# What kind of vulnerabilities are present?

- Are there known *Common Vulnerabilities and Exposures (CVE)* of the identified technologies? → see cve.org, cvedetails.com, nvd.nist.gov
- Is the system vulnerable to common types of attacks?

**CVSS v3.1 Severity and Metrics:**
**Base Score:** 9.8 CRITICAL
**Vector:** AV:N/AC:L/PR:N/UI:N/S:U/C:H/I:H/A:H
**Impact Score:** 5.9
**Exploitability Score:** 3.9

**Attack Vector (AV):** Network
**Attack Complexity (AC):** Low
**Privileges Required (PR):** None
**User Interaction (UI):** None
**Scope (S):** Unchanged
**Confidentiality (C):** High
**Integrity (I):** High
**Availability (A):** High

⚠ CVE-2022-22965 is in the CISA Known Exploited Vulnerabilities Catalog

CISA vulnerability name:
Spring Framework JDK 9+ Remote Code Execution Vulnerability

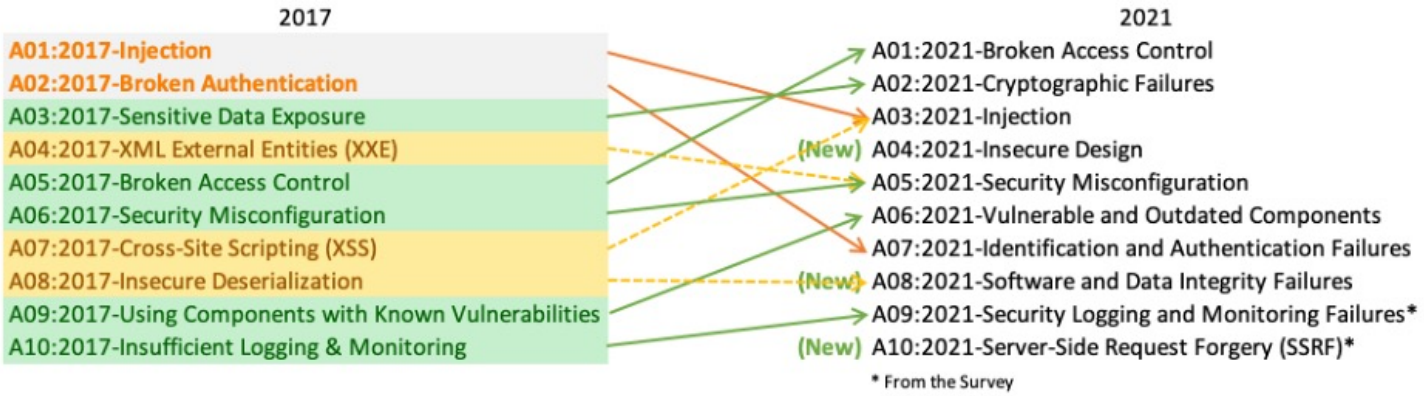
CISA required action:
Apply updates per vendor instructions.

CISA description:
Spring MVC or Spring WebFlux application running on JDK 9+ may be vulnerable to remote code execution (RCE) via data binding.

Added on 2022-04-04 Action due date 2022-04-25



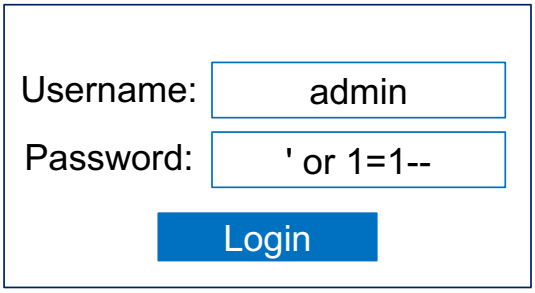
# OWASP Top 10 vulnerabilities



# Injection: Unintended functionality

```
db.execute("SELECT * FROM users WHERE
username = 'admin' AND password = '' or
1=1--'")
```

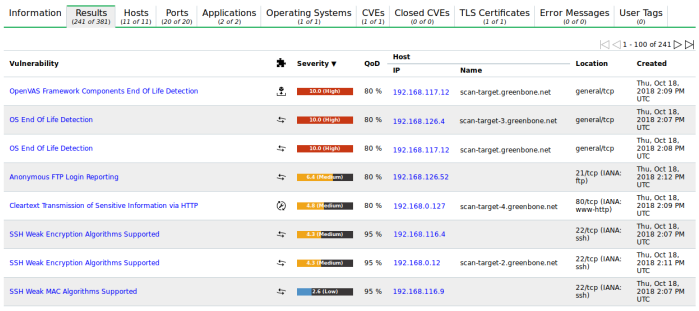
→ Authenticates if the user with the username ”admin” exist, regardless of the password

