Approaches to Sharing Blended Data in a 21st Century Data Infrastructure Public Workshop

Privacy Confidentiality Tradeoff

Julia Lane

Evidence Act

ACDEB
Advisory Committee on Data for Evidence Building

Advisory Committee on Data for Evidence Building: Year 2 Report

October 14, 2022

This section provides background on the Evidence Commission; the Evidence Act; the evolving evidence-building ecosystem; and ACDEB's purpose, progress, and promise.

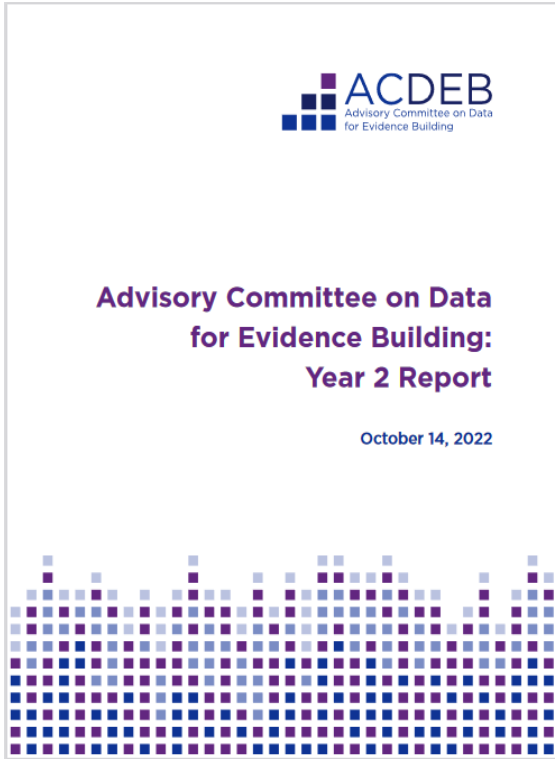
The Commission on Evidence-Based Policymaking

The Evidence Commission was created by the Evidence-Based Policymaking Commission Act of 2016. With widespread bipartisan support, the law signaled the President's and Congress's commitment to building capacity to produce higher-quality evidence for decisionmaking. In its 2017 report, the Evidence Commission articulated a vision around data and evidence generation and use as a routine and critical function of government. The report features 22 recommendations to improve data access, modernize privacy and confidentiality protections, strengthen evidence-building capacity, and establish an NSDS to support governmentwide evidence building.

The Foundations for Evidence-Based Policymaking Act

In 2018, Congress passed the Evidence Act, addressing half of the Evidence Commission's recommendations. The law established new legal expectations for openness and accessibility, building a framework where leaders across the government work together to coordinate data and evidence needs and uses. This includes aligning data from various sources, such as survey data and administrative data, with differing degrees of protection needed, from open to confidential, for different purposes, including producing statistical estimates and informing the administration and evaluation of government programs for certain shared purposes. Box 1 defines key terms and definitions.

Advisory Committee Recommendations



Measure and report data value. The production of value (or "utility") is inherent to the core responsibilities of statistical agencies and, as such, is critical for the NSDS. There are several dimensions of value—broadly, adherence to democratic and equitable values and providing value to the public and, more specifically, value of the data assets, value of NSDS capabilities, and value of the data service itself. The NSDS should model an approach to measure and report on the value of each of these aspects, including the following actions:

- ***Produce an NSDS data inventory with usage statistics.*** The NSDS should develop and maintain a publicly available inventory of NSDS data assets in keeping with Evidence Act requirements for agency data inventories. While not a full measure of value, as a baseline, this inventory should include usage statistics. To support a more seamless experience for users, the NSDS data inventory should model the format and content, including detailed metadata, that could be used to harmonize other data inventories and catalogs.
- ***Develop concrete measures of value.*** The NSDS should develop and publish concrete measures of value, including exploring ways to measure the impact and the value of evidence for different stakeholders.

