

---

**Project Information: *Applying Theoretical Advances in Privacy to Computational Social Science Practice***

---

**Principal Investigator (lead organization):**

Micah Altman  
Massachusetts Institute of Technology  
E25-131, 77 Massachusetts Ave.  
Cambridge, MA 02139  
(585) 466-4224, [escience@mit.edu](mailto:escience@mit.edu)

**Grantee Organization:** Massachusetts Institute of Technology

**Amount Requested:** \$846,704

**Requested Start Date:** April 1, 2015

**Requested End Date:** September 30, 2017

**Principal Investigators (sub-awardees):**

Mercè Crosas, Urs Gasser, Gary King, Salil Vadhan

---

**Project Goal**

---

We are motivated by the opportunities in social science created by massive new sources of data and developments in data analysis and sharing, and by the threat that privacy concerns pose to realizing the full potential of social science research. By leveraging ongoing multidisciplinary collaborations and theoretical advances in computation, statistics, law, and social science, the proposed project aims to improve reuse and replicability in empirical social science. Our goal is to develop and extend integrated privacy-preserving tools for enabling access to and use and disclosure of social science data, supported by rigorous computational, institutional, and legal foundations.

---

**Objectives**

---

- Analyzing the institutional and stakeholder incentives for managing research data privacy and the policy consequences of implementing new computational and legal privacy tools and concepts.
  - Designing a blueprint for securing large-scale confidential archival data in the Dataverse repository.
  - Exploring applications of our new computational and legal privacy tools to massive data and selected uses cases, including online education data, human subjects research data, and economic data protected by NDAs.
  - Expanding research collaborations to engage with other differential privacy and privacy law experts, ongoing data privacy and dissemination efforts at MIT and Harvard, and several related Sloan projects.
- 

**Proposed Activities**

---

- Engagement of post-doctoral researchers and visiting scholars in the development of game-theoretic models, data analysis, algorithm development, and legal, institutional, and stakeholder analysis.
  - Events such as workshops, working meetings, or focus groups with relevant stakeholders, which might include social science researchers, representatives from edX, commercial firms, and IRBs at MIT, Harvard, and others.
- 

**Expected Products**

---

- Additional practical computational tools for improving privacy and utility of data dissemination.
  - Revised metadata schemas, legal agreements, and draft regulatory language developed based on our research collaborations and study of the data privacy use cases.
  - Integration of legal and computational privacy tools with the Dataverse repository.
  - Blueprints for adding support for massive data and the data privacy use cases to the Dataverse repository.
  - Theoretical advances in our understanding of privacy, informed in part by the data privacy use cases studied.
  - Publications in peer-reviewed journals in law, computer science, applied statistics, and policy.
- 

**Expected Outcomes**

---

We expect this project not only to advance the state of data privacy research, but to enable practical improvements in the replicability of, access to, and confidentiality of big data produced by academic, commercial, and government entities.