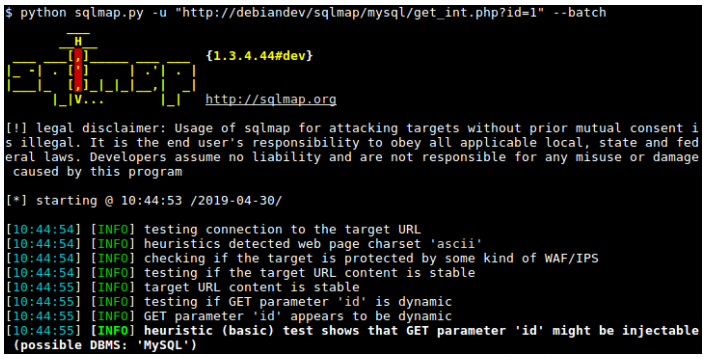





# How to find vulnerabilities

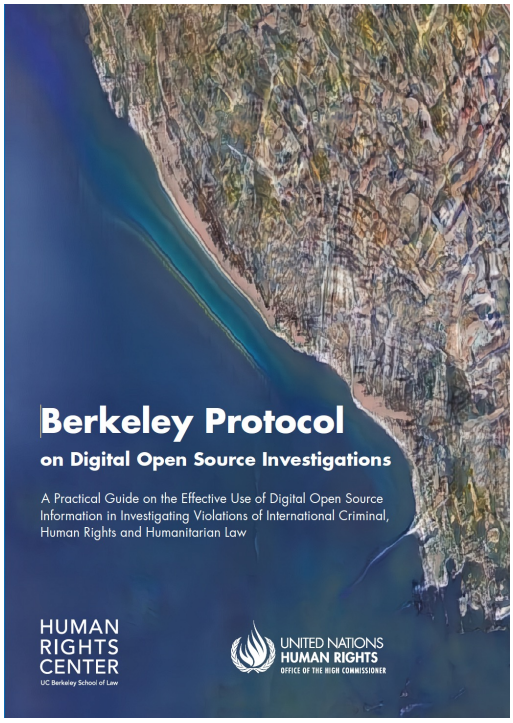
- CVE/exploit databases
- System scans for known CVEs (e.g. via OpenVAS) → can be invasive
- Scans for common vulnerability types (e.g. SQL injections via sqlmap) → can be invasive
- Threat intelligence, e.g. shared knowledge of previous attacks
- Manual inspection → requires most skill



# Open-source intelligence (OSINT)

- Open-Source Intelligence (OSINT): ”intelligence produced by collecting, evaluating and analyzing *publicly available information* with the purpose of answering a specific intelligence question”
- A subset of HUMINT (human intelligence) in general, which includes non-public information sources
- Passive (no direct contact to the target)






# OSINT use cases

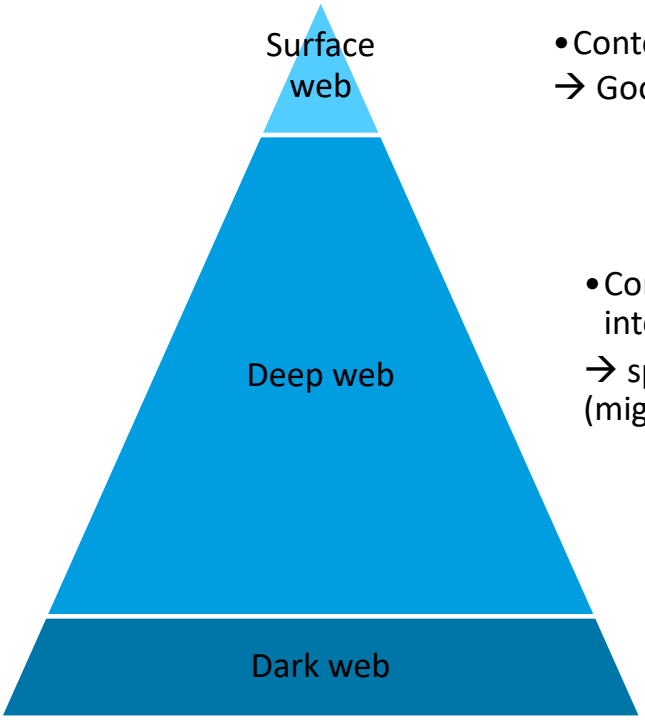
- For good
  - Intelligence/law enforcement
  - Activism, incl. war crimes/human right violations investigations
  - Fact checking
  - Threat analysis
- For bad: cyberattacks incl. social engineering, stalking, persecution of political opponents or selected population groups…
- Ethical and legal aspects are important → see Berkeley protocol

bellingcat

trace labs
CROWDSOURCED OPEN SOURCE INTELLIGENCE

TRACE AN OBJECT
STOP CHILD ABUSE

# Where can information be found?




- Content indexed by search engines
→ Google (Bing, etc.) searches

- Content not indexed by search engines, e.g. internal web pages

→ specialised approaches for searching (might require additional access)

- Content only accessible anonymously via specific software (e.g. Tor browser)

