多様な情報価値を共創する
情報大航海プロジェクト

平成21年度情報大航海プロジェクト
共通技術開発
「パーソナル情報保護・解析基盤」

情報大航海プロジェクトにおけるプライバシ保護

山口(繁富)利恵

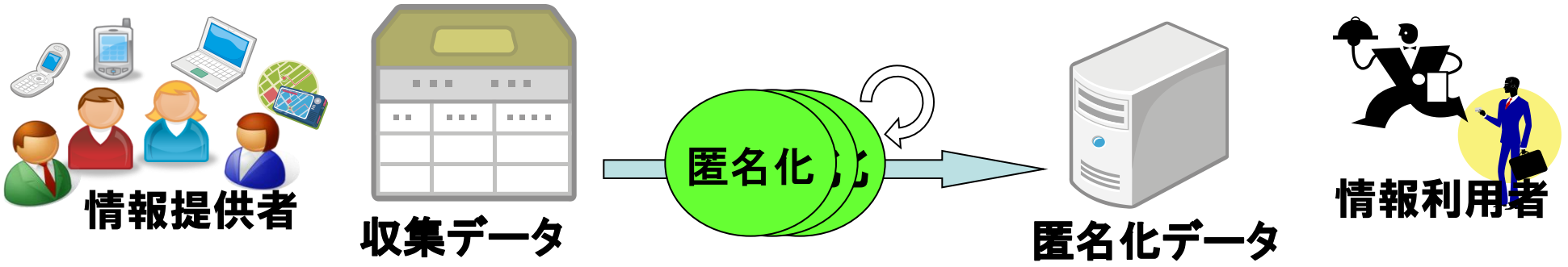
財団法人日本情報処理開発協会　株式会社三菱総合研究所　独立行政法人産業技術総合研究所
日本電信電話株式会社　NTTソフトウェア株式会社　ジェイエムテクノロジー株式会社　有限責任中間法人PUCC
NTTアドバンステクノロジ株式会社　ビジネスマイニング研究センター有限責任事業組合

匿名性のアプローチ

多様な情報価値を共創する
情報大航海プロジェクト

情報大航海における匿名化のアプローチ

- 情報を収集し、その情報を利活用する際のプライバシ保護の**取り扱い基準**を考えましょう
  - この基準に沿えば、利用者（情報提供者）に対し、**安心を提供**できるような枠組みを。
  - 情報大航海の中では、あくまでも**集合匿名化**に特化して考察。
- 集合匿名化の基準として、k 匿名性を採用
  - 技術的に評価をするためには、k 匿名性が**現状取り得ることのできる基準として、適切である**
  - k 匿名性を満たしたデータであれば，個人情報が保護されたこととする
- この指標を適用するデータに適用するのかを**基盤利用マニュアル**で示す
- k 匿名性を満たした情報を実現する枠組みの検討（**確認・認証の枠組み**）が今後必要



基準の定義

多様な情報価値を共創する
情報大航海プロジェクト

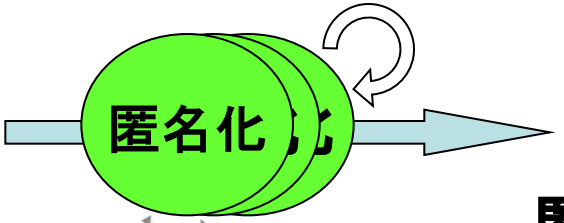
- 母集団一意性を満たすために技術基準が必要
  - ここでは、「ものさし」として，k 匿名性を利用
- 匿名化データをどのように実現するのか
  - k 匿名化を満たすために、複数の手法が存在し、その中でも、k 匿名化手法を選択
  - 実現するためには、安全性の整理も必須

匿名化したあとのデータの性質についての基準

上の基準を満たすためにどのような技術を利用すべきかを考える

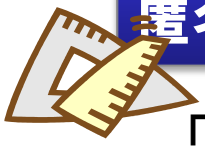
情報提供者

収集データ



匿名化データ

情報利用者



匿名化基準の必要性

「ものさし」としての基準が必要

- ここでは、「ものさし」として、k 匿名性を利用。

実現方法の選定

匿名化データをどのように実現するのか

- k 匿名性を満たすために、複数の手法が存在し、その中でも、k 匿名化手法を選択。

評価・認証

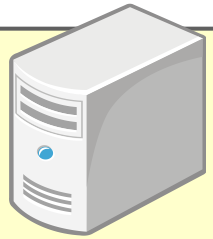
正しく匿名化データを作成したのか確認

- 正しくものさしをつかい、実現して、k 匿名性を満たしたデータをつくったかどうかの評価を行い、認証など、確認が必要。

k 匿名化の実現状況

多様な情報価値を共創する
情報大航海プロジェクト

- k 匿名性を満たしている情報にもいろいろな段階がある.
  1. 匿名化データをどのように準識別子の取り方を考えても, k 匿名性を満たしている状態
  2. ある準識別子に注目すると, k 匿名性を満たしている状態
  3. 技術的には、2.と同義となる場合もあるが、一般的に、個人情報保護法において、個人情報を保護した情報であるという状態



