

CSCI E-11

Class 7

March 10, 2026

Identity and Privacy



Harvard
Extension School
HARVARD DIVISION OF
CONTINUING EDUCATION

BBC: TikTok won't use end-to-end encryption.

TikTok won't protect DMs with controversial privacy tech, saying it would put users at risk

6 days ago

Share  Save 

Joe Tidy

Cyber correspondent, BBC World Service



A newly proposed quantum algorithm might break RSA faster than Shor's.

DATA PROTECTION

Quantum Decryption of RSA Is Much Closer Than Expected

For decades, the quantum threat to RSA and ECC encryption has been tied to Shor's algorithm and the assumption that we would need million-qubit quantum computers to make it practical. A newly announced algorithm challenges that assumption and suggests the breaking point could arrive far sooner than expected.



By [Kevin Townsend](#) | March 3, 2026 (9:00 AM ET)



<https://www.securityweek.com/quantum-decryption-of-rsa-is-much-closer-than-expected/>

Preprint

Article

This version is not peer-reviewed.

A Novel Hybrid Quantum Circuit for Integer Factorization: End-to-End Evaluation in Simulation and Real Quantum Hardware

Jesse Van Griensven Thé ^{*}, Victor Oliveira Santos , Bahram Gharabaghi 

<https://www.preprints.org/manuscript/202510.1649>

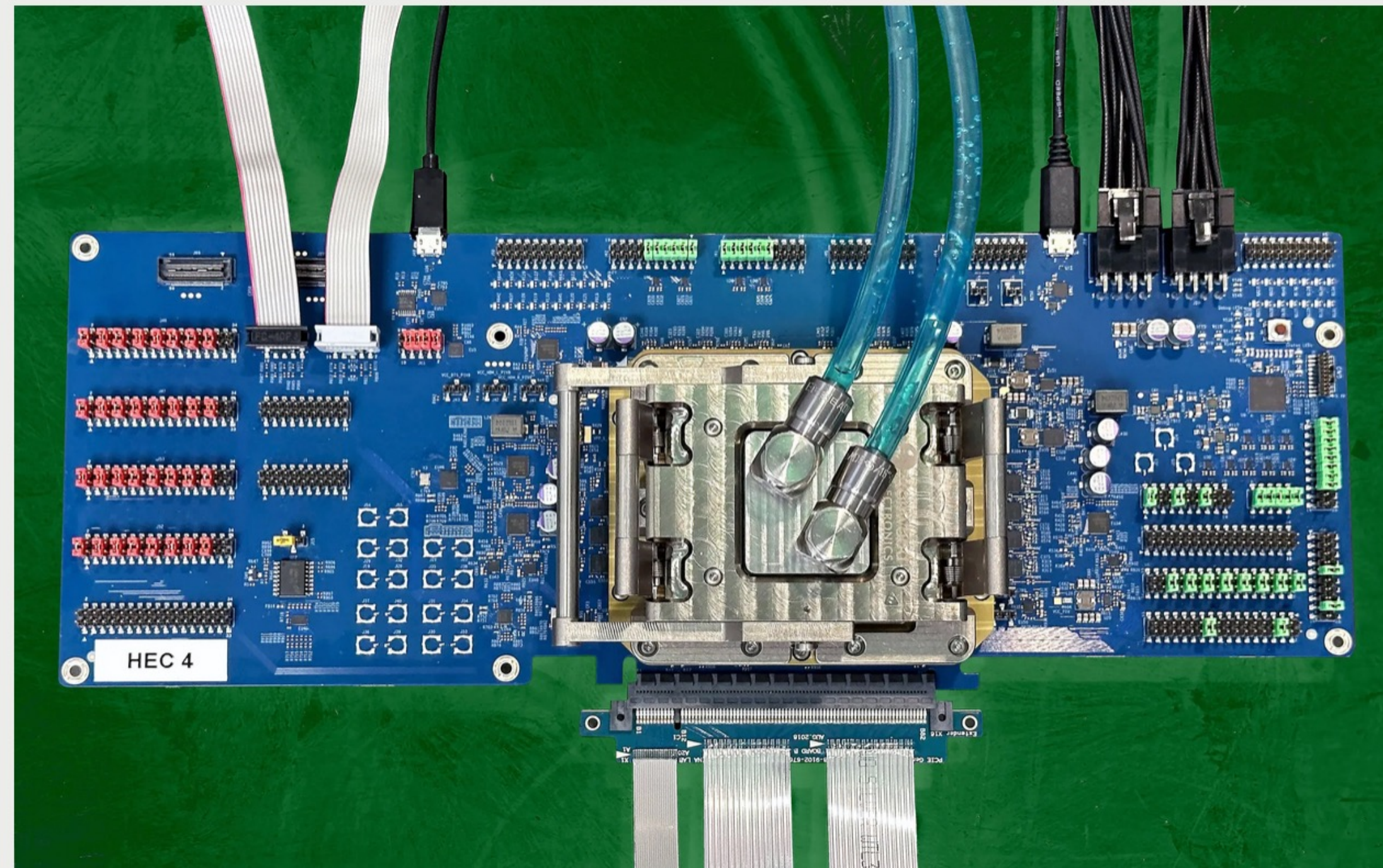
per-qubit or transform-only efficiency claim. **On quantum hardware, JVG reduced the required runtime from 67.8 s to 2 s, and the quantum gate counts by over 98%.** Projection for

Intel claims breakthrough on Fully Homomorphic Encryption (thanks, Burhan!)

Intel Demos Chip to Compute With Encrypted Data › Fully homomorphic encryption chip speeds operations 5,000-fold

BY SAMUEL K. MOORE | 8 HOURS AGO | 6 MIN READ | 

Samuel K. Moore is IEEE Spectrum's semiconductor editor.



Intel's prototype FHE chip can keep a secret. INTEL

<https://spectrum.ieee.org/fhe-intel>

Paper: The Complex Path to Quantum Resistance

security

1 of 28

TEXT ONLY



IS YOUR ORGANIZATION PREPARED?

The Complex Path to Quantum Resistance

ATEFEH MASHATAN AND DOUGLAS HEINTZMAN

T

here is a new technology on the horizon that will forever change the information security and privacy industry landscape. Quantum computing, together with quantum communication, will have many beneficial applications but will also be capable of breaking many of today's most popular cryptographic techniques that help ensure data protection—in particular, confidentiality and integrity of sensitive information. These techniques are ubiquitously embedded in today's digital fabric and implemented by many industries such as finance, health care, utilities, and the broader information communication technology (ICT) community. It is therefore imperative for ICT executives to prepare for the transition from quantum-vulnerable to quantum-resistant technologies.

This transition will be particularly complex, time-

acmqueue | march-april 2021 65

Authentication



Multipass.

Identity authentication relies on establishing that a claim matches an enrollment.



Identity Enrollment

Market-leading biometric solutions

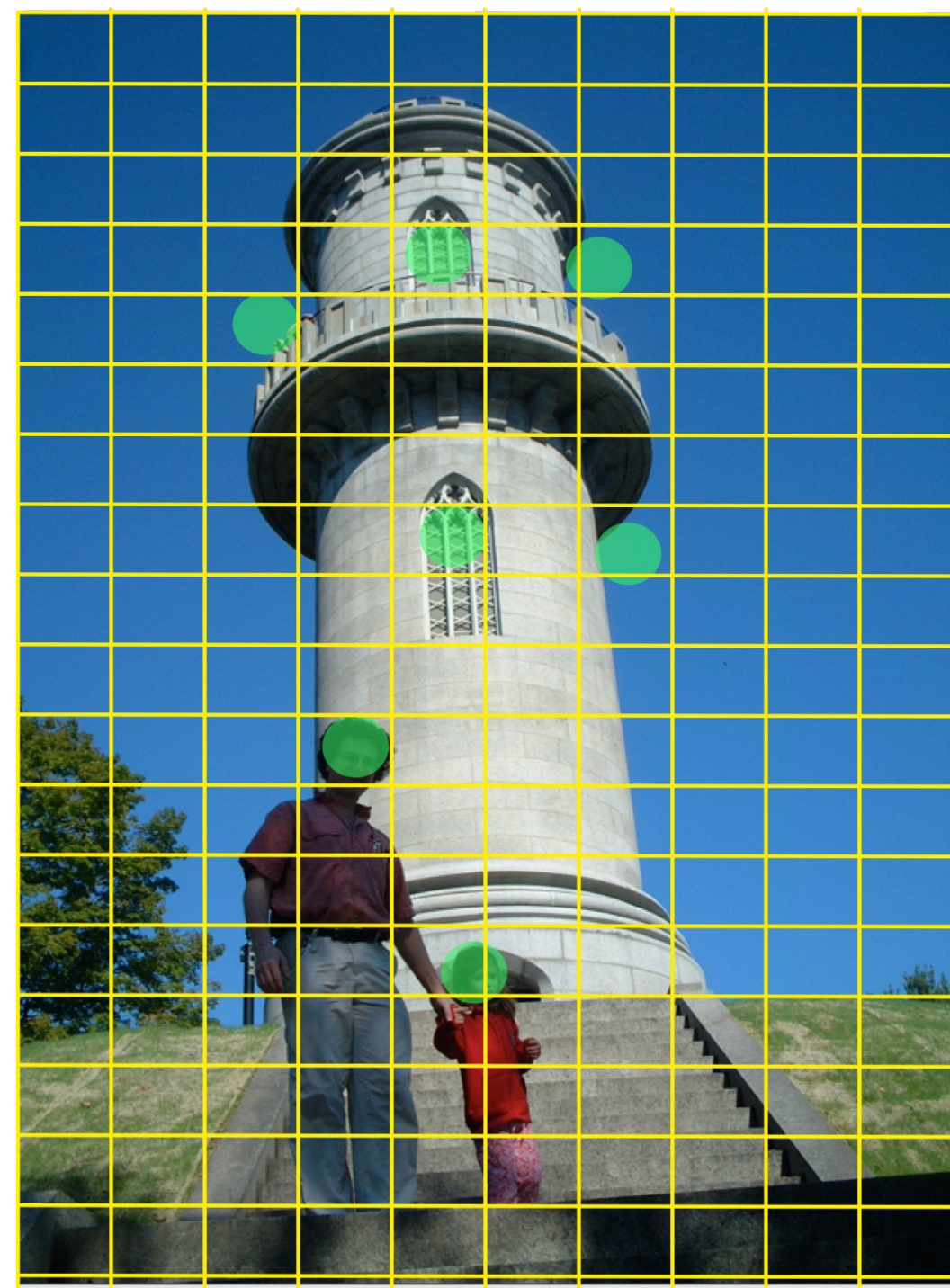
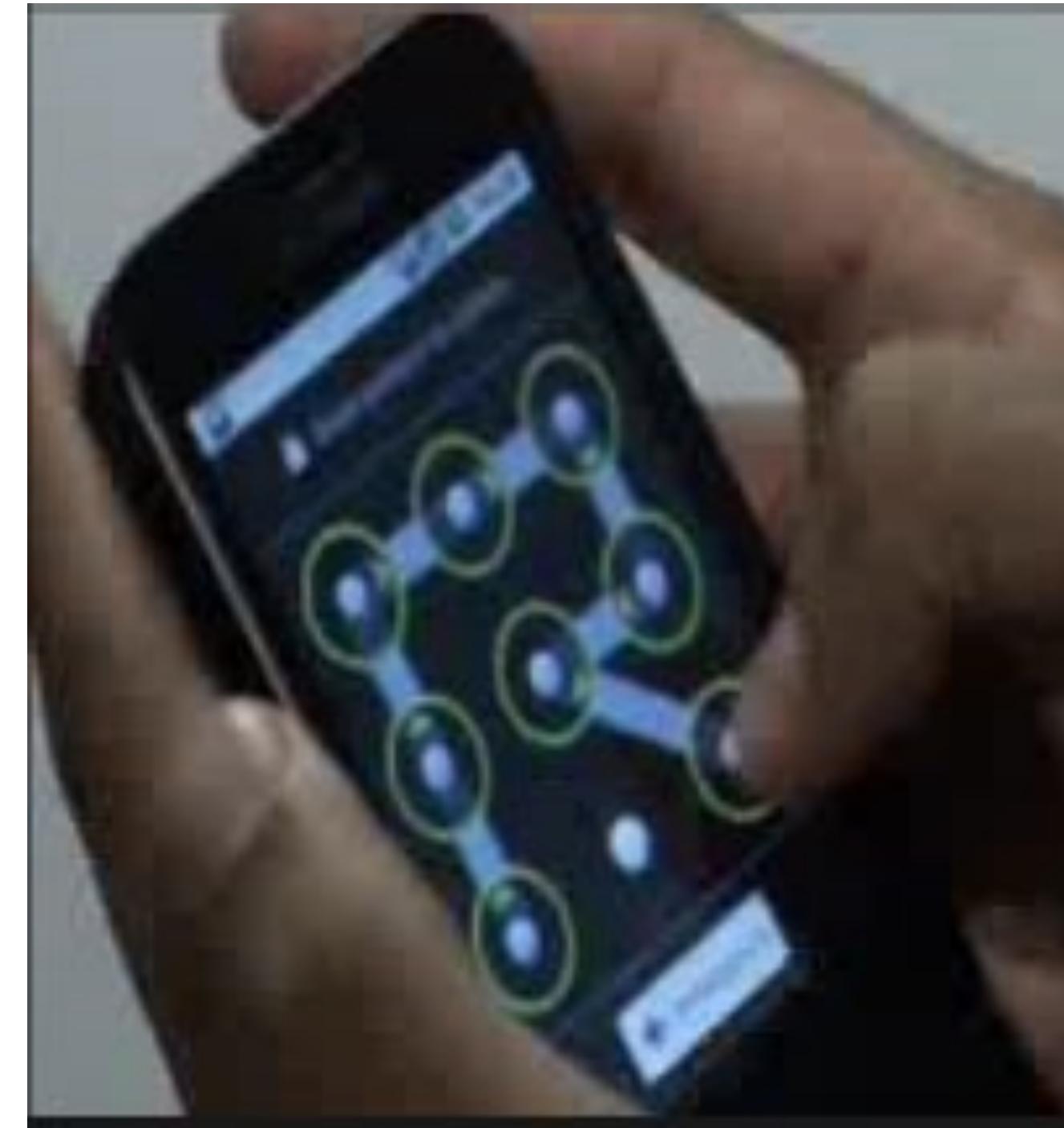
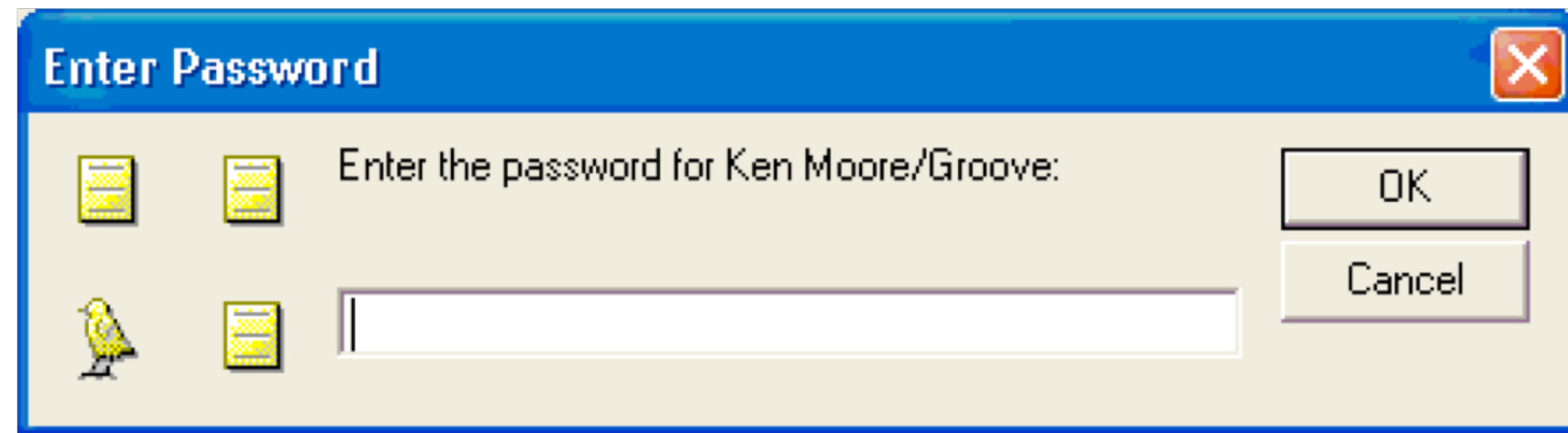
Fast, easy and secure enrollment solutions

Our trusted and long-lasting partnership with the U.S. government and decades of know-how in biometric security uniquely positions IDEMIA to provide a range of innovative enrollment solutions.

Learn how we can help make the enrollment process a breeze, whether online or in-person.



Authentication 1: Prove you know a secret (“something you know”).



Passwords can be guessed. “Strength” isn’t the point; good passwords are unique.

Rank	Password	Count of password uses
1	123456	21,627,656
2	admin	21,030,012
3	12345678	8,274,408
4	123456789	5,673,712
5	12345	3,950,777
6	password	3,545,119
7	Aa123456	2,520,728
8	1234567890	1,418,939
9	Pass@123	1,210,039
10	admin123	1,087,247
11	1234567	1,084,354
12	123123	1,060,563
13	111111	990,391
14	12345678910	988,396
15	P@ssw0rd	770,658
16	Password	755,709
17	Aa@123456	735,141
18	admintelecom	585,620
19	Admin@123	579,512
20	112233	576,908

Rank	2019 ^[15]
1	123456
2	123456789
3	qwerty
4	password
5	1111111
6	12345678
7	abc123
8	1234567
9	password1
10	12345
11	1234567890
12	123123
13	000000
14	lloveyou
15	1234
16	1q2w3e4r5t
17	Qwertyuiop
18	123
19	Monkey
20	Dragon

Rank ↕	Password ↕
1	123456
2	123456789
3	12345678
4	password
5	qwerty123
6	qwerty1
7	111111
8	12345
9	secret
10	123123
11	1234567890
12	1234567
13	000000
14	qwerty
15	abc123
16	password1
17	iloveyou
18	11111111
19	dragon
20	monkey

Passwords can be stolen.

META

Keeping Passwords Secure

March 21, 2019

LISTEN TO ARTICLE



Pedro Canahuati, VP Engineering, Security and Privacy

As part of a routine security review in January, we found that some user passwords were being stored in a readable format within our internal data storage systems. This caught our attention because our login systems are designed to mask passwords using techniques that make them unreadable. We have fixed these issues and as a precaution we will be notifying everyone whose passwords we have found were stored in this way.



YOUR DATA FOR ORGANISATIONS RESOURCES WHO WE ARE [NEWS AND MEDIA](#) DATA PROTECTION OFFICERS

News & Media

Contact details

Press releases

Latest news

Consultations

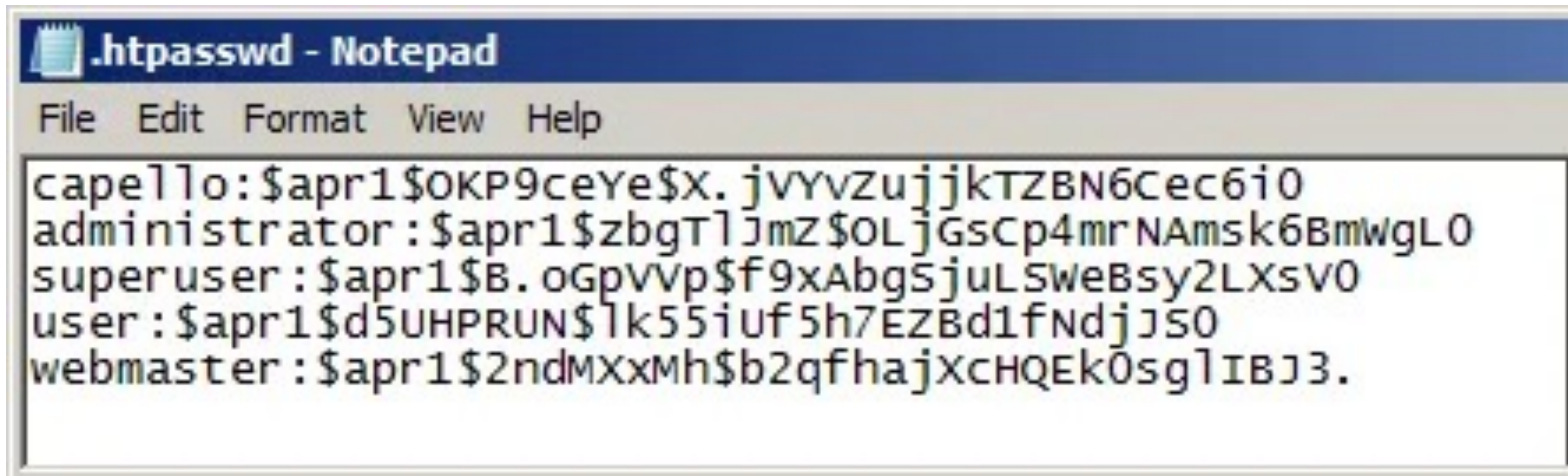
[Home](#) » [News & Media](#) » [Press releases](#)

Irish Data Protection Commission fines Meta Ireland €91 million

27th September 2024

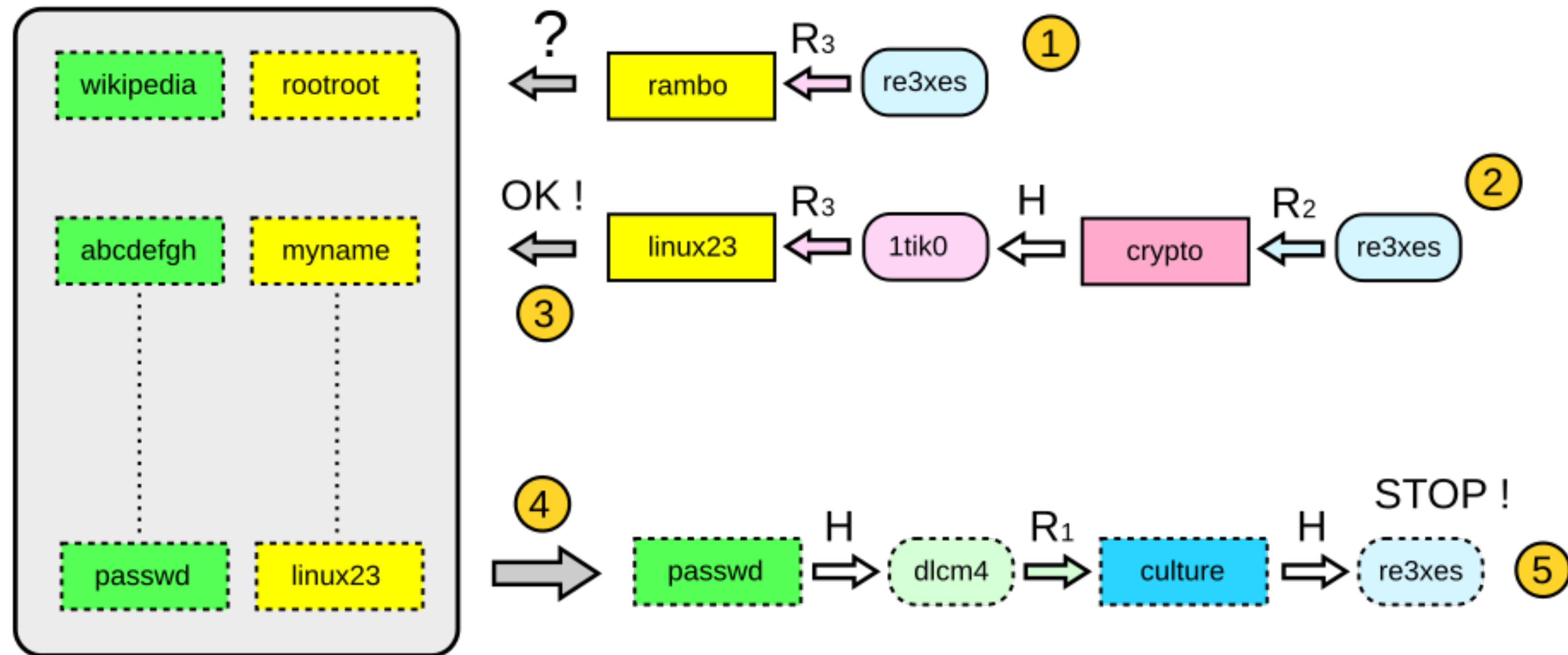
The Data Protection Commission (DPC) has today announced its final decision following an inquiry into Meta Platforms Ireland Limited (MPIL). This inquiry was launched in April 2019, after MPIL notified the DPC that it had inadvertently stored certain passwords of social media users in ‘plaintext’ on its internal systems (i.e. without cryptographic protection or encryption).

To protect password files, encrypt or hash each password.



```
.htpasswd - Notepad
File Edit Format View Help
capello:$apr1$OKP9ceYe$X.jvYvZujjkTZBN6Cec6i0
administrator:$apr1$zbgTlJmZ$OLjGScp4mrNAmsk6BmwgL0
superuser:$apr1$B.oGpVVp$f9xAbgSjuLSweBsy2LXsV0
user:$apr1$d5UHPRUN$lk55iuf5h7EZBd1fNdjJ50
webmaster:$apr1$2ndMXxMh$b2qfhajXcHQEk0sglIBJ3.
```


But attackers can use “rainbow tables” to attack hashed passwords.



https://en.wikipedia.org/wiki/Rainbow_table

“Salting” passwords can raise the cost of rainbow tables.

Hashing and Salting Alice's Password

User: Alice

Password: `farm1990M00`

Salt: `f1nd1ngn3m0`

Salted input: `farm1990M00f1nd1ngn3m0`

Hash (SHA-256): `07dbb6e6832da0841dd79701200e4b179f1a94a7b3dd26f612817f3c03117434`

Hashing and Salting Bob's Password

User: Bob

Password: `farm1990M00`

Salt: `f1nd1ngd0ry`

Salted input: `farm1990M00f1nd1ngd0ry`

Hash (SHA-256): `11c150eb6c1b776f390be60a0a5933a2a2f8c0a0ce766ed92fea5bfd9313c8f6`

Can you defend against attacks on password-based authentication?

NIST Special Publication 800-63B

Digital Identity Guidelines

Authentication and Lifecycle Management

June 2017

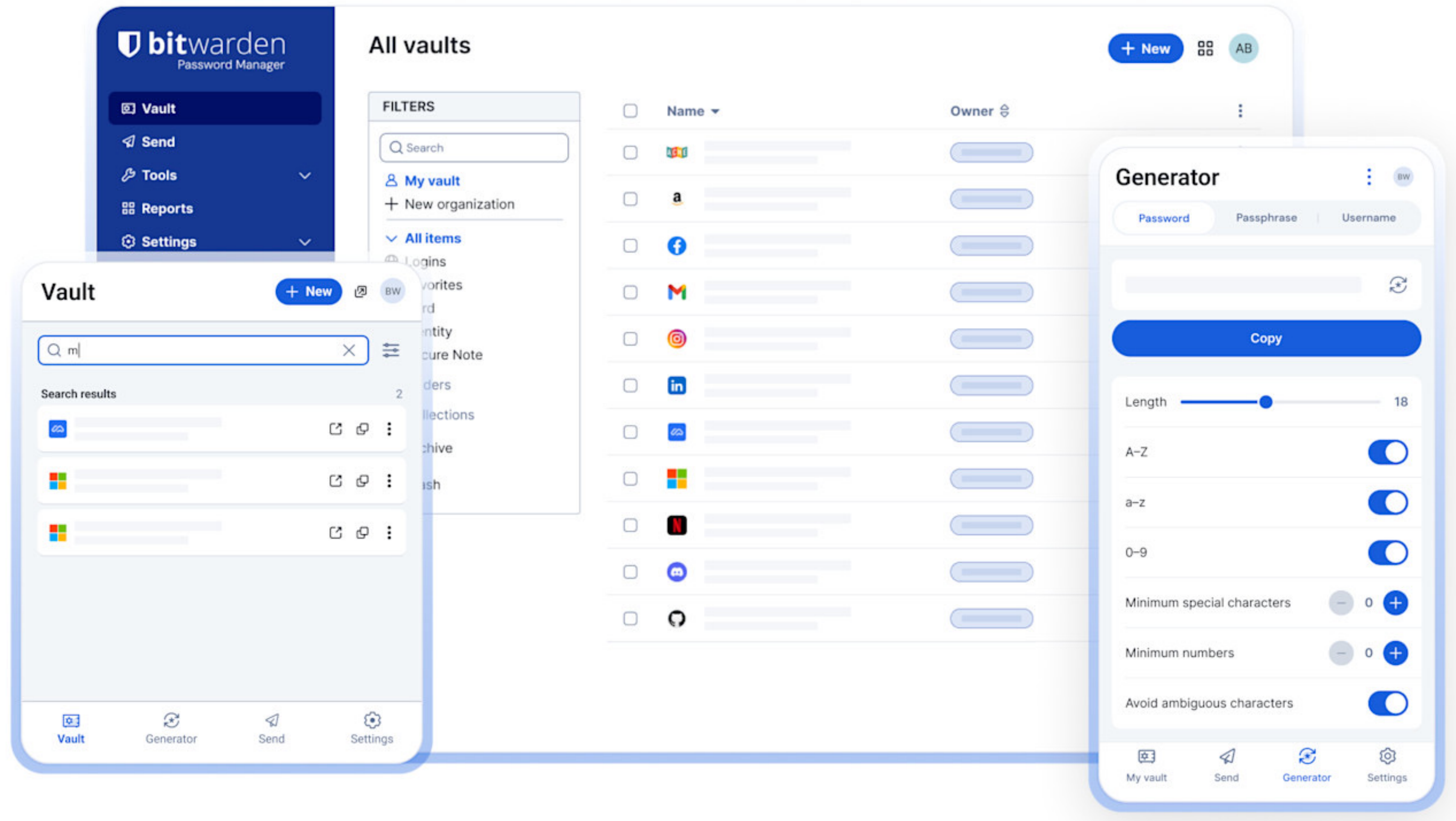
<https://doi.org/10.6028/NIST.SP.800-63b>

**NIST SP 800-63 Digital Identity
Guidelines**

July 2025

<https://pages.nist.gov/800-63-4/>

Password managers create and remember unique (strong) passwords for you.



