Disclosure, Privacy Models, and Privacy Mechanisms

Vicenç Torra

January 2024

Umeå University, Sweden

V. Torra (2022) A guide to data privacy, Springer (Chapter 3)

Outline

- Introduction
- Disclosure risk measures
  - Attribute disclosure risk measures
  - Identity disclosure risk measures

Introduction to the introduction

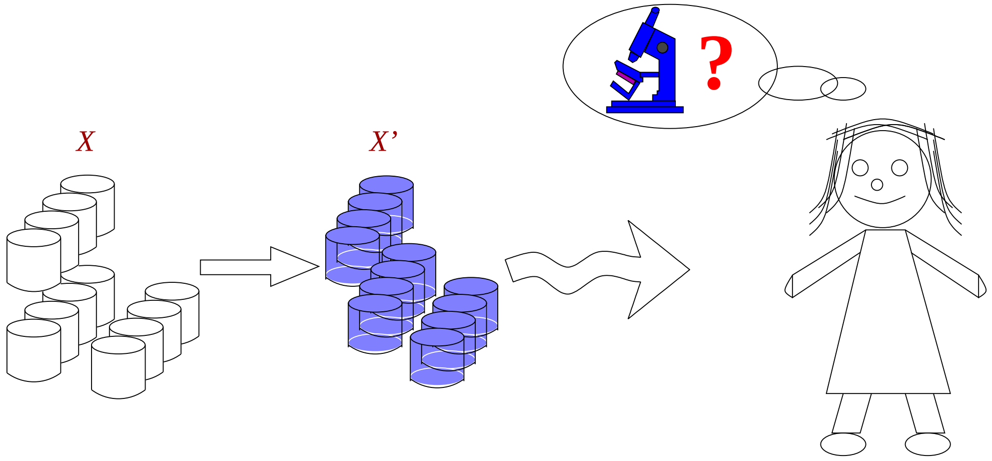
Disclosure, risk measures, and privacy models

- Disclosure, risk measures, and privacy models
- Protection mechanisms
  - Data protection mechanisms,
  - Privacy-preserving machine learning

Protection mechanisms need to be clearly disassociated of privacy models

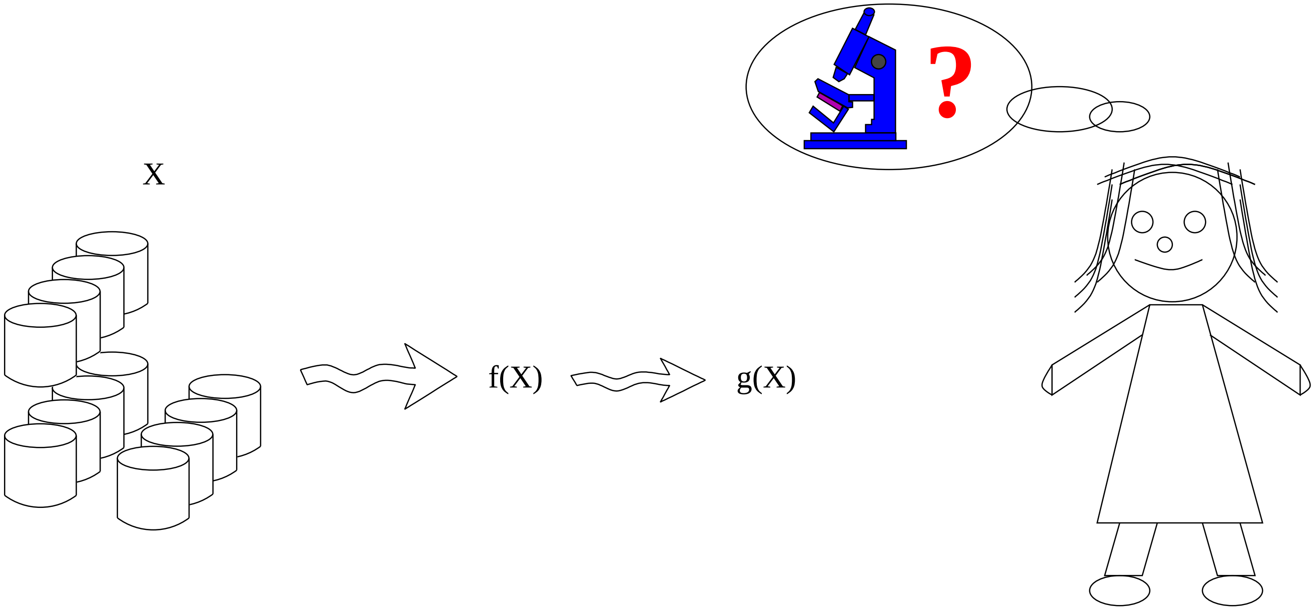
Disclosure, risk measures, and privacy models

- Privacy for data: data sharing, data publishing



Disclosure, risk measures, and privacy models

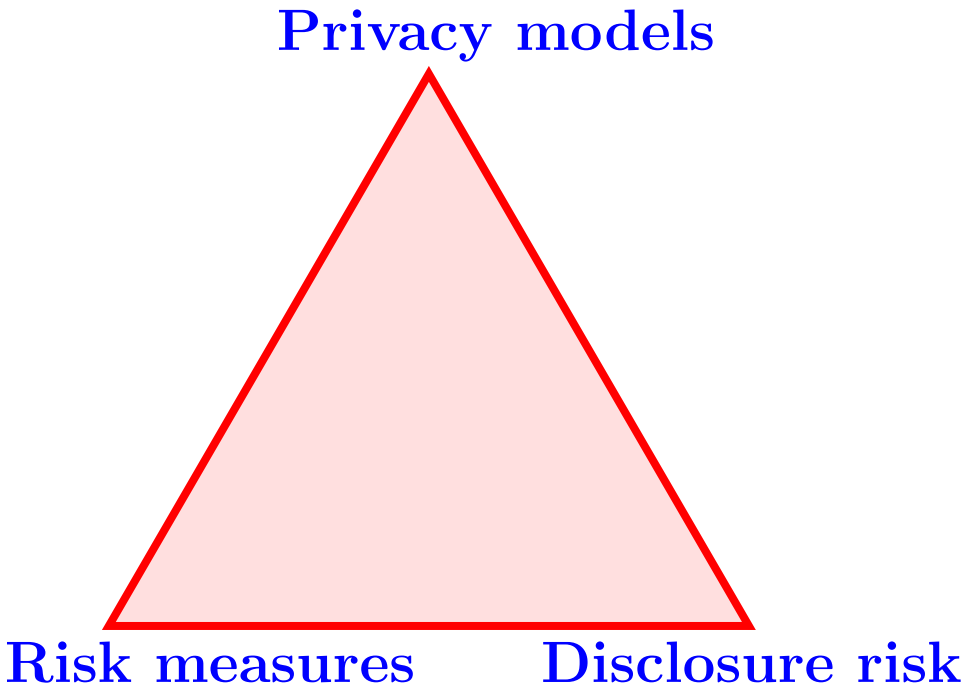
- Privacy for computations



Introduction

Disclosure, risk measures, and privacy models

- Three strongly related concepts
  (and as we will see linked to possible attacks)



