Vol.2019-IS-150 No.2
2019/11/23

パーソナルデータの同意取得と一般化の考察

柿崎 淑郎[1,a)]　稲村 勝樹[1]　松井 加奈絵[1]

概要：パーソナルデータは，個人情報，行動履歴や購買履歴などの個人に関する属性情報を含む，個人に関わる情報である．パーソナルデータの多くを個人情報が占めるが，個人情報保護法により，個人情報を第三者提供するには，本人同意が原則であり，あるいは，個人情報を匿名加工情報に加工することで同意不要とするかのいずれかとなる．本人同意を取得する場合，許可または不許可の二択であることがほとんである．対して，匿名加工情報は個人が特定できないように一般化などの加工を行うため，情報の精度が失われる．端的に言えば，第三者提供される個人情報は無加工の情報か，匿名加工されて精度が失われた情報のいずれかということになる．本稿では，これらの中間に位置づけられる粒度制御可能な属性情報加工手法と提供手法について検討し，パーソナルデータの一般化と同意取得への影響について考察する．

Consideration of Obtaining Consent and Generalization for Personal Data

Yoshio KAKIZAKI[1,a)]　Masaki INAMURA[1]　Kanae MATSUI[1]

1. はじめに

個人情報の保護に関する法律（個人情報保護法）[*1]は 2003 年に成立，施行された個人情報の取り扱いに関する法律である．個人情報保護法は OECD8 原則 [1] に対応し，個人情報の適正な取得，利用目的の特定，利用目的の制限，利用目的の通知，第三者提供の制限などが個人情報取扱事業者の義務として定められている．つまり，個人情報の取得，目的外利用，第三者提供などは本人同意が原則である．

個人情報保護法は 2015 年の改正によって，匿名加工情報が定義された．匿名加工情報は特定の個人を識別することができないように個人情報を加工し，当該個人情報を復元できないようにした情報である．そのため，匿名加工情報は個人情報ではないため，適切な加工，安全管理措置，公表義務，識別行為の禁止の 4 つの義務を守ることで，本人同意を得ることなく，匿名加工情報を第三者に提供することができる．

[1] 東京電機大学
Tokyo Denki Uniersity
[a)] kakizaki@mail.dendai.ac.jp

[*1] https://elaws.e-gov.go.jp/search/elawsSearch/elaws_search/lsg0500/detail?lawId=415AC0000000057

個人情報の取得に際して，本人同意を得る場合，許可あるいは不許可の二択であることが多い．本人同意によって個人情報が提供される場合，取得される情報の値は無加工であることがほとんどである．無加工の情報から匿名加工情報までの間には，中間的な加工状態が存在するが，匿名加工情報に至るまでは個人情報であり，それらの取得にも本人同意が必要になる．

個人情報よりも広範な概念としてパーソナルデータがあり，行動履歴や購買履歴，ヘルスケア情報などの個人に関する属性情報やプライバシ情報も含まれている．これらプライバシに関わる情報の場合，利用者はある程度ぼやかした情報ならば提供に同意できることも少なくない．例えば，自分がいる場所の天気予報を知りたいとしたときに，天気予報サービスに対して，GPS（Global Positioning System）による精緻な位置情報は提供したくないが，市区町村レベルの位置情報なら提供しても構わないという判断があり得る．このように，パーソナルデータの取得においては，本人同意が原則であるとともに，プライバシ保護の観点からは，どの程度の精緻さで提供するかも重要な視点となる．

本稿では，プライバシ保護の観点から，無加工のパーソナルデータと匿名加工情報の中間に位置づけられる粒度

表 1 パーソナルデータの例

属性				
識別子	準識別子		センシティブ属性	
ID	性別	年齢	病歴	年収
1001	男	45	心疾患	750
1002	男	42	ガン	640
1003	女	48	ガン	650
1004	女	41	心疾患	550

制御可能な属性情報加工手法と提供手法について検討し，パーソナルデータの一般化と同意取得への影響について考察する．

2. パーソナルデータ

パーソナルデータは，SNS のプロファイルなどの個人によって作成され明示的に共有される自発的データ（Volunteered data），位置情報などの個人の行動による観測データ（Observed data），自発的または観測データから分析される推定データ（Inferred data）に分類される [5]．