Password managers may not be as perfectly secure as they claim.



MATT BURGESS

ANDY GREENBERG

LILY HAY NEWMAN

SECURITY

FEB 21, 2026 6:38 AM

Security News This Week: Password Managers Share a Hidden Weakness

Zero Knowledge (About) Encryption: A Comparative Security Analysis of Three Cloud-based Password Managers

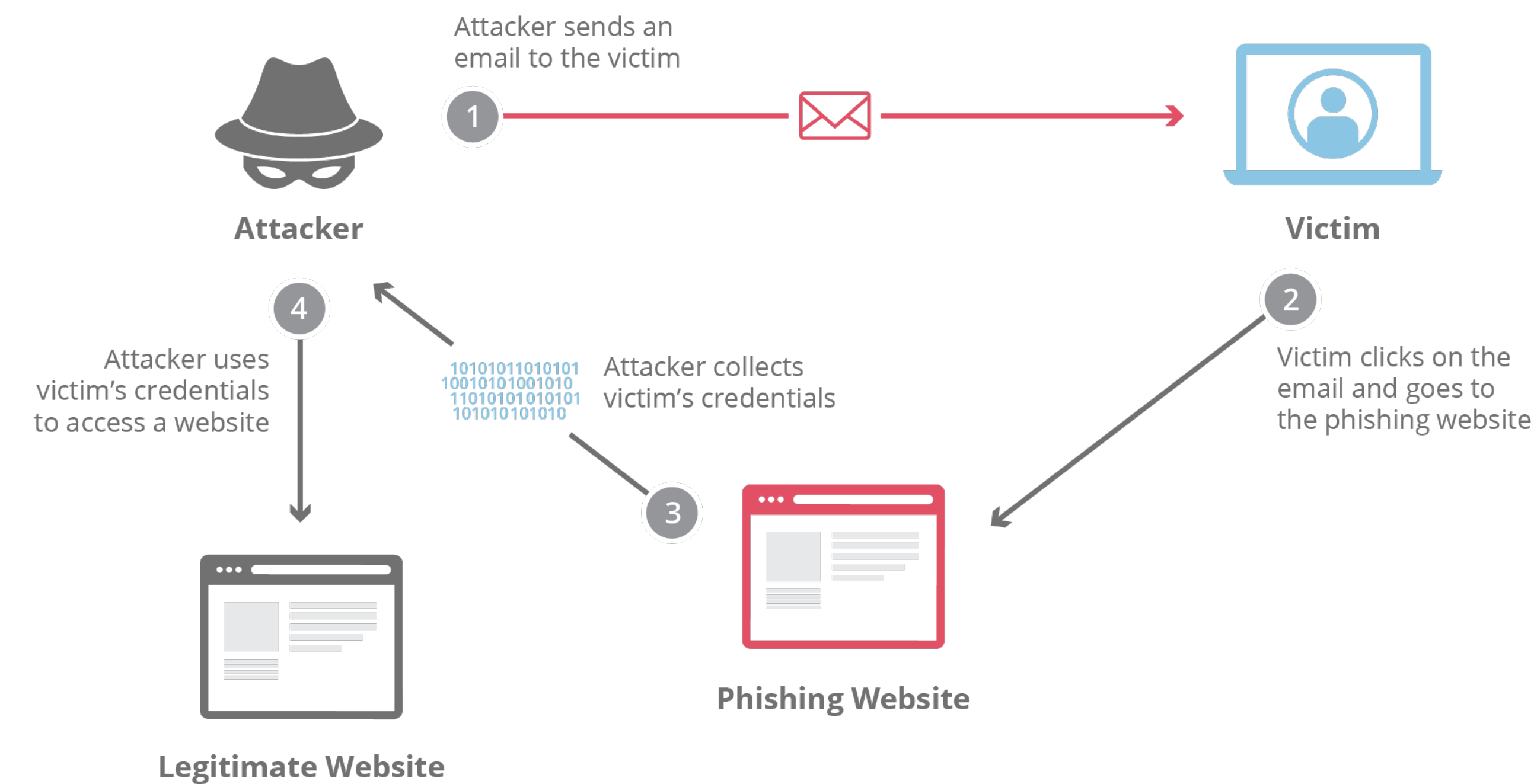
Matteo Scarlata ¹, Giovanni Torrisi ^{2†}, Matilda Backendal ^{2†}, and Kenneth G. Paterson ¹

¹ *Department of Computer Science, ETH Zurich, Switzerland*

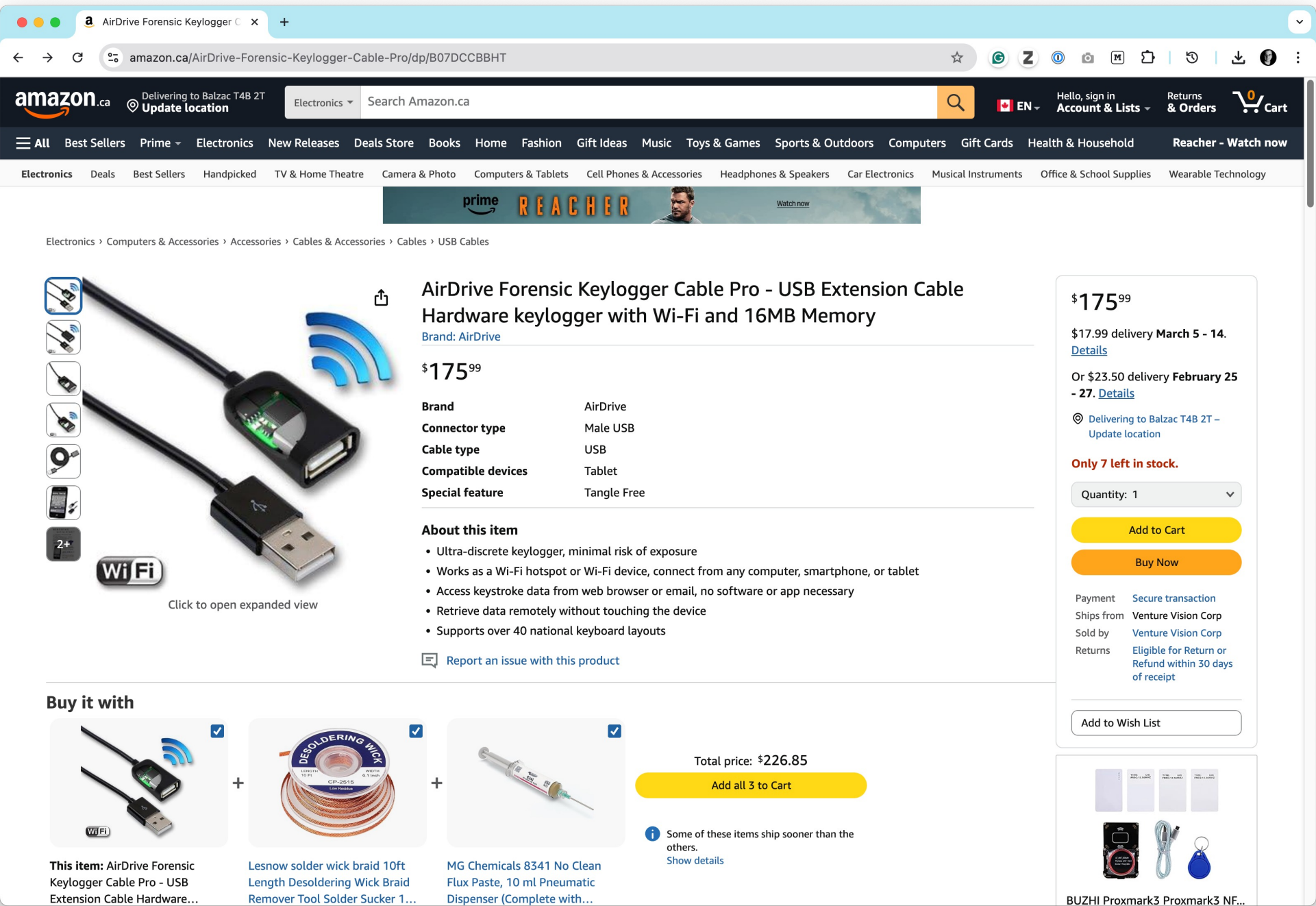
² *Faculty of Informatics, Università della Svizzera italiana (USI), Switzerland*

Class	Ref	Name	Cause	Impact	Interaction	Mitigations
Key Escrow	BW01	Malicious Auto-Enrolment	Lack of Key Auth, Key Substitution	Full vault compromise †	1 join	PKA, SC
	BW02	Malicious Key Rotation	Key Substitution	Full vault compromise †	1 rotation	SC
	BW03	Malicious KC Conversion	Lack of Key Auth, Key Substitution	Full vault compromise †	1 dialog	PKA, SC
Item-Level Vault Encryption	LP01	Malicious Password Reset	Lack of Key Auth	Full vault compromise †	1 login	PKA, SC
	BW04	Unprotected Item Metadata	Lack of Auth Enc	Read and modify metadata	-	AE, AD
	BW05	Item Field Swapping	Lack of Key Sep	Field and item swapping	-	KS
	BW06	Icon URL Item Decryption	Lack of Key Sep	Loss of confidentiality †	1 open	KS
	BW07	Remove KDF Iterations	Lack of Auth Enc	No brute-force protection	1 login	AD
	LP02	Item Field Swapping	Lack of Auth Enc	Field and item swapping	-	KS, AE
	LP03	Icon URL Item Decryption	Lack of Key Sep	Loss of confidentiality †	1 open	KS, AE
	LP04	Remove KDF Iterations	Lack of Auth Enc	No brute-force protection	1 login	AD
	LP05	Malleable Vault	Lack of Auth Enc	Loss of vault integrity	-	AE
	LP06	Unprotected Item Metadata	Lack of Auth Enc	Read and modify metadata	-	AD, AE
Sharing	DL01	Transaction Replay	Lack of Key Sep	Loss of vault integrity	-	KS
	BW08	Organisation Injection	Lack of Key Auth	Add users to arbitrary orgs	1 sync	SC
	BW09	Organisation Overwrite	Lack of Key Auth, Key Substitution	Organisation compromise †	1 join	PKA, SC
Backwards Compatibility	LP07	Sharing Key Overwrite	Lack of Key Auth	Shared vault compromise †	1 join	PKA
	DL02	Sharing Key Overwrite	Lack of Key Auth	Shared vault compromise †	1 join	PKA
	BW10	Disable Per-Item Keys	Lack of Auth Enc	Downgrade key hierarchy	-	KS, AD
	BW11	User Key Overwriting	CBC Support	Loss of confidentiality †	2 logins	AE
	BW12	Downgrade to Legacy	CBC Support	Full vault compromise †	2 logins	AE
	DL03	Item Injection	CBC Support	Loss of vault integrity	10 ⁴ syncs	AE
	DL04	Remove KDF Iterations	CBC Support	No brute-force protection	10 ⁴ syncs	AE
	DL05	CBC-Only Downgrade	CBC Support	Loss of confidentiality †	10 ⁵ syncs	AE
	DL06	Lucky 64	CBC Support	No brute-force protection	10 ⁴ syncs	AE

It can be easier to attack passwords at the point that the user enters it.

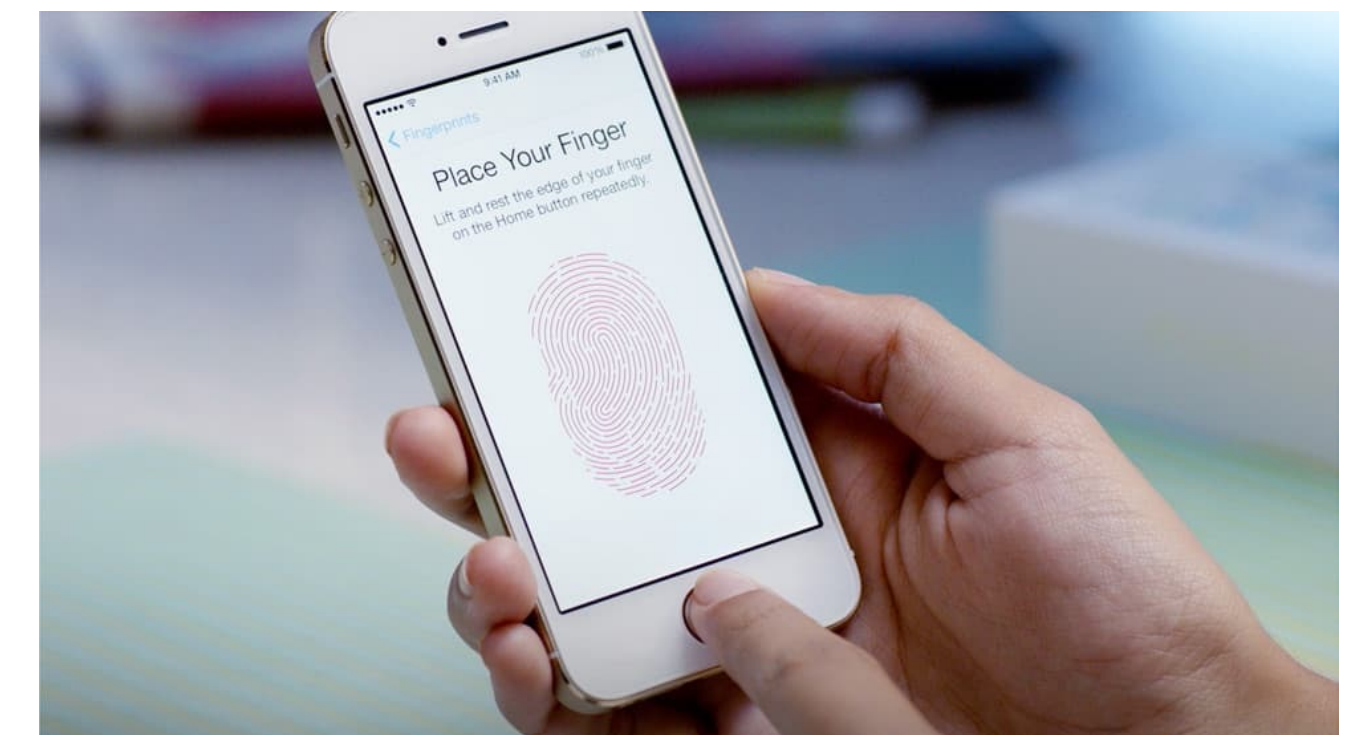


Phishing



Keylogging

Authentication 2: Something you *have* or something you *are*.



Something you ... forgot, lost, or used to be.

Sign in

[I don't have an account](#)

Email

Password

Your password is incorrect or this account doesn't exist. Please [reset your password](#) or check [this article](#).

Sign in

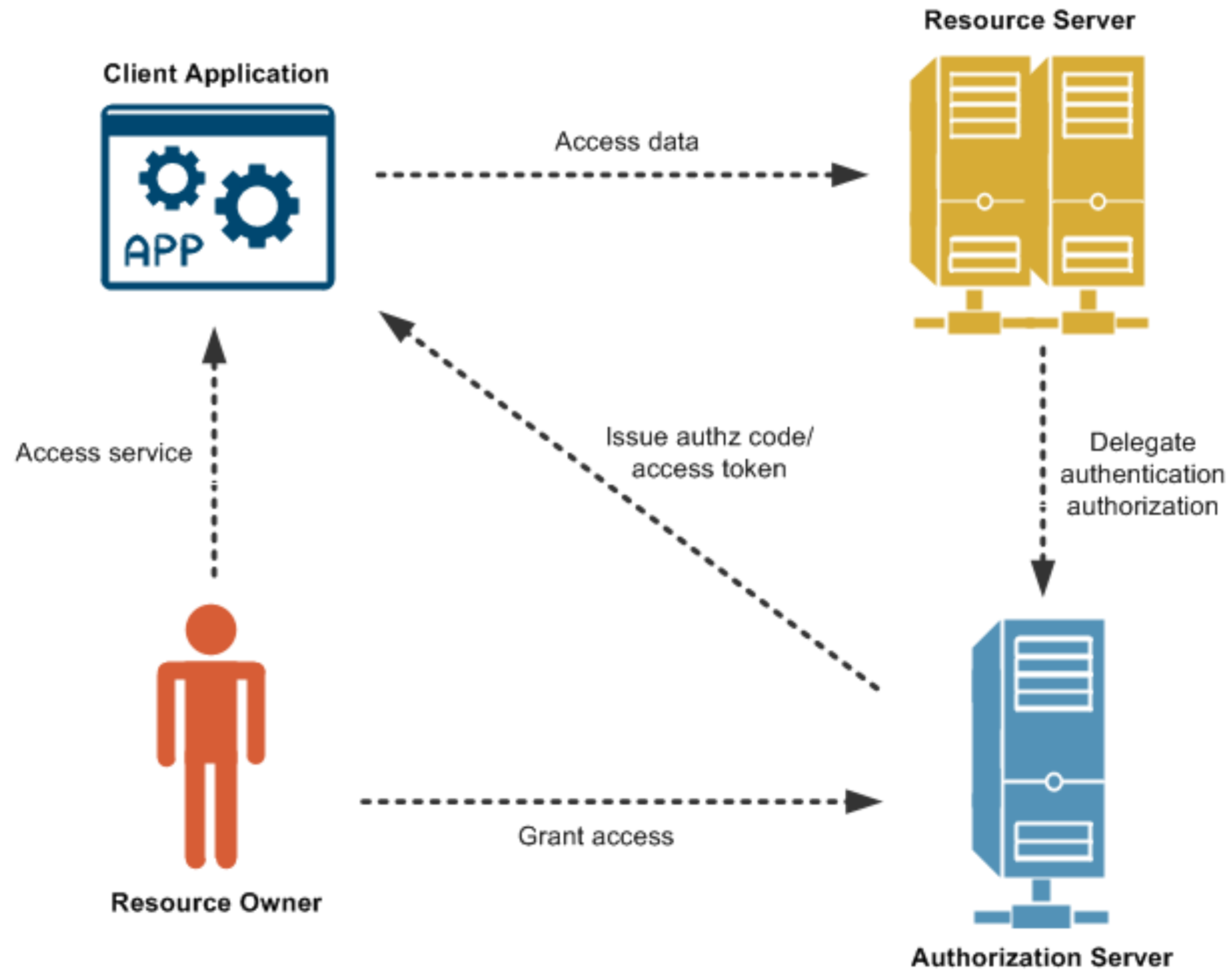
[Can't sign in?](#)



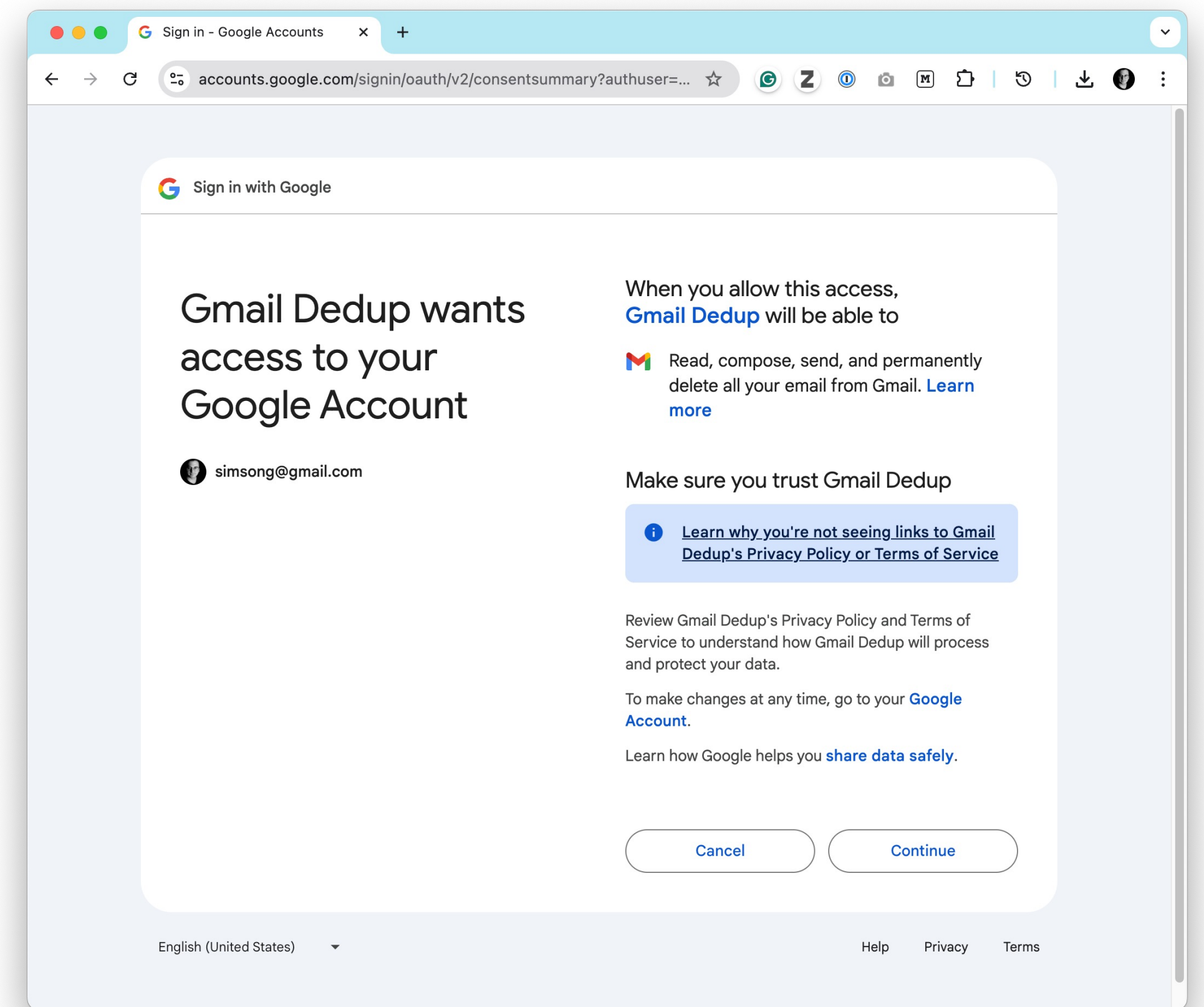
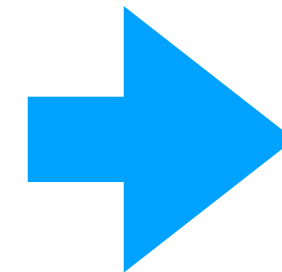
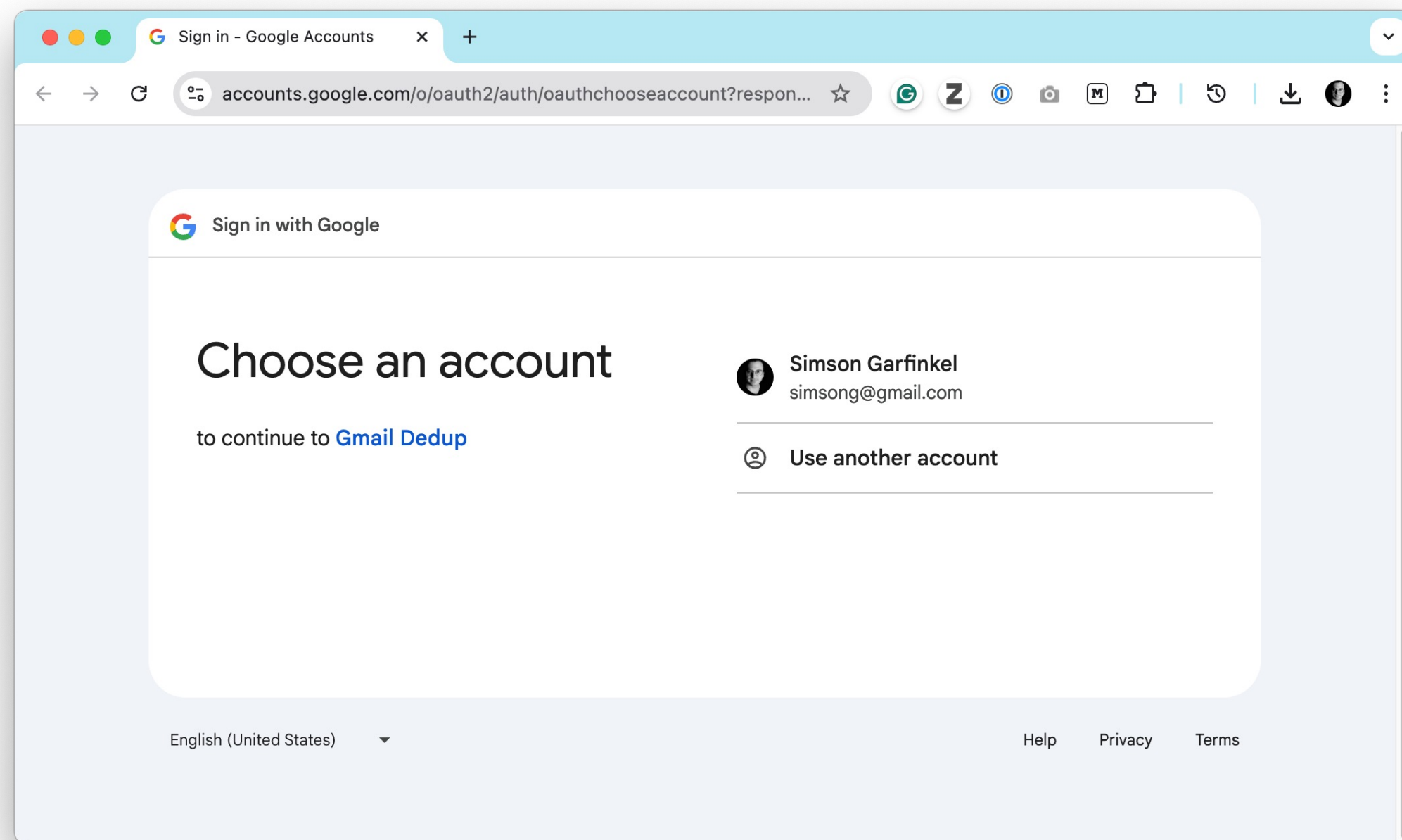
Cryptographic Approaches to Authentication



