ISSUE BRIEF

US POLICY

Privacy Enhancing Technologies:

A State Education Agency Landscape Analysis

February 15, 2025

FUTURE OF PRIVACY FORUM

Authored By:

Future of Privacy Forum

1350 Eye Street NW, Ste 350, Washington D.C. 20005
www.fpf.org

AEM Corporation

11951 Freedom Dr. 20190 Ste 1100, Reston, VA, 20190
www.aemcorp.com

Table of Contents

Executive Summary

In today's data-driven world, the need for data privacy and security has never been greater, especially in the education sector, where sensitive student and institutional data are crucial for operational and accountability purposes. The increasing demand to leverage this data for critical analysis and research questions, while preserving privacy and security, has led some education agencies to consider Privacy Enhancing Technologies (PETs). These technologies promise to secure data sharing and analysis to improve learning outcomes. However, due to insufficient awareness and resources, many state education agencies (SEAs) do not consider PETs as an option.

This landscape analysis evaluates the organizational readiness and critical use cases for PETs within SEAs and the broader education sector. It provides an overview of PET adoption, current data privacy challenges, and considerations for enhancing data protection measures. The analysis highlights the need to raise awareness surrounding PETs, provide targeted assistance, and establish a greater sense of community among SEAs.

Key findings include:

- PETs are not one-size-fits-all solutions but are evolving tools aimed at enabling the sustainable utility of data without sacrificing confidentiality or security.
- There is a significant gap in technical knowledge relating to PETs.
- There is a lack of awareness of relevant use cases surrounding PETs among practitioners.
- Successful PET implementation requires substantial investment in infrastructure, technical capabilities, and ongoing training.
- Legal and regulatory requirements complicate PET adoption, with institutions often cautious about deployment due to a lack of clarity and formal guidance.

Recommendations based on our analysis include establishing a shared vocabulary, creating trusted introductory resources, and curating relevant use cases to raise collective awareness about the capabilities and limitations of PETs. It also suggests developing a PET readiness model, focusing on core capabilities, and providing targeted technical assistance to support sustainable PET adoption and implementation. By engaging in concrete, forward-looking efforts, the SEAs will become poised to leverage their data assets while effectively ensuring critical privacy protections.

Introduction

With the advancement of digital tools and infrastructure, the proliferation and usage of data across the education sector have accelerated over the past several years. From applications designed to support instruction to standard compliance reporting, education data present an area of great opportunity for exploring how those assets can inform practice, investment, and problem-solving.

However, this growth has brought considerable privacy and security risks. Those responsible for safeguarding sensitive student information are handling more data than ever while facing increased pressure to answer difficult data questions, improve transparency among partners and communities, and modernize their systems.

Additionally, the rapid deployment and integration of artificial intelligence (AI) technologies to process, store, and transfer confidential information have raised red flags. These overwhelming data demands, limited state agency capacity, and increasing regulatory scrutiny highlight the urgent need for robust privacy safeguards for this sensitive data.

One response to these circumstances involves the adoption of Privacy Enhancing Technologies, or PETs. These technologies enable education agencies and institutions to securely share, analyze, and ultimately utilize student data to improve educational and organizational outcomes. Despite the advances PETs offer to State Education Agencies (SEAs) in utilizing the data they steward, a gap persists in applying these technologies and realizing their potential benefits.

To better understand the ways PETs are or are not being integrated by SEAs, FPF has worked with AEM Corporation to conduct a landscape analysis, including an overview of current PET adoption, current data privacy challenges, and considerations for enhancing data protection measures. The findings in this document highlight the need to raise awareness among SEAs of what PETs are and what they are not, the range of available types of PETs, their potential use cases, and considerations for the effective adoption and sustainable implementation of these technologies. SEAs need a better understanding of how PETs could augment data privacy and security ecosystems and where limitations exist. PET implementation can boost community trust and enhance data analysis benefits when done with intention. Proactively investing in PETs and applying best practices offers an opportunity to address privacy issues and prepare for secure future data use.

