（パーソナル）データ利活用におけるセキュリティ

明治大学

菊池　浩明

菊池 浩明Hiroaki Kikuchi

- 明治大学総合数理学部 教授
  - 大学院教務主任 / 情報基盤本部 副本部長
  - 理化学研究所革新知能統合研究センター客員研究員
  - JPCERT/CC代表理事
- 略歴
  - 明治大学卒，富士通研究所，東海大学情報通信学部教授、カーネギーメロン大学計算機科学科訪問研究員などを経て，2013年より明治大学
  - 電子情報通信学会 ICSS研究会専門委員会顧問
  - 情報処理学会 CSEC研究会顧問
  - 情報ネットワーク法学会理事
- 外部委員
  - JISAプライバシーマーク審査委員会・委員長
  - NICT映像センサー使用大規模実証実験検討委員会・委員長
  - 内閣官房IT戦略室　パーソナルデータに関する検討会技術検討WG・構成員
  - 内閣府　AMED・医療情報基盤担当室　　政策参与

報道

HOME | 脅威 | 脆弱性 | 情報漏えい | インシデント | 不正アクセス | 報告書 | 講演 | 業界

› 個人情報漏えい › 個人情報漏えい › 不正DDoS攻撃 › ACCS アク

メールアカウントへの不正アクセスが再発、個人情報流出だけでなく迷惑メール送信の踏み台にも（明治大学）

明治大学は10月24日、同学が発行するメールアカウント3件に不正アクセスがあり、該当アカウントからの迷惑メールの送信及びメール送受信データの一部がダウンロードされ同学学生をはじめとする個人情報の流出が判明したと発表した。

明治大学は10月24日、同学が発行するメールアカウ
に不正アクセスがあり、該当アカウントからの迷惑
送信及びメール送受信データの一部がダウンロード
学学生をはじめとする個人情報の流出が判明したと
た。

1件目は7月28日に、同学専任教員1名のメールアカ
第三者から不正アクセスを受け、該当アカウントか
5,677件の迷惑メールが送信され、また該当アカウン
受信データと添付書類がダウンロードされ、個人情
部に流出したことが判明したというもの。

(。・ω・)ゞ ！！１１月末迄 ScanNetSecurity創刊２０周年

2018-10-30

明治大学の不正アクセスはまた続く可能性が高い

つぶやいてみた。 | 不正アクセス事件

明治大学が再び不正アクセスを受けた可能性があると記事を書いたら、やはり不正アクセスを発表していました。この件に関する明治大学の公式発表を拝見しましたが、直感的には、再発の可能性は高い気がします。

明治大学で1147人の個人情報漏洩、同じ手口で2度目

明治大学は2018年10月24日、メールシステムに不正アクセスを受けていたと発表した。同大学の教職員など3人のメールアカウントが第三者に乗っ取られ、メールの送受信データから3アカウント合計...

NIKKE

ビッグデータの活用

健康医療情報の解析

- 背景：健康データの利活用進む
  - 住友生命「Vitality健康プログラム」
    - » 健康増進（健康診断，運動量）への取組に応じて，保険料の割引
- データの突合
  - 1生活習慣（ライフログ，歩数，血圧など）
  - 2. 疾病（医薬品，診療行為）

第三者提供の方法

方法	備考	法	個人情報	要配慮個人情報
1. 同意	オプトアウト（通知，又は本人が容易に知り得る状態に置くとき）	23条2項	○	×
	例外規定（生命，身体，財産の保護，児童など同意が困難な場合）	23条1項	○	○
2. 委託・共同利用	第三者に該当しない．	23条5項	○	○
3. 匿名加工	公表義務	37条		
	加工基準，安全管理措置	36条，保護委員会規則		

2017年5月30日

- 改正個人情報保護法の全面施行

朝日新聞　２０１７年５月２９日　朝刊　２ページ　東京本社

ビッグデータの活用 後押し

改正個人情報保護法 あす施行

情報匿名化し提供 可能に

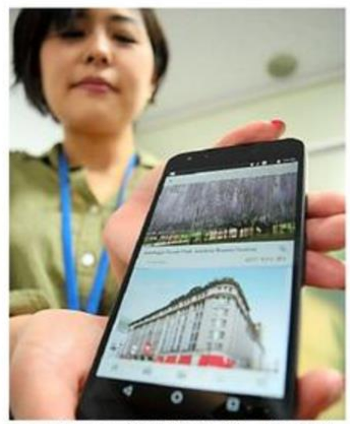
改正個人情報保護法が30日、全面施行される。ビッグデータの利活用を後押しするのが狙いの一つで、個人を特定できないように加工した情報であれば自由に流通させられるようになる。一方、保護規制が強化され、小規模事業者も法の適用対象となることに戸惑いが広がっている。

改正個人情報保護法の主なポイント

特定できないよう個人情報を加工し、復元もできなくする

新設 匿名加工情報 外部提供の本人同意は不要	名前 ○○○○ 生年月日 XXXX年○月○日	識別できる記述 全部または一部を削除
	年齢 (例)116歳 病歴 症例数が極めて少ない	推定できる特異な記述 削除
	マイナンバー XXXX XXXX XXXX 運転免許証の番号 XXXX XXXX XXXX	個人識別符号 すべて削除
新設 要配慮個人情報 原則、本人同意を義務づけ	人種や信条、病歴、犯歴	差別や偏見につながるため 区別
新設	不正な利益を図る目的での個人情報の盗用や流用に罰則 「個人情報保護委員会」の設置。監視・監督などの権限を一元化	
適用対象に	取り扱う個人情報が5千人分以下の中小企業や自治会など	