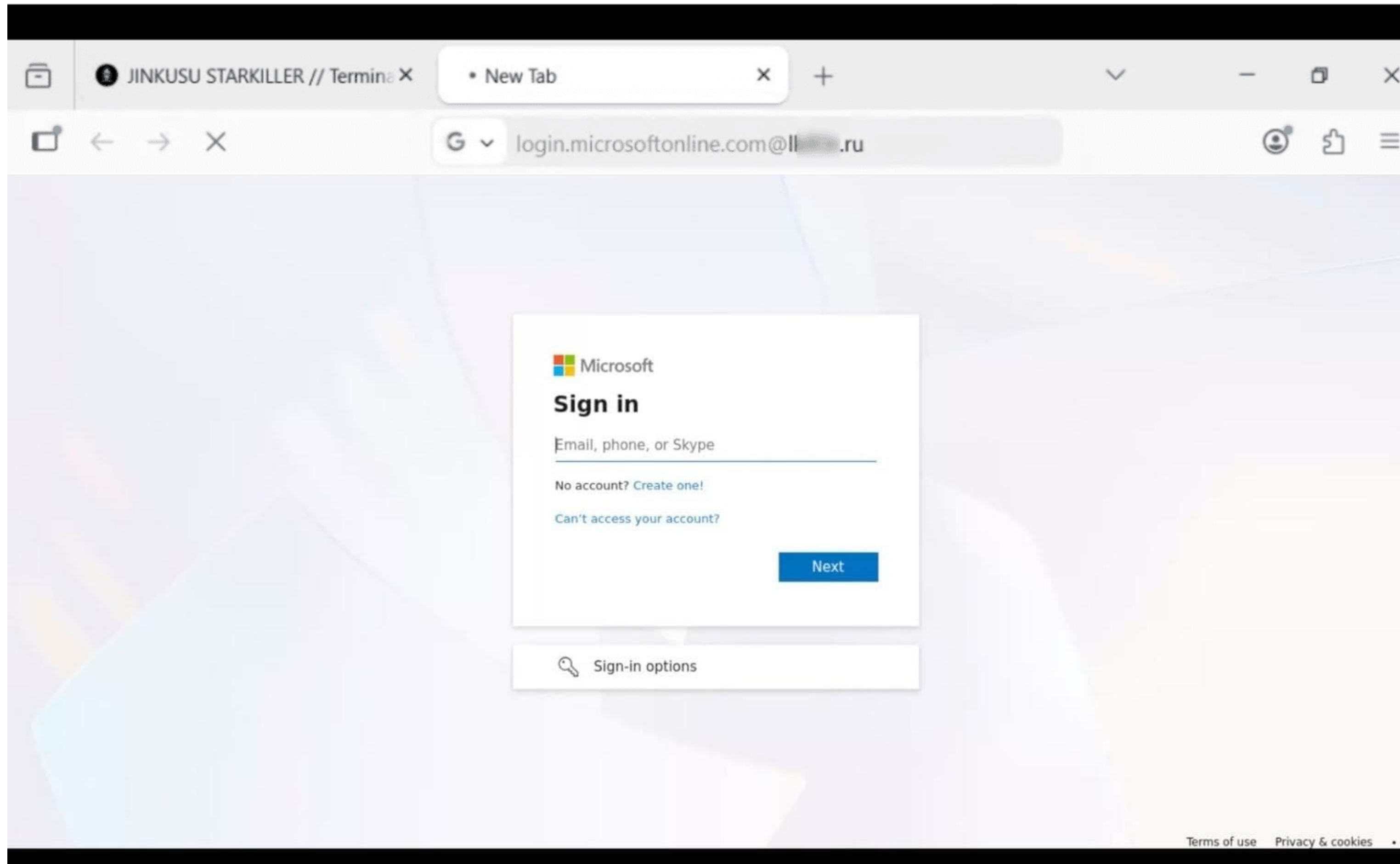
OAuth proxy authentication



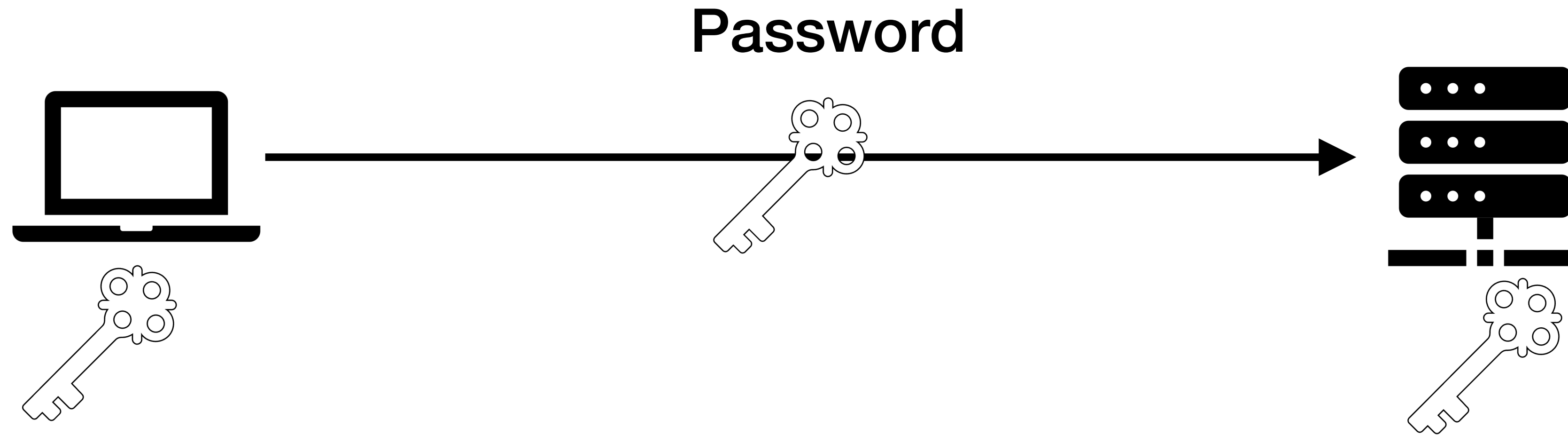
OAuth Example — Gmail Deduplicate



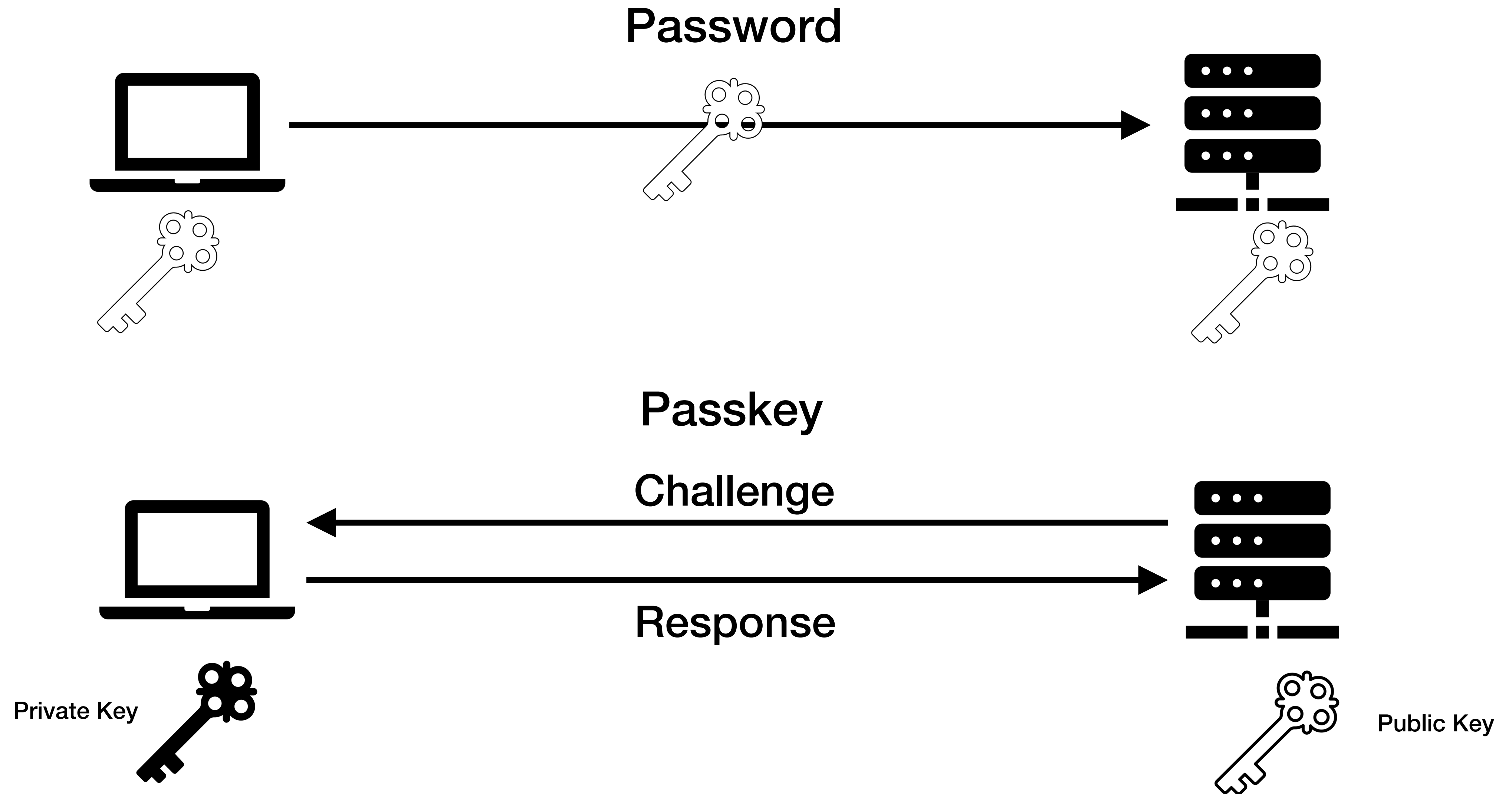
Attacker-in-the-middle and permission attacks on OAuth



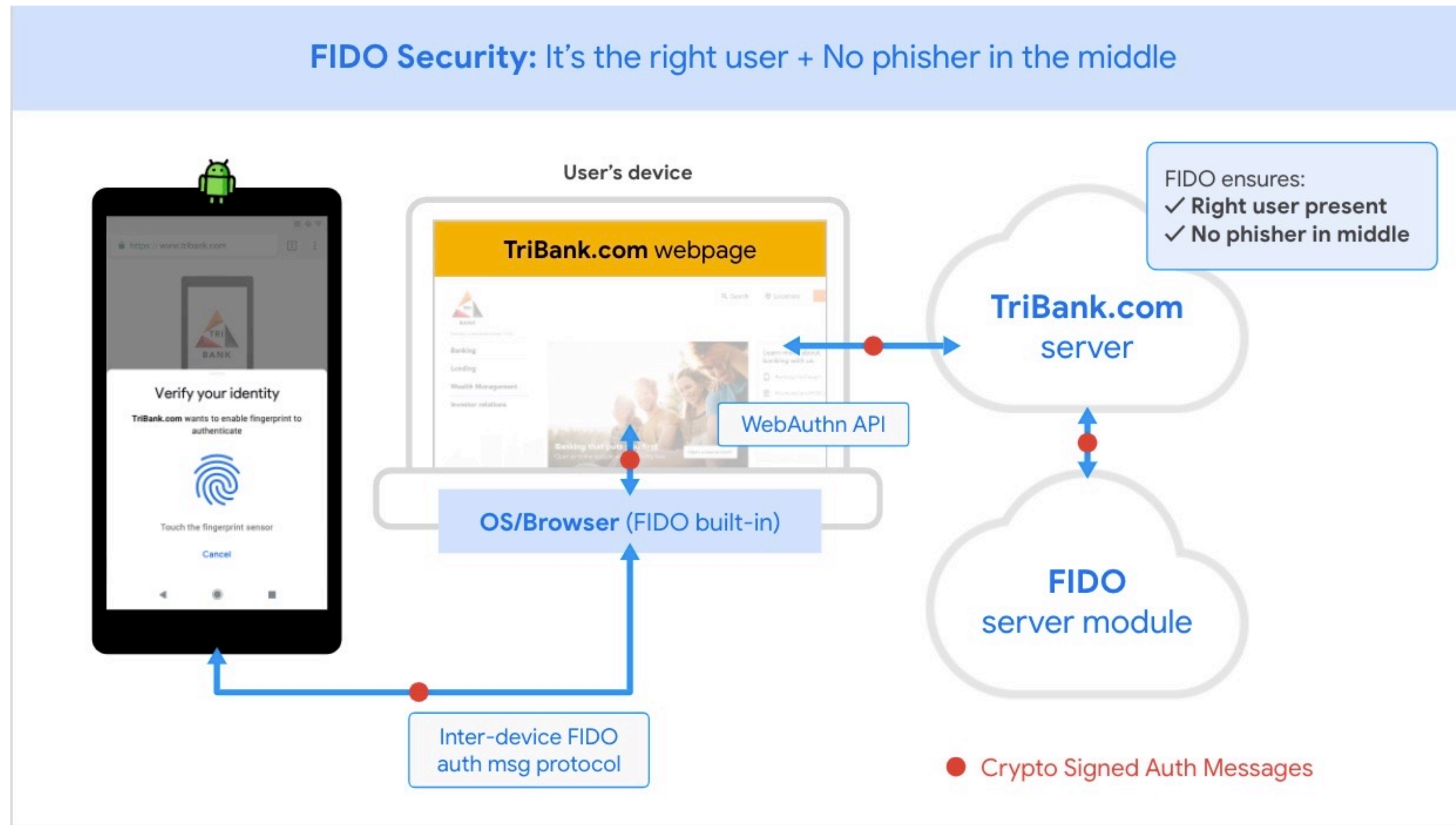
Passwords require storing, transmitting, and sharing a secret.



Passkeys require only creating and storing a secret locally.



Passkeys can limit phishing attacks.



Governments are beginning to issue electronic representations of identities.



**Digital. Secure.
Convenient.**

Your mobile driver's license
on your phone

CA DMV WALLET

mDL

Status: Active

mDL Info

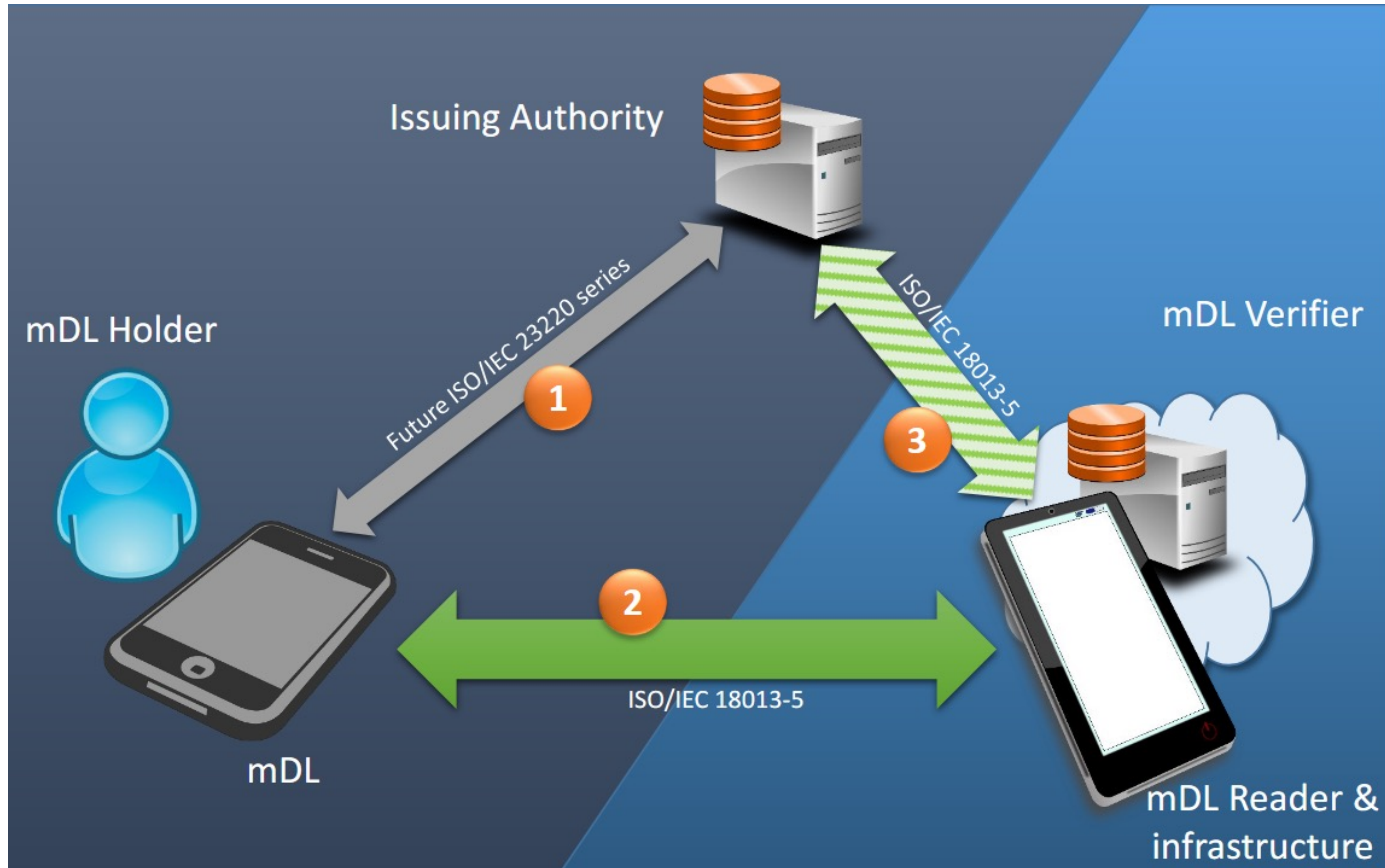
Issued by	Date Source
California DMV	DMV
Customer Identifier	DOB
1900XXXX	1984-11-19
ISS	
Feb 21, 2023	
EXP	
Feb 19, 2027	
Name	
Stan Ishudler	

Swipe left for QR Code

DMV Wallet

mDL

Government identity architecture: individual, enrollment, claim, verifier.



China approves Cyberspace IDs as part of federated identity scheme

🕒 May 27, 2025, 3:52 pm EDT | [Lu-Hai Liang](#)

As part of a federated identity scheme, the ID can be used across multiple government and private online services. Authorities say the single credential reduces the amount of personal information people need to share online and claim that this should improve cybersecurity and strengthen online privacy.

<https://www.biometricupdate.com/202505/china-approves-cyberspace-ids-as-part-of-federated-identity-scheme>

Human rights groups worry about implications.

A statement from Chinese Human Rights Defenders ([CHRD](#)) and [Article 19](#) calls this into question, claiming the system will “further constrict anonymity” in an already heavily policed digital environment. It takes particular issue with the policy requiring users to register through the National Online Identity Authentication App, developed by the MPS, using their national identification card and facial recognition.

There are privacy concerns about mDLs and related systems.



About ▾ Initiatives Insights & Research ▾ Contact 🔍

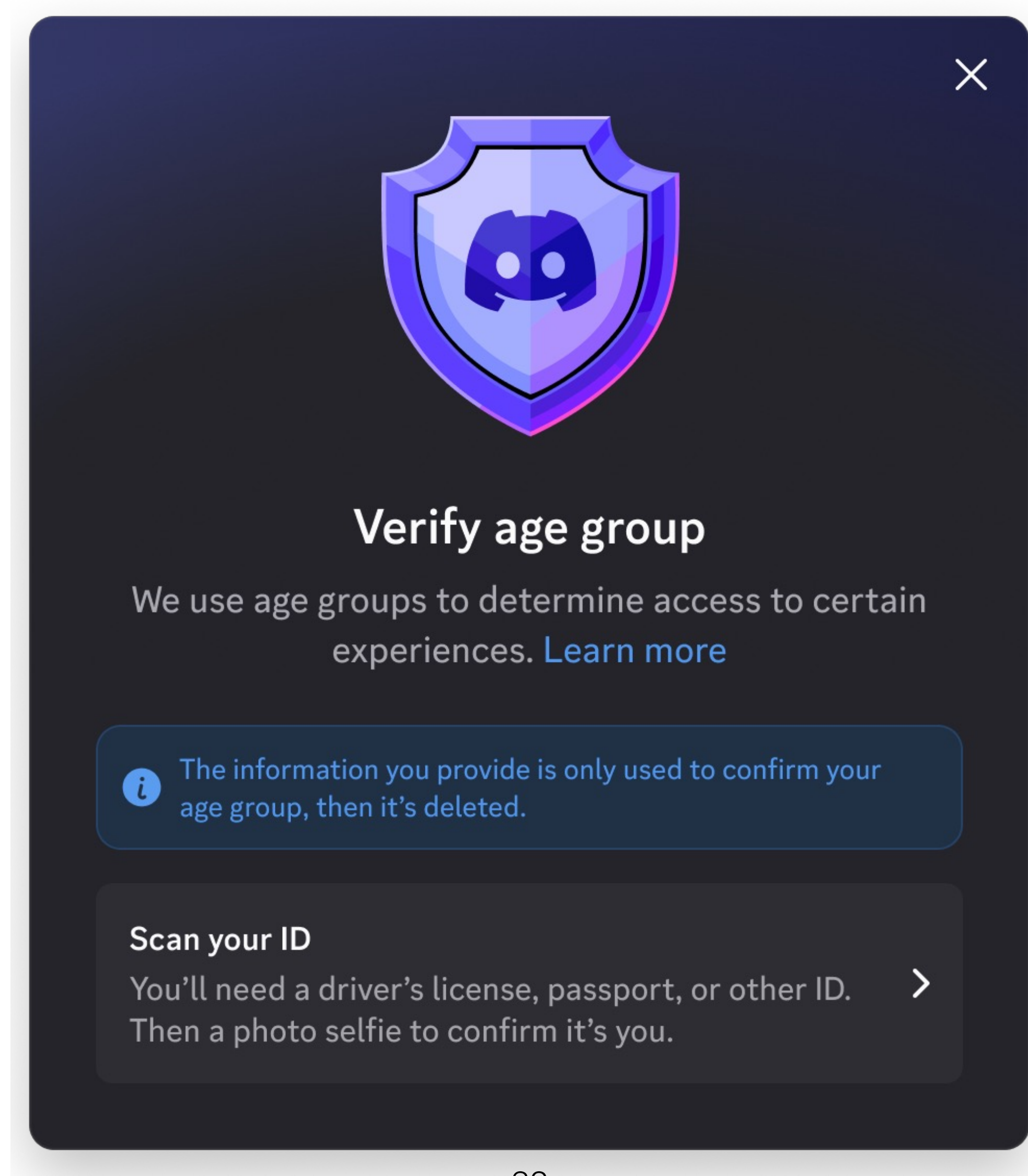
BLOG POST • IDENTITY

A New Approach to Address Concerns About Overuse of Digital IDs

October 30, 2025 | Jeremy Grant & Zack Martin

<https://www.centerforcybersecuritypolicy.org/insights-and-research/a-new-approach-to-address-concerns-about-overuse-of-digital-ids>

Authentication against government identities



You'll never guess what happened next!

Cyber Security

Cyber Security News

Data Breach

Vulnerability News

Discord Data Breach – 1.5 TB of Data and 2 Million Government ID Photos Extorted

By [Guru Baran](#) - October 9, 2025

Persona functionality and API documentation

Age verification vendor Persona left frontend exposed, researchers say

by Pieter Arntz | February 20, 2026

```
personal identity:
  - full legal name (including native script)
  - date of birth, place of birth
  - nationality, sex, height

address:
  - street, city, state, postal code, country

government document:
  - document type and number
  - issuing authority
  - issue and expiration dates
  - visa status
  - vehicle class/endorsements/restrictions

media:
  - FRONT PHOTO of ID document (URL)
  - BACK PHOTO of ID document (URL)
  - SELFIE PHOTO (URL + byte size)
  - VIDEO of identity capture (URL)

metadata:
  - entity confidence score
  - all verification check results with pass/fail reasons
  - capture method used
  - timestamps (created, submitted, completed, redacted)
```


An open letter against age assurance technology

Joint statement of security and privacy scientists and researchers on Age Assurance Executive summary

We are writing in response to the worldwide initiatives to introduce age assurance technologies to implement access control to internet services. We share the concerns about the negative effects that exposure to harmful content online has on children, and we applaud that regulators dedicate time and effort to protect them. However, we fear that, if implemented without careful consideration of the technological hazards and societal impact, the new regulation might cause more harm than good.

<https://csa-scientist-open-letter.org/ageverif-Feb2026>

Even anonymity/pseudonymity may not be enough.

arsTECHNICA

SECTIONS ▾FORUM | SUBSCRIBE | 🔍 | SIGN IN

👤 RIP PSEUDONYMITY

LLMs can unmask pseudonymous users at scale with surprising accuracy

Pseudonymity has never been perfect for preserving privacy. Soon it may be pointless.

DAN GOODIN – MAR 3, 2026 7:30 AM | 219



➦ Credit: Getty Images

Extract

🔍 @anon_user42

Struggling with CS224N at Stanford this quarter. Senior year is brutal 😞

🔍 @anon_user42

Walking Biscuit through Dolores Park missing Portland rain today 🌧️

▼ LLM

Profile: @anon_user42

Stanford CS (NLP), ~22 y/o, from Portland

Lives in SF, dog named "Biscuit"

Search

Profile: J. Chen

Stanford CS '24, age 22, lives in SF

From Portland. Dog "Biscuit"

ML Eng @ Nexus AI (SF startup)

▼ Search

🔍 @anon_user42

Stanford CS, SF, "Biscuit"

0.94

🔍 @cs_grad_pdx

Stanford CS, Portland

0.71

----- top-k -----

🔍 @ml_anon99

MIT '24, SWE

0.42

Reason

J. Chen

Stanford CS '24

SF, from Portland

Dog "Biscuit", ...

▼ LLM

Stanford CS '24 ✓ SF/Portland ✓ Age 22 ✓

Both own a dog named "Biscuit"

All 5 attributes align. No contradictions.

✓ @anon_user42 = J. Chen

Confidence: 94%

Large-scale online deanonymization with LLMs

Simon Lermen*
MATS

Daniel Paleka*
ETH Zurich

Joshua Swanson
ETH Zurich

Michael Aerni
ETH Zurich

Nicholas Carlini
Anthropic

Florian Tramèr
ETH Zurich

Abstract

We show that large language models can be used to perform at-scale deanonymization. With full Internet access, our agent can re-identify Hacker News users and Anthropic Interviewer participants at high precision, given pseudonymous online profiles and conversations alone, matching what would take hours for a dedicated human investigator. We then design attacks for the closed-world setting. Given two databases of pseudonymous individuals, each containing unstructured text written by or about that individual, we implement a scalable attack pipeline that uses LLMs to: (1) extract identity-relevant features, (2) search for candidate matches via semantic embeddings, and (3) reason over top candidates to verify matches and reduce false positives. Compared to classical deanonymization work (e.g., on the Netflix prize) that required structured data, our approach works directly on raw user content across arbitrary platforms. We construct three datasets with known ground-truth data to evaluate our attacks. The first links Hacker News to LinkedIn profiles, using cross-platform references that appear in the profiles. Our second dataset matches users across Reddit movie discussion communities; and the third splits a single user’s Reddit history in time to create two pseudonymous profiles to be matched. In each setting, LLM-based methods substantially outperform classical baselines, achieving up to 68% recall at 90% precision compared to near 0% for the best non-LLM method. Our results show that the practical obscurity protecting pseudonymous users online no longer holds and that threat models for online privacy need to be reconsidered.

1 Introduction

For decades, it has been known that individuals can be uniquely identified from surprisingly few attributes. Sweeney’s seminal work demonstrated that 87% of the U.S. population could be uniquely identified by just zip code, birth date, and gender [34]. Narayanan and Shmatikov showed that

anonymous Netflix ratings could be linked to public IMDb profiles using only a handful of movie preferences [24], while De Montjoye et al. [6] proved that four spatiotemporal points are enough to uniquely identify 95% of individuals in mobile phone datasets. Despite these attacks, pseudonymous online accounts (Reddit throwaways, anonymous forums, review profiles, etc) have remained largely unaffected by deanonymization attempts. The reason is simple: applying such attacks in practice has required structured data amenable to algorithmic matching or substantial manual effort by skilled investigators reserved for high-value targets [13].

Deanonymization is a two-step process at heart, involving profiling an anonymous person from their posts, and then matching them to a known identity. It’s well-known that large language models (LLMs) can infer personal attributes from text on online forums [8, 29, 38]. Given this, it makes sense to ask: how good are LLMs at full end-to-end deanonymization, and is this a practical threat to pseudonymous accounts?

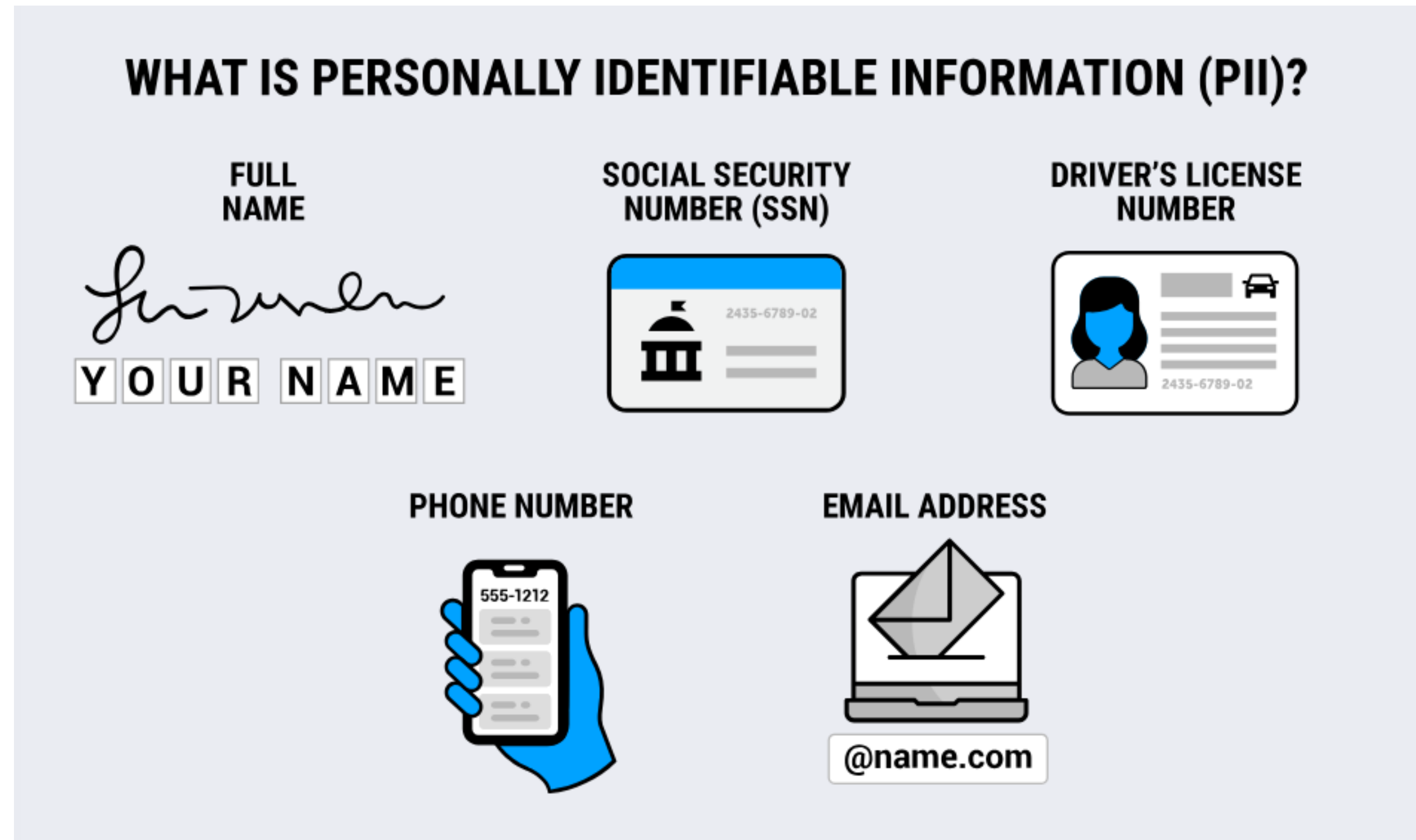
Our contributions. We demonstrate that LLMs fundamentally change the picture, enabling fully automated deanonymization attacks that operate on unstructured text at scale. We show this by phrasing deanonymization as a matching problem and showing LLMs can perform all steps needed to match accounts: extract identity-relevant signals from arbitrary text, efficiently search over millions of candidate profiles, and reason about whether two accounts belong to the same person. We show that the practical obscurity that has long protected pseudonymous users (the assumption that deanonymization, while theoretically possible, is too costly to execute broadly) no longer holds.

We validate this thesis in three deanonymization settings: (1) matching an online account to its real identity; (2) linking an identity to an unknown pseudonymous account; and (3) linking pseudonymous accounts of the same person across different platforms or time periods. These settings capture distinct threat models (e.g., doxxing of an online account, a stalker targeting a victim, or an adversary consolidating a user’s activity) and pose different technical challenges.

