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

NESTGate Realizing Personal Data Protection by k-anonymization Technology

● 森沢宜広　　● 松根伸志

あらまし

個人情報やプライバシー情報（位置情報・経路情報・購買履歴など）を含むパーソナルデータから，特定の個人を識別できないように保護する匿名化技術が注目されている。NESTGateは，富士通研究所のk-匿名化技術を実装したパーソナルデータ保護ツールである。k-匿名化は，情報の中に存在する個人情報に対し，同じ属性を持つ人が少なくともk人以上存在するように情報を加工する技術で，世界中で研究が行われている。NESTGateは，k-匿名化を業務の中で容易に扱えるユーザーインターフェースを備え，大量の情報の中から個人を特定することが困難になるように情報を加工する。機微情報を含む個人情報を処理するケースでは，情報の取扱いそのものが制限される場合がある。このようなケースにおいては，情報にアクセスできる人を識別し情報へのアクセスをコントロールする必要が生じる。認証機能は，情報へのアクセスを制限するアクセス制御を提供する。また，ジョブ管理機能によりジョブの実行状態を監視し，同時に複数の匿名化処理が平行して実行されないようコントロールされる。NESTGateは，クラウドコンピューティングや個人情報を扱う業際システムに迅速に組み込み，活用ができる製品となっている。

本稿では，パーソナルデータを扱う上での留意点，およびNESTGateの機能について述べる。

Abstract

Anonymization technologies are attracting attention since they ensure data including personal and privacy information (such as location information, route information, and purchase history) cannot be used to identify a specific individual. NESTGate is a personal data protection tool implementing Fujitsu Laboratories' k-anonymization technology. k-anonymization, which is a technology to process personal information with the same attribute that exists among at least *k* individuals total information, is being studied all over the world. NESTGate is equipped with a user interface that makes it easy to handle k-anonymization in business and obfuscates information to make it difficult to identify an individual from of a large volume of information. In cases where personal information including sensitive data is processed, handling of the information itself may be restricted. In such cases, in order to control access to the information, the need to identify individuals authorized to access it arises. An authentication feature provides access control that restricts access to the information. A job management feature monitors the state of work execution and provides a control to prevent multiple anonymization processes from being simultaneously executed. NESTGate is a product that can be quickly integrated into cloud computing and business interface systems that handle personal information. This paper describes points to note when handling personal information and the features provided.

ま え が き

クラウドコンピューティングの普及や業際ビジネスの広がり，スマートフォン，タブレット端末などの通信デバイスの多様化により，従来自社システムに蓄積してきた情報が，企業・組織の外部にも保存・蓄積され，活用される時代になってきている。多種多様，かつ膨大な情報を収集・分析するビッグデータ技術の進展により，新たな市場の形成が期待されている。

一方，「個人情報の保護に関する法律」（通称：個人情報保護法）が施行されてから，2015年で10年となった。改正法案が2015年第189回通常国会で可決・成立したことを受け，個人情報の取扱いに関する国民の目は一層厳しいものとなってきており，個人情報の漏えいは計り知れないダメージを企業に与えている。これらの二つの動向から，個人情報やプライバシー情報を含むパーソナルデータを適切に保護し，活用していく技術がこれまで以上に求められる時代になってきていると言える。

パーソナルデータを含む多種多量の情報から特定の個人が識別されるリスクを低減する技術として，匿名化技術が注目されている。匿名化とは，加工後の情報が特定の個人に戻ることがないように情報を加工することをいう。匿名化には複数のアルゴリズムが存在する。どのアルゴリズムを用いて情報を加工するかは，情報の利用目的や公開範囲に応じて，情報を所有し活用する側の判断に委ねることが一般的である。

日常の業務で匿名化を必要とする企業において，そのアルゴリズムを理解し，情報を加工する業務をシステム化することは容易ではない。そのため，特定の属性を削除したり，情報を塗りつぶしたりするなど，情報そのものを無意味なものに加工して活用することが日常的に行われている。しかし，情報の活用という視点で見た場合，このような加工方法では情報の再利用ができなくなっていた。そこで，情報の再利用を可能とする匿名化技術として，k-匿名化が注目されている。k-匿名化は，元の情報をなるべく保持しながら，個人特定のリスクを低減できるよう情報を加工する技術である。