This landscape analysis was designed to gain insight into current PET usage across SEAs by examining existing literature and research, including academic papers, reports, and case studies that contributed to a broad picture of the state of PET implementation.

To contextualize these findings, we collected qualitative data via focus groups of state education leadership, staff, and partners and interviews with industry leaders and subject matter experts. The findings presented here identify successes, suggest approaches to the challenges shared by participants, and shed light on the common points of confusion and barriers to adoption that participants discussed.

Overview: Privacy Enhancing Technologies

Definition

Privacy Enhancing Technologies, or PETs, are a collection of tools and methods designed to safeguard individual privacy by protecting sensitive personal data, concealing individual characteristics, or preventing the unintended disclosure of confidential information.[1] These evolving technologies aim to ensure data privacy and protection while maintaining the utility of results yielded from analyses.[2]

Although no universally accepted definition of PETs exists,[3] traditional data privacy techniques in education, such as the use of trusted third parties, access controls, and statistical disclosure limitations (e.g., aggregation, rounding, and cell suppression), are not recognized as PETs. Education agencies commonly utilize such techniques to ensure authorized access to and proper de-identification of student data. PETs, however, take a more technologically advanced approach and may supplement or even replace conventional methods.

Types: Input vs. Output

PET classification varies across agencies and sectors. Such technologies may be classified based on various characteristics (e.g., technical application, sophistication, use case), each possessing unique strengths and limitations. While the potential effectiveness of these technologies may be realized in similar institutional, organizational,

[1] Heurix, J., Zimmermann, P., Neubauer, T., & Fenz, S. (2015). A taxonomy for privacy enhancing technologies. Computers & Security, 53, 1-17.
[2] The Royal Society, "From Privacy to Partnership: The Role of Privacy Enhancing Technologies in Data Governance and Collaborative Analysis," January 2023.
[3] Shen, Y., & Pearson, S. (2011). Privacy enhancing technologies: A review. Hewlett Packard Development Company.

or technical conditions, it is essential to note that PETs within and across classifications are not interchangeable for all use cases.

For this analysis, based on existing educational research, we recognize the PET categorization of input privacy and output privacy.[4] **Input privacy** refers to methods to mitigate unauthorized access or inappropriate use when accessing or sharing data. **Output privacy** relates to methods used to minimize the risk of re-identification in data analysis results or data products built from the data set (i.e., figures, tables). Below are common types of PETs:[5]

Input Privacy

- Homomorphic Encryption: A method enabling encrypted 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.
- Trusted Execution Environment (TEE): Secure virtual computing spaces that enable the execution of code and access to data in an isolated manner, detached from the rest of the system. This isolated processing protects against unauthorized access, ensuring the confidentiality and integrity of sensitive data,[6] also known as a secure enclave.
- Secure Multiparty Computation: 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.
- Federated Learning: 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.
- Zero-Knowledge Proof: 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.

[4] O'Hara, Amy, & Straus, Stephanie (2022). Privacy Preserving Technologies in Education. Massive Data Institute, Georgetown University.
[5] Information Commissioner's Office, UK (2023). Privacy Enhancing Technologies. 1.0.5
[6] Adams, S., Gray, S., Massey, A., & van Eijk, R. (2024). Confidential Computing and Privacy - Policy Implications of Trusted Execution Environments (p. 2). Future of Privacy Forum.

Output Privacy

- Differential Privacy: 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 in large data sets.
- Synthetic Data: Artificial data created by data synthesis algorithms that mimic the patterns and statistical properties of real data, including personal data, producing comparable results to those obtained from analyzing the original data set.

The following section discusses the strengths and limitations of these PETs, which any organization considering adoption should consider.

Strengths

Within the education sector, discussions surrounding the utilization and efficacy of PETs to protect sensitive information while preserving the ability to derive valuable insights have gained traction. With educational providers increasingly leveraging digital platforms and analytics to drive improvements in learning outcomes, PETs provide a means through which personal data can be kept secure without impacting the utility of those data. This means that the data can be used to answer essential questions without compromising their security and privacy, which builds trust and transparency between parents, students, and educational institutions.

