パーソナルデータのプライバシー保護を実現する匿名化・暗号化技術

Anonymization and Encryption Technologies to Protect Privacy of Personal Data

● 伊藤孝一　● 小暮　淳　● 下山武司　● 津田　宏

あらまし

ビッグデータ分析や，今後のIoTの発展によりデータは更に増加することが見込まれる。そして，そこから集めたデータを個人にひも付いた情報（パーソナルデータ）として活用し，新たなビジネス創出につなげることが期待されている。一方，サイバー攻撃や内部不正による大規模な個人情報の流出，プライバシーへの配慮がないサービスの中止などの事例も続いている。また，マイナンバーや個人情報保護法の改正，EUデータ保護規制など，法的規制も強化されつつある。富士通研究所では，これらの法案や規制に準拠したプライバシー情報を匿名化・暗号化する技術を開発している。

本稿では，パーソナルデータを安全に扱うために，富士通研究所で開発したk-匿名化などの匿名化技術や，準同型暗号などの暗号技術について紹介する。

Abstract

The volumes of data in society are anticipated to further increase for reasons such as expansion of big data analysis and future development of the Internet of Things (IoT). There are high hopes that it will be possible to utilize the data gathered in these processes as information linked with individuals (personal data) and create new businesses. Meanwhile, there have been a series of cases of large-scale leakage of personal information due to cyber attacks and internal fraud, along with discontinuance of services that did not pay sufficient attention to privacy. In addition, legal regulations are being tightened as seen in the "My Number (Individual Number)" system, amendment of the Personal Information Protection Act and EU Data Protection Regulation. Fujitsu Laboratories has been developing technologies for anonymization and encryption of privacy information in conformity with these measures and regulations. This paper describes the anonymization technologies such as k-anonymization and encryption technologies such as homomorphic encryption developed by Fujitsu Laboratories for safe handling of personal data.

ま え が き

ビッグデータ分析による新たなビジネス創出に続き，今後IoTの発展により更にデータの増加が見込まれ，ビジネスにおいてデータ利活用の重要度がますます高まると考えられる。中でも，個人情報やプライバシー情報（位置情報・行動履歴・購買履歴などパーソナルデータ）の活用は，医療，金融，小売りなどの分野で大きく期待されている。一方，パーソナルデータはセンシティブな情報を含む場合もあり，取扱いを誤るとビジネスに大きな打撃を与えることがある。

例えば，2013年7月にJR東日本がSuicaの乗車履歴データを匿名化して販売しようとした事例[(1)]が知られている。本事例においては，事業者に対して多くの利用者から「勝手に売るな」「行動を監視されているようで気持ち悪い」「乗車履歴は個人情報ではないのか」などの声が上がり，事実上のサービス停止に追い込まれた。

このような事例の影響により，パーソナルデータの取扱いに関するルールを明確化するために，内閣IT総合戦略本部は「パーソナルデータに関する検討会」を設置し，2014年6月に「パーソナルデータの利活用に関する制度改正大綱」案を策定した。この大綱に基づき，2015年に個人情報保護法が改正され，2017年には施行される見込みである。

改正版の個人情報保護法には多くの特徴がある。中でもパーソナルデータを「匿名加工情報」に加工し，一定の制約（データ提供先におけるデータからの個人再特定禁止）のもと，本人の同意なしでデータを第三者に提供することが認められている点が大きな特徴である。匿名加工情報は，特定の個人の識別はもちろん，復元ができないように加工された情報である。このような加工を実現するための技術として「匿名化技術」が注目され，その重要度は今後更に高まると予想される。

本稿では，EU，米国など世界の代表的な地域のパーソナルデータに関する法規制動向，および日本の個人情報保護法の改正とその要点について述べる。また，プライバシー保護と利活用を両立するための匿名化技術に関する解説を行い，富士通研究所の匿名化技術を紹介する。更に，匿名化技術だけでは対応できない分野への取組みとして，富士通研究所の準同型暗号技術を紹介する。