Disclosure

Disclosure

- **Def 3.1**
  - Disclosure takes place when intruders take advantage of the observation and analysis of a release to improve their knowledge on some item of interest.

- Release: data, statistics, data-driven machine learning model, output from a running model (e.g., an LLM)

Disclosure

- **Def 3.1**
  - Disclosure takes place when intruders take advantage of the observation and analysis of a release to improve their knowledge on some item of interest.
- Release: data, statistics, data-driven machine learning model, output from a running model (e.g., an LLM)
- Intruder: the intruder attacks the release

Disclosure

- Disclosure. Attackers take advantage of observations to improve their knowledge on some confidential information about an IoI.
  ⇒ SDC/PPDM: Observe DB, Δ knowledge of a particular subject (the respondent in a database)
  - ○ Identity disclosure (entity disclosure). Linkability. Finding Mary in the database.
  - ○ Attribute disclosure. Increase knowledge on Mary's salary.
    also: learning that someone is in the database, although not found.

Disclosure

- Discussion.
  - Identity disclosure. Avoid.
  - Attribute disclosure. A more complex case. Some attribute disclosure is expected in data mining.

    At the other extreme, any improvement in our knowledge about an individual could be considered an intrusion. The latter is particularly likely to cause a problem for data mining, as the goal is to improve our knowledge. (J. Vaidya et al., 2006, p. 7.)

Disclosure

- Another dimension for disclosure.
    - Boolean vs. measurable condition
        - Boolean: Disclosure either takes place or not. Check, definition holds or not? This includes definitions based on a threshold.
        - Measurable: Disclosure is a matter of degree that can be quantified. Some risk is permitted.

Disclosure

- Another dimension for disclosure.
  - Boolean vs. measurable condition
    - ▷ Boolean: Disclosure either takes place or not. Check, definition holds or not? This includes definitions based on a threshold.
    - ▷ Measurable: Disclosure is a matter of degree that can be quantified. Some risk is permitted.
  - This has implication when selecting a protection mechanism
    - ▷ Boolean: focus on one $performance$ measure
      - ⋆ minimize information loss (max. utility), minimize execution time

Disclosure

- Another dimension for disclosure.
  - Boolean vs. measurable condition
    - ▷ Boolean: Disclosure either takes place or not. Check, definition holds or not? This includes definitions based on a threshold.
    - ▷ Measurable: Disclosure is a matter of degree that can be quantified. Some risk is permitted.
  - This has implication when selecting a protection mechanism
    - ▷ Boolean: focus on one *performance* measure
      - ⋆ minimize information loss (max. utility), minimize execution time
    - ▷ Measurable: focus on both *performance* and risk
      - ⋆ multiobjetive optimization problem

Disclosure

- Two dimensions. Privacy models / risk measures

	Attribute disclosure	Identity disclosure
Boolean	Differential privacy Result privacy Secure multiparty computation	k−Anonymity
Quantitative	Interval disclosure	Re−identification (record linkage) Uniqueness

Disclosure

- Attribute disclosure in clusters and cells
  - Aggregates do not avoid disclosure (recall motivating example)
  - Inferences depend on the type of summaries
    indistinguishable/all members of a cell/cluster have the same property

Disclosure

- Attribute disclosure in clusters and cells
  - Aggregates do not avoid disclosure (recall motivating example)
  - Inferences depend on the type of summaries
    indistinguishable/all members of a cell/cluster have the same property

- Types of attacks, for summaries of cells/clusters
  - External attack. An intruder is external to the cell, information is inferred from the analysis of the aggregate.
    All in a cell have money disorder, so, Mr Scrooge also
    We learn that Dona Obdúlia was in the psychiatric unit

Disclosure

- Attribute disclosure in clusters and cells
  - Aggregates do not avoid disclosure (recall motivating example)
  - Inferences depend on the type of summaries
    indistinguishable/all members of a cell/cluster have the same property

- Types of attacks, for summaries of cells/clusters
  - External attack. An intruder is external to the cell, information is inferred from the analysis of the aggregate.
    All in a cell have money disorder, so, Mr Scrooge also
    We learn that Dona Obdúlia was in the psychiatric unit
  - Internal attack. Leakage is caused by intruders (one or coalition) that use their own information to learn about the others.
    All people in the cell but one have eating disorders. Mr Scrooge know the disorder of all the others.
    Dona Obdúlia knows the total income of all other members.

Disclosure

- Attribute disclosure in clusters and cells. When?
  - Anonymization methods based on clustering and aggregation can suffer these types of attacks. E.g., microaggregation, generalization (databases), tabular data protection (aggregates on a few attributes).
  - Privacy models on this type of attacks. Models for tabular data. Variants of k-anonymity.

Disclosure

- Additional discussion.
  - Range of opinions. One extreme, privacy is impossible. On the other, risk is overestimated.
  - E.g., discussions by de Montjoye et al, replies, and replies to the replies.
  - Important point, uniqueness in a sample does not imply uniqueness in the population.

Disclosure

- Additional discussion.
  - Difficulty and success of data privacy depend on the object released (type of data or model) and the level of protection we want to achieve.
  - The discussion points out the importance of defining correctly disclosure risk, accurate ways to measure disclosure risk, and well-defined privacy models.
  - Measures permit us to evaluate and compare privacy mechanisms (masking methods, privacy-preserving machine learning algorithms as well as different instantiations with different parameters).
  - The selection of an appropriate mechanism is key.