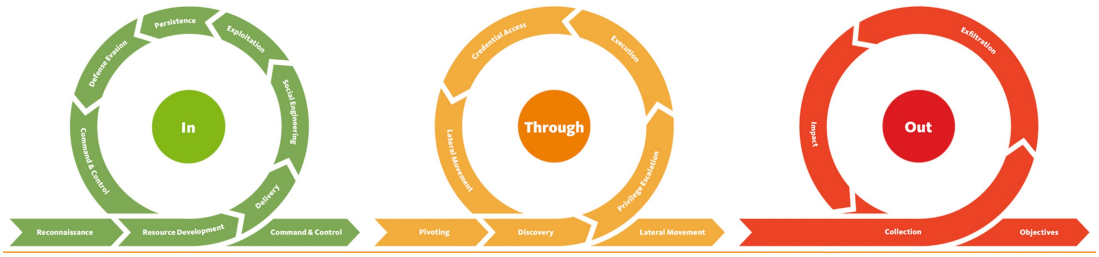
→ specific tools/search engines for searching



# Cyber Kill Chain alternatives

- Criticism: too high-level, all steps not always followed in that order

- Mitre ATT&CK: https://attack.mitre.org
  - 14 tactics with 188 related techniques
  - Aimed to provide more concrete descriptions of attacks
- Unified kill chain
  - 3 high-level stages: initial foothold, network propagation, acting on objectives
  - 18 tactics detailing the stages



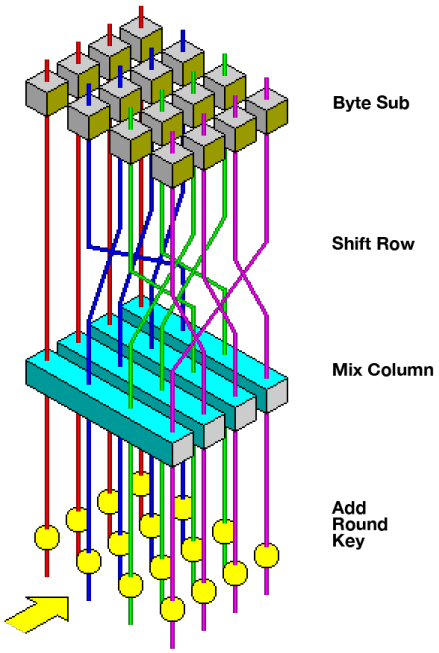

**Reconnaissance**

10 techniques

- Active Scanning (3)
- Gather Victim Host Information (4)
- Gather Victim Identity Information (3)
- Gather Victim Network Information (6)
- Gather Victim Org Information (4)
- Phishing for Information (3)
- Search Closed Sources (2)
- Search Open Technical Databases (5)
- Search Open Websites/Domains (3)
- Search Victim-Owned Websites

# Block ciphers

- Idea: map *fixed size* blocks of plaintext to *fixed size* blocks of ciphertext
- Variety of techniques to achieve *non-linearity*: permutation (substitution boxes), XOR, column mixing
- To avoid problems with key reuse: use different *initialization vectors*
- Larger message: different *modes* of combining the blocks (resulting in *stream cipher*)
- Notable examples: DES (broken), **AES (broken for some modes of operation)**

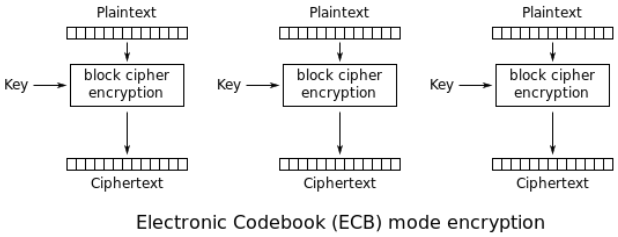



Source: Wikipedia



# AES: modes of operation

- Naive idea: encrypt each block separately → doesn't completely hide patterns in plaintext



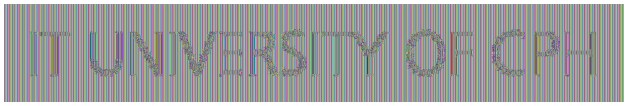

Electronic Codebook (ECB) mode encryption

Source: Wikipedia

- Other solutions: CBC (cipher block chaining), CFB (cipher feedback), CTR (counter) → see also https://csrc.nist.gov/projects/block-cipher-techniques/bcm/current-modes

Plaintext image


AES-ECB encrypted



AES-CBC encrypted

→ Try it yourself: https://github.com/robertdavidgraham/ecb-penguin

# Symmetric vs. Asymmetric cryptography

| Symmetric | Asymmetric | Common approach: hybrid encryption |
|---|---|---|
| • A secret key shared among the communication partners<br>• Key management: establishing keys, revoking keys, storing keys… | • A public key available to everyone<br>• A secret key known only to the receipient<br>• Key management: easier than in symmetric cryptography<br>• Mostly less efficient (in terms of storage/performance) than symmetric encryption | • Use asymmetric encryption to exchange a secret key<br>• Use the secret key for symmetric encryption |

**In all cases, establishing sender authenticity is needed**



# Summary: cryptographic techniques

| | Symmetric | Asymmetric |
|---|---|---|
| Principle | Same key for both parties | A pair of keys |
| Confidentiality (examples) | AES | RSA |
| Integrity (examples) | MACs | Digital signatures |
| Issues | Complex key management | Lower efficiency compared to symmetric cryptography |