本稿では，k-匿名化技術と本技術を採用したNESTGate（ネストゲート）の機能，および匿名化における留意点について述べる。

パーソナルデータと個人の特定リスク

2005年4月に施行された個人情報保護法では，5000人以上の個人情報をデータベースなどに保有して事業に用いている事業者は個人情報取扱事業者とされ，個人情報の有用性に配慮し，個人の権利・利益を保護することが求められていた。この法律の施行から2015年で10年が経過し，この間のICTの発展により，これまでのデータ利用形態では想定していなかった問題が浮き彫りとなってきた。2015年の第189回通常国会で，改正案となる「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」が提出され，9月に可決された。改正の主なポイントは以下のとおりである（**図-1**）。[1]

（1）個人情報定義の明確化
（2）適切な規律のもとで個人情報などの有用性を確保
（3）個人情報の保護を強化
（4）個人情報保護委員会の新設およびその権限
（5）個人情報取扱いのグローバル化
（6）そのほか

今回の改正では，特定の個人を識別できないように匿名加工した情報（匿名加工情報）については，企業の自由な活用を認めている。情報活用による経済の活性化が盛り込まれている点で，以前とは異なる配慮が見られる。これは，膨大なパーソナルデータを収集・分析するビッグデータ技術の利用を意識したものとなっている。

ここで，パーソナルデータを以下に定義する（**図-2**）。

個人情報保護法に規定する「個人情報（生存する個人に関する情報であって，当該情報に含まれる氏名，生年月日その他の記述等により特定の個人を識別することができるもの。他の情報と容易に照合することができ，それにより特定の個人を識別することができることとなるものを含む）」に限らず，個人が所有するスマートフォンに記録される位置情報や購買履歴など，それ自体では個人の特定はできないものの，特定の個人に結びつく情報として扱われているデータをいう。

1. 個人情報定義の明確化	・個人情報定義の明確化（身体的特徴などが該当） ・要配慮個人情報（仮称，いわゆる機微情報）に関する規定の整備
2. 適切な規律のもとで個人情報などの有用性を確保	・匿名加工情報（仮称）に関する加工方法や取扱いなどの規定の整備 ・個人情報保護指針の作成や届け出，公表などの規定の整備
3. 個人情報の保護を強化（名簿屋対策）	・トレーサビリティの確保（第三者提供に係る確認および記録の作成義務） ・不正な利益を図る目的による個人情報データベース提供罪の新設により個人情報データベースを提供する行為を処罰化
4. 個人情報保護委員会の新設およびその権限	・個人情報保護委員会を新設し，現行の主務大臣の権限を一元化
5. 個人情報の取扱いのグローバル化	・国境を越えた適用と外国執行当局への情報提供に関する規定の整備 ・外国にある第三者への個人データ提供に関する規定の整備
6. そのほか改正事項	・本人同意を得ない第三者提供（オプトアウト規定）の届け出，公表などの厳格化 ・利用目的の変更を可能とする規定の整備 ・取り扱う個人情報が 5000 人以下の小規模取扱事業者への対応

