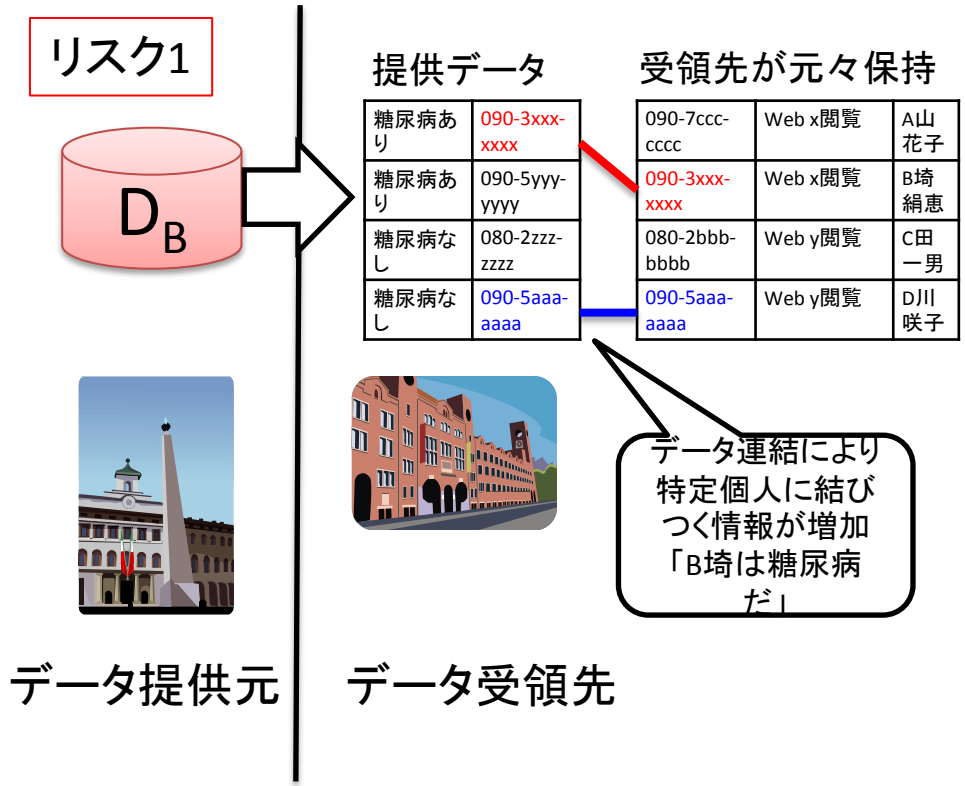
資料１－４

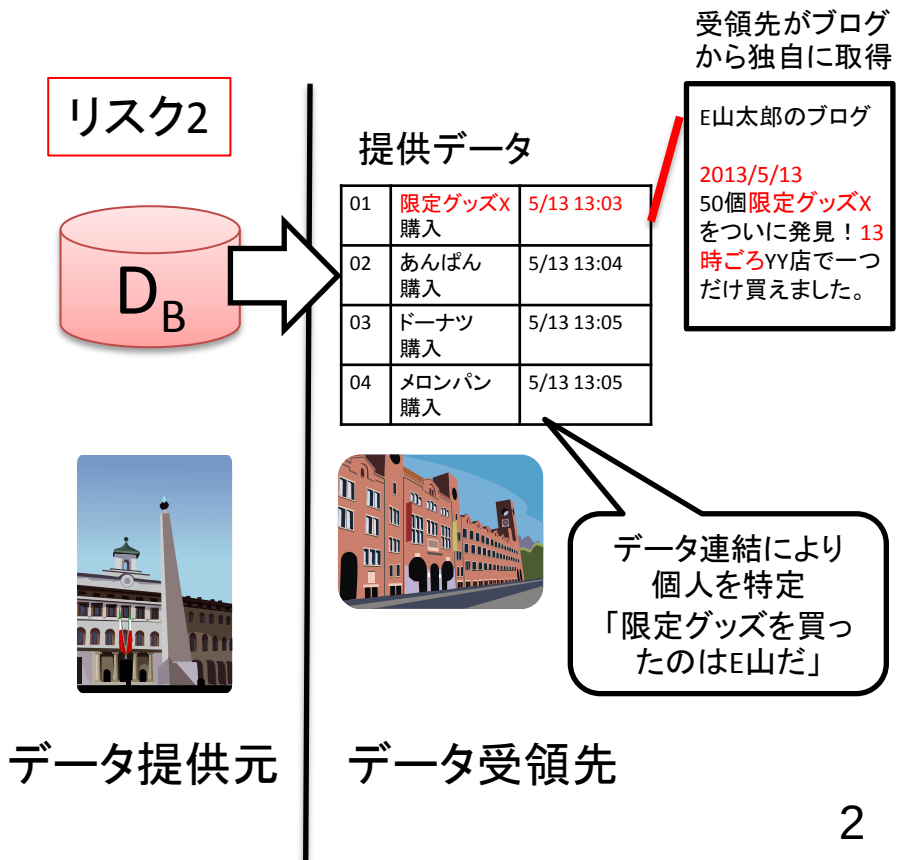
「準個人情報」及び「個人特定性低減データ」について

筑波大学 コンピュータサイエンス専攻
佐久間 淳

パーソナルデータ提供による特定のリスク

- リスク1. 識別情報を経由した特定
- リスク2. 識別情報を経由しない特定





共有可能性を持つ識別情報の取り扱いが重要

- リスク1(識別情報を経由する特定)を事前に防ぐことが必要
- 独立に収集された特定個人の情報が、事後的に結合することを防ぐためには?
  - 提供元、受領先が独立に収集可能(共有可能性)な個人を一意に識別する情報 は、事後的な結合を可能にする
  - 共有可能性を持つ識別情報を提供データから取り除く
- 共有可能性を持つ識別情報の例
  - 本人から取得：電話番号、メールアドレス、指紋、etc.
  - 本人の所有物から取得：macアドレス、位置情報(スマホ経由), etc.
- リスク2(識別情報を経由しない特定)を防ぐことは現実的に不可能
  - 事後的に被害を救済する措置が必要

「個人特定性低減データ」はどうあるべきか

- 「個人特定性低減データ」が提供されたことによって、受領先で、機械的かつ大規模な名寄せが発生しないこと
- 問題意識1. 識別情報を鍵として、多くの個人が機械的に特定されるべきではない
  - 共有可能性のある識別情報は個人特定性低減データから排除されるべき
- 問題意識2.識別情報を鍵とせず、少数の個人が偶然特定されることを防ぐことは技術的に非常に困難
  - このような特定リスクは個人特定性低減データでは考慮せず、被害があれば事後的に被害を救済する措置が必要

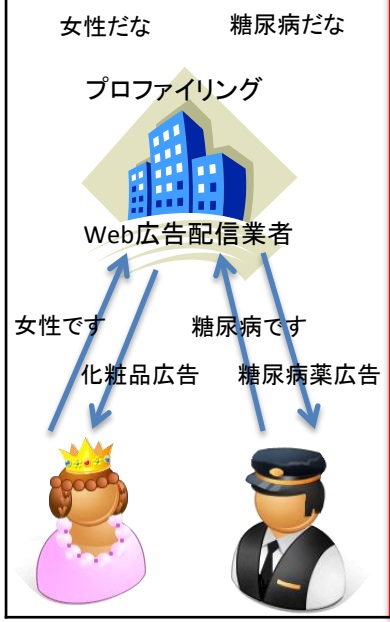
履歴情報による特定

- 個別の履歴は必ずしも個人を特定しないが、複数集まると個人を特定することができる可能性が高い