*Equal contribution.

arXiv:2602.16800v2 [cs.CR] 25 Feb 2026

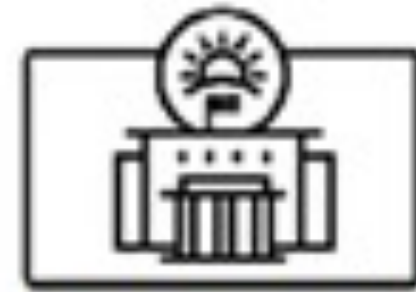
Privacy and Personal Information



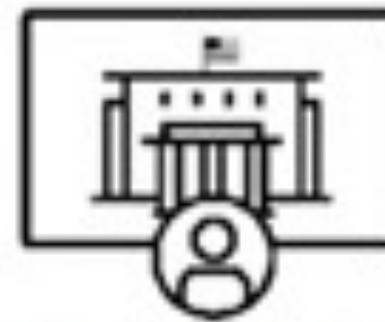
*These are illustrative examples only.

The Privacy Act of 1974 and the FIPPs

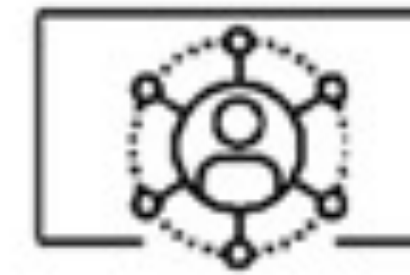
FAIR INFORMATION PRACTICE PRINCIPLES



TRANSPARENCY



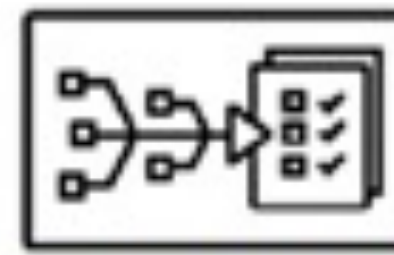
**INDIVIDUAL
PARTICIPATION**



AUTHORITY



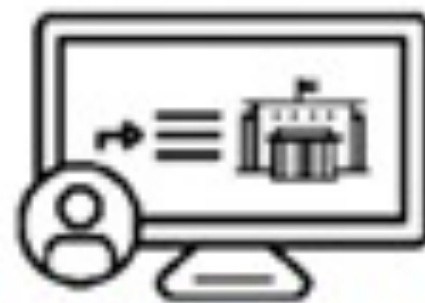
**PURPOSE SPECIFICATION
AND USE LIMITATION**



MINIMIZATION



**QUALITY AND
INTEGRITY**



**ACCESS AND
AMENDMENT**

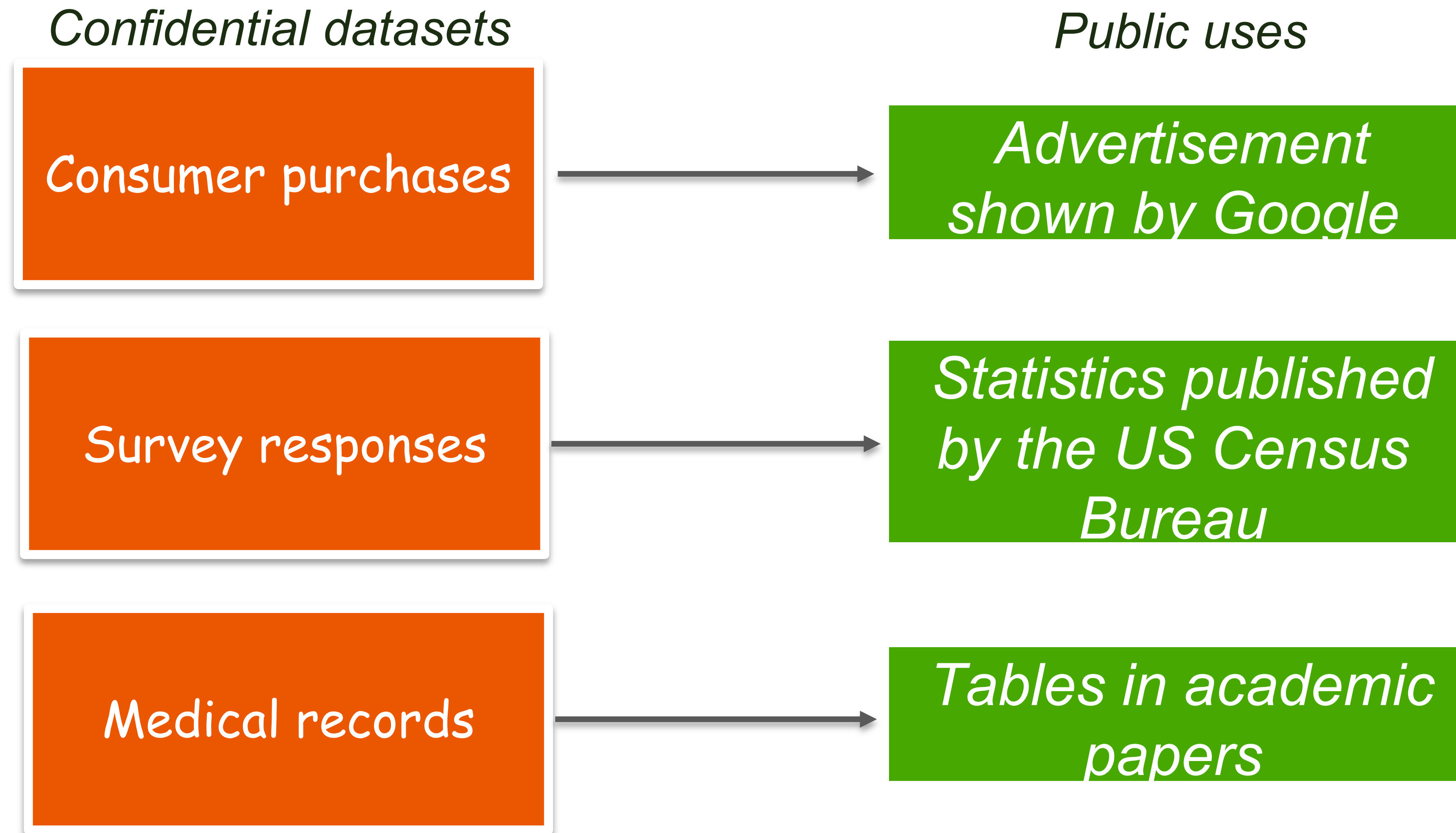


SECURITY

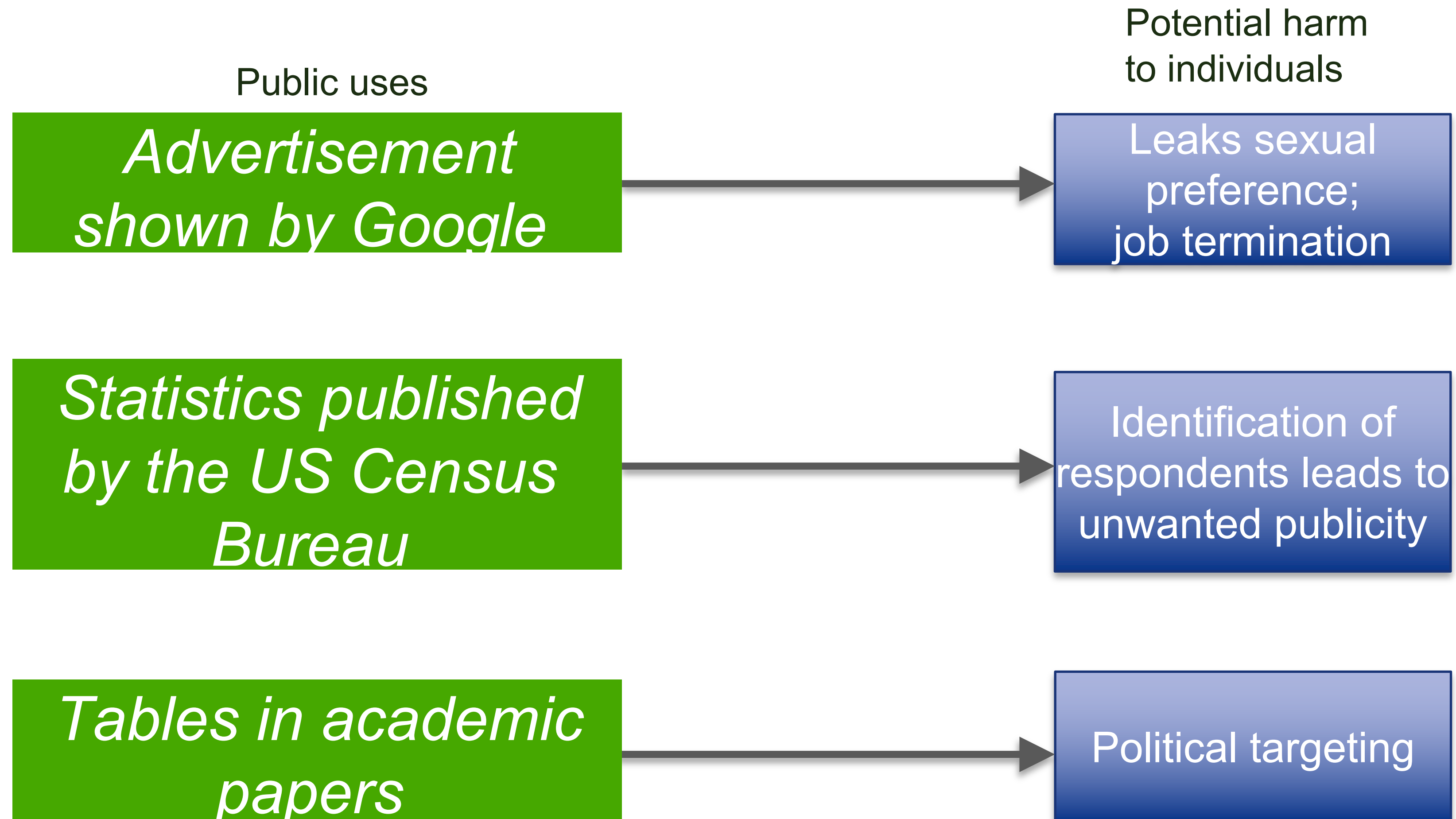


ACCOUNTABILITY

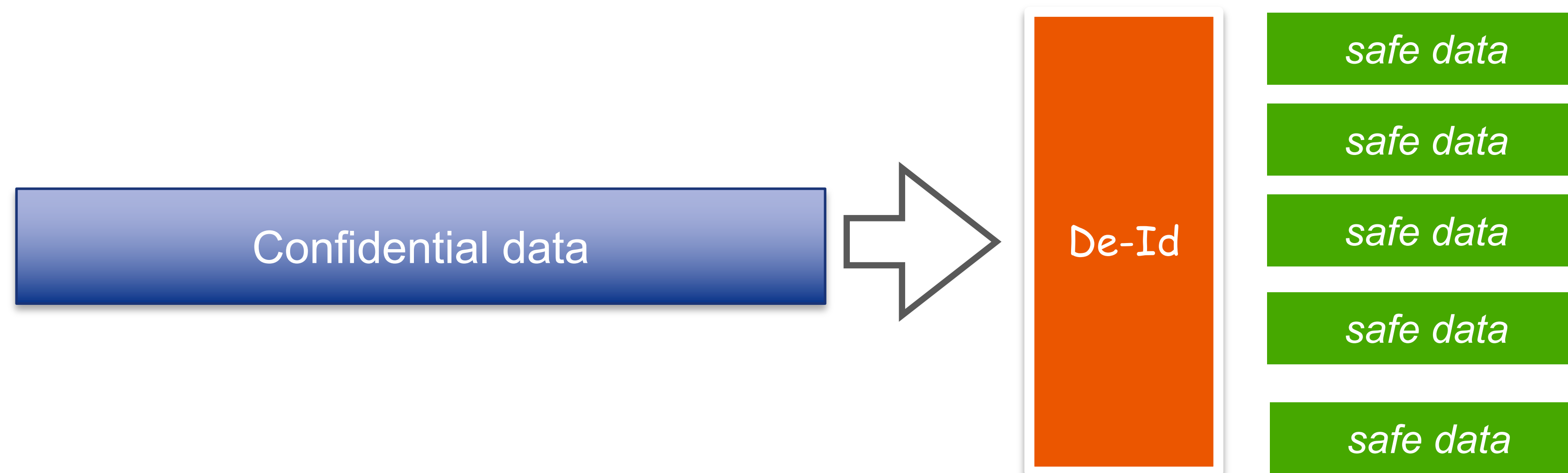
Many governments and businesses make *public use* of *confidential data*.



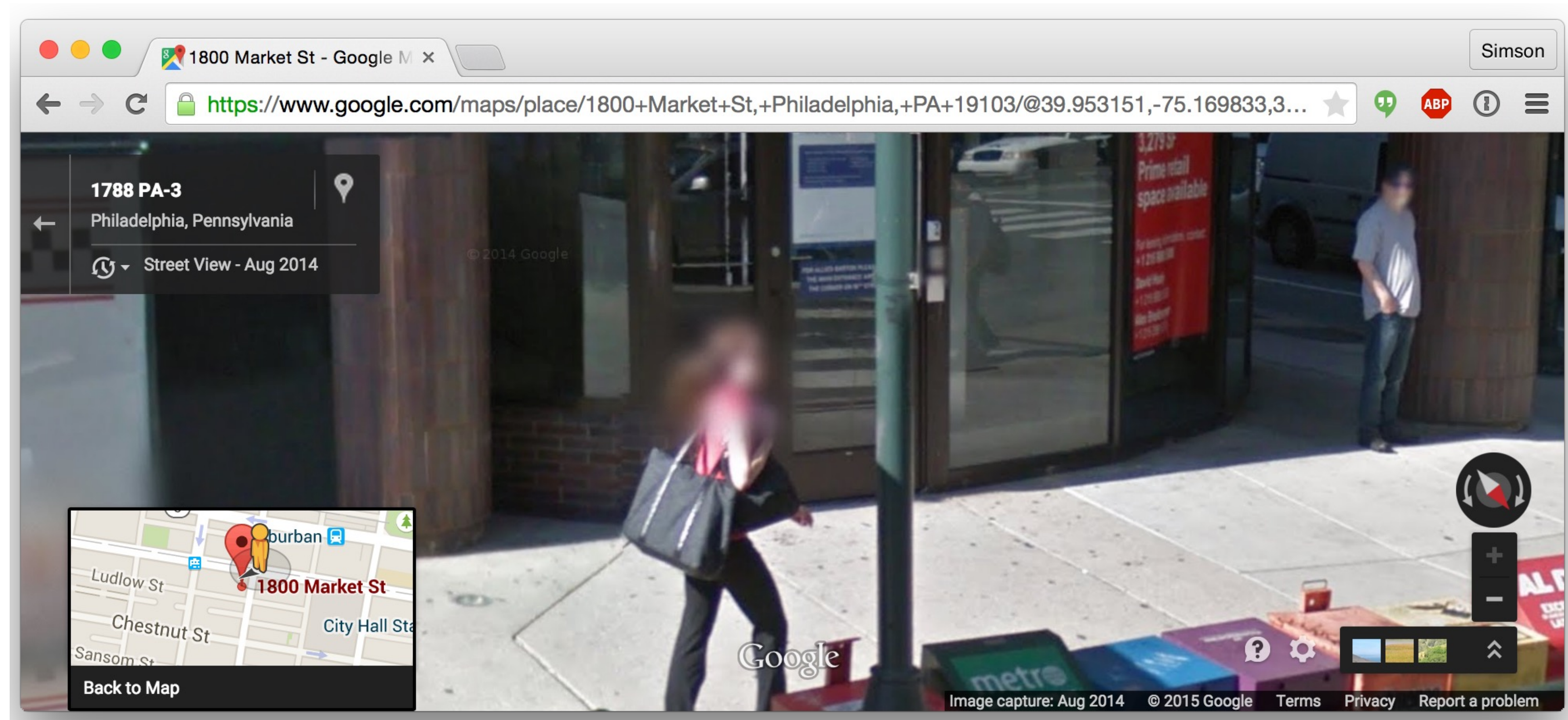
These outputs can result in harms to individuals.
Attackers can “reconstruct” the confidential database.



Input Controls – making data safe for statistics



De-Identification is a common input control.



An example of de-identified data: The Home Mortgage Disclosure Act.

A list of every mortgage issued in the US!

Location (census tract)

Property type

Purpose of loan

Applicant's sex, race, ethnicity, income

Download it while you can!

<https://www.consumerfinance.gov/data-research/hmda/>

The screenshot shows the CFPB's HMDA data page. The header includes the CFPB logo, navigation links (Consumer Education, Rules & Policy, Enforcement, Compliance, Data & Research, News), and a search bar. The main content area is titled "Mortgage data (HMDA)" and describes the data as the most comprehensive source of publicly available information on the U.S. mortgage market. It includes a grid of house icons. Below this, there is a section "About HMDA" explaining the act's purpose and a link to learn more. A "Download HMDA data" section provides links for 2007-2017 data and recent data. A "Mortgage trends" section features a bar chart icon and links to a 2023 data point and reports. A "Further Reading" sidebar on the right lists recent news items, including a 2024 article about upfront fees and a 2022 article about the 2021 market activity report.

Home Mortgage Disclosure Act

consumerfinance.gov/data-research/hmda/

An official website of the United States government

Español 中文 Tiếng Việt 한국어 Tagalog Русский العربية Kreyòl Ayisyen (855) 411-2372

CFPB Consumer Financial Protection Bureau

Search Submit a Complaint

Consumer Education Rules & Policy Enforcement Compliance Data & Research News

Mortgage data (HMDA)

HMDA data are the most comprehensive source of publicly available information on the U.S. mortgage market. Learn more about mortgage activity from these data or download the data for your own analysis.

About HMDA

The Home Mortgage Disclosure Act (HMDA) requires many financial institutions to maintain, report, and publicly disclose loan-level information about mortgages. These data help show whether lenders are serving the housing needs of their communities; they give public officials information that helps them make decisions and policies; and they shed light on lending patterns that could be discriminatory. The public data are modified to protect applicant and borrower privacy.

HMDA was originally enacted by Congress in 1975 and is implemented by Regulation C.

[Learn about complying with HMDA](#)

[Get resources for filers](#)

Download HMDA data

2007-2017 data	Recent HMDA data
Get loan-level data for your state or the nation as a whole.	Get HMDA data collected in or after 2017.
Download 2007-2017 data	See recent data and summaries
	Download data for a specific institution
	Browse HMDA data

Mortgage trends

Each year, millions of people apply for mortgages. Using HMDA data, we can learn what happened to the vast majority of those applications and compare that to previous years. If you want a summary, check out the Bureau's annual Data Point articles and accompanying tables.

- [Data Point: 2023 Mortgage Market Activity and Trends](#)
- [View reports using the HMDA data](#)

FURTHER READING

[Newsroom](#)

[CFPB Finds Americans are Paying Upfront Fees Seeking to Lower Interest Rates on Mortgages](#)

APR 05, 2024

[CFPB Annual Report of 2021 Mortgage Market Activity Reveals an End to the Refinancing Boom and an Increase in Home Purchase Loans](#)

SEP 19, 2022

[2021 HMDA Data on Mortgage Lending Now Available](#)

MAR 24, 2022

[View more](#)

There are (legally) no privacy issues with HMDA’s database

loan_amount_000s	co_applicant_sex_name	applicant_race_name_1	applicant_ethnicity_name	co_applicant_race_name_1	co_applicant_ethnicity_name
215	No co-applicant	White	Not Hispanic or Latino	No co-applicant	No co-applicant
225	No co-applicant	White	Not Hispanic or Latino	No co-applicant	No co-applicant
266	No co-applicant	White	Not Hispanic or Latino	No co-applicant	No co-applicant
320	No co-applicant	White	Not Hispanic or Latino	No co-applicant	No co-applicant
335	No co-applicant	White	Not Hispanic or Latino	No co-applicant	No co-applicant
342	No co-applicant	White	Not Hispanic or Latino	No co-applicant	No co-applicant
352	Female				
355	No co-applicant	White	Not Hispanic or Latino	No co-applicant	No co-applicant
355	No co-applicant			No co-applicant	No co-applicant
382	Male	White	Not Hispanic or Latino	White	Not Hispanic or Latino
399	No co-applicant	White	Not Hispanic or Latino	No co-applicant	No co-applicant
400	No co-applicant	White	Not Hispanic or Latino	No co-applicant	No co-applicant
404	No co-applicant	White	Not Hispanic or Latino	No co-applicant	No co-applicant
404	Female	White	Not Hispanic or Latino	White	Not Hispanic or Latino
404	Female	White	Not Hispanic or Latino	White	Not Hispanic or Latino
413	No co-applicant	White	Not Hispanic or Latino	No co-applicant	No co-applicant
416	No co-applicant	Asian	Not Hispanic or Latino	No co-applicant	No co-applicant
417	Female	White	Not Hispanic or Latino	White	Not Hispanic or Latino
417	Male	White	Not Hispanic or Latino	White	Not Hispanic or Latino
428	Female	White	Not Hispanic or Latino	White	Not Hispanic or Latino
444	Female				
450	No co-applicant	Asian	Not Hispanic or Latino	No co-applicant	No co-applicant
464	Female				
477	Female	White	Not Hispanic or Latino	White	Not Hispanic or Latino
486	No co-applicant	White	Not Hispanic or Latino	No co-applicant	No co-applicant
511	Female				

Problem: de-identified data can be re-identified.



NYC Taxi License Commission provided Chris Whong with *all the data*.

175 million trips

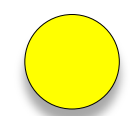
	A	B	C	D	E	F	G	H	I	J	K
1	medallion	hack_license	vendor_id	pickup_datetime	payment_type	fare_amount	surcharge	mta_tax	tip_amount	tolls_amount	total_amount
2	89D227B655E5C82AECF13C3F	BA96DE419E711691B944	CMT	1/1/13 15:11	CSH	6.5	0	0.5	0	0	7
3	0BD7C8F5BA12B88E0B67BED	9FD8F69F0804BDB5549F	CMT	1/6/13 0:18	CSH	6	0.5	0.5	0	0	7
4	0BD7C8F5BA12B88E0B67BED	9FD8F69F0804BDB5549F	CMT	1/5/13 18:49	CSH	5.5	1	0.5	0	0	7
5	DFD2202EE08F7A8DC9A57B0	51EE87E3205C985EF843	CMT	1/7/13 23:54	CSH	5	0.5	0.5	0	0	6
6	DFD2202EE08F7A8DC9A57B0	51EE87E3205C985EF843	CMT	1/7/13 23:25	CSH	9.5	0.5	0.5	0	0	10.5
7	20D9ECB2CA0767CF7A01564	598CCE5B9C1918568DEE	CMT	1/7/13 15:27	CSH	9.5	0	0.5	0	0	10
8	496644932DF3932605C22C79	513189AD756FF14FE670	CMT	1/8/13 11:01	CSH	6	0	0.5	0	0	6.5
9	0B57B9633A2FECD3D3B1944	CCD4367B417ED6634D9	CMT	1/7/13 12:39	CSH	34	0	0.5	0	4.8	39.3
10	2C0E91FF20A856C891483ED6	1DA2F6543A62B8ED934	CMT	1/7/13 18:15	CSH	5.5	1	0.5	0	0	7

- Every trip:
- Pickup date, time & GPS
 - Drop-off date, time & GPS
 - Fare & tip
 - Encoded medallion number.

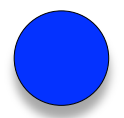
Some pick-up locations were from sensitive locations...
Some drop-offs were identifying....



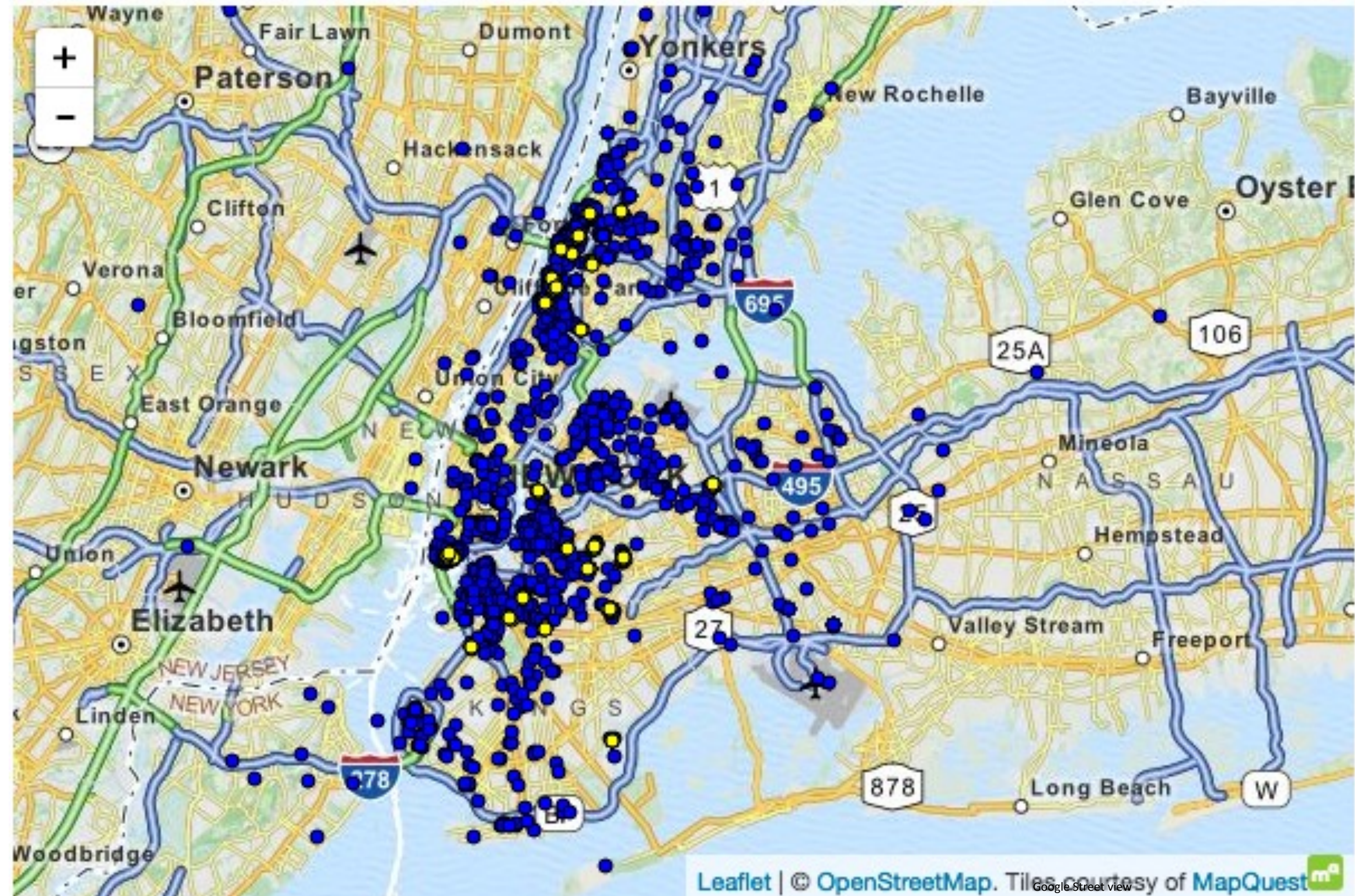
<http://content.research.neustar.biz/blog/differential-privacy/stripRaw.html>



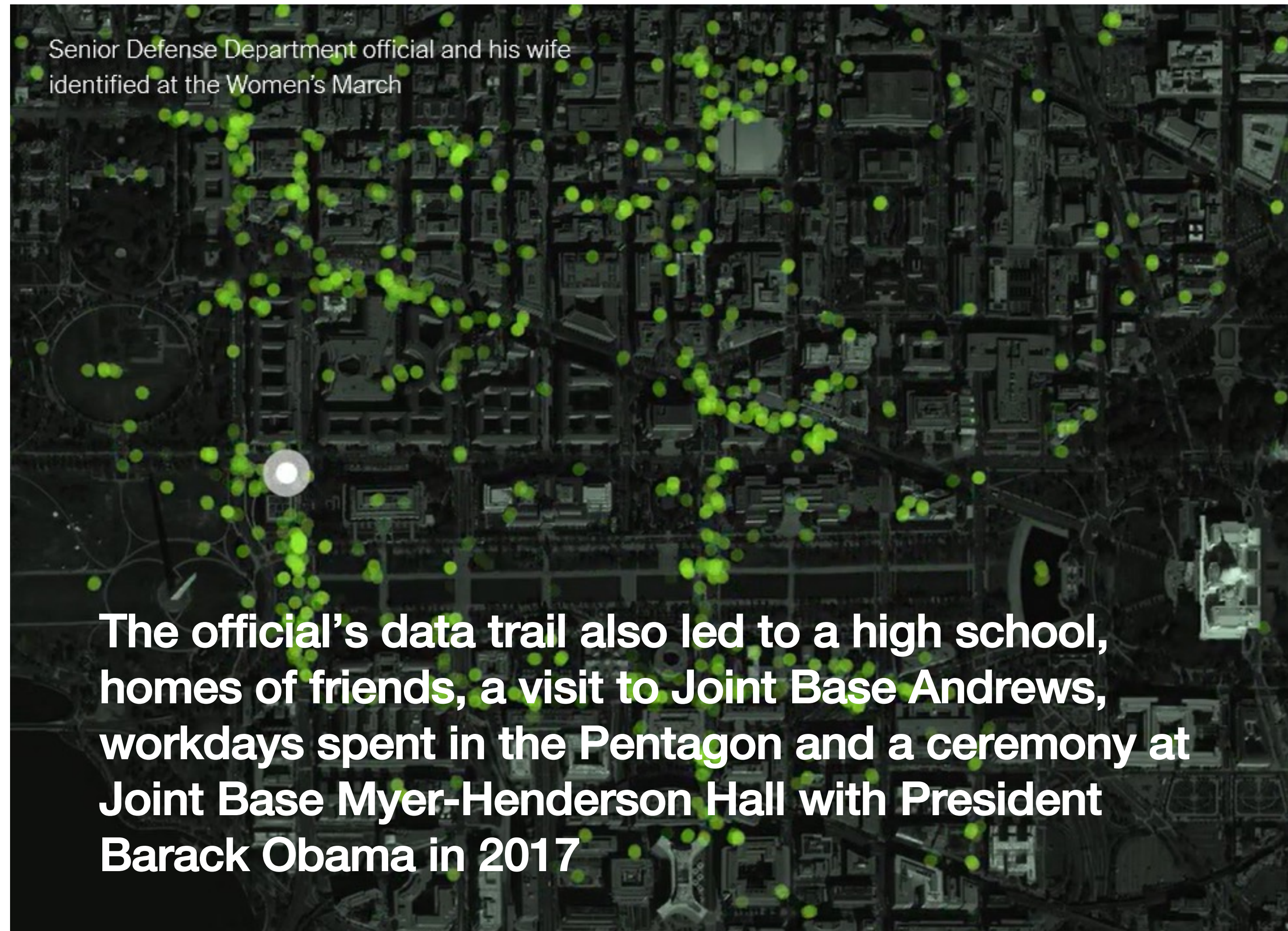
Frequent customer



Occasional customer



Location traces can uniquely identify you.



