

# Privacy Tools for Sharing Research Data



## Challenges for Sharing Sensitive Data

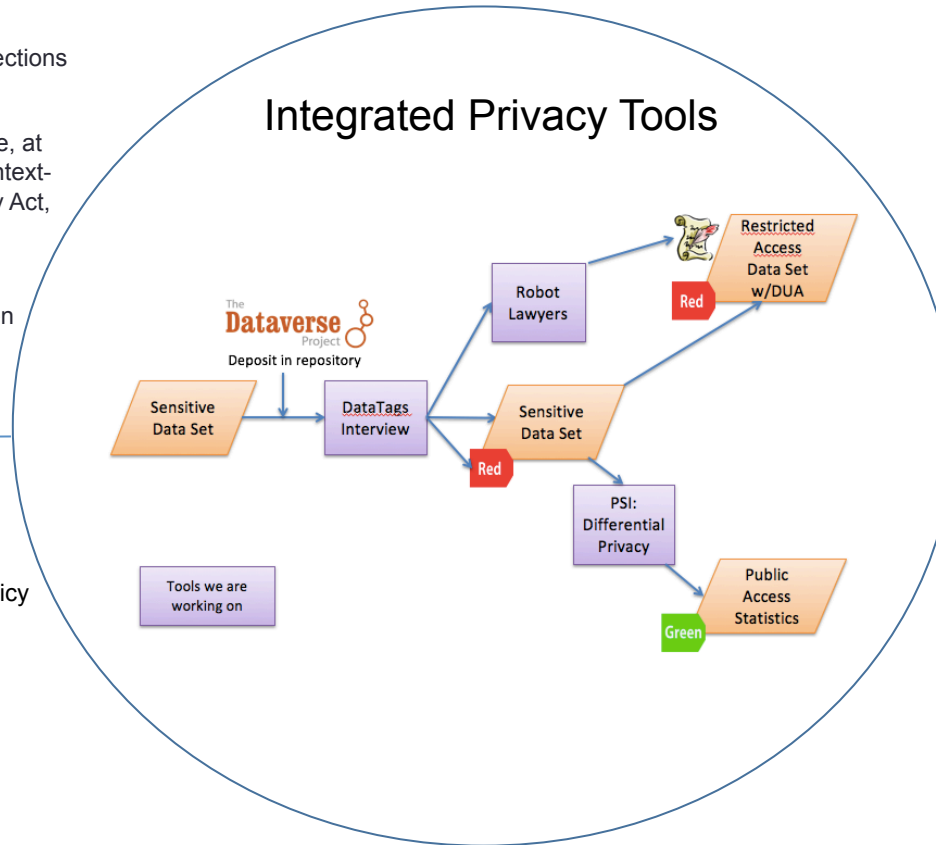
- **Difficulty of de-identification**  
Stripping “PII” usually provides weak protections and/or poor utility.
- **Complexity of law**  
Thousands of privacy laws in the US alone, at federal, state, and local levels, usually context-specific: HIPAA, FERPA, CIPSEA, Privacy Act, PPRA, ESRA, etc.
- **Inefficient process for obtaining restricted data**  
Can involve months of negotiation between institutions and original researchers.

## Solution:

An array of computational, legal, and policy tools to make privacy-protective data-sharing easier for researchers without expertise in privacy law/CS/stats:

- **Differential privacy** for enabling statistical analysis with strong privacy guarantees.
- **DataTags** for defining data handling policies.
- **Bridging law and CS** definitions of privacy.

## Integrated Privacy Tools



## Scientific Impact:

- Enable **wider and safer sharing of research data**.
- Advance science through **replication and new discoveries**.
- Numerous theoretical results illuminating **limits of differential privacy**.
- Differentially private **statistical inference**.
- Framework for **modern privacy analysis**

## Broader Impact:

- **Infrastructure for research** in social science and other human subjects research fields
- **Training in multidisciplinary research**:  $\approx 100$  students, postdocs, interns from law, computer science, social science, statistics
- **Policy impact**: White House Big Data Privacy Study, National Privacy Research Strategy, NIST 800-188 De-identifying Government Datasets, Federal Trade Commission
- **Numerous workshops and symposia organized**, including public symposium “Privacy in a Networked World” with 700+ registrants.

This project is a collaborative effort between Harvard's Center for Research on Computation and Society, Institute for Quantitative Social Science, Berkman Klein Center for Internet & Society, Data Privacy Lab, and MIT Libraries' Program on Information Science. It is funded by National Science Foundation Award CNS-1237235