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

Goal: Overview (but not exhaustive!)

- Background and Motivation
- Legal Landscape: Key Frameworks
- Translating Frameworks to Implementations
- Security and Breach Notification Requirements
- Cross-border Data Flows
- Technical-Legal Interplay

Story: Why Legal Considerations are Important

OPPORTUNITY INSIGHTS

Case Study: What counties/states in the U.S.A have better or worse economic/social mobility?

https://opportunityinsights.org/

Changing Opportunity: Class and Racial Gaps in Economic Mobility

EXPLORE THE PAPER

REVIEW KEY FINDINGS ↗
OPPORTUNITY ATLAS: NEW TREND DATA ↗

Story: Why Legal Considerations are Important

Case Study: What counties/states in the U.S.A have better or worse economic/social mobility?

Solution: Use statistical methods and quantitative social science to study the question.

https://opportunityinsights.org/

OPPORTUNITY INSIGHTS

Changing Opportunity: Class and Racial Gaps in Economic Mobility

Story: Why Legal Considerations are Important

"…There are several steps in our estimation process. We begin by combining three sources of [...] data linked by and housed at the Census Bureau (the 2000 and 2010 Decennial Census short forms; federal income tax returns for 1984, 1989, 1994, 1995, and 1998-2019; and the 2000 Decennial Census long form and the 2005-2019 American Community Surveys) to construct an analysis sample of Americans born between 1978-1992. We map these individuals back to the counties where they lived as children and measure their outcomes at age 27 (between 2005-2019). Parent and child income are measured using their percentile ranks in the national income distribution…."

Source: https://opportunityinsights.org/policy/frequently-asked-questions/

Story: Why Legal Considerations are Important

OPPORTUNITY INSIGHTS

Case Study: What counties/states in

the U.S.A

have better economic/social mobility?

First, collect raw data from IRS, Census Bureau.

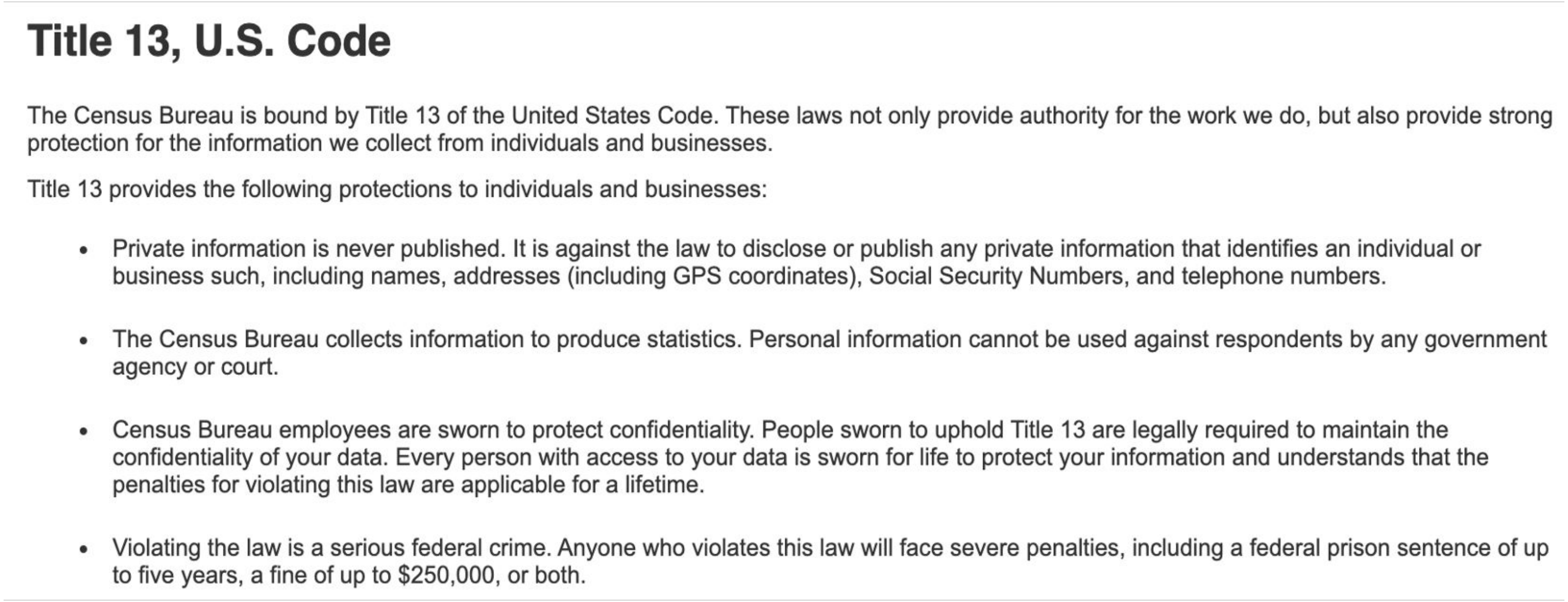
But Census Bureau needs to adhere to

Title 13.