Recommendation 1.6

OMB should adopt a risk-utility framework as the basis for standards on sensitivity levels, access tiers, and risk evaluations as part of the regulation on expanding secure access to CIPSEA data assets

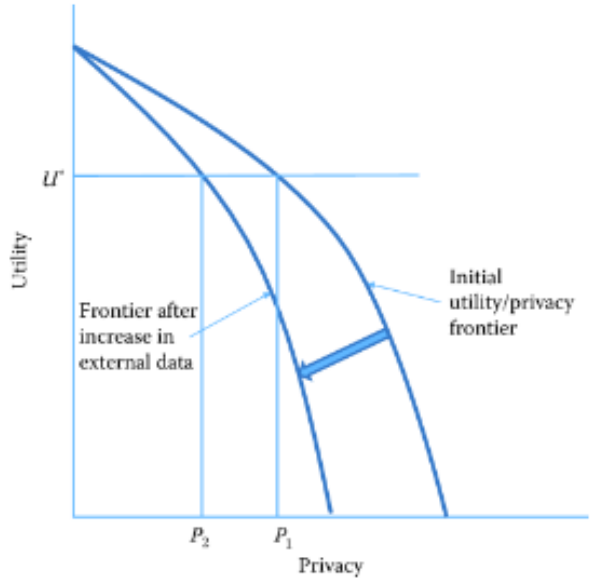


Figure 12.1: The privacy–utility tradeoff

https://textbook.coleridgeinitiative.org/chap-privacy.html

apnews.com/article/technology-government-and-politics-privacy-8ab4f67fbec19fd57365bd58ffde0e4d

AP U.S. News World News Politics Sports Entertainment Business Technology Health Science Oddities Lifestyle Photography

Mike Schneider
I cover census, demographics, Florida and related topics.
MikeSchneiderAP
mschneider@ap.org

Click to copy

RELATED TOPICS

Technology

The U.S. Census Bureau on Wednesday said it was putting on hold plans to apply by 2025 a controversial method for protecting the privacy of participants in its most comprehensive survey of Americans after facing pushback from prominent researchers and demographers.

The science doesn't yet exist to apply differential privacy algorithms to the annual American Community Survey, which covers more than 40 topics ranging from income, internet access, rent, disabilities and language spoken at home, the statistical agency said in a blog post.

"It's also not clear that differential privacy would ultimately be the best option," wrote Donna Daily, chief of the American Community Survey Office.

The Census Bureau embraced using differential privacy algorithms for the first time with the release last year of the first round of 2020 census data, and it had said the method would be applied to American Com
intentional errors to data
the smallest geographies,

ncsl.org/research/redistricting/differential-privacy-for-census-data-explained.aspx

LEGISLATORS & STAFF RESEARCH MEETINGS & TRAINING NCSL IN D.C. NEWS

units in each census block and the number and type of each group quarters unit in each census block are also to be kept invariant. In 2010 and previous decades, all these were kept "invariant" along with most data at the census block level, with the exception of race. All other data, including total population numbers for lower geographic units and demographic characteristics, will vary to some extent this decade.

Differential privacy will mean that, except at the state level, population and voting age population will not be reported as enumerated. And, race and ethnicity data are likely to be farther from the "as enumerated" data than in past decades, when data swapping was used to protect small populations. (In 2010, at the block level, total population, total housing units, occupancy status, group quarters count and group quarters type were all held invariant.) This may raise issues for racial block voting analyses.

While differential privacy is intended to protect confidentiality for respondents, it has implications for smaller subpopulations. For instance, the National Congress of American Indians notes, "The implementation of differential privacy could introduce substantial amounts of noise into statistics for small populations living in remote areas, potentially diminishing the quality of statistics about tribal nations."

Five Safes: Four of them mitigate risk

1. Safe People

Training; clearance; indemnification

2. Safe Settings

Physical protection (new options)