<https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html>

Year of birth, ZIP code, gender

State	#ZIPs	Population	RANGE								
			%population	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
VT	243	562,758	98%	100.0%	100.0%	80.1%	100.0%	100.0%	100.0%	100.0%	100.0%
VA	820	6,184,493	87%	88.2%	91.6%	71.9%	75.5%	82.7%	97.8%	100.0%	100.0%
WA	484	4,866,692	92%	94.6%	100.0%	82.8%	82.5%	87.2%	100.0%	100.0%	100.0%
WV	655	1,792,969	97%	96.7%	96.4%	90.2%	95.7%	96.4%	100.0%	100.0%	96.5%
WI	714	4,891,452	92%	88.9%	97.7%	77.6%	86.4%	92.6%	100.0%	100.0%	100.0%
WY	141	453,588	98%	100.0%	100.0%	79.2%	100.0%	100.0%	100.0%	100.0%	100.0%
USA	29,343	248,418,140	87%	85.8%	90.2%	75.0%	75.1%	87.0%	97.8%	99.0%	95.3%

L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

<https://dataprivacylab.org/projects/identifiability/paper1.pdf>

Output controls
Because input controls are not enough!

Every published statistic is a constraint on the confidential data.
These constraints can be mathematically *solved*.

Student Scores
(Hidden variables)

S1	S6
S2	S7
S3	S8
S4	S9
S5	S10

Your Score

Class Average = 98 (published)

Implies:

$$\frac{(S1 + S2 + S3 + S4 + S5 + S6 + S7 + S8 + S9 + S10)}{10} = 98$$

If

S10=80

and

$0 \leq S_n \leq 100$

then:

S1..S9 = 100

“The Fundamental Law of Information Recovery”

Every statistical release based on confidential data leaks some aspect of the confidential data on which it is based.

If the results of enough queries are released, the entire confidential database is eventually revealed.

or more simply...

“Overly accurate answers to too many questions will destroy privacy in a spectacular way.”

- [The Algorithmic Foundations of Differential Privacy](#) by [Cynthia Dwork](#) and [Aaron Roth](#).
Foundations and Trends in Theoretical Computer Science. Vol. 9, no. 3–4, pp. 211-407, Aug. 2014. [DOI:10.1561/04000000042](https://doi.org/10.1561/04000000042)

Sometimes “too many questions” is “1 question” (as evidenced here)

What if the teacher said the class average is between 98 and 100?

This is called **generalization** or **coarsening**


$$\text{Class Average} = 98 \dots 100$$

But wait —

Those other 9 students all got 100!

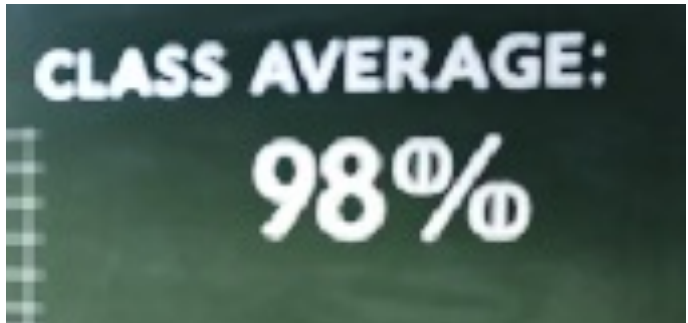
They are all friends and get together at lunch.

Now they can compute bounds on your grade:

$$98 \leq \frac{900 + (\textit{your grade})}{10} \leq 100$$

They know you got between an 80 and 100!

Differential privacy adds noise.
Noise creates uncertainty about the confidential data.

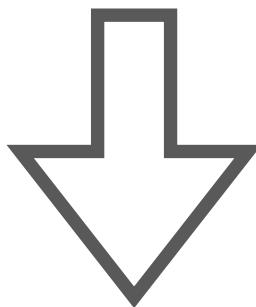


+ noise →

	0-60	61-70	71-80	81-90	91-100
West Side kids	10	15	75	95	65
East Side kids	15	45	125	0	0

+ noise →

	0-60	61-70	71-80	81-90	91-100
West Side kids	8	16	76	92	68
East Side kids	12	48	121	3	1



database
reconstruction

Original confidential
database



database
reconstruction

Incorrect
reconstruction

A common form of differential privacy:
Add a random value (x) to each statistic before it is published.

What we compute:

$$\text{Class Average} = 98 \text{ (published)} + \text{noise} = 96.5$$

What we publish:

$$\text{Class Average} = 98 \text{ (published)} + x = 96.5$$


The more noise that's added, the more privacy the students get!
And the worse the statistics...

Differential Privacy is tunable...

The “privacy loss” of differential privacy is a tunable parameter: $0 \leq \epsilon \leq \infty$

There’s no “right” value for ϵ — it’s a policy decision

More privacy loss = more accurate statistics.

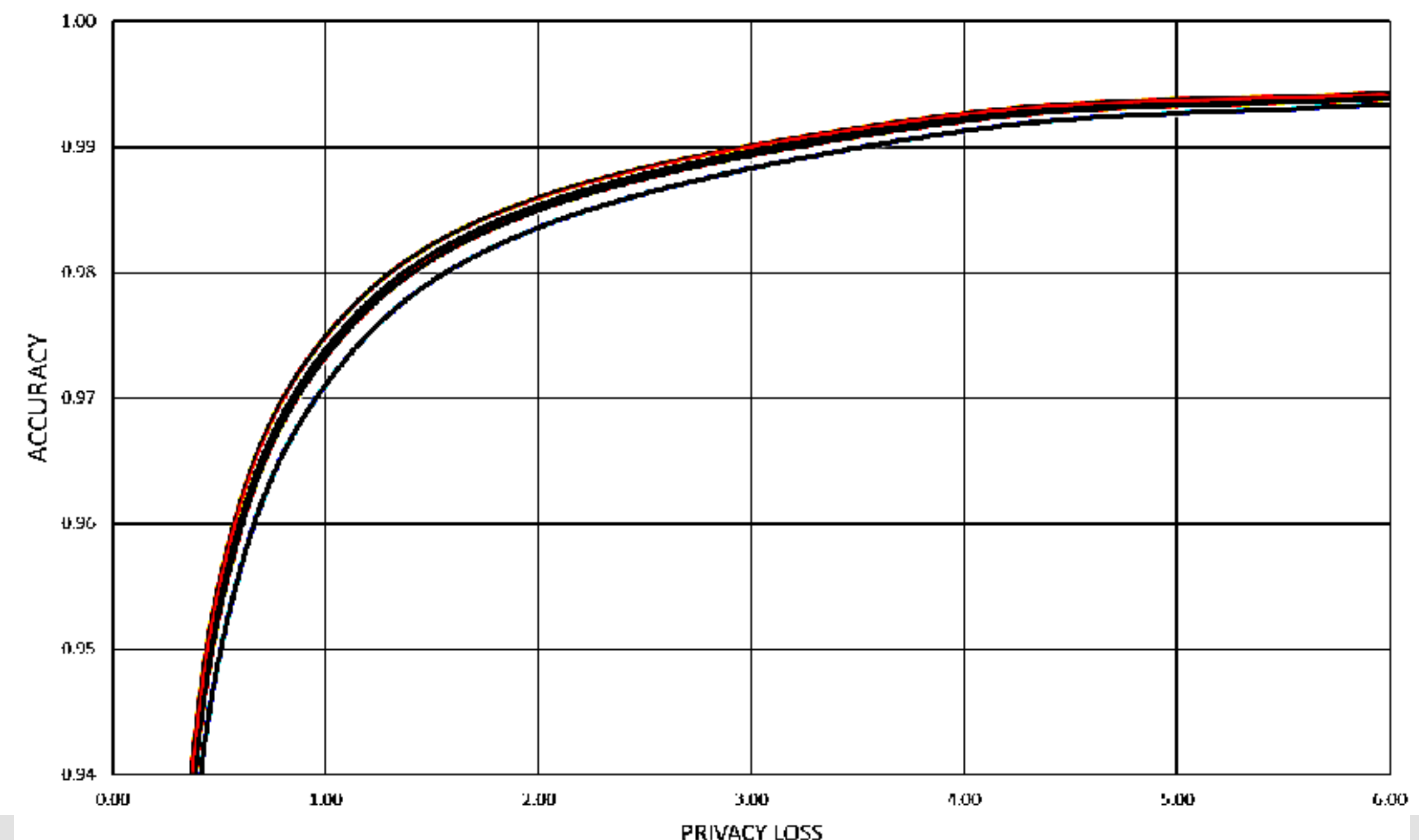
(Public good) vs. (personal privacy protection)

DP tells data users the accuracy of data.

Accuracy is hidden by other approaches.

*More “privacy loss”
More accuracy*

*Less “privacy loss”
Less accuracy*



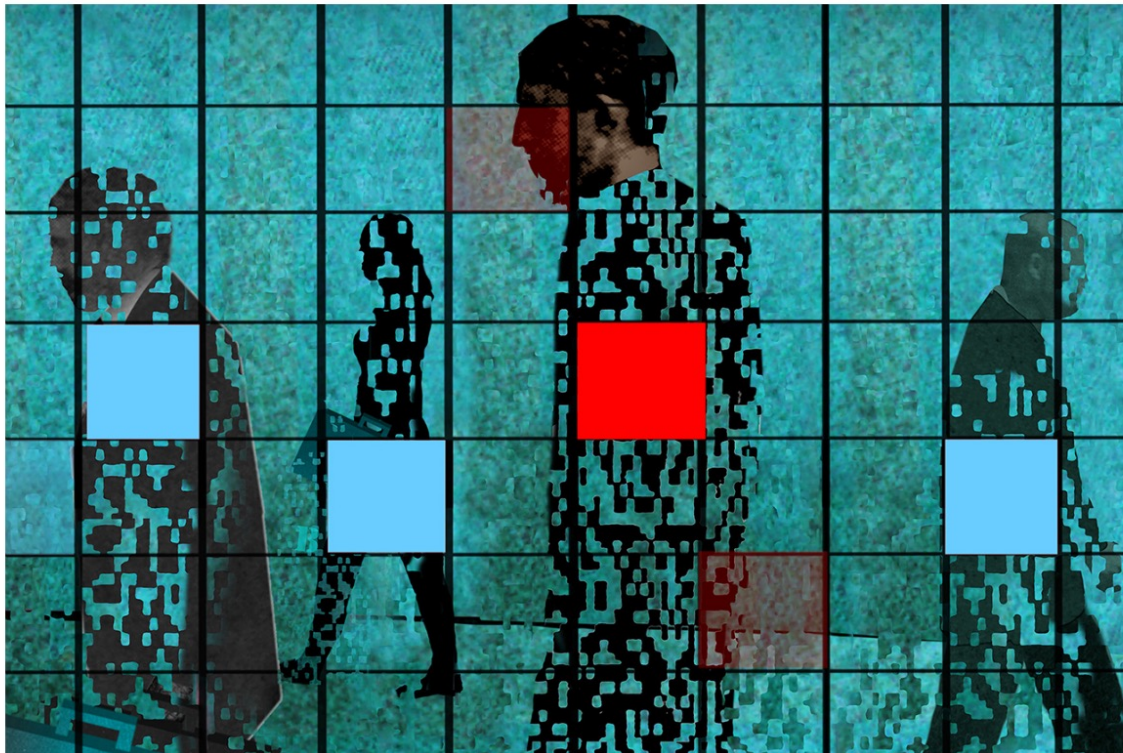
There’s a lot more to say...

Preserving Privacy While Sharing Data

Differential privacy can safeguard personal information when data is being shared, but it requires a high level of expertise.

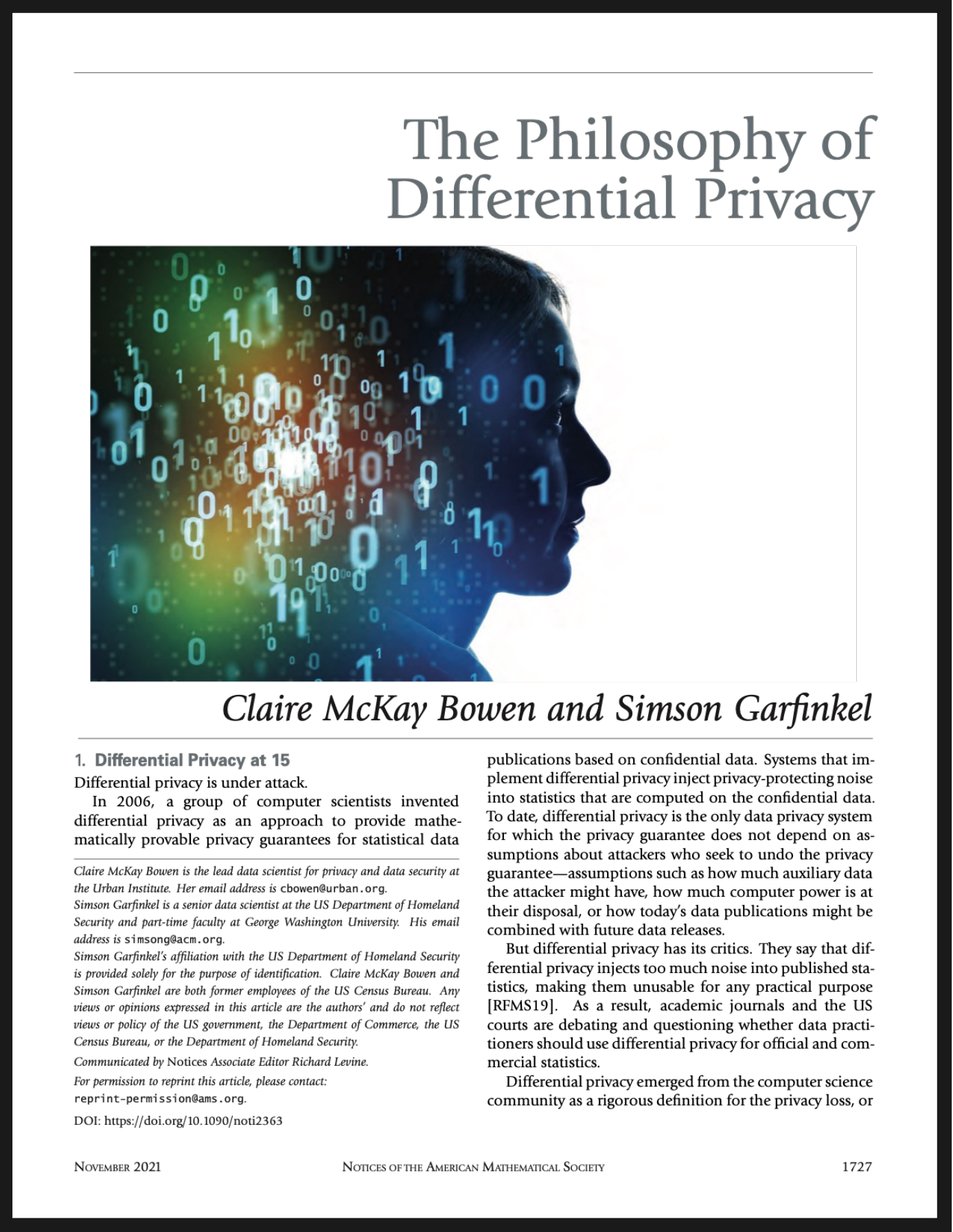
Simson L. Garfinkel and Claire McKay Bowen • April 26, 2022 Reading Time: 9 min

SUBSCRIBE PERMISSIONS AND PDF SHARE

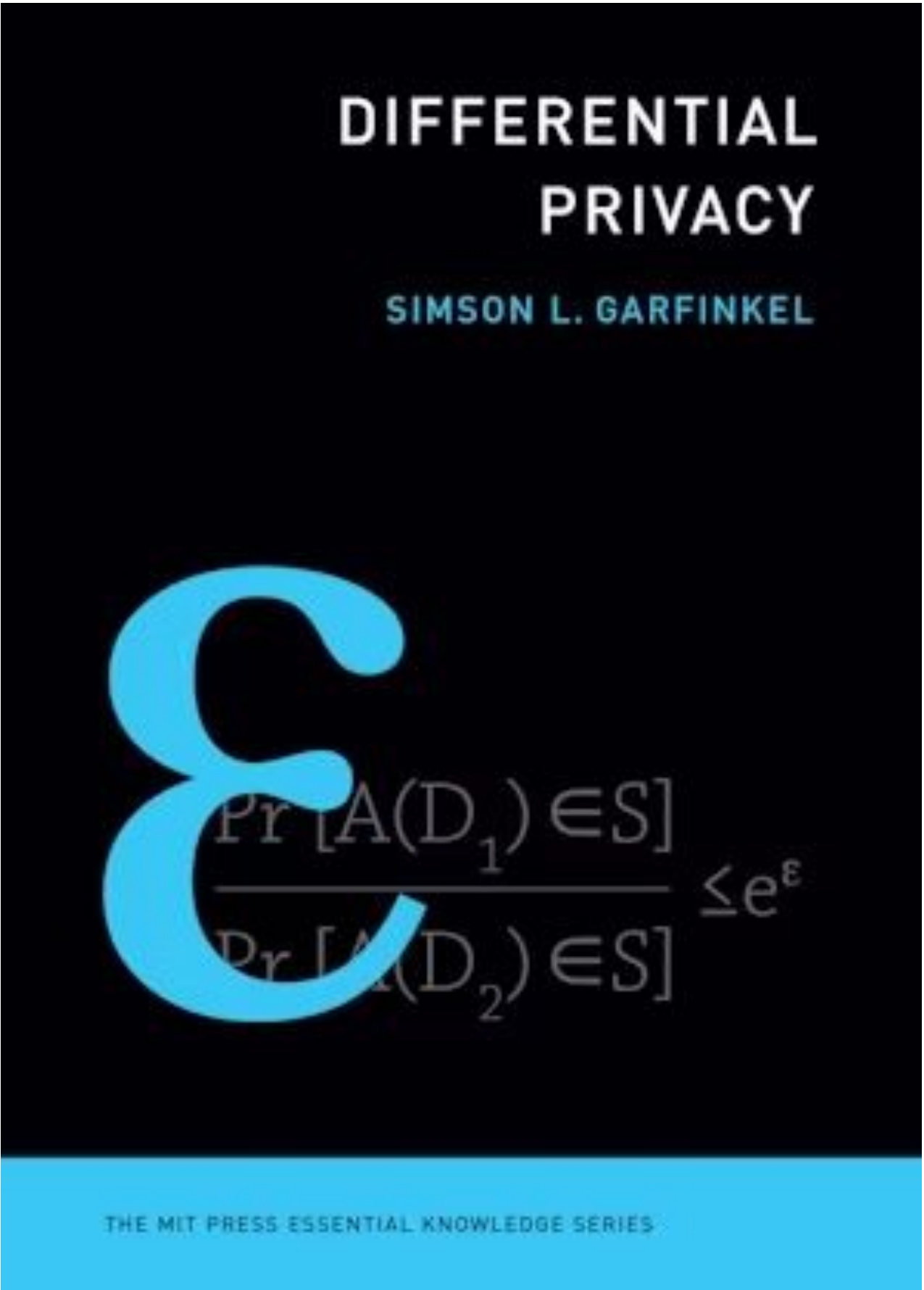


Michael Morgenstern/theispost.com

Garfinkel & Bowen
MIT Sloan Management Review
April 26, 2022

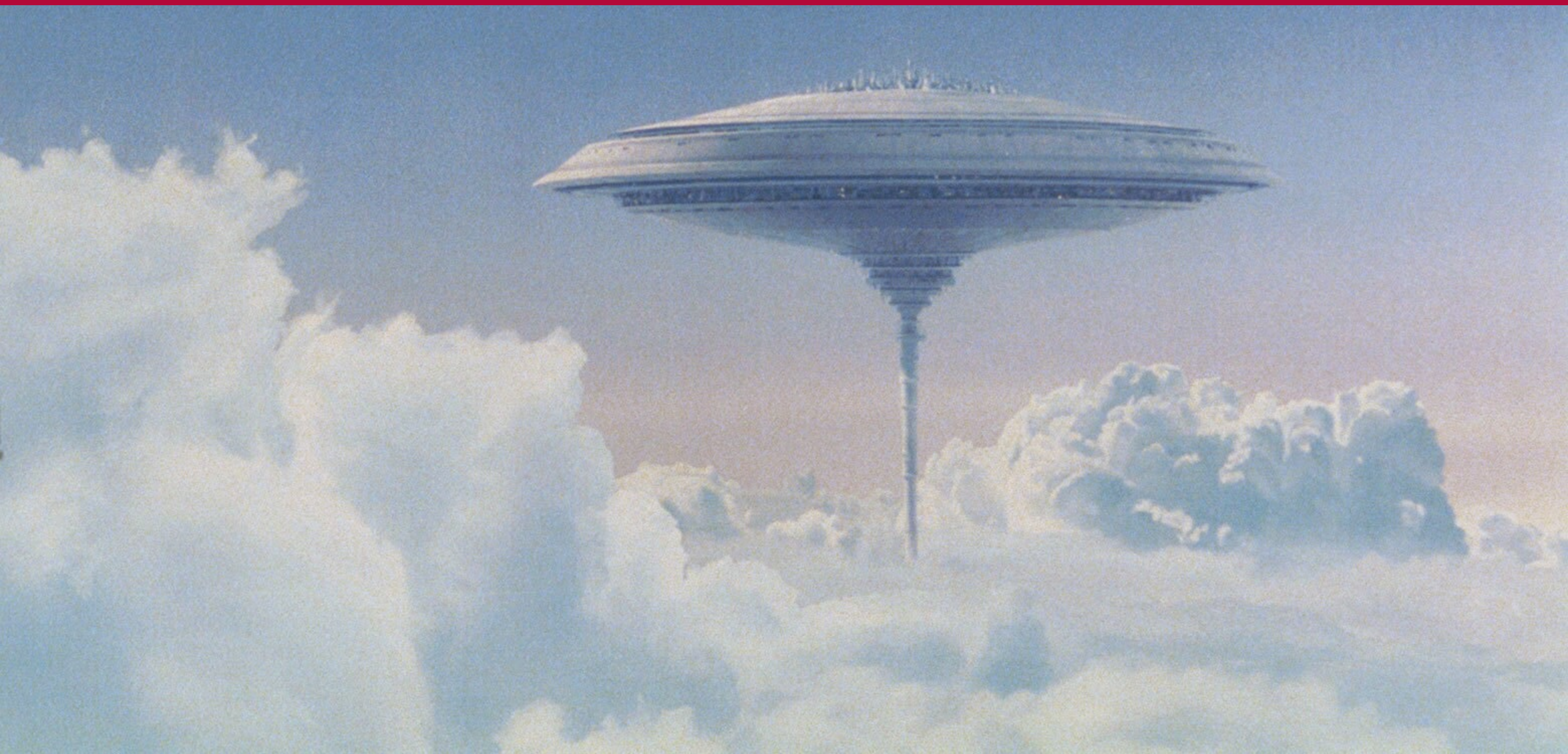


“The Philosophy of Differential Privacy”
Bowen & Garfinkel
Notices of the American Mathematical Society
November 2021



Differential Privacy
Garfinkel
MIT Press
March 25, 2025. (Open Access)

Working in the cloud is dangerous if you can't trust the host.



Secure Enclaves

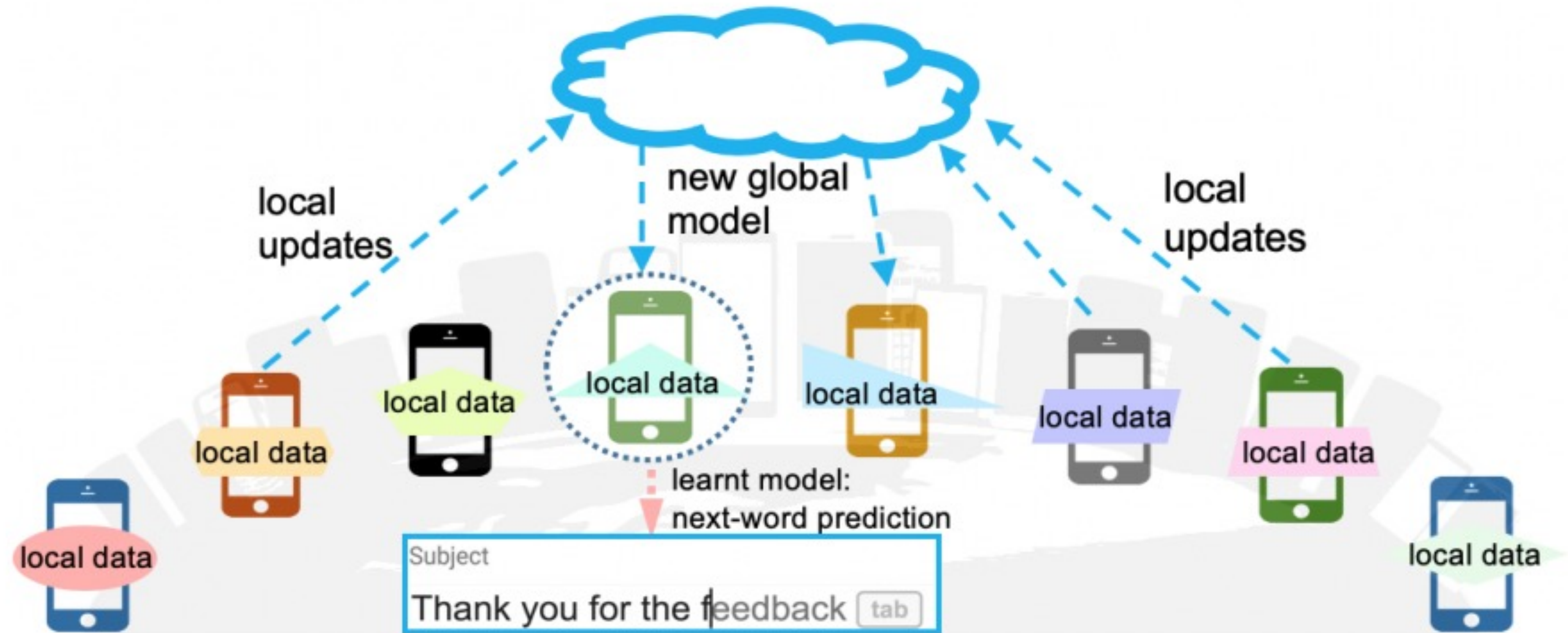
Overview of some prominent TEEs.