情報を自由に流通させられるようになる



スマホ用アプリ「TRAVEL JAPAN Wi-Fi」の画面。担当者は「改正法の施行で新たなビジネスが生まれる」と期待する＝東京都中央区

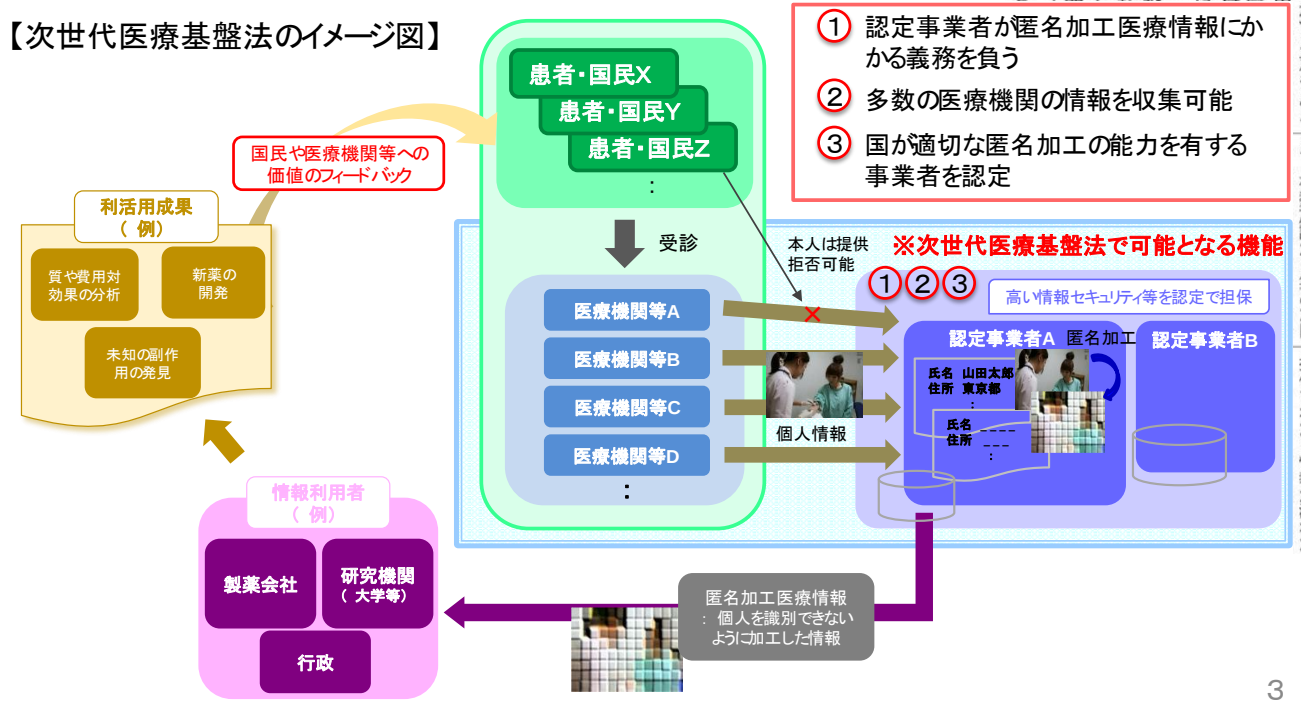
個人特定 懸念なお

スマートフォンのアプリを立ち上げると、近くの百貨店や観光情報が出てくる。KDDIのグループ会社「ワイヤ・アンド・ワイヤレス」が訪日外国人向けに提供する「TRAVEL JAPAN Wi-Fi」だ。全国20万カ所以上

の公衆無線LANが無
使え、観光地や店舗情
どの広告も表示する。
位置情報の取得への
が利用の前提で、累計
ンロード数は約200
個人の行動履歴ではな
端末IDから得られる
や訪問先、滞在時間な
ビッグデータを「統計
で丸めて」自治体など
売している。
責任者の川名義輝マ
ジャーは改正法で新た
場が生まれるのを期待
る。本人を特定できな
う情報を「匿名加工」
ば、より細かいデータ
てコンサルティング会
どからの需要があるか

2018年5月11日

- 次世代医療基盤法施行
  - ❑2019年春認定予定




朝日新聞　２０１８年４月２２日　朝刊　３ページ　東京本社

「医療ビッグデータ」提供へ始動

カルテなど集めて匿名化 企業・研究機関に

国が後押し 漏洩には懸念

診療録（カルテ）や検査データなど個人の医療情報を集めて企業や研究機関に提供する新制度が5月に始まる。国が認定した民間事業者が病院などから実名で集約した情報を匿名化して「医療ビッグデータ」として提供する。情報の漏洩や悪用を懸念する声もあるが、副作用の開発、病気の仕組みは施行する法」で新制度ば今秋には業者が出す報の管理についてたしていく。

る。事業開始時に年間10万人以上の情報を集められる体制を基本とし、すでに大学病院など複数の医療機関の情報を扱っている組織が想定されている。

認定事業者は、カルテや検査報告書、学校や職場の健康診断結果を収集できる。複数施設から集めた同

と期待される。検査画像を人工知能に学習させることで早期診断、早期治療につながる可能性がある。

こうした構想は、事業者が膨大な情報を集めやすい環境が前提になる。そのため、患者本人が拒否しなければ同意したとみなし、病院や診療所は事業者に情報を提供できる。ただし本人が気づかないうちに情報が提供されたり情報が漏れた

https://www.kantei.go.jp/jp/singi/kenkouiryou/jisedai_kiban/houritsu.html

2018年12月21日

- 日本IT団体連盟
  - 情報銀行推進委員会
  - 認定委員会
  - 認定申請受付開始
    - » 認定申請ガイドブック Ver 1.0
    - » モデル約款，モデル契約
  - 2019年春認定予定

https://www.tpdms.jp/

3つの第三者提供 (菊池まとめ)