For example, the input privacy measure of **homomorphic encryption** allows computation on encrypted data,[7] while the output privacy measure of **differential privacy** injects controlled noise into data sets or query results to protect individual identities while providing accurate aggregate insights.[8] These two technologies work together to protect personal data from storage to transmission and processing stages for analytic work. Such PETs also enable more secure collaboration and research across institutions and departments.

Other examples include secure multiparty computation and federated learning. **Secure multiparty computation** lets multiple parties jointly compute functions over a dataset without revealing the actual inputs.[9] While utilizing **federated learning**, shared AI models can be trained on decentralized data so that institutions can contribute to and benefit

[7] Hallman, R. A., Diallo, M. H., August, M. A., & Graves, C. T. (2018, March). Homomorphic Encryption for Secure Computation on Big Data. In IoTBDS (pp. 340-347).
[8] Bowen, C. M., & Garfinkel, S. (2021). Philosophy of differential privacy. Notices of the American Mathematical Society, 68(10), 1727-39.
[9] Alghamdi, W., Salama, R., Sirija, M., Abbas, A. R., & Dilnoza, K. (2023). Secure multi-party computation for collaborative data analysis. In E3S Web of Conferences (Vol. 399, p. 04034). EDP Sciences.

from collective insights without the actual flow of raw data from one environment to another.[10] Provided that such disclosure is permitted under federal and state student privacy laws, these PETs may be particularly useful in academic environments where collaboration leads to significant improvements in educational outcomes and contributions.

Further, as trust among students and stakeholders is highest when they are sure their information is secured, **PETs can offer a chance to innovate in the education sector**.[11] Because PETs can provide security and privacy infrastructure that allows testing of new analytic methodologies while maintaining privacy, institutions can work with advanced technologies to improve instruction, learning, and organizational outcomes. So, while the sector continues to evolve towards data-driven approaches, PETs stand out as fundamental elements that can help make education data ecosystems secure, efficient, and effective by respecting and protecting the privacy of all stakeholders.

Limitations

PETs offer significant benefits to the education sector, but specific PETs differ in their application and implementation. One such difference is the **difficulty and cost of deploying and maintaining these technologies**.[12] Many PETs require advanced technical capabilities and investment in infrastructure, which may be a crucial factor for an institution operating on a constrained budget or with low resource potential. Such conditions would make PET adoption challenging, especially for organizations with an IT or data support shortage or those needing more budgets to upgrade technology.

California's Cradle-to-Career (C2C) is actively engaged in plans to utilize PETs, including both input and output privacy measures, to further ensure equitable, secure data access and analysis.

Ready accessibility challenges also limit the use of PETs for education, as the additional layers of security and privacy make analysis more complex for those not experienced in working with these technologies.[13] This may mean staff require additional training and time, which might

[10] Li, L., Fan, Y., Tse, M., & Lin, K. Y. (2020). A review of applications in federated learning. Computers & Industrial Engineering, 149, 106854.
[11] OECD (2023). Privacy enhancing technologies. Topics, Policy sub-issue.
[12] Information Systems Audit and Control Association (2024). Exploring Practical Considerations and Applications for Privacy Enhancing Technologies. Resources, Whitepaper.
[13] Information Commissioner's Office, UK (2023). Privacy Enhancing Technologies. 1.0.5

otherwise be directed elsewhere. As a result, if not carefully managed, PET implementation may inadvertently hinder the data-driven insight they are designed to enable. Any organization considering implementation of PETs should carefully plan for the "learning curve" these technologies may present for their staff.

Another barrier to adoption is the **challenge of integrating PETs into existing systems**. Most educational institutions depend on systems and software that may be cumbersome or incompatible with new PETs. Integrating these newer technologies into established infrastructures can be extremely difficult because it involves not only technical hurdles but also resources that may be stretched by the demands of daily operations.