自発的データは自らが明示的に共有しているデータではあるが，プライバシ問題を起こしかねない．観測データは監視カメラやセンサ機器などから収集されるデータであり，IoT の普及に伴い，膨大かつ多岐にわたるデータが記録されているが，観測されていること自体に個人が気づかないことも少なくない．そして，推定データはこれらの自発的データと観測データから推定されるものであり，その精度は推定に用いるデータ量などに依存するが，その推定データが存在するあるいは推定されていること自体に個人が気づくことは明示されない限り難しい．そのため，プライバシの観点から，個人に関する推定データを推定されることを防ぐためには，推定データのオプトインを拒否するか，自発的データと観測データを抑制しなければならない．

パーソナルデータは表 1 に示すように，属性と属性値の表形式で表現され，1 レコードはある個人 1 名のデータを表す．属性のうち，その属性値が直接的に個人を特定できる属性を識別子（Identifier）という．例えば，名前，電話番号，メールアドレスなどが識別子となる．識別子ではない属性のうち，他の属性と組み合わせることで，識別子と同様に個人を特定できるような属性を準識別子（Quasi-Identifier; QI）という．準識別子の例として，性別，生年月日，郵便番号があり，これら 3 つの属性が組み合わさることで，アメリカ合衆国において 87.1%の個人を識別することができる [2]．個人が特定された状態で，公開されることが望ましくない属性をセンシティブ属性（Sensitive Attribute; SA）という．

2.1 属性値の一般化

属性値の一般化は，属性値をより抽象度が高い上位の値や上位の概念に置き換えることで行われる．例えば，購買情報が“メロンパン，鮭おにぎり，サイダー”のとき，“パン，おにぎり，炭酸飲料”としたり，“食料品，食料品，飲料品”としたりして，一般化を行う．他の例として，年齢が 35 の属性を一般化するとき，0 歳から 10 歳刻みに一般化すると“31〜40”となり，20 歳以上か否かで一般化すると“20 歳以上”となり，1 の位を四捨五入すると“40”となる．このような一般化は一般化階層木で表現される．一般化階層木は一般化前後での属性値の関係を階層的に示すが，どのように一般化するか，あるいは，抽象度をどのように定めるかは，一般化階層木の作成者に依存する．

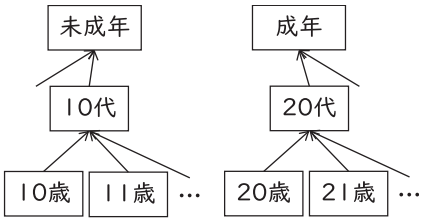

図 1 値一般化階層の例

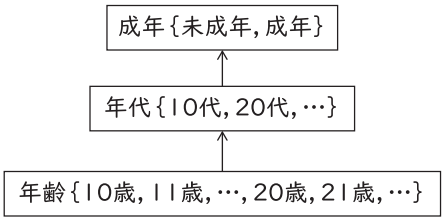

図 2 ドメイン一般化階層の例

属性値を一般化する方法として，値一般化階層（Value Generalization Hierarchy; VGH）とドメイン一般化階層（Domain Generalization Hierarchy; DGH）がある [6]．図 1 に値一般化階層の例，図 2 にドメイン一般化階層の例を示す．値一般化階層はある属性値を抽象化したとしても，他の属性値が抽象化されない．つまり，値一般化階層では異なる抽象度の属性値が混在しうる．一方で，ドメイン一般化階層ではある属性値を一般化する場合，すべての属性値が同じ抽象度に抽象化される．

属性値を一般化する時に，同じ属性値であれば全て同じように抽象化する大域的符号化（global recoding）と，同じ属性値であっても異なる抽象化を行う局所的符号化（local recoding）がある．例えば，GPS に基づく位置情報において，緯度経度の情報を 10 進数で表して，ある桁で丸める抽象化を行うと，山間部などで人口密度の差が大きいような場合に，人口密度が低い地域の抽象度を高くするなどの対応を行うのが，局所的符号化となる．