	匿名加工医療情報	匿名加工情報	情報銀行
法制度	「次世代医療基盤法」 2018年5月11日施行	「個人情報保護法」 2017年5月30日施行	「情報信託機能の認定に係る指針 ver1.0」2018年5月
対象	医療情報	個人情報	個人情報 （共同利用）
病歴等の第三者提供	○ （丁寧な）オプトアウト	× 要配慮個人情報はオプトイン	×/○ その都度本人同意
転々流通	× 契約がGLで推奨	○ 公表義務	
加工事業者	認定	任意	認定
識別のための照合	× 禁止行為	× 法38条禁止行為	
作成基準	施行規則18条 1－5項	施行規則19条 1－5項	認定申請ガイドブック

匿名加工情報取扱事業者例

企業	個人情報	対象	公示日	
日経	性別，年代，業種，役職	日経ID		https:/
イオン銀行	性別，年代，業種，役職年代，借入，返済実績	カードローン顧客	2019/1/18	https:/
DeSCヘルスケア	性別，生年月，レセプト，健診	健康保険組合		https:/
三井住友海上	性別，生年，健診，ストレスチェック	従業員		https:/
ツクイ	性別，年齢，要介護度	介護サービス利用者		https:/
CRD協会（中小企業信用リスク情報DB)	責務者，業種，生年，信用情報	会員（金融機関）		http://w
イズミ	性別，年代，購買実績	会員		https:/
IQVIA	診療科，患者数，レセプト	医師（アンケート）		https:/ policy-
平和堂	年齢，住所，販売データ	会員	2018/8/28	http://w
ブロードマイン	住所，生年，年収，保険商品	顧客		https:/
私立函館病院	病名，処方，診療，DPC,レセプト	患者		https:/

匿名加工情報例

ヘルスケアに特化した合弁会社

DeSC Healthcare

住友商事
デジタルヘルス・
サービス企画・運用経験

DeNA
国内最大級のユーザー規模
（約5,000万会員）を持つ
インターネットサービスの
開発・運用実績

- データ種類
  - 属性情報（性別，生年月）
  - ライフログ（歩数，体重，血圧，血糖）
  - サービス利用状況（参加履歴，記事閲覧数）
  - レセプト（合計点数，診療月，傷病，医薬品，診療行為）
- データ量
  - 数十万人 x 5年分

「公表」の例

DeSC Healthcare

匿名加工情報の作成と提供について

当社は、健康保険組合、共済組合もしくは国民健康保険組合（以下、併せて「組合」といいます。）又は当社のサービスを利用する個人から取得した情報について、特定の個人を識別すること及び個人情報を復元することができないよう適正な措置を講じた上で継続的に匿名加工情報を作成し、当該匿名加工情報を第三者に継続的に提供いたします。

本公表における「個人情報」及び「匿名加工情報」の用語の意義は、個人情報の保護に関する法律に従うものとし、当社は同法に基づき匿名加工情報の作成及び第三者に対する提供を行うものとします。

作成及び第三者に提供する匿名加工情報に含まれる個人に関する情報の項目

- 属性データ（性別、生年月、続柄、資格取得日、資格喪失日）
- レセプトデータ
- 健診・検診データ
- ライフログデータ（歩数／体重／血圧／血糖など）
- その他当社提供サービスの利用履歴データ（イベント参加状況、記事閲覧数など）

匿名加工情報の提供方法

暗号化及びパスワード保護をした電子ファイルをオンラインストレージサービスにて送信又はDVD等の外部記憶媒体にて送付

リスクとセキュリティ技術

総務省指針のセキュリティ基準

2）情報セキュリティ等② 具体的基準

項目	内容
情報セキュリティマネジメントの確立	・経営層（トップマネジメント）は情報セキュリティマネジメントに関してリーダーシップ、コミットメントを発揮すること ・情報セキュリティマネジメントの境界及び適用可能性を明確にし、適用範囲を決定すること ・情報セキュリティリスクアセスメントのプロセスを定め、適用すること、リスク分析、評価、対応を行うこと
情報セキュリティマネジメントの運用・監視・レビュー	・情報セキュリティマネジメントに必要な人・資源・資産・システムなど準備、割り当て、確定すること ・定期的なリスクアセスメントや、内部監査などを実施することで、情報セキュリティマネジメントの適切性、妥当性及び有効性を継続的に改善すること
情報セキュリティマネジメントの維持・改善	・情報セキュリティマネジメントを適切・継続的に維持していくこと ・不適合が発生した場合、不適合の是正のための処置を取ること、マネジメントの改善など行うこと
情報セキュリティ方針策定	・情報セキュリティ方針を策定し、経営層、取り扱う従業員層への周知、必要に応じた方針の見直し、更新
情報セキュリティ組織	・責任者の明確化、組織体制を構築 ・情報セキュリティに関する情報を収集、交換するための制度的枠組みに加盟すること ・経営層は従業員への情報セキュリティ方針及び手順に従った管理の実施を担当者の明確化
ティ	・情報セキュリティの意識向上，教育及び訓練の実施
資産管理	・重要な情報処理施設に関連する資産の洗い出し、特定し、適切な保護の責任を定めること ・国内にデータセンターを保有していること、又はそれと同等の管理が可能な委託先データセンターを確保していること 外部クラウドを活用する場合には当該クラウド利用契約上の情報セキュリティ要件などで担保されていることを示すこと（例：JIS Q 27017「JIS Q27002 に基づくクラウドサービスのための情報セキュリティ管理策の実践の規範」） ・情報を取り扱う媒体等から情報を削除・廃棄が必要となった場合にそれが可能な体制もしくは仕組みを有すること ・対象となる事業で扱う情報が他事業と明確に区分され管理されていること ※なお、外部クラウドなど活用する場合や、委託を行う場合に相手方事業者との間で、裁判管轄を日本の裁判所とすること、準拠法を日本法とすることを合意しておくこと
技術的セキュリティ	（アクセス制御） ・アクセス制御に関する規定を策定し、対応すること（例：アイデンティティ管理システムの構築、アクセス制御方針の実装） ・情報にアクセス権を持つ者を確定し、それ以外のアクセスの制限を適切に行うこと （暗号） ・情報の機密性、真正性、完全性を保護するため暗号の適切で有効な利用をすること ・電子政府推奨基準で定められている暗号の採用や、システム設計の確認など対応すること

（一般的な）情報漏洩や不正アクセスに対するセキュリティ基準と相違ない

匿名加工とは?