3. Safe Data

Protect core elements (hashing, suppression_

4. Safe Outputs

SDL techniques

Five safes: Three of them refer to utility tradeoff

1. Safe Projects

utility measure

2 Safe People

trust and quality measure

3. Safe Outputs

Risk vs. Harm

Technology options

Administrative Data Research Facility

5 Safes Framework

- Projects
- People
- Settings
- Data
- Exports

Governance

Enterprise Data Catalog - Traditional Metadata Management with Rich Context

Disclosure Review for Exports

Data Stewardship

Manage Projects, People, Datasets, & Agreements

Transparency

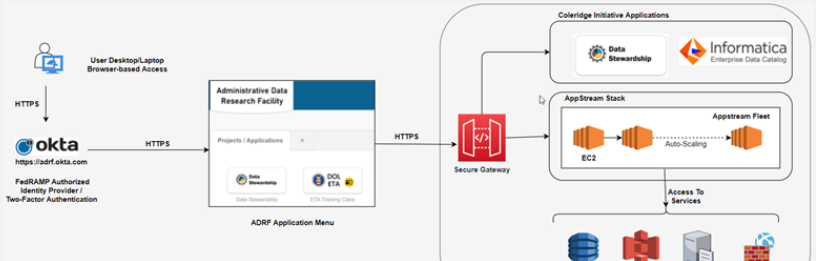
Usage Based Pricing Model

User Accessible Usage Statistics

Flexible Performance Options

Technology

Secure Remote Access, FedRAMP Authorized GovCloud



Assessing the Value of Public Data Assets

Costs and Risks

- Acquisition, collection, curation, protection, storage
- Risk of disclosure, re-identification, and reputation to the agency

Reward (or Utility) – more anecdotal

- Real value of these data to society, researchers and policy makers is yet to be determined
- Public provided data and information are special goods and offer a special challenge

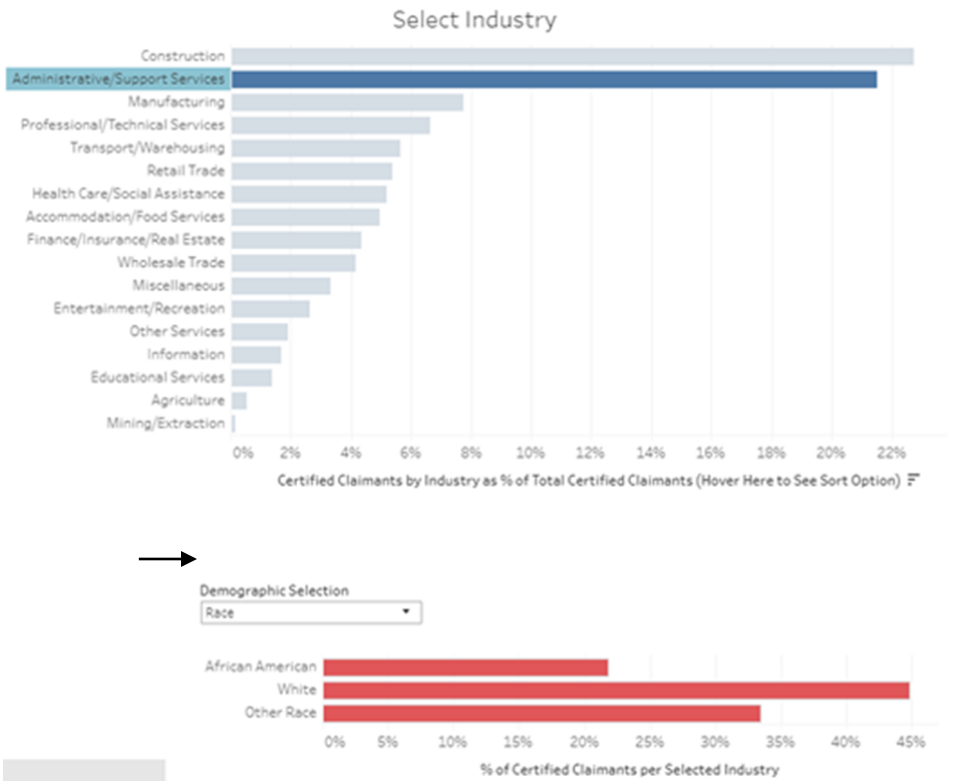
Project in progress

- Develop an approach for evaluating the value of publicly available datasets and the potential value of free public access to these data
- Start with a proof of concept developing the basic methodology and then applying it to two distinct data sets
- Infometrics approach based on information theory

