), **https://www.oecd.org/en/publications/enhancing-access-to-and-sharing-of-data_276aaca8-en.html.**

[2] U.S. Government Accountability Office (GAO), *Consumer Data: Increasing Use Poses Risks to Privacy* (GAO, 2022), **https://www.gao.gov/products/gao-22-106096**.

[3] Ellen Nakashima, "Hacks of OPM Databases Compromised 22.1 Million People, Federal Authorities Say," *Washington Post*, July 9, 2015, **https://www.washingtonpost.com/news/federal-eye/wp/2015/07/09/hack-of-security-clearance-system-affected-21-5-million-people-federal-authorities-say/**.

[4] "Equifax Data Breach," Electronic Privacy Information Center, **https://archive.epic.org/privacy/data-breach/equifax/**; Jenny Kleeman, "DNA Testing: What Happens If Your Genetic Data Is Hacked?," BBC, February 12, 2024, **https://www.bbc.com/future/article/20240212-dna-testing-what-happens-if-your-genetic-data-is-hacked**.

[5] "Public Interest Privacy Legislation Principles," National Consumers League, November 13, 2018, **https://nclnet.org/privacy_leg_principles/**.

[6] Zeynep Tufekci, "Here Are the Digital Clues to What Musk Is Really Up To," *New York Times*, February 21, 2025, **https://www.nytimes.com/2025/02/21/opinion/musk-doge-personal-data.html**.

[7] Jacob Leibenluft, *"DOGE" Access to Treasury Payment Systems Raises Serious Risks* (Center on Budget and Policy Priorities, 2025), **https://www.cbpp.org/research/federal-budget/doge-access-to-treasury-payment-systems-raises-serious-risks**.

[8] "DOGE Is Putting the Country's Data and Computing Infrastructure at Risk, HKS Expert Argues," Harvard Kennedy School, February 19, 2025, **https://www.hks.harvard.edu/faculty-research/policy-topics/science-technology-data/doge-putting-countrys-data-and-computing**.

[9] U.K. Information Commissioner's Office (ICO), *Chapter 5: Privacy-Enhancing Technologies (PETs)* (ICO, 2022), **https://ico.org.uk/media/about-the-ico/consultations/4021464/chapter-5-anonymisation-pets.pdf**.

[10] Sydney Saubestre, *What's the Value of Privacy?* (New America, 2024), **https://www.newamerica.org/oti/briefs/whats-the-value-of-privacy/**.

[11] United Nations (UN) BigData, *The PET Guide: The United Nations Guide on Privacy-Enhancing Technologies for Official Statistics* (UN Committee of Experts on Big Data and Data Science for Official Statistics, 2023), **https://unstats.un.org/bigdata/task-teams/privacy/guide/2023_UN%20PET%20Guide.pdf**.

[12] Centre for Data Ethics and Innovation (CEDI), *Privacy-Enhancing Technologies Adoption Guide* (CDEI, 2021), **https://cdeiuk.github.io/pets-adoption-guide/**.

[13] Organization for Economic Cooperation and Development, "Emerging Privacy-Enhancing Technologies," *OECD Digital Economy Papers*, no. 351 (2023), **https://www.oecd.org/en/publications/emerging-privacy-enhancing-technologies_bf121be4-en.html**.

[14] Danielle K. Citron and Daniel Solove, "Risk and Anxiety: A Theory of Data Breach Harms," *Texas Law Review*, 96 (2018): 737–786, **https://scholarship.law.bu.edu/faculty_scholarship/616/**.

[15] "As Internet User Numbers Swell Due to Pandemic, UN Forum Discusses Measures to Improve Safety of Cyberspace," United Nations Department of Economic and Social Affairs, **https://www.un.org/en/desa/internet-user-numbers-swell-due-pandemic-un-forum-discusses-measures-improve-safety-cyberspace**.

[16] Sean Lyngaas, "Chinese Hackers Breached US Government Office That Assesses Foreign Investments For National Security Risks," CNN, January 10, 2025, **https://www.cnn.com/2025/01/10/politics/chinese-hackers-breach-committee-on-foreign-investment-in-the-us/index.html**.

