データ駆動型社会を加速する個人データ流通基盤の信頼性向上技術

Technologies for Improving Reliability of Personal Data Distribution Platforms to Accelerate Data-Driven Society

山岡　裕司　　小倉　孝夫　　小栗　秀暢　　伊藤　孝一

あらまし

世界中でデータ活用に期待が高まる中，日本政府は特に個人の関与の下で個人データ流通・活用を進める事業である情報銀行を推進しており，情報銀行の信頼性を確保するため，民間に認定制度を定めるよう促している。民間はそれを受けて，情報銀行事業を審査・認定する事業を開始した。こうした世の中の動向を鑑みて富士通研究所は，データ漏えいのリスクを定量化することでマネジメント可能にするプライバシーリスク評価技術，および分散した個人データに対する本人同意を集約管理可能な同意管理技術を開発した。これらの技術によって，個人データを安全に流通させることが可能となり，プライバシーに配慮したデータ駆動型社会の実現に寄与するものと考える。

本稿では，プライバシーリスク評価技術および同意管理技術について，技術の概要，適用事例，および評価結果を述べる。

Abstract

As expectations for data utilization are rising around the world, the Japanese government is moving ahead with its information bank project, which promotes the distribution and utilization of personal data with the involvement of individuals, in particular. In order to ensure the credibility of these information banks, the government is encouraging the private sector to establish certification systems. In response, the private sector has launched projects to review and certify information bank businesses. In view of this trend, Fujitsu Laboratories has developed privacy risk assessment technology that facilitates management by quantifying data leakage risks and consent management technology capable of aggregating and managing individual consent for distributed personal data. These technologies allow for secure distribution of personal data and contribute to the realization of a data-driven society with consideration given to privacy. This paper describes the privacy risk assessment technology and consent management technology by presenting outlines of the technologies, examples of their application, and evaluation results.

1. まえがき

世界中でデータ活用に期待が高まる中，日本政府はSociety 5.0構想[1] などを策定し，特に個人データの流通を推進している。例えば，総務省は個人の関与のもとでデータ流通・活用を進める事業である情報銀行を推進し，その信頼性を確保するために民間が認定制度を設けるよう促している。それを受けて，一般社団法人日本IT団体連盟が，2018年末に情報銀行事業を審査・認定する業務を開始した。[2]

富士通は，情報銀行を実現する基盤としてPersoniumサービス[3] を提供している。更に，それを実現する技術である分散型PDS（Personal Data Store）の実用化などの活動が産業の発展に寄与したと認められ，第66回 電気科学技術奨励賞を受賞している。

情報銀行は個人データを流通させるため，一般にその運用リスクは大きいとされる。情報の安全管理措置に加え，データ提供先からの漏えい事故の対策も必須である。更に認定制度においては，データ提供先からの漏えい事故などによる個人への損害賠償責任は情報銀行が負う運用を求めている。

こうした世の中の動向を踏まえて富士通研究所では，長期間安定的に情報銀行の運用を継続するために必要な個人データ漏えいリスクの定量化やマネジメントを支援するプライバシーリスク評価技術，[4] および分散した個人データの同意情報を集約管理可能な同意管理技術を開発した。

本稿では，プライバシーリスク評価技術および同意管理技術について，技術の概要，適用事例，および評価結果について述べる。

2. プライバシーリスク評価技術

本章では，プライバシーリスク評価技術の概要，および適用事例について述べる。

2.1 技術の概要

プライバシーリスク評価技術は，パーソナルデータが暴露するプライバシーのリスクを定量化・金額化する技術である（図-1）。従来不足していた，匿名加工後のデータの特定性（匿名性の低さ）を算定するモデルおよび高速算定技術が要であり，一般的な性能のパソコンを使用した場合，100万人規模の実際のデータセットを約1時間で算定できる十分な実用性を確認している。

情報銀行では個人データの漏えいについて，従来は漏えい時の損害賠償額の定量的な目安がなかったため，リスク低減措置が行われず潜在的なリスクを抱えてしまっていたという問題を，この技術は解決できる。

