Privacy models: Summary

Vicenç Torra

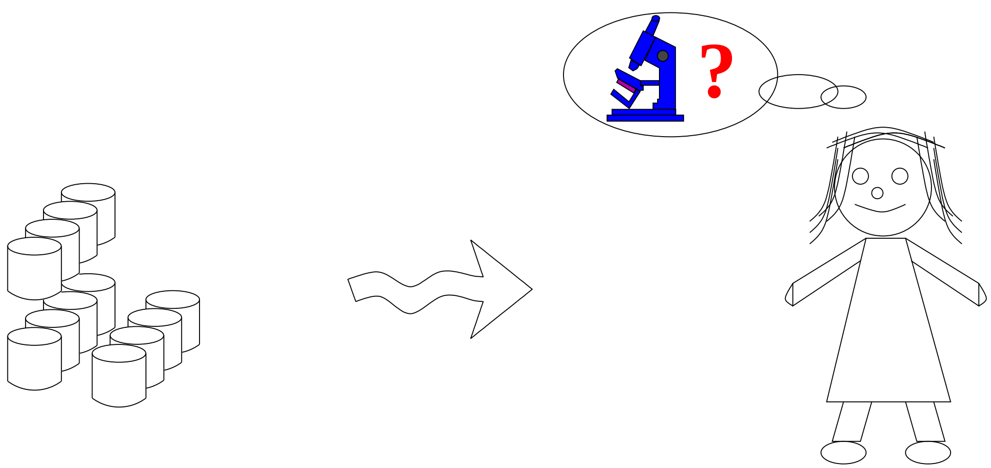
November, 2022

Umeå University, Sweden

Outline

- Privacy models

Privacy models



Privacy models

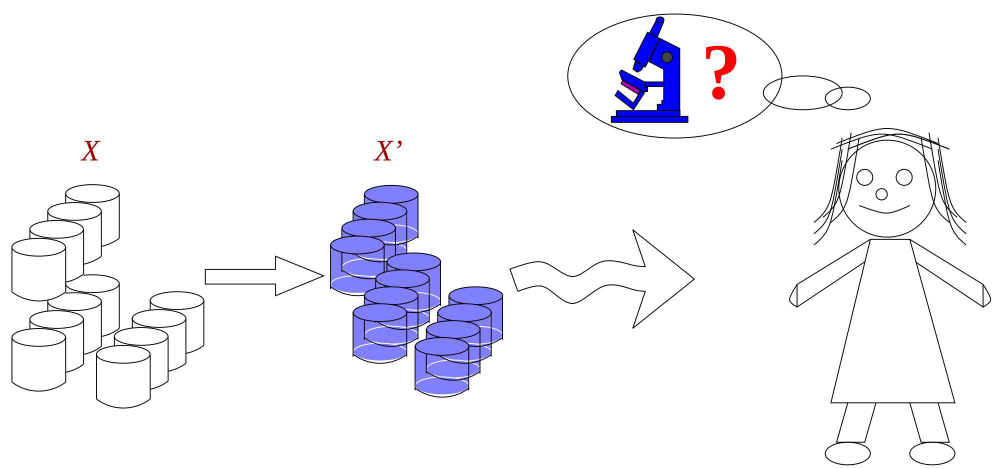
Privacy models. A computational definition for privacy. Examples.

- **Reidentification privacy.** Avoid finding a record in a database.
- **k-Anonymity.** A record indistinguishable with $k-1$ other records.
- **Secure multiparty computation.** Several parties want to compute a function of their databases, but only sharing the result.
- **Differential privacy.** The output of a query to a database should not depend (much) on whether a record is in the database or not.
- **Result privacy.** We want to avoid some results when an algorithm is applied to a database.
- **Integral privacy.** Inference on the databases. E.g., changes have been applied to a database.
- **Homomorphic encryption.** We want to avoid access to raw data and partial computations.