Legal and regulatory requirements associated with the application of PETs are also complicating factors.[14] While these technologies stand to play an active role in enhancing privacy and compliance in the future, the lack of clear guidance and standards can confuse potential adopters. Educational institutions may remain cautious about fully deploying certain PETs due to worries about non-compliance or legal consequences until clear standards and guidelines are established.

PETs can potentially improve various aspects of privacy and data security in the education sector while enabling those data for value-added analyses. Still, the implementation challenges must be addressed. Substantial costs, technical complexity, accessibility issues, integration difficulties, and unclear regulations all contribute to how and why organizations decide to pursue these technologies. Addressing these challenges requires stakeholders, policymakers, and educational institutions to work together to ensure that PETs can represent a net gain to the education data ecosystem regarding providing both privacy and data utility in the long term.

The State of States

Given the above backdrop, we aimed to better understand PET awareness, usage, challenges, and needs among state education agency data practitioners nationwide. Through a purposive sample of state data leaders and industry experts, we used a focus group approach to investigate three significant domains related to PETs:

1. Understanding and Awareness
2. Implementation Considerations

[14] United Nations (2023). United Nations Guide on Privacy-Enhancing Technologies for Official Statistics. United Nations Committee of Experts on Big Data and Data Science for Official Statistics, New York.

3. Constraints

Understanding and Awareness

Across focus groups, participants were asked to indicate their understanding and awareness of PETs, whether out in the space or within their respective agencies and institutions. Confusion and uncertainty were consistently expressed about what constitutes a PET and if certain privacy safeguards currently in place would be categorized as a PET. Most of these techniques were determined to be statistical disclosure limitation methods, such as aggregation, rounding, and cell suppression, or trusted data intermediaries or trusted third parties. Many of these active approaches were reported to be custom solutions. Those with limited PET experience or knowledge expressed interest in learning more about relevant use cases and applications.

The most widely discussed PETs of interest were secure enclaves (or TEEs), secure hashing, differential privacy, encryption, and synthetic data. Experience with these technologies varied. In multiple instances, secure enclaves and secure hashing were indicated as operational or in development in existing statewide longitudinal data and P-20W+ systems. In additional instances, encryption was reported to have been implemented specifically to translate students' unique IDs, either independently or through a TTP.

Washington's Education Research & Data Center (ERDC) has effectively layered a secure enclave and traditional SDLs, like suppression and top and bottom-coding, to safeguard sensitive data.

Implementation Considerations

Adopting technological advancements at scale in any industry can increase efficiencies and protections while also exposing institutions to internal pressures and external risks. PET implementation in the education sphere is no exception. The participants in this analysis shared implementation lessons learned related to the **integration, usability, and effectiveness of PETs**.

The motivation and rationale for deploying such technologies in states were diverse. Legislation and leadership influence as drivers were notable catalysts in instances of successful or in-progress implementation. Participants emphasized that this backing

further reinforced team expectations in alignment with their agencies' strategic goals, thereby ensuring successful adoption.

In one instance of state legislative funding, the adopted privacy-enhancing methods could only partially satisfy the initial need due to the potential exposure of sensitive data to trusted third parties or external users. However, despite this mismatch between legislative expectations and technical reality, the implemented technologies have become valuable, trusted avenues for secure data sharing. This underscores the complementary nature of PETs in a broader data ecosystem while providing a caveat to adopters that non-trivial work remains to bridge the gap between understanding PETs and their perceived effectiveness.

Further, participants emphasized the value of assessing, securing, and continually monitoring the breadth of resources required to engage in effective change. One participant, recounting timeline-related barriers in developing a systematic data suppression process, stressed the value and benefit of operational transparency when considering such technologies. In response to issues with a TTP's ability to execute the work, another shared that substantial time, effort, and energy had to be invested in the planning and training processes for the data teams who were part of the implementation.

Participants shared that existing privacy, infrastructure, and data governance frameworks may be strengthened by PETs only if implemented thoughtfully and deliberately. Furthermore, all participants highlighted the importance of assessing current processes that may be impacted before adoption. It was advised that PET implementation and maintenance necessitated the modification of several existing practices and operations, representing substantial change management efforts.