図-1　個人情報保護法の改正のポイント

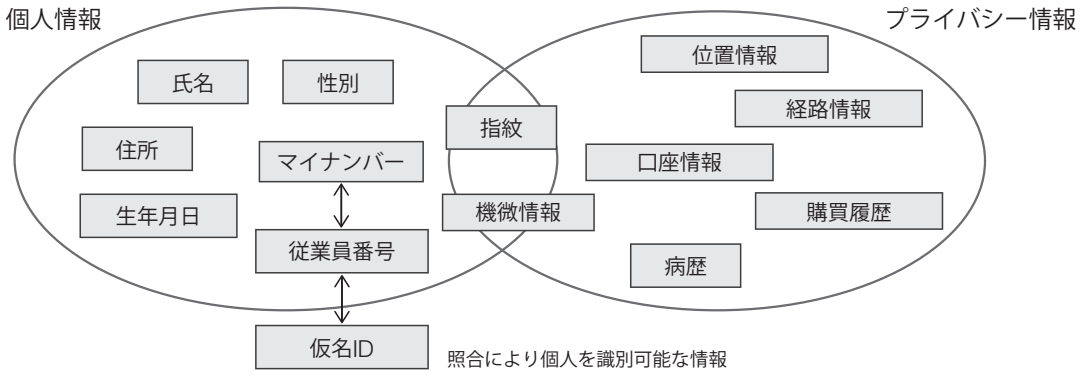

図-2　パーソナルデータ

次に，パーソナルデータにおける個人の特定リスクについて説明する。

図-3①は，個人情報（氏名）を削除した疾病履歴情報である。一般には，このように個人情報（氏名）を削除するような加工が施されている場合，この情報が個人情報を含んでいるとはみなされない。しかし，本図の中に65歳の男性である中村さんが入っていることを知っている者が見た場合，中村さんのレコードを一意に特定できてしまう。この例では，個人情報保護法に規定しているように氏名を削除しているが，図中のほかの属性を照合することで個人が特定されてしまうリスクが存在することを示している。

このように，個人情報とは直接関係のない属性を組み合わせることで個人が特定されるリスクを回避し，個人のプライバシーにも配慮したいというニーズに応えるため，PPDM（プライバシー保護データマイニング）の研究が行われており，ビッグデータの加工への適用が期待されている。PPDMでは，個人が特定されるリスクを低減する匿名化技術が注目されており，k-匿名化はこの技術の一つである。

図-3②は，同図①にk-匿名化処理を施したものである。この例では，年齢のフィールドを1歳単位から10歳単位に加工し直すことで，①で特定されていた中村さんのレコードは一意性がなくなり，特定できなくなっている。

k-匿名化は，属性を組み合わせてもk人未満にはレコードが特定できないよう情報を加工する技術であり，様々な研究機関が研究を行っている。

①氏名の削除

氏名	電話番号	性別	年齢	疾患	年収（万円）
山本	03-...	男	40	糖尿病	600
中村	03-...	男	65	糖尿病	400
小山	03-...	男	60	糖尿病	500
山本	03-...	男	43	糖尿病	700
佐藤	03-...	男	48	なし	800
田中	03-...	女	26	糖尿病	400
中本	03-...	女	22	なし	300

②k-匿名化

	電話番号	性別	年齢	疾患	年収（万円）
山本	03-...	男	40代	糖尿病	600
中村	03-...	男	60代	糖尿病	400
小山	03-...	男	60代	糖尿病	500
山本	03-...	男	40代	糖尿病	700
佐藤	03-...	男	40代	なし	800
田中	03-...	女	20代	糖尿病	400
中本	03-...	女	20代	なし	300

図-3　k-匿名化の例

NESTGateの匿名化処理

k-匿名化は，パーソナルデータの活用において注目されている技術であり，NESTGateはこの技術を実装したパッケージソフトウェア製品である。NESTGateは，先に述べたk-匿名化に加え，様々な活用シーンを考慮し「匿名データの作成・提供に係るガイドライン」（総務省2010年2月制定）[(2)]に記載されている統計法に基づく匿名化処理も実装している。

本章では，NESTGateが実装する匿名化処理の概要を説明する。なお，各匿名化処理は，単独または組み合わせて使用できる。

（1）k-匿名化

NESTGateが実装するk-匿名化アルゴリズムは，次の3種類となる。

・一般化による匿名化

QI（Quasi-Identifier：準識別子）を一般化し，k-匿名化を行うアルゴリズムである。準識別子とは，識別子（氏名など個人を直接特定できる情報）ではないが，組み合わせることで個人を特定する可能性がある情報をいう。図-3の例では，年齢，性別がQIとなる。

一般化による匿名化では，QIの曖昧化により情報を保護する。

・情報の存在性を優先した匿名化

QIにノイズを加え，情報の存在性を優先する匿名化であり，情報の欠損はないが，QIの値は元情報と異なる値となる。

・項目欠損化による匿名化

QIを欠損させることで特定性をなくす匿名化であり，QIの利用はできなくなるが，個人特定リスクは少なくなる。

上記の3種類のアルゴリズムは，いずれもk値を指定することで同じ属性の行が少なくともkレコード以上になるように情報を加工する。その際，情報の存在性を優先するのか，欠損により情報を保護することを優先させるのかを利用者が選択する。

