



# Data Privacy

## L01 - Data Privacy and Regulations



# Defining Privacy



- Wikipedia → “ability of an individual or group to **seclude themselves or information about themselves**, and thereby express themselves selectively”
- Merriam-Webster → “freedom from **unauthorized intrusion**” , “the quality or state of **being apart from company or observation**”

# Defining Privacy

- Private information sometimes equated to **personal data / personally identifiable information (PII)**
  - “(1) any **information that can be used to distinguish or trace an individual's identity**, such as name, social security number, date and place of birth, mother's maiden name, or biometric records;
  - (2) any other **information that is linked or linkable to an individual**, such as medical, educational, financial, and employment information.” [1]

**What should be considered private?**

# Privacy Principles

- **Minimization**: what you collect, who has access to it, and how long you keep it
- **Choice**: user's option to share information
- **Access**: user's right to review, correct and possibly delete the data you hold
- **Transparency**: disclosure of what, how and with whom data will be shared



# Privacy vs Security

- While privacy and security go hand in hand, they are not the same concepts
- Security is about protecting both the information you choose to share and that you choose not to share from getting into the hands of others



***Can you have security without privacy?***

***What about privacy without security?***



DuckDuckGo ✓  
@DuckDuckGo



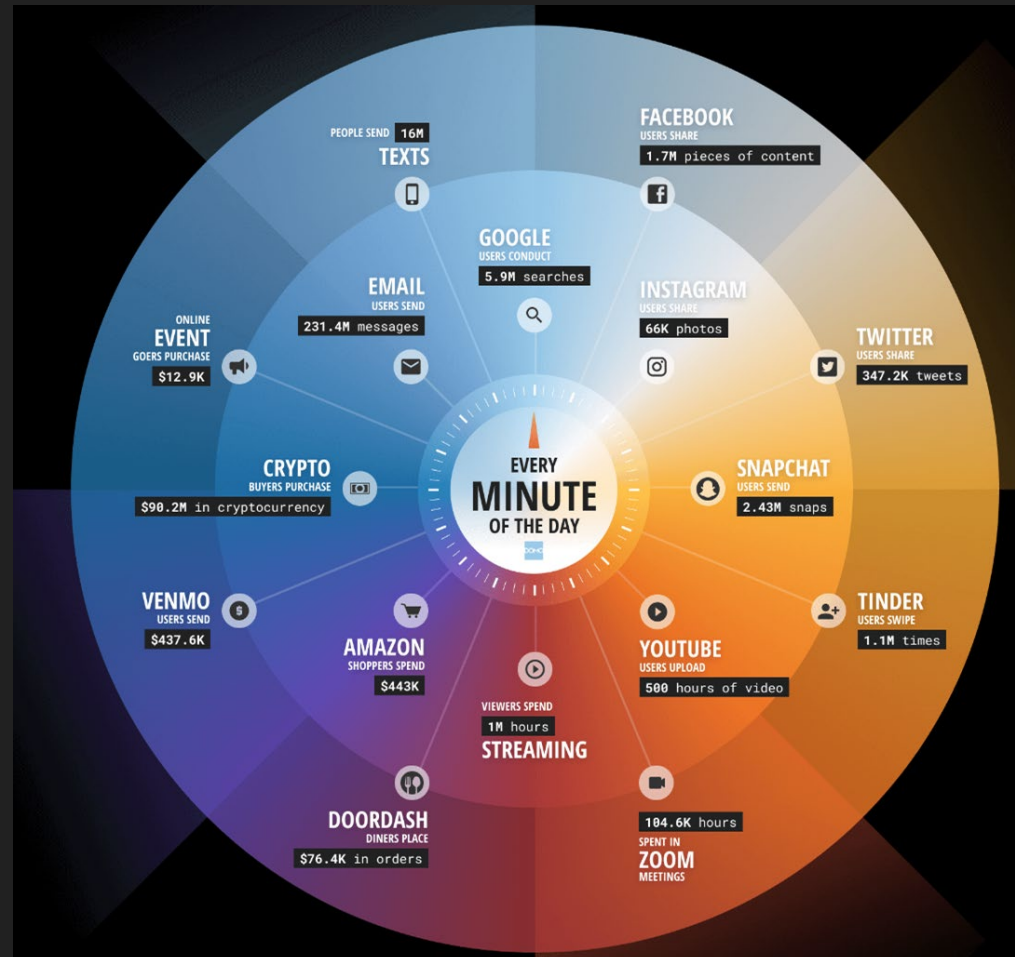
Replying to @DuckDuckGo

Security without privacy is like a house made of bullet-proof glass. No-one's getting inside but your personal life is still on display.