各国の動向

EU，米国，および日本のパーソナルデータに関する法規制動向の特徴を述べる。

（1）EU

パーソナルデータ保護に関する規制として，「EUデータ保護指令」を1995年に採択し，この指令に基づきEU各国での国内法化が進んだ。しかし，国ごとの互換性がないため，多国籍企業のビジネス上で負担が大きいことが問題となっていた。この問題を解消するため，EU全体の包括規制として2012年1月に「EUデータ保護規則案」が提案され，2015年内の欧州議会可決，2017年の施行を目指している。パーソナルデータの定義は広く（生体情報，位置情報，IPアドレスを含む），違反があった場合，監督機関である欧州データ保護機関によって最大で売上げの5％の罰金が科せられる。また，データの海外移転に関しても制約が設けられるなど，非常に厳格な保護を行っているのが特徴である。

（2）米国

パーソナルデータ保護に関する統一的な法律はなく，プライバシー上の問題が懸念されるデータ活用を行う個別事例に対して，連邦取引委員会（FTC）による調停が行われている。例えば，DVDのレンタル履歴から高い精度のレコメンデーションを行っていたNetflix Prizeが，視聴履歴に関するプライバシー上の懸念により2010年に中止された[(2)]が，その調停はFTCが行っている。米国は，一般的にはプライバシー保護よりデータ活用を重視していたが，2015年3月にホワイトハウスから消費者プライバシー権法案を公開するなど，プライバシー保護を重視する方向にシフトしてきていると考えられる。

（3）日本

2005年4月に個人情報保護法が施行されたが，個人情報の定義やその取扱いに関する規則が曖昧な上，個人情報を全体として監督する組織もなかった。

上記したEU，米国，日本の法規制動向に関する比較を**表-1**に示す。それぞれの地域に共通する傾向は，プライバシーに関する法規制強化であり，利用者が安心してデータを預けられる制度を実現しようとしていることである。

表-1　EU，米国，日本のパーソナルデータに関する法規制内容の比較

	EUデータ保護規則（欧州）	消費者プライバシー権法（米国）	改正個人情報保護法（日本）
パーソナルデータの定義	生体情報や位置情報，IPアドレスを含む	生体情報やIPアドレス，保険番号，車体番号を含む	生体情報を含む
利用目的の変更	後から変更不可	後から変更する場合，明示同意によるオプトインが必要	「変更前の利用目的と関連性を有すると合理的に認められる範囲」で変更可
機微情報	機微情報（sensitive personal data）：宗教，政治的信条，人種など	差別的効果（disparate impact）：人種，宗教，同性愛者など	要配慮個人情報：人種，信条，社会的身分，病歴，犯罪被害，前科・前歴など
オプトアウト	原則不可	原則不可 （FTC承認があった場合例外）	要配慮個人情報と海外データ移転は不可（オプトインが原則）
本人同意なしの第三者提供	原則不可（本人同意が原則）	非識別化データ（De-identified data）は本法の規制対象外	・匿名加工情報（個人識別できる記述を削除）は提供可能 ・第三者提供の公表と提供先での個人再識別禁止の二種類の義務
利用範囲	公的利益，もしくは当事者の正当な利益がある場合以外は，同意取得時の利用範囲を越えられない	コンテクストに沿わないパーソナルデータの利用分析は，FTC監査・承認が得られた場合のみ可能	オプトイン，オプトアウト（要配慮個人情報，海外データ移転などでは不可），匿名加工情報の3種
プライバシー評価	プライバシー影響評価（PIA）など	プライバシーアセスメントの定期実施	特になし
自己情報コントロール権	あり（訂正権，削除権，拒否権）	あり（同意撤回権，削除権）	あり（消去権，開示請求権，訂正権，提供停止権）
データの海外移転	制限あり	制限あり	制限あり
監督機関	欧州データ保護委員会	州検事総長，FTC	・個人情報保護委員会（特定個人情報保護委員会を改組） ・立ち入り検査