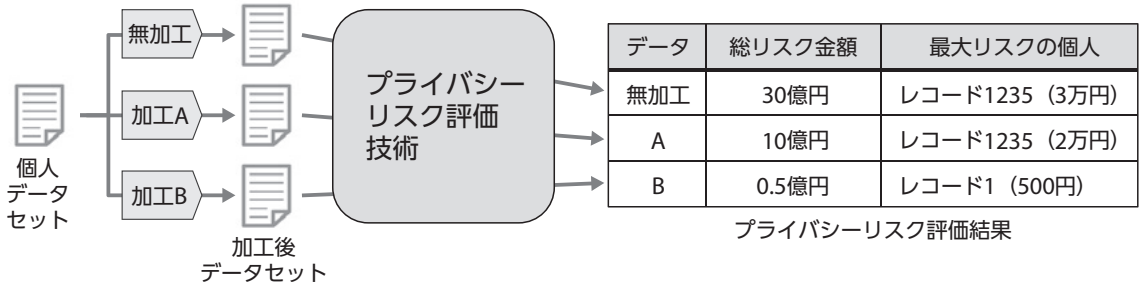
2.2 適用事例と評価

プライバシーリスク評価技術の新たな適用事例として，自治体データ，および医療データの二つについて述べる。

(1) 自治体データ

自治体におけるプライバシーリスク評価の適用事例として，大津市と富士通が共同で実施したデータ分析の実証実験がある。[5]

官民データ活用推進基本法の成立以降，多くの自治体がデータの利活用推進計画の作成を試みている。しかし，庁内でのデータの移動や加工が制限



データ	総リスク金額	最大リスクの個人
無加工	30億円	レコード1235（3万円）
A	10億円	レコード1235（2万円）
B	0.5億円	レコード1（500円）

プライバシーリスク評価結果

図-1　プライバシーリスク評価技術

されているため，複数の部署が保有するデータを結合して，自由な分析や議論を行うことができなかった。

それに対して大津市と富士通は，データを「FUJITSU ビジネスアプリケーション NESTGate 匿名化 V1」[6] を用いて匿名化し安全性を高めることで，出産・育児に関する15種類のデータベースを庁内から収集して結合し，実際の政策に活用することにした。しかし，データを匿名化しただけでは，自治体の担当者はデータの利用可否を判定できない。加工後のデータのリスクが十分に小さいことを見える化し，エビデンスとして残す活動は，市民への説明責任のある自治体運営において不可欠である。

そこで，プライバシーリスク評価ツールによって，各種の匿名化データを評価した（図-2）。この評価では，国籍/性別，住所，誕生日の三つの個人データに対して，以下のような匿名加工を行い，それぞれを組み合わせてリスクを評価した。

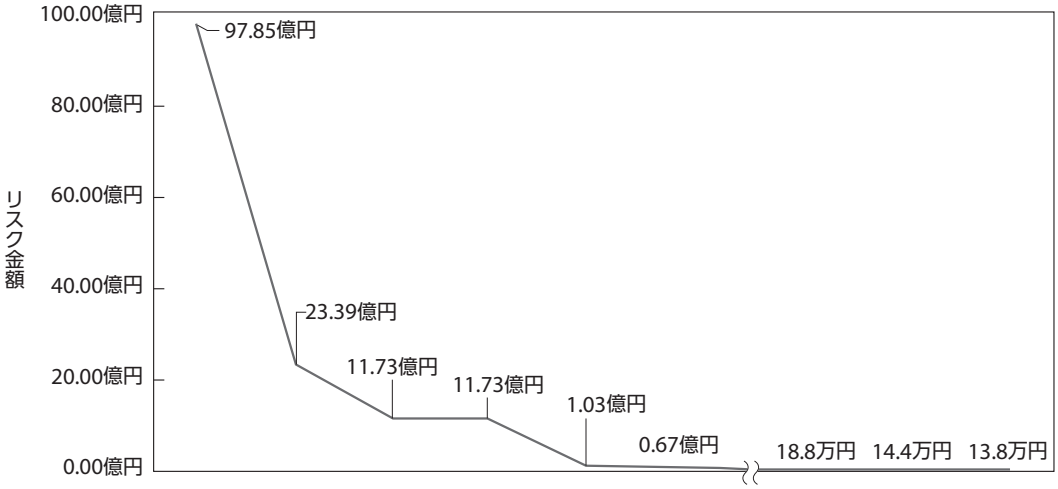
・国籍/性別 不明者の扱い
　あり：不明者を含む
　なし：不明者を削除する
・住所
　地域：大津市の地域（231地域）
　学区：大津市の学区（37学区）
　地区：庁内で定義した九つの地域群分類
・誕生日
　元データ：生年月日（加工しない）
　日付加工：生年＋月まで利用
　学制：小学校/中学校/高校/大学で分類

