Prof. David Evans

Data Privacy Expert

Ex. No.
PXRD007
1:20-cv-03010-APM
1:20-cv-03715-APM

REDACTED FOR PUBLIC FILING

Assignment

1. Evaluate **privacy risks** with the proposed sharing of User-side Data, Ads Data, and Search Data
2. Assess whether **privacy-enhancing technologies** can mitigate those privacy risks while still sharing useful information
3. Respond to the reports of Google's privacy expert

REDACTED FOR PUBLIC FILING

Key Opinions

1. There are well-established **privacy-enhancing techniques** that can be used to protect sensitive information.

2. Many organizations, including Google, **safely release sensitive data** by using privacy-enhancing techniques.

3. Google can share the data at issue in a way that **assures privacy while providing utility**.

REDACTED FOR PUBLIC FILING

Google's Expert Agrees Data Can Be Shared

Chris Culnane, PhD

Google's Expert
Principal & Consultant
Castellate Consulting Ltd.

Q. Dr. Culnane, you believe that it is possible for Google to share what you call the DOJ search data by applying privacy-enhancing techniques to achieve suitable privacy safeguards, don't you?

A. Yes.

Source: Culnane (Google Expert) Dep. at 11:8–12.

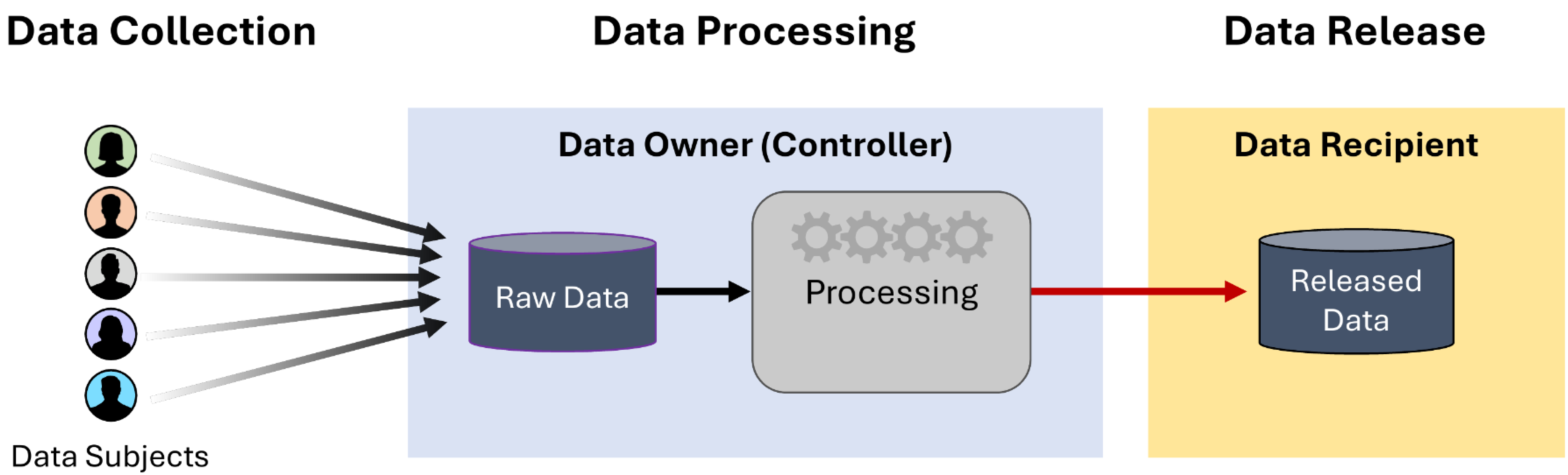
REDACTED FOR PUBLIC FILING

Experts' Disagreement

What Dr. Culnane Claims	My Opinion
"In the Search Context, Only Frequency Thresholds Provide Indistinguishability."	There are many well-established **privacy-enhancing techniques**, and the remedy should **use techniques appropriately to assure privacy while providing high utility.**

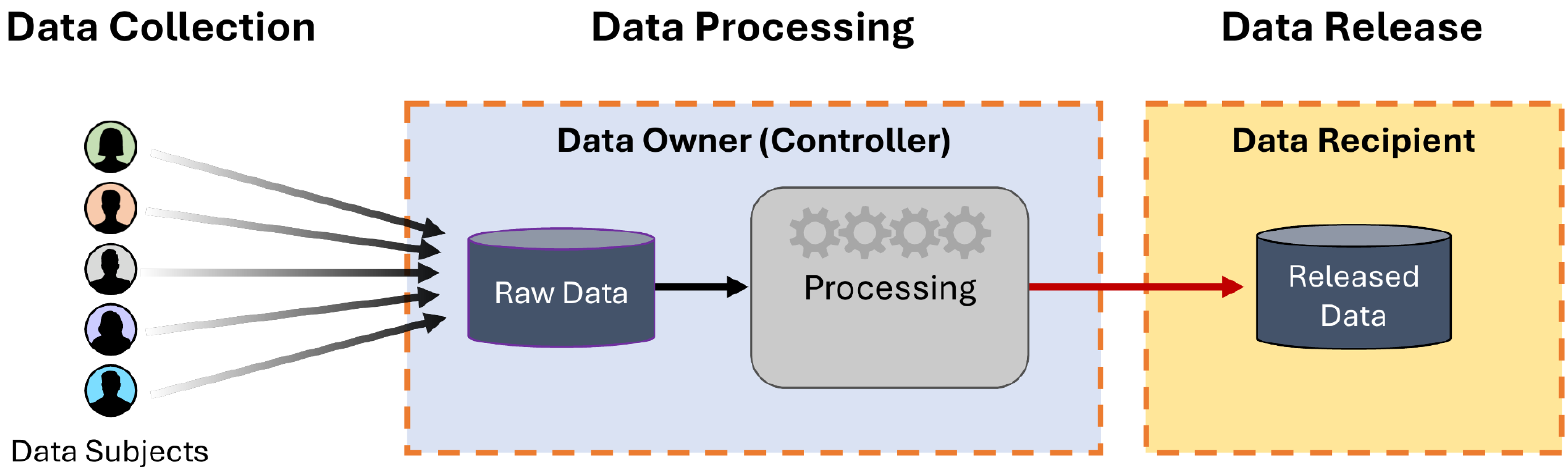
REDACTED FOR PUBLIC FILING

What is Data Privacy?



REDACTED FOR PUBLIC FILING

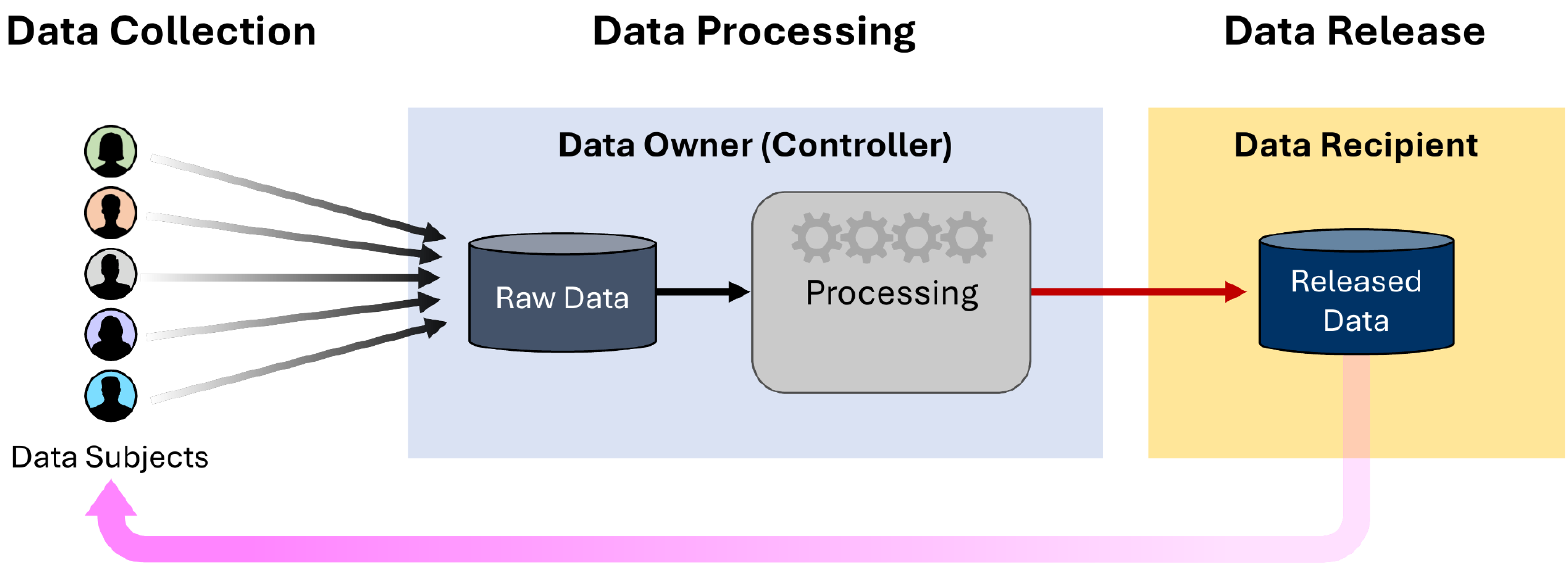
What is Data Privacy?



Data security: preventing **unintended releases** of data

REDACTED FOR PUBLIC FILING

What is Data Privacy?

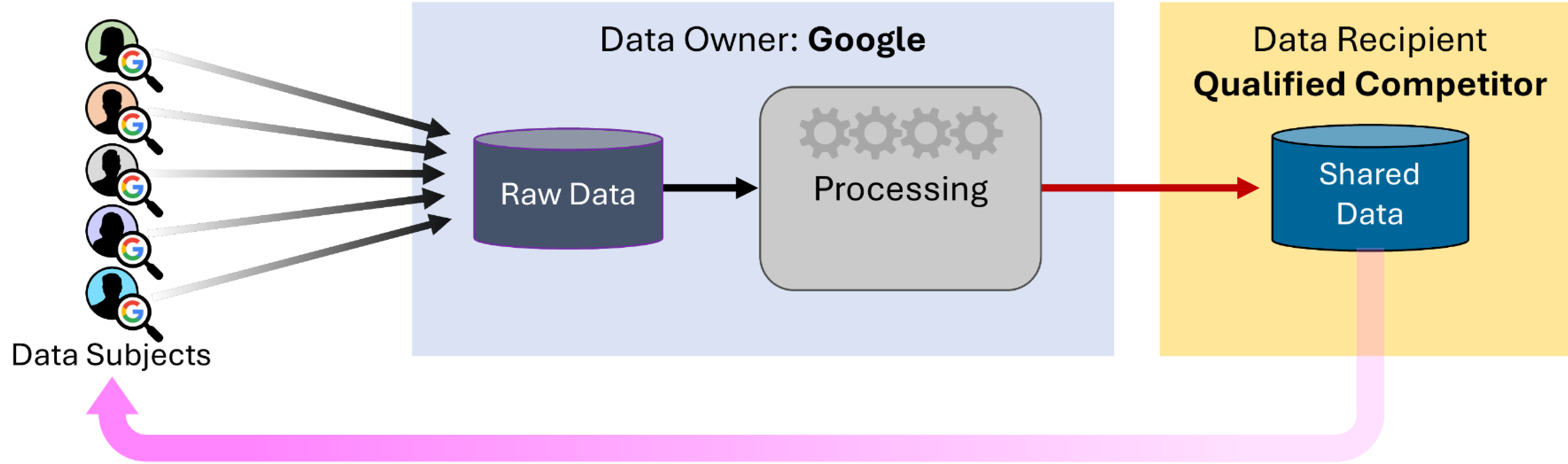


Data privacy: preventing **unintended disclosure** of sensitive information from **intentionally released data**

REDACTED FOR PUBLIC FILING

Data Privacy for Proposed Data Sharing

Data Collection | **Data Processing** | **Data Release**



Data privacy issue: potential for **disclosure** of sensitive information from **shared data** and mitigations to share safely

REDACTED FOR PUBLIC FILING

The Data at Issue

User-side Data
RPFJ Sections VI.A, C, & D

Search Index Data
RPFJ Section VI.A

Ads Data
RPFJ Sections VI.E & F

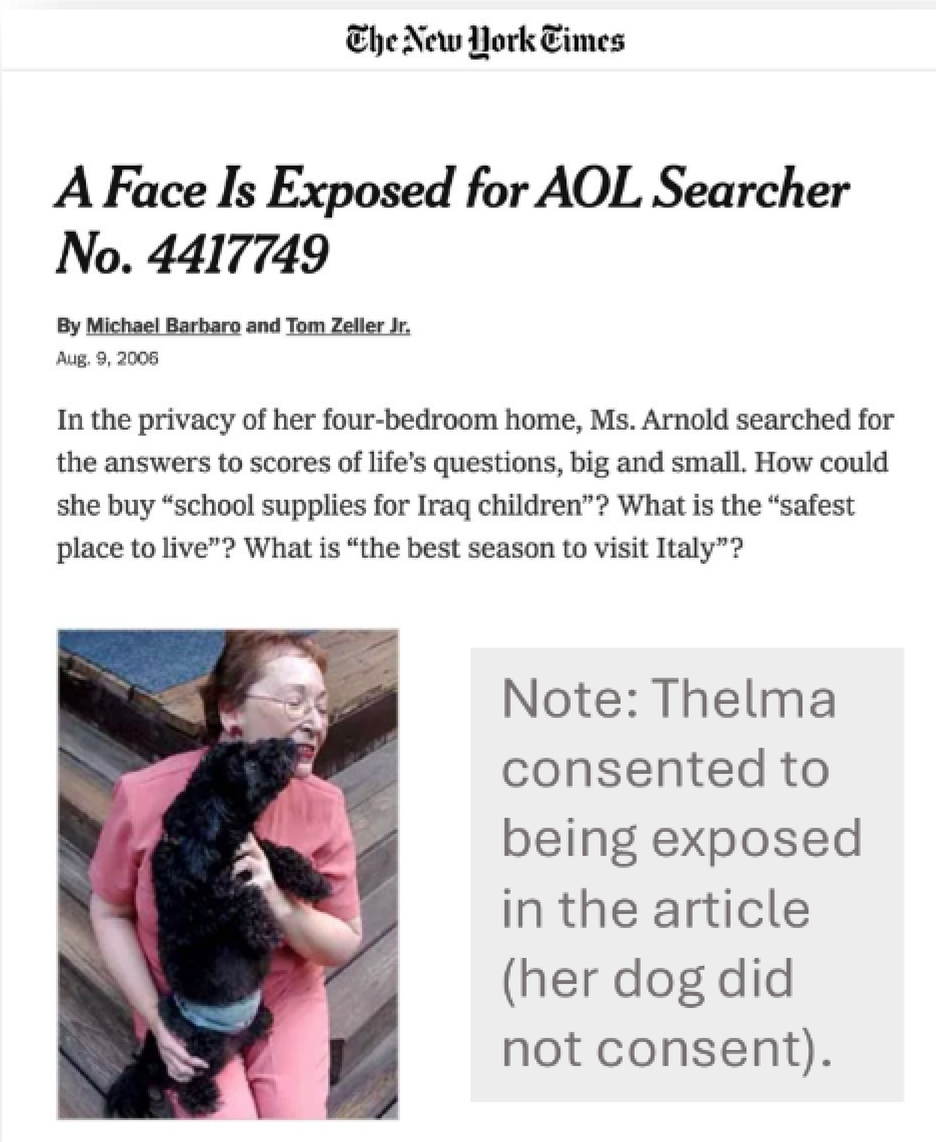
Submitted queries
Clicked-on links
Time looking at results
Hovering over a link
User location
User device
Ranking signals
...