表-2　日本の個人情報保護法の改正前後の比較

	現行個人情報保護法	改正個人情報保護法
個人情報	・氏名，生年月日そのほかの記述により特定個人を識別可 ・ほかの情報と照合して特定個人を識別可	同じ
個人識別符号	なし	・身体の一部特徴を符号化したもの（指紋，顔認証データ） ・役務，購入，書類に付される符号（旅券番号，免許証番号，携帯電話番号など）
機微情報	規定なし	・要配慮個人情報として定義：人種，信条，社会的身分，病歴，犯罪被害，前科・前歴など ・オプトインが原則
利用目的	「変更前の利用目的と相当の関連性を有すると合理的に認められる範囲」で変更可	「変更前の利用目的と関連性を有すると合理的に認められる範囲」で変更可
個人情報取扱い事業者	5000人以上を扱う事業者	制限なし
第三者提供	事前の本人同意が原則 （例外：オプトアウト，共同利用，委託）	・匿名加工情報（個人識別できる記述を削除）に関して，第三者提供の公表義務と提供先での再識別禁止義務のもとで可能
海外提供	国内と同様	海外提供時の条件あり
監督機関	主務大臣 報告，命令，団体の認定	・個人情報保護委員会（特定個人情報保護委員会を改組） ・立ち入り検査
罰則	事業者に対して行政処分	・データベース提供罪（従事者）

日本の個人情報保護法改正とその要点

本章では，日本の改正版個人情報保護法[3]における重要なポイントについて解説する。改正前と改正後の比較を**表-2**に示す。主な差異は以下のとおりである。

(1) 個人情報の定義

従来の個人情報に加えて，身体に関する情報などを含む「個人識別符号」と，人種・信条などを含む「要配慮個人情報」の二つを定義している。通常の個人情報と異なり，個人識別符号は匿名加工情報による第三者提供は不可であり，要配慮個

人情報は取得や提供時のオプトイン（事前同意）が必須という違いがある。

（2）利用目的

利用目的の変更に際し，従来の「利用目的と相当の関連を有する」の「相当」が削除された。ただし，どこまで緩和されるかは明確ではない。

（3）個人情報取扱い事業者

従来の「5000人以上を扱う事業者」の制限が廃止され，従来対象とされていなかった小規模事業者も対象となる。

（4）第三者提供

特定の個人を識別できる情報を削除する加工を行った匿名加工情報に関して，第三者提供することを公表することで，オプトアウト（本人同意なしの第三者提供）が可能となった。ただし，提供先はデータからの個人再識別禁止義務を負う。

（5）海外提供

改正前は，国境を越えた運用に関する制約が特に規定されていなかったが，改正後は一定以上の保護水準にあると認められる国，もしくは組織以外へ提供する場合，オプトインが必須となる。

（6）監督機関

今回の改正により，個人情報保護委員会が監督機関として新たに立ち上げられた。匿名加工情報の分野別の匿名加工基準の作成や，個人情報取扱い事業者に対する立ち入り検査，助言・指導を行う。

（7）罰則

改正により，データベース提供罪などの新たな罰則が追加された。

全般的に規制強化となっている中，匿名加工情報の第三者提供が特に注目される。これは，第三者提供事実の公表義務や提供先での個人再識別禁止義務のもと，本人同意なしの第三者提供を認めている点であり，これによりパーソナルデータの円滑な利用の促進が期待される。

匿名化技術

匿名加工情報に加工するには，複数の手法がある。ただし，匿名加工情報を生成する手法に求められる匿名加工基準については，分野ごとに個人情報保護委員会が今後策定する予定である。

したがって，匿名加工情報に加工するための手法は今後の決定を待つ必要があるが，匿名化技術と呼ばれる個人を特定できないようにデータを加工する技術が有力な候補と考えられている。匿名化技術に関しては，過去に多くの研究がなされ様々な手法が知られており，[4] 手法ごとにメリットとデメリットがある。本章では，代表的な匿名化技術である「仮名化（かめいか）」と「k-匿名化」の2種類の技術を紹介する。また，これらの技術に対する富士通研究所の取組みと，富士通研究所の匿名化技術を搭載した富士通のソリューションである「NESTGate」について紹介する。

