

事業者が匿名加工情報の具体的な作成方法を
検討するにあたっての参考資料
（「匿名加工情報作成マニュアル」） Ver1.0

平成28年8月
経済産業省

目 次

1. はじめに.....	1
2. 用語の定義.....	2
3. 加工方法に係る検討の手順.....	3
1) ユースケースの明確化.....	3
2) 識別子、属性、履歴の仕分け.....	5
3) 個人識別に係るリスクの抽出.....	7
4) 個人識別に係るリスクを踏まえた加工方法の検討.....	9
4. 匿名加工情報の加工の事例.....	15
4.1 電力利用データの事例.....	15
4.2 購買データの事例.....	23
4.3 移動データの事例.....	28

参考資料

匿名加工方法に係る用語説明.....	43
--------------------	----

1. はじめに

個人情報の保護を図りつつ、近年の飛躍的な情報通信技術の進展に対応したパーソナルデータの円滑な利活用を促進することにより、新産業・新サービスの創出等による活力ある経済社会の実現等を目的に、平成 27 年 9 月、個人情報保護法が改正された。

この改正では、特定の個人を識別することができないように加工された「匿名加工情報」を新たに定義し、本人の同意に代わる一定の条件の下、ビッグデータをはじめとするパーソナルデータの自由な利活用を認めることにより、新産業・新サービスが創出できる環境を整えることとされた。

匿名加工情報を作成する際の加工の程度は、個人データを取り扱う事業の内容や利用形態等によって判断されるべきものであり、一律の基準は存在しない。個人情報保護法においても、認定個人情報保護団体の指針等が定められ、実用に供されることが想定されている。

この「匿名加工情報作成マニュアル」は、具体的なユースケースを用いて、匿名加工情報を作成するための具体的な手順や方法について、学界、産業界、消費者団体等における有識者の方々からの意見を踏まえて、事業者の検討の参考にすべく作成したものである。また、これが最終版ということではなく、ユースケースの蓄積等に応じ、適宜改訂されていくべきものである。なお、本マニュアルにおいて、匿名加工情報取扱事業者に課される規律（匿名加工情報に係る公表、識別行為の禁止、安全管理措置 等）についての言及は、必要最小限にとどめており、個人情報保護法の法文の解釈を示すものではない。

今後、業界団体、企業、認定個人情報保護団体等が、匿名加工情報や匿名加工情報に係るガイドライン等を作成するにあたり、本マニュアルが有効に活用され、「匿名加工情報」による新産業・新サービスの創出に繋がっていくことを期待する。

2. 用語の定義

識別子

個人データを構成する情報であって、単体で個人を特定する可能性のある情報。

例：氏名、会員番号 等¹

注）個人情報に該当しないデータであっても、識別子に該当する場合（例：アカウント ID、端末 ID 等）があり、個人識別リスクを踏まえて判断する。

属性

個人データを構成する情報であって、経時的にデータが積み重ねられることのない情報で、単体では個人を特定することができないものの、他の属性との組み合わせや外部の情報との照合によって、個人を特定する可能性のある情報。

例：性別、年齢、郵便番号、家族構成 等

履歴

個人データを構成する情報であって、個人の行動の履歴を蓄積し、経時的にデータが積み重ねられる情報で、一般に単体では個人を特定することができないものの²、他の属性との組み合わせや外部の情報との照合によって個人を特定する可能性のある情報。

例：ウェブの閲覧履歴、購買の履歴 等

仮 ID

匿名加工情報の作成に当たり復元することのできる規則性を有しない方法により ID 等の識別子を他の記述等に置き換えた情報で、履歴と属性とを結びつけるものである。

図表 1 識別子、属性、履歴の関係

識別子	属性	履歴
アカウントID 氏名	性別 年齢	購買品1 購買品2 ...

