Privacy and Security in Distributed Data Markets

Daniel Alabi, Sainyam Galhotra, Shagufta Mehnaz, Zeyu Song, Eugene Wu

SIGMOD 2025 Tutorial

Part 4: Regulatory Considerations

Agenda

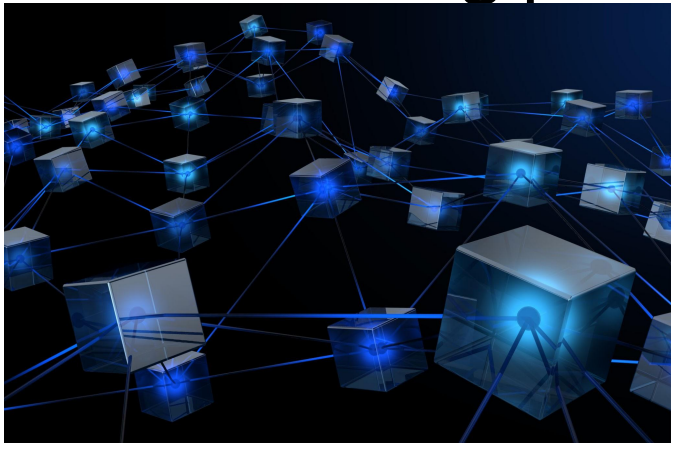
- Background and Motivation
- Legal Landscape: Key Frameworks
- Data Ownership and Control
- Consent, Purpose Limitation, and Transparency
- Security and Breach Notification Requirements
- Cross-border Data Flows
- Technical-Legal Interplay

Why Legal Considerations Matter

- Legal frameworks define data rights and duties
- Distributed data markets complicate governance
- Legal compliance is essential for trust and adoption
- Liability concerns for data market platforms

Compliance in Distributed Data Markets

- Federated learning in healthcare
- Blockchain-based data exchanges
- *Key challenge*: Trust among parties with differing incentives



Legal Foundations (Global Overview)

- GDPR (EU)
- CCPA/CPRA (California), HIPAA (US), Title 13 (US)
- APPI (Japan), LGPD (Brazil), others
- Varying consent, data definitions, cr
rules

Title 13 and U.S. Census Bureau

Title 13, U.S. Code

The Census Bureau is bound by Title 13 of the United States Code. These laws not only provide authority for the work we do, but also provide strong protection for the information we collect from individuals and businesses.

Title 13 provides the following protections to individuals and businesses:

- Private information is never published. It is against the law to disclose or publish any private information that identifies an individual or business such, including names, addresses (including GPS coordinates), Social Security Numbers, and telephone numbers.
- The Census Bureau collects information to produce statistics. Personal information cannot be used against respondents by any government agency or court.
- Census Bureau employees are sworn to protect confidentiality. People sworn to uphold Title 13 are legally required to maintain the confidentiality of your data. Every person with access to your data is sworn for life to protect your information and understands that the penalties for violating this law are applicable for a lifetime.
- Violating the law is a serious federal crime. Anyone who violates this law will face severe penalties, including a federal prison sentence of up to five years, a fine of up to $250,000, or both.

Source: https://www.census.gov/history/www/reference/privacy_confidentiality/title_13_us_code.html

Data Ownership and Stewardship

- Ambiguity in data and model ownership
- Data Controller vs. Data Processor roles
- Tension between legal rights and cryptographic control

Informed Consent and Purpose Limitation

- *Legal*: Data minimization and purpose restriction
- *Technical*: Enforcing purpose in distributed systems
- **Example**: Purpose-bound smart contracts

Transparency and Auditability

- *Legal*: User access and audit rights
- *Technical*:

1. Differential privacy audits
2. Audits of Zero-Knowledge proofs
3. Blockchain log audits

Security Obligations

- Requirements for ‘appropriate’ security
- Breach notification mandates (e.g., GDPR Art. 33)
- *Techniques*: Secure computation, encryption

Cross-Border Data Transfer Challenges

- Data localization laws
- GDPR adequacy and conflict-of-law
- Federated approaches to mitigate legal friction

Legal-Technical Interplay

- Differential Privacy ≠ Legal Anonymization
- Encryption: Legal access vs. secure storage
- ZKPs for provable compliance

Differential Privacy: A Primer for a Non-Technical Audience

*Alexandra Wood, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, James Honaker, Kobbi Nissim, David R. O'Brien, Thomas Steinke & Salil Vadhan**

ABSTRACT

Differential privacy is a formal mathematical framework for quantifying and managing privacy risks. It provides provable privacy protection against a wide range of potential attacks, including those

* Alexandra Wood is a Fellow at the Berkman Klein Center for Internet & Society at Harvard University. Micah Altman is Director of Research at MIT Libraries. Aaron Bembenek is a PhD student in computer science at Harvard University. Mark Bun is a Google Research Fellow at the Simons Institute for the Theory of Computing. Marco Gaboardi is an Assistant Professor in the Computer Science and Engineering department at the State University of New York at Buffalo. James Honaker is a Research Associate at the Center for Research on Computation and Society at the Harvard John A. Paulson School of Engineering and Applied Sciences. Kobbi Nissim is a McDevitt Chair in Computer Science at Georgetown University and an Affiliate Professor at Georgetown University Law Center; work towards this document was completed in part while the Author was visiting the Center for Research on Computation and Society at Harvard University. David R. O'Brien is a Senior Researcher at the Berkman Klein Center for Internet & Society at Harvard University. Thomas Steinke is a Research Staff Member at IBM Research – Almaden. Salil Vadhan is the Vicky Joseph Professor of Computer Science and Applied Mathematics at Harvard University.

This Article is the product of a working group of the *Privacy Tools for Sharing Research Data* project at Harvard University (http://privacytools.seas.harvard.edu). The working group discussions were led by Kobbi Nissim. Alexandra Wood and Kobbi Nissim are the lead Authors of this Article. Working group members Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, James Honaker, Kobbi Nissim, David R. O'Brien, Thomas Steinke, Salil Vadhan, and Alexandra Wood contributed to the conception of the Article and to the writing. The Authors thank John Abowd, Scott Bradner, Cynthia Dwork, Simson Garfinkel, Caper Gooden, Deborah Hurley, Rachel Kalmar, Georgios Kellaris, Daniel Muise, Michel Reymond, and Michael Washington for their many valuable comments on earlier versions of this Article. A preliminary version of this work was presented at the 9th Annual Privacy Law Scholars Conference (PLSC 2017), and the Authors thank the participants for contributing thoughtful feedback. The original manuscript was based upon work supported by the National Science Foundation under Grant No. CNS-1237235, as well as by the Alfred P. Sloan Foundation. The Authors' subsequent revisions to the manuscript were supported, in part, by the US Census Bureau under cooperative agreement no. CB16ADR0160001. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the Authors and do not necessarily reflect the views of the National Science Foundation, the Alfred P. Sloan Foundation, or the US Census Bureau.

Research Questions

- Can we cryptographically enforce legal policies?
- What counts as legally sufficient anonymization?
- Consent revocation

  in distributed systems?

Case Study in Health Analytics

- Health data federation across countries
- *Legal constraints*: HIPAA + GDPR
- *Risks*: Breach, cross-jurisdiction enforcement

Takeaways

- Legal frameworks shape privacy/security protocols
- Legal compliance ≠ technical privacy
- Must align PETs with regulatory requirements
- Learn about compliance from lawyers!!!