（1）仮名化

図-1に示すように，実名を仮名に置き換える（例えば，実名と無関係な仮IDなど）ことで，個人の特定を防ぐ。そのメリットは，同一人物の追跡が可能な点であり，例えば，仮名化した後も同一人物の健康状態の経過を追跡できる。一方，デメリットは，性別，年齢など，間接的に個人を特定できる属性の組合せから，レコードに対応する個人を特定できる可能性があることである。例えば，図-1の場合「男性, 103歳」に該当する個人が「田中 平助」さんであることを知っている人は，どのレコードが田中さんに対応しているかを特定できてしまう。

（2）k-匿名化

仮名化は，前述のようにレコードに対応する個人を特定できる可能性があるが，この問題を回避

氏名	性別	年齢	疾患
田中 平助	男	103	糖尿病
吉田 貴子	女	83	がん
山本 洋介	男	23	肺炎
鈴木 三郎	男	26	肺炎

仮名化

	仮ID	性別	年齢	疾患
田中	ID23052	男	103	糖尿病
吉田	ID47353	女	83	がん
山本	ID98124	男	23	肺炎
鈴木	ID83041	男	26	肺炎

図-1　仮名化の例

するk-匿名化という技術が知られている。k-匿名化技術は，間接的に個人特定可能な属性を準識別子（QI：Quasi-identifier）として定義することで，同一のQIの組合せが必ずk人以上存在するようデータを加工し，レコードに対応する個人特定を防ぐ。k＝2におけるk-匿名化の例を**図-2**に示す。年齢と性別をQIとして定義することで，同一（年齢，性別）の組合せが二人以上となり，どのレコードが田中さんであるか特定できない。

（3）富士通研究所の匿名化技術

富士通研究所では，匿名化を実現するための技術として，情報ゲートウェイ（以下，情報GW）技術[5]とk-匿名化ライブラリを開発した。

・情報GW

クラウドなど組織外部のサービス利用時に，情報GW（**図-3**）を介して外部とデータをやり取りすることで，外部サービスでは元データが分からないよう秘匿化し，利用時には秘匿化されたデータを元データに復元できる。秘匿化として仮名化を用いることで，仮名化を情報GWで一括処理し，既存システムに対する修正を最小限に抑えられる。

・k-匿名化ライブラリ

省メモリ，高速，安全性を備えた独自k-匿名化ライブラリの開発に成功した。図-2に示すとおり，k-匿名化は近い関係にあるレコード同士をグループ化するため，レコードごとに独立した処理を行うことができない。つまり，k-匿名化処理は入力データを主記憶装置上に読み込む必要があるが，入力データサイズが大き過ぎると主記憶装置上に格納しきれないため，処理速度が遅くなる。この