Disclosure

- Structure of the slides.
  - Disclosure risk measures
  - Privacy models

Disclosure

- Structure of the slides.
  - Disclosure risk measures
  - Privacy models

- Similar types of disclosure risk appears in different contexts and similar solutions may be needed

- and different types of risk may appear in the same problem and we need to combine different protection mechanisms

Disclosure

- Structure of the slides.
  - Disclosure risk measures
  - Privacy models

- Similar types of disclosure risk appears in different contexts and similar solutions may be needed

- and different types of risk may appear in the same problem and we need to combine different protection mechanisms
  - That's why this explanation of disclosure/privacy models is *independent* of the protection mechanisms

Disclosure

- Types of disclosure according to what is disclosed.
  1. Attribute disclosure (Section 3.1.3)
  2. Identity disclosure (Section 3.1.2)
- Privacy models and measures of disclosure.
  1. Measures for attribute disclosure (Section 3.2)
  2. Measures for identity disclosure (Section 3.3)
  3. Privacy as a measurable condition
     (a) Uniqueness (Section 3.3.1)
     (b) Re-identification (Section 3.3.2)
        i. Data integration: schema and data matching
        ii. Record linkage algorithms: distance-based and probabilistic RL
        iii. Generic vs. specific record linkage algorithms
  4. Privacy as a Boolean condition
     (a) k-Anonymity (Section 3.4.2)
     (b) Differential privacy (Section 3.4.6)
     (c) Secure multiparty computation (Section 3.4.10)
     (d) Interval disclosure (Section 3.2.1)

Risk measures for attribute disclosure

Attribute disclosure for numerical data

- Difficulties of assessing attribute disclosure
  - We saw:

    At the other extreme, any improvement in our knowledge about an individual could be considered an intrusion. The latter is particularly likely to cause a problem for data mining, as the goal is to improve our knowledge. (J. Vaidya et al., 2006, p. 7.)
  - For a data-driven model it is difficult to know in what extent we are measuring attribute disclosure and in what extent we are discussing about the quality of the model. The same applies to data releases. Are we producing data of good quality or are we causing attribute disclosure?

Attribute disclosure for numerical data

- Difficulties of assessing attribute disclosure
  - Ideally, let M be a data-driven model for attribute A with an **excellent generalization capability and no overfitting**. Then, the replacement of values $A(x)$ by $M(x)$ does not imply attribute disclosure.

Attribute disclosure for numerical data

Algorithm 7 Rank-based interval disclosure: $rid(X, V, V', x, p)$.

```
Data: $X$: Original file; $V$: Original attribute; $V'$: Masked attribute;
$x$: record; $p$: percentage

Result: Attribute disclosure for attribute $V'$ of record $x$

begin
    $R(V) :=$ Rank data for attribute $V'$
    $i :=$ position of $V'(x)$ in $R(V)$
    $w := p \cdot |X|/2/100$ (establish the width of the interval)
    $I(x) = [R[\min(i - w, 0)], R[\max(i + w, |X| - 1)]]$ (interval for $x$)
    $rid := V(x) \in I(x)$
    return $rid$
end
```

Attribute disclosure for numerical data

Algorithm 8 Standard deviation-based interval disclosure: $sdid(X, V, V', x,$

Data: X: Original file; V: Original attribute; V': Masked attribute; x: record; p: percentage

Result: Attribute disclosure for attribute V' of record x

```
begin
    $sd(V) :=$ standard deviation of $V$
    $sdid := |V(x) - V'(x)| \leq p \cdot sd(V)/100$
    return $\underline{sdid}$
end
```

Attribute disclosure for categorical data

- Categorical case: compare $V(x)$ and $V'(x)$
- Simple attribute disclosure risk ($SADR$).

$$SADR(X, X') = \frac{|\{x \in X | V(x) = V'(x)\}|}{|X|}.$$

Attribute disclosure

- Comparison between $V(x)$ and $V'(x)$ is right?
- If x is changed to x', *equality* $V(x) = V'(x)$ can be missleading

Attribute disclosure: model-based (categorical)

```
Algorithm 9 Model-based attr disclosure (for $x$): $mbd(C_{tr}, A, V, V', x)$.
  Data: $C_{tr}$: training set; $A$: algorithm to build a classification model;
  $V$: original attribute; $V'$: masked attribute; $x$: an example
  Result: attribute disclosure for $x$
  begin
    Take the dataset $C_{tr}$ as input and construct a model of the attribute
    $V$ using the algorithm $A$. Let this model be denoted by $M_{C_{tr},A}$
    $V'(x) := M_{C_{tr},A}(x)$. That is, the application of the classifier $M_{C_{tr},A}$
    to the data $x$
    $ad := (V'(x) = V(x))$
    return $\underline{ad}$
  end
```

Attribute disclosure: model-based (categorical)

```
Algorithm 10 Attribute disclosure for algorithm $A$: $adr(X, X', A)$.
  Data: $X$: original set; $X'$: masked data set; $A$: for classifiers
  Result: attribute disclosure risk
  begin
    $C_1, \dots, C_k$ a partition of $X$
    $C_i^{Tr}$ as in cross-validation. So, $C_i^{Tr} = \cup_{j \neq i} C_j$
        $adr_{A,C,X} := \frac{\sum_{i=1}^{k} |\{x | x \in C_i \text{ and } mbd(C_i^{Tr}, A, V, V', x)\}|}{|X|}.$
    Note that here $V(x)$ refers to the true value of attribute $V$ of $x$ (i.e., according to the original data set)
    return $\underline{adr_{A,C,X}}$
  end
```

Attribute disclosure: model-based (numerical)

- Numerical: Same as for categorical, but with intervals

$$ad := |V(x) - V'(x)| \leq p \cdot sd(V)/100$$

Attribute disclosure: absent attributes[1] (categorical)

```
Algorithm 11 Data-driven model-based attr. disclosure $ddmbar(X, X', A)$
  Data: $X$: original set; $X'$: masked data set; $A$: data mining algorithm
  Result: attribute disclosure risk in [0,1]
  begin
    Build a data-driven model $V$ from $X$ using algorithm $A$
    Build a data-driven model $V'$ from $X'$ using algorithm $A$
    Compute attribute disclosure as follows
        $$DDMBAR_A = \frac{|\{x|V(x) = V'(x')\}|}{|X|}$$
    return $\underline{DDMBAR_A}$
  end
```

[1] Attributes not in the database

Attribute disclosure: discussion

- Boundaries attribute disclosure / data of good quality: difficult
- Hints: (need to understand the data!)
  - A good estimate requiring a large number of input attributes may be ok, but a good estimate from a few can be problematic. Avoid good inferences from small number of attributes.