匿名加工する前の元データの総リスク値は約98億円であるのに対して，最も強い加工を行った場合（国籍/性別：全てなし，住所：地区，誕生日：学制）の値は約14万円となり，元データの0.001％までリスクを減少させたことを具体的に示した。

この結果を基に大津市は，元データの0.1％以下のリスクであるデータは庁内の簡易な申請で利用できる，というポリシーを定めた。このような安全性の定量化は，庁内におけるパーソナルデータの有用性と危険性を共有する効果もある。そのため，データ利活用の推進には非常に有益であると，関係者から評価をいただいた。

(2) 医療データ

2016年1月に，「がん登録等の推進に関する法律」が施行された。この法律のもと，がんの診療・経過などに関する情報を集め，保管・整理・分析する仕組みが，全国がん登録[7] である。これによって，がんに関する情報を国立がん研究センターがデータベース化し，国や都道府県が利用できるようになっている。また，匿名化された情報は，がんに係る研



国籍/性別不明	あり	あり	国籍なし	全てなし	あり	国籍なし	…	あり	国籍なし	全てなし
住所	地域	地域	地域	地域	地域	地域	…	地区	地区	地区
誕生日	元データ	日付加工	日付加工	日付加工	1年	1年	…	学制	学制	学制

図-2　大津市における匿名化の効果

究調査を希望する研究者に提供される。富士通は国立がん研究センターとの共同研究で，匿名化された情報の利用を想定して，プライバシーリスク評価技術の評価を行った。

がん登録という特殊な集団の情報の個人識別リスクについて，匿名化された「地域がん登録」の研究目的提供データに対し，プライバシーリスク評価技術を用いて，匿名化レベルとプライバシーリスクの関係性を評価した。[8] この評価では，年齢，住所，診断日の三つの個人データに対して，以下のような匿名加工を行い，それぞれを組み合わせてリスクを評価した。

・年齢
　10歳階級：年齢を10年単位に変換
　10歳階級＋若年/高齢まとめ：10年単位＋20歳以下，90代以上を集約
・住所
　市町村：県内の市町村名
　医療圏：政府が定義する医療圏（県内4分類）
・診断日
　日単位：診断日（無加工）
　年単位：診断日を年まで加工

