

ISSUE BRIEF

U.S. Policy

# Privacy Enhancing Technologies for State Education Agencies

An Overview for Practitioners

May 2026

## AUTHOR

**Jim Siegl**, Senior Fellow, FPF  
and  
AEM Corporation

### About This Document

This document provides a strategic framework for State Education Agencies (SEAs) and Statewide Longitudinal Data Systems (SLDS) to modernize data protection by moving privacy controls directly into the computational workflow. It categorizes a suite of Privacy-Enhancing Technologies (PETs)—including Differential Privacy, Synthetic Data, and Secure Multi-Party Computation—based on their specific roles in reducing raw data exposure during sharing, processing, and publication.

**Primary audiences:** State Education Agency staff, SLDS Staff.

---



**The Future of Privacy Forum (FPF)** is a non-profit organization that serves as a catalyst for privacy leadership and scholarship, advancing principled data practices in support of emerging technologies. Learn more about FPF by visiting [fpf.org](https://fpf.org).



All FPF materials that are released publicly are free to share and adapt with appropriate attribution. Learn more at [creativecommons.org](https://creativecommons.org).

# Table of Contents

<b>Executive Summary</b>	<b>4</b>
<b>Introduction</b>	<b>5</b>
<b>Privacy Risks in SEA/SLDS Environments</b>	<b>5</b>
<b>What PETs Can and Cannot Do</b>	<b>6</b>
A Note on Small Districts and Rare Subgroups	8
<b>Types of PETs</b>	<b>8</b>
Differential Privacy (DP)	8
Synthetic Data	9
Federated Learning (FL)	9
Trusted Execution Environments (TEEs)	10
Secure Multi-Party Computation (SMPC)	10
Homomorphic Encryption (HE)	11
Zero-Knowledge Proofs (ZKPs)	11
<b>Using PETs in SEA/SLDS</b>	<b>12</b>
<b>SEA/SLDS Use Cases</b>	<b>13</b>
Synthetic Data	13
Federated Learning	13
<b>Conclusion</b>	<b>14</b>

## *Executive Summary*

State Education Agencies (SEAs) and Statewide Longitudinal Data Systems (SLDS) link student records across programs, grade levels, and agencies to support accountability, research, and policy decisions. That linkage creates substantial value — but it also increases privacy risk, because linked data can identify students even when direct identifiers are removed. The challenge is not simply protecting data at rest or in transit; it is enabling analysis, collaboration, and public reporting while limiting how much student-level information is exposed in the process.

Privacy-Enhancing Technologies (PETs) offer practical tools for meeting that challenge. Unlike traditional disclosure avoidance approaches — suppression, rounding, and manual de-identification — PETs shift protections to the point of computation and sharing, reducing how often student-level data must be copied, moved, or distributed. They enable cross-agency computation without pooling raw records, more usable public dashboards, and research partnerships that don't require distributing student-level extracts.

But PETs are not cost-free. Each approach involves a tradeoff between privacy protection and analytical precision. Differential Privacy introduces noise that grows more distorting for small groups. Synthetic data may misrepresent rare populations. SMPC and TEEs constrain which analyses can be run. Selecting a PET means making a methodological decision — one that should be matched to the specific workflow, documented, and disclosed where outputs are published or shared. This is particularly consequential for small districts and rare subgroups, where the populations most at risk of re-identification are often those for whom PETs perform least well analytically.

This document describes the PETs most relevant to SEA and SLDS environments, explains what each can and cannot do analytically and operationally, and provides use case guidance for matching PET selection to specific workflows. PETs complement rather than replace strong governance — data minimization, least-privilege access, auditing, and enforceable agreements. The goal is a meaningfully smaller exposure surface, better-controlled outputs, and analytical workflows that agencies can stand behind.