Attribute disclosure: discussion

- Boundaries attribute disclosure / data of good quality: difficult
- Hints: (need to understand the data!)
  - A good estimate requiring a large number of input attributes may be ok, but a good estimate from a few can be problematic. Avoid good inferences from small number of attributes.
  - Not all attributes are equally relevant. Confidential and sensitive attributes need to be identified & used for disclosure risk analysis.

Attribute disclosure: discussion

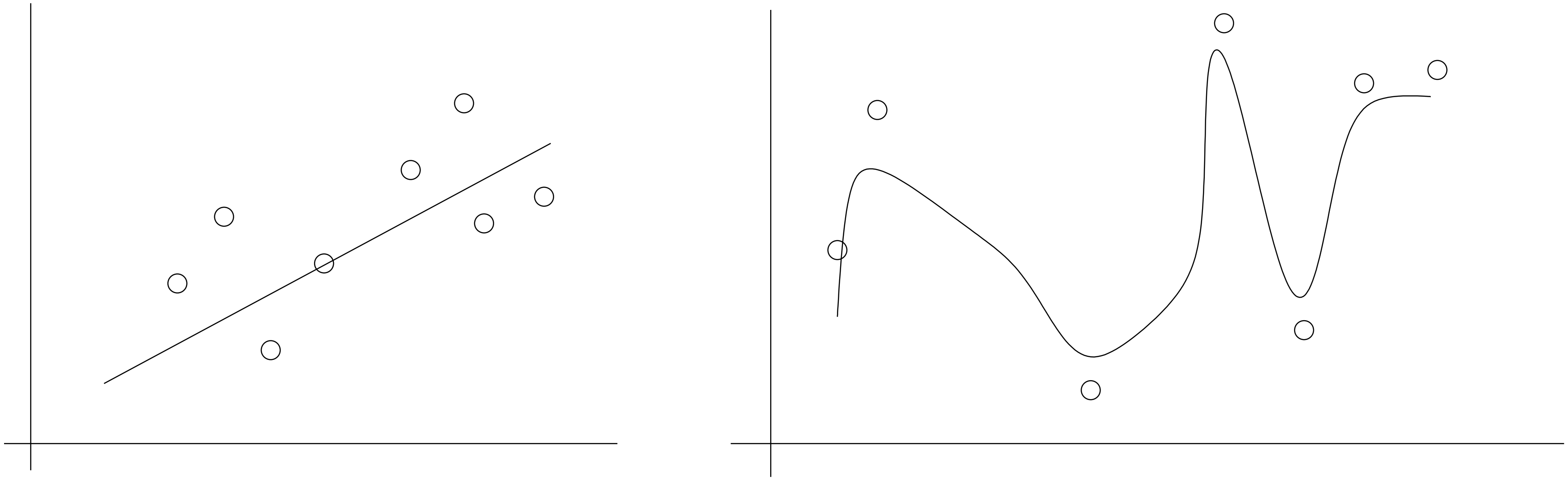
- Boundaries attribute disclosure / data of good quality: difficult
- Hints: (need to understand the data!)
  - A good estimate requiring a large number of input attributes may be ok, but a good estimate from a few can be problematic. Avoid good inferences from small number of attributes.
  - Not all attributes are equally relevant. Confidential and sensitive attributes need to be identified & used for disclosure risk analysis.
  - Not all information is always directly available in a database. It may be the case that sensitive information can be inferred from the database but it is not present there. E.g., observance of religious holidays or political orientation may not be an attribute in the database, but the database may contain enough information to estimate these attributes with good accuracy. Attribute disclosure analysis needs to consider whether these cases are relevant.

Attribute disclosure: discussion

- Another perspective:
  - In machine and statistical learning, models are expected to generalize data and avoid over-fitting. When a model generalizes correctly and there is no over-fitting, any inference for a particular individual x is due to general properties and not to its particularities. In contrast, bad generalization and over-fitting may imply that inferences are due to memorization and to learning particular features of certain records. When we require good data utility from a machine learning perspective, attribute disclosure should avoid detecting general information found in the data and focus on detecting these particular features of individuals.
  - This has connections with membership attacks, that, in short, try to detect records that are known to have been used in training a model, and they are detected because they are somehow distinguishable from more common ones.

Attribute disclosure: discussion

- If the model is a good generalization, (almost) coincidence may be ok.
- Compare:



Attribute disclosure: Through Membership Inference Attacks

- For data-driven models m
- Given x, was x in the training model?
- Is my data used to build m?
- Idea:
    - We build a classifier mia

Attribute disclosure: Membership Inference Attacks

Algorithm Model for membership inference attack: $mia(C_{tr}, A)$.

Data: D^i: data sets for building shallow models (each D^i partitioned into training and testing D_i^{tr}, D_i^{te}); A: algorithm to build shallow models

Result: Classifier for membership inference attack

```
begin
    $sm_i = A(S_i^{tr})$ for all $i = 1, \ldots, k$
    tuples $= \emptyset$
    for $i = 1, \ldots, k$ do
        forall the $x \in D_i^{tr}$ do
            $tuples = tuples \cup \{(x, sm_i(x), training)\}$
        end
        forall the $x \in D_i^{te}$ do
            $tuples = tuples \cup \{(x, sm_i(x), no - training)\}$
        end
    end
    $mia =$ build-classifier($tuples$)
    return mia
end
```

Attribute disclosure: Membership Inference Attacks

- Once we have the mia classifier, we define the membership inference attack attribute disclosure risk as

$$miaAR = \frac{|\{x|mia(x) = training\}|}{|X|}$$

(correctly identified members)

- In general, we can use performance measures (recall, precision, F1-score)

Risk measures for identity disclosure

Identity disclosure

- Re-identification. Estimation of correct re-identifications. Theoretically or empirically.
- Uniqueness. Probability that rare combinations in the protected data are also rare in the population.

Identity disclosure: Uniqueness

- **Uniqueness.** Risk is defined as the probability that rare combinations of attribute values in the protected data set are indeed rare in the original population.

Identity disclosure: Uniqueness

- **Uniqueness.** Risk is defined as the probability that rare combinations of attribute values in the protected data set are indeed rare in the original population.
  - Suitable for sampling ($\rho(X)$ is a subset of X).
  - For masked data, the same combination will not appear.