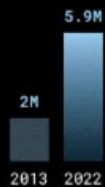
11:16 AM · Jul 26, 2017 · Twitter Web Client



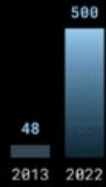
# Today's Privacy Challenge



# DATA NEVER SLEEPS 1.0 VS. 10.0



GOOGLE  
USER QUERIES



YOUTUBE  
HOURS UPLOADED



INSTAGRAM  
PHOTOS SHARED



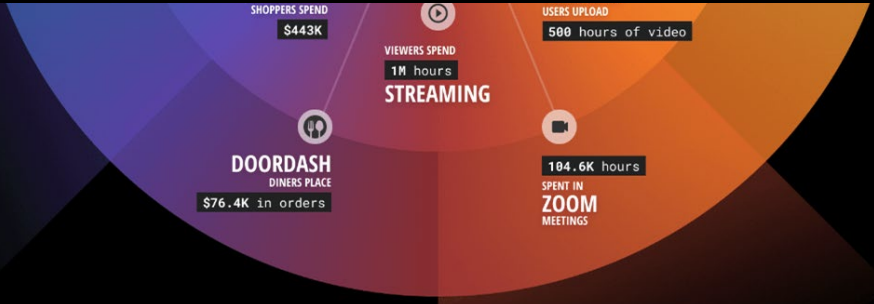
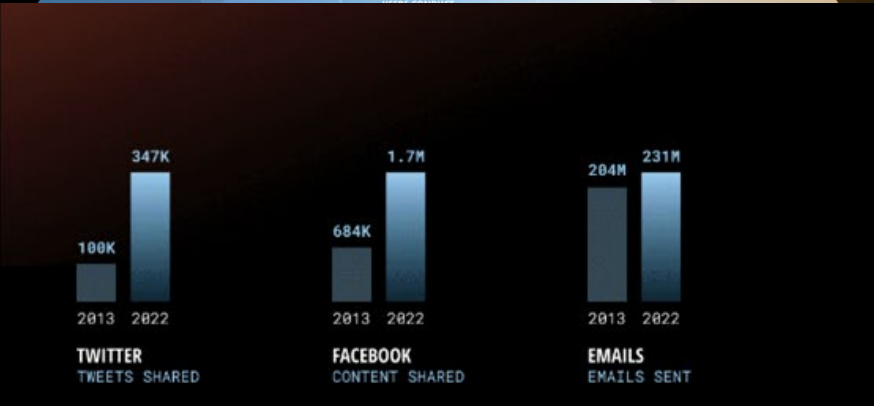
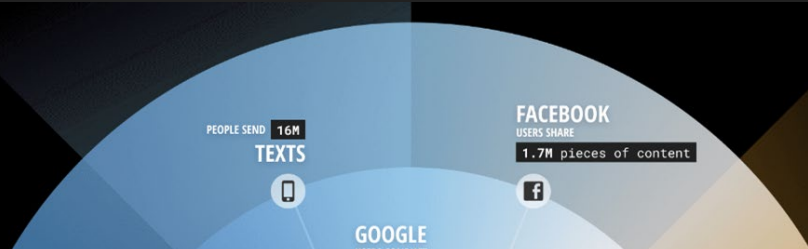
TWITTER  
TWEETS SHARED



