**Motivation**

**Computational Social Science**
The potential: massive new sources of data and ease of sharing will revolutionize social science.

**Challenges for Sharing Sensitive Data**

Complexity of Law
- Thousands of privacy laws in the US alone, at federal, state, and local levels, usually context-specific.
- HIPAA, FERPA, CIPESA, Privacy Act, PPRA, ESRA.

Difficulty of Deidentification
- Stripping "PII" usually provides weak protections and/or poor utility.

Inefficient Process for Obtaining Restricted Data
- Can involve months of negotiation between institutions, original researchers.

**Vision**

**Target: Data Repositories**

**Approach: Integrated Privacy Tools**

**Bridging Law & CS Definitions of Privacy**

*Argue that Differential Privacy Satisfies FERPA and other privacy laws via two arguments:*

1. **A technical argument supported by a technical argument**
   - The FERPA privacy standard is relevant for analyses computed with DP

2. **A technical argument supported by a legal argument**
   - Differential privacy satisfies the FERPA privacy standard
   - FERPA allows dissemination of de-identified information sufficient to show that DP satisfies outcome that is not identifiable
   - Extract a mathematical definition of privacy from FERPA and provide a mathematical proof that DP satisfies this definition

**Other Accomplishments**

- Many theoretical results illustrating the limits of differential privacy (lower bounds, algorithms, hardness results, attacks).
- Theoretical and empirical work bridging differential privacy & statistical inference (confidence intervals, hypothesis testing, Bayesian posterior sampling).
- Framework for modern privacy analysis: catalogue privacy controls, identify information uses, threats, and vulnerabilities, and design data programs that align these over data lifecycle.

**Co-PIs & Senior Personnel**

- Debbie Yehia, co-PI, CRCS & Georgetown
- James Honaker, Sr. Researcher, CRCS
- Michal Alon, co-PI, MIT
- Steve Chong, co-PI, CRCS
- Marco Gaboardi, University of Buffalo
- David O'Brien, Sr. Researcher, Berkman Klein Center
- João Viana, co-PI, CRCS
- Edoardo, co-PI, IQSS
- Steve Chong, co-PI, CRCS
- Marco Gaboardi, University of Buffalo
- David O'Brien, Sr. Researcher, Berkman Klein Center
- Letanya Swainley, co-PI, IQSS
- Edward Arel, co-PI, Harvard Data Dept
- Gary King, co-PI, IQSS
- Maria GobBIARDI, University of Buffalo
- David O'Brien, Sr. Researcher, Berkman Klein Center

**Goals of PSI**

• General-purpose: applicable to most datasets in repository.
• Automated: no differential privacy expert optimizing algorithms for a particular dataset or application.
• Tiered access: DP interface for wide access to rough statistical information, helping users decide whether to apply for access to raw data (cf. Census PUMS vs RDCs).

**Privacy Budgeting Interface**

*Privacy Definition: effect of each individual must be "hidden"*

*Digital Privacy Interface:"Meets all Theory and Practical Definitions of Differential Privacy (DP) and satisfies ε=4.500000, δ=0.000001, 2016.*

**Integration w/Statistical Tools for Social Science**

*PSI (Ψ): a Private data plateau*

*PSI (Ψ) at Privacy (TPDP)*

*Integrated for research in social science and other human subjects research fields.*

*Training in multidisciplinary research: = 100 students, postdocs, interns from law, computer science, social science, statistics.*


*Numerous workshops and symposia organized, including public symposium "Privacy in a Networked World" with 700+ registrants.*

*New journal "Technology Science" utilizing DataTags*

*Open-access pedagogical materials on data privacy for many audiences.*

**Disclaimer:**

*Data Owners:
- Kevin Wang, Obasi Alexandra Wood*

*Data Owner formalization and design data programs that align these over data lifecycle.

*Bayesian posterior sampling). Statistical inference (confidence intervals, hypothesis testing, privacy (lower bounds, algorithms, hardness results, attacks).*

*Bayesian inference (confidence intervals, hypothesis testing, privacy (lower bounds, algorithms, hardness results, attacks).*

*Bayesian inference (confidence intervals, hypothesis testing, privacy (lower bounds, algorithms, hardness results, attacks).*

*Bayesian inference (confidence intervals, hypothesis testing, privacy (lower bounds, algorithms, hardness results, attacks).*

*Bayesian inference (confidence intervals, hypothesis testing, privacy (lower bounds, algorithms, hardness results, attacks).*

*Bayesian inference (confidence intervals, hypothesis testing, privacy (lower bounds, algorithms, hardness results, attacks).*

*Bayesian inference (confidence intervals, hypothesis testing, privacy (lower bounds, algorithms, hardness results, attacks).*

*Bayesian inference (confidence intervals, hypothesis testing, privacy (lower bounds, algorithms, hardness results, attacks).*