> *In my realm, awareness of PETs is the biggest problem. Even learning about what other states have done and are custom, would be helpful.*
>
> *- State Data Privacy Manager*

Constraints

The most prevalent impediments to PET adoption and implementation were a lack of awareness and resources, whether time, funding, or capacity. In all instances, privacy safeguards were operational; however, participants were unclear if the protective

measures already constituted a PET or how implementation would impact their existing infrastructure and staffing.

Implementation experiences with PETs and custom solutions ranged from abandoned to emerging and sustainable. Across successful implementations, leadership buy-in and legislation significantly contributed to securing the necessary resources. All participants with successful implementations emphasized the need for time and resources to upskill and reskill staff to ensure stable adoption and maintenance. When PET adoption was abandoned or ongoing, participants stressed the technical capacity and personnel resources required for ownership and sustainability, regardless of upfront cost.

> *You need technical capacity to do the work but there is also how much bandwidth there is in house to sustain things that are built. It's a major concern. Even with a grant or cost is zero, we still need capacity to support putting in place, rollout, and maintenance.*
>
> *-State Chief Data Officer*

Perceptions of these methods also varied, though hesitancy persisted regarding the utility of such technologies to effectively and accurately produce results in response to data demands. For those with preexisting knowledge, a negative association was observed between the level of PET sophistication and the degree of comfort in adoption. Higher levels of sophistication generally resulted in lower levels of expressed comfort. The techniques discussed in this case were differential privacy and synthetic data.

Differential Privacy

Participants expressed concerns about the level of "noise" (protection) infused with data representing unique populations and the resulting utility and communicability of the results. The level of noise applied is often independent of the data set, leading to less accurate data when dealing with smaller populations. The overwhelming consensus of those with preexisting knowledge of differential privacy was that its viability is better realized on larger datasets. Thus, the value of the yielded results may be maintained. Additionally, as differential privacy relies on complex algorithmic logic to obscure attributes, communicating analytic results to audiences without this technical expertise requires an ongoing and significant investment (e.g., time, technical capacity, expertise).

Synthetic Data

Regarding synthetic data, participants shared responses similar to those about differential privacy. Skepticism about the utility of synthetic data was predominantly a

result of the high level of effort required to produce them (i.e., depending on the data set, extensive time and resources are necessary to generate comparable results properly) and the quality of the results (i.e., whether synthetic data on populations, rather than samples, provides actionable insights to influence policy). However, the consistently expressed level of interest in improving understanding of synthetic data and its use cases was notable. Participants indicated having observed an appetite amongst states and institutions for increased knowledge sharing and collaboration explicitly related to synthetic data.

Practitioners lack a body of technical knowledge and relevant use cases surrounding these tools. The lack of awareness of existing solutions, whether open source or at cost, coupled with the complexity and perceived trade-off of more sophisticated methods, often results in adoption being deprioritized or abandoned entirely.

The **Kentucky Center for Statistics (KYSTATS)** uses generated synthetic data for training purposes with the added benefit of allowing for analyses developed using synthetic data to be re-run on the original data. This internal validation process promotes greater insight into the accuracy and utility of synthetic data.

Structural Considerations

A common refrain was that non-PETs were currently being adequately leveraged to support agencies' needs. Participants **shared concerns about PETs' credibility, return on investment to implement, and necessity**. In multiple instances, participants noted that the benefits of such technologies could likely be realized through other means, such as capacity building, or outweighed entirely by investing in different areas.

Ideology and autonomy were also identified as prevalent constraints. The political and historical landscape were major factors when pursuing implementations that have statewide benefits and repercussions. Participants shared that getting such efforts off the ground is difficult without the necessary knowledge and relevant use cases to garner leadership buy-in.