# Example: Initial recoding

| ID | Sex | ZIP code |
|---|---|---|
| 1 | female | 1150 |
| 2 | male | 1259 |
| 3 | female | 2300 |
| 4 | male | 2100 |
| 5 | male | 1368 |
| 6 | female | 2100 |

| ID | Sex | ZIP code |
|---|---|---|
| 1 | female | 1XXX |
| 2 | male | 1XXX |
| 3 | female | 2XXX |
| 4 | male | 2XXX |
| 5 | male | 1XXX |
| 6 | female | 2XXX |

Goal: k-anonymity with k = 2
Quasi-identifiers: sex, zip code

Generalisation: remove last 3 digits from the ZIP code → not enough



# Example: Additional supression

| ID | Sex | ZIP code |
|---|---|---|
| 1 | female | 1XXX |
| 2 | male | 1XXX |
| 3 | female | 2XXX |
| 4 | male | 2XXX |
| 5 | male | 1XXX |
| 6 | female | 2XXX |

| ID | Sex | ZIP code |
|---|---|---|
| 1 | female | * |
| 2 | male | 1XXX |
| 3 | female | 2XXX |
| 4 | male | 2XXX |
| 5 | male | * |
| 6 | female | 2XXX |

Supression: remove values related to sample unique records

→ Problems: records still identifiable
- Attacker knowing that Alice (female, 1150) is in the dataset → Alice is record 1
- No sense supressing ZIP code of record 1 if it is 2XXX

# Example: Alternative supression

| ID | Sex | ZIP code | f |
|---|---|---|---|
| 1 | female | 1XXX | 2 |
| 2 | male | 1XXX | 2 |
| 3 | female | * | 3 |
| 4 | male | 2XXX | 2 |
| 5 | male | * | 3 |
| 6 | female | 2XXX | 2 |

→ Achieving k-anonymity of k=2

* can be any value

→ Record 1 can have the same values of quasi-identifiers as record 3

→ Record 5 can have the same values as records 2, 4

→ Record 2 can have the same values as record 5 but **not** record 4



# Issues with l-diversity: Skewness attack

2-anonymity
2-diversity

| Sex | Age | Condition |
|---|---|---|
| female | 20-34 | Diabetes |
| male | 35-49 | Flu |
| female | 20-34 | Cold |
| female | 35-49 | HIV |
| male | 35-49 | Cold |
| female | 35-49 | Flu |

| Name | Sex | Age |
|---|---|---|
| Eve | female | 36 |

"0.8% of adults have HIV"

Given the released table, what is the probability of Eve having HIV?

→ Issue with uneven distribution of sensitive values

# Issues with l-diversity: Similarity attack

2-anonymity
2-diversity

| Sex | Age | Condition |
|---|---|---|
| female | 20-34 | Diabetes |
| male | 35-49 | Flu |
| female | 20-34 | Cold |
| female | 35-49 | Stomach cancer |
| male | 35-49 | Cold |
| female | 35-49 | Breast cancer |

| Name | Sex | Age |
|---|---|---|
| Eve | female | 36 |

What can you learn about Eve?



# Summary statistics: Protection methods

| Gender | Number of employees | Mean income (DKK/month) |
|---|---|---|
| Male | 3 | 60 000 |
| Female | 4 | 45 000 |

Supression

Noise addition

| Gender | Number of employees | Mean income (DKK/month) |
|---|---|---|
| Male | * | 60 000 |
| Female | * | 45 000 |

| Gender | Number of employees | Mean income (DKK/month) |
|---|---|---|
| Male | 3 | 50216.55 |
| Female | 4 | 44033.31 |

# Query-based release: Protection methods

| Age | Gender | ZIP | Travel | Illness |
|---|---|---|---|---|
| 29 | M | 2200 | No | Asthma |
| 21 | F | 2300 | Italy | Diabetes |
| 23 | F | 2300 | No | No |
| 25 | M | 2100 | Sweden | No |

- Select count of entries where:
  - Gender = ”female” AND
  - Age = 21 AND
  - Illness = ”no”
- Output: 0

Query restriction/auditing

Output: NA

Noise addition

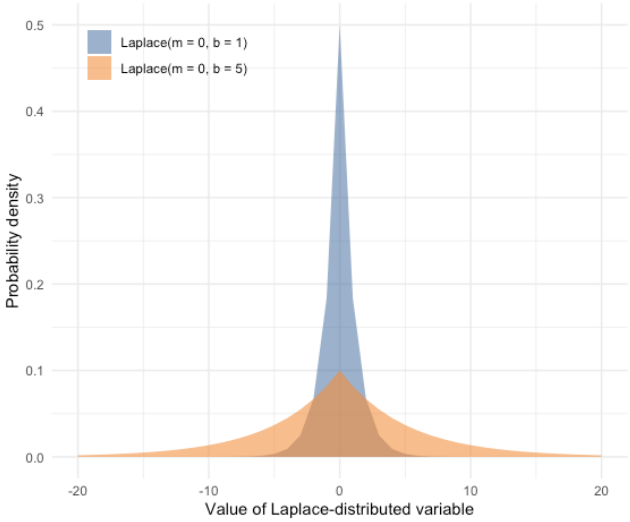
Output: 2

# Laplace mechanism

- Idea: add random noise $X \in \mathbb{R}$ to query output
- $X \sim$ Laplace$(m, b)$ as a random variable drawn from the Laplace distribution
  - Probability density function:

