PRIVACY TOOLS FOR SHARING RESEARCH DATA

CONTACT DETAILS:
NAME: Salil Vadhan, Michael Wojcik , Micah Altman
ORGANISATION: Harvard University , MIT
EMAIL: - salil@seas.harvard.edu, wojcik@seas.harvard.edu, escience@mit.edu

Objectives:

The Privacy Tools for Sharing Research Data project is a broad, multidisciplinary effort to help enable the collection, analysis, and sharing of personal data for research in social science and other fields while providing privacy for the data subjects. The project is led by the Center for Research on Computation & Society, the Institute for Quantitative Social Science, and the Berkman Center for Internet & Society at Harvard, and collaborators include Microsoft Research, and the Program on Information Science at the Massachusetts Institute of Technology. The project received seed funding from Google and is now supported primarily as a Frontier project in the NSF Secure and Trustworthy Cyberspace Program.

Bringing together computer science, social science, statistics, and law, the investigators seek to refine and develop definitions and measures of privacy and

data utility, and design an array of technological, legal, and policy tools for social scientists to use when dealing with sensitive data. The project will develop secure implementations of these algorithms and legal instruments, which will be made publicly available and used to enable wider access to privacy-sensitive data sets a through the Dataverse Network system.

On-going activities:

Privacy Tagging Infrastructure: The project is developing a "privacy tags" framework for automatically determining how to handle a privacy-sensitive dataset when it is deposited into a repository such as the dataverse network. A privacy tag is an iconic label assigned to a data set to assess its privacy risk and access level. A tag is generated as a result of an automated interview which gathers information on the contents of the data set and the agreements that apply to that data set, and checks them against privacy laws. The Dataverse Network will serve as an initial implementation of infrastructure to support privacy tags.

Legal Research: The legal team in engaged research efforts on three principal areas: producing a comprehensive database of federal and state privacy laws and regulations that bear on the sharing of research data, conducting an extensive review of literature and practice materials, and reviewing the current state of the art in contractual approaches to protecting privacy and assigning liability in data use

Policy. The project participated in a variety interdisciplinary activities focused on bridging legal and mathematical definitions of privacy and/or influencing policy processes regarding privacy.

Results:

Although the project is in its initial stages, it has produced over fifteen articles and reports that include work examining the limits of anonymization for specific databases, and the capabilities of disclosure limitation methods. The project has also produced educational material ranging including a a hands-on practical introduction to confidential data management for researchers. Selected articles and course material are available through on the project website.

In September, the project will be bringing together leading experts in computer science, social science, law and health sciences to review the state of the art in data privacy research, and to discuss how approaches from these different disciplines should be integrated in the context of real-world use cases that involve the management of confidential research data.

Over the next year, the project aims to release through the website a comprehensive database of federal and state privacy laws and regulations that bear on the sharing of research data. The project also aims to release software extensions for tagging and managing confidential data, through the open source Dataverse Network system.

URL: http://privacytools.seas.harvard.edu/