Recommendations

PETs are not one-size-fits-all, nor a 'magic wand.' Instead, they are an evolving suite of sophisticated solutions to enable the sustainable utility of data without sacrificing confidentiality or security. However, with a limited understanding of PETs among many practitioners and even fewer resources or opportunities for sharing knowledge about their relevance and application in education, these technologies are often overlooked in

favor of more traditional methods. This illuminates the need to **raise collective awareness, consider targeted assistance, and establish a sense of community**. The following sections lay out avenues to respond to those needs.

Raise Collective Awareness: *Out of sight, out of mind.*

To begin enabling PET adoption at scale, states call for a body of knowledge supporting the viability, usability, and sustainability of such methods at a practical level. Furthermore, states are often navigating in silos, hindering those leading data efforts from knowing (1) whether PETs are operational in their environment and (2) the specific function and benefit of adopted solutions. This disconnect may be partially addressed through a shared vocabulary surrounding PETs, a trusted body of resources, and a collection of commonly recognizable use cases.

Recommendation: Establish a Shared Vocabulary

There is a need to establish and reinforce consistency among PET terminology to streamline understanding. Regardless of role, interest level, or implementation status, states were more easily discouraged when we did not speak the same language.

Recommendation: Steward Trusted Introductory Resources

Agencies would benefit from a repository of high-level materials from trusted sources rooted in this common vocabulary, defining PETs and how they can be leveraged. This repository should include realized benefits, emerging risks, and role-based impact within education agencies and institutions. The included materials should be informative but digestible to non-expert audiences (i.e., one-pagers and infographics before technical briefs).

Recommendation: Collect and Curate Use Cases

Where available, relevant education use cases that examine real-world applications from adoption to sustainability should be identified and shared. While much PETs technology is not new, it is evolving. States expressed they do not want to be “left behind” or accrue more technical debt. In state education, spotlighting those currently experiencing the benefits of implementation is invaluable to securing buy-in.

It is particularly important to highlight use cases that encourage consideration of PETs as opposed to custom solutions. Custom solutions to ensure data privacy and confidentiality are widespread among states, as they are perceived as more manageable. There is a shared sense of comfort and predictability with

technologies like stored procedures and complex coding. Use cases can identify instances where the results yielded from traditional methods can be improved via PETs, noting open-source offerings' availability, transparency, and customizability.

Such a collection of use cases could highlight the opportunity and benefit of taking a multitiered position with privacy protections by layering traditional with more sophisticated methods. Emphasize PETs' complementary nature alongside other safeguards to further build trust and reinforce PETs as *a* solution rather than *the* solution.

> *The best PET would be statistical literacy. To the capacity point, I don't need a technology to better suppress my data. I have more use for an analyst who could crunch my data and deliver better service to my customers.*
>
> *-State Education Data Manager*

Consider Targeted Assistance: *What problem are we trying to solve?*

To fully adopt something new, users must have confidence in that technology. Considering the limited hours, dollars, and technical staff that an organization might have at its disposal, and the competing priorities they must juggle daily, PET adoption can quickly transition from interesting to impossible. States understand and appreciate the importance of a sustainable data governance model and robust security framework. However, state officials advised they need a better sense of how and where PETs fit. Therefore, a logical initial step is to assess states' current internal processes, structures, and architectures to inform the critical path to PET adoption and implementation.

Recommendation: Develop a PET Readiness Model

State officials would benefit from development of a guided tool to engage and assist states in assessing current institutional, organizational, and technical conditions to weigh potential implementation better. In recognition of the varying maturity of systems within and across states (i.e., the starting line is different for all), any such tool's components must be accessible and reflect industry best practices to determine the relevance and feasibility of adoption.

Recommendation: Focus on Capabilities

Success requires sustainability. PET adoption is one thing; ownership is another. Looking beyond buy-in, state practitioners are concerned about the sustainability of these technologies once in place, given persistent capacity challenges. This,

however, presents an opportunity to identify the core PET capabilities an organization can focus on to target their training efforts. By leveraging a PET Readiness Model and attending to the capabilities required to improve readiness, prospective and active implementers can clearly target their training efforts towards those that will best enable them.