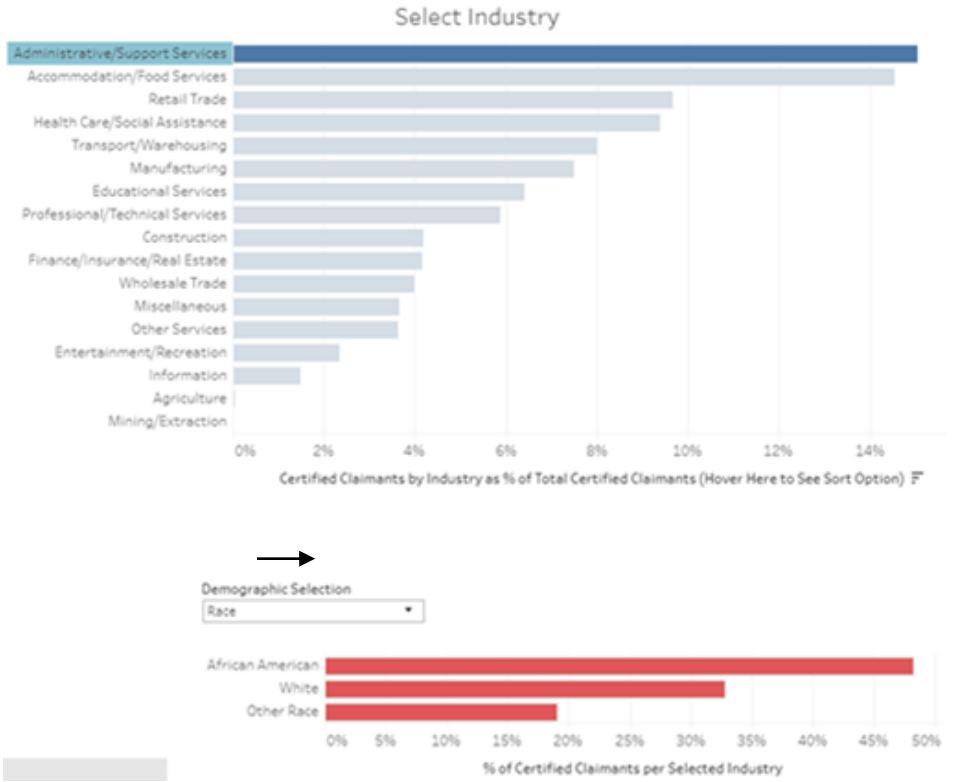
Spiro Stefanou, Administrator of Economic Research Service, USDA