2.2 パーソナルデータの匿名化

匿名化されたパーソナルデータの安全性指標の 1 つとして，k-匿名性がある．匿名化されたデータから $1/k$ 未満の確率でしか個人を識別できないとき，そのデータは k-匿名性を持つといい，パーソナルデータが k-匿名性を持つよう

表 2　パーソナルデータの匿名化例

属性				
識別子	準識別子		センシティブ属性	
ID	性別	年齢	病歴	年収
-	男	40 代	心疾患	750
-	男	40 代	ガン	640
-	女	40 代	ガン	650
-	女	40 代	心疾患	550

に加工することを k-匿名化という [3, 4]．k-匿名化はパーソナルデータを匿名加工情報に加工する手段の一つであり，加工されたデータに個人識別性がないことの根拠となり得る．

k-匿名化を実現するための処理として，属性値を削除する抑制，属性値を抽象化する一般化，属性値に一定範囲内の乱数を加えるノイズ付加などがある．k-匿名化では，識別子を削除し，準識別子を一般化などし，センシティブ属性がどの個人のデータであるかを特定できないようにする．

表 1 を 2-匿名化した例を表 2 に示す．

表 2 では識別子である ID の属性値を削除して抑制し，準識別子である年齢の属性値を一般化で抽象化した．これにより，準識別子が同値である“男，40 代”，“女，40 代”はいずれも 2 レコードあり，またセンシティブ属性の属性値を知ったとしても，どのレコードが目的の個人かを識別することができない．

このような属性値の一般化は，抽象度を低くすればプライバシ問題を引き起こしやすく，逆に抽象度を高くすれば情報自体の価値が失われるため，トレードオフの関係になっている．また，抽象度をどの程度まで高めれば十分にプライバシ保護されているという一般的な判断はできず，対象とする属性や同時に提供される属性，同時に提供される他者の情報によって変わってくる．そのため，属性値を抽象化して一般化したとしても，直ちに匿名加工情報とはならない．よって，匿名加工情報ではない限り，一般化された属性であっても，それらの取得にあたっては本人同意が必要となる．

3. 同意の取得

パーソナルデータの取得や提供についての同意は，同意しない人が申告するオプトアウト方式が広く用いられてきた．この場合，利用者によっては「知らなかった」「だまされた」のように感じることもあり，批判的な意見が根強くあった．近年においては，EU の一般データ保護規則（General Data Protection Regulation; GDPR）施行などに伴い，利用者のプライバシ意識がますます高まり，同意する人が申告するオプトイン方式が増え始め，利用者と事業者間での意識の相違が生じにくい状況になりつつある．

3.1 事業者の取り組み

NTT ドコモはパーソナルデータの関する同意内容を確認したり，変更したりできるパーソナルデータダッシュボードの提供する[*2] としている．また，ヤフーは利用データを基に算出した数値で各種サービスを優遇する信用スコアについて，当初の初期設定はオンであったものをオフとする方針を示し，利用者が個別に同意した場合のみ，信用スコアを作成する[*3] こととした．また，両社ともに，利用者がよりわかりやすいように，プライバシーポリシーを改定している．

これらはいずれも「個人の情報が勝手に第三者に提供されているのではないか」という疑念を払拭する取り組みであり，利用者に丁寧に説明し，確かな同意を得なければ，パーソナルデータを利活用したサービスは立ち行かないと判断したものと見られる．

OECD プライバシーガイドラインの個人情報取扱に関する基本原則 [1] においても，収集制限の原則（適正な取得）が示されており，近年のプライバシ意識の高まりにともって，基本原則に従った対応が増えてきた．

3.2 スマートフォンでの取り組み

例えば，スマートフォンの位置情報を利用するアプリケーションが，位置情報取得のパーミッションを要求する場合，利用者がアプリケーションに対して位置情報の精度（正確な位置情報かあるいはおおよその位置情報）を選択することはできない．また，アプリケーションが実行中ではないバックグラウンド状態でも位置情報取得を許可した場合，どのような頻度で位置情報が取得されているかはわからない．

この問題に対し，Apple の iOS13 では位置情報の取得について変更が加えられ，従来は“常に許可”，“App の使用中のみ許可”，“許可しない”だったのが，“App の使用中は許可”，“1 度だけ許可”，“許可しない”となった．また，アプリケーションがバックグラウンドで位置情報を取得しているとき，どのような位置情報が取得されたかを地図上に表示して可視化し，定期的に位置情報の取得許可を利用者に確認するようになった．