[17] Michele Gilman and Rebecca Green, "The Surveillance Gap: The Harms of Extreme Privacy and Data Marginalization," *NYU Review of Law and Social Change* 42, no. 2 (2018), **https://socialchangenyu.com/review/the-surveillance-gap-the-harms-of-extreme-privacy-and-data-marginalization**.

[18] U.K. Information Commissioner's Office, *Chapter 5*, **https://ico.org.uk/media/about-the-ico/consultations/4021464/chapter-5-anonymisation-pets.pdf**.

[19] Adita Karkera et al., *Bridging the Data Sharing Chasm* (Deloitte Insights, 2023), **https://www2.deloitte.com/us/en/insights/industry/public-sector/government-trends/2023/boosting-data-sharing-across-government.html**.

[20] Chris Sadler, *Protecting Privacy in Data Releases* (New America, 2020), **https://www.newamerica.org/oti/reports/primer-disclosure-limitation/**.

[21] Simon Fondrie-Teitler, "Keeping Your Privacy-Enhancing Technology (PET) Promises," *Office of Technology Blog*, Federal Trade Commission, February 1, 2024, **https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2024/02/keeping-your-privacy-enhancing-technology-pet-promises**.

[22] Privacy Technical Assistance Center, *Data De-identification: An Overview of Basic Terms* (U.S. Department of Education, 2013), **https://studentprivacy.ed.gov/sites/default/files/resource_document/file/data_deidentification_terms_0.pdf**.

[23] "Differential Privacy," Harvard University School of Engineering and Applied Sciences, **https://privacytools.seas.harvard.edu/differential-privacy.**

[24] "Differential Privacy," Harvard, **https://privacytools.seas.harvard.edu/differential-privacy**.

[25] "Differential Privacy for Census Data Explained," National Conference of State Legislatures, November 10, 2021, **https://www.ncsl.org/technology-and-communication/differential-privacy-for-census-data-explained**.

[26] Joseph Near, David Darais, and Kaitlin Boeckl, "Differential Privacy for Privacy-Preserving Data Analysis: An Introduction to Our Blog Series," *Cybersecurity Insights* (blog), National Institute of Standards and Technology, July 27, 2020, **https://www.nist.gov/blogs/cybersecurity-insights/differential-privacy-privacy-preserving-data-analysis-introduction-our**.

[27] Cynthia Dwork et al., "Calibrating Noise to Sensitivity in Private Data Analysis," *Theory of Cryptography: Third Theory of Cryptography Conference* (2006), **https://people.csail.mit.edu/asmith/PS/sensitivity-tcc-final.pdf**.

[28] "Understanding Differential Privacy," U.S. Census Bureau, 2020, **https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/differential-privacy.html**.

[29] Population Reference Bureau and the U.S. Census Bureau's 2020 Census Data Products and Dissemination Team, *Why the Census Bureau Chose Differential Privacy* (Census Bureau, March 2023), **https://www2.census.gov/library/publications/decennial/2020/census-briefs/c2020br-03.pdf**.

[30] Andi Wilson Thompson and Claire Park, *Privacy's Best Friend* (New America, 2020), **https://www.newamerica.org/oti/reports/privacys-best-friend/**.

[31] "What Is Encryption?," U.K. Information Commissioner's Office, **https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/security/encryption/what-is-encryption/**.

[32] Kerry A. McKay and David Cooper, "Guidelines for the Selection, Configuration, and Use of Transport Layer Security (TLS) Implementations," *NIST Special Publication* 800-52 (2019), **https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-52r2.pdf**.

[33] Office of the Chief Information Security Officer, *IT Security Procedural Guide: SSL/TLS Implementation CIO-IT Security-14-69* (General Services Administration, June 12, 2023), **https://www.gsa.gov/system/files?file=SSL-TLS-Implementation-%5BCIO-IT-Security-14-69-Rev-7%5D-06-12-2023.pdf**.