Recommendation: Engage in Universal Technical Assistance
State education agencies, post-secondary institutions, researchers, and partners must engage in virtual opportunities to connect on PETs. Outside of introductory content, officials at these institutions would benefit from shared discussion around potential topics such as the distinction between PETs and traditional disclosure avoidance techniques, open source versus at-cost solutions, and panel discussions highlighting lessons learned from the field.

Recommendation: Explore Targeted Opportunities
Once established, organizations should leverage the PET Readiness Assessment and the PET Capability Model to develop and offer targeted technical assistance for various audiences. These engagements should be organized around a user group (e.g., researchers, state privacy experts) and the shared capability or area of concern (e.g., data linkage barriers, data sharing challenges) to connect the dots and inform future offerings.

Establish a Sense of Community: *Take a people-first approach.*

Numerous existing initiatives and communication vehicles within and across education agencies aim to further collective innovation. As PETs continue to be tested and deployed in states, creating and leveraging space to bring more perspectives and expertise will help generate trust and a greater sense of community.

Recommendation: Establish a Community of Practice
Conversations about PETs are already happening within states. Still, a dedicated space is needed to further these conversations. Practitioners require opportunities to discuss adoption experiences, the prevalence of regulatory obligations on implementation, and the feasibility of additional technical integrations via AI. A community of practice or similarly operating structure would allow states to hear their PET experiences and interests from each other. This approach has proven valuable with state education agencies across parallel efforts.

Recommendation: Leverage Existing Networks

Decision fatigue is apparent in the context of PETs. A concerted effort should be made to increase the visibility of PETs and their application within and across ongoing efforts in order to lessen the impact on leadership and decision makers, including the Chief Information Officer Network, the Chief Privacy Officer Network, the P-20W+ Community of Innovation, and the SLDS teams.

Conclusions

Throughout this landscape analysis, it is abundantly clear that the successful adoption and implementation of privacy-enhancing technologies (PETs) hinges on increased fundamental awareness and understanding of these technologies. There is an appetite for these tools and methods in the education space. However, with adequate direction and guidance, from agency leadership and policymakers, interested agencies and institutions are subject to further growing pains. At the same time, those uninterested or opposed to these innovations are incentivized to learn more about them through exposure to successful use cases leveraging existing networks. Prioritizing the growth of a working knowledge on PETs across the education sector will allow for increased opportunities for information sharing, both strategically and organically.

The bottom line is that change is hard and should not be undertaken for its own sake. This analysis illuminates that the cost of inaction does not reliably move the needle in the presence of institutional, political, technical, and cultural barriers. As with adopting any major technological innovation, a catalyst is often required. By focusing efforts toward building PET competency in states through consistent, targeted engagement and technical assistance, a more informed and motivated user group emerges – one that *can* reliably move the needle.

Appendix: Use Cases

Maryland Synthetic Data Project

The Synthetic Data Project was an effort to determine whether synthetic data could be a solution to expand access to confidential data while protecting individual privacy. The Synthetic Data Project (SDP) team collaborated with the Maryland Longitudinal Data System (MLDS) Center to come up with goals that test the feasibility of creating and using synthetic data for MLDS Center research.[15] The four main goals are outlined below:

- **Goal 1**: Create three gold standard datasets (GSDSs) that cover K12 to postsecondary education, postsecondary education to the workforce, and K12 education to the workforce
- **Goal 2**: Generate multiple sets of synthetic data based on the GSDS
- **Goal 3**: Disseminate information about the MLDS Center's synthetic data via a summit for education and workforce researchers
- **Goal 4**: Examine the feasibility of using synthetic datasets for cluster-level inference analysis

Meeting these goals requires a deep understanding of the data structure and characteristics of the variables within the MLDS. Many external and internal resources were needed to determine if the synthetic data was usable. They found that the synthetic datasets created may not always be helpful for all research questions of interest. Researchers are interested in a variety of cohorts or groups based on the research project, and many MLDS data elements are variables with limited, discrete values.[16] The SDP team was paving the way, which meant there was a need to hire many resources with specific skill sets to fill the gap they needed to have in-house. This required hiring consultants, a challenging and costly process within state universities.