Identity disclosure: Uniqueness

Measures for identity disclosure: Uniqueness (categorical data/sampling)

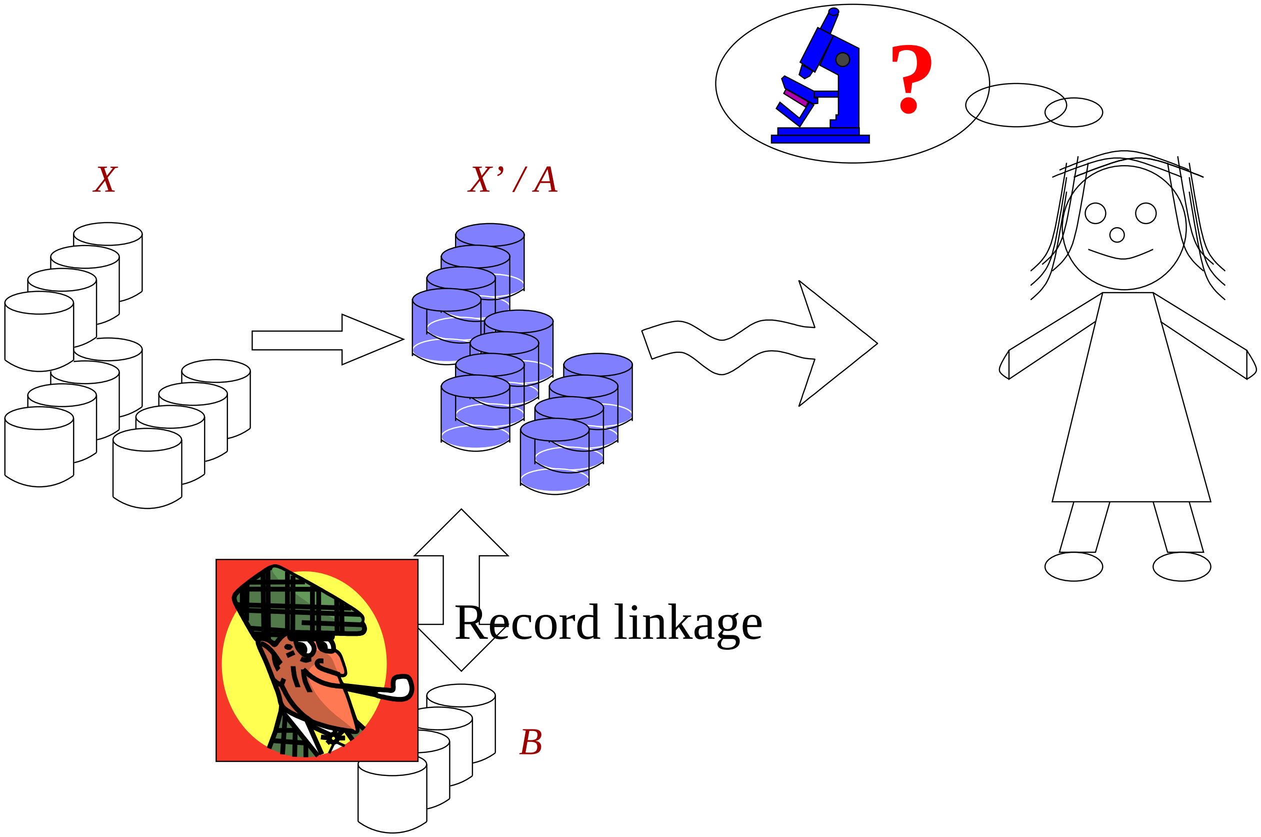
- **File-level uniqueness.** It is defined as the probability that a sample unique (SU) is a population unique (PU). The following expression has been used:

  $$P(PU|SU) = \frac{P(PU, SU)}{P(SU)} = \frac{\sum_j I(F_j = 1, f_j = 1)}{\sum_j I(f_j = 1)}$$

  where $j = 1, \ldots, J$ denotes possible values in the sample, F_j is the number of individuals in the population with key value j (frequency of j in the population), f_j is the same frequency for the sample and I stands for the cardinality of the selection.
- **Record-level risk uniqueness.** It is defined as the probability that a particular sample record is re-identified (recognized as corresponding to a particular individual in the population).

Identity disclosure

- **Privacy from re-identification**. Identity disclosure. Scenario:
  - A: File with the protected data set
  - B: File with the data from the intruder (subset of original X)



Identity disclosure

A scenario for identity disclosure: $X = id||X_{nc}||X_c$

- Protection of the attributes
  - **Identifiers.** Usually removed or encrypted.
  - **Confidential.** X_c are usually not modified. $X'_c = X_c$.
  - **Quasi-identifiers.** Apply masking method ρ. $X'_{nc} = \rho(X_{nc})$.

Identity disclosure

- **Privacy from re-identification**. Identity disclosure.
  - ○ A: File with the protected data set
  - ○ B: File with the data from the intruder (subset of original X)

How to establish the correct links between the two files?
Record linkage algorithms (used in e.g. database integration)

Identity disclosure

- **Privacy from re-identification**. Identity disclosure.
  - A: File with the protected data set
  - B: File with the data from the intruder (subset of original X)

  How to establish the correct links between the two files?
  Record linkage algorithms (used in e.g. database integration)

- Two main types.
  - Distance-based record linkage
  - Probabilistic record linkage

Identity disclosure

- **Re-identification**. Given $A = X' = \rho(X)$ and $B \subset X$, a measure:

$$Reid(B, A) = \frac{\sum_{b \in B} c(r(b), true(b))}{|B|}. \qquad (1)$$

where

- $true : B \to A$, for each record b (of the intruder) returns the correct record for re-identification,
- $r : B \to A$, models the re-identification algorithm.
  Note: In order to make the definition general, we consider that r returns a probability distribution on A. That is, given a record b in B, it assigns to each record a in A a probability of matching.
- c a function, with $c(r(b), true(b))$ we evaluate the result for each record in $[0, 1]$.