$$f(x|m,b) = 1/(2b)\, e^{\left(\frac{-|x-m|}{b}\right)}$$

  - Mean: $m$
  - Variance: $2b^2$






# Example: Summary statistics

Database $D$

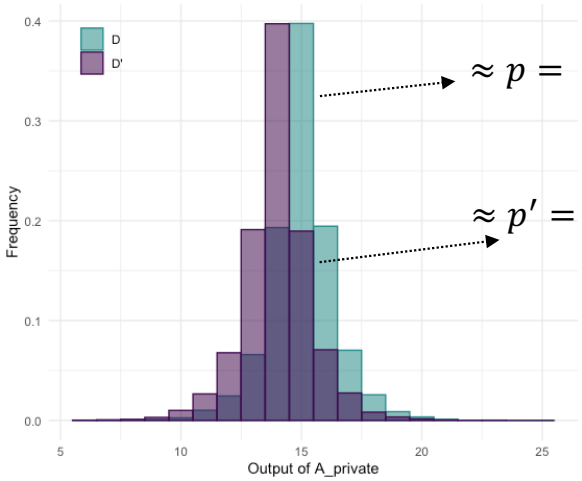
Alice

- $A_{private}(D) = A_{nonprivate}(D) + X$
- $X \sim Laplace\left(0, \frac{1}{\varepsilon}\right)$

→ $A_{private}$ is $\varepsilon$-differentially private!

- E. g. for $\varepsilon = 1$

Database $D'$




$$\approx p = P(A_{private}(D) \in (14.5, 15.5))$$

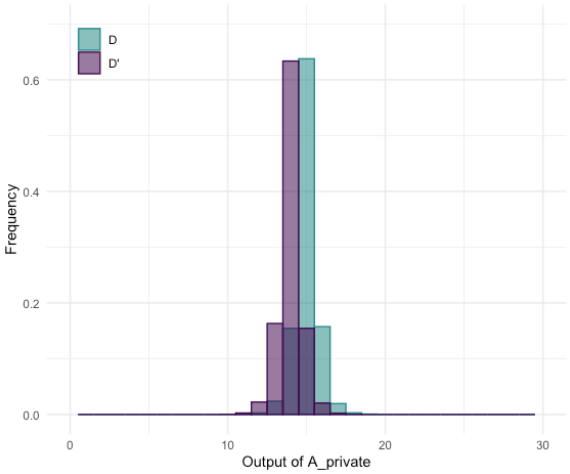
$$\approx p' = P(A_{private}(D') \in (14.5, 15.5))$$

$$\frac{p}{p'}, \frac{p'}{p} \leq e^1 \approx 2.71$$

# Example: Summary statistics

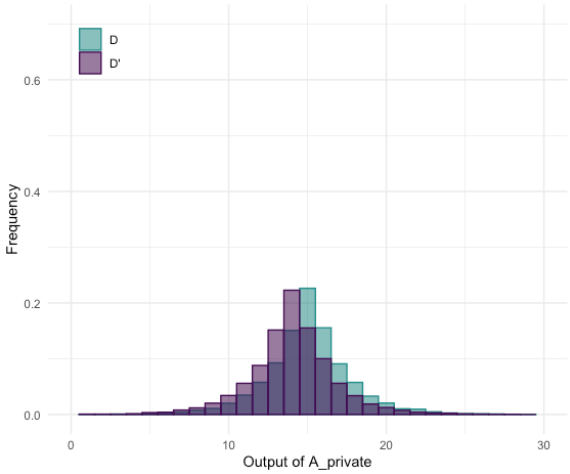
$\varepsilon = 2$: better utility (outputs less spread out), less privacy

$$\left(\frac{p}{p'}, \frac{p'}{p} \leq e^2 \approx 7.39\right)$$



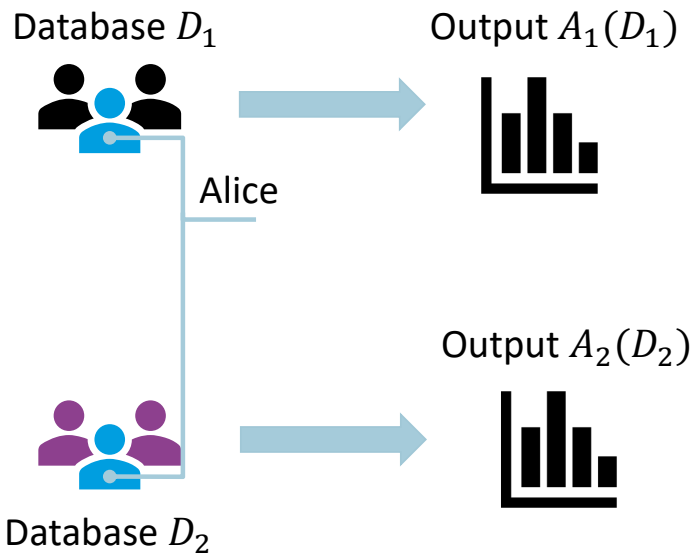

$\varepsilon = 0.5$: less utility (outputs more spread out), better privacy

$$\left(\frac{p}{p'}, \frac{p'}{p} \leq e^{1/2} \approx 1.64\right)$$






# Properties: Sequential Composition




$A_1$: $\mathcal{D} \longrightarrow \mathcal{S}$ is $\varepsilon_1$-differentially private