FACEBOOK  
CONTENT SHARED



EMAILS  
EMAILS SENT





| #  | Company           | Image recognition |                         |                     |
|----|-------------------|-------------------|-------------------------|---------------------|
|    |                   | Face recognition  | Environment recognition | Product recognition |
| 1  | Facebook          | •                 | •                       | •                   |
| 2  | Instagram         | •                 | •                       | •                   |
| 3  | Tinder            | •                 | •                       |                     |
| 4  | Grindr            | •                 | •                       |                     |
| 5  | Uber              | •                 |                         |                     |
| 6  | TikTok            | •                 | •                       | •                   |
| 7  | Strava            |                   |                         |                     |
| 8  | Spotify           |                   |                         |                     |
| 9  | Myfitnesspal      |                   |                         |                     |
| 10 | Clubhouse         | •                 |                         |                     |
| 11 | Credit Karma      |                   |                         |                     |
| 12 | Twitter           | •                 | •                       |                     |
| 13 | Airbnb            | •                 | •                       |                     |
| 14 | Lidl Plus         |                   |                         | •                   |
| 15 | American Airlines |                   |                         |                     |
| 16 | eBay              |                   |                         | •                   |
| 17 | Sleepcycle        |                   |                         |                     |
| 18 | Paypal            |                   |                         |                     |
| 19 | Slimming World    |                   |                         |                     |
| 20 | Whatsapp          |                   |                         |                     |

# What can companies tell from image recognition?

The personal data that recognition software helps companies collect from you



**Facial recognition**  
Recognises people and their key attributes



**Background recognition**  
Detects elements in shot, establishes environment



**Object recognition**  
Can identify an object or product within an image

| # | Company   | Face recognition | Environment recognition | Product recognition | Your contacts | Voice data/ recognition | Access to image library | Languages |
|---|-----------|------------------|-------------------------|---------------------|---------------|-------------------------|-------------------------|-----------|
| 1 | Facebook  | •                | •                       | •                   | •             | •                       | •                       | •         |
| 2 | Instagram | •                | •                       | •                   | •             | •                       | •                       | •         |
| 3 | TikTok    | •                | •                       | •                   | •             | •                       | •                       | •         |
| 4 | Twitter   | •                | •                       | •                   | •             | •                       | •                       | •         |
| 5 | Tinder    | •                | •                       |                     | •             |                         | •                       | •         |

Find the full report at [clario.co/blog/which-company-uses-most-data](https://clario.co/blog/which-company-uses-most-data)

| #  | Company           | Face recognition |
|----|-------------------|------------------|
| 1  | Facebook          | •                |
| 2  | Instagram         | •                |
| 3  | Tinder            | •                |
| 4  | Grindr            | •                |
| 5  | Uber              | •                |
| 6  | TikTok            | •                |
| 7  | Strava            | •                |
| 8  | Spotify           | •                |
| 9  | Myfitnesspal      | •                |
| 10 | Clubhouse         | •                |
| 11 | Credit Karma      | •                |
| 12 | Twitter           | •                |
| 13 | Airbnb            | •                |
| 14 | Lidl Plus         | •                |
| 15 | American Airlines | •                |
| 16 | eBay              | •                |
| 17 | Sleepcycle        | •                |
| 18 | Paypal            | •                |
| 19 | Slimming World    | •                |
| 20 | Whatsapp          | •                |

## How Companies Use Location Data

Here are just a few ways companies manipulate your location data.



### Income level

Determine your approximate disposable income based on neighborhood demographic data.



### Current location

Send real-time notifications to prompt you toward nearby shops or restaurants.



### School or workplace

Send restaurant push notifications right before your usual break times.



### Shopping habits

Send coupons or promo codes for stores you frequent.

## What can companies tell you about your location data?



Location recognition software can tell you



Location recognition software can tell you about your environment



Object recognition software can identify an object or product within an image

| on | Your contacts | Voice data/recognition | Access to image library | Languages |
|----|---------------|------------------------|-------------------------|-----------|
| •  | •             | •                      | •                       | •         |
| •  | •             | •                      | •                       | •         |
| •  | •             | •                      | •                       | •         |
| •  | •             | •                      | •                       | •         |