Data Google collects from users and uses to train models
(RankEmbed, NavBoost, Glue, and [redacted])

Source: UPX0004 at -.001 - .003; UPX0005 at -807, -832; UPX0868 at -603-604; GOOG-DOJ-34933211 at -226; [redacted]

REDACTED FOR PUBLIC FILING

Innocuous Data Can Reveal Sensitive Information


The New York Times

A Face Is Exposed for AOL Searcher No. 4417749

By Michael Barbaro and Tom Zeller Jr.
Aug. 9, 2006

In the privacy of her four-bedroom home, Ms. Arnold searched for the answers to scores of life's questions, big and small. How could she buy "school supplies for Iraq children"? What is the "safest place to live"? What is "the best season to visit Italy"?

Note: Thelma consented to being exposed in the article (her dog did not consent).

Linking

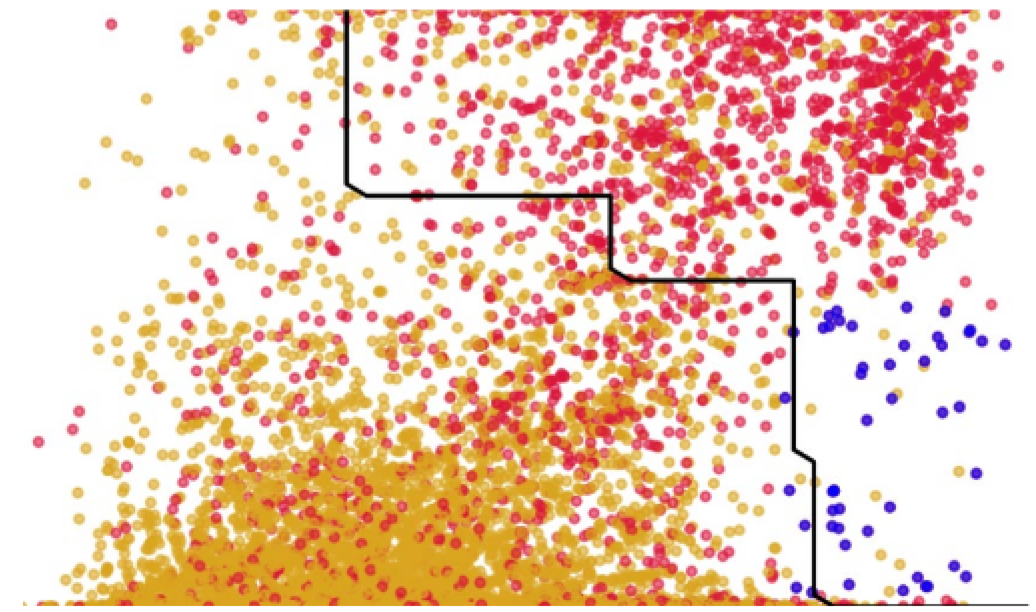
Aggregate Statistics

Block	Total	Race 1	...	Race 63	...
⋮	⋮	⋮	⋮	⋮	⋮
20394	712	0	...	82	...
20395	2316	3	...	27	...
⋮	⋮	⋮	⋮	⋮	⋮

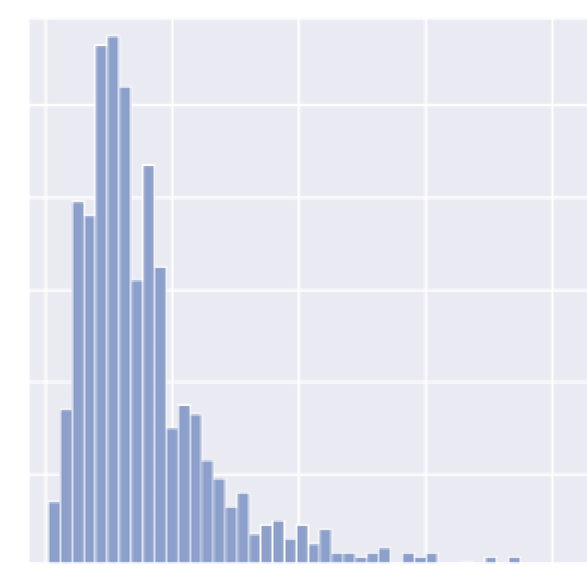
ID	Block	Race	Age	...
⋮	⋮	⋮	⋮	⋮
032	20394	7	23	...
033	20394	5	82	...
⋮	⋮	⋮	⋮	⋮

Individuals

Reconstruction



Attribute Inference Attacks



Membership Inference Attacks

Inference Attacks

Source: https://www.nytimes.com/2006/08/09/technology/09aol.html; Jayaraman, Bargav and David Evans. "Are Attribute Inference Attacks Just Imputation?". In: ACM CCS 2022. url: https://dl.acm.org/doi/10.1145/3548606.3560663; Duan, Michael, Anshuman Suri, et al. "Do Membership Inference Attacks Work on Large Language Models?". In: COLM 2024. url: https://arxiv.org/pdf/2402.07841.

REDACTED FOR PUBLIC FILING

Assessing Privacy Risk

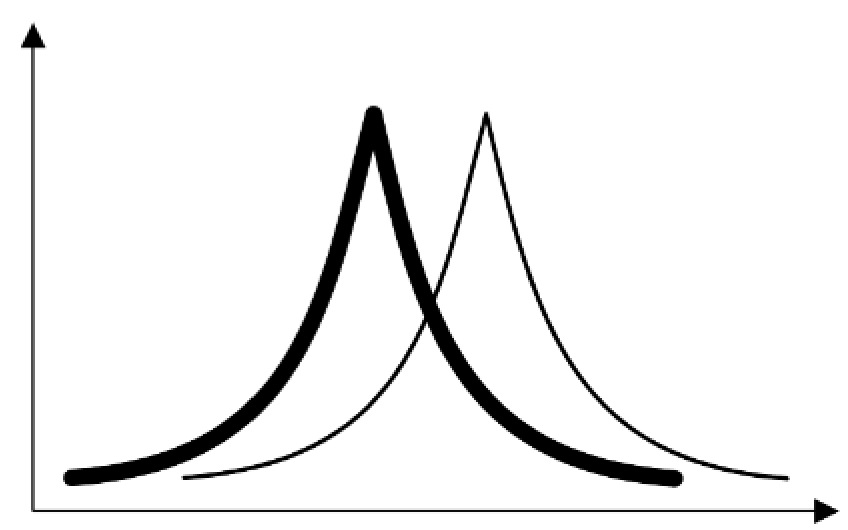
Until ~2000:

ad hoc privacy

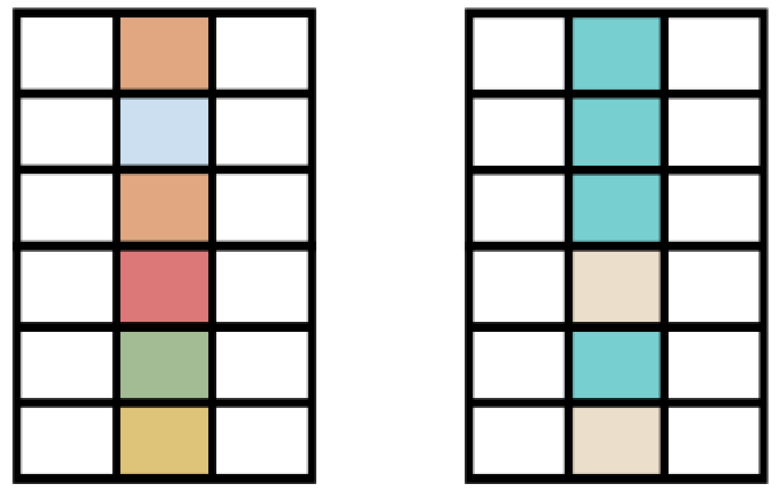
Trying things and hope they work

Today: formal privacy

Mathematical definitions of privacy and principled mechanisms for satisfying them



differential privacy

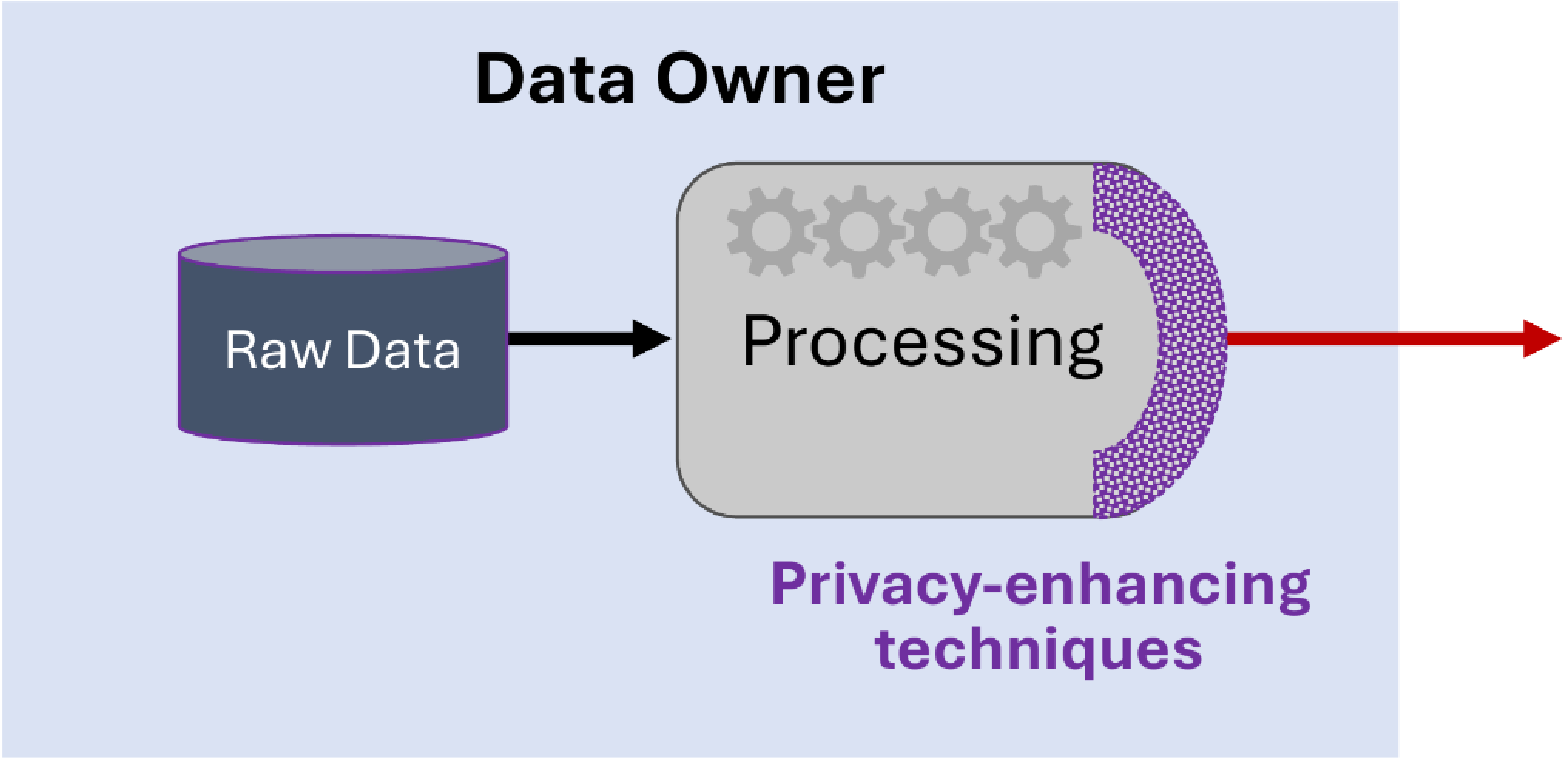


***k*-anonymity**

and hundreds of others...

REDACTED FOR PUBLIC FILING

Privacy-Enhancing Techniques (PETs)



Output useful data with **privacy assurances**

REDACTED FOR PUBLIC FILING

Opinion 1: Privacy-Enhancing Techniques Work

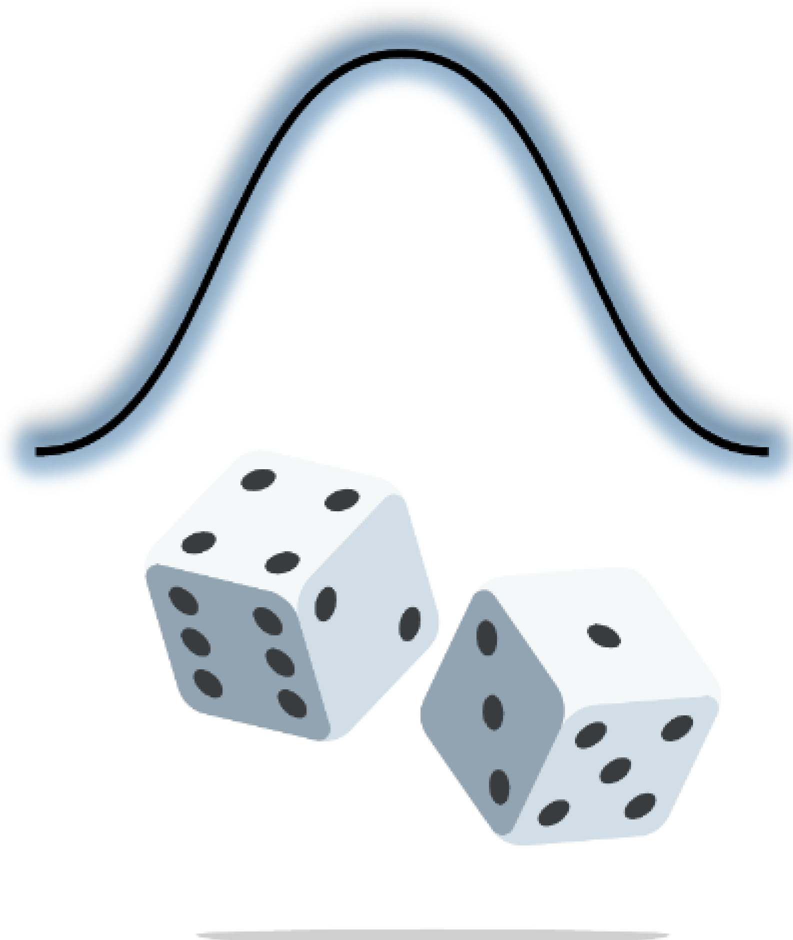
1. There are well-established **privacy-enhancing techniques** that can be used to protect sensitive information.

2. Many organizations, including Google, **safely release sensitive data** by using privacy-enhancing techniques.

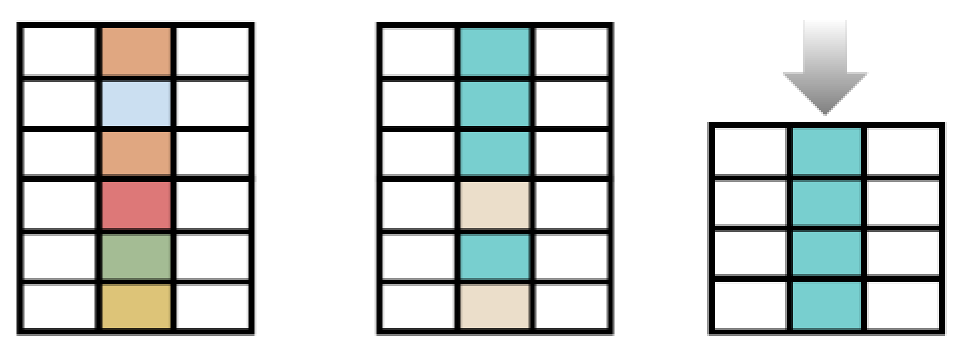
3. Google can share the data at issue in a way that **assures privacy** while **providing utility**.