¹ 住所は、市町村、地番及び住居番号等の全てが揃うと、個人を特定する可能性のある情報となる。

² ただし、音声対話履歴のように、「氏名・住所を順に聞き出す」内容を含むデータは、履歴であっても、単体で個人を特定する可能性がある。

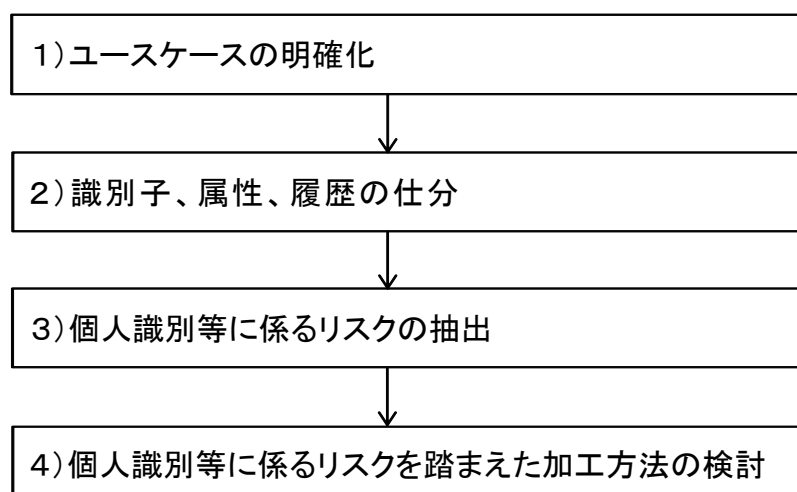
3. 加工方法に係る検討の手順

匿名加工情報を作成するための加工方法は、1) ユースケースの明確化、2) 識別子、属性、履歴の仕分け、3) 個人識別等に係るリスクの抽出、4) 個人識別等に係るリスクを踏まえた加工方法の検討 の4つの検討プロセスから構成される。

なお、匿名加工の要求レベルは、個人データを取り扱う事業の内容や利用形態等によって判断されるべきもので、普遍的な基準は存在しない。そのため、ケースバイケースで匿名加工の要求レベルについて検討する必要がある。

以下では、各検討プロセスの詳細について解説する。

図表 2 加工方法に係る検討プロセスの全体の流れ



1) ユースケースの明確化

まず、加工方法に係る最初の検討プロセスとして、ユースケースを明確化する。

ユースケースの明確化に当たっては、次の事項を確認し、図表 3 のとおりユースケースの概要を整理する。

1. 匿名加工情報の作成者における業務・サービスの概要
2. 匿名加工情報の作成に用いる個人情報データベース等のデータの項目、規模等
 - 個人情報データベース等によって識別される個人の人数
 - データの項目の内容（カテゴリなどの離散値、年齢などの整数値、支出などの連続値、燃費などの実数値など別、及びデータの項目の内容及び取り得る値の集合、最小値、最大値など）

3. 匿名加工情報に含める必要のあるデータの項目、規模（データ件数）³
4. 匿名加工情報の利用目的⁴

図表 3 ユースケースの概要

業務・サービス名	顧客属性(性・年齢・年収)に応じた購買データの販				
業務サービスの概要	保有する顧客の購買データを性・年齢・年収などの属性情報で分析を実施				
特定される個人の総数	●万人				
個人情報	データ種別	データ件数	データ項目	項目	備考
	顧客属性データ	●万件	契約者ID	11桁数値	
			性別	男女	分布
			生年月日	yyyymmdd	分布
			電話番号	10/11桁数値	分布
			勤務先	カテゴリ	分布
			年収	値	最大/最小/平均
			決済金融機関	名称	分布
			住所	名称	分布
	利用明細データ	●千万件	契約者ID	11桁数値	
			利用日	yyyymmdd	
			利用加盟店名	名称	
			購入商品	名称	
			支払金額	値	最大/最小/平均
匿名加工情報の利用目的	マーケティング用統計の作成				

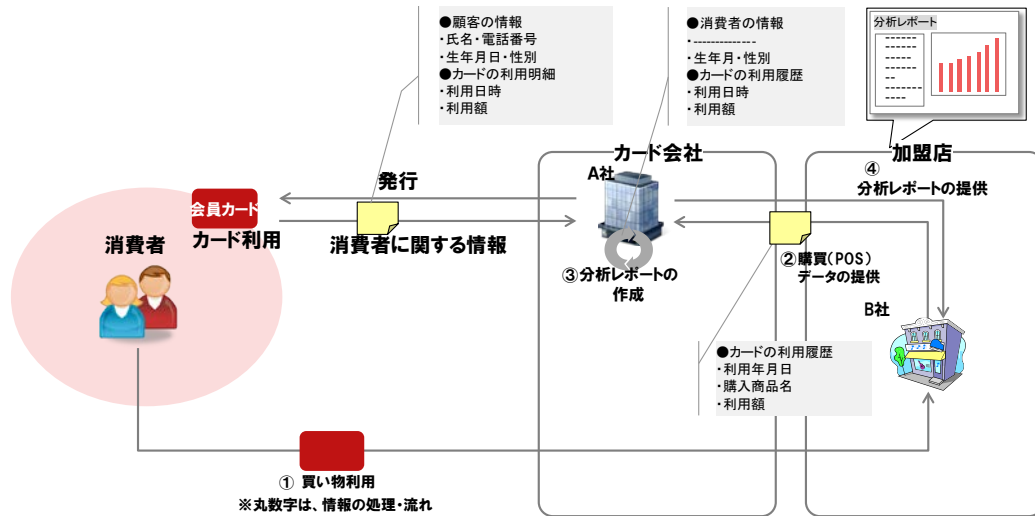
次に、ユースケースを具体的にイメージすることができるよう、以下について整理するのが望ましい。