[34] Karen Scarfone, Murugiah Souppaya, and Matt Sexton, "Guide to Storage Encryption Technologies for End User Devices: Recommendations of the National Institute of Standards and Technology," *NIST Special Publication* 800-111 (2007), **https://www.hhs.gov/sites/default/files/nist800111.pdf**.

[35] "Dropbox Account Safety: How Dropbox Keeps Your Files Secure," Dropbox, February 13, 2025, **https://help.dropbox.com/security/how-security-works**; "Default Encryption At Rest," Google Cloud, May 2024, **https://cloud.google.com/docs/security/encryption/default-encryption**.

[36] Chris Sadler, "Homomorphic Encryption Could Fix The Gaps In Our Data Security," *Open Technology Institute Blog*, New America, September 1, 2020, **https://www.newamerica.org/oti/blog/homomorphic-encryption-could-fix-gaps-our-data-security/**.

[37] "Fully Homomorphic Encryption," IBM, September 23, 2021, **https://www.ibm.com/support/z-content-solutions/fully-homomorphic-encryption/**.

[38] Syreen Banabilah et al., "Federated Learning Review: Fundamentals, Enabling Technologies, and Future Applications," *Information Processing & Management* 59, no. 6 (2022), **https://doi.org/10.1016/j.ipm.2022.103061**.

[39] Joseph Near and David Darais, "Protecting Trained Models in Privacy-Preserving Federated Learning," *Cybersecurity Insights* (blog), National Institute of Standards and Technology, July 15, 2024, **https://www.nist.gov/blogs/cybersecurity-insights/protecting-trained-models-privacy-preserving-federated-learning**.

[40] Mona Flores et al., "Federated Learning Used for Predicting Outcomes in SARS-COV-2 Patients," *Research Square* rs-3 (preprint, 2021), **https://pmc.ncbi.nlm.nih.gov/articles/PMC7805458/**.

[41] Dan Wang et al., "Federated Analytics: Opportunities and Challenges," *IEEE Network* 36, no. 1 (January/February 2022), **https://ieeexplore.ieee.org/abstract/document/9617564**.

[42] Daniel Ramage and Stefano Mazzocchi, "Federated Analytics: Collaborative Data Science Without Data Collection," *Google Research Blog*, May 27, 2020, **https://research.google/blog/federated-analytics-collaborative-data-science-without-data-collection/**.

[43] Abdul Majeed, "Attribute-Centric Anonymization Scheme for Improving User Privacy and Utility of Publishing E-Health Data," *Journal of King Saud University–Computer and Information Sciences* 31, no. 4 (Fall 2019), **https://doi.org/10.1016/j.jksuci.2018.03.014**.

[44] Kinza Yasar and Andrew Zola, "Definition: Hashing," Tech Target, May 2024, **https://www.techtarget.com/searchdatamanagement/definition/hashing**.

[45] Venkatakrishna Valleru and K. Suganyadevi, "Secure Hashing Algorithms for Protecting Sensitive Data in Cyber Environments," *15th International Conference on Computing Communication and Networking Technologies* (2024), **https://ieeexplore.ieee.org/abstract/document/10725272**.

[46] Sam Dupre, Lindsay Spell, and Paul Jung, *Geospatial Data Disclosure Avoidance and the Census: Select Topics in International Censuses* (U.S. Census Bureau, March 2022), **https://www.census.gov/content/dam/Census/programs-surveys/international-programs/stic/geospatial-disclosure.pdf**.

[47] Daniel Morales, Isaac Agudo, and Javier Lopez, “Private Set Intersection: A Systematic Literature Review,” *Computer Science Review* 49 (2023), **https://doi.org/10.1016/j.cosrev.2023.100567**.

[48] Aaron Segal, Bryan Ford, and Joan Feigenbaum, “Catching Bandits and Only Bandits: Privacy-Preserving Intersection Warrants for Lawful Surveillance,” *IEEE Symposium of Foundations of Computation Intelligence* (2014), **https://www.usenix.org/sites/default/files/conference/protected-files/foci14_slides_segal.pdf**.