REDACTED FOR PUBLIC FILING

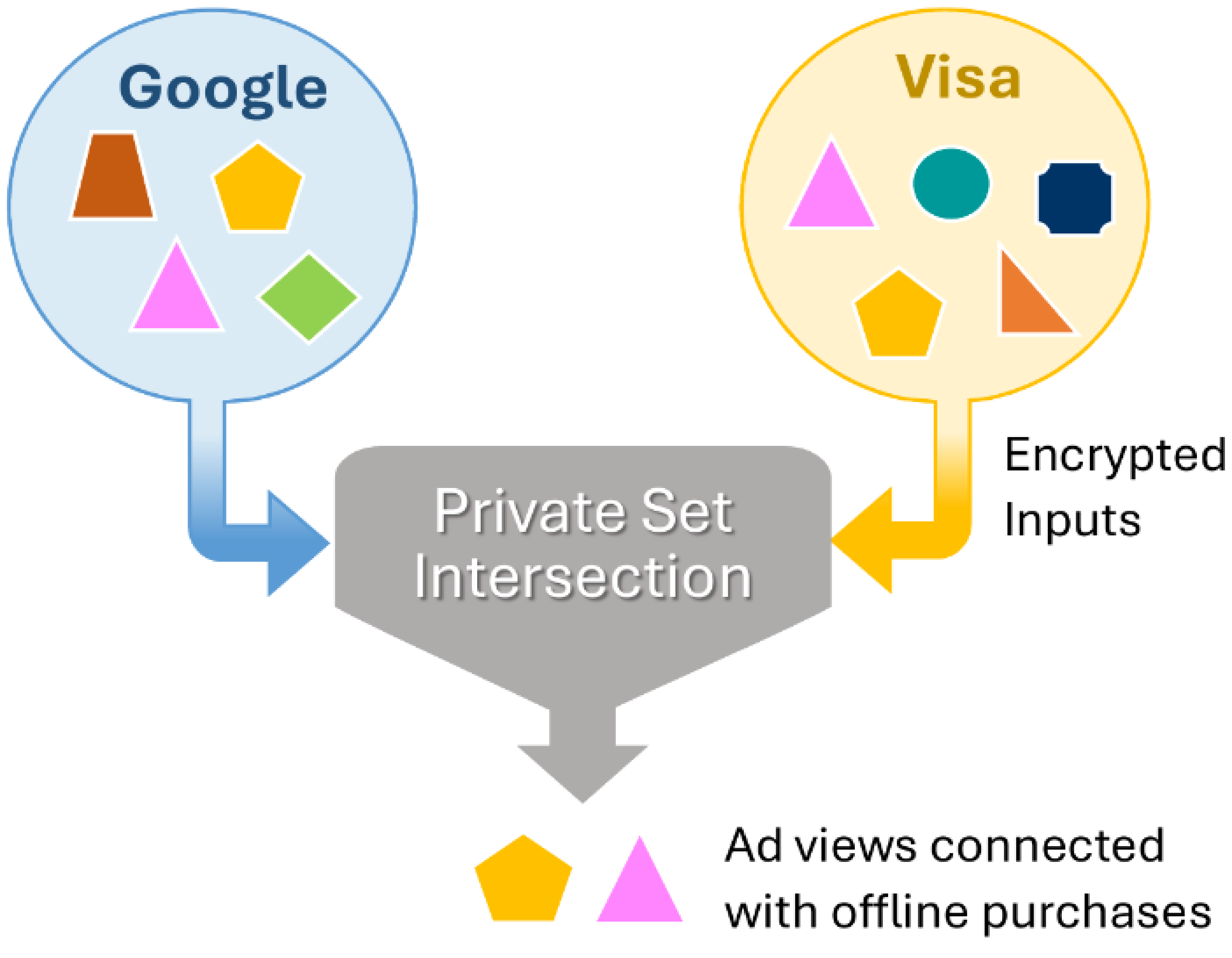
Broad Types of Privacy-Enhancing Techniques



Noise



Frequency Bounds

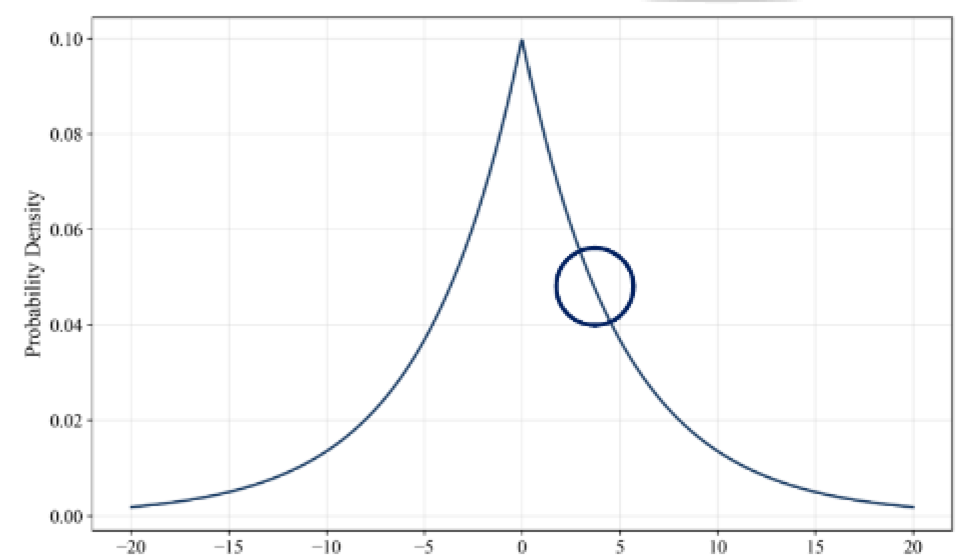


Cryptographic Methods

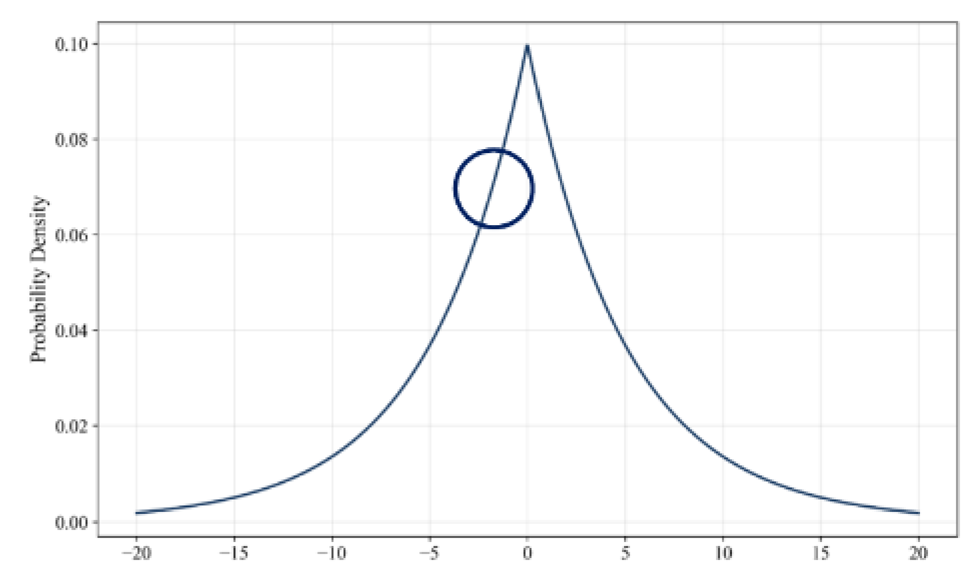
REDACTED FOR PUBLIC FILING

Noise for Privacy

Source Data | **Noise**  | **Released Data**

629 + 

= 631.52

629 + 

= 628.73

REDACTED FOR PUBLIC FILING

Differential Privacy

Gives a **mathematical bound** on exposure of individual's data

No assumptions needed about what is sensitive information, actual data, what adversary can do, what adversary already knows

$$\frac{\text{Probability of this output from dataset containing user}}{\text{Probability of this output from dataset } \textbf{without} \text{ user}} \leq \exp(\epsilon)$$

Privacy loss parameter (epsilon) provides precise control of **privacy–utility tradeoff**

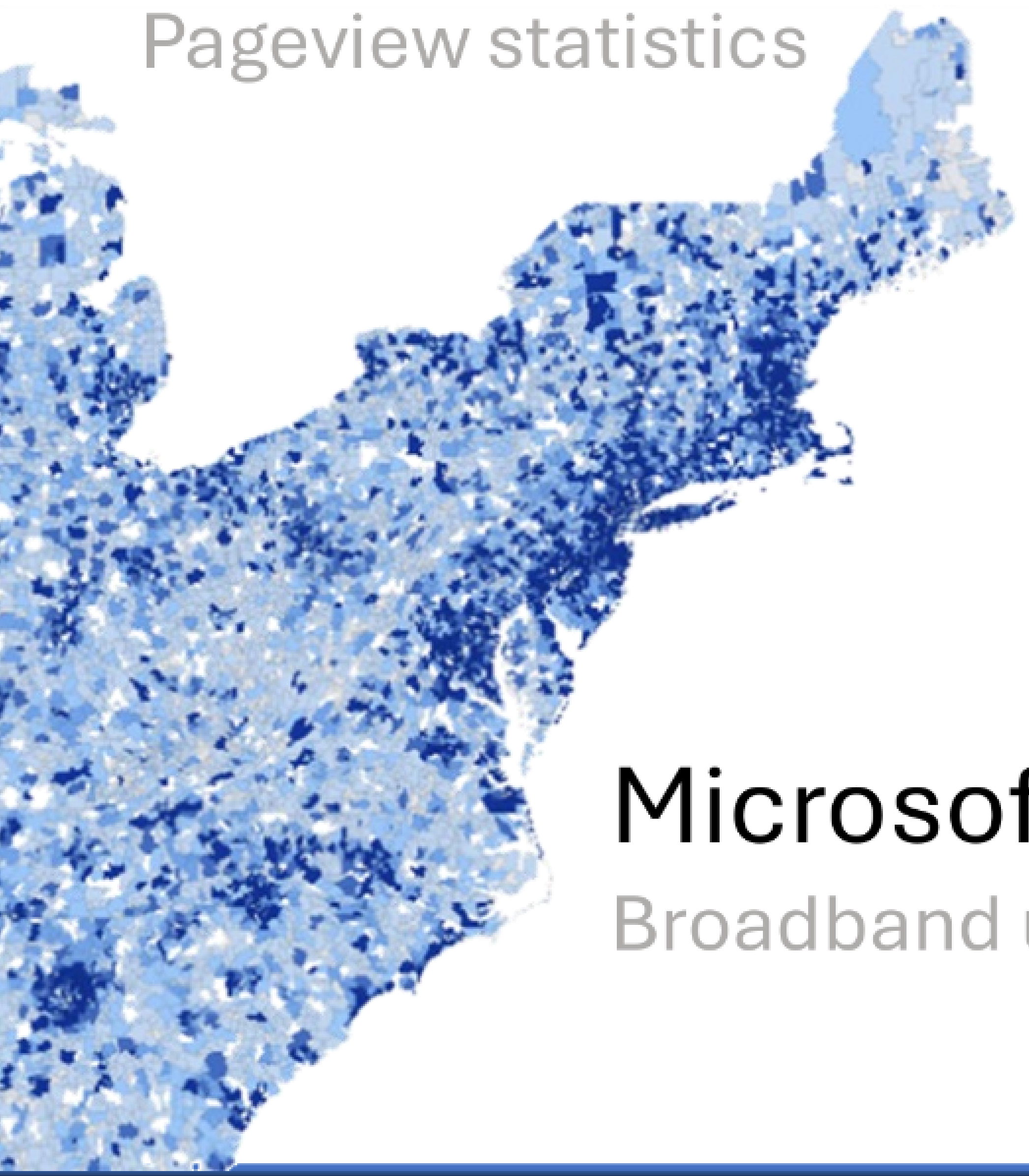
Source: Dwork, Cynthia, et al. "Calibrating Noise to Sensitivity in Private Data Analysis". In: Theory of Cryptography. Ed. by Shai Halevi and Tal Rabin. Vol. 3876. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, 2006, pp. 265–284. doi: 10.1007/11681878_14. url: https://link.springer.com/chapter/10.1007/11681878_14.