$A_2$: $\mathcal{D} \longrightarrow \mathcal{S}$ is $\varepsilon_2$-differentially private

$\Rightarrow (A_1, A_2)$: $\mathcal{D} \times D \longrightarrow \mathcal{S} \times S$ is $(\varepsilon_1 + \varepsilon_2)$-differentially private

→ **privacy budget of $A_1$ and $A_2$ combined is $\varepsilon_1 + \varepsilon_2$**

- Interpretation
  - Privacy guarantees decrease with future releases based on the same data
  - However: we can quantify an upper bound on privacy decrease

# Properties: Parallel Composition

Database $D_1$

Output $A_1(D_1)$

Output $A_2(D_2)$

Database $D_2$

$A_1$: $\mathcal{D} \longrightarrow \mathcal{S}$ is $\varepsilon_1$-differentially private

$A_2$: $\mathcal{D} \longrightarrow \mathcal{S}$ is $\varepsilon_2$-differentially private

$\Rightarrow (A_1, A_2)$: $\mathcal{D} \times D \longrightarrow \mathcal{S} \times S$ is max($\varepsilon_1$,$\varepsilon_2$)-differentially private on all $D_1, D_2$: $D_1 \cap D_2 = \emptyset$

→ Privacy budget is max($\varepsilon_1$,$\varepsilon_2$)

- Interpretation:
  - Analyses of disjoint datasets do not decrease privacy
  - Example: contingency tables, histograms…



# Properties: Group differential privacy

Database $D$

Analysis output $A(D)$

Database $D_k{}'$

Analysis output $A(D_k{}')$

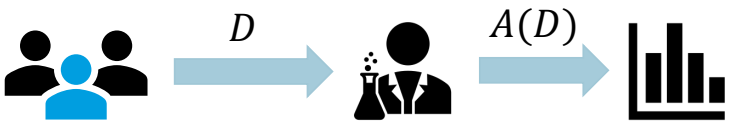
$A$: $\mathcal{D} \longrightarrow \mathcal{S}$ $\varepsilon$-differentially private

$D, D_k{}' \in \mathcal{D}$: databases that differ in **k records**

→ $A$ achieves $k\varepsilon$-differential privacy on $D, D_k{}'$:
$\forall\, S \in \mathcal{S}: P(A(D) = S) \leq e^{k\varepsilon} P(A(D_k') = S)$

Interpretation: possible to calculate disclosure level for groups (e.g. households)

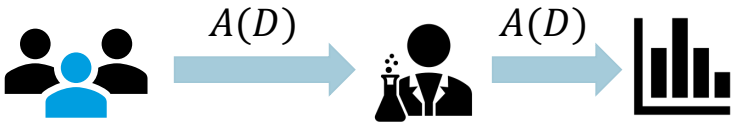
# Centralised vs. local model




Centralised model: curator gets non-anonymised data from the user

Local model

- Only anonymised data received by the curator
- Less risk of disclosure (e.g. due to insufficient protection of raw data)
- However, some degree of trust still required → still needs to be ensured that the data is anonymised