## Introduction

Privacy-Enhancing Technologies (PETs) offer a practical way to use and analyze student data while protecting their privacy. But PETs are not only a protective layer. They enable new capabilities for data use, such as running approved computations where the data resides; supporting safe self-service analytics; and reducing the need to extract data every time a question is asked. Rather than relying solely on [traditional statistical disclosure avoidance techniques](#) such as blurring, [suppression, and data masking](#), PETs reduce exposure by changing how data is shared, computed, and released. PETs help State Education Agency (SEA) and Statewide Longitudinal Data System (SLDS) teams answer important questions using student-level data with reduced risk to the data and reduced exposure. Beyond reducing exposure, PETs enable reusable workflows that run on updated data without generating new extracts.

Traditional statistical [approaches](#), such as de-identification, small-cell suppression, rounding, and manual disclosure review, typically act at the end of the process, after data has been extracted, copied, and exposed to additional systems and users. PETs shift protections earlier in the lifecycle and help to reduce the number of times student-level data must be copied, moved, or re-packaged when supporting analysis. PETs shift protection by enabling privacy techniques during use. They can support secure multi-party analysis and federated workflows, and they can enforce stronger output protections through formal methods such as Differential Privacy. These capabilities expand beyond collaboration options. Agencies can compute shared results across organizations without pooling raw inputs, and they can publish more insights while managing disclosure risk. This allows agencies to maintain analytic utility while shrinking the risks associated with data sharing and collaboration across organizations. PETs shift privacy protections from post-hoc review to the point of computation and sharing.

## Privacy Risks in SEA/SLDS Environments

SEA and SLDS environments are both high-value and highly connected. States must publish information for the public, including data for accountability and to support legitimate research. Many state workflows depend on disclosing student-level data to analysts, program staff, contractors, or external partners. PETs enable alternative collaboration patterns that can meet many analytic needs without distributing student-level extracts to every participant. Even when granting access for authorized purposes, each copy increases the risk of accidental sharing or weak storage practices.

Linking records across different systems, including early learning, K-12, postsecondary, and workforce, can create student profiles that provide abundant information about each student. This increases the risk of reidentifying an individual. It also increases demand for more granular analysis, which can be difficult to support safely using suppression-only approaches when

stakeholders need frequent views. The more attributes that are included in the workflows, the easier it is to identify someone, even if direct identifiers like name and student ID are removed. This is especially difficult to manage because traditional disclosure avoidance was not designed for ongoing, interactive analysis over time, a new capability enabled by PETs.

SLDSs often rely on managed services and vendor-built applications to process their data. That creates access to student data by third-party providers, introducing risk. PETs can change the vendor relationship, enabling specialized tools and services without broad access to raw data. They can also improve collaboration by allowing more parties to contribute to an analysis while ensuring only approved outputs leave the controlled environment. Tables, dashboards, and research releases can reveal sensitive information when cells are small or when multiple reports allow a reasonable person to infer a hidden value. In states with many small districts or subgroups, this risk persists.

PETs enable the following capabilities for SEAs and SLDSs:

- Improve public reporting while reducing the risk that students can be identified from small groups or unique combinations.
- Enable research and evaluation that avoids exporting raw data.
- Support cross-agency analyses with fewer data transfers.
- Work with vendors to limit access to sensitive student information.
- Improve incident response by reducing the reach of a breach or an accidental disclosure.
- Enable analysis by standardizing approved computations and output rules, supporting repeatability, auditability, and faster turnaround for routine requests.

## What PETs Can and Cannot Do

The primary job of PETs is to reduce data exposure. PETs can protect sensitive data while it is being processed, not just at rest or in transit. Depending on the technology used, this could mean performing analytics on encrypted data, ensuring each collaborating party learns only what they need to conduct their analysis, or enforcing output-level privacy so results remain useful without revealing any individual student's information. And instead of relying solely on a manual review at the end, output controls support self-service analytics by limiting what can be inferred across repeated queries. PETs also enable safer collaboration across entities by supporting joint analysis without full raw-data sharing.