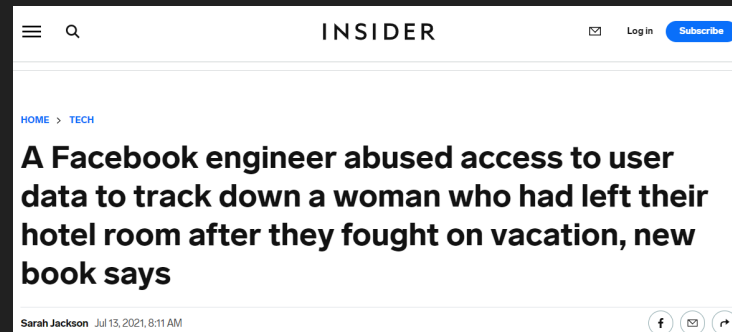
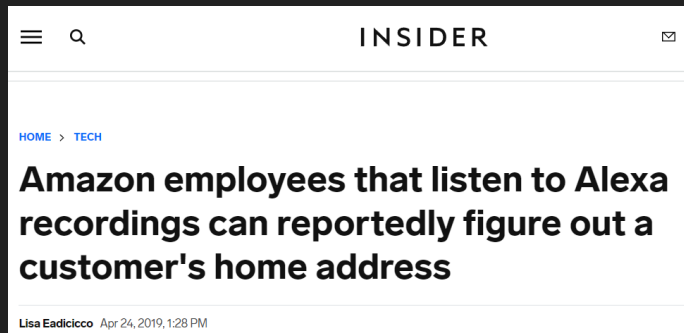
most-data

clario



...but, only algorithms access the data...

*Always?*

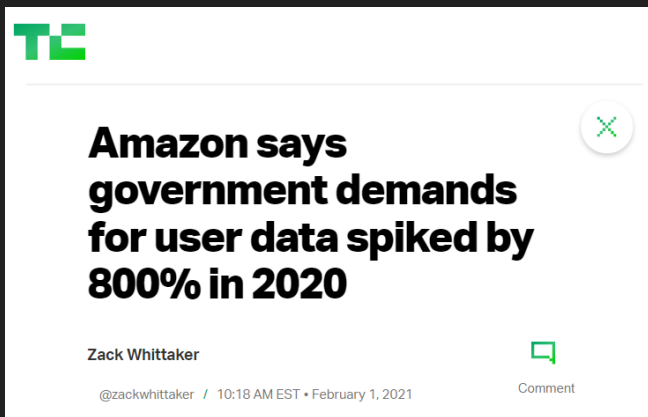


...but, only algorithms access the data...

Always?

...and they don't share data...

Never?



**TE**

**Amazon says government demands for user data spiked by 800% in 2020**

Zack Whittaker

@zackwhittaker / 10:18 AM EST • February 1, 2021

Comment



**THE VERGE** TECH ▾ REVIEWS ▾ SCIENCE ▾ ENTERTAINMENT ▾ MORE ≡ f t r

GOOGLE POLICY US & WORLD

## Google reportedly gave some users' data to Hong Kong authorities in 2020

*The company said last year it would stop responding to such requests*

By Kim Lyons | @SocialKimLy | Sep 11, 2021, 4:46pm EDT



**The New York Times**

## Cambridge Analytica and Facebook: The Scandal and the Fallout So Far

Revelations that digital consultants to the Trump campaign misused the data of millions of Facebook users set off a furor on both sides of the Atlantic. This is how The Times covered it.

...but, only algorithms access the data...

Always?

...and they don't share data...

Never?

...but they anonymize data before sharing!

*Yes (maybe if it's not a breach)...*

***Is anonymization enough?***

| Name       | Age | Condition |
|------------|-----|-----------|
| John Doe   | 59  | Cancer    |
| Mary Smith | 78  | Covid-19  |

# Case Study: The Netflix Prize



dannypeled.com

- 2007 -- Netflix competition to create the best collaborative filtering algorithm to predict what rating a user would give to a movie based on their previous ratings
- Netflix Prize had a huge positive impact in the research area on recommendations

# Case Study: The Netflix Prize

- Dataset: Ratings of almost 500K customers (more than 100M ratings in total)
- Netflix **anonymized** the dataset
  - From their FAQ “[...] *all customer identifying information has been removed; all that remains are ratings and dates.*”

*What could go wrong?*



# Case Study: The Netflix Prize

- Researchers from UT at Austin presented a **statistical de-anonymization attack** using IMDB as background knowledge [1]
- Able to **identify individual users** and determine potentially sensitive information (political views, religious views, or sexual orientation).
- Four Netflix users filled a class action lawsuit.
  - Netflix ended up reaching a settlement with the plaintiffs in 2010.
- Cancellation of projected sequel to the Netflix Prize