REDACTED FOR PUBLIC FILING

Widespread Acceptance and Use

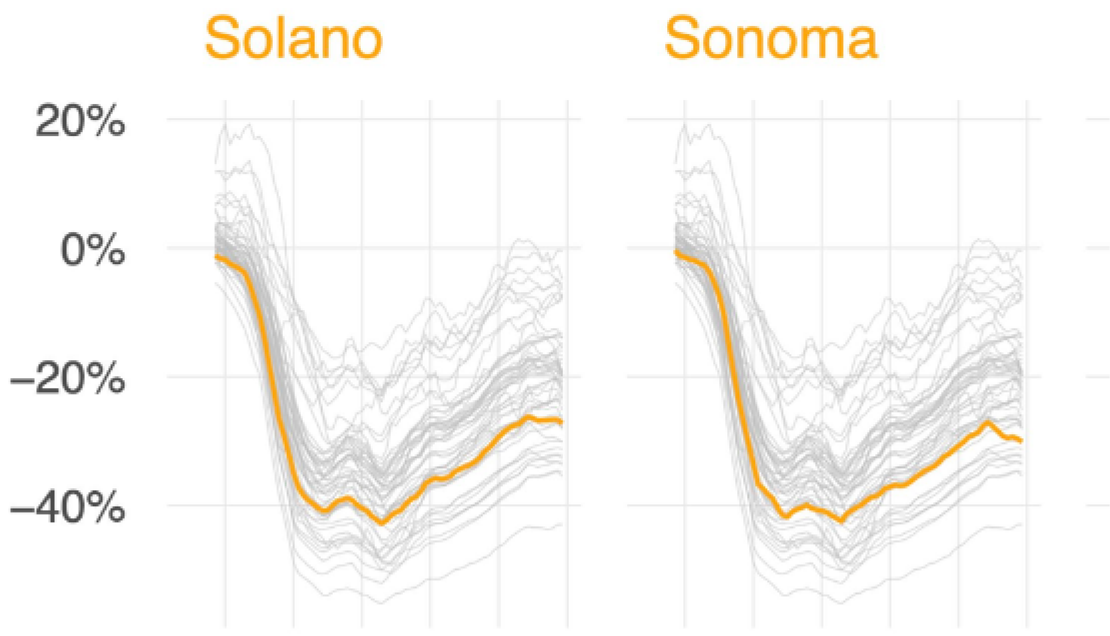
Wikimedia

Pageview statistics



Apple

Learning iconic scenes



Facebook

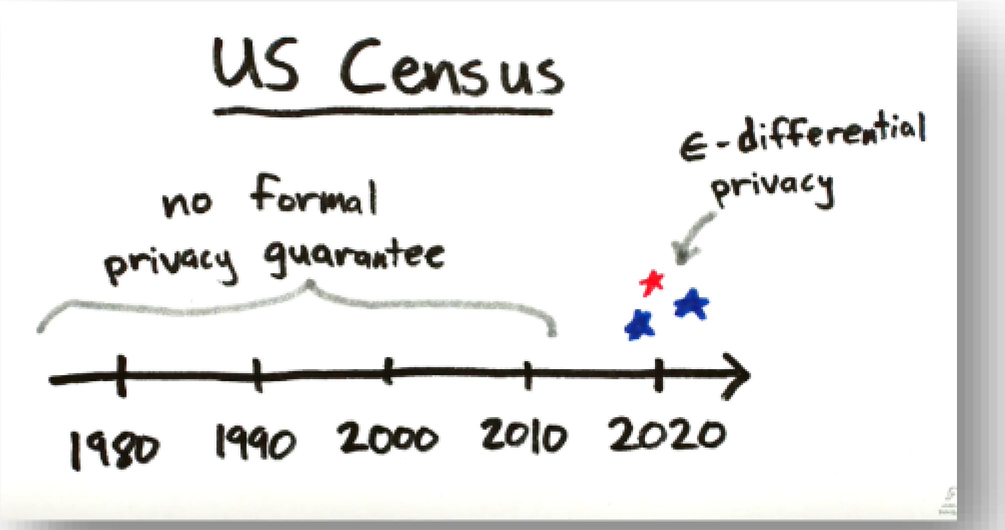
Movement dataset

Microsoft

Broadband usage

US Census Bureau

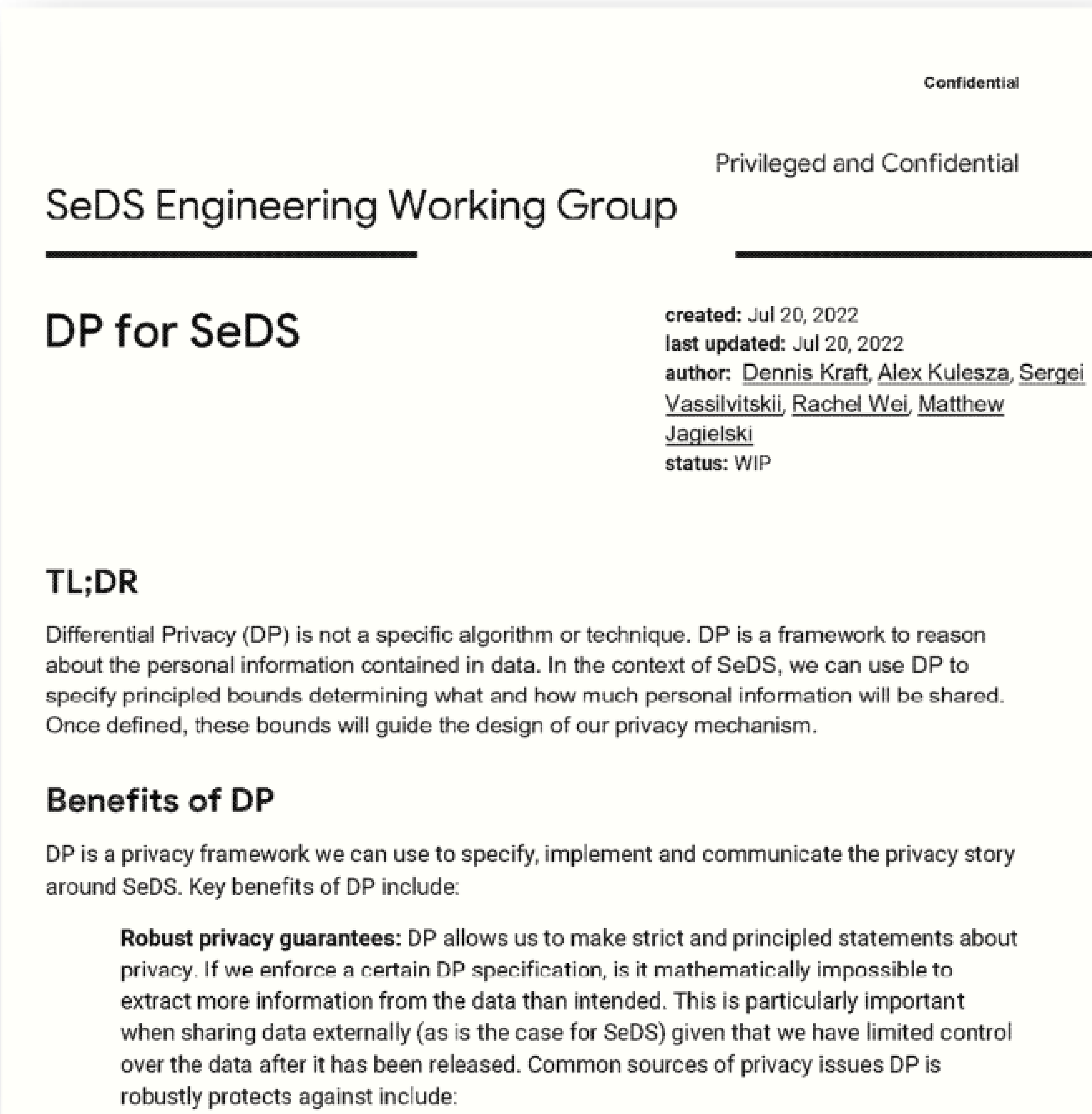
Redistricting data



Source: https://www.tmlt.io/casestudy/revealing-wikipedia-usage-data-while-protecting-privacy; https://github.com/microsoft/USBroadbandUsagePercentages/blob/master/assets/broadbandusagezipcode.png; https://machinelearning.apple.com/ research/scenes -differential-privacy; https://scontent-iad3-1.xx.fbcdn.net/v/t39.8562-6/240856930_242973511029218_5562704693675330708_n.pdf?_nc_cat=101&ccb=1-7&_nc_sid=e280be&_nc_ohc=5PyY0fYO49IQ7kNvwFYH-d0&_nc_oc=AdmPfkNb3J_Rp-QzysvDOLgV5cCtK7_wufNJdC2kayD9msa-F7dM_wTxAhR9_ut9_qHCxx-IM6hvFmS2hj5mjy-N&_nc_zt=14&_nc_ht=scontent-iad3-1.xx&_nc_gid=uAGCgBOL4wKO8N6V0ZUr9g&oh=00_AfGFhALrgWMZwus_U9LXz0z2PNTBtefV5-i8gOgaFxFRvg&oe=680A321B; https://www.youtube.com/watch?v=pT19VwBAgKA; https://machinelearning.apple.com/research/scenes-differential-privacy

REDACTED FOR PUBLIC FILING

Google Uses Differential Privacy (DP)



Confidential

Privileged and Confidential

SeDS Engineering Working Group

DP for SeDS

created: Jul 20, 2022
last updated: Jul 20, 2022
author: Dennis Kraft, Alex Kulesza, Sergei Vassilvitskii, Rachel Wei, Matthew Jagielski
status: WIP

TL;DR

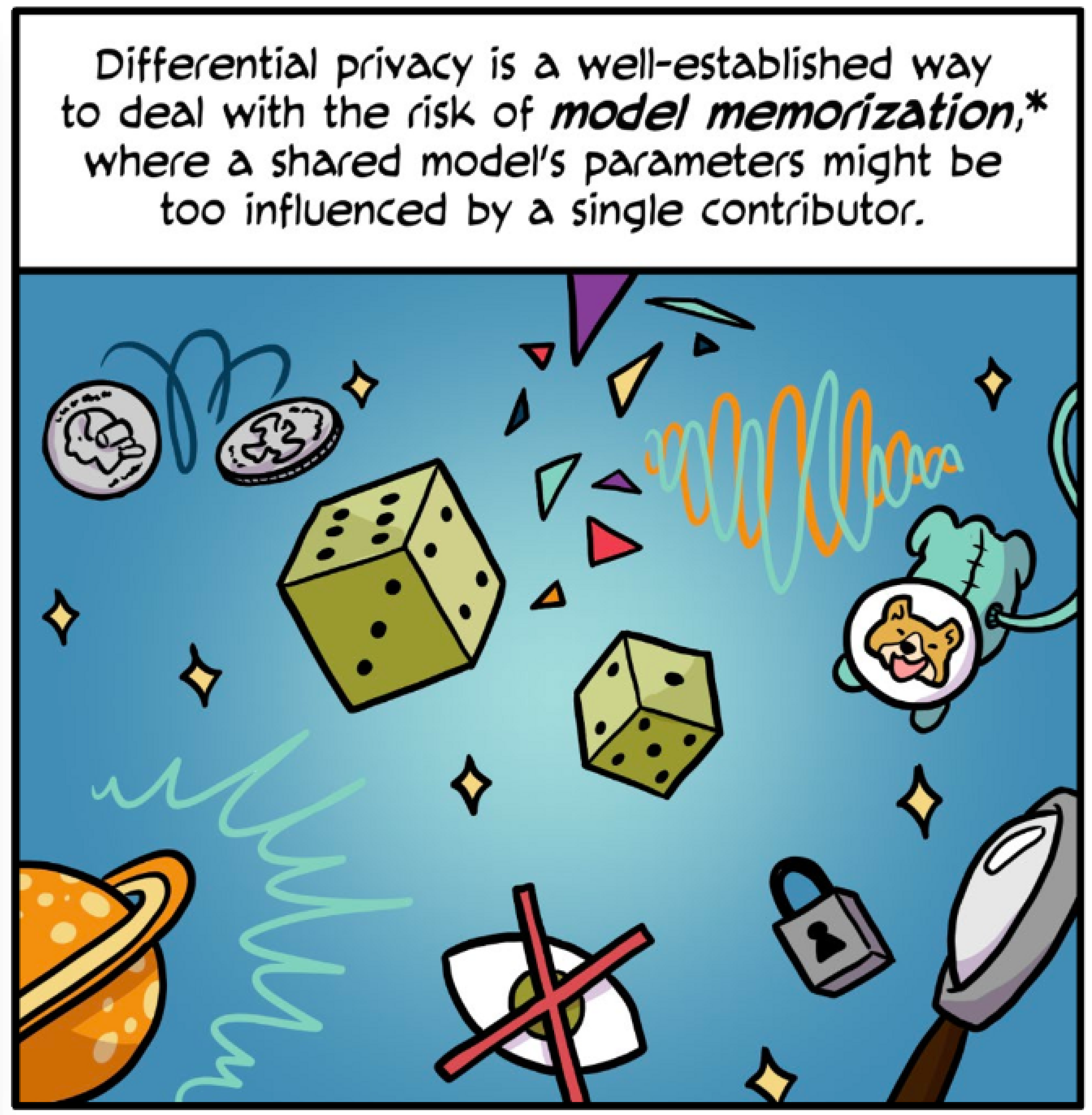
Differential Privacy (DP) is not a specific algorithm or technique. DP is a framework to reason about the personal information contained in data. In the context of SeDS, we can use DP to specify principled bounds determining what and how much personal information will be shared. Once defined, these bounds will guide the design of our privacy mechanism.

Benefits of DP

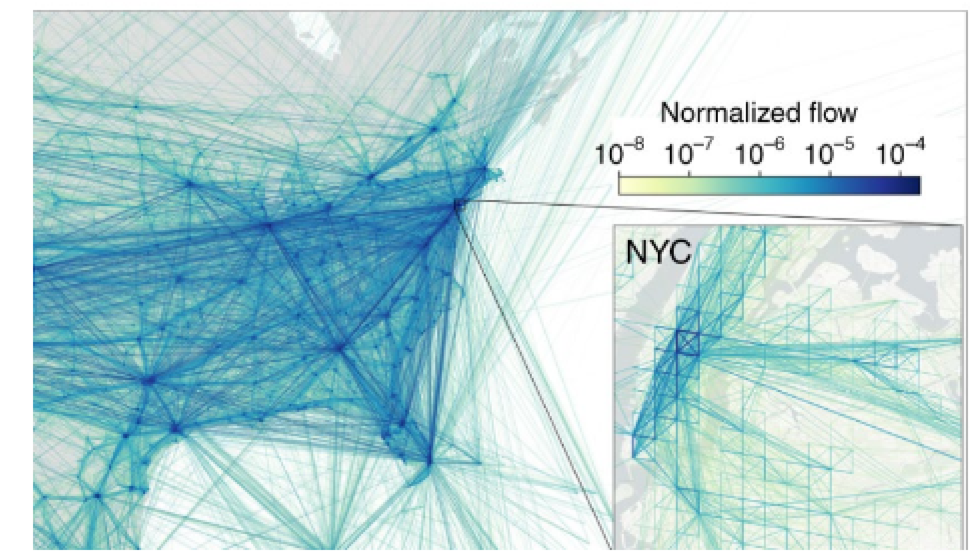
DP is a privacy framework we can use to specify, implement and communicate the privacy story around SeDS. Key benefits of DP include:

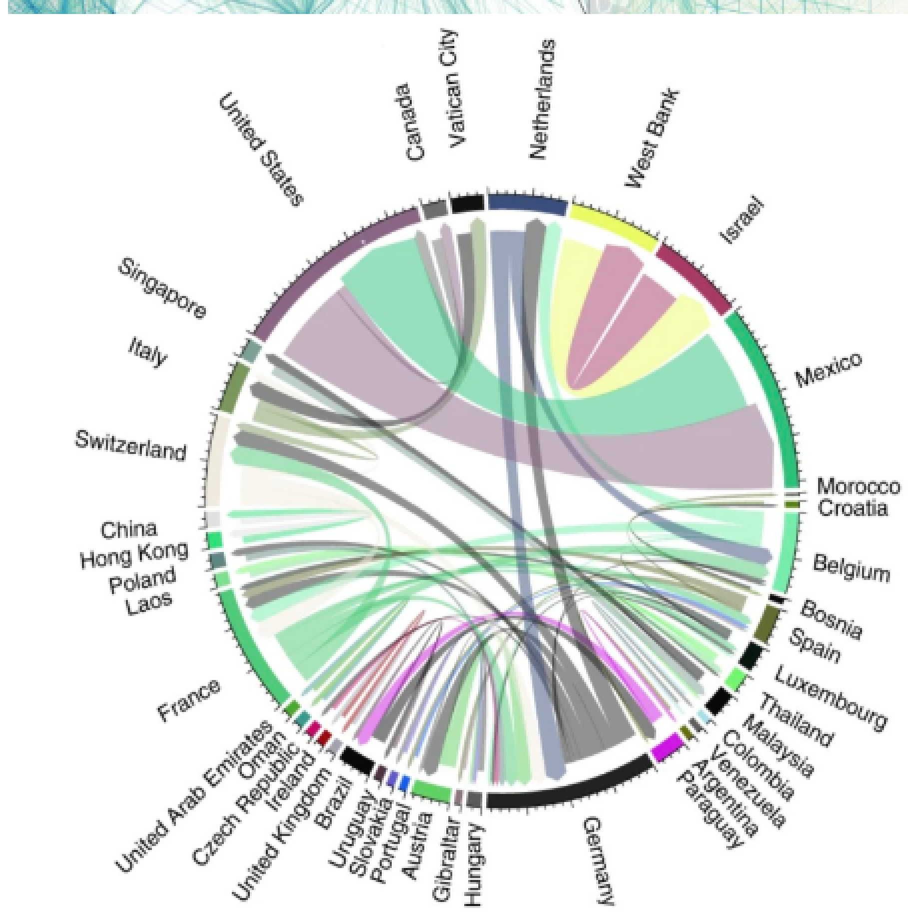
Robust privacy guarantees: DP allows us to make strict and principled statements about privacy. If we enforce a certain DP specification, is it mathematically impossible to extract more information from the data than intended. This is particularly important when sharing data externally (as is the case for SeDS) given that we have limited control over the data after it has been released. Common sources of privacy issues DP is robustly protects against include:

Internal Google Document



Google AI Comic





Variation in Mobility

REDACTED FOR PUBLIC FILING

K-anonymity Formal Privacy Definition

Privacy definition that requires that any released data record is **indistinguishable** from at least $k - 1$ other records.

Query	Location	Device
best Mexican food	(38.8977°, 77.0365°)	Pixel9a-Android15-v23.523

Indistinguishable

Query	Location	Device
best Mexican food	(38.8977°, 77.0365°)	Pixel9a-Android15-v23.523

Indistinguishable?

Query	Location	Device
best **m**exican food	(38.8977°, 77.036**4**°)	Pixel9a-Android15-v**22.173**

Indistinguishable?