- ・ ユースケースの全体イメージ
個人データの収集 → 個人データの加工による匿名加工情報の作成 → 匿名加工情報の利用・提供 という一連のデータの流れがわかるイメージ図
- ・ 取り扱う個人情報データベース等のデータサンプル

³ 利用目的等に照らして必要なデータの項目を絞っておくことでその後の加工方法の検討を効率的に行い得る。

⁴ あらかじめ利用目的を特定せずに複数の目的に対応できるように加工することも制限されているわけではない。

図表 4 ユースケースの全体イメージ



図表 5 匿名加工情報の作成に用いる個人情報データベース等のサンプルイメージ

①顧客属性データ												
契約者ID	氏名	性別	生年月日	電話番号	勤務先	年収	決済金融機関	住所				
								郵便番号	都道府県	市区町村	町名・丁目・番地	アパート・マンション・ビル
53012602	野村太郎	男性	1987年10月12日	03-222-5555	AA商事	450万円	みずほ	100-0005	東京都	千代田区	丸の内1-6-5	丸ノ内北ロビル
53597201	山田花子	女性	1990年5月23日	090-444-7777	BB保険	270万円	三井住友	116-0002	東京都	荒川区	荒川2-3-6	ハイツ荒川
81567824	佐藤一郎	男性	1968年8月19日	03-123-4567	C区役所	530万円	スルガ	111-1111	東京都	港区	六本木1-2-3	ビバリー六本木

②利用明細データ				
契約者ID	利用日	利用加盟店名	購入商品	支払い金額
53012602	2015年10月15日	丸の内店		12,000
53012602	2015年10月16日	荒川店		50,000
53012602	2015年10月17日	六本木店		250,000

2) 識別子、属性、履歴の仕分け

前プロセスで整理した取り扱う個人情報データベース等のサンプルデータにおいて、ケースに応じて必要なデータを抽出し、識別子、属性、履歴のいずれに該当するものの仕分けを行う⁵。

⁵ なお「識別子」「属性」「履歴」の区分は、適切に匿名加工を行うための便宜的なものであり、加工を行うデータの内容や構造、得ようとする結果との関係で仕分けを行う必要がある。そのため、データの項目名から一意に仕分けができるものではなく、適切な加工ができない、又は結果が得られない場合には仕分けを見直す必要が生じることも想定される。

仕分け作業は、識別子→属性→履歴の順に、次の各事項に留意して行う。

識別子

- データの項目のうち、単体で個人を特定する可能性のある情報を抽出する。
- アカウント ID や端末 ID など、個人情報に該当しないデータであっても、単体で個人を特定する可能性のある情報は、識別子として仕分けする。
- 住所は、詳細（番地や住居番号）まで含める場合は、個人を特定する可能性があるため、識別子として仕分けする。