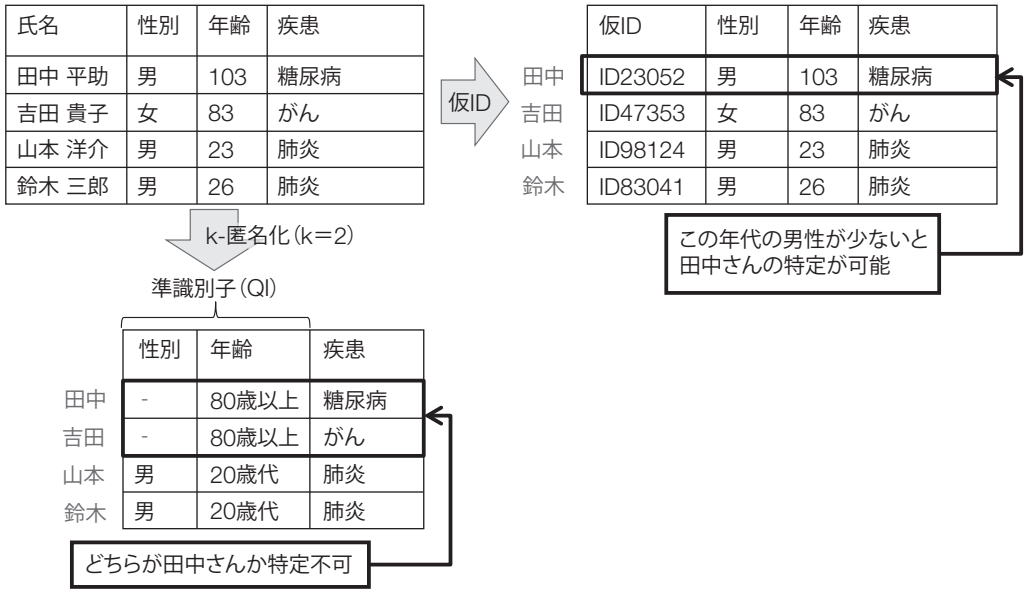

図-2　k-匿名化の例（k＝2）

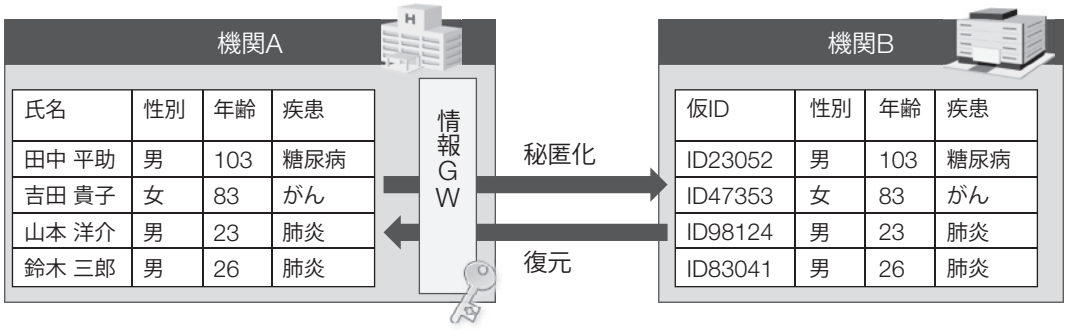

図-3　情報GWの概略

問題を解決するため，富士通研究所は独自のk-匿名化アルゴリズムを開発した。それによって，主記憶装置に読み込むデータのサイズを入力データより小さく（データ次第だが，およそ1/10 ～ 1/100）抑えることができ，搭載メモリ量が多くない安価なサーバやパソコンでも高速なk-匿名化処理を実現できる。更に，k-匿名化より高度な匿名化機能にも対応している。例えば，図-2のk-匿名化の結果では，山本さんが3，4行目のどのレコードであるか特定できないが，肺炎であることが分かる。本ライブラリには，「k-存在秘匿性」[6]と呼ばれるk-匿名化より高度なプライバシー保護を実現する機能が搭載されており，上記のようなk-匿名化では防げないプライバシー情報の漏えいを防ぐことができる。

・匿名化技術ソリューションNESTGate

前述の，情報GWとk-匿名化ライブラリを搭載した匿名化ソリューションとして，富士通はNESTGateを製品化している。NESTGateを用いることで，仮名化，k-匿名化処理，統計法ガイドライン[7]などの各種匿名化処理できる。**図-4**にNESTGateの適用イメージを示す。「秘匿化ソリューション（情報GW）」に対してデータを通すことで，規定ポリシーに従った自動的な仮名化処理を行うことができる。また，「k-匿名化ソリューション（バッチ処理）」を用いることで，富士通研究所で開発した高速なk-匿名化ライブラリや，統計法ガイドラインに従った匿名化処理を行うことができる。詳細は，参考文献（8）を参照されたい。

準同型暗号

匿名化技術は，個人を特定できないようにデータ自体を加工し情報量を低下させるため，データ分析活用の精度に悪影響を及ぼす場合がある。例えば，ゲノム分析などではゲノムデータ自体を加工しては分析ができない。富士通研究所では，そのような分野に適用するプライバシー保護技術である「準同型暗号」の開発を進めている。

準同型暗号は，入力データを秘匿したまま加算・乗算などの算術処理ができ，それらを組み合わせることで理論的には任意の処理が可能となる。

xを暗号化したデータを$\mathrm{Enc}(x)$と表した場合，この準同型暗号が持つ基本的な性質は以下の数式で表すことができる。

$$\begin{cases}\mathrm{Enc}(x)+\mathrm{Enc}(y)=\mathrm{Enc}(x+y)\\ \mathrm{Enc}(x)\times\mathrm{Enc}(y)=\mathrm{Enc}(x\times y)\end{cases}$$