情報の利用目的や公開範囲，情報の質（機微な情報など）により，これらのアルゴリズムのどれを使用するのかを判断する。

（2）ソート

匿名化処理後の行の並びが元情報と同一となっていた場合，元情報の並びを知っている者は，匿名化後のデータであっても容易に誰のものかを推測できてしまう。このようなリスクを回避するため，NESTGateでは匿名化処理の最後でソート処理を行うことを推奨している。一般的なソート処理は辞書順に並べ替えるものであるが，NESTGateにおけるソート処理は，同じ属性を持つ行同士を並べて出力するように行を並べ替える。これにより，行の並びが元情報と同一とならず，シャッフル効果を得られる。

（3）項目変換

ある属性の値を別の値に置換する。置換文字列として正規表現を指定できるため，多様な文字列の置換が可能となっている。例えば，以下のような変換が可能である。

・属性のトップコーディング

属性の値が設定したしきい値以上である場合に，別の値に置換する。例えば，年齢が50歳以上の場合，年齢の値を「50歳以上」に置換する。

・属性のボトムコーディング

属性の値が設定したしきい値以下である場合に，別の値に置換する。例えば，年齢が50歳以下の場合，年齢の値を「50歳以下」に置換する。

・属性のグルーピング

属性の値を設定したグループにまとめ，別の値に置換する。例えば，年齢の値を「10代」「20代」などに置換する。

（4）スワッピング

ある属性の値を別の行の値と交換する。交換する値は，元の値と交換する値に近いものが選択される。

（5）リサンプリング

元情報から行を間引く。存続させる行の割合と間引くレコードをランダムに指定することが可能となっている。

（6）誤差

指定した属性にランダムにノイズを加える。

（7）属性削除

元情報から行，または列を削除する。氏名など，直接個人に関連する属性である識別子の削除に使用することを想定している。

NESTGateの機能構成

NESTGateの機能は，**図-4**に示す機能ブロックに分けられる。ここでは，利用者が接するユーザーインターフェース，ジョブ管理機能，認証機能について説明する。

（1）ユーザーインターフェース

NESTGateの匿名化機能を操作するためのインターフェースを提供する。NESTGateは，匿名化を行うサーバとユーザーが操作するクライ

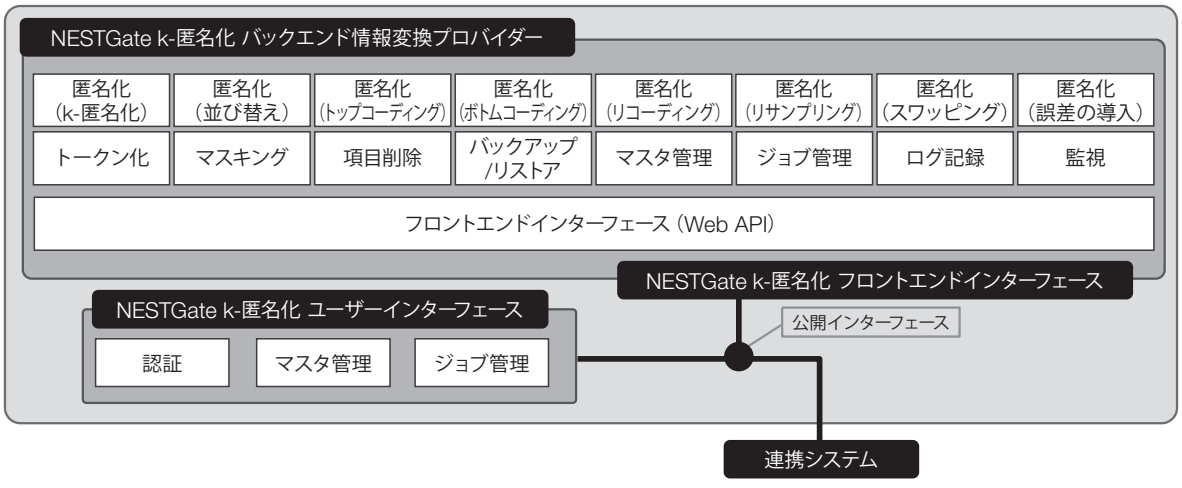

図-4　NESTGateの機能構成

アントの機能を提供する。サーバとクライアントの間は，HTTPメッセージベースの通信を行う仕組みとしており，メッセージプロトコルにREST（Representational State Transfer）を採用している。