プライバシーリスクは最も強い匿名化を行った場合，元データの48％まで低減できることを確認した（図-3）。

これによって，本技術はがん登録ユーザーの匿名化された情報でも，一般人の判断力や理解力による個人識別リスクについて，一定の判断基準を与えると評価された。

3. 同意管理技術

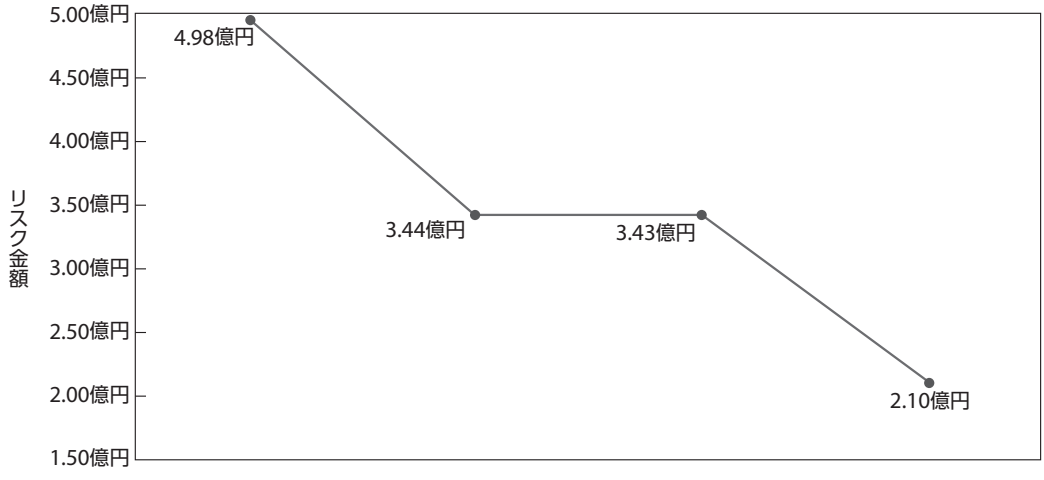
本章では，同意管理技術の概要，およびこの技術を用いた性能評価について述べる。

3.1 技術の概要

同意管理技術は，個人の同意に関する情報を集約管理しながら，各事業者で分散管理されている個人データを安全に流通させる技術である。

情報銀行は，データを中央サーバに集約管理して流通させることが想定されているが，個人データについてはこの枠組みの外部において，別のサービス事業者などが管理することが多い。例えば，運転履歴などの個人データは更新頻度が高く，集約せずに流通できることが望ましい。一方，個人データの活用には原則として本人の同意が必須であり，同意の結果は効率よく集約する方が良い。

このように，データ種別によって管理体系が異な



年齢	10歳階級	10歳階級	10歳階級＋高齢/若年まとめ	10歳階級
住所	市町村	市町村	市町村	医療圏
診断日	日単位	年単位	年単位	年単位

図-3　がん登録リスク評価結果（病名の機微性を高く設定した場合）

ることが問題である。そこで，アクセス管理プロトコルUMA（User Managed Access）[9] を，複数人の個人データを効率的に流通できるように拡張することで，この問題を解決した。

(1) 同意ポータル

同意ポータルは，データ提供事業者からデータ利用事業者へ個人データを提供する際に，データ提供事業者にデータを預けているユーザーの同意をデータ利用事業者の代理で取得し，その記録を管理するシステムである（図-4）。これによって，ユーザーは過去の同意記録，提供事業者，提供済み個人データの内容などを照会できる。更に，サービスの利用規約や同意画面は事業者間で共通的に利用でき，過去に同意した利用規約，同意内容の差分を明確化できるため，ユーザーの同意可否の判断が容易になる。

同意ポータルを活用することで，サービスを提供するデータ利用事業者は，利用規約，同意内容の文章や画面，同意記録システムなどを独自に開発する負担が軽減される。例えば，テレマティクス保険サービスの場合，データ利用事業者である保険会社が，データ提供事業者である自動車会社から個人の運転履歴データを取得して，保険料を算出する。このサービスにおいて，同意ポータルは個人との窓口になることで，同意取得が容易となる。

(2) 個人データ一括取得方式

年代，性別などの属性を切り口にした個人データの分析に向けて，同じ属性の個人データを一括取得する技術を開発した。アクセス管理プロトコル標準であるUMAでは，個人のデータをデータ提供事業者とデータ利用事業者などの異なる組織間で共有し，一か所でアクセスを制御する。UMAは，個人単位のデータ取得を目的として設計されている。そのため，例えば1万人におよぶ大規模なユーザーからデータを取得する場合，大量のメッセージが発生することが問題となる。

これに対して富士通研究所では，一回分のUMAプロトコルシーケンスで同じ属性のデータを一括取得することを実現した。同意ポータルは，同意済みの複数ユーザーのデータ群を仮想的に一人分のデータとして扱い，このデータ群に一括してアクセス権を設定する。これによって，データ利用事業者は個人データを一括で取得可能となる。

また，取得するデータの属性条件（年代，性別など）をデータリクエストのURI（Unified Resource Identifier）に記述することで，UMAプロトコルに