この準同型暗号を用いることで，二つのベクトルデータの類似度（ハミング距離）を暗号化したまま計算できる。また，秘匿したままでの文字列検索，すなわちデータやクエリキーワードを暗号化したまま検索処理できる。処理結果も秘匿される準同型暗号の特徴により，どんなキーワードを検索，あるいはヒットしたかも知られたくない，例えば製薬研究開発の遺伝子情報検索などへの応

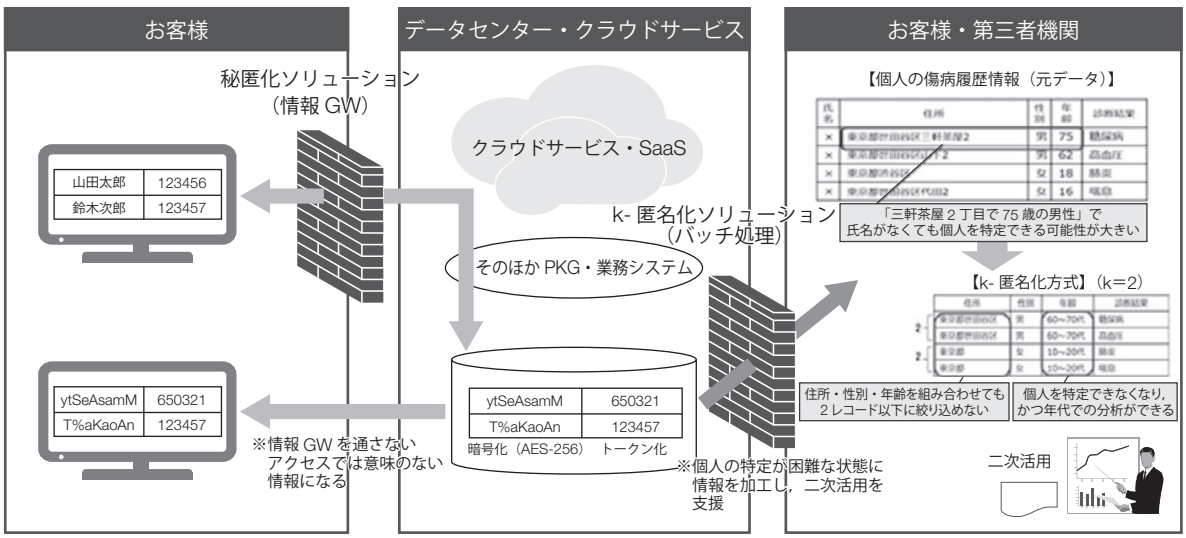

図-4　匿名化技術ソリューション「NESTGate」

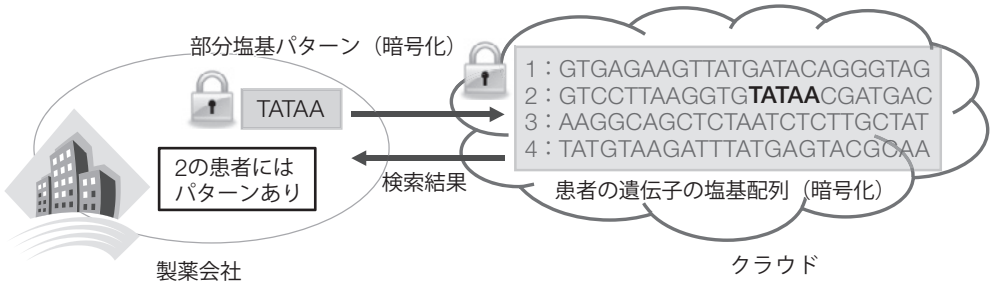

図-5　準同型暗号による秘匿検索技術

用が期待される。

準同型暗号技術の開発は，2009年にIBMによって発表された完全準同型暗号が契機となった。発表当初は複雑な暗号化処理に加え，暗号サイズの肥大化と，秘匿処理にかかる膨大な計算量から実用化には壁があった。その後，マイクロソフト社によってパラメーターなどの生成が容易で使い勝手の良い方法が提案された。また，富士通研究所がアルゴリズムレベルの並列処理により数千倍の処理性能向上を実現するなど，更なる改良が世界中で行われている。