[49] Sanjaikanth E. Vadakkethil Somanathan Pillai and Kiran Polimetla, “Enhancing Network Privacy Through Secure Multi-Party Computation in Cloud Environments,” *International Conference on Integrated Circuits and Communication Systems* (2024), **https://ieeexplore.ieee.org/abstract/document/10498662**.

[50] Reihaneh Torkzadehmahani et al., “Privacy-Preserving Artificial Intelligence Techniques in Biomedicine,” *Methods of Information in Medicine* 61 (June 2022), **https://pmc.ncbi.nlm.nih.gov/articles/PMC9246509**/.

[51] Trivellore E. Raghunathan, “Synthetic Data,” *Annual Review of Statistics and its Applications* (March 2021), 8:129–140, **https://doi.org/10.1146/annurev-statistics-040720-031848**.

[52] James Jordon et al., *Synthetic Data–What, Why, and How? (Royal Society*, 2022), **https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/Synthetic_Data_Survey-24.pdf**.

[53] Innovate U.K. (IUK) Business Connect, *Pandemic Response and Forecasting Technical Brief: Transforming Pandemic Response and Forecasting Through Federated Learning with End-to-End Privacy* (IUK Business Connect, 2022), **https://iuk-business-connect.org.uk/wp-content/uploads/2022/08/PETs-Prize-Challenges_-Public-Health-Technical-Brief-1.pdf**.

[54] Tab Warlitner, John O’Leary, and Sushumna Agarwal, “Data Tokenization for Government,” Deloitte Insights, November 21, 2019, **https://www2.deloitte.com/us/en/insights/industry/public-sector/chief-data-officer-government-playbook/2018/data-tokenization-for-government.html**.

[55] Antonio Muñoz et al., “A Survey on the (In)security of Trusted Execution Environments,” *Computers & Security* 129 (University of Malaga, June 2023), **https://doi.org/10.1016/j.cose.2023.103180**.

[56] Xiaoqiang Sun et al., “A Survey on Zero-Knowledge Proof in Blockchain,” *IEEE Network* 35, no. 4 (July/August 2021), **https://ieeexplore.ieee.org/abstract/document/9520375**.

[57] Jérôme Gorin, Martin Biéri, and Côme Brocas, “Demonstration of a Privacy-Preserving Age Verification Process,” Laboratoire d’Innovation Numérique de la CNIL, June 22, 2022, **https://linc.cnil.fr/demonstration-privacy-preserving-age-verification-process**.

[58] Centre for Data Ethics and Innovation, *Privacy-Enhancing Technologies Adoption Guide*, **https://cdeiuk.github.io/pets-adoption-guide**/.

[59] United Nations, *The PET Guide*, **https://unstats.un.org/bigdata/task-teams/privacy/guide/2023_UN%20PET%20Guide.pdf**.

[60] United Nations, *The PET Guide*, **https://unstats.un.org/bigdata/task-teams/privacy/guide/2023_UN%20PET%20Guide.pdf**.

[61] U.K. Information Commissioner’s Office, *Chapter 5*, **https://ico.org.uk/media/about-the-ico/consultations/4021464/chapter-5-anonymisation-pets.pdf**.

[62] U.K. Information Commissioner’s Office, *Chapter 5*, **https://ico.org.uk/media/about-the-ico/consultations/4021464/chapter-5-anonymisation-pets.pdf**.

[63] Rachel Cummings et al., “Advancing Differential Privacy: Where We Are Now and Future Directions for Real-World Deployment,” *Harvard Data Science Review* 6, no. 1 (2024), **https://doi.org/10.1162/99608f92.d3197524**.

[64] U.K. Information Commissioner’s Office, *Chapter 5*, **https://ico.org.uk/media/about-the-ico/consultations/4021464/chapter-5-anonymisation-pets.pdf**.

[65] J.M. Auñón et al., “Evaluation and Utilisation of Privacy-Enhancing Technologies—A Data Spaces Perspective,” *Data in Brief* 55 (August 2024), **https://doi.org/10.1016/j.dib.2024.110560**.