図 3 に Android におけるアプリの権限画面を示した．スマートフォンは生活に密着したデバイスであり，利用者のパーソナルデータが豊富に含まれている．スマートフォンアプリはスマートフォンの OS が提供する API を通して，スマートフォン内のデータやセンサにアクセスする．そのために様々な権限管理が行われており，図 3 にようにどのアプリケーションがどのような権限を利用しているかをすぐに確認でき，必要に応じてその許可を容易に取り消す

[*2] https://xtrend.nikkei.com/atcl/contents/watch/00013/00597/
[*3] https://tech.nikkeibp.co.jp/atcl/nxt/news/18/05924/

図 3 Android10 におけるアプリの権限

ことができるようになっている．見た目や操作性は異なるが，基本的には iOS でも同様である．

このように，プライバシ問題に対しては，どのような情報が誰にどのように使われているかを気づかせることが必要で，そのための一つとして可視化されることが増えてきた．

これによって，パーソナルデータ取得の適正化，権限取得の細分化などが進んでいるが，利用者の簡便さも考慮して，データやセンサにアクセスできるか否かというアクセス制御が主であり，アクセス許可した内容に対する制限や加工はできない．

4. 議論

利用者からパーソナルデータの取得同意を得るためには，利用目的の特定が原則だが，利用目的を特定しても利用者の同意が得られない場合はある．そのような場合，利用者はあらゆるパーソナルデータを提供したくない，利用者が利用目的に納得していないなどが考えられる．前者から同意を得るのは困難であるが，後者は改善の余地がある．

利用者はパーソナルデータを提供するときに，その対価を期待することがある．従来から行われている方法の一つにポイントサービスがある．一般的にはポイントカードが発行され，そのポイントカードの ID で利用者を識別する．ポイントカードをレジなどで提示することで，ポイントカードの ID と購買履歴などを紐付ける．事業者は購買履歴と利用者の属性を組み合わせてマーケティング等に活用し，その対価としてポイントカードへポイントの形で金銭的な還元を行う．利用者が提供しているデータに対する対価として適当であると納得しているかどうかは議論の余地がある．2016 年に NTT データ経営研究所が実施した「パーソナルデータに関する一般消費者の意識調査」[*4] によれば，次のような結果が報告されている．

- 企業が消費者のパーソナルデータを収集し，マーケティング等に利用していることについて，70%の回答者が「知っており不快である」または「知らなかったので不快である」と回答している
- 趣味・趣向，年齢・生年月日については，74%の消費者が，金銭やポイント等の対価を得る条件ならパーソナルデータを提供しても良いと回答している
- パーソナルデータを提供する対価として，「500 円以上 1000 円以下」[*5]の回答割合が最も高かった

パーソナルデータ提供の対価として，金銭やポイント等以外にも，利用者個々人に適応したサービスである，いわゆるパーソナライズドサービスも挙げられる．ここでは具体例として天気予報サービスを考える．天気予報サービスは利用者の位置情報を取得し，その位置情報に基づいて，その地域の天気予報を表示するサービスである．利用者の位置情報に基づくので，そのサービスにアクセスした時点でいた場所の天気予報が表示されるため，自らが自分のいる地域を選ぶ必要がなく，利便性が高い．しかし，ある利用者は「今いる位置情報に基づいて雨が降るかどうかを知りたい」と思うかもしれない．またある利用者は「日頃の移動履歴から通勤・帰宅の過程で傘が必要かを知りたい」と思うかもしれない．このような利用者にとっては，位置情報を天気予報サービスに提供していても，自らが期待するサービスが受けられていないため，パーソナルデータの提供対価として妥当ではないと考える．結果的には，利用者はそのサービスの利用をやめてしまったり，位置情報取得の同意を撤回したりすることとなる．

このように，同意取得の場合は，利用目的の特定に加えて，利用者がどのようなメリットを享受できるかを明確にすることも重要である．例に挙げた天気予報サービスであれば，次のような改善によって，利用者の同意が得やすくなる可能性がある．