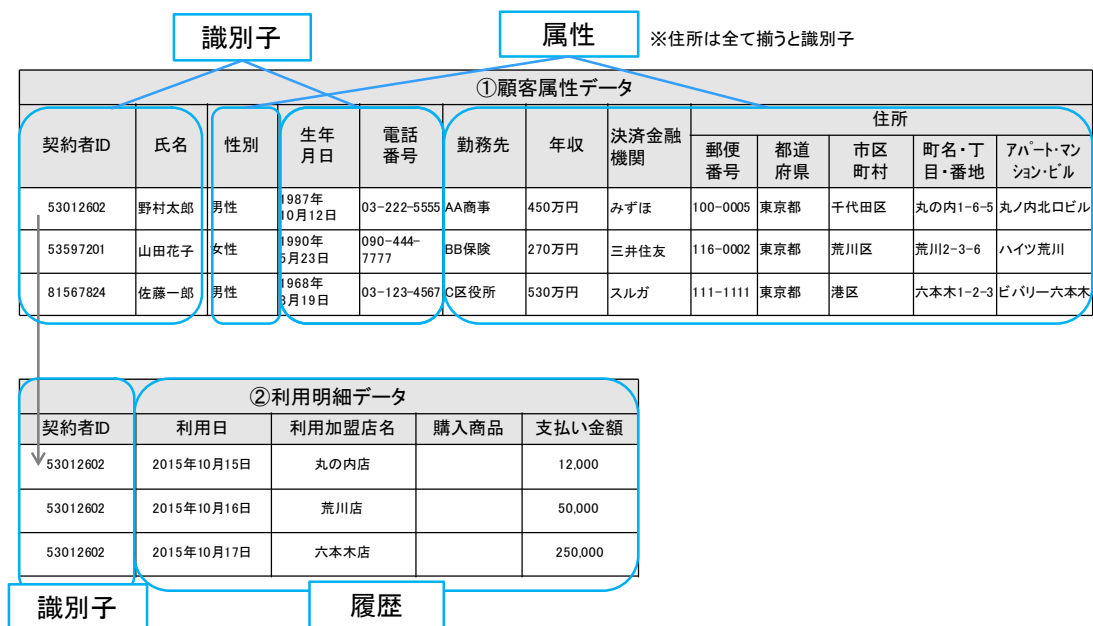
属性

- 識別子以外のデータの項目のうち、経時的にデータが積み重ねられることのない情報で、他の属性との組み合わせや外部の情報との照合によって、個人を特定する可能性のある情報を抽出する。
- 住所は、詳細（番地や住居番号）まで含めずに、一定の曖昧さを含むレベルまで利用する場合（例：都道府県と市町村まで）は、属性として仕分けする。

履歴

- データの項目のうち、ウェブの閲覧履歴、購買の履歴等、個人の行動の履歴を蓄積したものをはじめとする経時的にデータが積み重ねられる情報で、識別子等により属性と結びつけられている情報。一般に属性と組み合わせられ、若しくは外部情報との照合がなされない限り個人を特定する可能性のない情報が該当する。

図表 6 識別子、属性、履歴の仕分け作業のイメージ



3) 個人識別に係るリスクの抽出

前プロセスまでのユースケースの明確化や、識別子、属性、履歴の仕分け結果を踏まえ、個人識別に係るリスクを抽出する。

個人識別に係るリスクとしては、以下の①個人が特定されるリスクを挙げて評価することになるが、ユースケースに応じて以下の②データが他の情報と照合されるリスク及び③データを用いて本人へアプローチされるリスクについても評価することが望ましい場合もある⁶。②及び③も①に含めて評価し得る場合もあるが、ここではいずれも①を具体化したものとして別途評価対象として挙げた。

①個人が特定されるリスク

②データが他の情報と照合されるリスク⁷

③データを用いて本人へアプローチされるリスク⁸

個人識別に係るリスクの抽出作業は、識別子→属性→履歴の順に、次の各事項に留意して行う。

識別子

- ・ 個人識別のリスクが高いものである。
- ・ 原則として削除等を行うべきものであり（次プロセス参照）、氏名等、評価をするまでもなく削除すべきものもある。

図表 7 リスクが認められる場合（識別子）

①個人が特定されるリスク	識別子が削除等されていない場合
②データが他の情報と照合されるリスク	-
③データを用いて本人へアプローチされるリスク	住所等（メールアドレスを含む。）の識別子が削除等されていない場合

属性

- ・ 一般に、複数の種類の属性が組み合わせられると、個人識別のリスクは高くなる。具体的なユースケースを踏まえ、どの属性とどの属性が組み合わせられると個人識別のリスクが

⁶ 「個人の属性が推定されるリスク」については、要配慮個人情報や一般に他人に知られたくないような情報を推定する場合に限定することを想定し、対象とすべきという意見もある。4. 匿名加工情報の加工の事例のうち、電力利用データ及び購買データの事例では本リスクを対象としたが、移動データの事例では対象外とした。

⁷ 本リスクに対しては個人情報保護法第38条で手当されているとして、加工によりリスク処理する必要性に乏しいという意見もある。

⁸ 個人識別符号の該当性の判断要素の一つとして本人到達性が挙げられている（瓜生和久編著「一問一答 平成27年改正個人情報保護法」（商事法務、平成27年）14頁）。