- 電話番号やメールアドレスと異なり、それ自体がデータ解析の興味対象
  - 個人特定のリスクがあるからといって、一律に削除することは本来の趣旨から外れる
- 履歴情報に関する「個人特定性低減データ」作成には、特定が困難であるよう、抽象化や履歴長の制限が必要
- 履歴情報を完全に特定不可能であるように加工することは技術的には相当に困難
  - 抽象化や履歴長の制限を行ったとしても、低い確率で名寄せは発生しうると考え、事後的な救済策を用意すべき

プロファイリングにおけるパーソナル情報利用の類型化

プライバシ・機微情報の直接利用	プライバシ・機微情報のプロファイリングによる直接利用	プライバシ・機微情報のプロファイリングによる直接利用(受領データあり)	プライバシ・機微情報のプロファイリングによる間接利用(受領データあり)
個人から属性情報を得て、利用して広告	個人からWeb視聴履歴を得て、そこから直接属性情報を推定して広告	個人からWeb視聴履歴を得て、それと、属性情報を含む別のWeb視聴履歴を合わせて推定して広告	個人からWeb視聴履歴を得て、それと、属性情報を含まない別のWeb視聴履歴を含めて推定して広告
第三者データ利用無し 属性取得あり 属性推定あり(profiling)	第三者データ利用無し 属性取得なし 属性推定あり(profiling)	第三者データ利用あり 属性取得なし 属性推定あり(profiling)	第三者データ利用あり 属性取得なし 属性推定なし(profiling)

属性情報直接利用

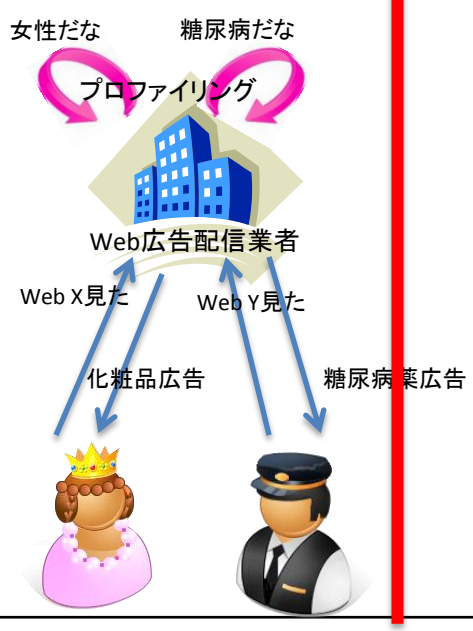


属性情報推定利用

属性情報推定無し

第三者からの受領データ

Web x	女性	糖尿病あり
Web x	女性	糖尿病あり
Web y	男性	糖尿病なし
Web y	女性	糖尿病なし



第三者からの受領データ　　広告配信業者はここは受領していない

Web x	Web a	女性	糖尿病なし
Web x	Web a	女性	糖尿病なし
Web y	Web b	男性	糖尿病あり
Web y	Web b	女性	糖尿病あり