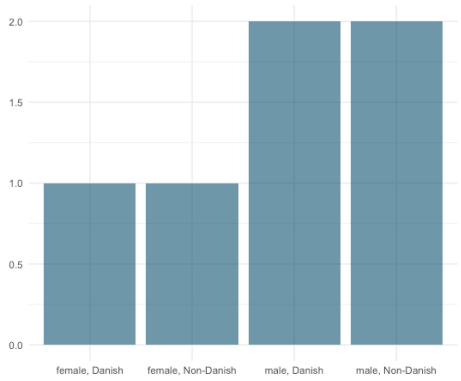


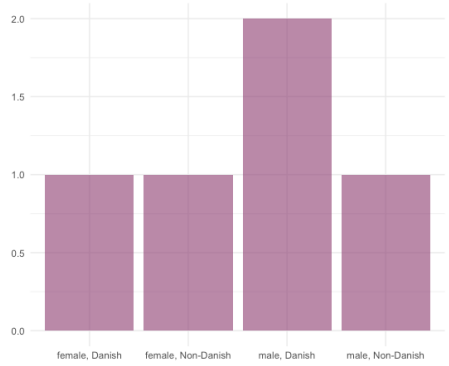



# Complex data

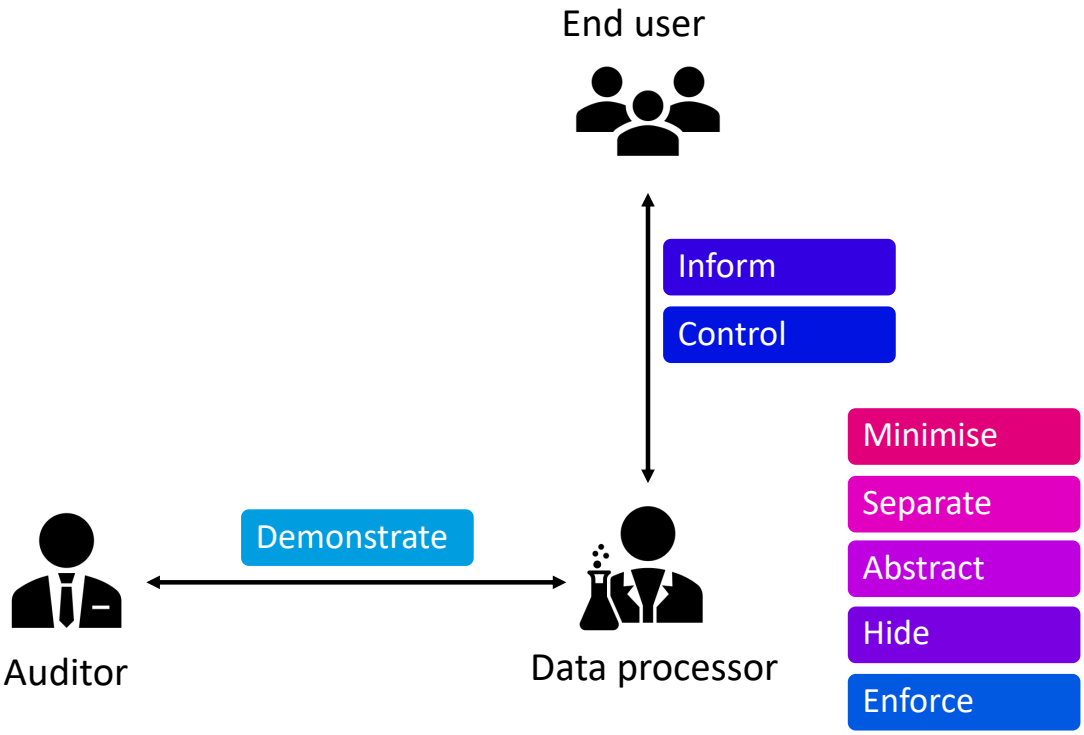
- Differentially private algorithms mostly for tabular and query-based releases → limited analytical use
- Microdata-based releases → either histographs (see readings) or local differential privacy
  - Hard to apply on high-dimensional data
  - Even harder on complex data (e.g. social network graphs)

→ Ongoing topic of research



# Privacy design strategies



# Structural challenges: Compexity of risk estimation

- Underestimation of risks due to biases
- But also: objectively hard to estimate risks
  - Reidentification risks: everything can be an identifer (recall: Netflix challenge)
  - Attribute disclosure risks: possible even without reidentification
  - Inferential disclosure: Which data can be used for predictive modeling?
- Aggregated effects of data sharing must be considered (”death by a thousand of cuts”)


Legal information


Public data


Purchase & behaviour data


Intimate facts



# Summary

- Human factors influence the security and privacy of systems
  - Security: the user has to be invloved, e.g. during authentication
  - Privacy: the user has to be involved, see inform & control privacy strategies
- Reasons for failures in security and privacy
  - Lack of motivation: security as non-primary task, decision making biases
  - Lack of ability: mental capabilities & structural challenges
- Possible mitigation approaches
  - Focusing on usability of the systems
  - Focusing on educating the user
  - Focusing on reducing user involvement
  - Human-centered security and privacy by design

aumento

# Legal Basis

Some of the take aways from the decision:

1. All legal basis's apart from consent must be interpreted restrictively!

**2. Sensitive data takes precedent:**

If you collect a set of data containing both sensitive data and non-sensitive data, in particular, collected *en bloc* without it being possible to separate the data items from each other at the time of collection, the processing of that set of data must be regarded as being prohibited, within the meaning of Article 9(1) of the GDPR, if it contains at least one sensitive data item and none of the derogations in Article 9(2) of that regulation applies.

# GDPR

aumento
ADVOKATFIRMA



**Responsibility of the controller**

1. Taking into account the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for the rights and freedoms of natural persons, the controller shall implement appropriate technical and organisational measures to ensure and to be able to demonstrate that processing is performed in accordance with this Regulation. Those measures shall be reviewed and updated where necessary.

2. Where proportionate in relation to processing activities, the measures referred to in paragraph 1 shall include the implementation of appropriate data protection policies by the controller.

3. Adherence to approved codes of conduct as referred to in Article 40 or approved certification mechanisms as referred to in Article 42 may be used as an element by which to demonstrate compliance with the obligations of the controller.

aumento
ADVOKATFIRMA
Side 32
03-01-2024

## The main principle relating to compliance when processing personal data

### Accountability

means that the controller shall be responsible for, and be able to demonstrate compliance with the 6 main principles

What does demonstrate mean?