Identity disclosure

- **Re-identification**. Examples for $Reid(B, A)$
  - Algorithm assigns to each record b a single record a_b in A

$$Reid_d(B, A) = \frac{\sum_{b \in B} |\{b | a_b = true(b)\}|}{|B|}.$$

  - Algorithm assigns to each record b an anonymity set in A denoted by A_b , we model the re-identification algorithm by the distribution $r(b)[a'] = 1/|A_b|$ for all $a' \in A_b$, and $r(b)[a'] = 0$ for all $a' \notin A_b$. Then, we define $c(r(b), true(b))$ as $r(b)[true(b)]$. Naturally, we will have that $c(r(b), true(b)) = r(b)[true(b)] \leq 1/|A_b|$. Then,

$$Reid_k(B, A) = \frac{\sum_{b \in B} r(b)[true(b)]}{|B|}.$$

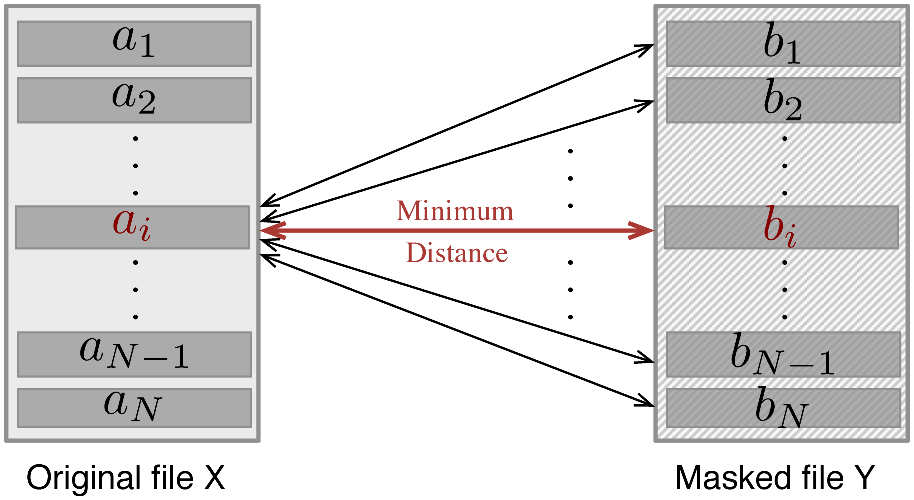
Identity disclosure

- **Re-identification**. Can intruders distinguish correct links from incorrect links?

$$K.Reid(B, A) = \frac{|K|}{|B|} \quad (2)$$

Identity disclosure

- Distance-based record linkage: $d(a,b)$ with $a \in A$ and $b \in B$.
  - Assign to the record at a minimum distance, ideally an intruder wants for a record i: $d(a_i, b_j) \geq d(a_i, b_i)$ for all j
    but due to masking we expect this does not happen



Identity disclosure

- **Algorithm** Distance-based record linkage

```
Input $A$: file; $B$: file
Output $LP$: linked pairs; $NP$: non-linked pairs
For $a \in A$
    b' $= \text{arg } \min_{b \in B} d(a, b)$
    $LP = LP \cup (a, b')$
    for  $b \in B$ such that $b \neq b'$
        $NP := NP \cup (a, b)$
    end for
end for
Return $(LP, NP)$
```

Identity disclosure

- Probabilistic record linkage: $d(a, b)$ with $a \in A$ and $b \in B$.
  - ○ Classification of pairs of records (a, b) in 3 classes
    Linked pair, non-linked, clerical pair
  - ○ How?
    - ▷ For each pair (a, b), an index is computed using the conditional probabilities
      - ⋆ $P(coincidence|Matching)$: coincidence between both records when there is matching
      - ⋆ $P(coincidence|Unmatching)$: coincidence between both records when there is no matching
    - ▷ Classification using thresholds

Identity disclosure

- Probabilistic record linkage: $d(a, b)$ with $a \in A$ and $b \in B$.
  - Computation of

    $P(coincidence|Matching)$ and

    $P(coincidence|Unmatching)$:

Identity disclosure

- Probabilistic record linkage: $d(a, b)$ with $a \in A$ and $b \in B$.
  - Computation of
    $P(coincidence|Matching)$ and
    $P(coincidence|Unmatching)$:
    - ▷ Using EM algorithm

Identity disclosure

- Probabilistic record linkage: $d(a, b)$ with $a \in A$ and $b \in B$.
  - Computation of
    $P(coincidence|Matching)$ and
    $P(coincidence|Unmatching)$:
    - ▷ Using EM algorithm
  - Computation of thresholds

Identity disclosure

- Probabilistic record linkage: $d(a, b)$ with $a \in A$ and $b \in B$.
  - Computation of
    $P(coincidence|Matching)$ and
    $P(coincidence|Unmatching)$:
    - ▷ Using EM algorithm
  - Computation of thresholds
    - ▷ From the probabilities of false positive/negative
      $P(Linkedpair|Unmatching)$
      $P(Nonlinkedpair|Matching)$

Identity disclosure

A scenario for identity disclosure. Reidentification

- Flexible scenario for identity disclosure
  - A protected file using a masking method
  - B (intruder's) is a subset of the original file.

Identity disclosure

A scenario for identity disclosure. Reidentification

- Flexible scenario for identity disclosure
  - A protected file using a masking method
  - B (intruder's) is a subset of the original file.
    $\rightarrow$ intruder with information on only some individuals

Identity disclosure

A scenario for identity disclosure. Reidentification

- Flexible scenario for identity disclosure
  - ○ A protected file using a masking method
  - ○ B (intruder's) is a subset of the original file.