Query	Location	Device
mexican **restaurant**	(38.8977°, 77.0365°)	Pixel9a-Android15-v**21.083**

REDACTED FOR PUBLIC FILING

How to Satisfy *K*-anonymity

Source data (k=**1**)

Query	Location	Device
best Mexican food	(38.8977°, 77.0365°)	Pixel9a-Android15-v23.523
best Mexican food	(38.8977°, 77.0365°)	Pixel9a-Android15-v23.523
best **m**exican food	(38.8977°, 77.036**4**°)	Pixel9a-Android15-v**22.173**
mexican **restaurant**	(38.8977°, 77.0365°)	Pixel9a-Android15-v**21.083**

Record removal

Released data (k=**2**)

Query	Location	Device
best Mexican food	(38.8977°, 77.0365°)	Pixel9a-Android15-v23.523
best Mexican food	(38.8977°, 77.0365°)	Pixel9a-Android15-v23.523
best **m**exican food	(38.8977°, 77.036**4**°)	Pixel9a-Android15-v**22.173**
mexican **restaurant**	(38.8977°, 77.0365°)	Pixel9a-Android15-v**21.083**

How to Satisfy *K*-anonymity with Utility

Source data
(k=**1**)

Query	Location	Device
best Mexican food	(38.8977°, 77.0365°)	Pixel9a-Android15-v23.523
best Mexican food	(38.8977°, 77.0365°)	Pixel9a-Android15-v23.523
best **m**exican food	(38.8977°, 77.036**4**°)	Pixel9a-Android15-v**22.173**
mexican **restaurant**	(38.8977°, 77.0365°)	Pixel9a-Android15-v**21.083**

Generalization

Suppression

Released data
(k=**3**)

Query	Location	Device
best Mexican food	**DC 20500**	Pixel9a-Android15
best Mexican food	**DC 20500**	Pixel9a-Android15
best **M**exican food	**DC 20500**	Pixel9a-Android15
mexican **restaurant**	DC 20500	Pixel9a-Android15-v**21.083**

REDACTED FOR PUBLIC FILING

Better Generalization Improves Utility

Source data
(k=**1**)

Query	Location	Device
best Mexican food	(38.8977°, 77.0365°)	Pixel9a-Android15-v23.523
best Mexican food	(38.8977°, 77.0365°)	Pixel9a-Android15-v23.523
best **m**exican food	(38.8977°, 77.036**4**°)	Pixel9a-Android15-v**22.173**
mexican **restaurant**	(38.8977°, 77.0365°)	Pixel9a-Android15-v**21.083**

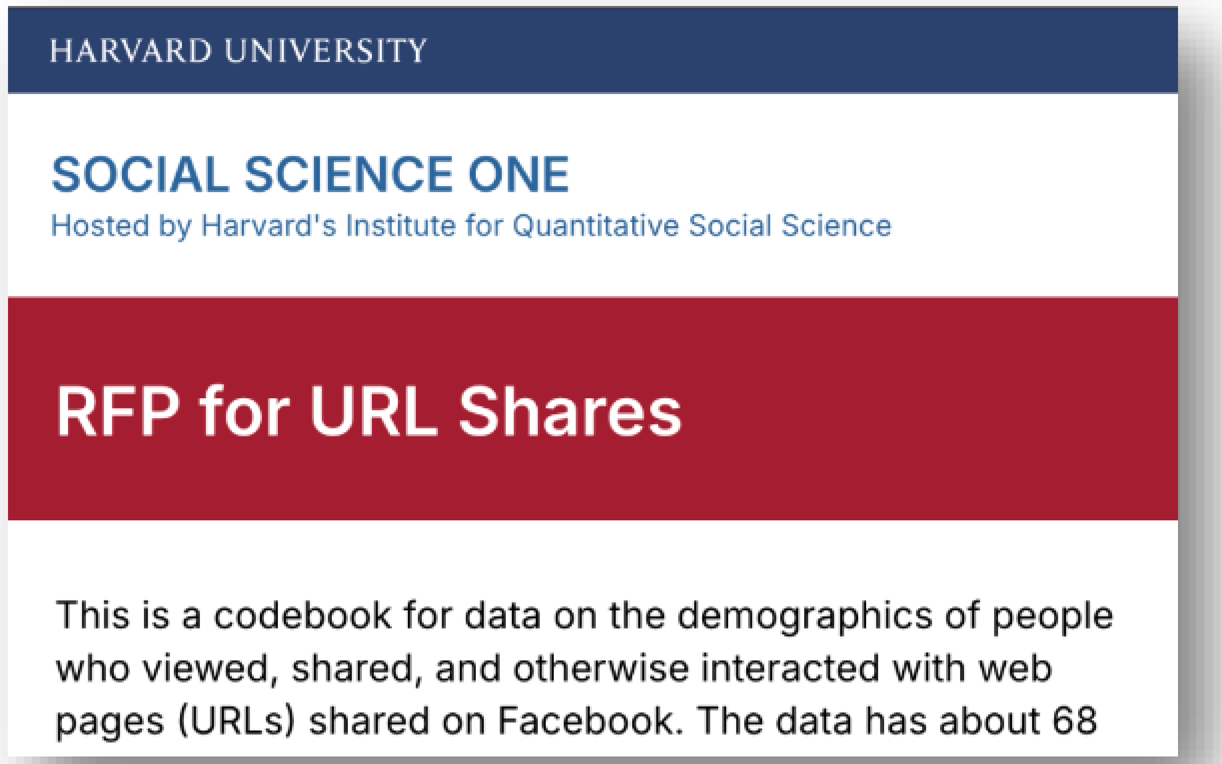
Generalization

Suppression

Released data
(k=**4**)

Query Intent	Location	Device
Mexican restaurant	**DC 20500**	Pixel9a-Android15
Mexican restaurant	**DC 20500**	Pixel9a-Android15
Mexican restaurant	**DC 20500**	Pixel9a-Android15
Mexican restaurant	**DC 20500**	Pixel9a-Android15

Example uses Generalization for *K*-anonymity



HARVARD UNIVERSITY

SOCIAL SCIENCE ONE
Hosted by Harvard's Institute for Quantitative Social Science

RFP for URL Shares

This is a codebook for data on the demographics of people who viewed, shared, and otherwise interacted with web pages (URLs) shared on Facebook. The data has about 68

U.S. CENTERS FOR DISEASE CONTROL AND PREVENTION

CDC
Public Use Data

Generalization
Partial Suppression
L-diversity

Cloudflare
Validating Leaked Passwords

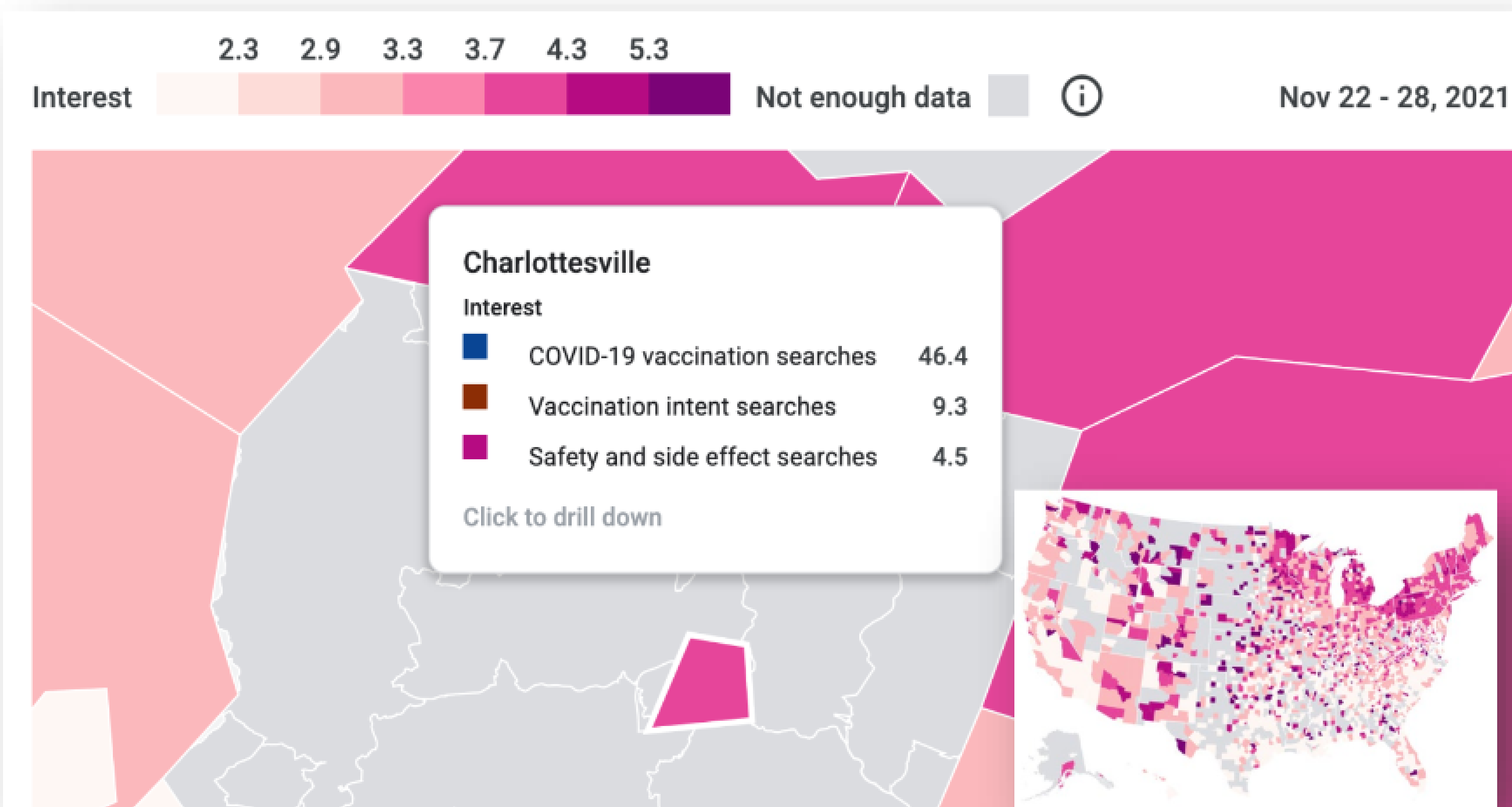
Generalization
Partial Suppression

Facebook
URLs Dataset

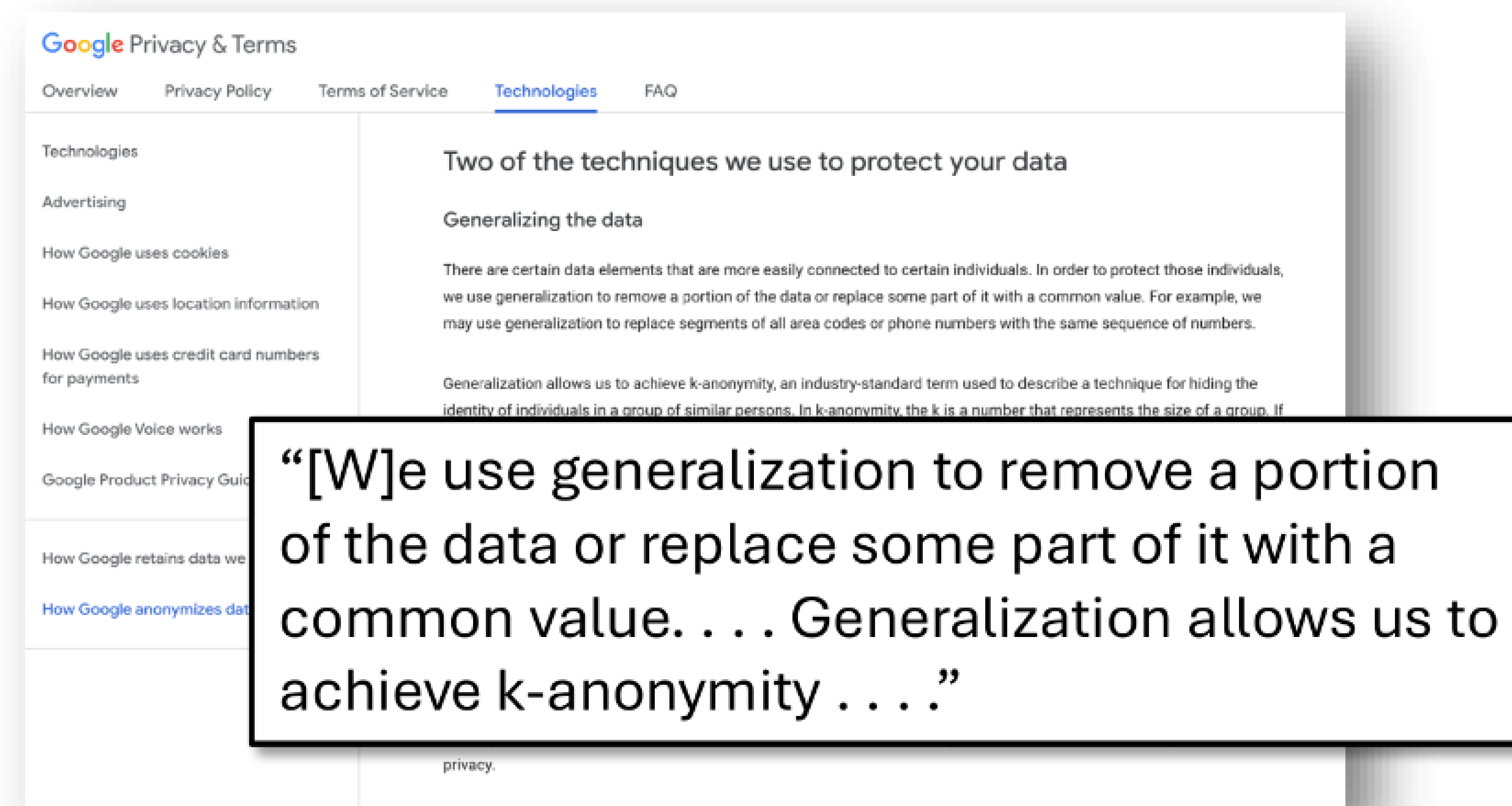
Generalization
Partial Suppression
Differential Privacy

REDACTED FOR PUBLIC FILING

Uses of Generalization for Privacy at Google



COVID-19 Vaccination Search Insights
Generalization (Geographic, Time, Grouping search queries)



Google’s Privacy Policy
Generalization for k-anonymity

REDACTED FOR PUBLIC FILING

Google's Data Sharing Implementation For DMA

Anonymity Set A (Size 5)		
Query	**Country**	**Device**
News today	DE	Mobile
News today	DE	Desktop
News today	DE	Desktop
News today	DE	Mobile
News today	DE	Mobile

Anonymity Set A.1 (Size 2)		
Query	**Country**	**Device**
News today	DE	Desktop
News today	DE	Desktop

Anonymity Set A.2 (Size 3)		
Query	**Country**	**Device**
News today	DE	Mobile
News today	DE	Mobile
News today	DE	Mobile

No Field Suppression
No Generalization
No Spell-Correcting Queries
No Grouping by Query Intent

Google's Experts' Report on DMA
(Dr. Culnane and Prof. Rubenstein)

21. Google identified three additional recovery mechanisms and is working on implementing them. These mechanisms require significant engineering work to develop and will therefore not be ready for the initial dataset, but Google expects to introduce them for the second quarterly release of its Art. 6(11) dataset.

22. First, Google has developed a privacy-safe way to release additional data about low-volume queries. For queries that typically fail to meet the m-threshold for a given country, Google will apply the threshold at an EEA-wide level, and report combined statistics across the EEA instead of country-level data for many queries that do not support finer country-level data.

Generalization by combining all countries

23. Second, Google Search automatically corrects some typos and misspellings in user queries, showing the user results for the corrected query. Before anonymization, Google will replace "typo" queries that were automatically corrected for the results shown to the user with their corrected versions.