個人情報

菊池浩明	48歳	4月20日	メロンパン	125円
菊池浩明	48歳	4月23日	ウーロン茶	150円
中川裕志	51歳	4月23日	ウーロン茶	150円
村上隆夫	23歳	4月23日	メロンパン	125円

匿名化

匿名加工情報
（非個人情報）

300		4月20日	パン	120円
300		4月23日	飲料	150円
100		4月23日	飲料	150円
500		4月23日	パン	120円

課題1: リスクの多様性

菊池浩明	48歳	4月20日	メロンパン	125円
菊池浩明	48歳	4月23日	ウーロン茶	150円
中川裕志	51歳	4月23日	ウーロン茶	150円
村上隆夫	23歳	4月23日	メロンパン	125円

匿名化

再識別

2. **識別**
(linking)

300		4月20日	パン	120円
300		4月23日	飲料	150円
100		4月23日	飲料	150円
500		4月23日	パン	120円

菊池浩明?

中川裕志?

同一者?

1. **特定**
(singling out)

違法
38条識別行為の禁止

3. **属性推定**
(inference)

メロンパン?

適法
（プロファイリング権利で規制？）

課題2. 万能な匿名加工はない

- 多用なビッグデータに対して，共通して適用可能な加工技術が未確立

加工 ＼ リスク	特定 Single-out	識別	属性推定
削除	部分的	部分的	部分的
仮名化	No	部分的	No
一般化	No	部分的	部分的
k-匿名性	Yes	部分的	No
ℓ-多様性	No	No	Yes
差分プライバシー	Yes	Yes	部分的

ISO/IEC DIS 20889:2018, Table A.1から

課題3. 安全性の曖昧な基準

規則	対象	例
(1)	個人情報の全部又は一部を削除すること	氏名を削除
(2)	個人識別符号の全部を削除すること	マイナンバーを削除
(3)	個人情報と当該個人情報に措置を講じて得られる情報とを連結する符号を削除する	管理用 ID を 削除
(4)	特異な記述等を削除	116歳を90歳以上に置換
(5)	個人情報データベース等を構成する他の個人情報に含まれる記述等との差異その他の当該個人情報データベース等の性質を勘案し、その結果を踏まえて適切な措置を講ずること	・自宅住所が推定できる位置情報削除 ・購入者が限定される特殊商品を一般カテゴリーに置換

「個人情報保護ガイドライン（匿名加工情報編）」より抜粋

k-匿名化

1. オリジナル

M		T
A大学	24歳	コーヒー
A大学	24歳	お茶
B大学	28歳	お茶
C総研	35歳	コーヒー
C総研	35歳	コーヒー

2. 一般化（所属）

M		k	T
大学	24歳	2	コーヒー
大学	24歳		お茶
大学	28歳	1	お茶
企業	35歳	2	コーヒー
企業	35歳		コーヒー

k=1

(大学，28歳）で
一意に識別できる
(single-out)

3. 一般化（年齢）

M		k	T
大学	20代	3	コーヒー
大学	20代		お茶
大学	20代		お茶
企業	30代	2	コーヒー
企業	30代		コーヒー

k=2

（所属，年齢）でどの
グループも少なくとも
2人以上属する

PWSCUP匿名加工コンテスト

	2015	2016	2017	2018
開催	10/21-22 長崎	10/11-12 秋田	10/23-24 山形	10/23-24 長野
参加者数	13チーム (20名)	15チーム (42名)	14チーム (43名)	14チーム
データセット	疑似ミクロデータ （世帯消費額）	UCI Dataset “Online Retail” (購買履歴)		
属性数	25	11 （顧客4属性＋履歴7属性）		履歴(5属性)
顧客数	8,333	400	500	1000
履歴数	なし	18,524	44,917	73,021
テーマ	摂動化	履歴	仮名長	一般化

本戦 (2018年10月24日Hメトロポリタン長野)

