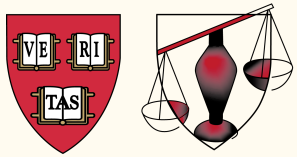
Framework for a Modern Privacy Analysis

David O'Brien & Alexandra Wood
Berkman Center for Internet & Society



Presentation for the Privacy Tools Project NSF Site Visit
October 19, 2015

with Micah Altman, MIT Libraries & Brookings Institution
Salil Vadhan, SEAS
Urs Gasser, Berkman Center

Background

- A rough framework emerged from policy commentaries we previously submitted to federal agencies (e.g., HHS, OSHA, FTC, OSTP)
- We refined the framework through participation in the Berkeley Center for Law & Technology symposium *Open Data: Addressing Privacy, Security, and Civil Rights Challenges* (April 2015)
- Research summarized in a new article to be published in the *Berkeley Technology Law Journal* this fall

Article overview

- Responds to efforts by government agencies to increase the amount of data they release to the public while taking steps to protect privacy
- Explores data privacy and utility in examples of real-world data releases
- Dissects and applies recent advances in data privacy from computer science, statistics, and law
- Proposes a framework for a modern privacy analysis informed by concepts we are investigating through the Privacy Tools project

Methodology

Use case analysis of four broad categories of government data releases

- Freedom of information and Privacy Act requests
- Traditional public and vital records
- Official statistics
- E-government and open government initiatives

Focusing on three dimensions

- Types of information released
- Standards for making release decisions
- Privacy interventions in use

Gaps identified in government data releases

1. Most agencies address privacy by withholding or redacting records that contain certain pieces of directly or indirectly identifying information.

2. Agencies lack formal guidance for choosing among and implementing privacy interventions in specific cases.

3. Similar privacy risks (or even identical data) are treated differently by different government actors.

These gaps demonstrate the need for
a more systematic approach to privacy analysis.

Framework for a modern privacy analysis

Modeled on information security and lifecycle approaches

1. Developing a catalog of privacy controls
2. Identifying information uses, threats, and vulnerabilities
3. Designing data releases by aligning uses, vulnerabilities, and threats with controls—at each stage of the information lifecycle (collection/acceptance, transformation, retention, access/release, post-access)

Catalog of privacy controls

- Procedural, technical, educational, economic, and legal means for enhancing privacy—at each stage of the information lifecycle

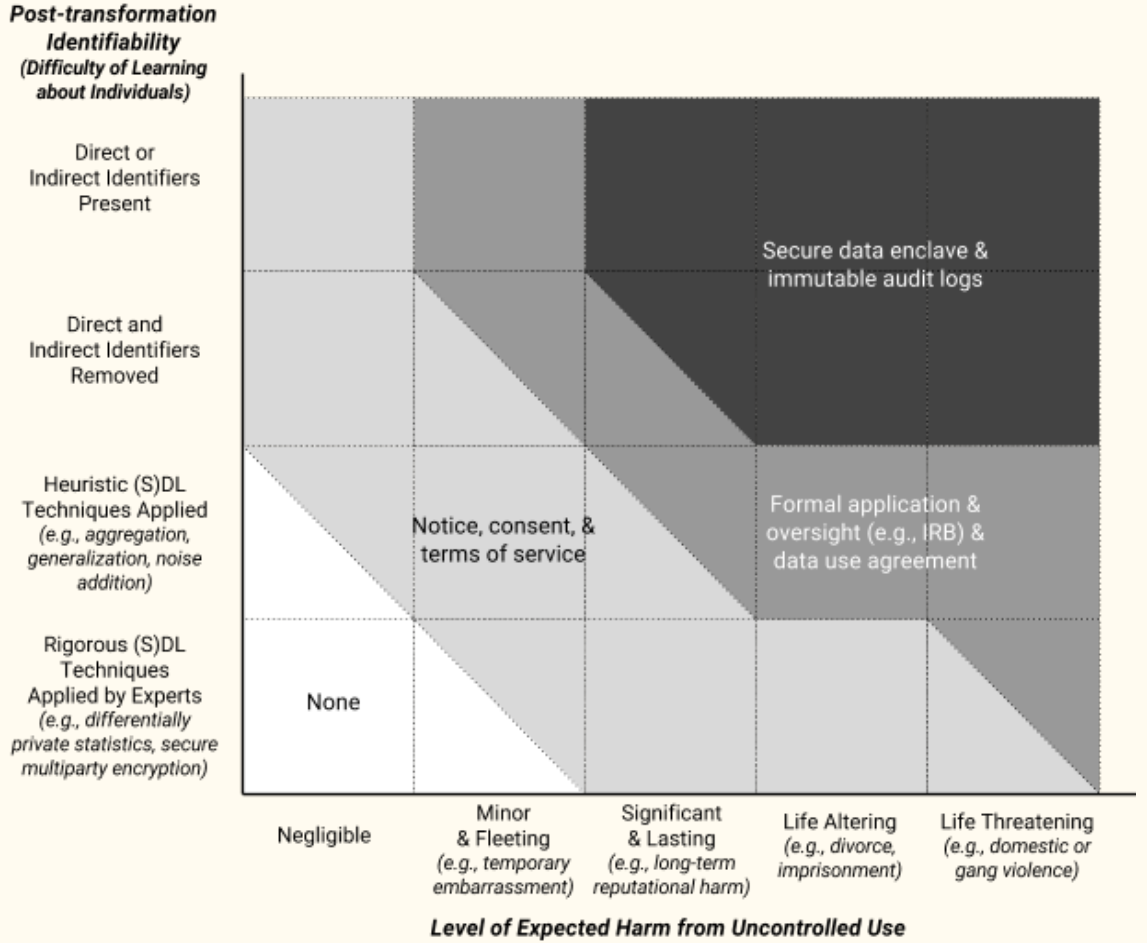
	Procedural	Economic	Educational	Legal	Technical
Access/Release	Access controls; Consent; Expert panels; Individual privacy settings; Presumption of openness vs. privacy; Purpose specification; Registration; Restrictions on use by data controller; Risk assessments	Access/Use fees (for data controller or subjects); Property rights assignment	Data asset registers; Notice; Transparency	Integrity and accuracy requirements; Data use agreements (contract with data recipient)/ Terms of service	Authentication; Computable policy; Differential privacy; Encryption (incl. Functional; Homomorphic); Interactive query systems; Secure multiparty computation