However, PETs are not a substitute for strong data governance and operational controls. They reduce exposure of PII, but they do not fix poor decision-making, weak policies, or careless operations. Agencies still need clear rules about allowable questions, approved workflows, and how results will be reviewed and communicated. PETs do not eliminate the need for least-privilege access, data-sharing agreements, audit logging, incident response, or disclosure review. If an organization over-collects, retains data indefinitely, misconfigures access, or allows broad internal use, PETs cannot prevent failures caused by weak administrative safeguards or

oversight. In other words, PETs can help minimize risks of analytics and sharing, but overall privacy still depends on governance, clear accountability, and secure implementation.

PETs provide more robust protection against student reidentification. They change what is exposed, when it is exposed, and what an attacker can learn from it. This also increases what agencies can safely publish and analyze because privacy protections can support repeatable releases without increasing disclosure risk. Traditional disclosure avoidance usually focuses on protecting data after it has already been extracted or at the point of publication. PETs directly aim to mitigate those risks and can keep records in the SEA/SLDS environment while also allowing partners to receive only approved aggregates or model outputs, reducing the blast radius — the scope of exposure — in the event of a breach and lowering the opportunity for linkage. This approach can also improve utility by enabling more informative outputs without using heavy suppression that removes analytic value.

Every PET involves a tradeoff between privacy protection and analytical fidelity, and that tradeoff is not uniform across methods. Differential Privacy introduces statistical noise that reduces precision — particularly for small groups, rare subgroups, and small districts, which are common in state education data. Synthetic data preserves aggregate patterns but may misrepresent tail distributions and rare populations. SMPC and TEEs preserve computational accuracy but constrain which analyses can be run and introduce operational overhead that limits iteration. Federated Learning reduces raw data exposure but produces models that may be less accurate than those trained on centralized data, particularly when district data structures are inconsistent.

For SEA and SLDS teams, the relevant question is not whether a tradeoff exists but whether it is acceptable for the specific workflow. A public accountability dashboard can tolerate controlled noise better than a program evaluation that will inform budget decisions. A system integration test can use synthetic data that wouldn't be appropriate for a longitudinal equity analysis. Matching the PET to the analytical requirement — and documenting that choice — is part of responsible governance and implementation.

PETs reduce the need to distribute student-level data at all. Traditional methods often rely on removing direct identifiers, generalizing fields, and suppressing small cells before sharing extracts. But in education data, small districts or rare subgroups can still uniquely describe students. PETs can protect computation, not just in the released dataset or final report. Traditional disclosure avoidance doesn't address this, so even if the final output is reviewed, the risk accumulated during processing can still be high.

PETs handle the linkable data in SEA/SLDS contexts better than traditional disclosure avoidance, which assumes you can remove a small set of identifiers and be safe. PET approaches can limit what is learned and what leaves the system because protections operate at the computation and output layer. But SLDS data is rich and longitudinal, which increases uniqueness and the risk of linkage. PETs are stronger methods against reidentification because they minimize the release of features that make individuals unique in the first place.

## A Note on Small Districts and Rare Subgroups

The populations that most need privacy protection in state education data are often those for whom PETs perform least well analytically. Small districts, low-incidence disability categories, and rare demographic combinations are both the most re-identifiable in traditional reporting and the most likely to be distorted by noise-based methods like Differential Privacy or underrepresented in synthetic data.

This doesn't mean PETs shouldn't be used for these populations — it means agencies should be especially careful about how PET outputs for these groups are interpreted and communicated. Results for small groups under DP should be accompanied by appropriate uncertainty language. Synthetic data used in contexts where rare populations matter should be validated specifically for those populations. In some cases, traditional suppression combined with strong governance may still be the more defensible approach for the smallest cells, even when DP is used for the broader dataset.