Once the synthetic datasets were implemented, administration was required to track downloads and use of the platform storing the artificial data. This needed to be factored in to maintain and follow best practices. Stakeholder involvement was critical to test and ensure the GSDs and synthetic data were usable throughout the project. In addition, the SDP team encountered pushback from some MLDS Center Governing Board members,

[15] Maryland Longitudinal Data System Center. (2024). *Synthetic Data Project*. Maryland.gov.
[16] Henneberger, A., Wooley, M. E., Gillaspy, K., & Stapleton, L. (2024, September). Maryland's Synthetic Data Project Outcomes. SLDS.ed.gov.

specifically related to concerns about losing control over research and analysis being conducted using state data.

In conclusion, synthetic data was found to be a viable strategy for connecting researchers to SLDSs while ensuring data privacy. The SDP team suggests a detailed review of state and federal law before creating synthetic data. Leader buy-in is critical to starting and sustaining a project of this magnitude. Knowing these lessons from Maryland can help states prepare and get ahead of facing some of these challenges as they plan to explore synthetic data as a solution.

U.S. Census Bureau

The U.S. Census Bureau must balance accurate data collection while maintaining effective privacy protections. These two goals can be challenging to achieve because increased data usability can inadvertently increase the disclosure risk for an individual's identity, while strengthening privacy protections may affect data accuracy. The Bureau has applied disclosure limitation techniques such as data swapping since 2000 to prevent disclosure of individual respondents. However, with increased computing power and access to external data from other sources, the danger of reidentification urged the Census Bureau to use enhanced privacy protection.

Differential privacy (DP) is the disclosure avoidance method that the Bureau is shifting to today.[17] DP injects "noise" or controlled randomness into datasets in such a way that anonymizes individuals yet retains aggregate accuracy at higher levels. The Bureau has issued demonstration datasets treated with DP, allowing comparisons with data previously published. Responses have identified inconsistencies in household data and increased variability in rural areas compared to urban areas. Longitudinal studies and smaller racial groups are the most affected, adding complications as to how to bring in such privacy methods while ensuring that the data remains usable.

The Nebraska Statewide Workforce & Educational Reporting System (NSWERS)

The Nebraska Statewide Workforce & Educational Reporting System (NSWERS) is a collaborative data system that tracks students from preschool through the workforce. Established in 2020, NSWERS integrates data from Nebraska's K-12 education, higher

[17] *Understanding differential privacy*. (2024). Census.gov.

education institutions, and workforce agencies, including over 1 million records on student outcomes across various stages of education and employment.

NSWERS began exploring synthetic data – artificially generated data that mimics actual data – to address challenges in data sharing, privacy, and collaboration. NSWERS tested synthetic data through pilot projects, starting with datasets on high school graduation rates and time to employment.[18] These datasets include predictors like high school GPA and postsecondary GPA. To ensure the synthetic data accurately reflects relationships in the original data, NSWERS uses evaluation tools from the Urban Institute, including metrics for statistical consistency and relationships between variables.

Key lessons from the pilot include the importance of technical assistance, the benefits of open-source tools, and the need for leadership support. NSWERS also faced challenges, such as defining use cases, preparing data for synthesis, and handling complex data structures spanning multiple domains (K-12, postsecondary, and workforce).

Ultimately, NSWERS aims to expand the use of synthetic data to make its education-to-workforce data more accessible and useful for internal and external stakeholders. The project highlights the potential of privacy-preserving technologies to improve collaboration while maintaining data security.

[18] O'Hara, A., Straus, S., & Deschapelles, C. (2024). *Synthesizing workforce and education data using an open source tool: lessons learned*. Massive Data Institute, Georgetown University, McCourt School of Public Policy.

1350 EYE STREET NW | SUITE 350 | WASHINGTON, DC 20005

info@fpf.org | FPF.ORG