Identifying threats, vulnerabilities, and utility

- **Threats** are defined broadly as potential adverse circumstances or events that could cause harm to a data subject as a result of inclusion of the subject's data
- **Harms** are defined as injuries sustained by data subjects as a result of a threat being realized
- **Vulnerabilities** are defined as characteristics that increase the likelihood that threats will be realized
- **Utility** is defined broadly as the analytical value of the data

Guide to selecting privacy controls

Illustrating how to choose privacy controls that are consistent with the uses, threats, and vulnerabilities at each lifecycle stage



Application of the framework

- Analyzing two specific data release cases
  1. Public release of workplace injury and illness records
  2. Boston and Seattle municipal open data portals
- Identifying gaps and misalignments between the intended uses, threats, vulnerabilities, and controls at each information lifecycle stage
- Discussing how privacy controls can be better tailored to the specific threats and vulnerabilities, as well as the needs of different types of users

Aligning uses, threats, and vulnerabilities

Tiered modes of access with embedded review, accountability, and redress mechanisms could bring gains in both privacy and utility if properly implemented and support a broad range of uses across different types of data

- Public access to contingency tables and visualizations
- Intermediate access through a privacy-aware model server for interactive analysis
- Restricted access to minimally redacted data through a virtual data enclave under the terms of a DUA

Related work

- Micah Altman, Alexandra Wood, David R. O'Brien, Salil Vadhan, and Urs Gasser, **Towards a Modern Approach to Privacy-Aware Government Data Releases**, 30 *Berkeley Technology Law Journal* __ (2015, forthcoming).

- Micah Altman, David R. O'Brien, Salil Vadhan, and Alexandra Wood, **Comments to the White House Office of Science and Technology Policy Re: Big Data Study: Request for Information** (2014).

- David R. O'Brien, Jonathan Ullman, Micah Altman, Urs Gasser, Michael Bar-Sinai, Kobbi Nissim, Salil Vadhan, and Michael Wojcik, **When is Information Purely Public?**, Working Paper (2015).

- Alexandra Wood, David R. O'Brien, Micah Altman, Alan Karr, Urs Gasser, Michael Bar-Sinai, Kobbi Nissim, Jonathan Ullman, Salil Vadhan, Michael Wojcik, **Integrating Approaches to Privacy Across the Research Lifecycle: Long-Term Longitudinal Studies**, Working Paper (2014).

Preprints available from http://privacytools.seas.harvard.edu

Future plans

- Upcoming presentation at NYU-Berkeley *Conference on Responsible Use of Open Data: Government and the Private Sector* (November 2015)
- Utilizing and building on framework in other works in progress (e.g., Privacy in Long-term Longitudinal Studies article and pedagogical document on privacy, de-identification, and differential privacy)
- Opportunities to expand on framework in future policy commentaries, including current Common Rule NPRM