高くなるものかについて評価する。

- ・ データセットに対して、属性の値についても評価し、極めて頻度の小さい値がある場合は、個人識別のリスクが高まる可能性があるものとして抽出する。
- ・ 他の情報と照合されるリスクは、市販のデータベースなど、データセットと照合可能な外部情報が存在するかを確認する。例えば、住宅地図などの外部情報を用いることによって、住所から切り出した属性と照合される可能性について評価することなどが該当する。さらに匿名加工情報の提供先となる事業者を限定する場合においては、当該事業者が照合可能なデータを有しているかどうかについて評価する⁹。
- ・ ダイレクトメールや訪問販売等によって本人へアプローチされるリスクは、主に位置に関する情報が該当し、自宅や職場の地域が属性を組み合わせることで絞り込まれる場合、リスクとして認識する必要がある。

図表 8 リスクが認められる場合（属性）

①個人が特定されるリスク	・ 複数の種類の属性が組み合わせられる場合 ・ 極めて頻度の小さい値がある場合
②データが他の情報と照合されるリスク	・ 照合可能な外部情報が存在する場合 ・ 提供先事業者が照合可能なデータを有している場合
③データを用いて本人へアプローチされるリスク	・ 自宅や職場等の本人が所在する確率の高い場所が、属性を組み合わせることで絞り込まれる場合

履歴

- ・ 特異な値や傾向を持つ履歴は、個人識別性が高いものとして特異値に係るリスクとして認識する。値や傾向が特異であるかは、利用目的やデータの性質、データセットにおける位置付けや割合等を斟酌して個別具体的に判断せざるを得ないが、例えば、非常に高額な商品や希少な商品を購入した場合の購入履歴などが該当する。
- ・ 一個人に対する長期間のデータ蓄積、又は大量のデータ蓄積によって、個人の識別性が生じる可能性がある場合は、データ規模、期間に係るリスクとして認識する。
- ・ 履歴を継続的に提供する場合、履歴を照合することで、突合できる場合も考えられるため、履歴の継続的提供を想定するユースケースの場合は、リスクとして認識する。
- ・ 位置情報に関する情報が個人を識別する可能性のある粒度の場合、個人が識別されるリスクとして認識する。例えば、詳細な GPS 情報の履歴を蓄積すると、自宅や職場の場所が識別されるリスクがある。また、位置に関する情報を含むことによって、本人にアプローチされるリスクが生じる場合がある。

⁹ 提供先となる事業者が照合可能なデータを有しているかどうかまでは、評価が困難であるという意見もある。

図表 9 リスクが認められる場合（履歴）

①個人が特定されるリスク	・ 特異な値や傾向を有する場合 ・ 一個人に対する長期間又は大量のデータ蓄積がある場合 ・ 位置に関する情報が、個人識別性のある精度の場合 ・ 履歴を継続的に提供する場合
②データが他の情報と照合されるリスク	・ 履歴を継続的に提供する場合
③データを用いて本人へアプローチされるリスク	・ 位置に関する情報を含む場合

4) 個人識別に係るリスクを踏まえた加工方法の検討

個人識別に係るリスクを踏まえ、リスク要因を除去するための適切な加工方法を検討する。

前プロセスで抽出した個人識別に係るリスクについて、識別子→属性→履歴の順に、次の各事項に留意して適切な対処を検討する。

識別子

識別子については、図表 10 に記載のとおり、それ自体を個人データから削除又は復元することのできる規則性を有しない方法で他の記述等に置き換える（仮名化）。それだけでは個人識別に係るリスクを十分に低減できない場合は、識別子に紐付く属性や履歴の加工処理も必要となる。

図表 10 個人識別に係るリスクを踏まえた識別子の加工方法の例

個人識別に係るリスク	加工方法
①個人が特定されるリスク ③データを用いて本人へアプローチされるリスク	・ 原則として、識別子¹⁰を削除又は復元することのできる規則性を有しない方法により他の記述等に置き換える（仮名化）。 ・ 属性に仕分けできるレベルに落とすために識別子を一部削除等する。 ・ 履歴と照合するための仮 ID を例外的に用いる場合、仮 ID の要件としては以下が考えられる。