# Case Study: AOL Search Data [1]

- **Dataset:** 20M search queries for 650K users from 2006
- **Why released:** allow researchers to understand search patterns
- **How anonymized:** user identifiers removed
  - All searches from same user linked by an arbitrary identifier
- **Attacks:** many successful attacks identified individual users
  - Ego-surfers: people typed in their own names
  - Zip codes and town names identify an area
  - NY Times identified 4417749 as 62yr old GA widow [2]
- **Consequences:** CTO resigned, two researchers fired
  - Well-intentioned effort failed due to inadequate anonymization



[1] Slide extracted from Cormode, G., & Srivastava, D. (2009, June). Anonymized data: generation, models, usage. SIGMOD (pp. 1015-1018).

[2] <https://www.nytimes.com/2006/08/09/technology/09aol.html>

# Data Privacy Regulations



# A Long Time Ago

- The history of privacy rights and regulations starts long before the era of the Internet and Smartphones.
- The Supreme Court has found that the U.S. Constitution (1789) does provide for a right to privacy in its First, Third, Fourth, and Fifth amendments [1]
- 1890, Warren and Brandeis articulated the **Right to Privacy** [2] ("**right to be let alone**"): newspapers are the primary sources of "*the unwarranted invasion of individual privacy*" and urged that the courts "*protect the privacy of private life.*"

[1] <http://law2.umkc.edu/faculty/projects/ftrials/conlaw/rightofprivacy.html>

[2] L. Brandeis and S. Warren. *The right to privacy*. *Harvard law review*, 4(5), 1890.



*“[the Constitution] conferred, as against the Government, the **right to be let alone** - the most comprehensive of rights and the right most valued by civilized men.”*

Justice Brandeis, dissenting in *Olmstead v. US* (1928)

# Data Protection Regulations

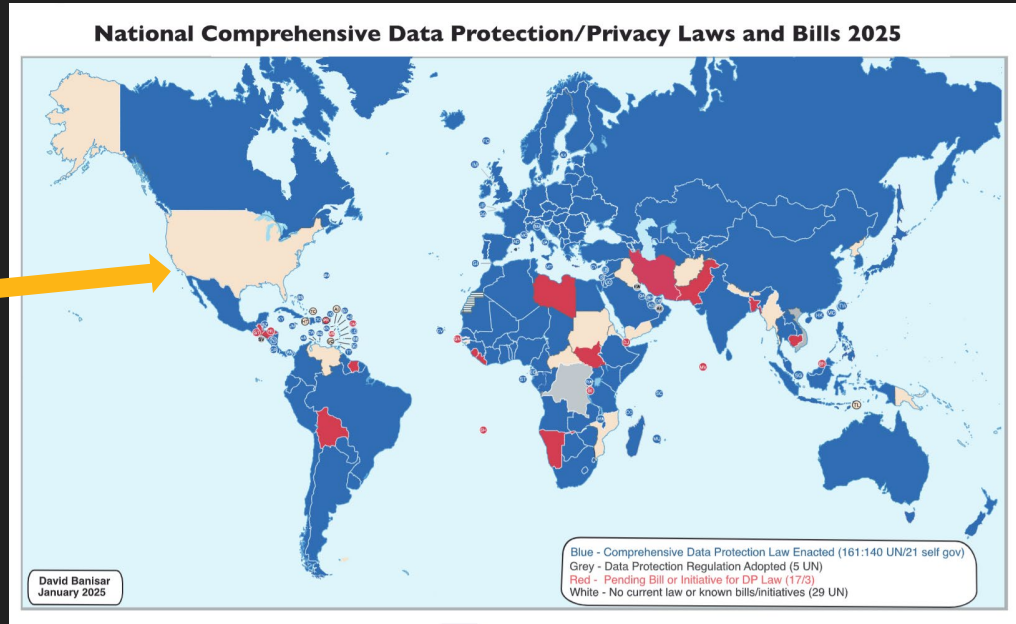
- Technological advances (e.g., computers) enabled governments and corporations to **capture and retain large amounts of individuals data** since the 1950s.
- Data protection/privacy regulations have been enacted.
- Federal state of Hesse (Germany) passed the first data protection law in the world in 1970 [1].