サービス利用時のみ位置情報取得を許可 天気予報サービ

[*4] https://www.nttdata-strategy.com/aboutus/newsrelease/161122/
[*5] ただし選択肢の中でこの項目が最も高額

スにアクセスした地点の天気予報を表示します．

一定時間毎に位置情報取得を許可　天気予報サービスにアクセスした地点の天気予報を表示します．通勤時間帯，帰宅時間帯にいる地点の天気予報に基づいて降雨予報を提供します．

位置情報取得を拒否　全国の天気予報が表示します．

“一定時間毎に位置情報取得を許可”は，提供されるサービスにメリットを感じて同意する利用者，自宅や職場を知られたくないと感じて同意しない利用者などに判断が分かれると思われる．しかしながら，それがパーソナライズドサービスであり，利用者自らがパーソナルデータの提供対価として，納得するかどうかである．

また別の視点から，位置情報などの観測データから分析される推定データの問題がある．ある利用者は自宅や職場の位置を知られたくないので，その情報を提供していないと仮定する．一方で，位置情報を活用したパーソナライズドサービスは便利に感じており，位置情報の定期的な提供を行っていたとする．定期的な位置情報の取得から，平日や休日の移動履歴などから自宅と職場の位置は比較的容易に推定される．先に挙げた天気予報サービスの場合，必要なのは自宅や職場の具体的な場所ではなく，大まかなエリアである．そのため，実際には天気予報サービスは，精度の高い GPS による位置情報を取得しながらも，そこから導かれるエリア情報だけを利用しているかもしれない．しかし，利用者にとっては，精度の高い位置情報を提供しており，そこに感覚のギャップが生じている．

このような例であれば，利用者のスマートフォンで精度の高い GPS による位置情報を取得したとしても，天気予報サービスに提供する際に，スマートフォン側で抽象化し，エリア情報だけが提供されればよい．つまり，事業者が利用者の属性値を取得して必要な粒度に抽象化するのではなく，利用者自身が自らが納得するように属性値を抽象化してから事業者に提供する仕組みが必要である．

GDPR では，個人が自らのパーソナルデータのコントロール権（自己情報コントロール権）を持つべきであるとしている．このような自己情報コントロール権を容易に行使できるようになることで，適切なパーソナルデータの利活用とパーソナライズドサービスの実現は達成される．

これらに関係するプラットフォームとして情報銀行がある．個人との契約に基づいて，その個人のデータを管理するのが情報銀行であり，情報銀行の利用者は自らのデータを提供する条件を指示し，情報銀行はその指示に従って第三者である事業者にデータを提供するかどうかを判断する．情報銀行でも同様に，利用者が自らのデータを第三者に提供したとして，その対価を享受できない，あるいは実感できなければ提供を取り下げると考えられ，データ提供とその対価としてのサービス（金銭等も含む）のバランス感と納得感が重要になる．これに関連して，経済産業省は情報信託機能の認定に係る指針 ver1.0 [*6] を公表している．情報銀行の取り組みよって，パーソナルデータの管理，利用などが進むことが期待されている．

参考文献

[1] OECD: The OECD Privacy Framework, https://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf (2013). accessed Oct. 20, 2019.
[2] Sweeney, L.: Simple Demographics Often Identify People Uniquely, https://dataprivacylab.org/projects/identifiability/paper1.pdf (2000). accessed Oct. 20, 2019.
[3] Sweeney, L.: Achieving K-anonymity Privacy Protection Using Generalization and Suppression, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 10, No. 5, pp. 571–588 (online), DOI: 10.1142/S021848850200165X (2002).
[4] Sweeney, L.: k-Anonymity: A Model for Protecting Privacy, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 10, No. 5, pp. 557–570 (online), DOI: 10.1142/S0218488502001648 (2002).
[5] World Economic Forum: Personal Data: The Emergence of a New Asset Class, http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf (2011). accessed Oct. 20, 2019.
[6] 村本俊祐，上土井陽子，若林真一：データを極小歪曲し k-匿名性を保持したデータに変換するプライバシー保護アルゴリズム，日本データベース学会 Letters，Vol. 6, No. 1, pp. 97–100 (2007).

[*6] https://www.meti.go.jp/press/2018/06/20180626002/20180626002-2.pdf