研究目的

- 有用性高く安全な匿名加工方式の追求

(1) 具体的な加工方法標準

- 加工方法はユースケース依存
- プライバシーの度合いはデータ依存

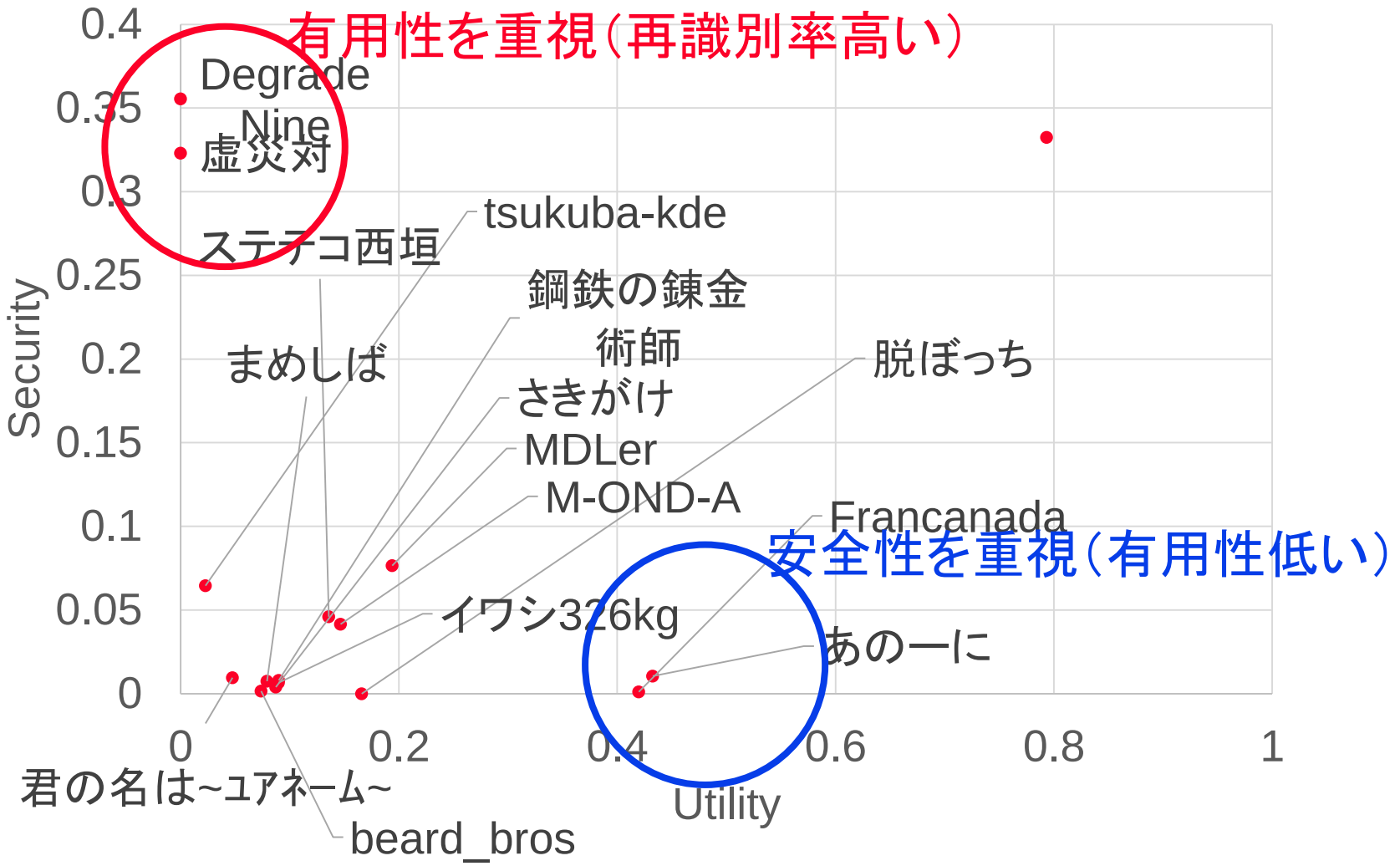
(2) 攻撃者の仮定

- 攻撃者が持つ知識をどう仮定するか?
- 自動的に評価できないか?

