Modernizing Education Data Systems through Privacy Enhancing Technologies (PETs)

Stephanie Straus, M.Ed.
Amy O'Hara, Ph.D.
McCourt School of Public Policy
Georgetown University

FCSM October 25, 2023

Session D-3: New Perspectives and Methods on Privacy and Disclosure Control

Education Data Ecosystem

What Protects Privacy Now?

- Lockdowns (don’t trust anyone)
- Trusted third parties (limited/earned trust)
- Contracts (licenses, NDAs, MOUs)
- Statistical disclosure controls
  - Rounding, swapping, suppression, etc.

What Are Privacy Enhancing Technologies (PETs)?

- Cryptographic techniques that increase data protection while allowing for greater data utility
- Can enhance how data are analyzed and/or published
- Can complement or replace other statistical disclosure limitation methods

PETs are safer and more secure ways to analyze, link, and share data

Two Aspect of Privacy

- Input privacy
  - Data access or sharing challenges
  - Reduces risks of unauthorized access or inappropriate use

- Output privacy
  - Results of data analysis, such as information in tables or graphs
  - Reduces the risks of re-identification of data subjects

Is Anyone Using PETs?

Table 3. PPTs in Education

PPT Type	Project	Ongoing?
secure hashing	Birth through Eight Strategy for Tulsa (BEST) with Tulsa Public Schools and Oklahoma Policy Institute Asemio	Yes
secure hashing	Oregon Integrated Client Services Data Warehouse (ICS) Oregon Departments of Education, Health, Human Services, and others	Yes
secure hashing	SILK hash encoding with Administration for Children and Families and education agency partners Georgia Policy Labs	Yes
secure hashing	Hashed matching algorithm for Virginia Longitudinal Data System* University of Virginia Social and Decision Analytics Lab	No
SMC	Boston Women's Workforce Council gender pay gap study Boston University	Yes
SMC	National Post-secondary Student Aid Study-Federal Student Aid linkage demonstration National Center for Education Statistics (NCES) and Georgetown University	No
SMC	Virginia Longitudinal Data System - Defense Advanced Research Projects Agency (DARPA) demonstration State Council of Higher Education for Virginia	Yes

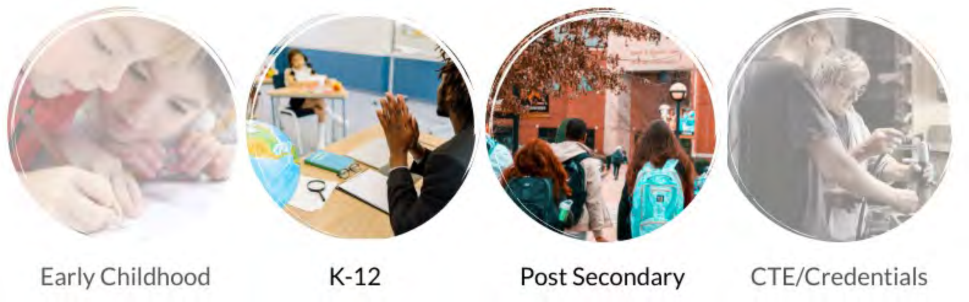
Table 3. PPTs in Education

PPT Type	Project	Ongoing?
TEE	Federated data model with joins on demand for Education Providers Silicon Valley Regional Data Trust	No
TEE	Secure data enclave for research access to student school district records on student social, emotional, academic, and physical well-being Character Lab, University of Pennsylvania	Yes
TEE	Education Research Data Center, WA Departments of Children, Youth, and Families, State Board of Education, and others State of Washington	Yes
TEE	LearnLab's DataShop, world's largest repository of learning interaction data, spun off from Cognitive Tutors program LearnLab, Carnegie Mellon University	Yes
TEE	Secure virtual data enclave for research access to NCES Restricted Use Files Institute of Education Sciences, Coleridge Initiative	Yes
TEE	Secure virtual enclave for research access to safety, health, and outcomes data on children Children's Data Network, University of Southern California	Yes
DP	Post-Secondary Employment Outcomes (PSEO) Census Bureau, multiple post-sec. institutions	Yes
DP	College Scorecard IRS, Statistics of Income Division, Department of Education, Tumult Labs	Yes

Conducted 40 stakeholder interviews to identify existing and abandoned projects

Most Common PETs

- Institutions were using:
  - Secure multiparty computation
  - Secure hashing
  - Secure enclave/trusted execution environment
  - Differential privacy
  - Synthetic data



Secure Multiparty Computation (SMC)

Secure multiparty computation (SMC): the process by which two distrusting parties jointly compute a research query on their datasets, without ever seeing the other's underlying data, through encryption.	
☑ only aggregate results released ☑ descriptive statistics ☑ finds overlap in datasets ☑ no trusted third party sees data	☒ time-consuming ☒ limited in operations/statistics ☒ requires careful data preparation ☒ does not address output privacy

- Education examples: Estonia, Virginia, our NCES demonstration
- Other examples: Boston Women's Workforce, Allegheny County Department of Human Services demonstration, DARPA and IARPA investments

Differential Privacy (DP)

Differential privacy (DP): a method for obscuring identities or attributes in the underlying record-level data by infusing results/statistics with noise.	
☑ reduces re-identification risks for individuals or groups in the data (i.e., students, programs) ☑ provides a formal privacy guarantee (can guard against threats known today and those in the future) ☑ useful for known queries	☒ challenging to implement on low levels of geography or unique population groups without adding a lot of noise ☒ tradeoff between privacy and accuracy – as you add more "noise" (protection) you move further from true values ☒ no input privacy

- Education examples: U.S. Census Bureau's Post-Secondary Employment Outcomes, College Scorecard
- Other examples: Census 2020, Google, Apple, Facebook

What Are Barriers to PET Deployment?

Legal	Institutional	Technical	Cultural
• Actual legal barriers • Perceived/claimed legal barriers • Data sharing agreements • Not enough Yes lawyers	• Politics/political cycle • Protectionism • Inertia/no demand for change • No resources to test/implement • Lack of expertise • No guidance from feds	• Untested/unavailable tools • Untrusted software • No standards • Slow compute • Skills gap	• Lack of PET understanding • Lack of examples • Lack of trust in service providers • No incentive to change

Next Steps

- PET information and training sessions
- Develop guidance
- Field Building
- Demonstration projects
  - Synthetic data with Nebraska Statewide Workforce & Educational Reporting System
  - Secure query system with IRS
- In discussion: Secure enclave and multiparty computation with state ed. agencies


GEORGETOWN UNIVERSITY
McCourt School of Public Policy
MASSIVE DATA INSTITUTE
amy.ohara@georgetown,edu
stephanie.straus@georgetown.edu
Read our report,
Privacy Preserving Technologies in Education