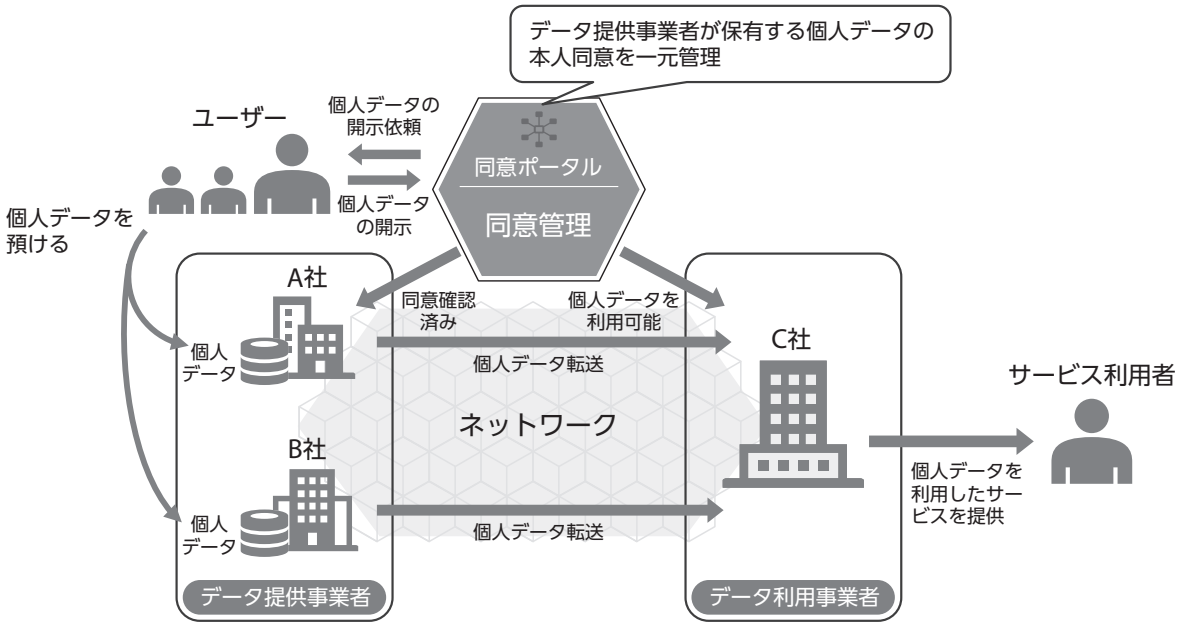


図-4　同意ポータル

おけるデータ利用事業者側のインターフェースの互換性に配慮した。これによって，データ利用事業者はUMAによる個人単位のデータ取得と一括でのデータ取得の両方が利用可能となる。例えば，UMAをサポートするオープンソースを用いて，様々なデータ利活用サービスの開発が可能となる。

3. 2 性能評価

同意管理技術の評価として，テレマティクス保険サービスを想定したシステムを試作し，個人データ一括取得方式の性能を評価した。このサービスでは，保険会社が車の運転履歴を利用できるように，契約の際に事前に同意を取得することを前提とした。

性能評価の環境について，一台のサーバ(注)上に各事業者が使用可能な同意ポータルを構築し，別のサーバに測定ツールを構築した。測定ツールから対象サーバへデータ取得のHTTPリクエストを実行し，応答時間を計測した。

性能評価では，以下の三つの方式でデータ利用事業者に見立てた測定ツールからリクエストし，200人の個人データを取得するまでの時間を測定した。個人データのサイズは，10 KB，100 KB，1 MBと変化させた。

・一括取得方式：1回のリクエストで200人分の個人データを取得する。
・シーケンシャル方式：測定ツール側からリクエストを送信し，応答を受信した後に次のリクエストを送信する。これを，200回繰り返す。
・パラレル方式：測定ツール側から10ミリ秒おきに200リクエストを送信する。

性能測定の結果，一括取得方式では，シーケンシャル方式やパラレル方式と比較して，6倍～236倍の高速化効果が得られた（図-5）。

これらの結果を基に，1万人の運転履歴データ（一人あたり約100 KB想定）を一括取得するための時間を算出した。10 Gbpsの伝送速度を前提とした場合，UMAプロトコル処理に約800ミリ秒（実験値），個人データの転送に約3秒（10 Gbpsの伝送速度から算出）を要するため，約3.8秒で1万人分のデータ取得が可能である。