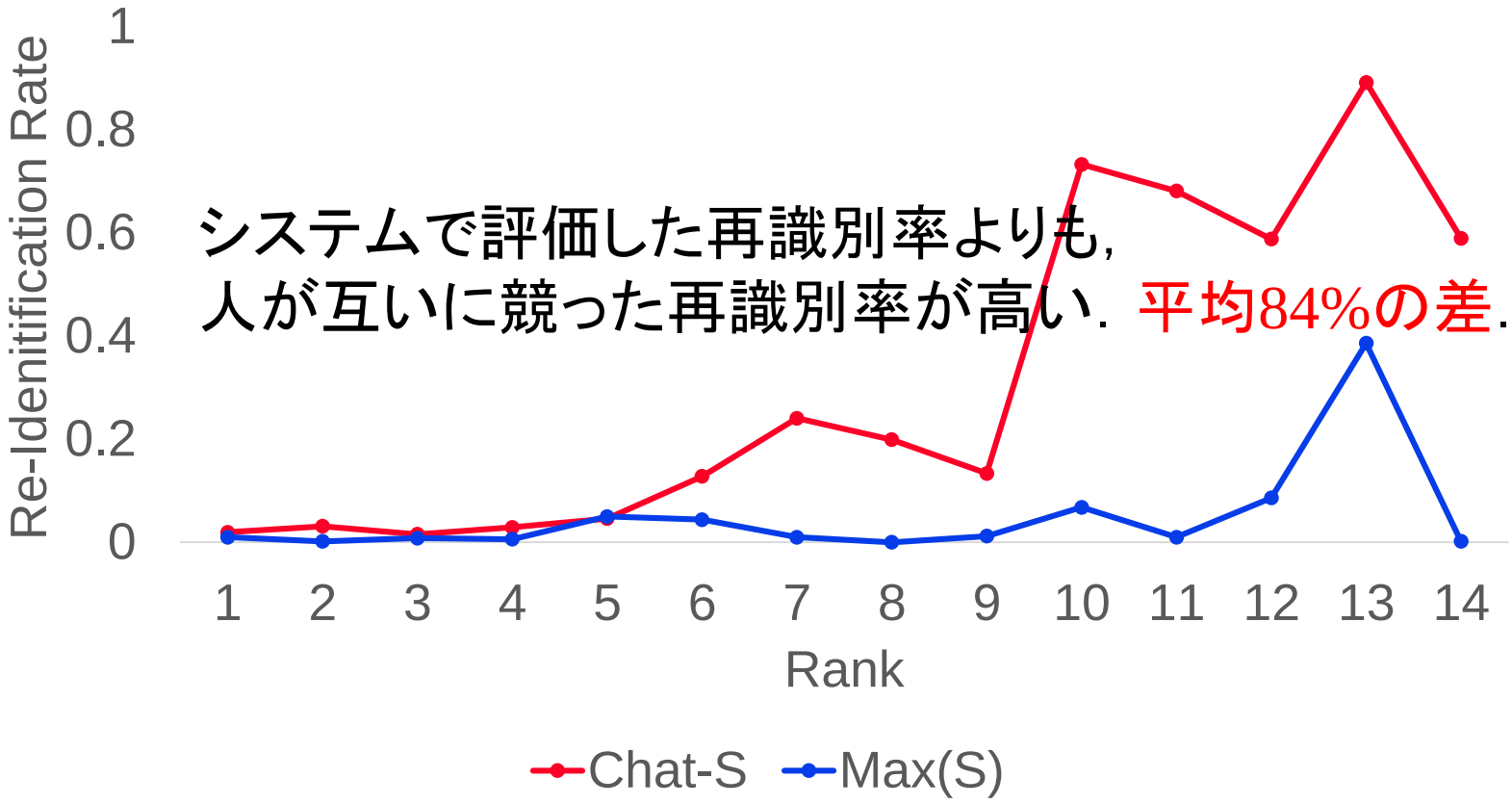
(3) 長期間の履歴データ

- データ分割と仮名の制御

有用性と安全性の関係


有用性を重視（再識別率高い）
安全性を重視（有用性低い）
Degrade
Nine
虚災対
ステテコ西垣
tsukuba-kde
鋼鉄の錬金
術師
まめしば
さきがけ
MDLer
M-OND-A
脱ぼっち
Francanada
イワシ326kg
あのーに
君の名は~ユアネーム~
beard_bros
Security
Utility
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.2
0.4
0.6
0.8
1

再識別率



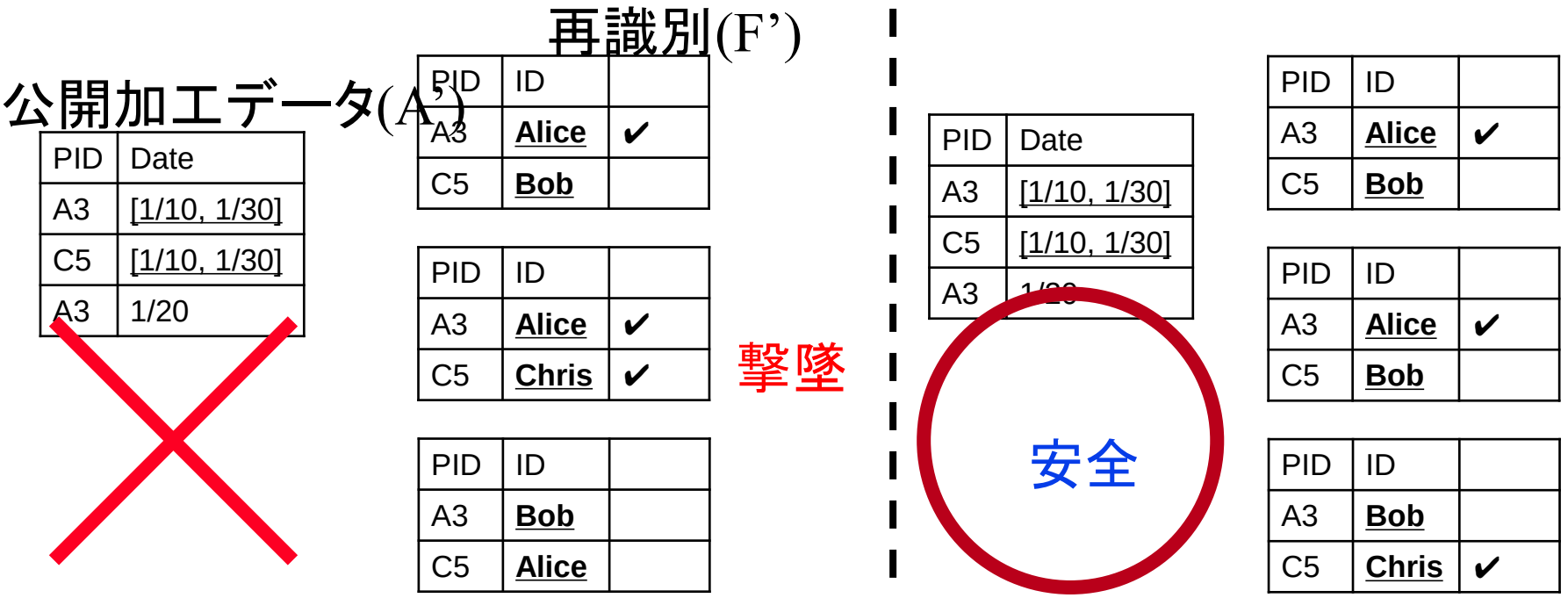
システムで評価した再識別率よりも，
人が互いに競った再識別率が高い．平均84%の差．

特徴: 安全性評価 「撃墜」

- 安全性: 仮名とIDの対応の**当てにくさ**で評価

  当てやすい証拠となる再識別が1つでも存在

  ➔「撃墜」 **安全でない**と判断



本戦の結果 (再識別)