    → intruder with information on only some individuals

    → intruder with information on only some characteristics

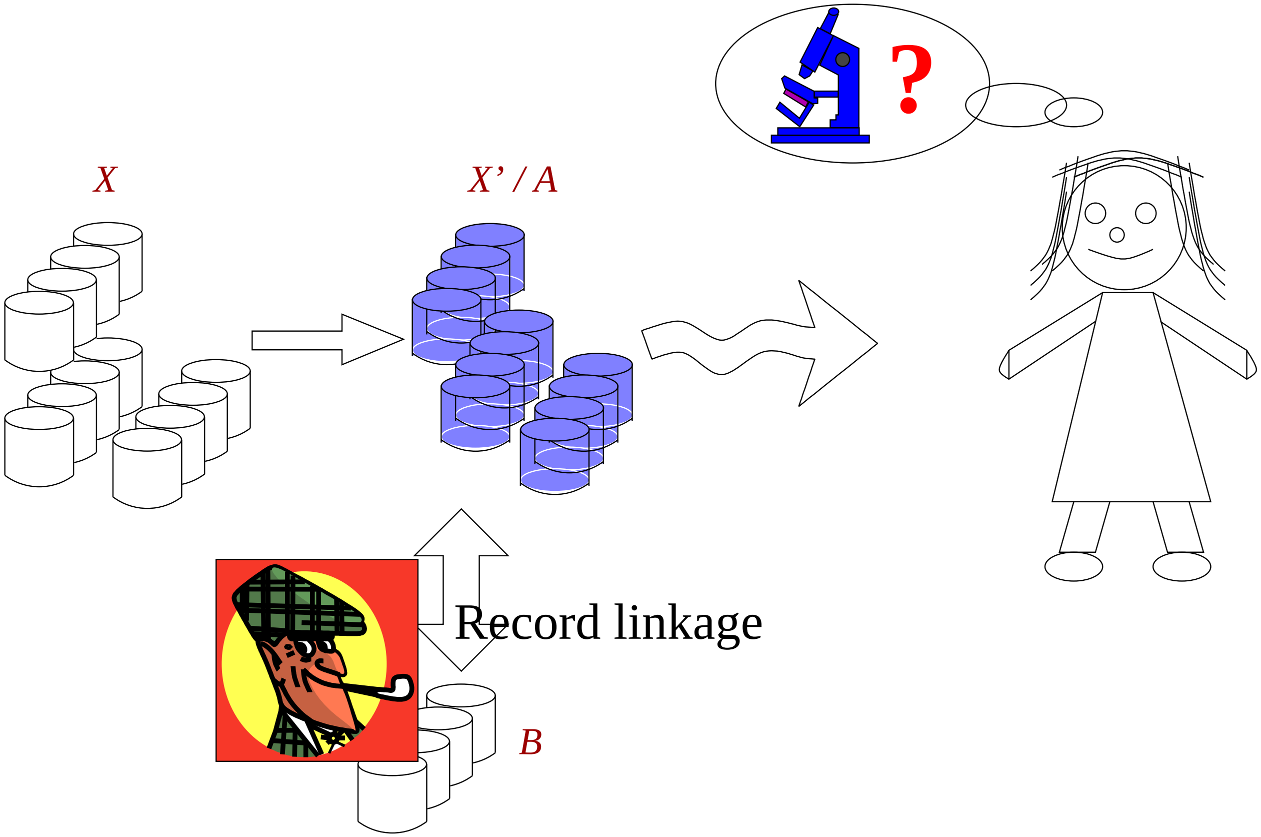
Identity disclosure

A scenario for identity disclosure. Reidentification

- Flexible scenario for identity disclosure
  - ◦ A protected file using a masking method
  - ◦ B (intruder's) is a subset of the original file.
    $\rightarrow$ intruder with information on only some individuals
    $\rightarrow$ intruder with information on only some characteristics
  - ◦ But also,
    - ▷ B with a schema different to the one of A (different attributes)
    - ▷ Other scenarios. E.g., synthetic data
    - ▷ Other type of data: graph data
      (reidentifying people in a social network)

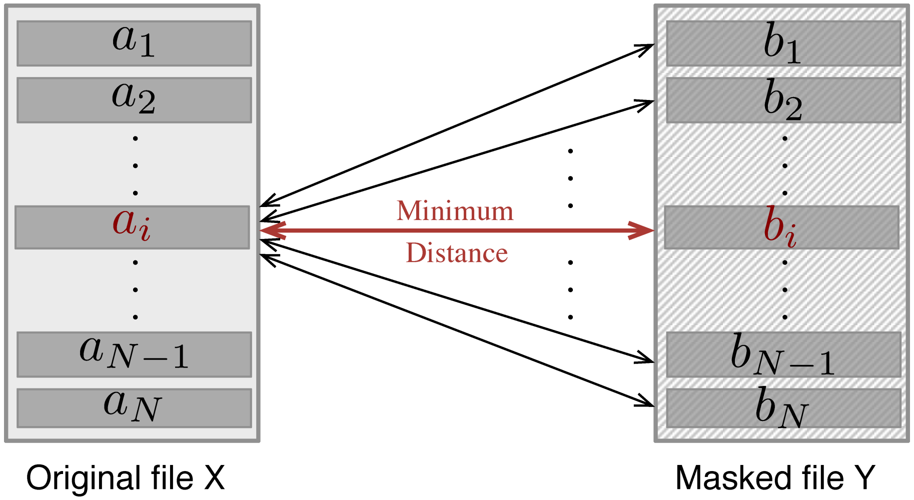
Identity disclosure

- **Privacy from re-identification**. Worst-case scenario (maximum knowledge) to give upper bounds of risk:
  - transparency attacks (information on how data has been protected)
  - largest data set (original data)
  - best re-identification method (best record linkage/best parameters)



Identity disclosure

- **Privacy from re-identification**. Worst-case scenario.
  - ML for distance-based record linkage parameters. (A and B aligned)
    - Goal: as many correct reidentifications as possible: for each record i: $d(a_i, b_j) \geq d(a_i, b_i)$ for all j



- 🟥 $d(a_i, b_j)$ as average/sum of attribute/variable distances

$$\mathbb{C}_p(\mathit{diff}_1(a_i, b_j), \ldots, \mathit{diff}_n(a_i, b_j))$$

Identity disclosure

- **Privacy from re-identification**. Worst-case scenario.
  - ML for distance-based record linkage parameters. (A and B aligned)
    - ▷ Goal: as many correct reidentifications as possible. But, if error for a_i: $K_i = 1$ and $d(a_i, b_j) + CK_i \geq d(a_i, b_i)$ for all j

      where d is an aggregated distance $d(a, b) = \mathbb{C}_p(\mathit{diff}_1, \ldots, \mathit{diff}_n)$:
    - ▷ Formally,

$$\mathbb{C}_p(\mathit{diff}_1(a_i, b_j), \ldots, \mathit{diff}_n(a_i, b_j)) + CK_i \geq \mathbb{C}_p(\mathit{diff}_1(a_i, b_i), \ldots, \mathit{diff}_n(a_i, b_i))$$