図-5に秘匿検索の利用イメージを示す。クラウド上には，患者の塩基配列を暗号化した状態で格納してある。製薬会社などが特定の部分塩基を持つ患者を検索する場合，暗号化した部分塩基をクラウドに送信することで，該当する患者を検索できる。これに対して準同型暗号を利用することで，患者の塩基データを保護するとともに，機密情報に該当し得る製薬会社が知りたい部分塩基データも第三者に知られることなく，検索処理を実行可能である。富士通研究所では更なる処理性能や使い勝手の向上，および利用シーンの拡大を目指し，実用化に向けた研究開発を進めている。

む　す　び

本稿では，パーソナルデータ利活用に関する法制度の動向を解説し，改正個人情報保護法における匿名加工情報に加工する富士通研究所の匿名化技術，および匿名化技術だけでは対応できない分野に適用できる準同型暗号技術について解説した。富士通の匿名化技術製品NESTGateを用いることで，パーソナルデータを安全に利活用する匿名化技術を，既存システムに対する修正を最小限に抑えた上で導入できる。

今後は，利用分野や目的に応じて，プライバシー保護と利活用の最適なバランスを実現する匿名化手法，および準同型暗号技法の評価を実データを用いて進める。また，国内のみならず海外の法制度にも対応した，安全なパーソナルデータ利活用を実現するプライバシー保護技術の研究開発を進める予定である。

参考文献

(1) 日経ビジネス：Suica履歴販売の失策．2014年7月23日．
http://business.nikkeibp.co.jp/article/opinion/20140718/268916/?rt=nocn

(2) New York Times：Netflix Cancels Contest After Concerns Are Raised About Privacy. 2010年3月12日.
http://www.nytimes.com/2010/03/13/technology/13netflix.html?_r=0

(3) 内閣官房 国会提出法案（第189回 通常国会）：個人情報の保護に関する法律及び行政手続における特定の個人を識別するための番号の利用等に関する法律の一部を改正する法律案．2015年3月10日．
http://www.cas.go.jp/jp/houan/150310/siryou4.pdf

(4) WP216：Opinion 05 2014 on Anonymisation Techniques．2014年4月10日．
http://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/3070528

(5) 津田 宏ほか：安全なクラウド連携のためのデータセキュリティ．FUJITSU, Vol.62, No.5, p.531-537 (2011).
http://img.jp.fujitsu.com/downloads/jp/jmag/vol62-5/paper10.pdf

(6) 山岡裕司ほか：k-存在秘匿性：k-匿名性を拡張した実際的なプライバシーモデル．2015年暗号と情報セ

キュリティシンポジウム，3C4-3，2015年1月.
（7）総務省：匿名データの作成・提供に係るガイドライン．2011年3月28日.
http://www.stat.go.jp/index/seido/pdf/35glv3.pdf
（8）森沢宜広ほか：k-匿名化技術によりパーソナルデータ保護を実現するNESTGate．FUJITSU，Vol.67，No.1，p.34-40（2016）.
http://www.fujitsu.com/jp/documents/about/resources/publications/magazine/backnumber/vol67-1/paper06.pdf

著者紹介

伊藤孝一（いとう　こういち）
知識情報処理研究所
データ・プライバシー保護プロジェクト 所属
現在，匿名化技術の研究開発に従事。

下山武司（しもやま　たけし）
知識情報処理研究所
データ・プライバシー保護プロジェクト 所属
現在，暗号技術の研究に従事。

小暮　淳（こぐれ　じゅん）
応用研究センターライフイノベーション研究所 所属
現在，電子通貨および暗号の研究開発に従事。

津田　宏（つだ　ひろし）
知識情報処理研究所
データ・プライバシー保護プロジェクト 所属
現在，情報漏えい対策の研究開発に従事。