Privacy models

Privacy models. A computational definition for privacy. Publish a DB

- **Reidentification privacy.** Avoid finding a record in a database.
- **k-Anonymity.** A record indistinguishable with $k - 1$ other records.
- **k-Anonymity, l-diversity.** l possible categories
- **Interval disclosure.** The value for an attribute is outside an interval computed from the protected value: values different enough.
- **Result privacy.** We want to avoid some results when an algorithm is applied to a database.



Privacy models

Privacy models. A computational definition for privacy. Publish a DB

- Modify DB X to obtain a DB X' compliant with the privacy model.

Original DB X:

Respondent	City	Age	Illness
DRR	Barcelona	30	Heart attack
ABD	Barcelona	32	Cancer
COL	Barcelona	33	Cancer
GHE	Tarragona	62	AIDS
CIO	Tarragona	65	AIDS
HYU	Tarragona	60	Heart attack

Published DB X':

——	City	Age	Illness
—	Barcelona	30	Cancer
—	Barcelona	30	Cancer
—	Barcelona	30	Cancer
—	Tarragona	60	AIDS
—	Tarragona	60	AIDS
—	———	–	———

Privacy models

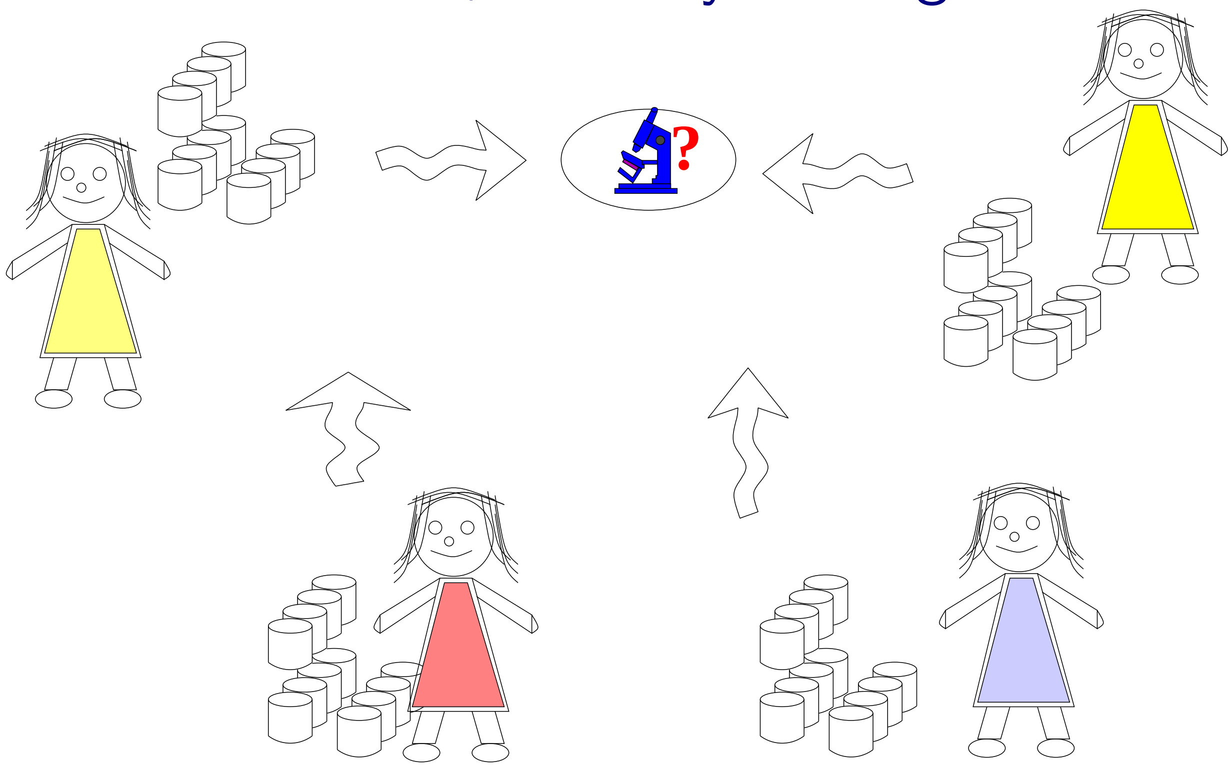
- Difficulties
  Naive anonymization does not work, highly identifiable data, high dimensional data

- Examples of successful reidentification attacks
  Sweeney analysis of USA population, data from mobile data, shopping cards, film ratings

Privacy models

Privacy models. A computational definition for privacy. Share a result

- **Secure multiparty computation.** Several parties want to compute a function of their databases, but only sharing the result.