TEE	Target ISA	Security Features						
TPM/vTPM	Multiple Targets	◐	●	●	●	—	—	—
Intel TXT	x86_64	●	●	●	●	—	—	—
Intel SGX	x86_64	●	●	●	—	—	—	—
ARM TrustZone	ARM	●	—	◐	●	●	◐	—
Sanctum/KeyStone/ MultiZone	RISC-V	●	●	●	●	—	◐	●
AMD SEV	AMD x86	●	●	●	—	—	—	—

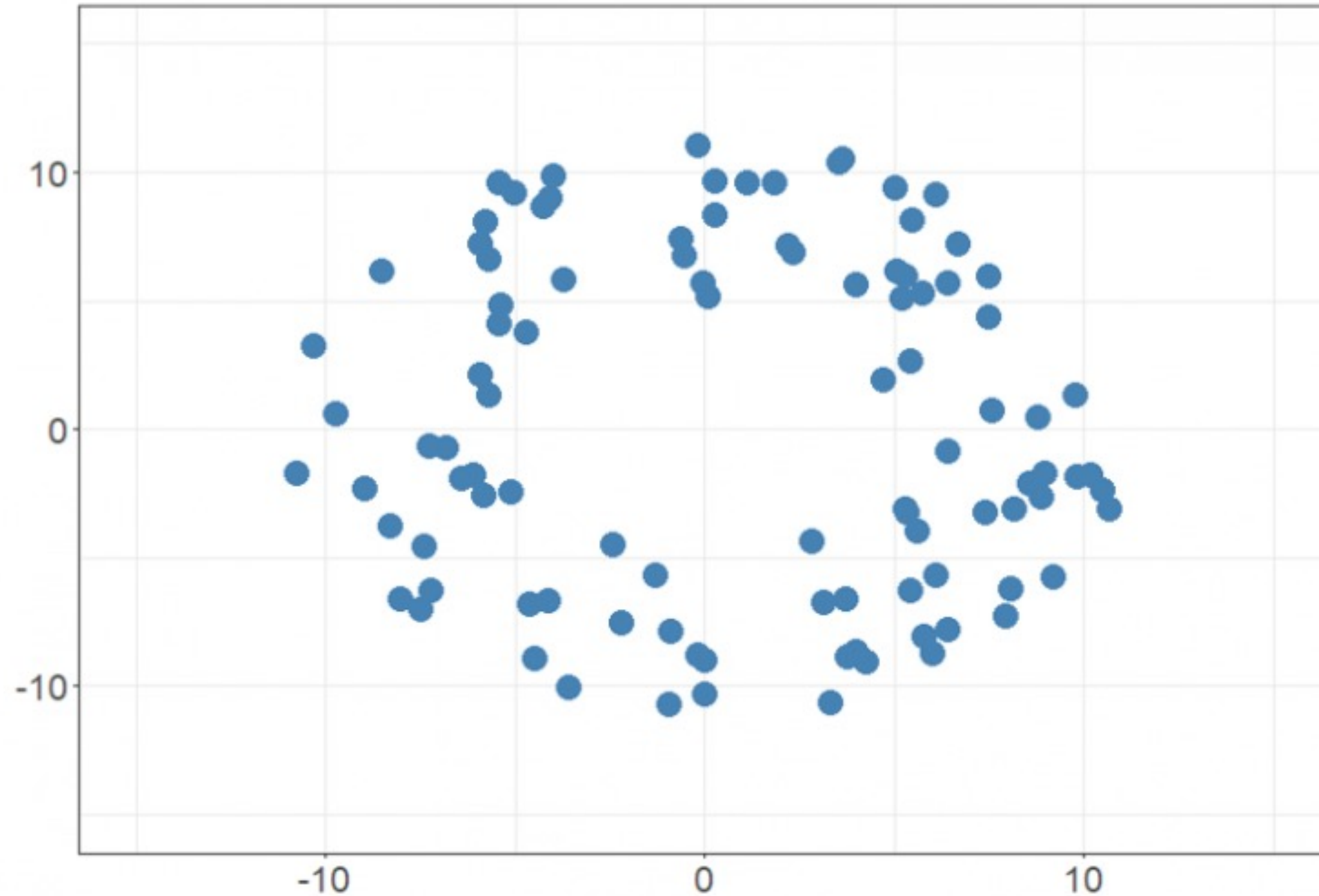
- provides property
- ◐ supports partially
- does not support

In-enclave protection
Remote attestation
Sealing
Secure boot
Isolated peripherals
Side-channel resistance
Open source hardware

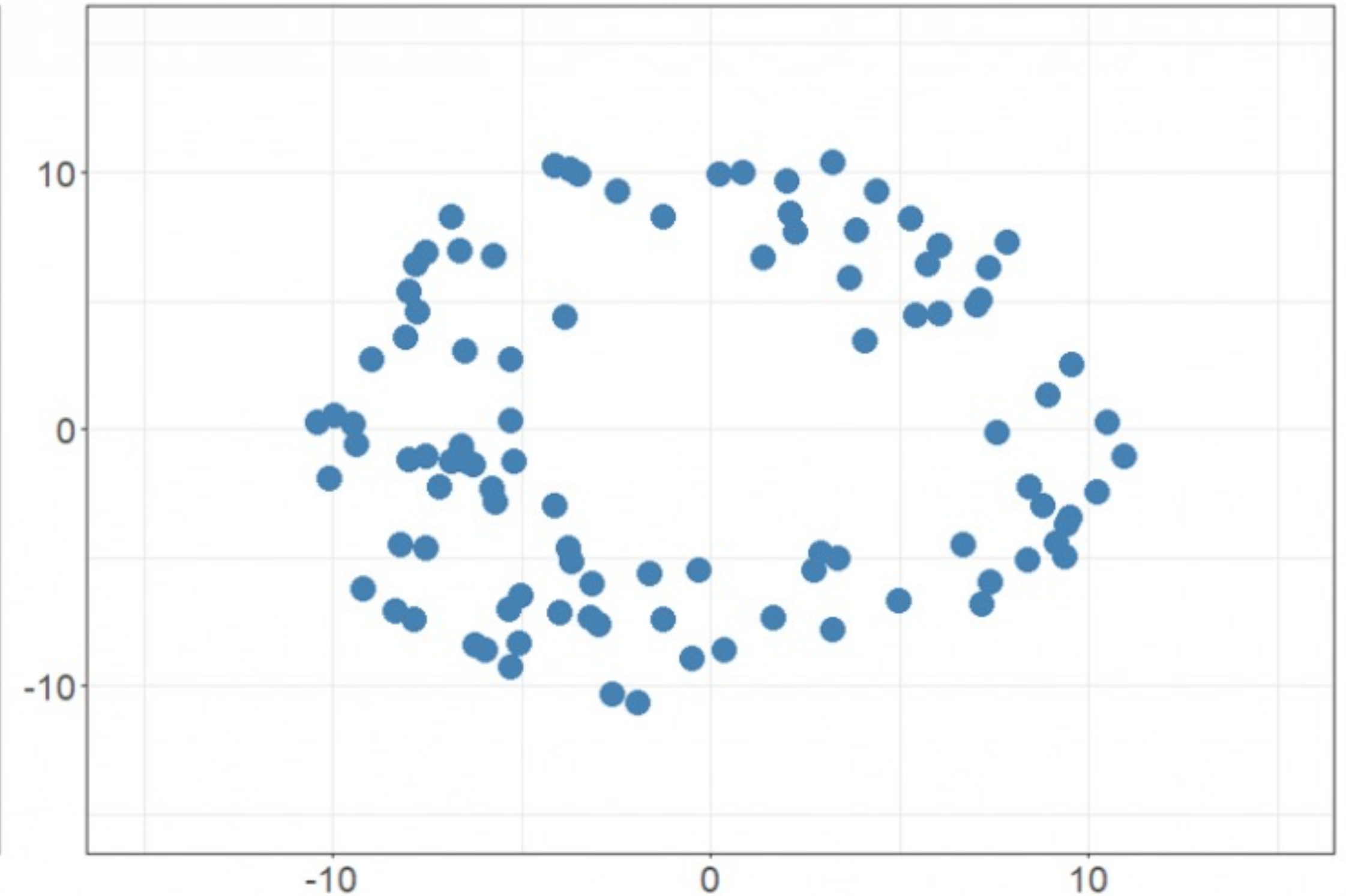
Federated Learning



Synthetic Data



Original data



Synthetic data

The synthetic data retains the structure of the original data but is not the same

Lab 4 — Due yesterday!

Virtual machines are graded

Grades are synched to Canvas.

If you got less than a 5 — figure out why.

If you got a zero — something is wrong. Please contact us.

aws

Search

[Option+S]

United States (N. Virginia)

Simson Harvard (5867-9448-3136)

Simson Harvard

Certificate Manager

CloudFormation

DynamoDB

Secrets Manager

IAM

S3

Cloud9

CloudWatch

Lambda

EC2

Route 53

Amazon Simple Email Se...

IAM Identity Center

Console Home

Info

Reset to default layout

Add widgets

Recently visited

Info

Route 53

Billing and Cost Management

Amazon Simple Email Service

Amazon WorkMail

EC2

CloudWatch

Lambda

IAM Identity Center

Secrets Manager

IAM

DynamoDB

CloudFormation

Service Quotas

Certificate Manager

View all services

Applications (0)

Info

Create application

Region: US East (N. Virginia)

Select Region

us-east-1 (Current Region)

Find applications

< 1 >

Name

Description

Region

Originati.

No applications

Get started by creating an application.

Create application

Go to myApplications

Welcome to AWS

Getting started with AWS

Learn the fundamentals and find valuable information to get the most out of AWS.

Training and certification

Learn from AWS experts and advance your skills and knowledge.

AWS Builder Center

Learn, build, and connect with builders in the AWS community.

AWS Health

Info

Open issues

0

Past 7 days

Scheduled changes

0

Upcoming and past 7 days

Other notifications

0

Past 7 days

Cost and usage

Info

Current month

\$4.45

35%

Forecasted month end

\$12.59

34%

Savings opportunities

Not enabled

Cost (\$)

Oct 25

Nov 25

Dec 25

Jan 26

Feb 26

Mar 26

Month (Year)

WorkMail

Route 53

Virtual Private Cloud

Secrets Manager

EC2 - Other

Others

Feb 26

WorkMail	\$4.00
Route 53	\$1.34
Virtual Private Cloud	\$1.45
Secrets Manager	\$1.69
EC2 - Other	\$0.55
Others	\$0.36

Feb 26

Mar 20

s Manager

Lambda

<

Dashboard

Applications

Functions

▼ Additional resources

Capacity providers [New](#)

Code signing configurations

Event source mappings

Functions (3)

Last fetched 3/10/2026, 4:06:19 PM

Actions ▼

Create function

< 1 > ⚙

☐	Function name ▼	Description ▼	Package type ▼	Runtime ▼	Type ▼	Last modified ▼
☐	home-app-stage-E11HomeFunction-K9TirhT4ZoN7	-	Zip	Python 3.13	Standard	2 weeks ago
☐	home-app-prod-E11HomeFunction-9BWcNJ0jnNq4	-	Zip	Python 3.13	Standard	2 weeks ago
☐	leaderboard-LeaderboardFlaskFunction-wLJEx0EJOZJp	-	Zip	Python 3.13	Standard	3 weeks ago

Invocations



Count

1.13K

567

0

03/04

03/06

03/08

03/10

— Invocations [sum: 10,862]

Duration



Milliseconds

21.43K

10.72K

0

03/04

03/06

03/08

03/10

— Minimum [1.34] — Average [338] — Maximum [21,432]

Throttles



Count

1

0.5

0

03/04

03/06

03/08

03/10

— Throttles [max: 0]

Total concurrent executions



Count

27

13.5

0

03/04

03/06

03/08

03/10

— Concurrent executions [max: 27]

Recent invocations



#	: Timestamp	: RequestId	: LogStream	: DurationInMS	: BilledDurationIn...	: MemorySetInMB	: MemoryUsedInMB
▶ 1	2026-03-10T20:06:19.832Z	7508abd1-9e32-4956-b63a-6552c6cd54a8	2026/03/10/[\$LATEST]1ddb28d2521844a7b0920e600354bda0	279.58	280.0	256.0	149.0
▶ 2	2026-03-10T20:01:19.958Z	76959336-3211-47cb-ab6d-0df078de7561	2026/03/10/[\$LATEST]1ddb28d2521844a7b0920e600354bda0	276.81	277.0	256.0	149.0
▶ 3	2026-03-10T19:56:19.893Z	bc0f6cea-d90a-4e0a-b0fa-dc0460927bce	2026/03/10/[\$LATEST]1ddb28d2521844a7b0920e600354bda0	305.87	306.0	256.0	149.0
▶ 4	2026-03-10T19:51:19.775Z	e4864ef0-8e76-4d95-bc3f-6283181dc966	2026/03/10/[\$LATEST]1ddb28d2521844a7b0920e600354bda0	279.46	280.0	256.0	149.0
▶ 5	2026-03-10T19:46:19.786Z	10051435-31fd-4755-b764-4cf05f1329e4	2026/03/10/[\$LATEST]1ddb28d2521844a7b0920e600354bda0	262.12	263.0	256.0	149.0
▶ 6	2026-03-10T19:41:19.837Z	e6b32de9-2271-4f49-b142-5904068fbb8f	2026/03/10/[\$LATEST]1ddb28d2521844a7b0920e600354bda0	269.57	270.0	256.0	149.0
▶ 7	2026-03-10T19:36:19.679Z	6a9f897a-9ccc-40a6-a370-478dbe82aca9	2026/03/10/[\$LATEST]1ddb28d2521844a7b0920e600354bda0	242.82	243.0	256.0	148.0
▶ 8	2026-03-10T19:34:48.237Z	56394ccf-35bf-4da6-a427-e99e2209881f	2026/03/10/[\$LATEST]1ddb28d2521844a7b0920e600354bda0	2.04	3.0	256.0	148.0
▶ 9	2026-03-10T19:34:48.040Z	c2e08669-23a2-4fc3-a68d-802a1be8cffd	2026/03/10/[\$LATEST]1ddb28d2521844a7b0920e600354bda0	1.87	2.0	256.0	148.0

Most expensive invocations in GB-seconds (memory assigned * billed duration)



#	: Timestamp	: RequestId	: LogStream	: BilledDurationIn...	: MemorySetInMB	: BilledDurationInGBSeconds
▶ 1	2026-03-07T02:57:33.964Z	3e5a14f2-e002-4d7b-b191-79158d903a2a	2026/03/07/[\$LATEST]53256f761e8945229faa8f03845369f3	21433	256	5.3582
▶ 2	2026-03-08T11:47:32.276Z	4d9150e4-593c-5d19-959b-32c33d397765	2026/03/08/[\$LATEST]da5fa65ec2fe42bdb4afc4b2f4f92dea	14180	256	3.545
▶ 3	2026-03-10T19:22:08.260Z	02bbbc86-1a23-5757-9771-deac6bb1c4e0	2026/03/10/[\$LATEST]7766a880b2fc48aab7bd3197891830c3	13456	256	3.364
▶ 4	2026-03-10T19:22:13.076Z	7a40bb0d-3ee8-5e0a-a95f-d9274f7f1dbd	2026/03/10/[\$LATEST]39ddcb3a5ea24291b29130e4716bdb28	13384	256	3.346
▶ 5	2026-03-10T19:22:08.023Z	89ccea1e-e2a9-57a7-b25d-8834933b1b02	2026/03/10/[\$LATEST]6380519c6cd448848e46fb0e34a9b813	13332	256	3.333
▶ 6	2026-03-09T18:51:11.616Z	65931260-d5ff-528e-8940-7049cd2eebe5	2026/03/09/[\$LATEST]65258aac948d490fb067b59d4d9f1e7f	13314	256	3.3285
▶ 7	2026-03-09T18:51:06.811Z	f927c99c-95b0-5d29-a4a1-abb58d7d0442	2026/03/09/[\$LATEST]7275bc59cdb84c9ba7daf867028a54a2	13303	256	3.3258
▶ 8	2026-03-08T11:47:22.584Z	a3fcf3b4-8de6-5e1e-b6c6-9193ee1aa462	2026/03/08/[\$LATEST]f4e08f5cfba0403ebb31a7590baf6821	13252	256	3.313
▶ 9	2026-03-09T18:51:07.368Z	0ea7c127-35a6-50f6-9dd4-a30e2d6bb99d	2026/03/09/[\$LATEST]5369c28fcc814db2a529126d27c8bf81	13219	256	3.3047

Failed invocations (execution errors, runtime exceptions, timeouts)



#	: Timestamp	: RequestId	: LogStream	: ErrorMessage
▶ 1	2026-03-10T19:22:29.437Z	f24da8c1-2741-5b14-af6c-caeb56d6f8a3	2026/03/10/[\$LATEST]39ddcb3a5ea24291b29130e4716bdb28	[ERROR] 2026-03-10T19:22:29.437Z f24da8c1-2741-5b14-af6c-caeb56d6f8a3 OS...
▶ 2	2026-03-10T19:22:29.437Z		2026/03/10/[\$LATEST]39ddcb3a5ea24291b29130e4716bdb28	2026-03-10 19:22:29,437 ERROR [e11.e11core e11ssh.py:57 __init__] OSErro...
▶ 3	2026-03-10T19:22:12.853Z	7a40bb0d-3ee8-5e0a-a95f-d9274f7f1dbd	2026/03/10/[\$LATEST]39ddcb3a5ea24291b29130e4716bdb28	[ERROR] 2026-03-10T19:22:12.853Z 7a40bb0d-3ee8-5e0a-a95f-d9274f7f1dbd OS...
▶ 4	2026-03-10T19:22:12.853Z		2026/03/10/[\$LATEST]39ddcb3a5ea24291b29130e4716bdb28	2026-03-10 19:22:12,853 ERROR [e11.e11core e11ssh.py:57 __init__] OSErro...
▶ 5	2026-03-10T19:22:07.913Z	02bbbc86-1a23-5757-9771-deac6bb1c4e0	2026/03/10/[\$LATEST]7766a880b2fc48aab7bd3197891830c3	[ERROR] 2026-03-10T19:22:07.912Z 02bbbc86-1a23-5757-9771-deac6bb1c4e0 OS...
▶ 6	2026-03-10T19:22:07.913Z		2026/03/10/[\$LATEST]7766a880b2fc48aab7bd3197891830c3	2026-03-10 19:22:07,912 ERROR [e11.e11core e11ssh.py:57 __init__] OSErro...
▶ 7	2026-03-10T19:22:07.791Z	89ccea1e-e2a9-57a7-b25d-8834933b1b02	2026/03/10/[\$LATEST]6380519c6cd448848e46fb0e34a9b813	[ERROR] 2026-03-10T19:22:07.791Z 89ccea1e-e2a9-57a7-b25d-8834933b1b02 OS...
▶ 8	2026-03-10T19:22:07.791Z		2026/03/10/[\$LATEST]6380519c6cd448848e46fb0e34a9b813	2026-03-10 19:22:07,791 ERROR [e11.e11core e11ssh.py:57 __init__] OSErro...
▶ 9	2026-03-10T19:22:07.117Z		2026/03/10/[\$LATEST]daecbbe95ff748f483f248118400e2cf	2026-03-10 19:22:07,117 ERROR [e11.grader grader.py:151 discover_and_run...

<

▶

0

ity

ls (APM)

nitoring

ing

Log events

You can use the filter bar below to search for and match terms, phrases, or values in your log events. [Learn more about filter patterns](#)

Clear

1m

30m

1h

12h

Custom

UTC timezone

Display

▶	Timestamp	Message
		There are older events to load. Load more .
▶	2026-03-10T19:22:29.437Z	[ERROR] 2026-03-10T19:22:29.437Z f24da8c1-2741-5b14-af6c-caeb56d6f8a3 OSError e=timed out hostname=13.58.77.49 port=22 username=ubun...
▶	2026-03-10T19:22:29.437Z	2026-03-10 19:22:29,437 ERROR [e11.e11core e11ssh.py:57 __init__] OSError e=timed out hostname=13.58.77.49 port=22 username=ubuntu
▶	2026-03-10T19:22:29.437Z	2026-03-10 19:22:29,437 INFO [e11.home api.py:336 api_grader] summary={'lab': 'lab4', 'passes': [], 'fails': ['cannot connect by ssh...
▶	2026-03-10T19:22:29.437Z	[INFO] 2026-03-10T19:22:29.437Z f24da8c1-2741-5b14-af6c-caeb56d6f8a3 summary={'lab': 'lab4', 'passes': [], 'fails': ['cannot connect...
▶	2026-03-10T19:22:29.460Z	2026-03-10 19:22:29,460 INFO [e11 e11_common.py:410 add_grade] add_grade to dynamodb.Table(name='e11-users') user=user_id='0aee15c1-...
▶	2026-03-10T19:22:29.460Z	[INFO] 2026-03-10T19:22:29.460Z f24da8c1-2741-5b14-af6c-caeb56d6f8a3 add_grade to dynamodb.Table(name='e11-users') user=user_id='0ae...
▶	2026-03-10T19:22:29.586Z	[INFO] 2026-03-10T19:22:29.586Z f24da8c1-2741-5b14-af6c-caeb56d6f8a3 send_email2 to=['sgarfinkel@fas.harvard.edu', 'simsong@gmail.co...
▶	2026-03-10T19:22:29.586Z	2026-03-10 19:22:29,586 INFO [e11 e11_common.py:492 send_email2] send_email2 to=['sgarfinkel@fas.harvard.edu', 'simsong@gmail.com'] ...
▶	2026-03-10T19:22:29.588Z	END RequestId: f24da8c1-2741-5b14-af6c-caeb56d6f8a3
▶	2026-03-10T19:22:29.588Z	REPORT RequestId: f24da8c1-2741-5b14-af6c-caeb56d6f8a3 Duration: 10210.41 ms Billed Duration: 10211 ms Memory Size: 256 MB Max Memor...
▶	2026-03-10T19:26:19.502Z	START RequestId: ca64e160-92a6-42e2-b35c-9faa32c355d4 Version: \$LATEST
▶	2026-03-10T19:26:19.503Z	[INFO] 2026-03-10T19:26:19.503Z ca64e160-92a6-42e2-b35c-9faa32c355d4 log_level=INFO req method='GET' path='/' action='None' source_i...
▶	2026-03-10T19:26:19.512Z	2026-03-10 19:26:19,503 INFO [e11.home home.py:640 lambda_handler] log_level=INFO req method='GET' path='/' action='None' source_ip=...
▶	2026-03-10T19:26:19.873Z	END RequestId: ca64e160-92a6-42e2-b35c-9faa32c355d4
▶	2026-03-10T19:26:19.874Z	REPORT RequestId: ca64e160-92a6-42e2-b35c-9faa32c355d4 Duration: 371.09 ms Billed Duration: 372 ms Memory Size: 256 MB Max Memory Us...

No more records within selected time range *Auto retry paused.* [Resume](#)



ity
ls (APM)
nitoring
ing

Log events

You can use the filter bar below to search for and match terms, phrases, or values in your log events. [Learn more about filter patterns](#)



Actions

Start tailing

Create metric filter

Filter events - press enter to search

Clear 1m 30m 1h 12h Custom

UTC timezone

Display



▶	Timestamp	Message
		There are older events to load. Load more.
▶	2026-03-10T19:22:29.437Z	[ERROR] 2026-03-10T19:22:29.437Z f24da8c1-2741-5b14-af6c-caeb56d6f8a3 OSError e=timed out hostname=13.58.77.49 port=22 username=ubun...
▶	2026-03-10T19:22:29.437Z	2026-03-10 19:22:29,437 ERROR [e11.e11core e11ssh.py:57 __init__] OSError e=timed out hostname=13.58.77.49 port=22 username=ubuntu
▶	2026-03-10T19:22:29.437Z	2026-03-10 19:22:29,437 INFO [e11.home api.py:336 api_grader] summary={'lab': 'lab4', 'passes': [], 'fails': ['cannot connect by ssh...
▶	2026-03-10T19:22:29.437Z	[INFO] 2026-03-10T19:22:29.437Z f24da8c1-2741-5b14-af6c-caeb56d6f8a3 summary={'lab': 'lab4', 'passes': [], 'fails': ['cannot connect...
▶	2026-03-10T19:22:29.460Z	2026-03-10 19:22:29,460 INFO [e11 e11_common.py:410 add_grade] add_grade to dynamodb.Table(name='e11-users') user=user_id='0aee15c1-...
▶	2026-03-10T19:22:29.460Z	[INFO] 2026-03-10T19:22:29.460Z f24da8c1-2741-5b14-af6c-caeb56d6f8a3 add_grade to dynamodb.Table(name='e11-users') user=user_id='0aee...
▶	2026-03-10T19:22:29.586Z	[INFO] 2026-03-10T19:22:29.586Z f24da8c1-2741-5b14-af6c-caeb56d6f8a3 send_email2 to=['sgarfinkel@fas.harvard.edu', 'simsong@gmail.co...
▶	2026-03-10T19:22:29.586Z	2026-03-10 19:22:29,586 INFO [e11.home home.py:640 lambda_handler] send_email2 to=['sgarfinkel@fas.harvard.edu', 'simsong@gmail.com'] ...
▶	2026-03-10T19:22:29.588Z	END RequestId: f24da8c1-2741-5b14-af6c-caeb56d6f8a3
▶	2026-03-10T19:22:29.588Z	REPORT RequestId: f24da8c1-2741-5b14-af6c-caeb56d6f8a3 Duration: 10210.41 ms Billed Duration: 10211 ms Memory Size: 256 MB Max Memor...
▶	2026-03-10T19:26:19.502Z	START RequestId: ca64e160-92a6-42e2-b35c-9faa32c355d4 Version: \$LATEST
▶	2026-03-10T19:26:19.503Z	[INFO] 2026-03-10T19:26:19.503Z ca64e160-92a6-42e2-b35c-9faa32c355d4 log_level=INFO req method='GET' path='/' action='None' source_i...
▶	2026-03-10T19:26:19.512Z	2026-03-10 19:26:19,503 INFO [e11.home home.py:640 lambda_handler] log_level=INFO req method='GET' path='/' action='None' source_ip=...
▶	2026-03-10T19:26:19.873Z	END RequestId: ca64e160-92a6-42e2-b35c-9faa32c355d4
▶	2026-03-10T19:26:19.874Z	REPORT RequestId: ca64e160-92a6-42e2-b35c-9faa32c355d4 Duration: 371.09 ms Billed Duration: 372 ms Memory Size: 256 MB Max Memory Us...

No more records within selected time range Auto retry paused. [Resume](#)

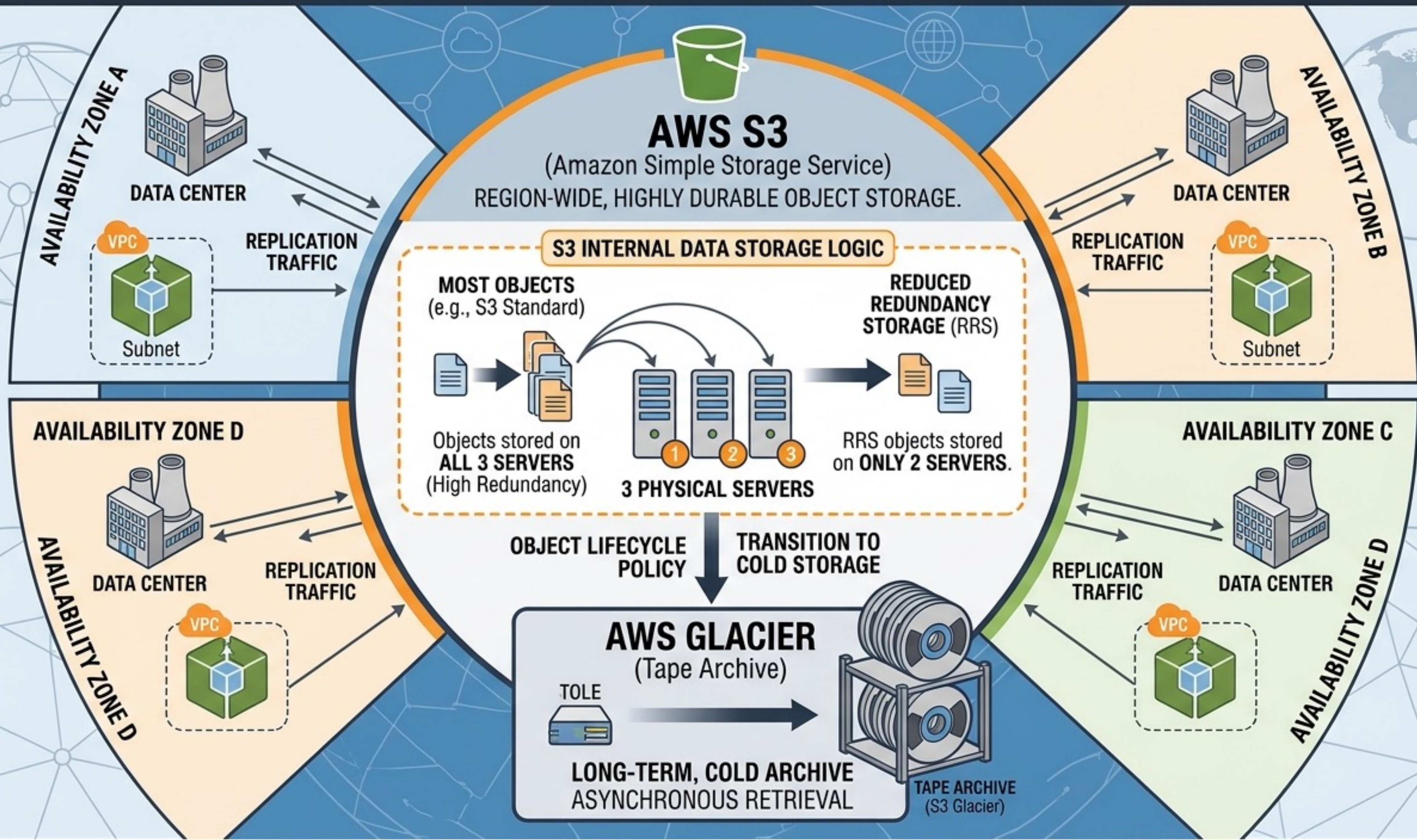
Lab 5 — Due March 30

Objectives for this lab:

- Learn about S3, Amazon's Simple Storage service. We'll use S3 to store images.
- Learn how we allow the JavaScript client to upload directly to S3 with a “signed post,” and how we allow it to download directly with a “signed get.”



AWS REGIONAL ARCHITECTURE: S3 DATA REDUNDANCY AND LIFECYCLE



The CAP Theorem (Brewer's theorem): Consistency, Availability, and Partition Tolerance: Pick Two!