# Regulations Today (January 2025)

- Over 166 countries, self-governing jurisdictions, territories have adopted national laws and almost 30 countries and jurisdictions have pending bills or initiatives [1]

20 states:

California, Colorado, Connecticut, Delaware, Florida,\* Indiana, Iowa, Kentucky, Maryland, Minnesota, Montana, Nebraska, New Hampshire, New Jersey, Oregon, Rhode Island, Tennessee, Texas, Utah, and Virginia

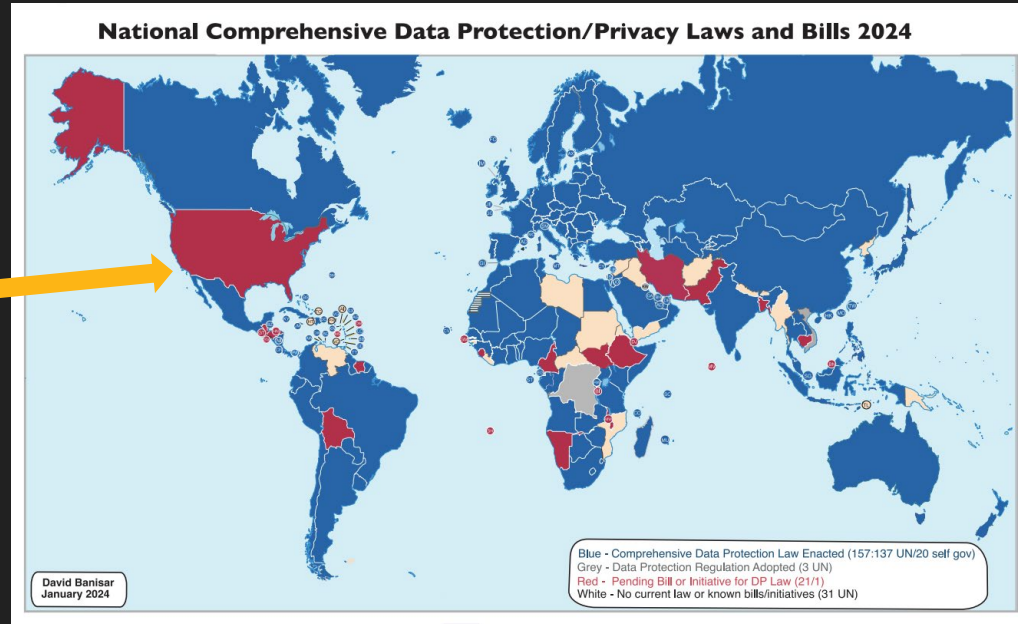


[1] National Comprehensive Data Protection/Privacy Laws and Bills 2024, D Banisar - Privacy Laws and Bills, 2025

# Regulations Today (January 2024)

- Over 160 countries, self-governing jurisdictions, territories have adopted national laws and almost 30 countries and jurisdictions have pending bills or initiatives [1]