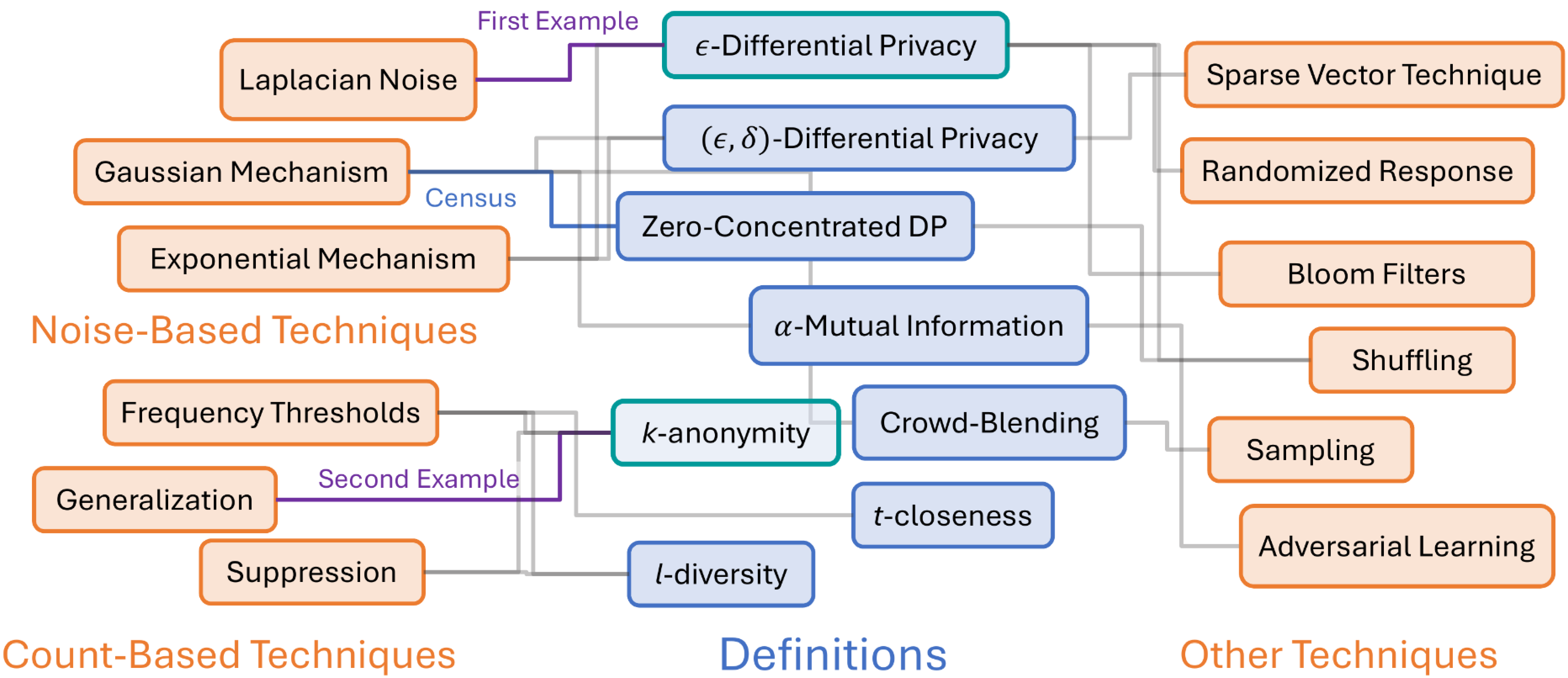
Generalization by fixing "typo" queries

24. Third, Google has developed an additional mechanism to "map" certain low-frequency queries that Search does not automatically correct (*e.g.*, [mssql

Google's Second Response to European Commission
(January 2024, 1¼ years after DMA)

REDACTED FOR PUBLIC FILING

Many Formal Privacy Definitions And Principled Techniques



REDACTED FOR PUBLIC FILING

Opinion 2: PETs Can Be Used To Safely Release Useful Data

1 There are well-established **privacy-enhancing techniques** that can be used to protect sensitive information.

2 Many organizations, including Google, **safely release sensitive data** by using privacy-enhancing techniques.

3 Google can share the data at issue in a way that **assures privacy** while **providing utility.**

REDACTED FOR PUBLIC FILING

Selecting Appropriate Privacy-Enhancing Techniques

Properties of the **source data**	Uses of the **released data**
• Type and amount	• Amount required
• Granularity	• Granularity needed
• Dimensionality	• Correlations used
• Sensitivity	• Accuracy thresholds
• Update frequency	• Sharing frequency
…	…
Disclosure Risk	**Data Utility**

REDACTED FOR PUBLIC FILING

Selecting Privacy-Enhancing Techniques for Data at Issue

Slide 37 from Google's Opening Statement

Plaintiffs' Privacy Expert Offers No Opinion

David Evans, PhD.
DOJ Expert

A. There are many ways to protect text data, and one way is to use the frequency-based method to achieve a definition similar to K-anonymity.

Q. That is what you propose should be done here?

A. I don't make any proposal as to what should be done here. I just speak to the availability of many different privacy-enhancing techniques that could be used to satisfy the requirements of the RPFJ.

Evans (DOJ) Trial Tr. 130:10-22

Google
RDXD-01.037

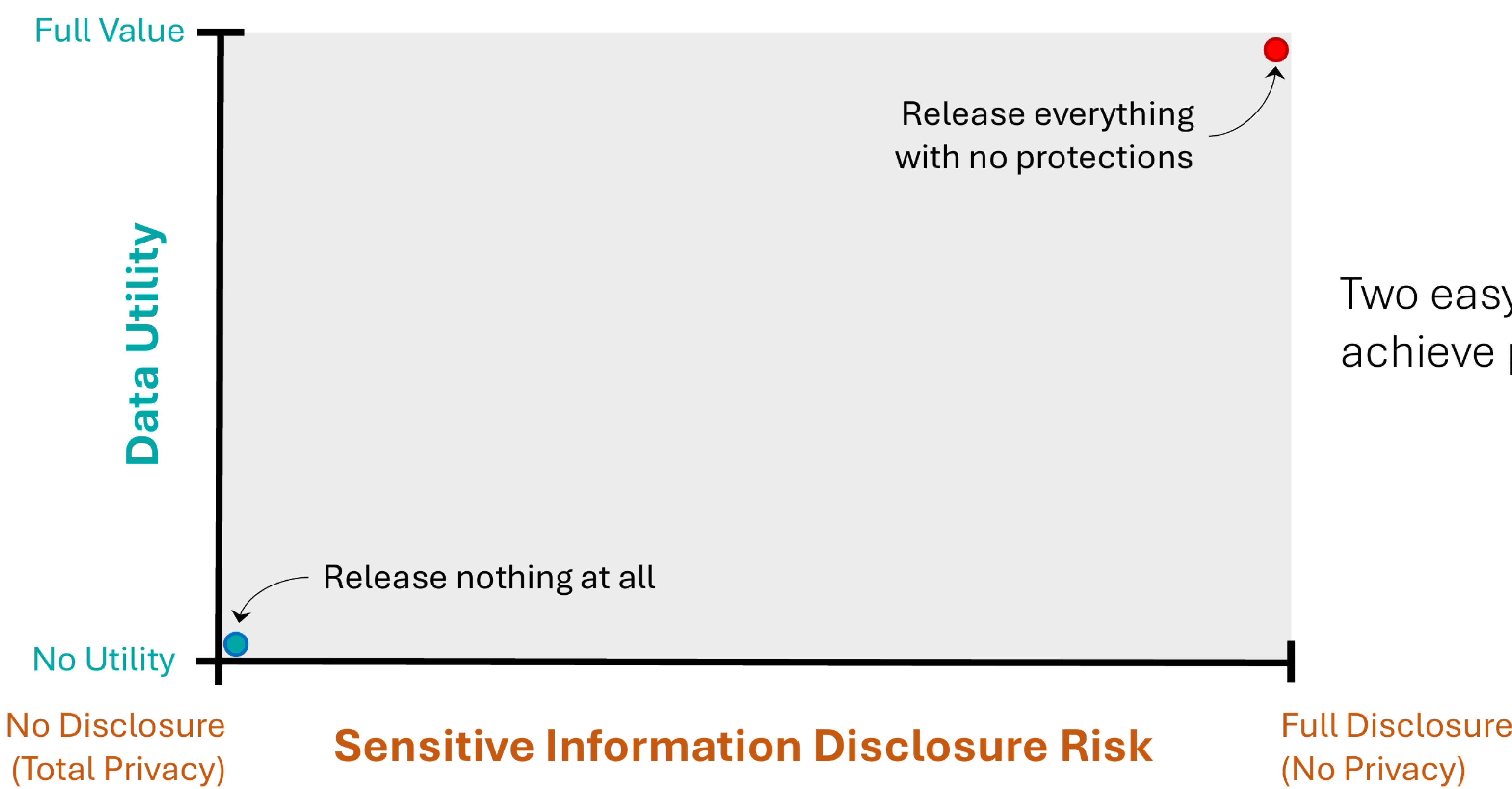
REDACTED FOR PUBLIC FILING

Privacy–Utility Tradeoff

Source: Adapted from Opening Report Figure 1: Privacy – utility tradeoff curve.

REDACTED FOR PUBLIC FILING

Privacy–Utility Tradeoff

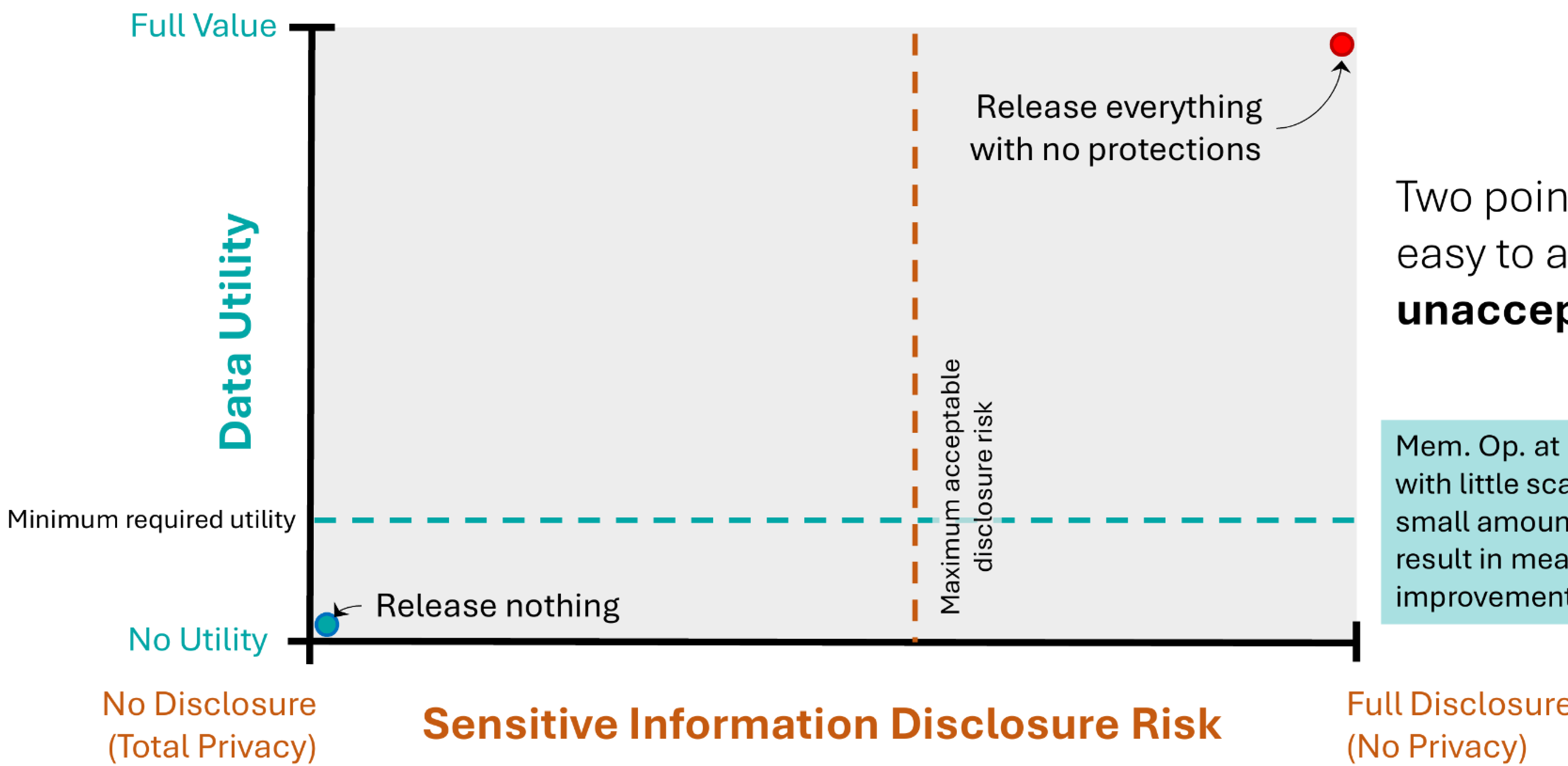


Two easy-to-achieve points

Source: Adapted from Opening Report Figure 1: Privacy – utility tradeoff curve.

REDACTED FOR PUBLIC FILING

Privacy–Utility Tradeoff



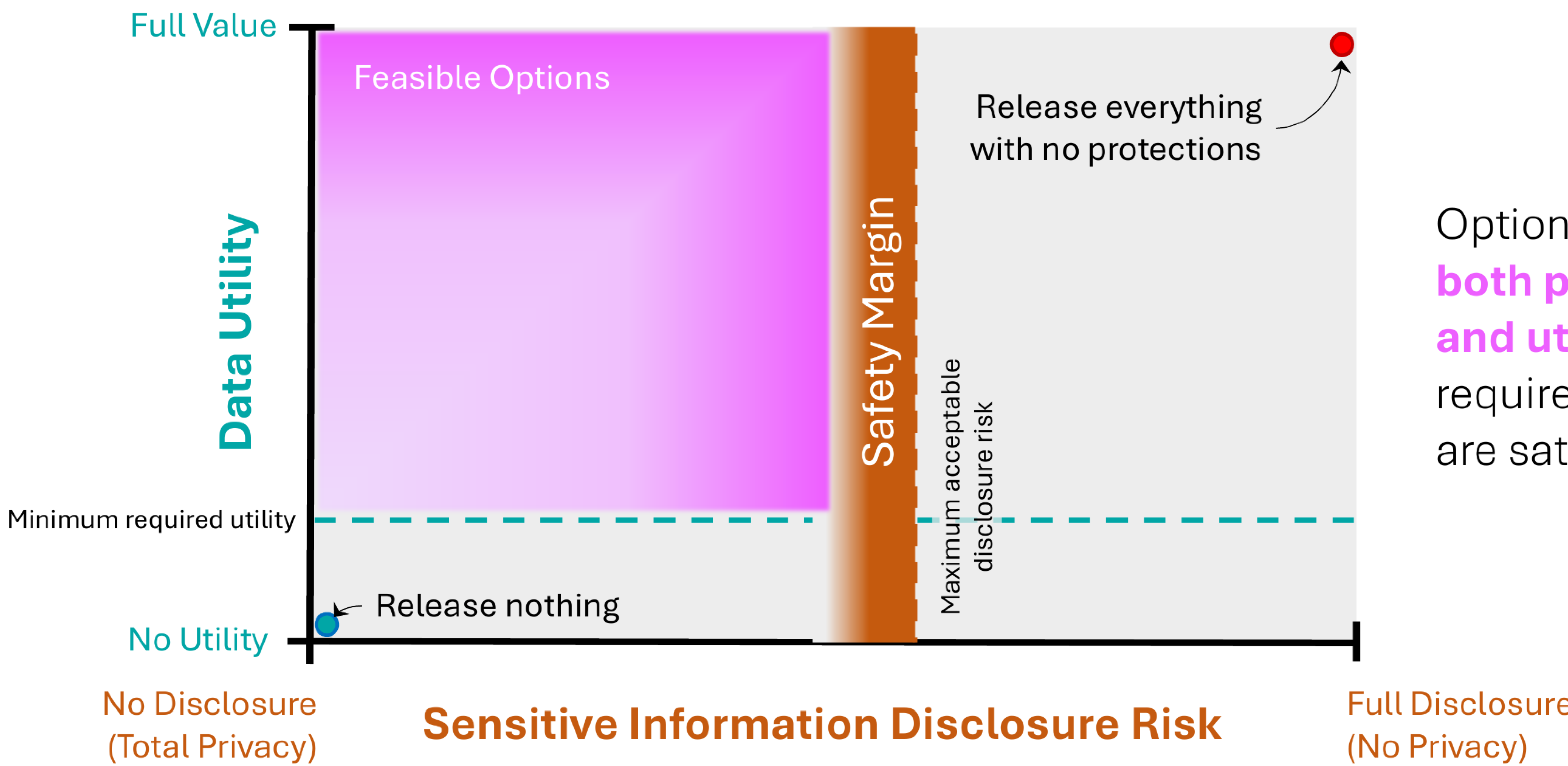
Two points are easy to achieve but **unacceptable**

Mem. Op. at 106. "For GSEs with little scale, even a small amount of data can result in meaningful improvements."