Consistency

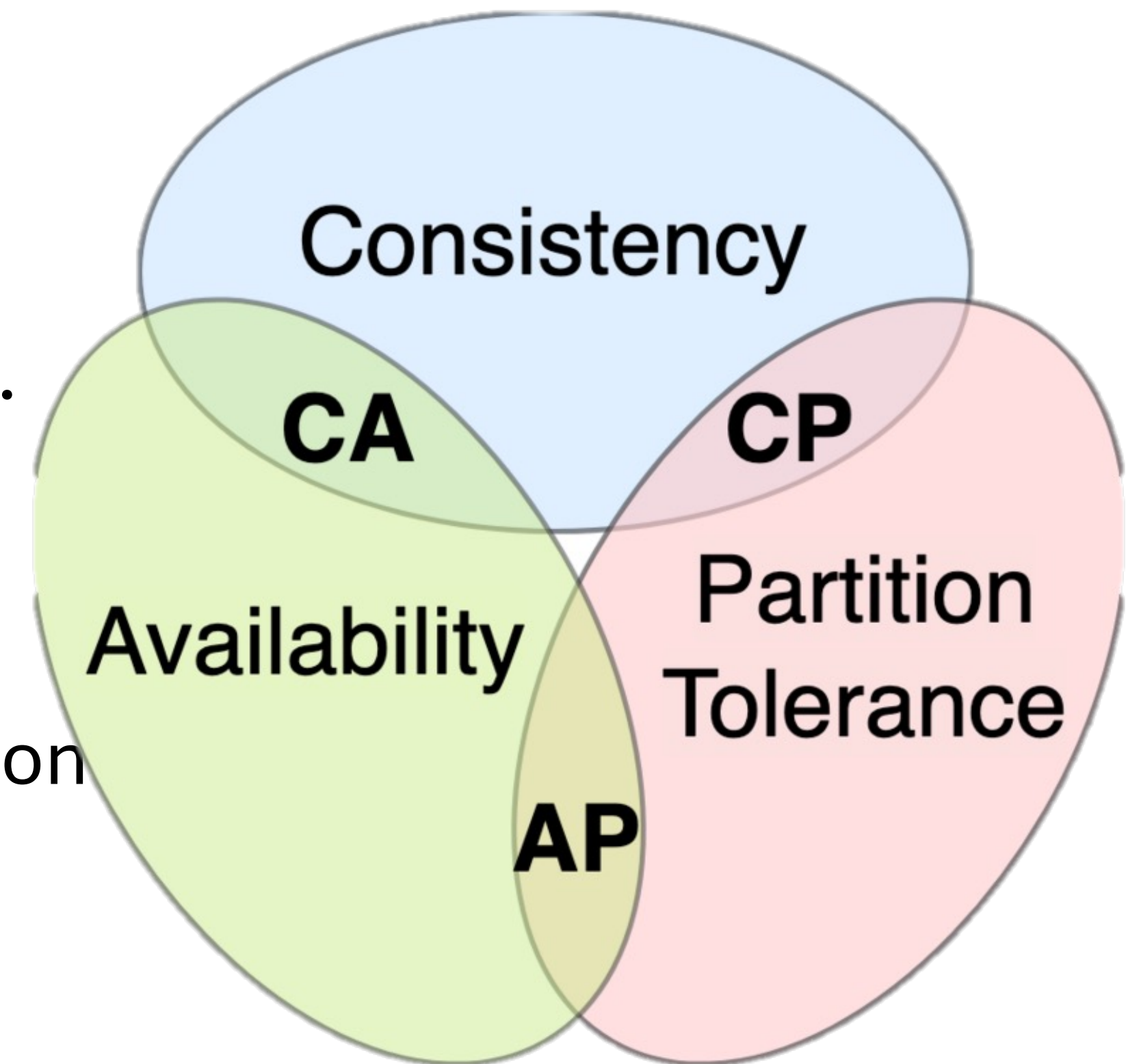
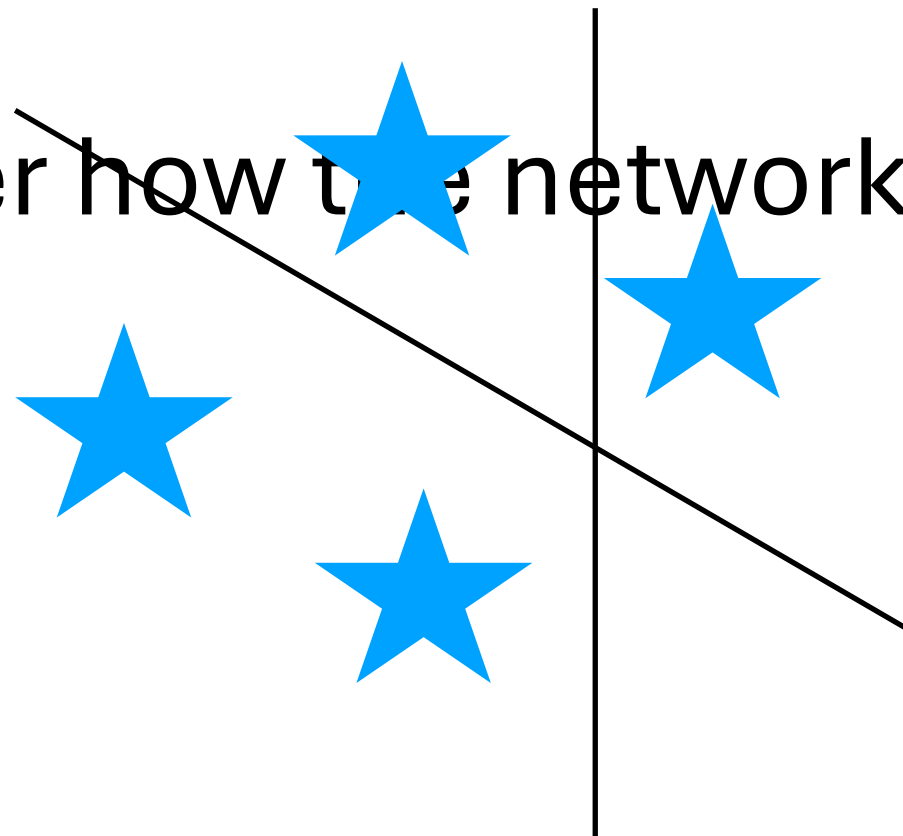
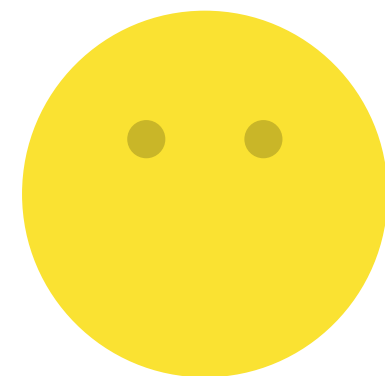
- Every write from every client has an error or success
 - Every read from every client receives the same data or the same error
- Every write must be replicated to every node

Availability

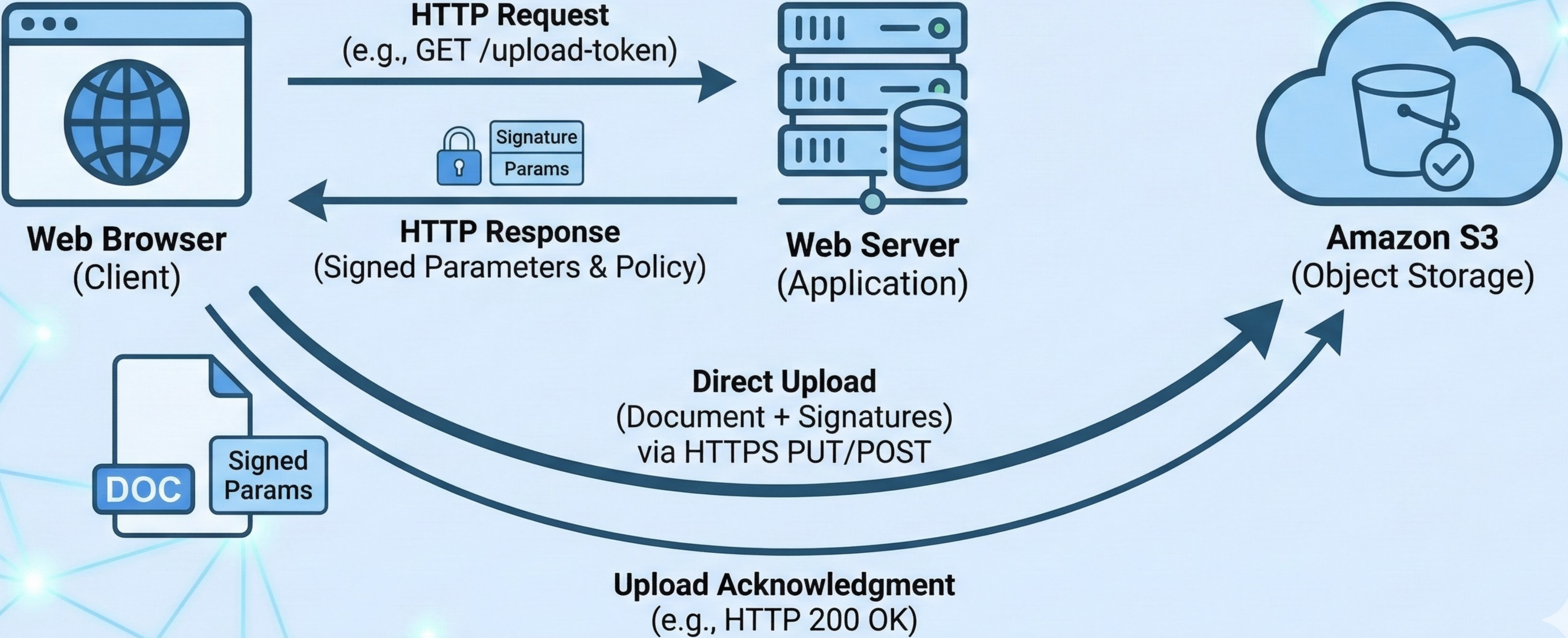
Every request to a non-failing node must result in a response.

Partition Tolerance

The system must operate no matter how the network is partitioned



Secure Direct-to-S3 Document Upload Workflow via Signed Parameters



Direct Upload (Document + Signatures) via HTTPS PUT/POST

HTTPS POST Request to Amazon S3

HTTPS Header

```
POST /<bucket-name> HTTP/1.1
Host: s3.amazonaws.com
Content-Type: multipart/form-data; boundary=-----WebKitFormBoundary...
Content-Length: <size>
Authorization: AWS4-HMAC-SHA256 Credential=..., SignedHeaders=..., Signature=...
x-amz-date: <timestamp>
```

HTTPS Body (Multipart/Form-Data)

```
-----WebKitFormBoundaryData-m-data-----
Content-Disposition: form-data; name="key"
<object-key>

-----WebKitFormBoundaryData-m-data-----
Content-Disposition: form-data; name="AWSAccessKeyId"
<access-key-id>

-----WebKitFormBoundaryData-m-data-----
Content-Disposition: form-data; name="policy"
<base64-encoded-policy>

-----WebKitFormBoundaryData-m-data-----
Content-Disposition: form-data; name="signature"
<calculated-signature>

-----WebKitFormBoundaryData-m-data-----
Content-Disposition: form-data; name="file"; filename="document.doc"
Content-Type: application/msword
<binary-file-data>

-----WebKitFormBoundaryData-m-data-----
```



Amazon S3
(Object Storage)



lab5/server/

```
0 Nov 10 14:54 __init__.py
192 Oct 28 06:25 __pycache__/
26 Nov 11 05:04 apikey.py@ -> ../../lab_common/apikey.py
22 Nov 11 05:04 db.py@ -> ../../lab_common/db.py
9156 Jan 13 17:02 image_controller.py
6380 Jan 13 21:18 image_validate.py
25 Nov 10 14:54 main.py@ -> ../../lab4/server/main.py
39 Nov 10 14:54 message_controller.py@ -> ../../lab4/server/message_controller.py
33 Nov 10 14:54 schema_lab4.sql@ -> ../../lab4/server/schema_lab4.sql
746 Dec 9 08:01 schema_lab5.sql
23 Nov 11 05:04 static@ -> ../../lab_common/static
26 Nov 11 05:04 templates@ -> ../../lab_common/templates
```


Making a presigned post

```
def make_presigned_post(s3_bucket,s3key):  
    """Return the S3 presigned_post fields"""  
    return s3_client.generate_presigned_post(  
        Bucket = s3_bucket,  
        Key = s3key,  
        Conditions = [  
            { "Content-Type": JPEG_MIME_TYPE }  
        ],  
        Fields = { "Content-Type": JPEG_MIME_TYPE},  
        ExpiresIn = 120, # in seconds  
    )
```


Making a presigned post

```
def make_presigned_post(s3_bucket,s3key):  
    """Return the S3 presigned_post fields"""  
    return s3_client.generate_presigned_post(  
        Bucket = s3_bucket,  
        Key = s3key,  
        Conditions = [  
            { "Content-Type": JPEG_MIME_TYPE },  
            [ "content-length-range", 1, MAX_IMAGE_SIZE_BYTES]  
        ],  
        Fields = { "Content-Type": JPEG_MIME_TYPE},  
        ExpiresIn = 120, # in seconds  
    )
```



```

/*
 * Uploads a image using a presigned post. See:
 * https://aws.amazon.com/blogs/compute/uploading-to-amazon-s3-directly-from-a-web-or-mobile-application/
 * https://boto3.amazonaws.com/v1/documentation/api/latest/guide/s3-presigned-urls.html
 *
 * Presigned post is provided by the /api/post-image call (see below)
 */
function upload_image_post(imageFile) {
  const formData = new FormData();
  formData.append("api_key", (document.querySelector("#api-key") || {}).value || "");
  formData.append("api_secret_key", (document.querySelector("#api-secret-key") || {}).value || "");
  formData.append("message", (document.querySelector("#message") || {}).value || "");
  formData.append("image_data_length", imageFile.size);

  setText("#status-message", "Requesting signed upload...");
  fetch("api/post-image", { method: "POST", body: formData })
    .then((r) => {
      if (!r.ok) {
        setText("#status-message", `Error: ${r.status} ${r.statusText}`);
        return Promise.reject(new Error(r.statusText));
      }
      return r.json();
    })
    .then((obj) => {
      console.log("api/post-image returned obj=", obj);
      if (obj.error) throw new Error(obj.error);

      setText("#status-message", `Uploading image ${obj.image_id}...`);
      const uploadFormData = new FormData();
      for (const [k, v] of Object.entries(obj.presigned_post.fields)) {
        uploadFormData.append(k, v);
      }
      uploadFormData.append("file", imageFile); // order matters

      const ctrl = new AbortController();
      const timeout = setTimeout(() => ctrl.abort(), UPLOAD_TIMEOUT_SECONDS * 1000);
      return fetch(obj.presigned_post.url, {
        method: "POST",
        body: uploadFormData,
        signal: ctrl.signal,
      }).finally(() => clearTimeout(timeout));
    })
    .then((uploadResponse) => {
      if (!uploadResponse.ok) {
        setText("#upload_message", `Error uploading image status=${uploadResponse.status} ${uploadResponse.statusText}`);
        return Promise.reject(new Error(uploadResponse.statusText));
      }
      setText("#status-message", "Image uploaded.");
      const imageEl = document.querySelector("#image-file");
      if (imageEl) imageEl.value = ""; // clear selected file
    })
}

```


This Week's Reading

- Mark Russinovich, *et al.* 2021. "Toward confidential cloud computing. CACM, June 2021, pp. 54–61.
<https://doi.org/10.1145/3453930>

Optional:

- Mark Russinovich. 2023. Confidential Computing: Elevating Cloud Security and Privacy. CACM, January 2024, pp. 52–53.
<https://doi.org/10.1145/3624577>
- UN Handbook on Privacy-Preserving Computation Techniques, BigData UN Global Working Group, <https://unstats.un.org/bigdata/task-teams/privacy/UN%20Handbook%20for%20Privacy-Preserving%20Techniques.pdf>

For Next Week:

- Enjoy your break!

For the Week After Break:

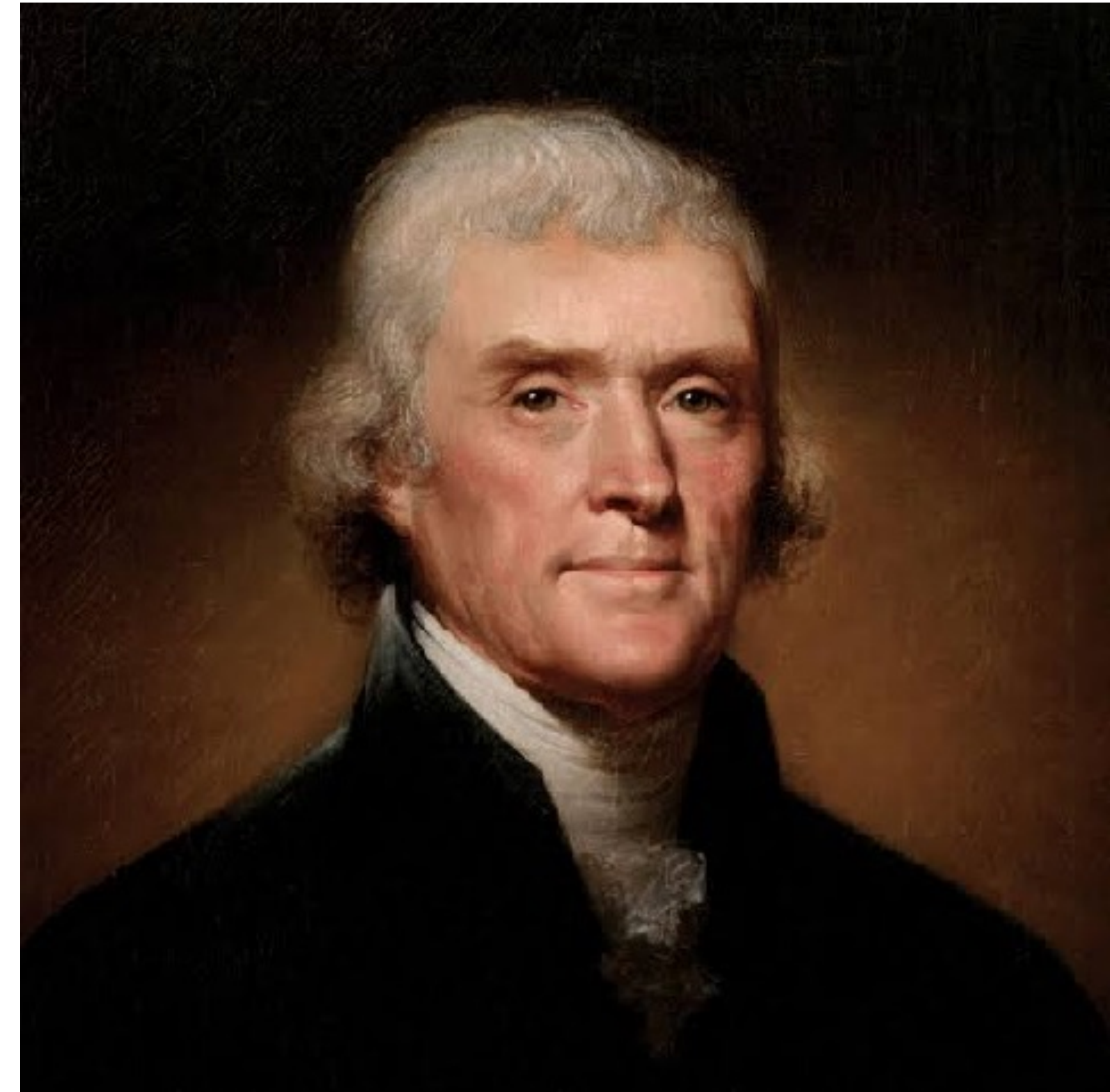
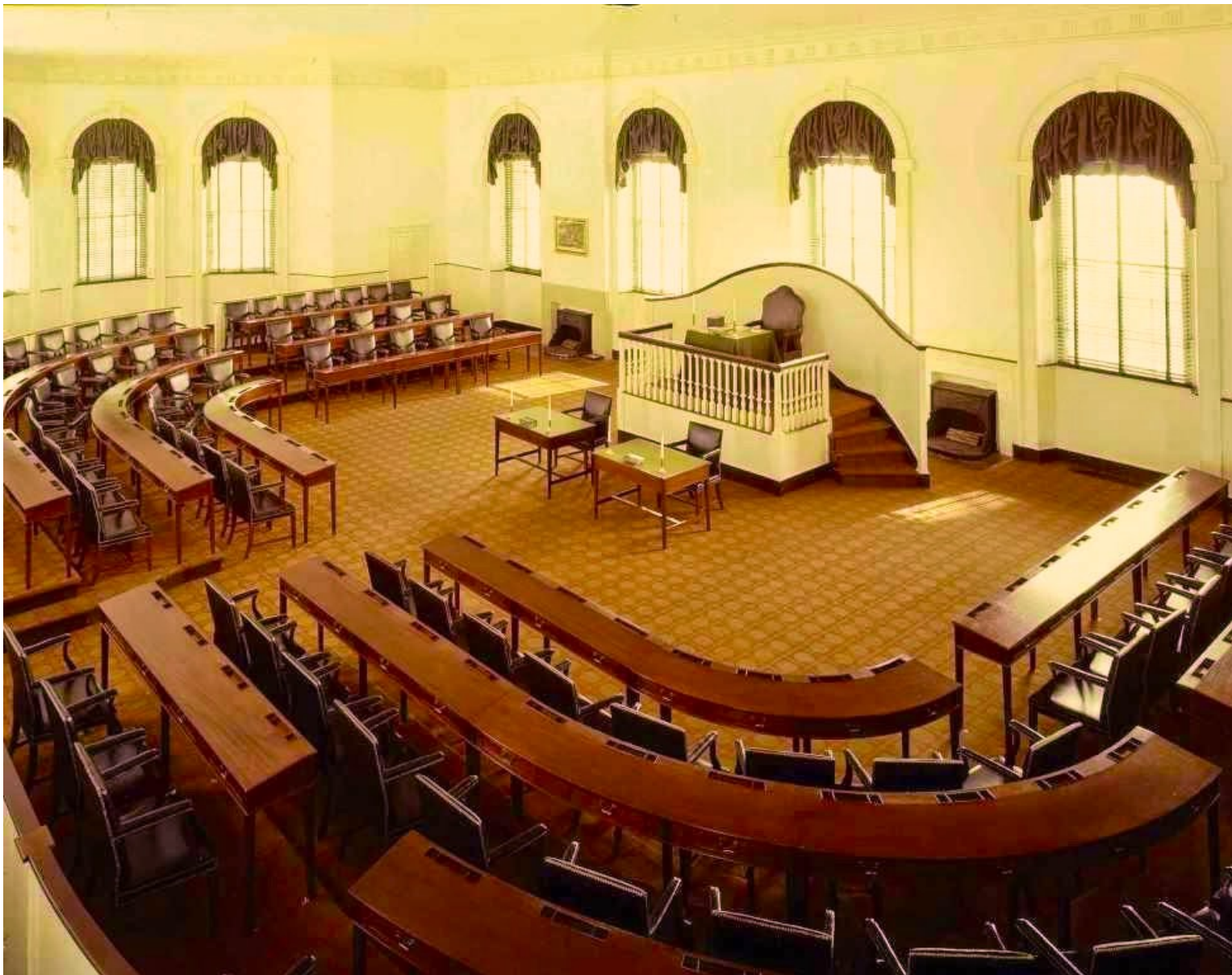
- Read Russinovich
 - Comment and respond on Perusall
 - Take the Reading Quiz on Canvas
- Work on Lab 5
 - Due March 30
 - DO NOT POSTPONE
- Participate in the Ed Discussion Board
- Enjoy the warm weather!

How we brought DP to the 2020 Census

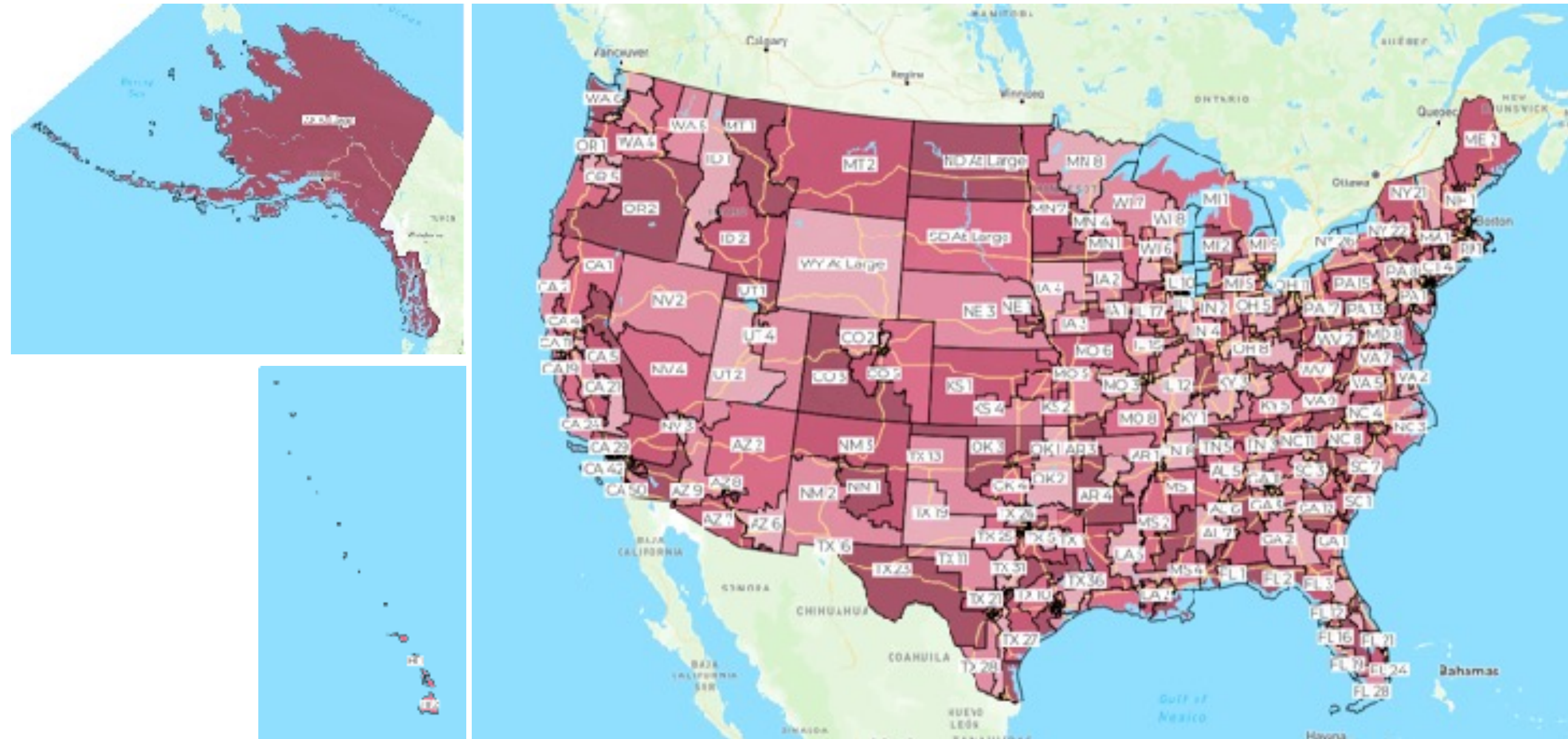
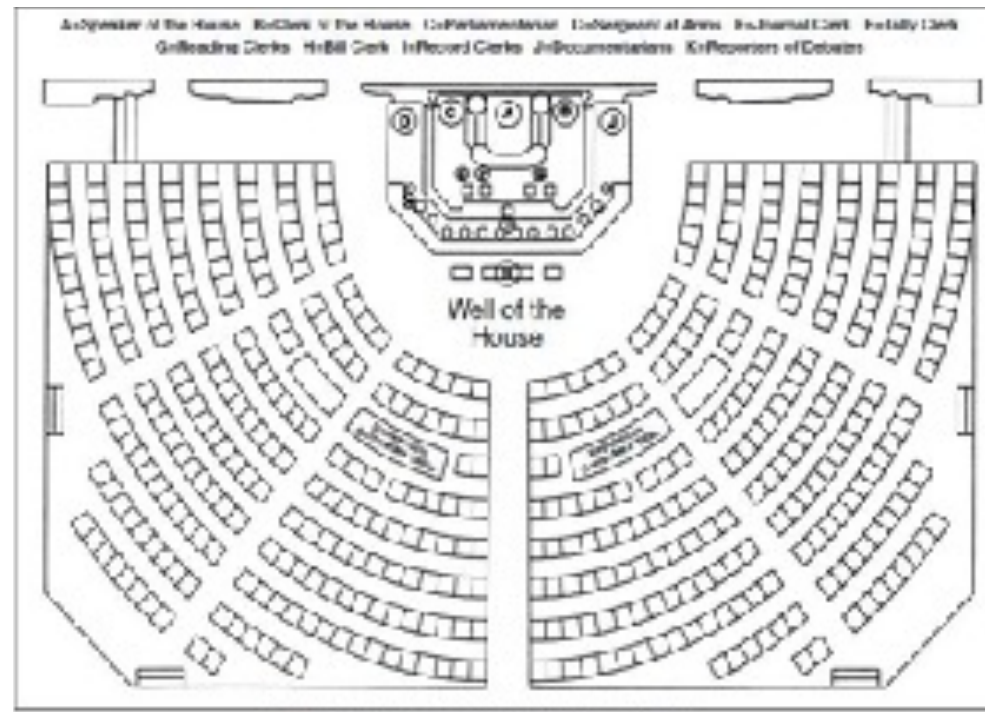
The US Census is the world's longest running statistical program.

First US Census:
1790

Purpose:
Apportion the US House of Representatives



Thomas Jefferson
Primary author, US Declaration of Independence
First US Secretary of State
First US Patent Commissioner (reviewed every patent)
Oversaw first US Census



Each congressional district elects a member to the House of Representatives.

There have been 435 seats since 1912

Each state elects 2 senators

The 2020 Census asked every household in the US to fill out a survey.



Counting everyone in your household can shape your future.