## Types of PETs

### Differential Privacy (DP)<sup>1 2</sup>

- **What it is:** A method of data anonymization that relies on the injection of "noise" to protect the identification of sensitive, individual data. Differential privacy is commonly used for large datasets.
- **How it works:** Adds random "noise" limiting how much a student can affect the result. The output is nearly the same whether or not a student's record is included.
- **Example:** A statewide attendance dashboard publishes chronic absenteeism by subgroup; DP helps protect small districts and rare subgroups from being reverse-engineered.
- **Considerations:** DP introduces noise that grows more distorting as group sizes decrease.

---

<sup>1</sup> National Institute of Standards and Technology. *Guidelines for Evaluating Differential Privacy Guarantees* (NIST SP 800-226).

<https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-226.pdf>

<sup>2</sup> U.S. Census Bureau. *Comparing Differential Privacy With Older Disclosure Avoidance Methods*.

<https://www.census.gov/library/fact-sheets/2021/comparing-differential-privacy-with-older-disclosure-avoidance-methods.html>

For states with many small districts or rare subgroups, this means some results may be too imprecise to support reliable conclusions at fine levels of disaggregation. The epsilon parameter governs this tradeoff and should be set before analysis runs, not calibrated post-hoc to achieve a desired result. There is no universal standard for epsilon in education reporting; agencies should document the value used and the rationale.

## Synthetic Data<sup>3</sup>

- **What it is:** Data that mimics the structure, patterns, and statistical properties of real student data but does not contain information from any actual student records.
- **How it works:** A model is trained on real student data to learn underlying patterns and relationships. It then generates new, artificial records that preserve those patterns without reproducing real student information.
- **Example:** An SLDS team provides vendors with synthetic data extracts for integration testing. Vendors can use this data to build and validate data pipelines and system integrations without accessing sensitive production student data.
- **Considerations:** Synthetic data does not automatically provide formal privacy guarantees — its protective value depends on the generation method and how well fidelity was validated. It is generally unsuitable for analyses requiring precise estimates, rare population representation, or exact counts. States with many small districts or high proportions of students in rare subgroups should validate synthetic datasets specifically for those populations before use in decision-relevant workflows, not just for aggregate statistical properties.

## Federated Learning (FL)<sup>4</sup>

- **What it is:** A method that allows multiple parties to train AI models on their data and combine identified patterns into a more accurate "global" model without sharing their data.
- **How it works:** Each state uses its own data on an AI model, and the data stays within that environment. The updates are sent out and combined with other states' updates to produce an improved shared model, which is then sent back to the state.
- **Example:** Districts train a statewide early warning model locally; the SEA aggregates updates to produce a shared model without pooling student-level training data.
- **Considerations:** Synthetic data does not automatically provide formal privacy guarantees — its protective value depends on the generation method and how well fidelity was validated. It is generally unsuitable for analyses requiring precise estimates, rare

---

<sup>3</sup> Seeman, J., Williams, A. R., & Bowen, C. M. (2025, January). *Synthetic Data for the Nebraska Statewide Workforce & Educational Reporting System*. Urban Institute. <https://files.eric.ed.gov/fulltext/ED673557.pdf>

<sup>4</sup> Fachola, C., Tornara, A., Bermolen, P., Capdehourat, G., Etcheverry, L., & Fariello, M. I. (2023). *Federated Learning for Data Analytics in Education*. *Data*, 8(2), 43. <https://www.mdpi.com/2306-5729/8/2/43>

population representation, or exact counts. States with many small districts or high proportions of students in rare subgroups should validate synthetic datasets specifically for those populations before use in decision-relevant workflows, not just for aggregate statistical properties.