Privacy models

Privacy models. A computational definition for privacy. Share a result

- Compute

$$f(DB_1, DB_2, DB_3, DB_4)$$

  without sharing DB_1, DB_2, DB_3, DB_4
- Example: national age mean of hospital-acquired infection patients (hospitals do not want to share the age of their infected patients!)

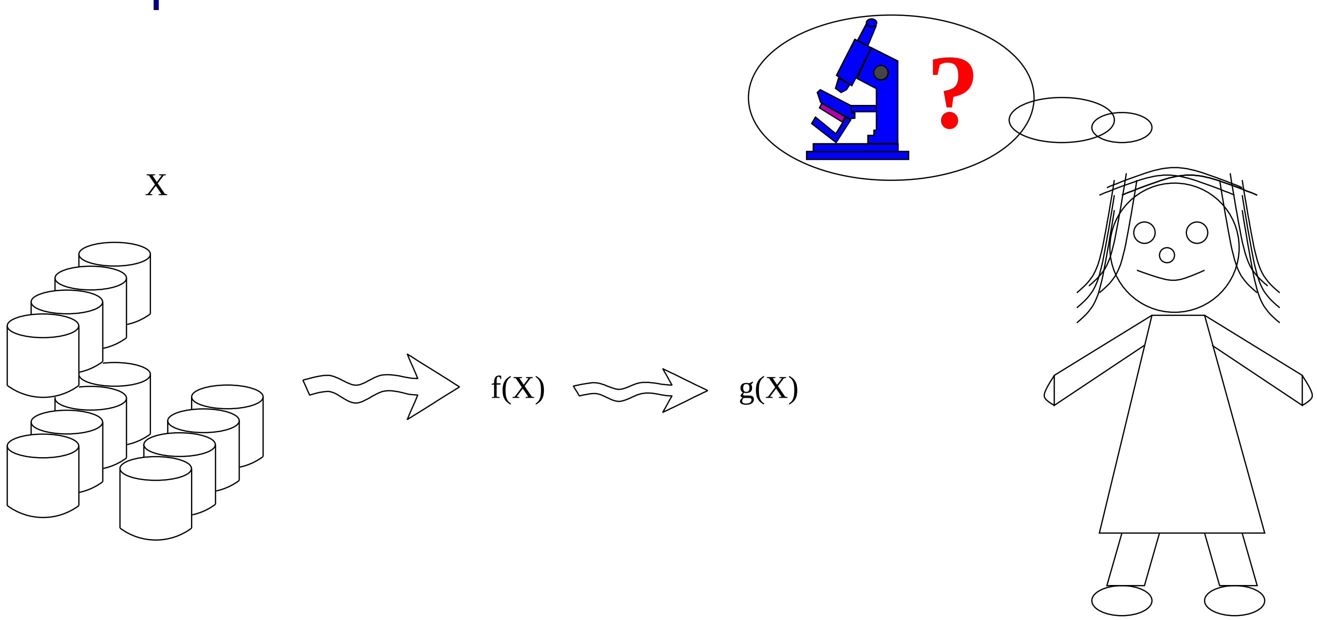
Privacy models

- Difficulties
  Distributed approach (no trusted-third party) – computational cost of solutions

Privacy models

Privacy models. A computational definition for privacy. Compute result

- **Differential privacy.** The output of a query to a database should not depend (much) on whether a record is in the database or not.
- **Integral privacy.** Inference on the databases. E.g., changes have been applied to a database.
- **Homomorphic encryption.** We want to avoid access to raw data and partial computations.



Privacy models

- Difficulties. A simple function can give information on who is in the database
  - E.g., mean salary