What else do I need to know?
Responding to the 2020 Census is:

> **Easy**

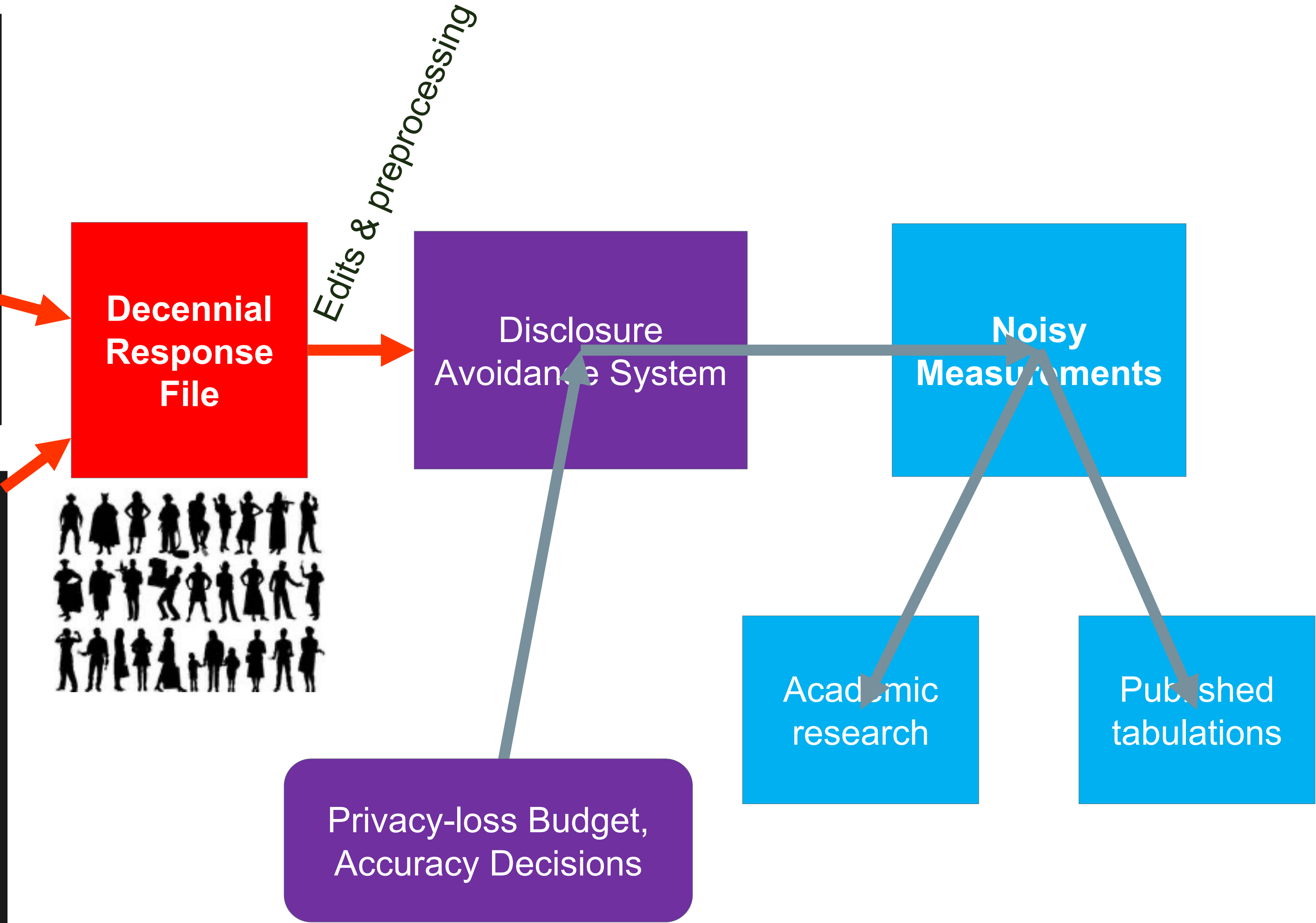
Every household in the United States will receive a notice to complete the census in early 2020. You can complete the form online, by phone, or by mail.

> **Safe**

Your personal information is confidential, is protected by law, and can never be used to identify you. It can never be shared with law enforcement agencies or your property manager.

> **Important**

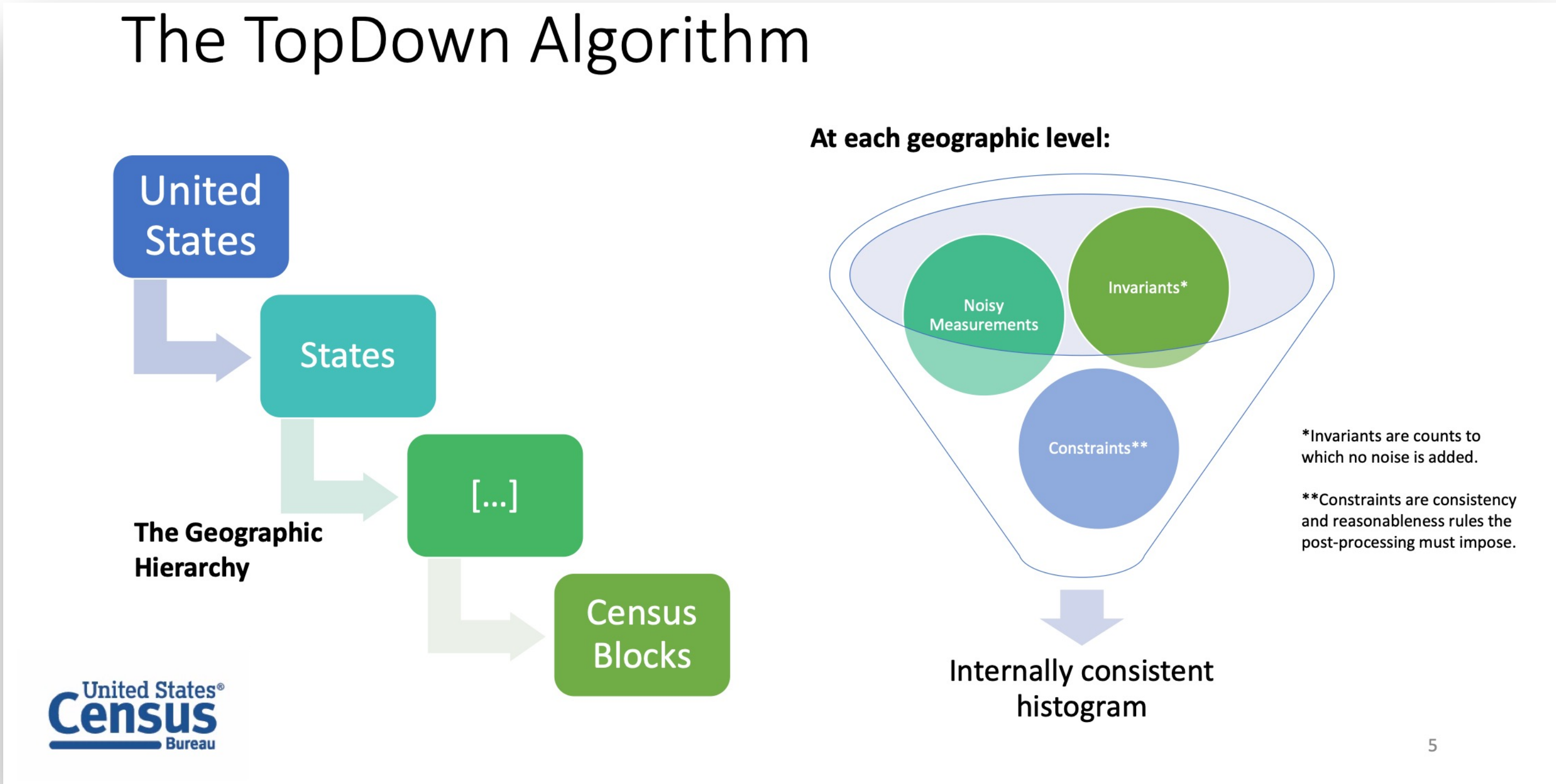
Businesses and leaders in your community will use the data collected in the census to make decisions about where to build new buildings, revitalize old ones, open stores, create jobs, and more.



This was implemented using the TopDown Algorithm

Statistics computed at each geographic layer.

Accurate statistics for each layer determined by higher layer.



<https://www2.census.gov/about/training-workshops/2023/2023-06-15-nmf-presentation.pdf>

The Census Bureau ultimately released *multiple data products* for the 2020 census.

“Noisy Measurement Files”

Raw data with noise added.

- Fractional numbers*
- Negative numbers*

Four sets of NMFs:

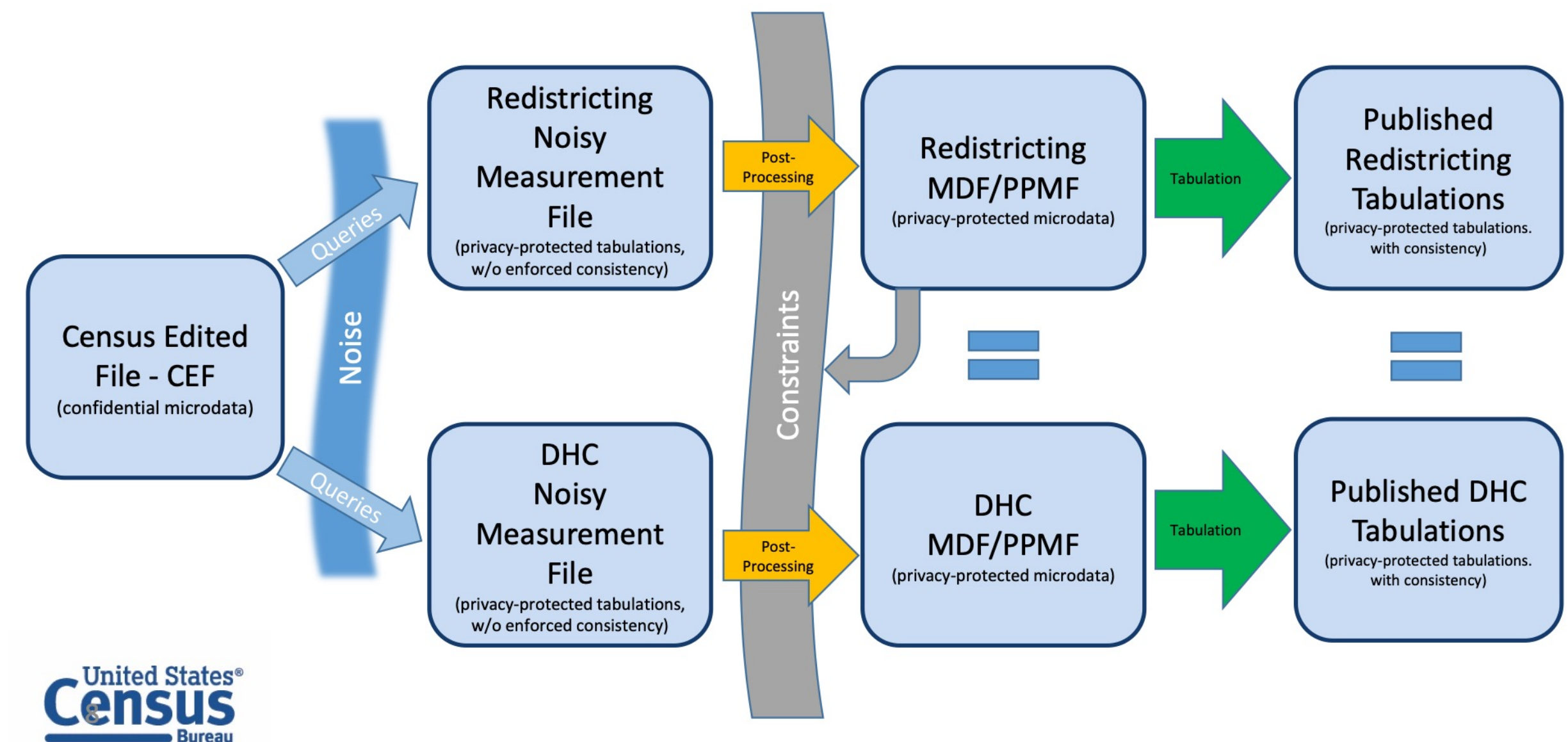
Block-by-block Redistricting files

- PL-94-171 NMF People*
- PL-94-171 NMF Housing Units*

Larger geographies with more demographic attributes.

- DHC NMF People*
- DHC NMF Housing Units*

Noisy Measurement Files (NMFs), Privacy-Protected Microdata Files (PPMFs), Published Tabulations



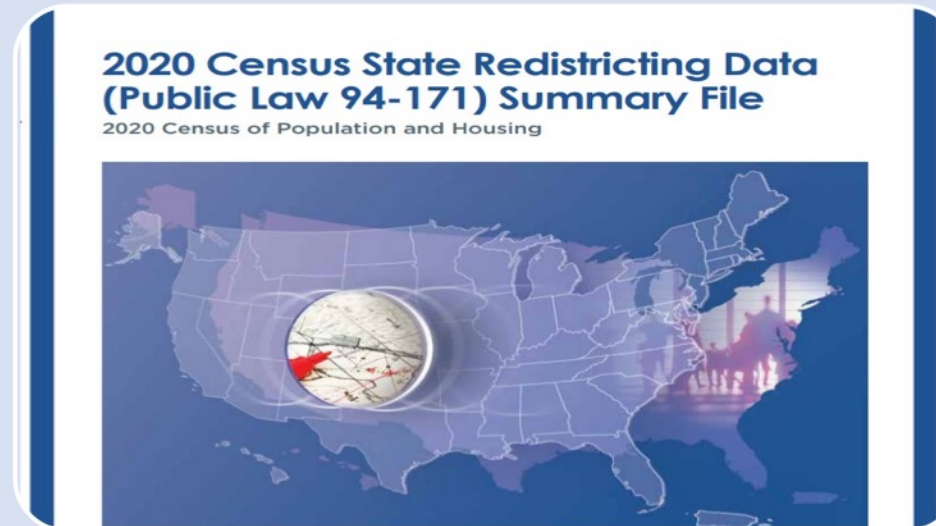
<https://www.census.gov/data/academy/webinars/2023/noisy-measurement-files.html>

Privacy Protected Microdata Files

Published tabulations

Multiple data products creates confusion...

Should I Use the NMF, the PPMF, or the Tabulations?



2020 Census Redistricting and DHC Tabulations

- Official 2020 Census Statistics
- Higher Accuracy (feature of TDA)
- Does include bias due to post-processing



2020 Census PPMF

- 100% microdata file
- Consistent with published tabulations
- Useful for special tabulations and microdata analysis



2020 Census NMF

- Can be used to produce unbiased estimates and confidence intervals
- Can be used to evaluate alternate post-processing mechanisms
- Research product

But this wasn't the real problem

The Census 2020 DP Backlash

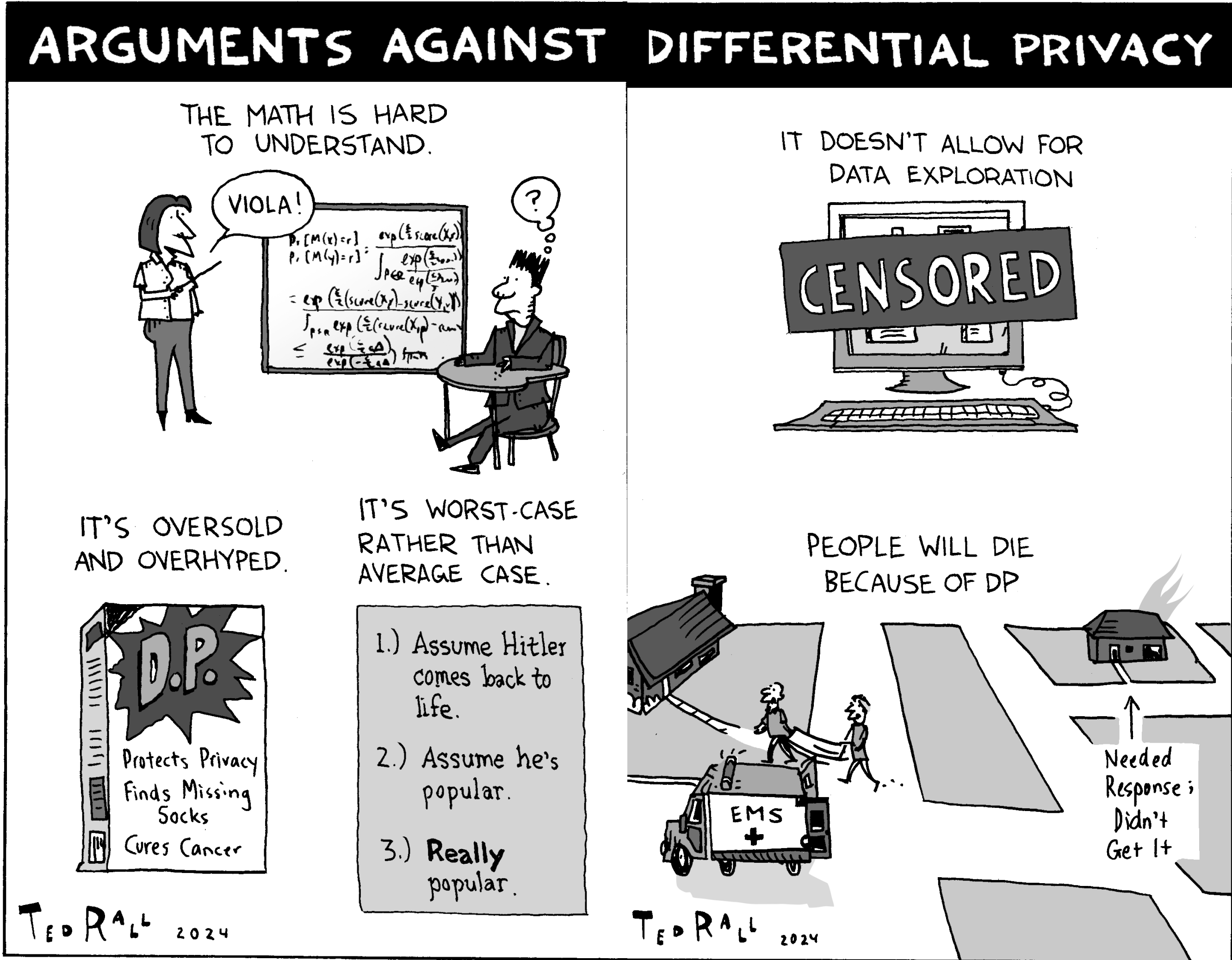
Differential privacy was not widely known or understood.

Many data users wanted highly accurate data reports on small areas.

- Some were anxious about the intentional addition of noise.
- Some were concerned that previous studies done with swapped data might not be replicated if they used DP data.

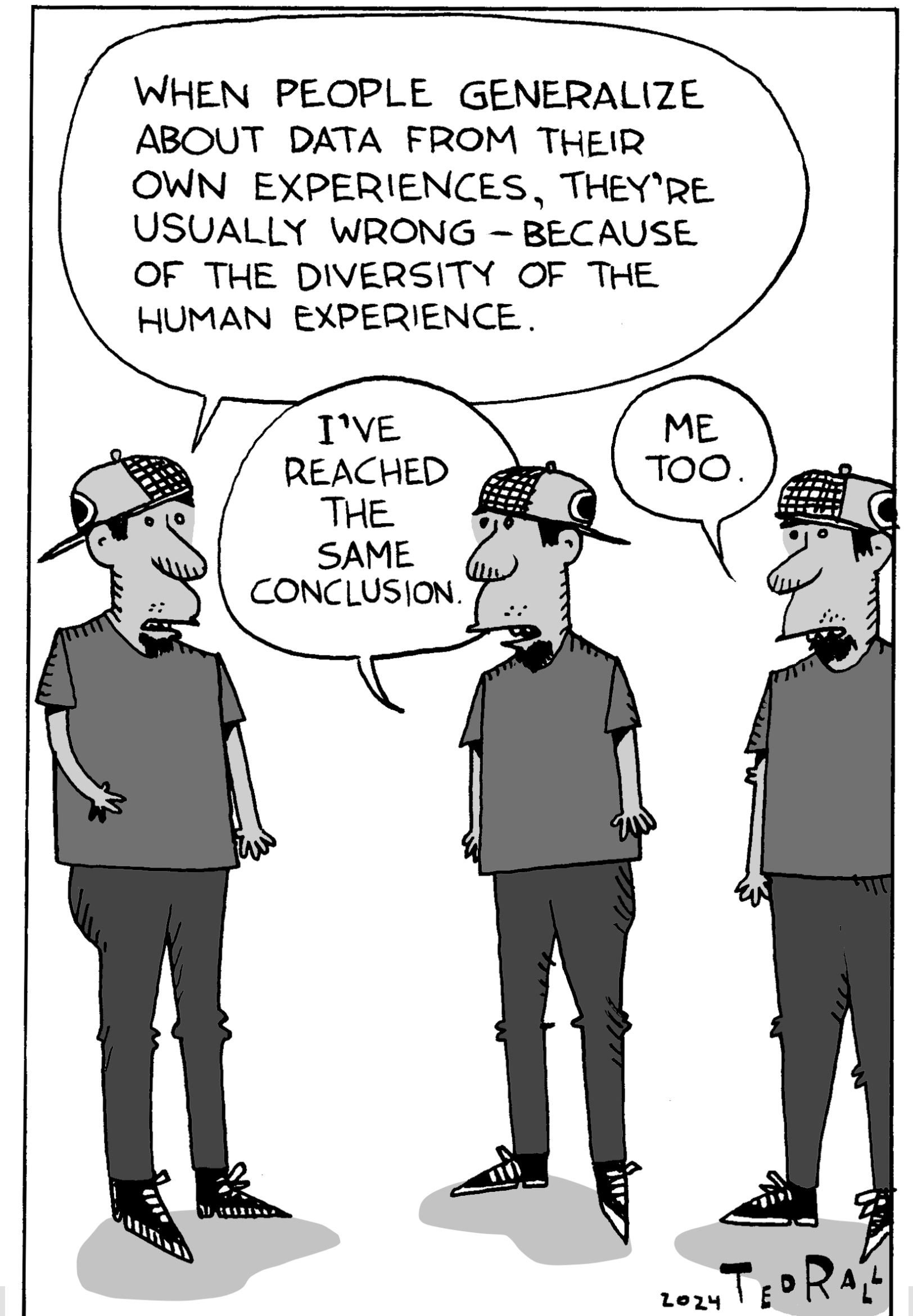
Many data users believed they require access to Public Use Microdata.

Users in 2000 and 2010 didn't know the error introduced by swapping and other protections applied to the tables and PUMS.



There is a lot more diversity in the US population than people realize.

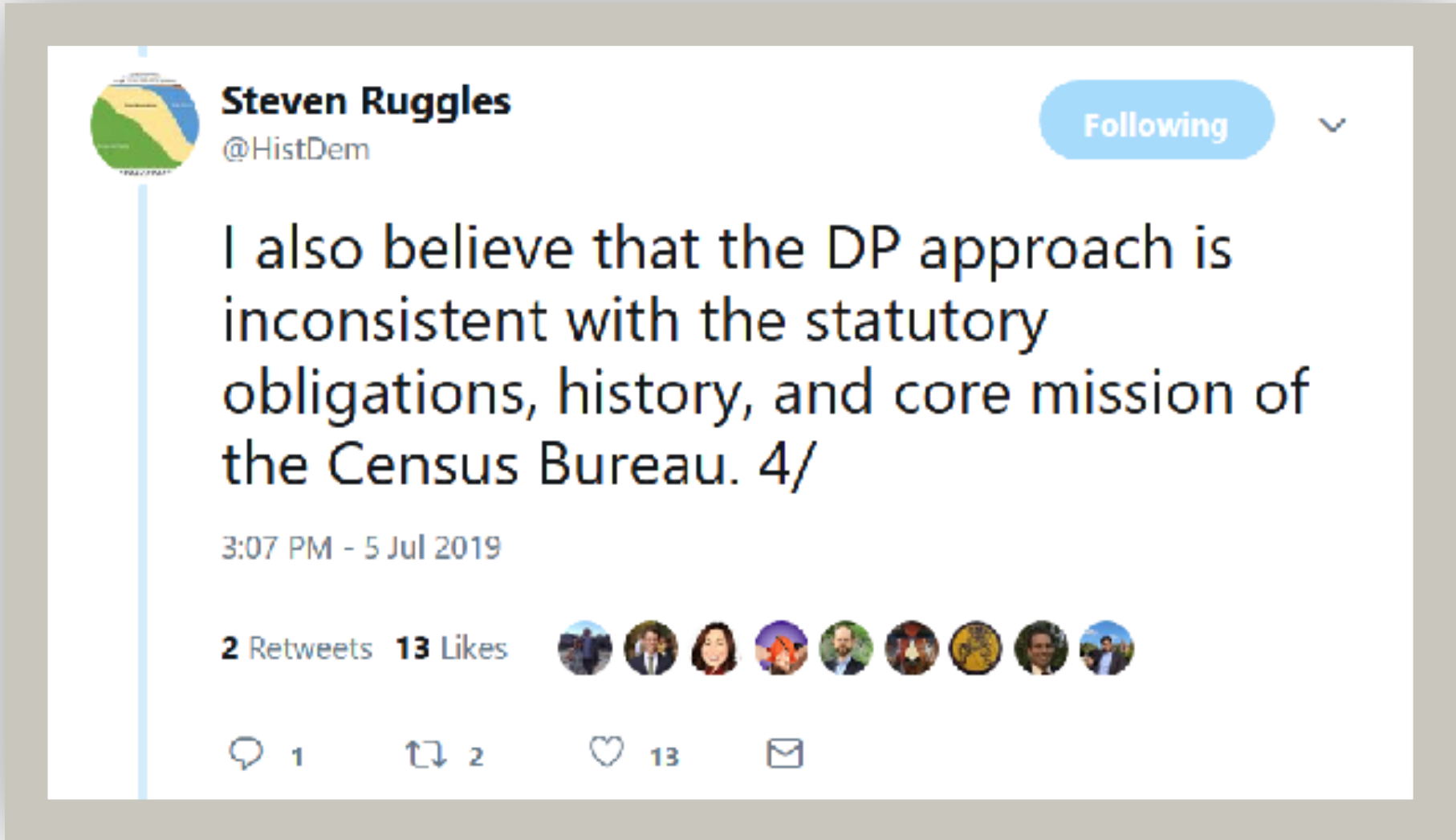
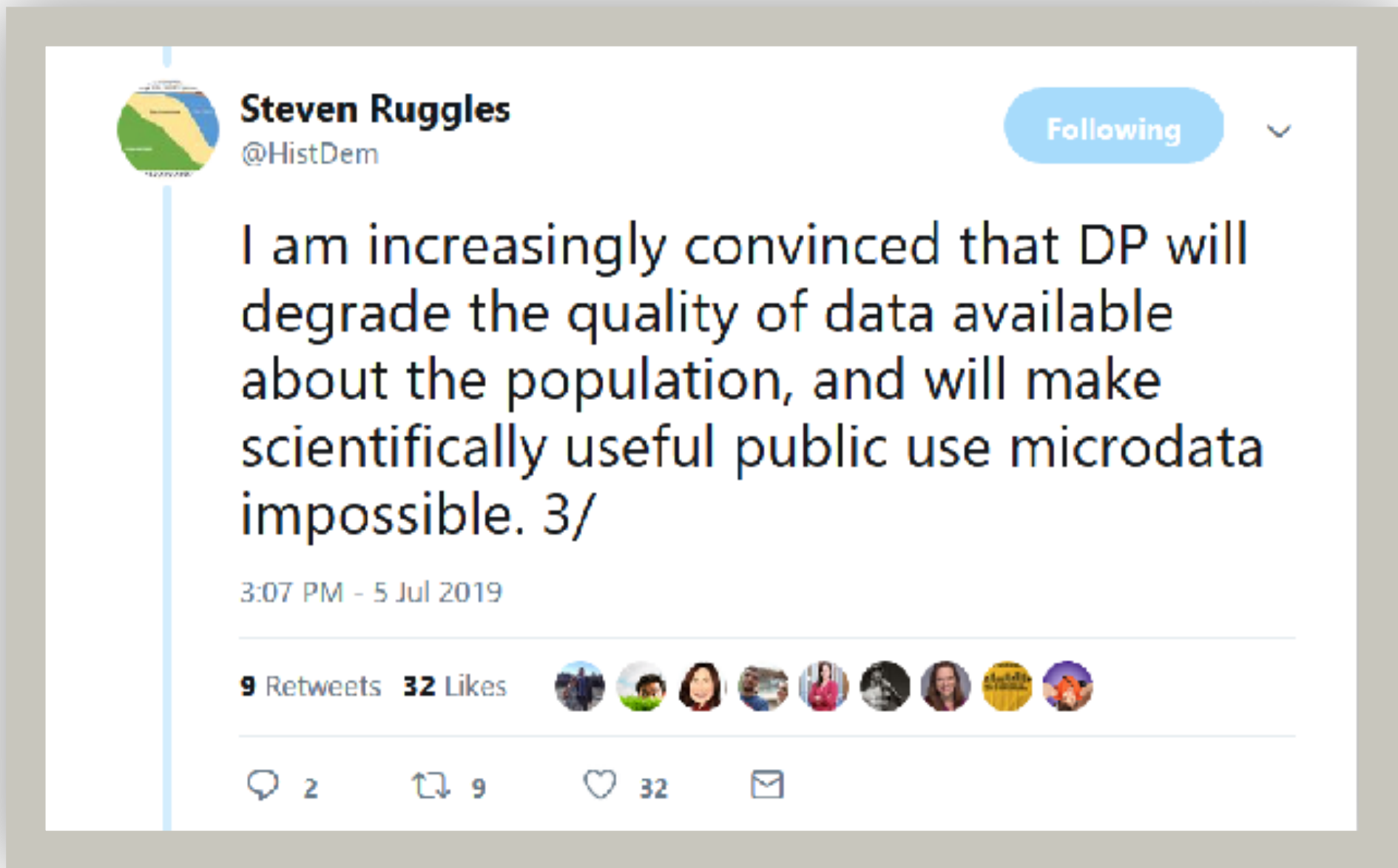
“About 57 percent of the 2010 Census population were ‘unique’ at the smallest census geography, block level, meaning they were the only people in their block with a specific combination of sex, age (in years), race (any of the 63 possible Office of Management and Budget race combinations), and Hispanic/Latino ethnicity” (McKenna 2018).



Dr. Steven Ruggles voiced concerns about DP in the 2020 Census.

Ruggles made his career distributing public use data.

IPUMS — Integrated Public Use Microdata Series, is the world's largest level population data.



Steven Ruggles 5 Jul 2019

Ruggles organized a petition with 4000+ signers asking for no DP in 2020 Census.
Ultimately DP was used, but delayed for the American Community Survey.

Errors in the 2020 Census were blamed on DP.

An article in The New York Times stated that DP was responsible for allocating 13 adults and one child to Census Block 1002 in downtown Chicago, a block that “consists entirely of a 700-foot bend in the Chicago River” (Wines 2022).

In fact, the TopDown algorithm implemented a constraint such that “the number of householders (person one on the questionnaire) cannot be greater than the number of housing units” (J. Abowd et al. 2022).

- Likely answer: Error in geography file
- Unlikely answer: house boat