CCPA in  
California

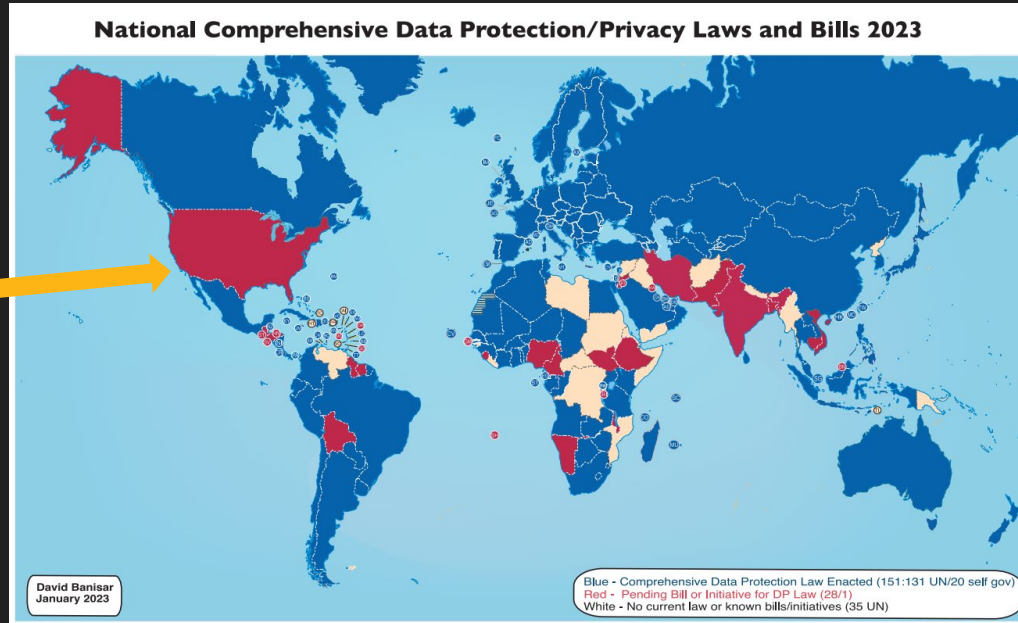




# Regulations Today (January 2023)

- Over 150 countries, self-governing jurisdictions, territories have adopted national laws and almost 30 countries and jurisdictions have pending bills or initiatives [1]

CCPA in California

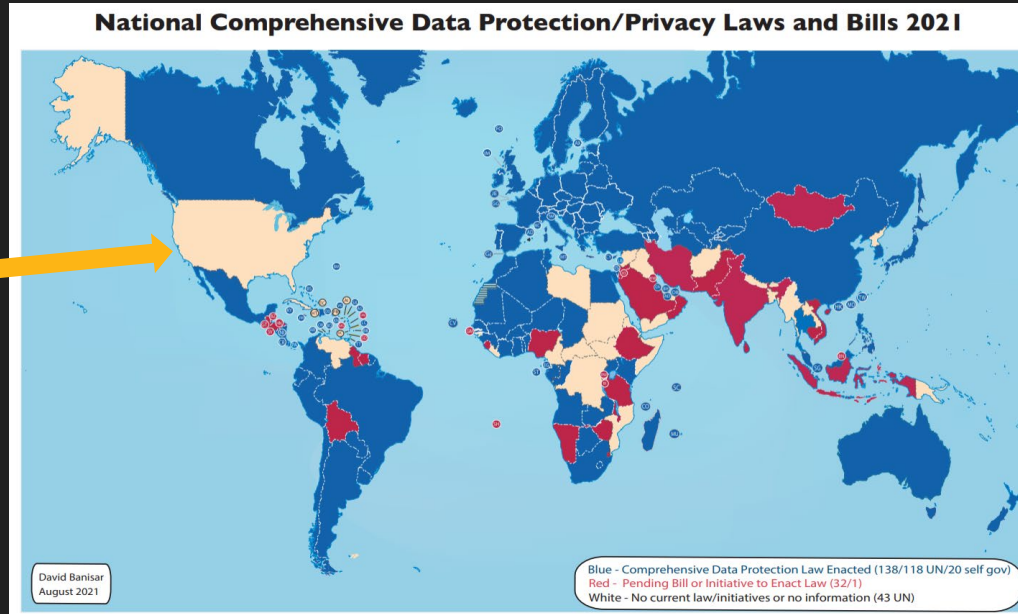


[1] National Comprehensive Data Protection/Privacy Laws and Bills 2023, D Banisar - Privacy Laws and Bills, 2023

# Regulations (August 2021)

- Over 140 countries, self-governing jurisdictions, territories have adopted national laws and almost 30 countries and jurisdictions have pending bills or initiatives [1]

CCPA in  
California





# European General Data Protection Regulation (GDPR)



*Slides from Martin Degeling*

# GDPR

- In effect starting May 25th, 2018
- Unified data protection rules in all 27 member states of the European Union
- Replaced previous Data Protection Directive
  - Unlike Data Protection Directive, GDPR is a law (i.e., other, national laws are not required)
- Is applicable to all services offered within the EU (regardless of where the company is located)

# New and Expanded Rights

- Right to be informed
- Right of access
- Right to rectification
- Right to erasure
- Right to restriction
- Right to data portability
- Right to object
- Right to prevent automated processing, including profiling