## Trusted Execution Environments (TEEs)<sup>5</sup>

- **What it is:** A Trusted Execution Environment (TEE) is a [hardware-enforced isolated environment](#) where code and data are protected from access by the rest of the system — including the operating system, hypervisor, and system administrators. Data is decrypted and processed only within the protected environment; nothing outside it can read the contents while computation is running.
- **How it works:** An SEA and a partner each upload their data to secure storage, encrypted, but it is decrypted only within the enclave. It is designed to prevent most users from accessing raw student data.
- **Example:** A state offers a confidential research workspace where researchers submit approved jobs that run in a TEE-backed environment; results are released after disclosure review.
- **Considerations:** TEE-backed environments preserve computational accuracy — analysts work on real data within the enclave — but outputs still require disclosure review before release. The enclave protects data during processing; it does not automatically protect what leaves the environment. Agencies operating TEE-based research platforms should have output review procedures that are as rigorous as those applied to traditional data extracts.

## Secure Multi-Party Computation (SMPC)<sup>6</sup>

- **What it is:** A technique allowing multiple parties to process their combined data without any party needing to share all its information with the others. This approach minimizes the risk of exposing sensitive information.
- **How it works:** Instead of sharing raw student data with each other, each party keeps its input secret and only shares protected pieces that are useless on their own. The group can still get the correct final result, but no one can see anyone else's underlying data.
- **Example:** An SEA and workforce agency compute employment outcomes for program participants using SMPC, producing the metric while keeping each side's underlying records private.
- **Considerations:** SMPC requires that the computation be fully specified before it runs. Agencies accustomed to iterative or exploratory analysis will find this constraining. If the

---

<sup>5</sup> Pei, M., H. Tschofenig, D. Thaler, and D. Wheeler. 2023. *Trusted Execution Environment Provisioning (TEEP) Architecture* (RFC 9397). <https://www.rfc-editor.org/rfc/rfc9397>

<sup>6</sup> 1. Information Commissioner's Office, "Secure multiparty computation (SMPC)," ICO. <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/privacy-enhancing-technologies/what-pets-are-there/secure-multiparty-computation-smpc/>.

research question changes after the protocol is established, the computation must be renegotiated and rerun. This makes SMPC best suited to recurring, well-defined analyses rather than one-time exploratory work.

## Homomorphic Encryption (HE)<sup>7</sup>

- **What it is:** A method enabling computations to be conducted on encrypted data without needing to decrypt it first. The decrypted results match what would have been obtained by performing the calculations on the original unencrypted data
- **How it works:** An SEA encrypts student data, and a third party runs analytics without ever seeing the raw student records, computing directly on the encrypted data and returning encrypted results. Only the institution can decrypt the outputs.
- **Considerations:** The decrypted results match what would have been obtained by performing the calculations on the original unencrypted data
- **Example:** A state sends encrypted eligibility data to a service provider to compute scholarship qualification totals; the provider returns encrypted results that only the SEA decrypts.

## Zero-Knowledge Proofs (ZKPs)<sup>8 9</sup>

- **What it is:** A method that allows one party (the prover) to demonstrate to another party (the verifier) that a particular statement is true without disclosing any information beyond the statement's truth.
- **How it works:** A prover takes a secret piece of information and uses cryptography to generate a short proof showing it is true without revealing the secret. A verifier checks the proof, and if it verifies, accepts the claim even though they never saw the underlying data.
- **Example:** A state credential platform lets students prove pathway completion to an employer without disclosing full transcript history.

---

<sup>7</sup> Cheon, Jung Hee, Andrey Kim, Miran Kim, and Yongsoo Song. 2022. "Introduction to Homomorphic Encryption and Schemes." In *Protecting Privacy through Homomorphic Encryption*, 3–28. Springer. [https://link.springer.com/chapter/10.1007/978-3-030-77287-1\\_1](https://link.springer.com/chapter/10.1007/978-3-030-77287-1_1)

<sup>8</sup> Silde, Tjerand, and Akira Takahashi. 2024. "Zero Knowledge Proofs: Challenges, Applications, and Real-world Deployment." NIST Workshop on Privacy-Enhancing Cryptography. <https://csrc.nist.gov/csrc/media/presentations/2024/wpec2024-3b1/images-media/wpec2024-3b1-slides-a-kira-tjerand--ZKP-Overview.pdf>