*: 再提出ペナルテ
**: 書式エラー

ID	チーム名	01	02	03	05	06	07	08	09	10	11	12	13	14	撃墜
01	鋼鉄の錬金術師	-	0(2/7)	0(4/7)	0(5/7)	1(7/7)	0(2/7)	0(4/7)	1(7/7)	0(6/7)	0(5/7)	1(7/7)	0(6/7)	0(2/7)	3
02	**paddy**	1(8/8)	-	0(0/8)	0(4/8)	1(8/8)	0(0/8)	0(4/8)	1(8/8)	0(4/8)	0(6/8)	1(8/8)	0(4/8)	0(6/8)	**4**
03	ほぼぼっち	0(4/8)	0(1/7)	-	0(4/8)	1(7/7)	0(5/7)	0(4/8)	1(8/8)	0(4/8)	0(2/8)	1(7/7)	0(4/8)	0(2/8)	3
05	Unhackable Anonymity	0(3/7)	0(1/7)	0(3/7)	-	1(7/7)	0(3/7)	0(5/7)	1(7/7)	0(3/7)	0(4/7)	1(7/7)	0(3/7)	0(3/7)	3
06	westlab	0(2/7)	0(3/7)	0(5/7)	0(5/7)	-	0(1/7)	0(4/7)	1(7/7)	0(5/7)	0(4/7)	1(7/7)	1(7/7)	0(5/7)	3
07	unicorn	**	**	0(2/8)	0(4/8)	0(0/7)	-	0(2/8)	1(7/7)	1(8/8)	0(2/8)	**	0(6/8)	**	2
08	ステテコ西垣2	0(9/11)	0(2/7)	0(2/11)	0(4/11)	1(184/184)	0(4/7)	-	1(338/338)	0(2/11)	0(3/11)	1(480/480)	0(7/11)	0(7/11)	2.9*
09	筑波大学	0(3/7)	0(1/7)	0(4/7)	0(2/7)	0(1/7)	0(2/7)	0(3/7)	-	0(5/7)	0(3/7)	1(7/7)	0(3/7)	0(1/7)	1
10	**MaeData**	0(2/7)	0(3/7)	0(3/7)	0(4/7)	1(7/7)	0(3/7)	0(4/7)	1(7/7)	-	1(7/7)	1(7/7)	0(5/7)	0(3/7)	**4**
11	匿名係長 NINJA HATTORI	0(0/8)	0(1/7)	0(4/8)	0(4/8)	1(184/184)	0(5/7)	0(4/8)	1(338/338)	0(4/8)	-	1(480/480)	0(4/8)	0(4/8)	3
12	X-面（仮）	0(4/7)	0(2/7)	1(7/7)	0(0/7)	0(2/7)	0(6/7)	0(0/7)	0(3/7)	0(3/7)	0(2/7)	-	0(3/7)	0(1/7)	1
13	あのーに	-	-	-	0(16/22)	-	-	-	**	**	-	-	-	**	0
14	先駆け	0(2/7)	0(1/7)	0(5/7)	0(0/8)	1(21/21)	0(4/7)	0(2/7)	1(21/21)	0(5/7)	0(2/7)	-	0(5/7)	-	2

結論

- 法整備の下，健康情報を含む個人情報の利活用が始まろうとしている
- 第三者提供には，本人同意，委託・共同利用，匿名加工の3種類（もっと？）がある
- 事業者の「認定」に係る課題.
  - 一般的な情報漏洩と不正アクセス対策以上の特有の評価は必要か.
- 匿名加工に係る課題.
  - 属性推定（プロファイリング）は法規制がない.
  - 有用性と安全性の間にトレードオフがあり，正確に再識別リスクを評価するのは困難である. コンテストでは再識別率に83%の差がある.

個人情報情報とプライバシー情報

	リピート分析用情報	匿名加工情報	属性・人流情報
個人情報	個人情報（識別符号）	匿名加工情報	非個人情報（プライバシー情報）
法的根拠	政令（施行令）	保護法	Q1-13
加工方法	カメラ画像から特徴量抽出	識別符号の削除，仮名化など	属性推定，履歴マッチング，カメラ画像は「即座に削除」
本人同意	要	不要	不要
（同意なき）第三者提供	否	可	可
作成記録と項目などの公表義務	NA	要（法36条）	なし？
（識別のための）照合	NA	禁止（法38条）	なし？

個人情報保護委員会Q&A

Q1－12 店舗にカメラを設置し、撮影した顔画像やそこから得られた顔認証データを マーケティング等の商業目的に利用することを考えています。個人情報保護法との関 係で、どのような措置を講ずる必要がありますか。

- A.本人を判別可能なカメラ画像やそこから得られた顔認証データを取り扱う場合、個人情報の利用目的をできる限り特定し、あらかじめ公表するか、又は個人情報の取得後速やかに本人に通知若しくは公表するとともに、当該利用目的の範囲内でカメラ 画像や顔認証データを利用しなければなりません。

Q1－13カメラ画像から抽出した性別や年齢といった属性情報や、人物を全身のシル エット画像に置き換えて作成した移動軌跡データ（人流データ）は、個人情報に該当 しますか。

- 個人情報とは、特定の個人を識別することができる情報をいいます。性別、年齢、又は全身のシルエット画像等による移動軌跡データのみであれば、抽出元の本人を判別可能なカメラ画像や個人識別符号等本人を識別することができる情報と容易に照合することができる場合を除き、個人情報には該当しません。

Privacy Risks with Facebook's PII-based Targeting: Auditing a *Data Broker*'s Advertising Interface

Giridhari Venkatadriy 1, Athanasios Andreoux 2, Yabing Liuy1, Alan Mislovey1, Krishna P. Gummadiz3, Patrick Loiseauz 4, Oana Goga 4
1 Northeastern University 2 EURECOM 3 MPI-SWS 4 Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG

Presented at IEEE S+P 2018

Data Broker

- Definition
  - ❑ businesses whose revenue model revolves around aggregating information about individuals from a variety of public and private sources.
- Example

IPGに売却
(2018)

Oracleに売却
(2014)

de facto data brokers

- Online services
  - Google
  - Facebook

  - they use the collected data to build powerful advertising services that have data on billions of users worldwide

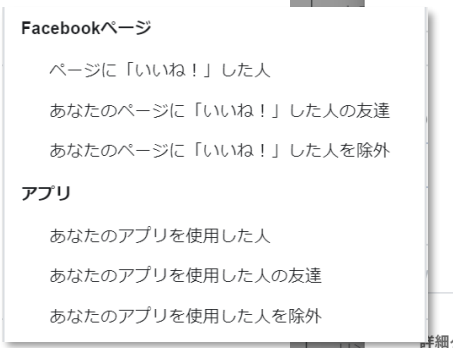
Custom Audiences

- Online Advertising platforms
  - ❑ advertisers to create custom audiences

Facebook	Custom Audiences
Twitter	Tailored Audiences
Google	Customer Match
Pinterest	Audiences
LinkedIn	Audience Match