匿名化データ

1. どの準識別子だとしても大丈夫

18000**	男	3*
18000**	男	3*
18000**	男	3*
18100**	女	4*
18100**	女	4*
18100**	女	4*
18000**	男	50以上
18000**	男	50以上
18000**	男	50以上
18000**	男	50以上

2. ある準識別子に注目すると、k 匿名性を満たしている

郵便番号	性別	年齢	趣味
18000**	男	3*	アニメ
18000**	男	3*	アニメ
18000**	男	3*	アニメ
18100**	女	4*	映画
18100**	女	4*	アニメ
18100**	女	4*	ドラマ
18000**	男	50以上	映画
18000**	男	50以上	ドラマ
18000**	男	50以上	ドラマ
18000**	男	50以上	時代劇

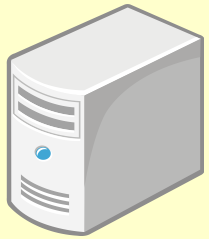
3. 識別子のみ削除してあるだけで、k 匿名性を一切満たしていない

勤務地	年齢	職業	年収
東京都港区西新橋2丁目	33	船乗り	1,000万円
東京都千代田区霞が関1丁目	21	放送局	2,000万円

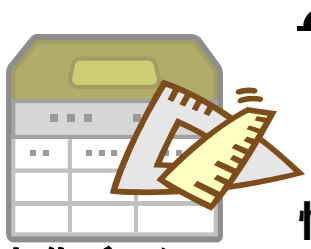
この条件を満たす人は一人しかいないが、3つの準識別子に注目すると、3-匿名性を満たしている

匿名化データの作り方

- k 匿名性を満たしたデータを定義したとしても、一意には決まらない
- 準識別子の取り方には複数の種類がある。
  - ものさしの当て方も、一意には決まらない
- 情報利用者にとって、それぞれに必要な情報が違うので、
  有用な匿名化方法を利用者自身が見つけなければならない。



匿名化データ



収集データ

情報利用者

どんな風にものさしを
あてようか？

同じデータを利用したとしても、有用性が変わってくるので、
利用者にとって、必要な取り方をしなければならない

郵便番号	性別	年齢	趣味
18000**	男	3*	アニメ
18000**	男	3*	アニメ
18000**	男	3*	アニメ
18100**	女	4*	映画
18100**	女	4*	アニメ
18100**	女	4*	ドラマ
18000**	男	50以上	映画
18000**	男	50以上	ドラマ
18000**	男	50以上	ドラマ
18000**	男	50以上	時代劇

趣味	郵便番号	性別	年齢
アニメ	1800001	男	31
アニメ	1800001	男	33
アニメ	1800030	男	35
アニメ	1810045	女	45
ドラマ	1800030	男	53
ドラマ	1800045	男	62
ドラマ	1810045	女	46
映画	18100**	女	4*
映画	18000**	男	50以上
時代劇	18000**	男	50以上

このデータは利用できない

基準を満たした実現方法の選択と今後の課題

多様な情報価値を共創する
情報大航海プロジェクト

- 匿名化データを実現するためにはどのようにしたらよいのか
  - 経験的なやり方として、名前・住所など個人を特定しうる情報を**削除**という手法は既に存在。
  - 情報を集めないような手法(暗号を利用した検索手法や、匿名認証手法など)は、今回の範囲外。 ← まず、**情報を収集している**、というところからスタート

- *k* 匿名性のデータを実現するための手法としては、様々存在する
  - 匿名化データを定義したとしても、実現手法は、**一つには決まらない**。
  - *k* 匿名化か母集団一意におけるリスク評価にするか
    - ここでの、*k* 匿名化か母集団一意か、というのは評価の軸(ものさし)としてではなく、**実現手法**なので、注意

→ 現状では、ある特定サービスについての二次利用処理については、
***k* 匿名化が適切**

- 正しく匿名化を行ったかどうかを確認する認証・制度が必要
  - ものさしを正しくあてることができたのか。
  - あてたものさしを正しく実現したのか。

→ 今後の課題として、認証制度などの**評価の枠組み**の検討が必要