Screw up, then employees go to jail!

Changing Opportunity: Class and Racial Gaps in Economic Mobility

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.

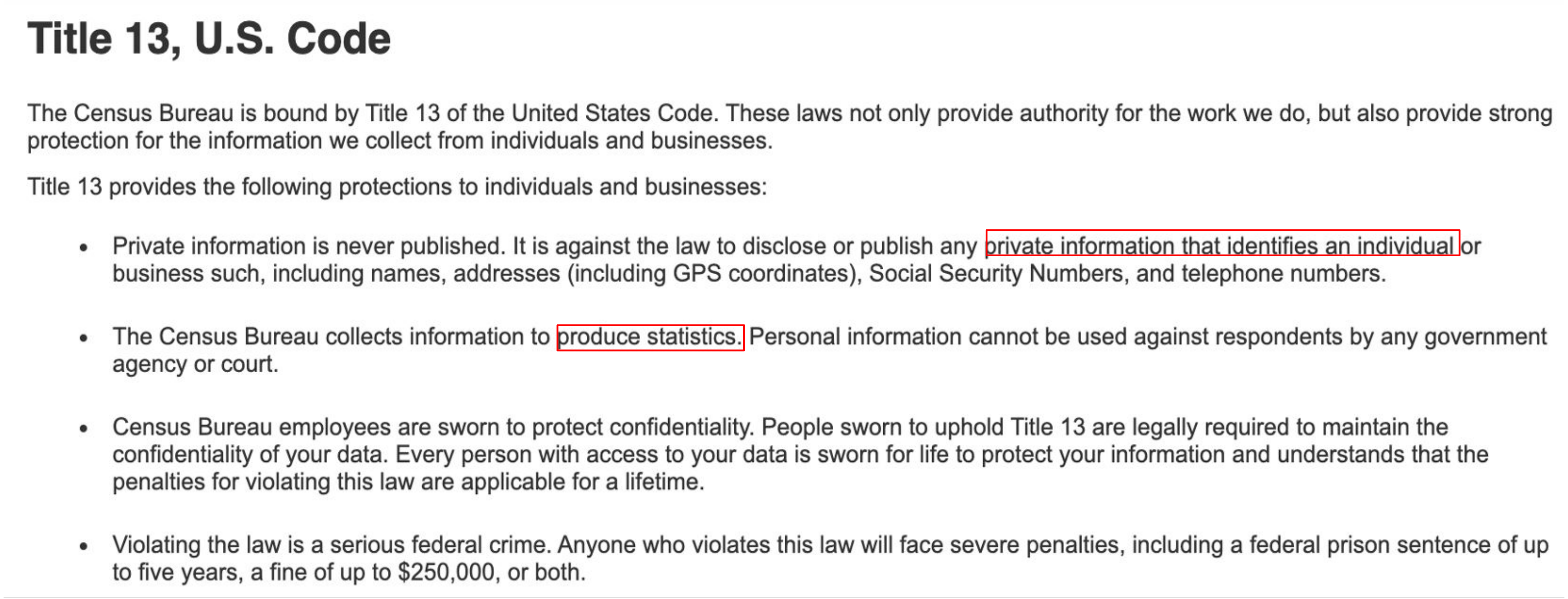
- The Census Bureau collects information to produce statistics. Personal information cannot be used against respondents by any government agency or court.