Site	Name	Email	Phone number	City or ZIP	State or Province	Birthday, Gender	Employer	Site user ID	Mobile advertiser ID	Min. Size
Facebook	✓	✓	✓	✓	✓	✓	✗	✓	✓	20
Instagram	✓	✓	✓	✓	✓	✓	✗	✓	✓	20
Twitter	✗	✓	✓	✗	✗	✗	✗	✓	✓	500
Google	✓	✓	✓	✓	✗	✗	✗	✓	✓	1,000
Pinterest	✗	✓	✗	✗	✗	✗	✗	✗	✓	100
LinkedIn	✗	✓	✗	✗	✗	✗	✓	✗	✓	100

Facebook「オーディエンス編集」



オーディエンス詳細:

- 地域:
  - 日本
- 年齢:
  - 18歳～65+歳
- 性別:
  - 男性
- 次の条件に一致する人:
  - 趣味・関心: ネットワークストレージまたはIT
- 趣味・関心の拡大:
  - オフ

V. Attacks

- A. De-anonymizing web visitors
  - determine if a particular user has visted website.
- B. Inferring a victim's PII
  - with victim's email address, infer a victim's other PII (phone number)
- C. en-masse De-anonymization
  - infer the phone numbers of all members of all visitors to a webpage.

two characteristics (how)

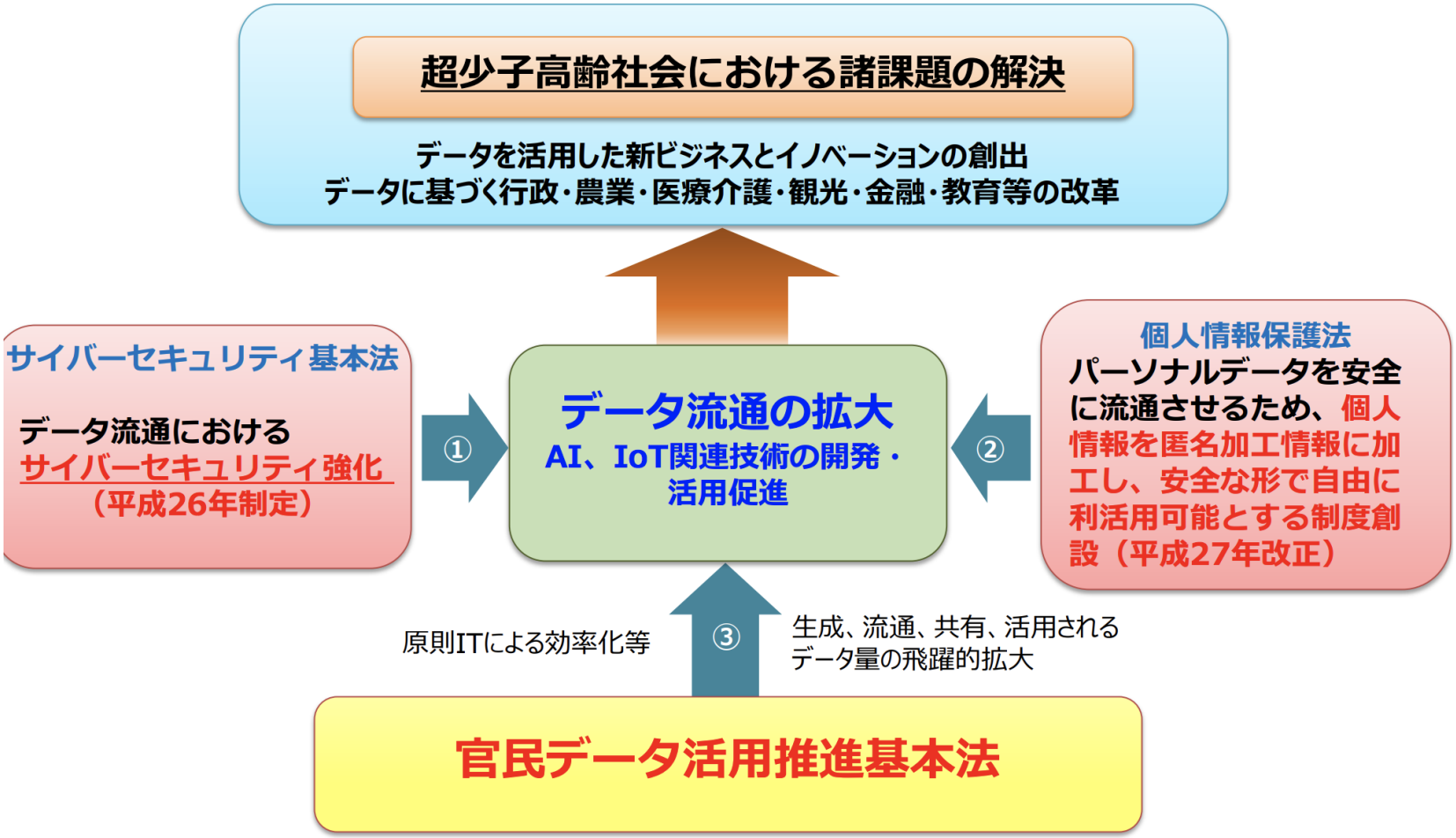
- 1. size statistics
  - reported and obfuscated using rounding (we negate the effect of rounding)
- 2. de-duplicates
  - multiple PII records refer to the same user when reporting the size (we determine whether two of PII refer to the same)

Evaluation

- Boston
  - (617 and 895) plus 7 digits
  - 140 lists of 1M numbers (30 minutes to upload)
- France
  - 6 or 7 digits
  - 82 lists of 10M numbers (over a period of a week)

官民データ活用基本法

官民データ活用推進基本法制定の背景



H29.3. 内閣官房IT総合戦略