¹⁰ メールアドレスは、所属組織や氏名を推知させるようなものではない場合であっても、識別子として削除等の対象とする。

	また、仮 ID を継続して用いる場合は、その生成方法、有効期間の両方に配慮が必要である。 ○仮 ID の要件 ・ 不可逆であること。 ・ 一定の期間を超えて、同一の個人に対する同一の仮 ID を使用しないこと。 ・ 異なる匿名加工情報を作成するときは、同一の個人に対し同一の仮 ID を使用しないこと。 ○仮 ID の生成方法 ・ 秘密の文字列を加えてハッシュ化¹¹する方法が有効である¹²。 ・ 対応表を十分に安全管理して一意な乱数を割当てることが有効である。 ○仮 ID の有効期間 ・ 仮 ID の有効期間は、属性、履歴も含めて考慮した個人識別性に係るリスクを踏まえ、適切な期間に設定すること（仮名制御）が不可欠である。 ・ 識別子の有効期間終了に伴い当該識別子を更新する場合や新たに匿名加工情報を作成する場合、同一の仮 ID を使用しないように配慮する。
--	---

属性

データの性質や利用目的等によっては識別子の加工だけでは個人識別に係るリスクを十分に低減できない場合もあり、そうした場合は、当該識別子に紐付く属性の加工も必要である。属性の加工方法の例は図表 11 に記載のとおりであるが、ケースに応じてこれらを適宜組み合わせることになる。4. 匿名加工情報の加工の事例も併せて参照されたい。

¹¹ ハッシュ化は、暗号化とは異なり、入力値から誰でも計算できる種類の処理である。不可逆的な（一方向の）変換ではあるが、識別可能な値を生成する。

¹² 仮 ID は、個人と必ずしも一対一対応ではなく、a)一人が複数の識別子に割り当てられる場合、b)複数人が一つの識別子に割り当てられる場合、などがありえる。

図表 11 個人識別に係るリスクを踏まえた属性の加工方法の例

個人識別に係るリスク	加工方法
①個人が特定されるリスク ・複数の種類の属性が組み合わされる場合 ・極めて頻度の小さい値がある場合	・どの属性とどの属性が組み合わせられると個人識別のリスクが高くなるかについて、組み合わせ数の算定等や、データセットに含まれる個人の数と属性数及び各属性の値の数により評価する。 ・データセットに対して、属性の値がどの程度の分布になっているのかについて評価する。 ・その上で、置換え、一般化、カテゴライズ化、特異値の削除、サンプリング等を行って、属性同士を組み合わせても、個人の識別リスクが低減できるレベルまで加工する。この際 k-匿名性を考慮する。 ○k-匿名化 ・データセットに含まれる属性の項目を勘案し、k-匿名性の指標を用いて個人識別に係るリスクを判定する。なお、k 値はデータの性質や規模、利用目的により判断する。 ・判定の結果として、k=1（一意に識別可能な属性）の値を有する属性の組み合わせであっても当該属性値の加工処理を必要としないこともあり得る。 ○一般化等 ・属性に、頻度の小さい値のあることが明らかな場合は、データの値域を広げる一般化、他の値と一体にしてカテゴライズ化を行う等の対応を行う。 ○サンプリング ・データセットから無作為にサンプリングを行うことで、個人識別に係るリスクを低減する。データセットにおける各属性の分布状況を踏まえ、サンプリング比率を調節する。サンプリングによって、k-匿名性が損

	なわれる可能性があるので、その適用順序に留意が必要である。
②データが他の情報と照合されるリスク ・照合可能な外部情報が存在する場合 ・提供先事業者が照合可能なデータを有している場合	・市販のデータベースなどの照合可能な外部情報の存在、提供先事業者が照合可能なデータを有しているかどうかについて評価する。例えば、住宅地図等の外部情報を用いることによって、住所から切り出した属性と照合される可能性について評価する。 ・上記外部情報等が存在する場合、容易に照合できないレベルまで 一般化 、カテゴライズ化を行う。
③データを用いて本人へアプローチされるリスク ・自宅や職場の地域が属性を組み合わせることで絞り込まれる場合	主に詳細な住所や要配慮情報にかかる場所等の位置に関する情報について、当該属性を削除あるいは 一般化 、 グルーピング 等を行って、リスクを低減する。
個人識別に係るリスク全般	上記の処理に加えて、履歴レコードを入れ替える（ スワッピング ）、値に ノイズ付加 を行うことで、さらに個人の識別リスクを低減することができる ¹³ 。

履歴

個人識別に係るリスクを十分に低減できない場合は、識別子や属性の加工に加えて、履歴の加工が必要となる場合もある。加工方法の例は図表 12 に記載のとおりであるが、これらを選択、組み合わせて加工を行う。

