

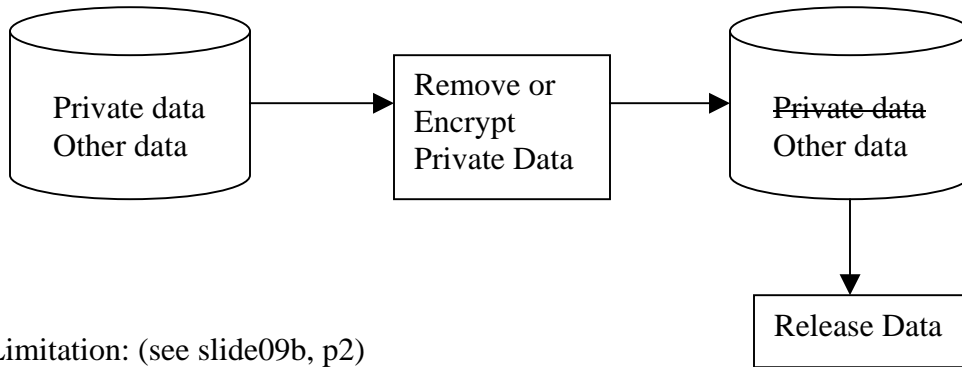
Generalizing Data to Provide Anonymity when Disclosing Information (1)

Topics:

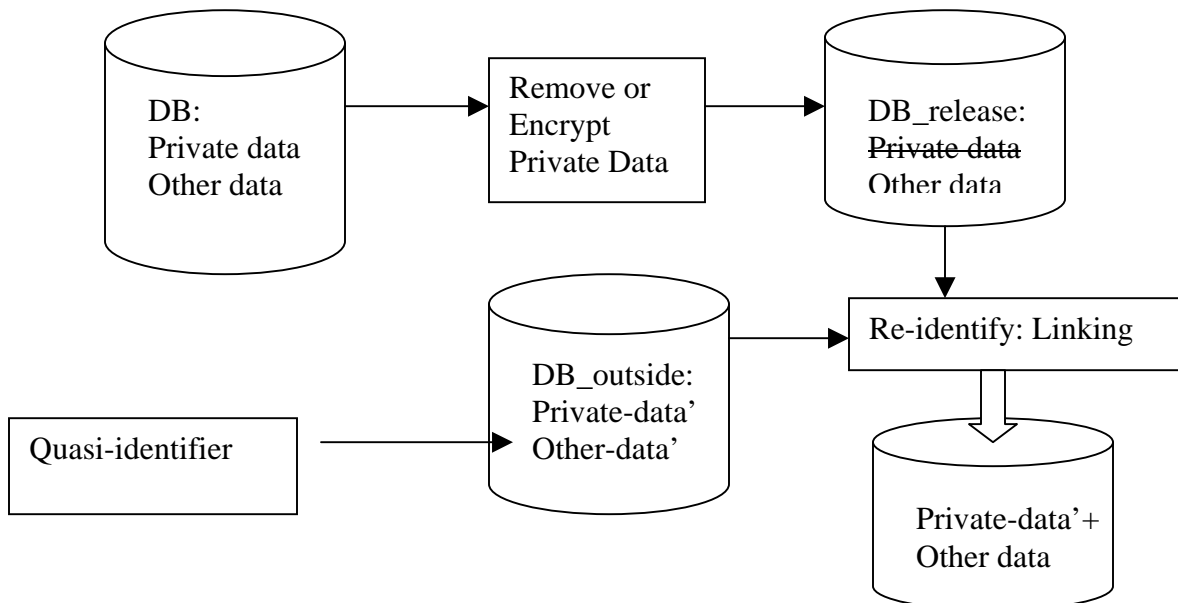
1. Problem description
 - a. How to formally define a problem?

Topic 1: Problem Description

- Motivation:
 - Protect individual-specific (private) data
 - e.g., name, address, phone number, SSN
 - Solution 1: no release of data
 - **Question 1:** Limitations – usage, business and research need shared info.
 - Story: IRB - INSTITUTIONAL REVIEW BOARD for RESEARCH INVOLVING HUMAN SUBJECTS
 - Solution 2: remove/encrypt private data -> anonymous info.



- Limitation: (see slide09b, p2)
 - **Re-identifying** anonymous data
 - Link to outside data (public data)
 - e.g., voting list - use DOB (12%); DOB+gender (29%); DOB+Zip (69%)



- Formal Definition. <- derived from motivational examples
 - From the example: DOB is the quasi-identifier:
 - **Exercise** (5 min): define quasi-identifier.
 - **My version 1:** $DB_release = DB_de-identified = PT$ (or Private Table)
 - $DB = (v_1, v_2, \dots, v_n, u_1, u_2, \dots, u_m)$
 - (v_1, \dots, v_n) are private attributes; (u_1, \dots, u_m) are public attributes
 - $DB_release = (u_1, \dots, u_m)$
 - $DB_outside = (subsetof(v_1, \dots, v_n), subsetof(u_1, \dots, u_m))$
 - Link $DB_outside(subsetof(u_1, \dots, u_m))$ with $DB_release(u_1, \dots, u_m)$ to obtain DB
 - **My version 2:** $DB_release = DB_de-identified = PT$ (or Private Table)
 - $DB = (v_1, v_2, \dots, v_n, u_1, u_2, \dots, u_m)$
 - (v_1, \dots, v_n) are private attributes; (u_1, \dots, u_m) are public attributes
 - $DB_release = (u_1, \dots, u_m)$
 - $DB_outside = (subsetof(v_1, \dots, v_n), subsetof(u_1, \dots, u_m))$
 - $DB_individual = Revover(DB_outside, DB_release, u_i)$
- Example: public attributes: $u_1 = DOB$, $u_2 = gender$, $u_3 = zip$
 u_1, u_2 , and u_3 in $DB_outside$ are linked to u_1, u_2 , and u_3 in DB.