Source: Adapted from Opening Report Figure 1: Privacy – utility tradeoff curve.

REDACTED FOR PUBLIC FILING

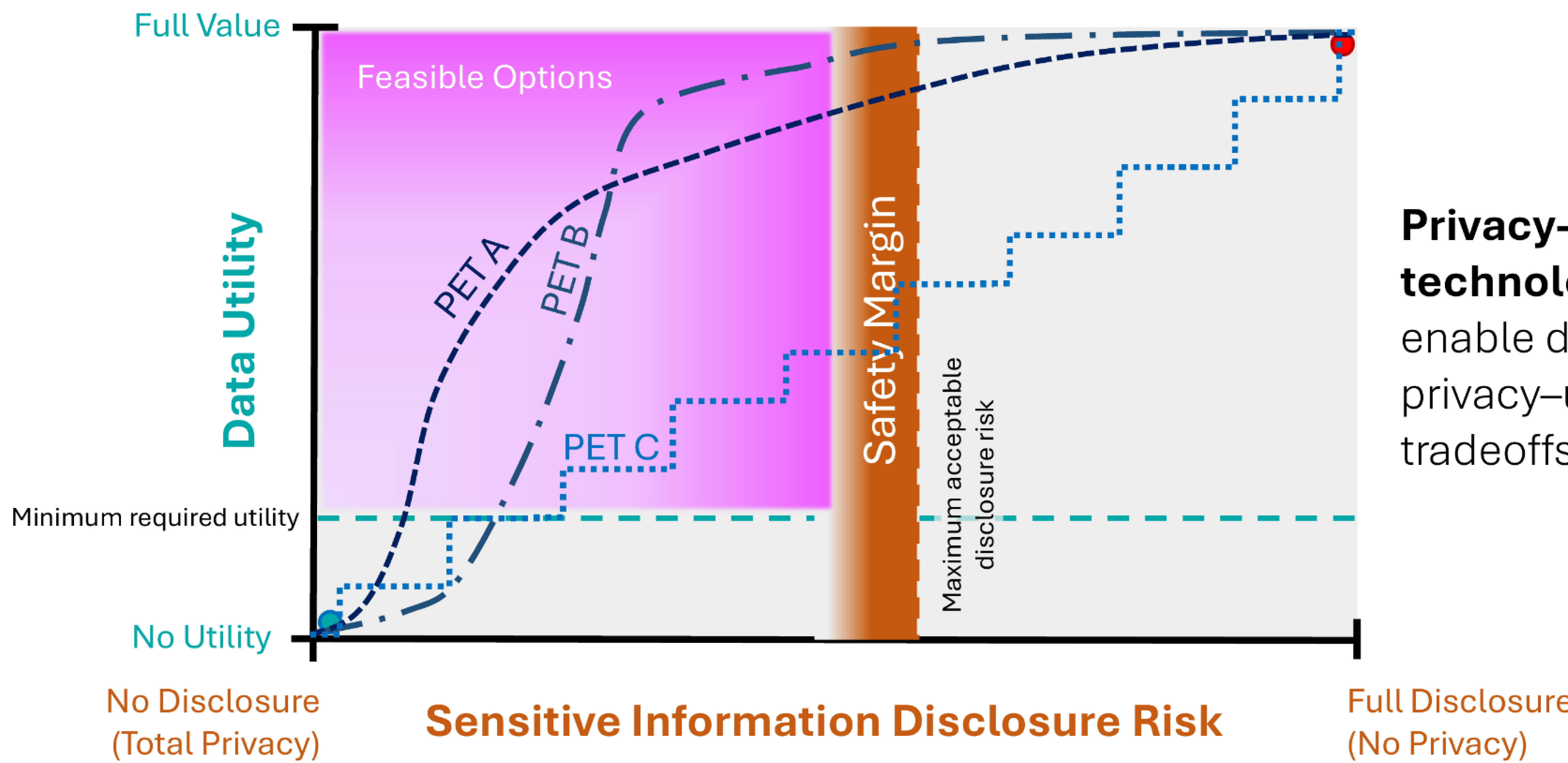
Privacy–Utility Tradeoff



Options where **both privacy and utility** requirements are satisfied

REDACTED FOR PUBLIC FILING

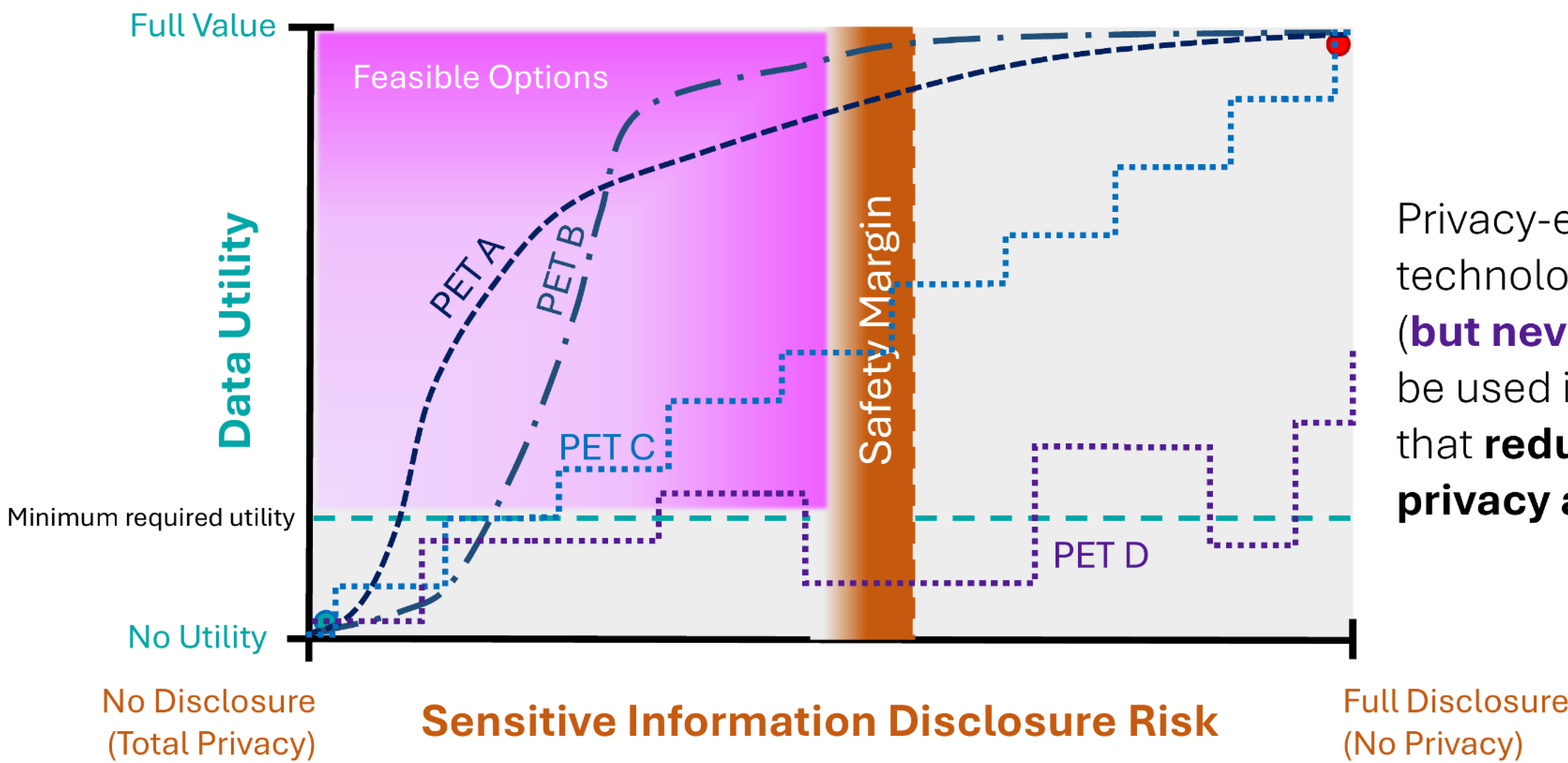
Privacy–Utility Tradeoff



Privacy-enhancing technologies enable different privacy–utility tradeoffs

REDACTED FOR PUBLIC FILING

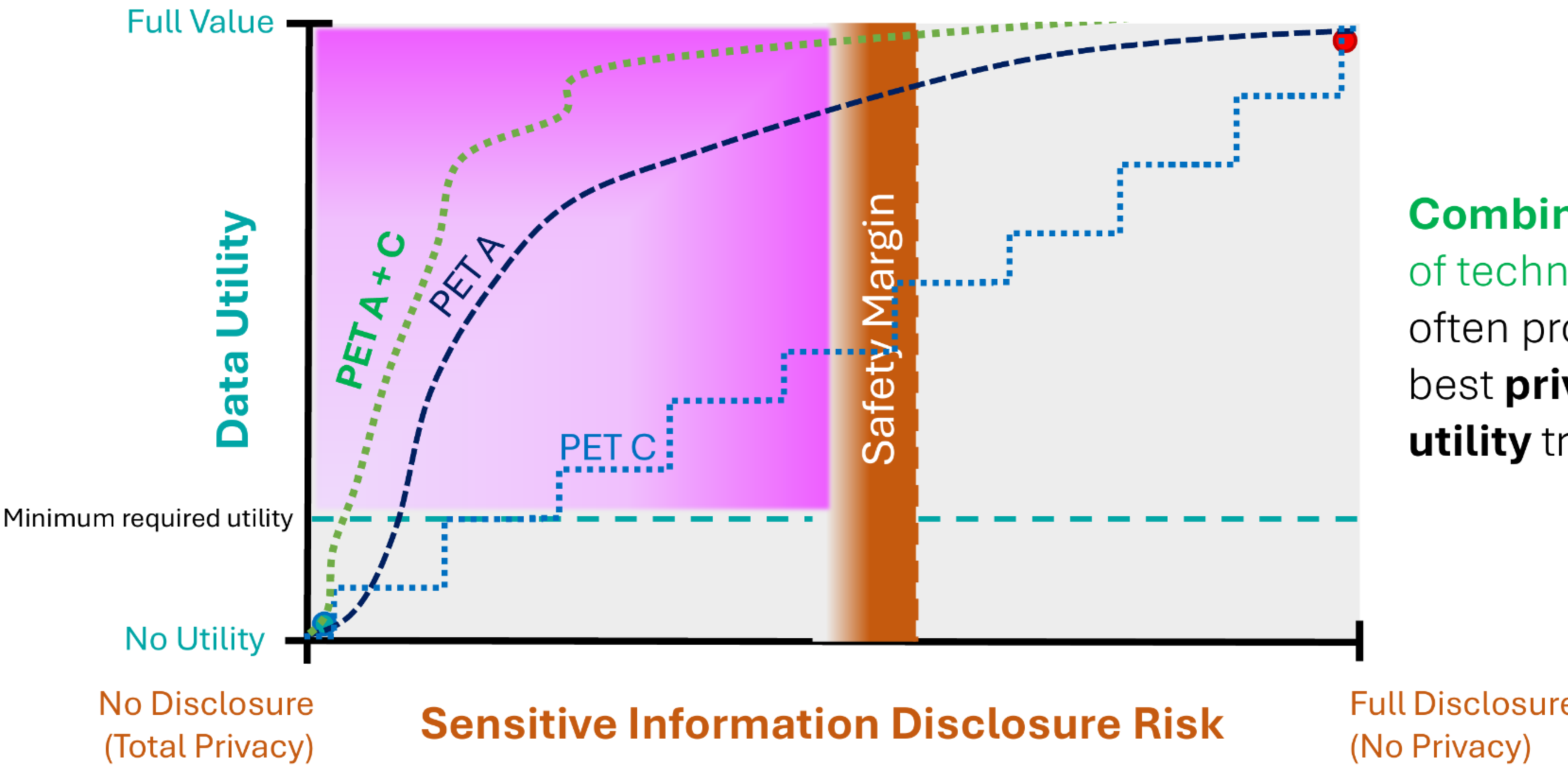
Privacy–Utility Tradeoff



Privacy-enhancing technologies could (**but never should**) be used in ways that **reduce both privacy and utility**

REDACTED FOR PUBLIC FILING

Privacy–Utility Tradeoff



Combinations of techniques often provide the best **privacy–utility** tradeoff

Source: Adapted from Opening Report Figure 1: Privacy – utility tradeoff curve.

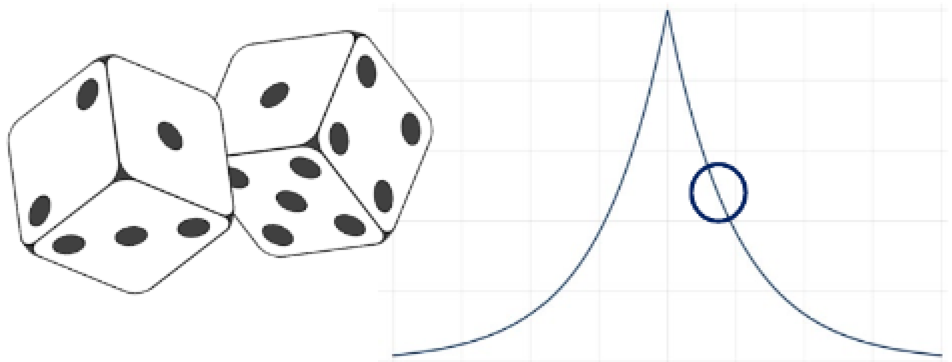
REDACTED FOR PUBLIC FILING

Example: Combining PETs

Query	Location
Mexican restaurant	DC 20500
resturant mexican	DC 20500
Mexican restuarant	DC 20500
mexican history	DC 20500

Generalization to select (query, location)

Differential Privacy Noise to release statistics



Query	Location	User Behaviors	
Mexican restaurant	DC 20500	Count	631.52
		Clicks	472.24
		Average Time (s)	2.24
		Abandoned	18.02
		districttaco.com	83.24
		dlenadc.com	45.29
		mividamexico.com	21.20
		...	

Source: Korolova, Aleksandra, Krishnaram Kenthapadi et al:. "Releasing search queries and clicks privately". In Proceedings of the 18th International Conference on World Wide Web. url: https://dl.acm.org/doi/10.1145/1526709.1526733.