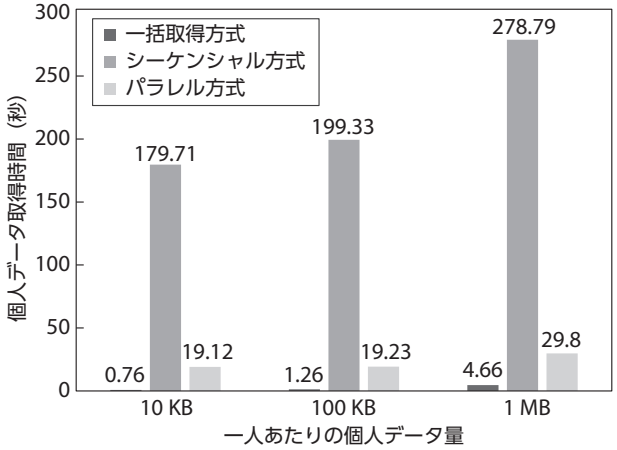


図-5　同意管理技術の性能評価

このように，UMAの拡張によって，複数人の個人データの効率的な流通が可能となる。

4. むすび

本稿では，個人データ流通事業の信頼性を高めるために富士通研究所が開発した，プライバシーリスク評価技術および同意管理技術について述べた。これらの技術によって，個人データの流通に同意しやすくなるだけでなく，流通先で新たなデータ分析が可能となる。こうした仕組みは，分析精度の向上にもつながることから，データ駆動型社会の実現に寄与するものと考える。

今後，本技術を実用化することで個人データ流通事業の信頼性向上を実現し，個人データの流通・活用を拡大させ，ひいては経済発展や社会的課題の解決につなげたい。

（注）　CPU：E5-2660 2.60 GHz，10コア，メモリ：64 GB。

参考文献

（1）内閣府：Society 5.0.
https://www8.cao.go.jp/cstp/society5_0/
（2）一般社団法人日本IT団体連盟：情報銀行推進委員会.
https://www.tpdms.jp/
（3）富士通：Personiumサービス.
https://jp.fujitsu.com/solutions/cloud/k5/function/

paas/personium/
(4) 山岡裕司：安全な社会をデジタルでストレスなく守るセキュリティ．FUJITSU，Vol.69，No.5，p.55-60 (2018)．
https://www.fujitsu.com/jp/documents/about/resources/publications/magazine/backnumber/vol69-5/paper08.pdf
(5) 大津市，富士通：大津市と富士通、ICT活用およびデータ分析分野における連携協定を締結．
https://pr.fujitsu.com/jp/news/2018/11/8.html
(6) 富士通：FUJITSU ビジネスアプリケーション NESTGate 匿名化 V1.
https://www.fujitsu.com/jp/solutions/business-technology/intelligent-data-services/bigdata/ba-solutions/nestgate/anony/index.html
(7) 国立研究開発法人国立がん研究センター：全国がん登録，登録情報の提供．
https://ganjoho.jp/reg_stat/can_reg/national/datause/general.html
(8) 日本がん登録協議会：がん登録・がん統計について．
http://www.jacr.info/about/registry.html
(9) Kantara Intiative：User Managed Access.
https://kantarainitiative.org/confluence/display/uma/

著者紹介

山岡　裕司（やまおか　ゆうじ）
（株）富士通研究所
セキュリティ研究所
データ・プライバシー保護技術の研究開発に従事。

小倉　孝夫（おぐら　たかお）
（株）富士通研究所
スーパーミドルウェア・ユニット
データ流通・利活用技術の研究開発に従事。

小栗　秀暢（おぐり　ひでのぶ）
（株）富士通研究所
セキュリティ研究所
データ・プライバシー保護技術の研究開発に従事。

伊藤　孝一（いとう　こういち）
欧州富士通研究所
セキュリティグループ
欧州におけるデータ保護技術の研究開発に従事。