Identity disclosure

- **Privacy from re-identification**. Worst-case scenario.
  - ML for distance-based record linkage parameters. (A and B aligned)
  - Goal: as many correct reidentifications as possible.
  - Minimize K_i: minimize the number of records a_i that fail
- Formalization:

$$
\begin{aligned}
&Minimize \sum_{i=1}^{N} K_i \\
&Subject\ to: \\
&\qquad \mathbb{C}_p(\mathit{diff}_1(a_i, b_j), \ldots, \mathit{diff}_n(a_i, b_j)) - \\
&\qquad\qquad - \mathbb{C}_p(\mathit{diff}_1(a_i, b_i), \ldots, \mathit{diff}_n(a_i, b_i)) + CK_i > 0 \\
&\qquad K_i \in \{0, 1\} \\
&\qquad \text{Additional constraints according to } \mathbb{C}
\end{aligned}
$$

Identity disclosure

- **Privacy from re-identification**. Worst-case scenario.
  - ML for distance-based record linkage parameters. (A and B aligned)
  - The case of the weighted mean ($\mathbb{C} = WM$)/Weighted Euclidean
  - Formalization:

$$d^2(a,b) = WM_p(\mathit{diff}_1(a,b), \ldots, \mathit{diff}_n(a,b))$$

with arbitrary vector $p = (p_1, \ldots, p_n)$ and
$\mathit{diff}_i(a,b) = ((a_i - \bar{a}_i)/\sigma(a_i) - (b_i - \bar{b}_i)/\sigma(b_i))^2$

Identity disclosure

- **Privacy from re-identification**. Worst-case scenario.
  - ML for distance-based record linkage parameters. (A and B aligned)
  - The case of the weighted mean ($\mathbb{C} = WM$)
  - Formalization:

$$
\begin{aligned}
Minimize\ & \sum_{i=1}^{N} K_i \\
Subject\ to: & WM_p(\mathit{diff}_1(a_i,b_j),\ldots,\mathit{diff}_n(a_i,b_j)) - \\
& \quad - WM_p(\mathit{diff}_1(a_i,b_i),\ldots,\mathit{diff}_n(a_i,b_i)) + C\ K_i > 0 \\
& K_i \in \{0,1\} \\
& \sum_{i=1}^{n} p_i = 1 \\
& p_i \geq 0
\end{aligned}
$$

Identity disclosure

- **Privacy from re-identification**. Worst-case scenario.
  - ML for DBRL parameters: Distances considered $\mathbb{C}$
    - ▷ Weighted mean.
      Weights: importance to the attributes
      Parameter: weighting vector $n = \#$ attributes

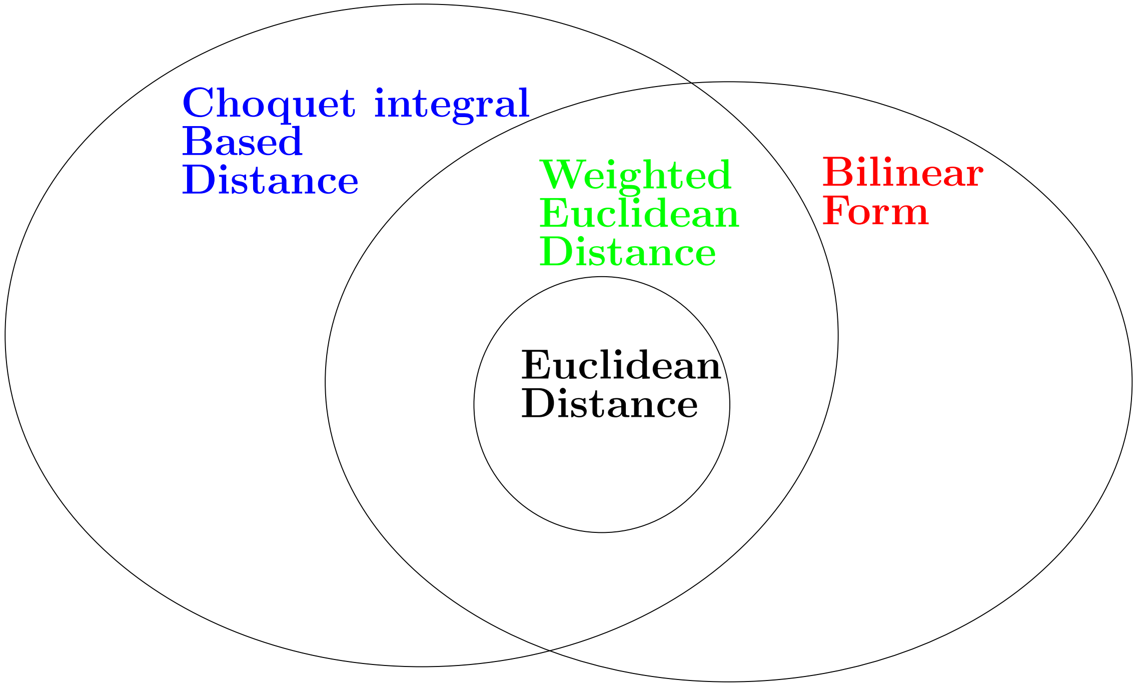
Identity disclosure

- **Privacy from re-identification**. Worst-case scenario.
  - ML for DBRL parameters: Distances considered $\mathbb{C}$
    - ▷ Weighted mean.
      Weights: importance to the attributes
      Parameter: weighting vector $n = \#$ attributes
    - ▷ OWA - linear combination of order statistics (weighted):
      Weights: to discard lower or larger distances
      Parameter: weighting vector $n = \#$ attributes
    - ▷ Bilinear form - generalization of Mahalanobis distance
      Weights: interactions between pairs of attributes
      Parameter: square matrix: $n \times n$ ($n = \#$ attributes)
    - ▷ Choquet integral.
      Weights: interactions of sets of attributes ($\mu : 2^X \to [0, 1]$)
      Parameter: non-additive measure: $2^n - 2$ ($n = \#$ attributes)

Identity disclosure

Distances used in record linkage based on aggregation operators

- Graphically



Bilinear form. Quadratic form that generalizes Mahalanobis distance. Choquet integral. A fuzzy integral w.r.t. a fuzzy measure (non-additive measure). CI generalizes Lebesgue integral. Interactions.

References

References

- Torra, V. (2022) Guide to data privacy, Springer.
- Shokri, R., Stronati, M., Song, C., Shmatikov, V. (2017) Membership inference attacks against machine learning models, arXiv:1610.05820v2.
- Winkler, W. E. (2004) Re-identification methods for masked microdata, Proc. PSD 2004, LNCS 3050 216-230.
- Elliot, M. (2002) Integrating file and record level disclosure risk assessment, in J. Domingo-Ferrer, Inference Control in Statistical Databases, LNCS 2316 126-134.

Thanks