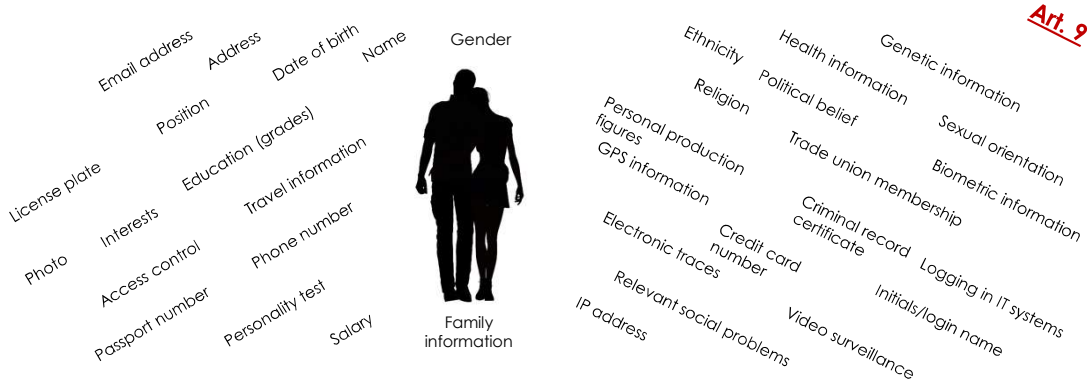

Documentation for the compliance.

Evidence in a criminal case?

No – the prosecutor has the burden of proof.

aumento
ADVOKATFIRMA
Side 35
03-01-2024

## Personal Data!



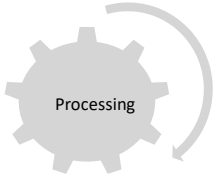
aumento
ADVOKATFIRMA



# Processing

Any activities to which personal data are exposed, from the start until the end, are comprised by the "data processing concept"

- Storage
- Blocking
- Search
- Disclosure
- Alteration
- Erasure
- Dissemination
- Collection
- Combination
- Recording
- Use




aumento
ADVOKATFIRMA



# Some of the GDPR definitions

Pseudonymisation means the processing of personal data in such a manner

that the personal data can no longer be attributed to a specific data subject

without

the use of additional information.

provided that such additional information is kept separately and

is subject to technical and organisational measures

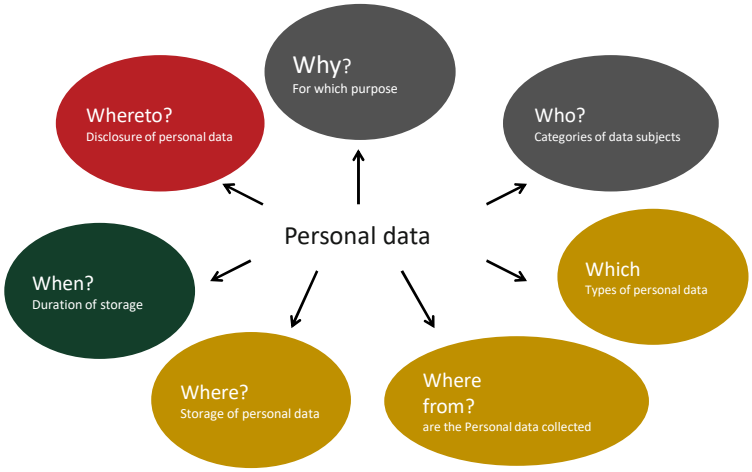
to ensure that the personal data are not attributed to an identified or identifiable natural person.

The main questions to ask oneself:

The seven "W"questions

aumento ADVOKATFIRMA

Side 74
03-01-2024




The fast track analyse of a data collection – all questions need to be answered.

# INDIVIDUAL RIGHTS

aumento ADVOKATFIRMA

Side 76
03-01-2024


*Own initiative*

*Request*

- The right to be informed (Art. 13 & 14)
- The right to access and rectification
- The right to object
- The right to erasure
- The right to data portability

aumento
ADVOKATFIRMA
Side 96
03-01-2024

# Territorial scope

What applies to controller or processor established outside of the Union?

Applies to personal data of *data subjects who are in the Uniion*

If the processing activities are related to:

the offering of goods or services,

for free or against payment.

Monitoring of data subjects behaviour as far as their behaviour takes place within the Union.

GDPR applies to the processing of personal data by a controller not established in the Union, but in a place where Member State law applies by virtue of public international law.

aumento
ADVOKATFIRMA
100

# Profiling

Profiling is:

ANY form of processing of personal data with the purpose of:

[1]evaluate [2]certain [3]personal [4]aspects

in particular (examples not an exhausting list.)

to analyse or predict (can be both – first you analyse and then you predict) *aspects* concerning
performance at work,
economic situation,
health,
personal preferences,
interests,
reliability,
behaviour,
location or movements.

Must observe the main principles in particular it should be fair and transparent