近年，Webベースのシステムが様々な業務で使用されるようになり，既存システムとの連携や統合も可能である。また，提供するインターフェースを使用して，ニーズに合わせて容易にカスタマイズが可能である。標準で提供するクライアント機能の実装は，マイクロソフト社のスマートクライアント技術を採用している。

（2）ジョブ管理機能

NESTGateが行う匿名化は，入力・出力ともにCSVファイルを対象としてバッチ処理する。匿名化処理は，同一サーバ上では並行動作は行わず1ジョブ単位で処理される。このため，匿名化を行うジョブはほかのジョブの処理中には待機状態となっており，こうしたジョブの実行を管理する機能を提供する。

ジョブ管理機能は，匿名化の要求を受け付け，匿名化処理を行い，結果をクライアントに返却するまでの一連のジョブの実行を管理する。また，ジョブのキャンセル機能なども併せ持っており，これらの機能は全てサーバが提供するWebサービスAPI（Application Programming Interface）を通じて行われている。

（3）認証機能

クライアントがNESTGateの匿名化機能をWebシステムなどから利用するケースでは，認証機能は非常に重要である。認証機能によって，正規の利用者のみがNESTGateの機能が使用できるようアクセス制御が行われる。また，匿名化をする情報に機微情報が含まれている場合，情報そのものへのアクセスが制限される場合がある。NESTGateの認証機能は，このようなケースにおいて情報へのアクセスコントロールを実現する。

匿名化の留意点

ICTの発展は，これからも人々の暮らしを便利に，豊かにしていくことが期待されている。一方，プライバシー情報を含むパーソナルデータの扱いについては，情報を所有する側が細心の注意を払って利用することが求められる。匿名加工情報の扱いについては，個人情報保護法の改正で，その取扱い規定が明確になる方向であるが，本稿で述べた匿名化技術は，情報の保護と活用の両面で役立つ技術であると考えている。

しかし，匿名加工した情報が再び識別されることがないという保証が完全にはできない事例もある。匿名化の難しさがよく分かる事例を以下に紹介する。

1997年，マサチューセッツ州が公開した匿名化後の医療情報から，州知事の情報が特定された。この事例では，医療情報は個人を特定する識別子である氏名を削除して公開されたが，一般販売されている投票者名簿との照合により，医療情報内の州知事の情報が名前とともに再識別可能となってしまった。この事例は，匿名化後の情報と外部に公開されている情報の照合により，個人が特定された有名な事例である。

NESTGateは，複数の匿名化アルゴリズムを搭載したパーソナルデータ保護ツールである。しかし，匿名化を行う際は，識別子や準識別子について，利用内容，公開範囲，漏えいリスクを十分検討した上で使用することをお願いしたい。

む　す　び

本稿では，パーソナルデータを利活用する際に必要となる匿名化や富士通が提供するNESTGateの機能について述べた。k-匿名化は，パーソナルデータを扱っていく上で注目されている技術であり，本技術を実装したNESTgateを是非活用いただきたい。

今後は，k-匿名化とともに研究が進められているl（エル）-多用性や準同型暗号などの技術の実用化を期待している。

参考文献

（1）個人情報の保護に関する法律.
http://law.e-gov.go.jp/htmldata/H15/H15HO057.html
http://www.cas.go.jp/jp/houan/189.html
（2）総務省：匿名データの作成・提供に係るガイドライン.
http://www.stat.go.jp/index/seido/pdf/35glv3.pdf

著者紹介

森沢宜広（もりさわ　よしひろ）
（株）富士通システムズ・ウエスト
ビジネスソリューション本部スマート
コンテンツソリューション事業部 所属
現在，セキュリティ分野を中心とした
ソリューションビジネスに従事。

松根伸志（まつね　しんじ）
（株）富士通システムズ・ウエスト
ビジネスソリューション本部スマート
コンテンツソリューション事業部 所属
現在，NESTGateの開発に従事。