- 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

Title 13 and U.S. Census Bureau



Title 13, U.S. Code

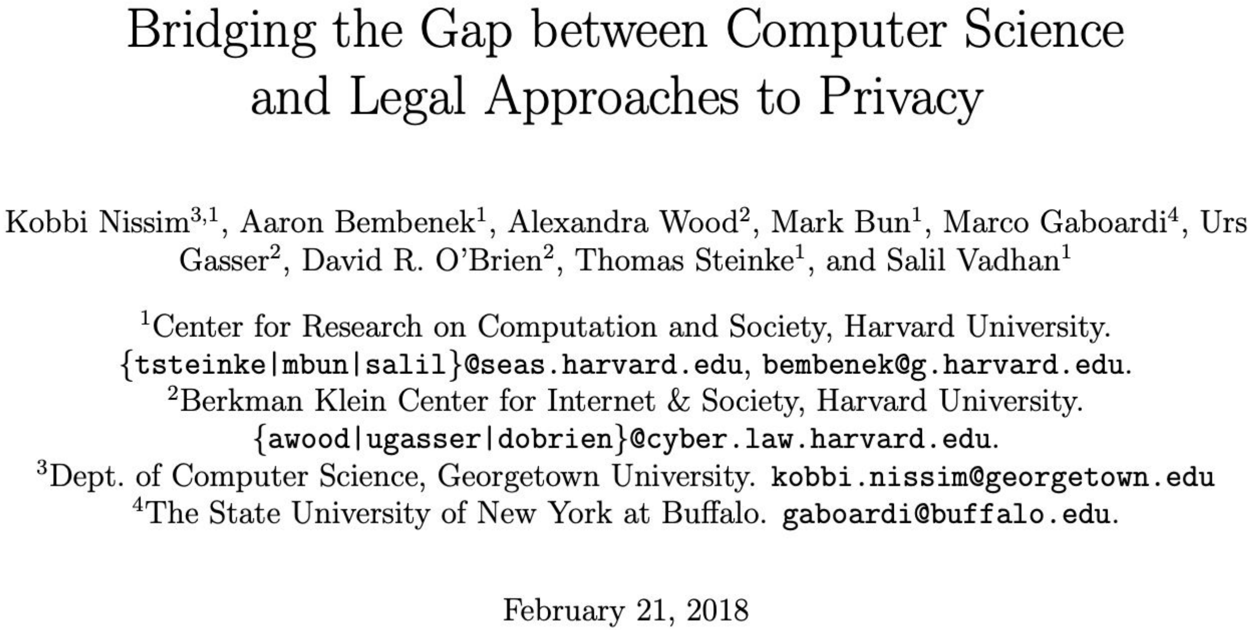
The Census Bureau is bound by Title 13 of the United States Code. These laws not only provide authority for the work we do, but also provide strong protection for the information we collect from individuals and businesses.

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

Bridging the Gap: Technical vs. Legal



Bridging the Gap between Computer Science and Legal Approaches to Privacy

Kobbi Nissim[3,1], Aaron Bembenek[1], Alexandra Wood[2], Mark Bun[1], Marco Gaboardi[4], Urs Gasser[2], David R. O'Brien[2], Thomas Steinke[1], and Salil Vadhan[1]

[1]Center for Research on Computation and Society, Harvard University.
{tsteinke|mbun|salil}@seas.harvard.edu, bembenek@g.harvard.edu.
[2]Berkman Klein Center for Internet & Society, Harvard University.
{awood|ugasser|dobrien}@cyber.law.harvard.edu.
[3]Dept. of Computer Science, Georgetown University. kobbi.nissim@georgetown.edu
[4]The State University of New York at Buffalo. gaboardi@buffalo.edu.

February 21, 2018

Bridging the Gap: Technical vs. Legal

“...the fields of law and computer science have generated different notions of privacy risks in the context of the analysis and release of statistical data about individuals…”

Source: Bridging the gap between computer science and legal approaches to privacy *(*Harv. JL & Tech.)

Bridging the Gap: Technical vs. Legal

“...this article articulates the nature of the gaps between legal and technical approaches to privacy in the release of statistical data about individuals. It also presents an argument that the use of differential privacy is sufficient to satisfy the requirements of the Family Educational Rights and Privacy Act of 1974 (FERPA), a federal law that protects the privacy of education records in the United States. This argument illustrates what may evolve to a more general methodology for rigorously arguing that technological methods for privacy protection satisfy the requirements of a particular information privacy law…”