履歴については、一般的に識別子又は属性と組み合わせられない限り個人を特定する可能性のない情報が該当するが、履歴であっても必ずしも無加工で利用できるものではないことに十分留意する必要がある。

¹³ 他にも、摂動化（ランダムイズや乱数加算）といった手法もある。

図表 12 個人識別に係るリスクを踏まえた履歴の加工方法の例

個人識別に係るリスク	加工方法
①個人が特定されるリスク ・ 特異な値や傾向を有する場合 ・ 一個人に対する長期間又は大量¹⁴のデータ蓄積がある場合	・ 履歴の記録期間を明らかにし、その期間の長さによるリスクを評価する。 ・ 特異値や特異な傾向を持つ履歴の削除（トップ（ボトム）コーディングを含む）、置換え、一般化、カテゴライズ化等を行う。 ・ 長期間又は大量のデータであり、かつ、それにより個人が特定されるリスクが高い場合、以下の加工を行い、適宜組み合わせる。 (1) リスクに応じて、履歴の期間、蓄積回数の上限を定め、上限を超える場合は仮 ID を更新するなどし、特定の個人が識別できないように加工する（仮名制御）。 (2) 日付のランダム化（乱数でずらす）やシフト化（一様に一定の日数でずらす）、あるいは日付の間隔のランダム化（順序を保持し、日数の間隔を乱数で一般化する）（日付・間隔加工）。 (3) 日付等履歴の削除（トップ（ボトム）コーディングを含む）、置換え、一般化、カテゴライズ化等を行う。 (4) さらにリスク低減のため、属性の k-匿名性を確保した上で、仮 ID 等のスワッピングやノイズ付加を行う。
①個人が特定されるリスク ・ 履歴を継続的に提供する場合 ②データが他の情報と照合されるリスク ・ 履歴を継続的に提供する場合	・ 定期的に仮 ID を更新、付け替える。 ・ 履歴の継続的な提供にあたり、履歴の期間が重ならないように処理をする、同じ個人の履歴を連続したデータとして提供しない等によって、履歴の照合ができないように処理をする。

¹⁴ 一個人における購買履歴や取引履歴が相当程度蓄積された場合は、必ずしも長期間のデータでなくても個人が特定されるリスクは高まると考えられる。

① 個人が特定されるリスク ・位置に関する情報が、個人識別性のある精度の場合 ③データを用いて本人へアプローチされるリスク ・位置に関する情報を含む場合	・位置に関する情報の全部又は一部の削除（起点・終点を落とす等）、一般化（精度を粗くする等）、又はグルーピング等する。
--	--

4. 匿名加工情報の加工の事例

4.1 電力利用データの事例

1) 概要

以下では、電力利用データを事例として匿名加工情報の加工事例を示す。

本事例で取り扱う電力利用データとは、スマートメーター、ネットワーク接続された分電盤等を通じて取得された、家庭の電力使用量に係る履歴データである。

我が国においては、2024 年度までに家庭の全世帯でスマートメーターが設置されることが予定されており、電力会社やアグリゲーター、HEMS サービス等、スマートメーター等を通じて電力利用データを取得することが可能となる。

これらの電力利用データは、電力使用状況の見える化による需要家への情報提供、電力使用パターンを踏まえたより具体的な節電アドバイス、電力利用状況の変化を踏まえた共働き世帯の子どもや独居老人の見守り、生活様態推計を踏まえたエリアマーケティング等、多様な用途に活用される可能性がある。

一方で、特異な値や傾向から個人が識別されるリスクも持つため、匿名加工情報への加工に当たっては一定のリスクアセスメントが必要であると考えられる。

また、スマートメーターの普及や電力利用データのサービスへの活用は、緒に就いたばかりであり、IoT 機器の家庭内への普及や電力小売全面自由化と相まって、今後様々なサービスが開発されていくことが想定されるため、その目的や具体的な用途に応じて、匿名加工情報への加工のあり方も変わりうることに留意が必要である。

2) 電力利用データ活用に係るパーソナルデータのサンプルイメージ

本項目で扱う、電力利用データ活用に係る匿名加工情報の作成に用いるパーソナルデータのサンプルイメージは下記を想定する。

図表 13 電力利用データのサンプルイメージ

基本情報

契約者ID

氏名

電話番号

電子メールアドレス

性別

生年

職業

郵便番号

都道府県

市区町村

町名・丁目・番地

アパート・マンション・ビル

住居情報

住宅区分

住居形態

構造

竣工年

延床面積

家族の人数

家族構成(職業、生年(西暦)、性別)