<sup>9</sup> Goldwasser, Shafi, Silvio Micali, and Charles Rackoff. 1989. "The Knowledge Complexity of Interactive Proof Systems." *SIAM Journal on Computing* 18 (1): 186–208. [https://people.csail.mit.edu/silvio/Selected%20Scientific%20Papers/Proof%20Systems/The\\_Knowledge\\_Complexity\\_Of\\_Interactive\\_Proof\\_Systems.pdf](https://people.csail.mit.edu/silvio/Selected%20Scientific%20Papers/Proof%20Systems/The_Knowledge_Complexity_Of_Interactive_Proof_Systems.pdf)

## Using PETs in SEA/SLDS

The right PET to use depends on the analyses being done, the data needed, and the sensitivity of that data. It also depends on the agency's capabilities, such as interactive reporting, cross-agency computation, or broader access to insights without access to raw data. When conducting cross-agency program evaluation, a common approach is for an SEA/SLDS team to create a student-level extract, remove direct identifiers, apply small-cell suppression rules, and enter into a data-sharing agreement. A PET such as secure multi-party computation or a trusted execution environment, can enable the agencies to compute agreed-upon outcomes without moving raw student data across agencies. Once defined, an approved secure computation can often be reused regularly to produce updated results with consistent protections and reduced administrative overhead. This reduces linkage risk and limits what any party learns beyond what had been approved.

Interactive dashboards and public reporting typically use suppression thresholds and complementary rounding to protect data. However, a PET like differential privacy is often better aligned because it supports managing cumulative disclosure risk across repeated queries. This enables more usable dashboards by replacing blanket suppression with controlled statistical protection that remains consistent even as users study the results over time.

To protect privacy, SEAs may suppress outcomes for small racial/ethnic groups, students with rare disabilities, or small schools. This often leaves stakeholders with incomplete or unusable results. PETs can help by enabling the safer release of useful information. Differential privacy can allow publication of subgroup statistics with calibrated noise rather than blanket suppression, and synthetic data can support analytical development while limiting exposure of rare combinations in the real data. Regardless of which PET an agency selects, each approach offers more robust protections than traditional methods, provides stronger defense against re-identification, and can standardize how analyses are run and how outputs are governed when implemented properly.

When selecting among these approaches, teams should also document the analytical implications of their choice. For workflows using differential privacy, the privacy budget (epsilon) and noise parameters should be recorded and, where outputs are published, disclosed. For synthetic data workflows, fidelity validation results — particularly for rare populations and small districts — should be reviewed before outputs are used for decision-making. For SMPC and TEE deployments, the constraint that analyses must be prespecified means that exploratory or iterative analytical needs should be identified and addressed before implementation begins, not after.

## SEA/SLDS Use Cases

PETs represent a meaningful advance in how SEAs and SLDSs can protect student data during use and analysis. They enable insights with stronger security and protection for PII than the methods currently in use. PETs also expand what agencies can do with data — supporting safer repeated analysis, controlled collaboration, and more usable outputs for decision-makers. Agencies can run cross-program analyses, evaluate interventions, and support research partnerships with less reliance on raw data transfers and manual de-identification. Integrating PETs into routine workflows embeds privacy protections at the operational level, reducing reliance on after-the-fact disclosure review. PETs can also enable new access models, such as allowing more stakeholders to explore approved results through dashboards or query tools while protections are enforced automatically and consistently.

### Synthetic Data

Synthetic data is artificially generated data that [mimics the statistical properties](#) of real-world datasets without containing any records from actual individuals. By preserving the **distributions, correlations, and patterns** in the original data, synthetic sets enable robust testing and analysis while significantly mitigating the risk of re-identification.

Synthetic extracts are particularly useful when teams need realistic test data but do not need real identities or exact counts for production decisions. For example, when an SLDS team collaborates with a vendor to build a new reporting tool, rather than sharing extracts, the agency can provide a synthetic dataset that mirrors their **data architecture and logic** without exposing protected student information.