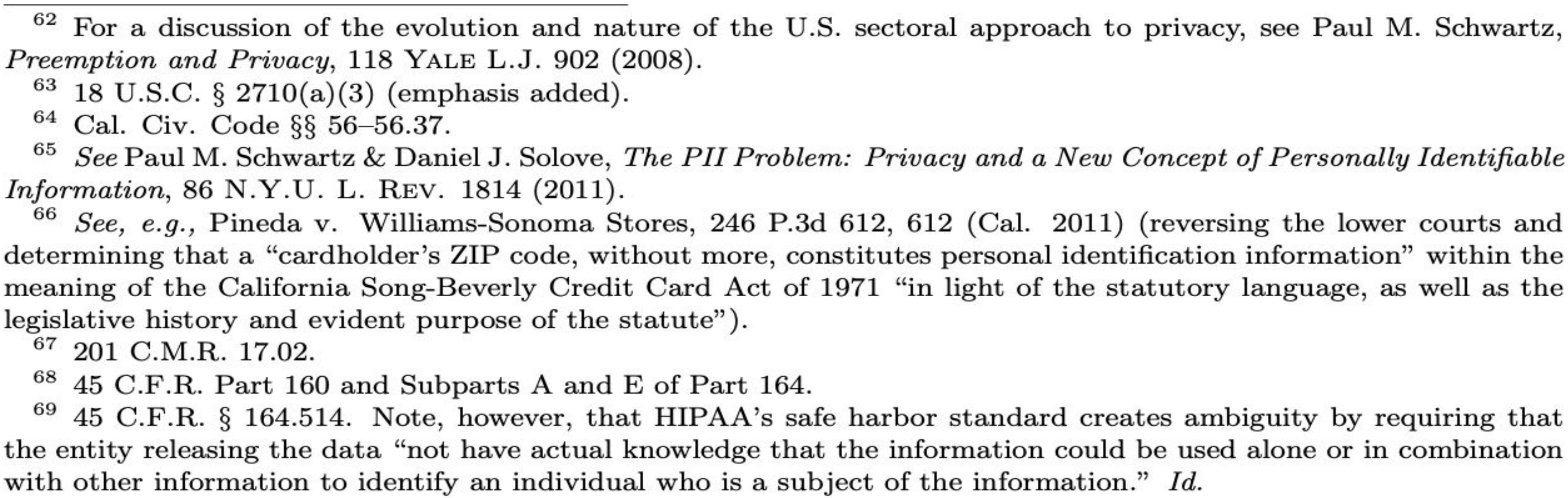
Source: Bridging the gap between computer science and legal approaches to privacy *(*Harv. JL & Tech.)

Bridging the Gap:Title 13 and U.S. Census Bureau

“... In this way, the mathematical proof demonstrates that the use of differential privacy is sufficient to satisfy a broad range of reasonable interpretations of FERPA, including interpretations that may be adopted in the future…”

Source: Bridging the gap between computer science and legal approaches to privacy *(*Harv. JL & Tech.)

Step 1: Interpret Privacy Law



[62] For a discussion of the evolution and nature of the U.S. sectoral approach to privacy, see Paul M. Schwartz, *Preemption and Privacy*, 118 YALE L.J. 902 (2008).

[63] 18 U.S.C. § 2710(a)(3) (emphasis added).

[64] Cal. Civ. Code §§ 56–56.37.

[65] *See* Paul M. Schwartz & Daniel J. Solove, *The PII Problem: Privacy and a New Concept of Personally Identifiable Information*, 86 N.Y.U. L. REV. 1814 (2011).

[66] *See, e.g.*, Pineda v. Williams-Sonoma Stores, 246 P.3d 612, 612 (Cal. 2011) (reversing the lower courts and determining that a “cardholder’s ZIP code, without more, constitutes personal identification information” within the meaning of the California Song-Beverly Credit Card Act of 1971 “in light of the statutory language, as well as the legislative history and evident purpose of the statute”).