REDACTED FOR PUBLIC FILING

Many Organizations Balance Privacy and Utility

An official website of the United States government Here's how you know

Census Bureau Sets Key Parameters to Protect Privacy in 2020 Census DHC. Settings Reflect Feedback-Driven Improvements

The U.S. Census Bureau's Data Stewardship Executive Policy Committee (DSEP) has selected the final settings and parameters for the Disclosure Avoidance System (DAS) for the 2020 Demographic and Housing Characteristics File (DHC), the next major release from the 2020 Census. Based on these settings, Census Bureau staff are now preparing the final data product for publication in May 2023.

The DHC includes, unchanged, the six tables that were released in August 2021 as part of the final Redistricting Data (P.L. 94-171) Summary File. It also includes tables that support the Demographic Profiles and general statistical descriptions of the population and housing in the United States and the Commonwealth of Puerto Rico.

Feedback-Driven Improvements

The selections, approved October 20, 2022, reflect feedback from data users on a series of demonstration data products and accompanying metrics that applied iterative versions

Engineering Blog

Privacy Preserving Single Post Analytics

Ryan Rogers
December 12, 2023

Authors: ***Ryan Rogers, Subbu Subramaniam, Lin Xu***

Contributors: **Mark Cesar**, **Praveen Chaganlal**, **Xinlin Zhou**, **Jefferson Lai**, **Jennifer Li**, **Stephanie Chung**, **Margaret Taormina**, **Gavin Uathavikul**, **Laura Chen**, **Rahul Tandra**, **Siyao Sun**, **Vinyas Maddi**, **Shuai Zhang**

Content creators post on LinkedIn with the goal of reaching and engaging specific audiences. Post analytics helps

Machine Learning Research

Highlight | April 14, 2025

Privacy

Understanding Aggregate Trends for Apple Intelligence Using Differential Privacy

At Apple, we believe privacy is a fundamental human right. And we believe in giving our users a great experience while protecting their privacy. For years, we've used techniques like differential privacy › as part of our opt-in device analytics program. This lets us gain insights into how our products are used, so we can improve them, while protecting user privacy by preventing Apple from seeing individual-level data from those

Source: https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/newsletters/key-parameters-set-2020-census-dhc.html; https://www.linkedin.com/blog/engineering/trust-and-safety/privacy-preserving-single-post-analytics; https://machinelearning.apple.com/research/differential-privacy-aggregate-trends.

REDACTED FOR PUBLIC FILING

Google Has Experience Balancing Privacy and Utility

Confidential

Privileged and Confidential

SeDS Engineering Working Group

DP for SeDS

created: Jul 20, 2022
last updated: Jul 20, 2022
author: Dennis Kraft, Alex Kulesza, Sergei Vassilvitskii, Rachel Wei, Matthew Jagielski
status: WIP

"Over the years, we have **gained valuable experience** with DP, how it **translates to privacy policy** and how to implement it technically. Moreover, we have developed a **mature set of tools** to deploy DP quickly and efficiently."

Robust privacy guarantees: DP allows us to make strict and principled statements about privacy. If we enforce a certain DP specification, is it mathematically impossible to extract more information from the data than intended. This is particularly important when sharing data externally (as is the case for SeDS) given that we have limited control over the data after it has been released. Common sources of privacy issues DP is robustly protects against include:

Internal Google Document

Differentially Private Stream Processing at Scale

Bing Zhang[1], Vadym Doroshenko[†1], Peter Kairouz[†3], Thomas Steinke[†2], Abhradeep Thakurta[†2], Ziyin Ma[1], Eidan Cohen[1], Himani Apte[1], Jodi Spacek[1]
[1]Google
[2]Google DeepMind
[3]Google Research
{zhangbing,dvadym,kairouz,steinke,athakurta,ziyinma,eidanch,himaniapte,jodes}@google.com

2 Jul 2024

ABSTRACT
We design, to the best of our knowledge, the first differentially private (DP) stream aggregation processing system at scale. Our system – *Differential Privacy SQL Pipelines (DP-SQLP)* – is built using a streaming framework similar to Spark streaming, and is built on top of the Spanner database and the F1 query engine from Google.

Towards designing DP-SQLP we make both algorithmic and systemic advances, namely, we (i) design a novel (user-level)

called Differential Privacy SQL Pipelines (DP-SQLP), and make algorithmic advances along the way to cater to the scalability needs of it. DP-SQLP is implemented using a streaming framework similar to Spark streaming [52], and is built on top of the Spanner database [12] and F1 query engine [43] from Google. We also present production applications with two use cases in Section 6. The first is a real world use case that deploys DP-SQLP in *Google Shopping* to generate streaming page-view counts. The second applies the streaming DP algorithm to *Google Trends*.

"In terms of **data utility** after adopting DP-SQLP, we were able to retain 59% of the page-view.... to **99.9% for pages with an average view rate of 60 views/hour**. When comparing noised impression counts with the raw counts, the **relative error** is around 11%.... to ensure **user level DP guarantee**, per day. We use $\varepsilon = 1$ for ...."

Proceedings of the VLDB Endowment, Vol. 17, No. 12 ISSN 2150-8097.
doi:XX.XX/XXX.XX

it has appeared in the most recent event time window, then this potentially leaks information. Naively,

[1]Our work is most closely related to [9]. We defer a full comparison to Section 1.2.

Google Research Paper

REDACTED FOR PUBLIC FILING

Opinion 3: Data at Issue Can Be Shared Safely

1. There are well-established **privacy-enhancing techniques** that can be used to protect sensitive information.

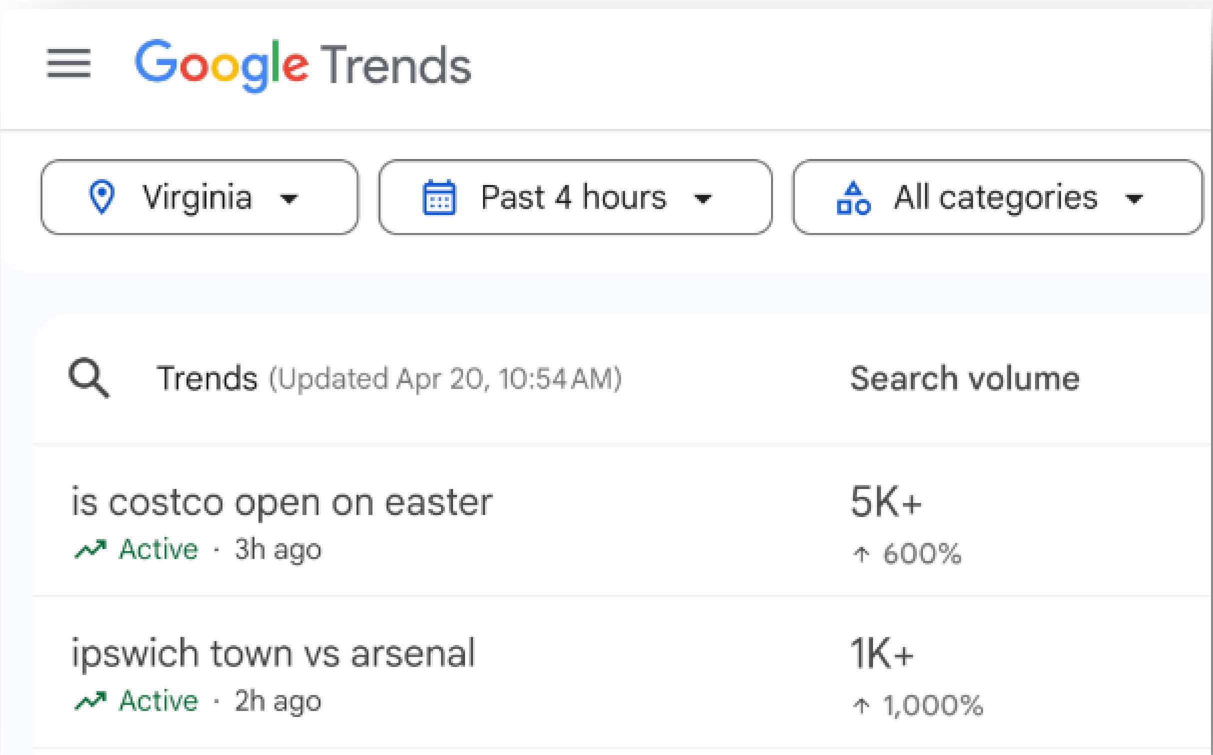
2. Many organizations, including Google, **safely release sensitive data** by using privacy-enhancing techniques.

3. Google can share the data at issue in a way that **assures privacy** while **providing utility**.

REDACTED FOR PUBLIC FILING

Google Currently Uses PETs to Release Similar Data

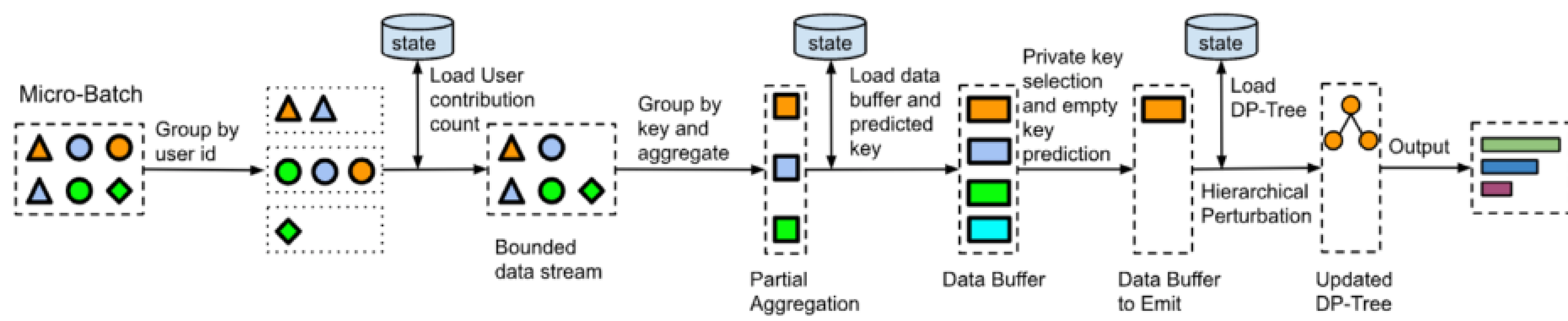
Search Queries



Google Trends

Covid Symptoms, Vaccination Insights, …

User Interactions



Google Shopping

Advertising Data

Private-Set Intersection, Analytics

Real-time

Google Trends, Google Shopping

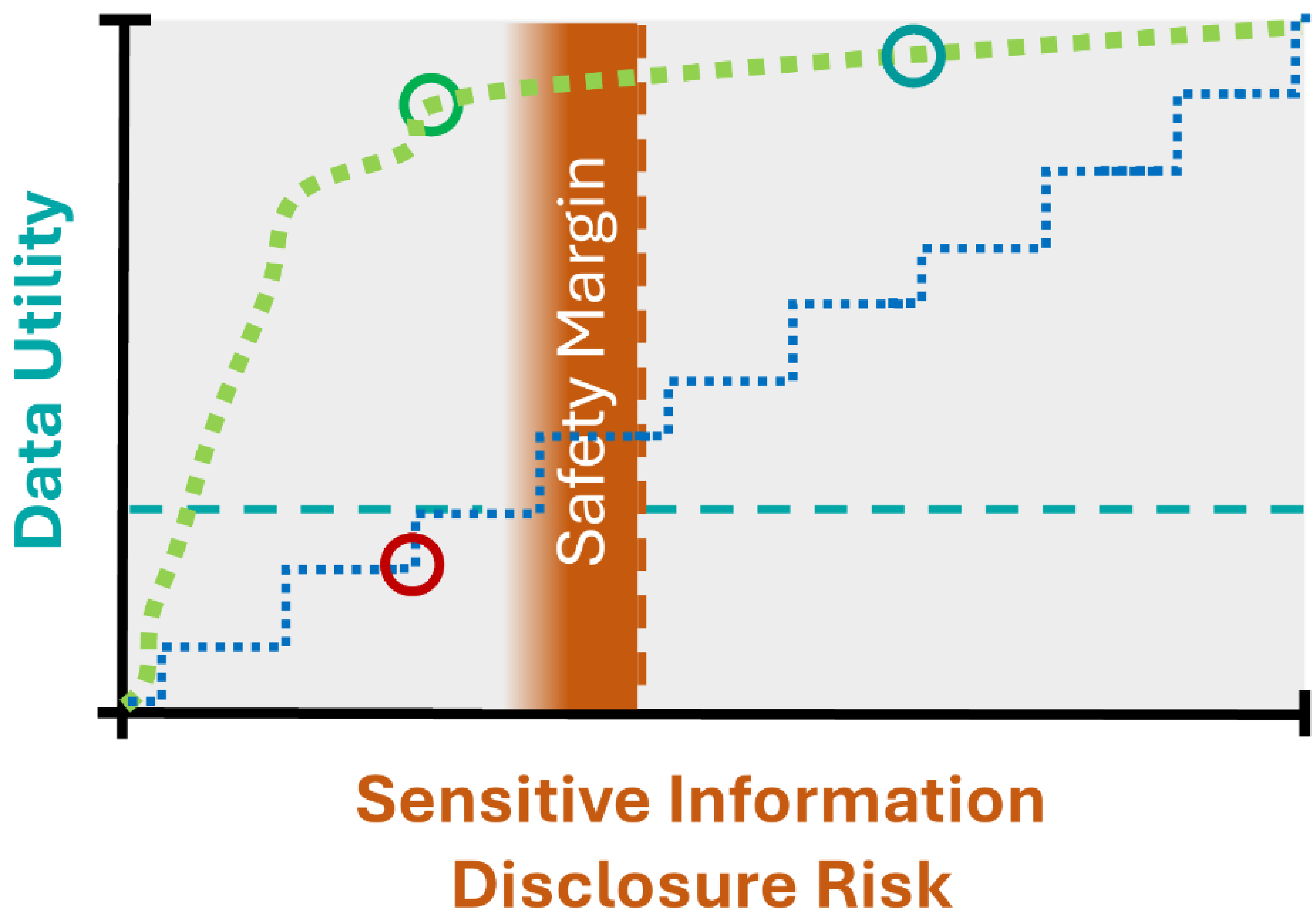
Enormous Scale

Plume (Trillions of records with DP)

Source: https://trends.google.com/trends/; https://pair-code.github.io/covid19_symptom_dataset/?country=US; Zhang, p. 8, url: https://arxiv.org/pdf/2303.18086; https://www.youtube.com/watch?v=gKehlyFK8lg

REDACTED FOR PUBLIC FILING

Implementing the Data Sharing Remedy



The **Technical Committee** with understanding of **intended uses** and **data content** can assess use of privacy-enhancing techniques and parameters for an appropriate **privacy–utility** tradeoff.

Source: Adapted from Opening Report Figure 1: Privacy-utility tradeoff curve.

REDACTED FOR PUBLIC FILING

Google's Expert Agrees: Data Can Be Shared Safely

Dr. Culnane's Deposition

Q. Dr. Culnane, you believe that it is possible for Google to share what you call the DOJ search data by applying privacy-enhancing techniques to achieve suitable privacy safeguards, don't you?

A. Yes.

Q. Do you have any opinion as to whether it is technologically feasible to share the DOJ search data as Plaintiffs describe in Plaintiffs' Proposed Final Judgment?

A. The subject of my report is looking at the ability to do that safely, so there is an opinion as -- if it is correctly protected, and in my view, if you protect personal data as opposed to PII, then you can anonymize the dataset. If you successfully do that, then you can protect privacy by doing that, yes.

Source: Culnane (Google Expert) Dep. 11:8–12, 21:17–22:5.

REDACTED FOR PUBLIC FILING

Conclusion

1. There are well-established **privacy-enhancing techniques** that can be used to protect sensitive information.

2. Many organizations, including Google, **safely release sensitive data** by using privacy-enhancing techniques.

3. Google can share the data at issue in a way that **assures privacy while providing utility.**

REDACTED FOR PUBLIC FILING