However, synthetic data does not automatically provide formal privacy guarantees, and its analytical validity depends on the generation approach. It is generally less suitable for analyses that require precise estimates, rare populations, or exact counts, and should be evaluated for both privacy risk and statistical fidelity before use.

Agencies should verify that synthetic extracts accurately represent the range of district sizes and demographic compositions in their system before sharing with vendors. A synthetic dataset calibrated to statewide averages may not adequately represent the smallest districts or rarest subgroups, which are often where data pipeline errors surface in production

### Federated Learning

When an SEAs or SLDS program is building or updating a statewide early warning model (to predict chronic absenteeism, course failure risk, dropout risk, etc.) without centralizing district-level student records, using Federated Learning may be useful. Districts may want to participate in a collaborative model, but may not be able to transfer training data to the state due

to local constraints or governance policies. The SEA or SLDS defines a shared model design and a standard set of input variables, such as recent attendance patterns, prior course performance, or enrollment context, along with training steps, performance measures, and rules for model use.

Each district trains the model locally on its own student data and sends only model updates back for aggregation. This can enable regular model refreshes and continuous improvement without requiring central collection of district training records. The aggregated model is developed and redistributed to districts until performance is stable. Federated Learning [reduces privacy risks](#) by limiting the amount of raw data shared or centralized. However, it still requires strong governance and technical mitigations as model updates may pose risk without appropriate safeguards. FL can also support model refreshes as new data arrives, allowing for continuous improvement without pooling training records.

Agencies deploying FL-based early warning models should evaluate model performance separately for large and small districts. Smaller districts contribute less training data and may receive a global model that performs worse for their student population than a locally trained one would.

## Conclusion

Governance and security basics are a necessary starting point to ensure the successful use of PETs. This would include mapping data flows, minimizing collection, enforcing the least privilege principle, and requiring strong authentication. PETs work best when an organization already has strong security practices in place — they complement, rather than replace, governance and operational controls. They can complement these rules by making certain protections enforceable through the system itself, which reduces reliance on handling these decisions as they arise. In addition, teams should plan for PET capability adoption by identifying which workflows would benefit most and defining how PET-based steps fit into existing approval and publication practices. Where applicable, agencies should also document standard PET parameters and operational practices (such as how privacy settings are selected, how outputs are validated, how results are explained to stakeholders, etc.) to ensure consistency and transparency.

Part of that documentation should address analytical tradeoffs explicitly. When an agency deploys differential privacy for public reporting, the epsilon value and noise configuration should be part of the operational record. When synthetic data is used for vendor testing, fidelity validation results should be retained. When SMPC is used for cross-agency analysis, the prespecified computation and any limitations on what could be asked should be documented alongside the outputs. These records serve both accountability and reproducibility — and they allow future analysts to understand not just what the data showed, but what the privacy protections cost analytically.

Over time, PETs can also enable more timely and useful analysis by reducing the need to re-create extracts and by supporting standardized workflows that can be rerun on new data with predictable protections. Adopting PETs should create improvements over time. The goal is not perfect privacy — it is a meaningfully smaller exposure surface, better-controlled outputs, and stronger incident readiness. These can also lead to:

- Reduced distribution of student-level data across programs, partners, and vendors.
- Fewer staff and vendors with access to sensitive data, especially outside controlled environments.
- Documented and enforced disclosure and output controls, supported by automation.
- Improved auditability and incident readiness, including clear logs, role boundaries, and recovery procedures.

SEAs and SLDSs can keep learning from data without having to spread it across every system. When PETs are used accurately and paired with strong governance, they help keep student data strongly protected, which is the ultimate goal of this work. When PETs are integrated into routine analytics, they can expand access to insights, improve reproducibility, and support responsible innovation while maintaining strong privacy protections.



Washington, DC | Brussels | Singapore

[FPF.org](http://FPF.org)