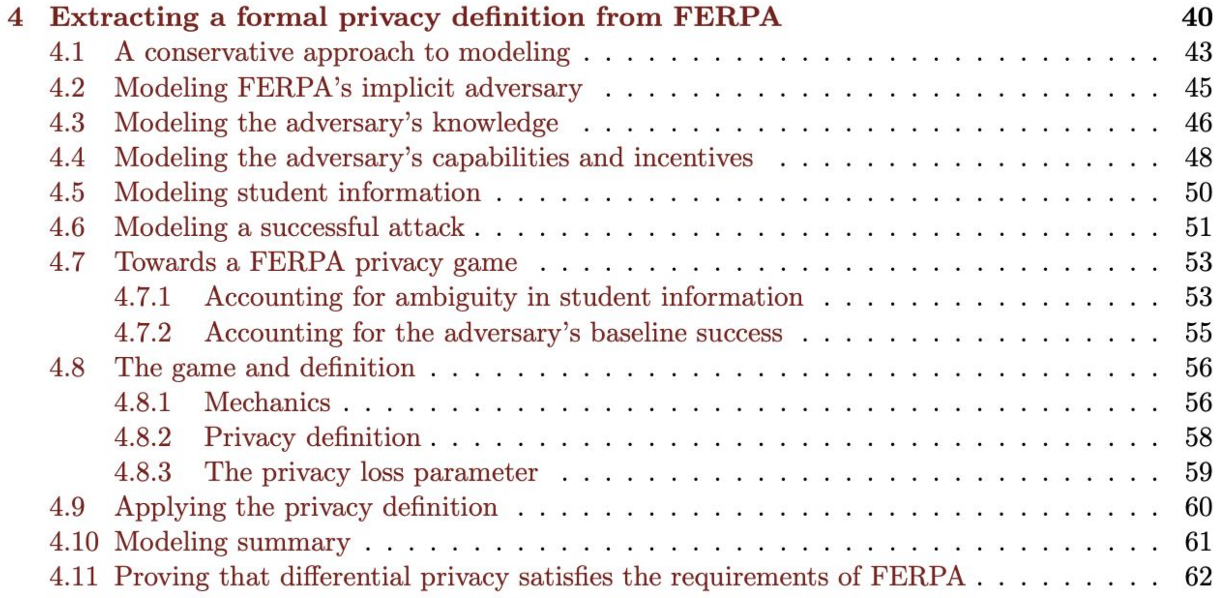
[67] 201 C.M.R. 17.02.

[68] 45 C.F.R. Part 160 and Subparts A and E of Part 164.

[69] 45 C.F.R. § 164.514. Note, however, that HIPAA’s safe harbor standard creates ambiguity by requiring that the entity releasing the data “not have actual knowledge that the information could be used alone or in combination with other information to identify an individual who is a subject of the information.” *Id.*

Source: Bridging the gap between computer science and legal approaches to privacy *(*Harv. JL & Tech.)

Step n+1: Translate into Technical Terms



Source: Bridging the gap between computer science and legal approaches to privacy (Harv. JL & Tech.)

Why Legal Considerations Matter

- Legal frameworks define data rights and duties
- Legal compliance is essential for trust and adoption
- Distributed data markets complicate governance
- Liability concerns for data market platforms (e.g., Opportunity Insights)

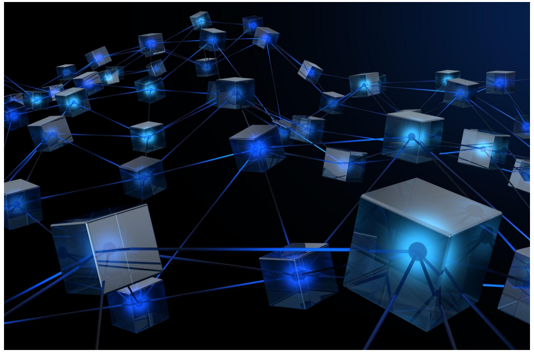
Compliance in Distributed Data Markets

- *Key challenge*: Trust among parties with differing incentives

Examples:

(1) Opportunity insights ⇔ Census Bureau (Title 13)

(2) Hospitals ⇔ Health Insurance Companies (HIPAA)



Legal Foundations (Global Overview)

- GDPR (EU)
- CCPA/CPRA (California), HIPAA (US), Title 13 (US)
- Varying consent, data definitions, cross-border rules

Further Examples of Translation (GDPR)

- *Legal*: Purpose limitation

"…Personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation')…"

- *Technical*: Can't collect data for academic research and sell to advertisers

Source: https://gdpr-info.eu/art-5-gdpr/

Further Examples of Translation (GDPR)

- *Legal*: Purpose limitation and data minimization

"…Personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation')…"

"…Personal data shall be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation');…"

- *Technical*: Can't ask for SSN when signing up for blog

Source: https://gdpr-info.eu/art-5-gdpr/

Further Examples of Translation (GDPR)

- *Legal*: Breach notification mandates (e.g., GDPR Art. 33)

"...In the case of a personal data breach, the controller shall without undue delay and, where feasible, not later than 72 hours after having become aware of it, notify the personal data breach to the supervisory authority competent in accordance with Article 55, unless the personal data breach is unlikely to result in a risk to the rights and freedoms of natural persons. Where the notification to the supervisory authority is not made within 72 hours, it shall be accompanied by reasons for the delay. The processor shall notify the controller without undue delay after becoming aware of a personal data breach..."

- *Technical*: Send notifications to supervisory authority

Source: https://gdpr-info.eu/art-33-gdpr/

Takeaways

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