# Data Protection by Default/Design

- Aka **Privacy by design**

*“Taking into account the **state of the art**, [...] the controller shall [...] implement appropriate technical and organisational measures [...] to implement data-protection principles, such as data **minimisation**, in an effective manner [...].”*

- Collecting data for unknown purposes is forbidden

# Data Portability

*“The data subject shall have the right to receive the personal data concerning him or her [...] in a structured, commonly used and machine-readable format”*

- Right to export data and transfer it to another data controller



# Right to be Forgotten

- Aka **Right to Erasure**

*“The data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay and the controller shall have the obligation to erase personal data without undue delay”*

- When purpose is fulfilled, consent is withdrawn, unless the controller demonstrates **compelling legitimate grounds** for the processing

# Automatic Decision-Making

- *“The data subject shall have the right **not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.**”*
- Need to identify cases to which this applies and make sure user intervention is possible
- Users have the right to challenge automatic decisions

# Regulatory Measures

- **High fines:** up to 20,000,000 EUR or up to 4% of the annual worldwide turnover
- **Data breaches** must be reported within 72 hours, data subjects must be notified
- **Need for data protection impact assessment**

## The biggest GDPR fines (2025)

1. Meta GDPR fine- €1.2 billion
2. Amazon GDPR fine – €746 million
3. Meta GDPR fine – €405 million
4. Meta GDPR fine – €390 million
5. TikTok GDPR fine- €345 million
6. LinkedIn GDPR fine – €310 million
7. Uber GDPR fine – €290 million
8. Meta GDPR fine – €265 million
9. WhatsApp GDPR fine – €225 million
10. Google LLC fine- €90 million

<https://dataprivacymanager.net/5-biggest-gdpr-fines-so-far-2020/>

| Country   | Date of Decision | Fine [€]  | Controller/ Processor         | Quoted Art.  | Type  | Source                                    |
|---|------------------|-----------|-------------------------------|--|---|---|
| <br>SPAIN            | 2024-11-13       | 500       | 4T OCIO Y CAFÉ 2009           | Art. 6 GDPR  | Insufficient legal basis for data processing                                      | <a href="#">link</a> <a href="#">link</a> |
| <br>SPAIN            | 2024-10-31       | 900       | RIVENDELL TECHNOLOGY, S.L.    | Art. 58 (2) GDPR   | Insufficient cooperation with supervisory authority                               | <a href="#">link</a>                      |
| <br>SPAIN            | 2024-11-06       | 1,000     | MINAS DE VALDECASTILLO, S.A.. | Art. 5 (1) c) GDPR, Art. 13 GDPR   | Non-compliance with general data processing principles                            | <a href="#">link</a>                      |
| <br>SPAIN            | 2024-11-22       | 220,000   | CARTONAJES BAÑERES, S.A       | Art. 15 GDPR, Art. 35 GDPR   | Insufficient technical and organisational measures to ensure information security | <a href="#">link</a>                      |
| <br>THE NETHERLANDS  | 2024-12-23       | 40,000    | Coolblue B.V                  | Art. 5 (1) a) GDPR, Art. 6 (1) GDPR  | Insufficient legal basis for data processing                                      | <a href="#">link</a> <a href="#">link</a> |
| <br>SPAIN            | 2024-09-26       | 1,300,000 | TELEFÓNICA DE ESPAÑA SAU      | Art. 5 (1) f) GDPR, Art. 5 (1) c) GDPR   | Non-compliance with general data processing principles                            | <a href="#">link</a>                      |
| <br>THE NETHERLANDS | 2024-11-26       | 4,750,000 | Netflix International B.V.    | Art. 5 (1) a) GDPR, Art. 12 (1) GDPR, Art. 13 (1) c), e), f) GDPR, Art. 13 (2) GDPR, Art. 15 (1) | Insufficient fulfilment of information obligations                                | <a href="#">link</a> <a href="#">link</a> |

<https://www.enforcementtracker.com/>

# Discussion

**Do we need Data Privacy Regulations?**

**Are DPRs enough?**

**Can we make sure that DPRs are updated?**

**How do we implement DPRs?**

...