Assessing the Value of Public Data Assets

Pre Covid-19 UI Claimants: African-Amer 22%



8/7/21 Covid-19 week UI Claimants: African-Amer 48%



Use Case
The Role of Multi-State Collaboratives to Support Evidence Building

George W. Putnam, LMI Director

IL Dept of Employment Security
August 30, 2021

Assessing the Value of Public Data Assets
Timely, local, actionable

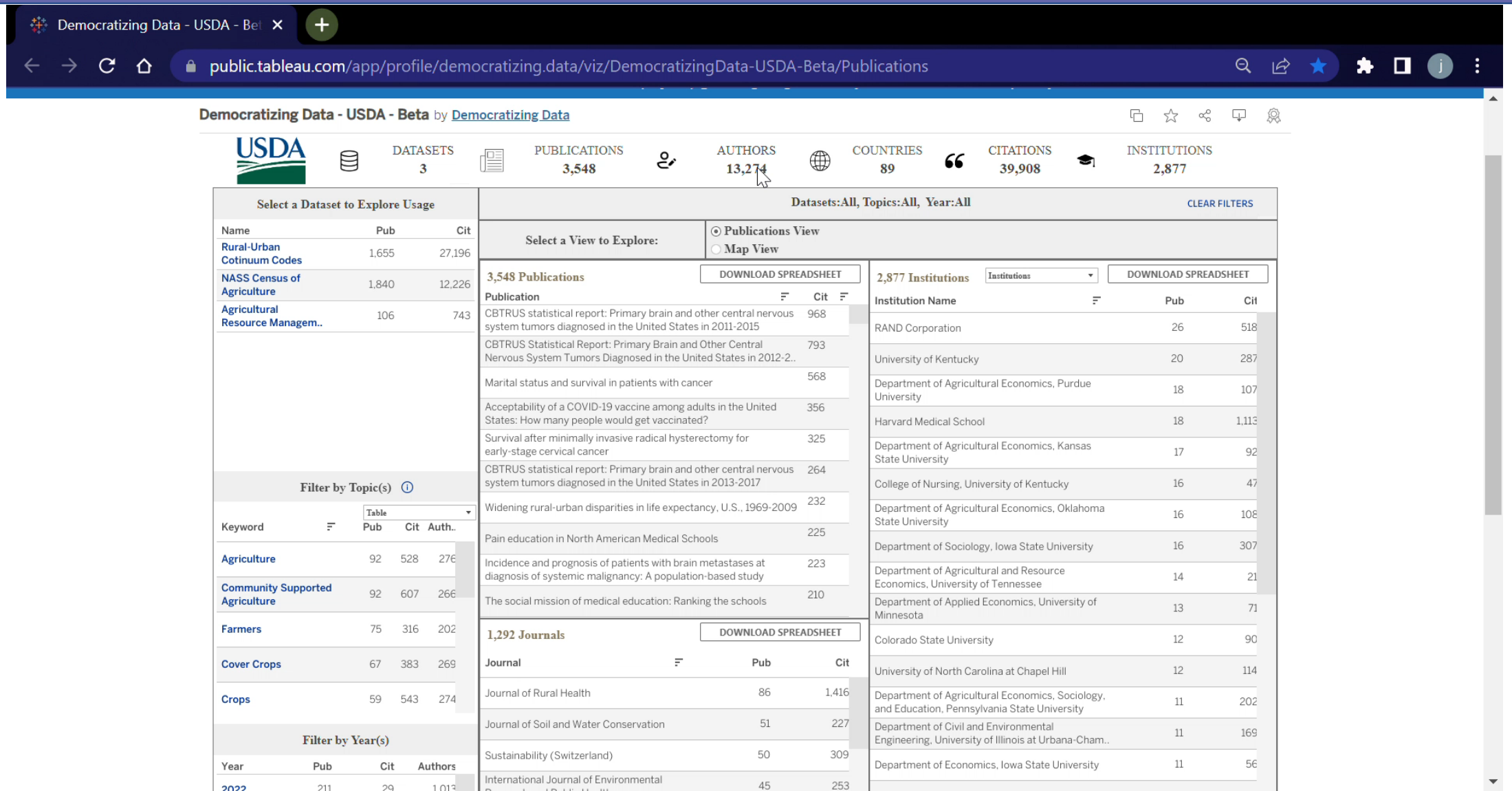
- Intelligence for statewide and local stakeholder needs
  - Governor's Office (policymaking), State agency staff (program administration), Local Workforce Boards (strategic resource allocation)
  - Synthetic/differentially private data unacceptable
  - Tiered access maximizes the content of local data patterns AND
    - Protects against the disclosure of individuals- confidential summary tabulations
    - Protects against the disclosure of firms- non-disclosure agreement with portal users
- Scalable
  - AR, IL, IN, MO, NJ and TN are building on the approach of the Unemployment to Reemployment portal
  - Common data model with repurposing of PROMIS UI claimant files
  - Shared development visualization code
  - Standardization of UI claimant measures

Use Case
The Role of Multi-State Collaboratives to Support Evidence Building

George W. Putnam, LMI Director

IL Dept of Employment Security
August 30, 2021

Assessing the Value of Public Data Assets Usage and reproducibility



Recommendation 1.6

OMB should adopt a risk-utility framework as the basis for standards on sensitivity levels, access tiers, and risk evaluations as part of the regulation on expanding secure access to CIPSEA data assets

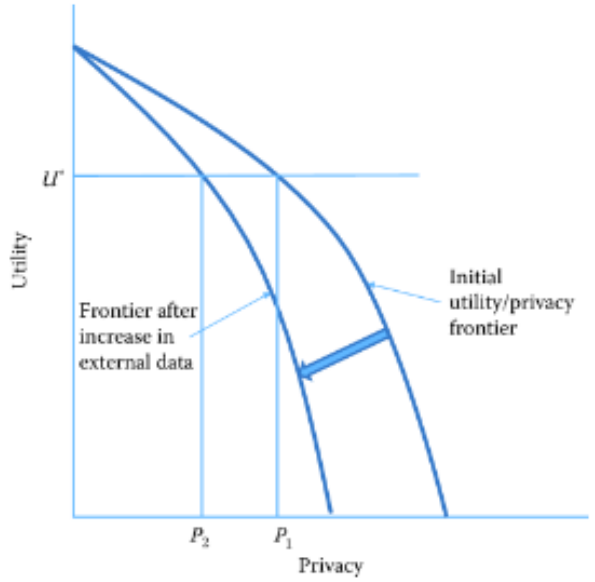


Figure 12.1: The privacy–utility tradeoff

https://textbook.coleridgeinitiative.org/chap-privacy.html