[66] Samaneh Mohammadi et al., “Balancing Privacy and Performance in Federated Learning: A Systematic Literature Review on Methods and Metrics,” *Journal of Parallel and Distributed Computing* 192 (October 2024), **https://doi.org/10.1016/j.jpdc.2024.104918**.

[67] P. Ram Mohan Rao, S. Murali Krishna, and A. P. Siva Kumar, “Privacy Preservation Techniques in Big Data Analytics: A Survey,” *Journal of Big Data* 5 (2018), **https://journalofbigdata.springeropen.com/articles/10.1186/s40537-018-0141-8**.

[68] Centre for Data Ethics and Innovation, *Privacy-Enhancing Technologies Adoption Guide*, **https://cdeiuk.github.io/pets-adoption-guide/**.

[69] “Understanding Differential Privacy,” U.S. Census Bureau, **https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/differential-privacy.html**.

[70] Rachel Cummings et al., “Advancing Differential Privacy,” **https://doi.org/10.1162/99608f92.d3197524**.

[71] J.M. Auñón et al., “Evaluation and Utilisation of Privacy-Enhancing Technologies,” **https://doi.org/10.1016/j.dib.2024.110560**.

[72] Elizabeth Renieris, “Why PETs (Privacy-Enhancing Technologies) May Not Always Be Our Friends,” *Ada Lovelace Institute* (blog), April 29, 2021, **https://www.adalovelaceinstitute.org/blog/privacy-enhancing-technologies-not-always-our-friends/**.

[73] General Data Protection Regulation, "Article 32: Security of Processing," European Parliament and Council of the European Union, 2016, **https://gdprhub.eu/Article_32_GDPR**.

[74] Christian Reimsbach-Kounatze, Taylor Reynolds, and Clarisse Girot, "Emerging Privacy Enhancing Technologies: Current Regulatory and Policy Approaches," *OECD Digital Economy Papers*, no. 351 (March 2023), **https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/03/emerging-privacy-enhancing-technologies_a6bdf3cb/bf121be4-en.pdf**.

[75] Centre for Information Policy (CIPL) Leadership, *Privacy-Enhancing and Privacy-Preserving Technologies: Understanding the Role of PETs and PPTs in the Digital Age* (CIPL, December 2023), **https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl-understanding-pets-and-ppts-dec2023.pdf**.

[76] David W. Archer et al., *UN Handbook on Privacy-Preserving Computation Techniques* (UN Global Working Group on Big Data, 2023), **https://unstats.un.org/bigdata/task-teams/privacy/UN%20Handbook%20for%20Privacy-Preserving%20Techniques.pdf**.

[77] U.K. Information Commissioner's Office, *Chapter 5*, **https://ico.org.uk/media/about-the-ico/consultations/4021464/chapter-5-anonymisation-pets.pdf**.

[78] Royal Society, From Privacy to Partnership: *The Role of Privacy-Enhancing Technologies in Data Governance and Collaborative Analysis* (Royal Society, January 2023), **https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/from-privacy-to-partnership.pdf**.

[79] Royal Society, *From Privacy to Partnership*, **https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/from-privacy-to-partnership.pdf**.

[80] "Privacy-Enhancing Technologies," Massive Data Institute at Georgetown University's McCourt School of Public Policy, **https://mdi.georgetown.edu/privacy-enhancing-technologies/**.

[81] "Software Cybersecurity for Producers and Purchasers: Software Supply Chain Security Guidance Under Executive Order (EO) 14028 Section 4e," National Institute of Standards and Technology, May 5, 2022, **https://www.nist.gov/itl/executive-order-14028-improving-nations-cybersecurity/software-cybersecurity-producers-and**.

[82] Nancy Levesque, "FPF Launches Effort to Advance Privacy-Enhancing Technologies, Convenes Experts, and Meets With White House," *Future of Privacy Forum* (blog), July 9, 2024, **https://fpf.org/blog/future-of-privacy-forum-launches-effort-to-advance-privacy-enhancing-technologies/**.

[83] "Privacy-Enhancing Technologies Prize Challenges," U.K. Centre for Data Ethics and Innovation, March 2023, **https://petsprizechallenges.com/**.