A12345

野村太郎

03-222-5555

abcd@xmail.com

男性

1987

会社員

100-0005

東京都

千代田区

丸の内1-6-5

丸ノ内北ロビル

戸建

核家族

木造

3年

112㎡

4

妻(専業主婦、1986年、女性)
息子(幼稚園、2011年、男性)
娘(未就学、2014年、女性)

A67890

佐藤一郎

090-444-7777

xyz@fb.co.jp

女性

1990

公務員

116-0002

東京都

荒川区

荒川2-3-6

ハイツ荒川

マンション

両親同居

RC

25年

85㎡

2

父(会社員、1960年、男性)
母(専業主婦、1962年、女性)

B12345

山田太郎

03-123-4567

efgh@yahoo.co.jp

男性

1968

無職

111-1111

東京都

港区

六本木1-2-3

ビバリー六本木

独居

RC

7年

300㎡

0

-

識別子

属性

※住所は全て揃うと識別子

電力

契約者ID

年月日・時分

電力使用量

A12345

2015年10月31日 00:00

〇〇Wh

A12345

2015年10月31日 00:01

〇〇Wh

A12345

2015年10月31日 00:02

〇〇Wh

A12345

2015年10月31日 00:03

〇〇Wh

A12345

2015年10月31日 00:04

〇〇Wh

家電

契約者ID

推定購入年月

推定購入商品

A12345

2015年10月

照明

A12345

2015年10月

エアコン

A12345

2015年10月

冷蔵庫

A12345

...

識別子

履歴

識別子

属性

なお、履歴たる電力利用データについては、その取得のあり方によって、瞬時値（取得時の瞬間的な電力利用量（kW））や 30 分値（スマートメーターで標準的に取得可能な 30 分ごとの累計電力使用量（kWh））など様々である。

3) 具体的な匿名加工情報の作成方法

以上を踏まえて、具体的なサービスを想定した匿名加工情報の作成方法について以下に示す。

下記はあくまで、仮想的に想定したサービスにおいて、匿名加工のあり方を検討したものであり、サービスの詳細内容によって、取得するデータの種類や加工のあり方は変わらうことに留意が必要である。

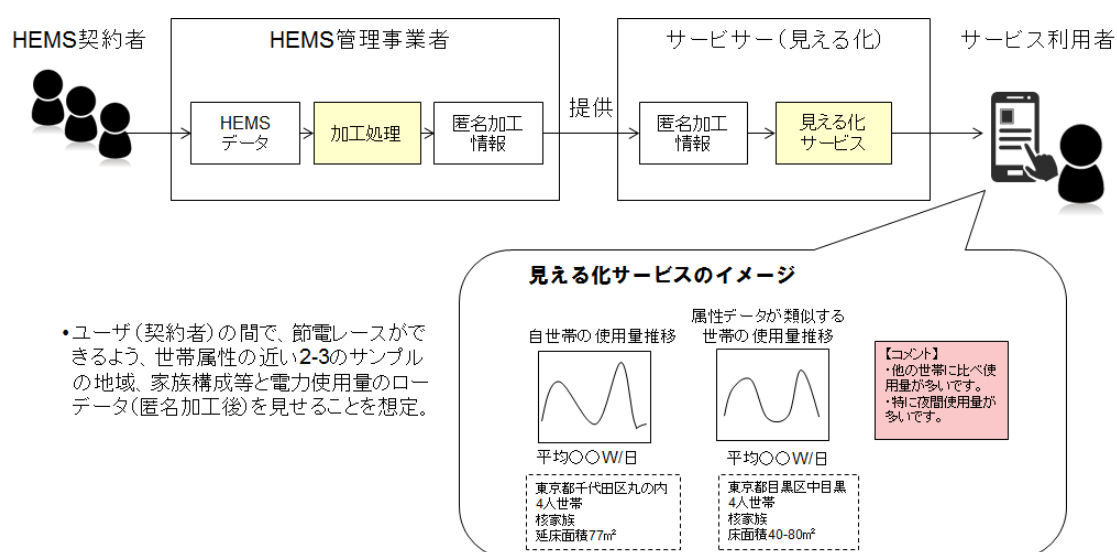
例えば、サービスの内容によって、属性の加工の粒度をより粗くする代わりに長期に亘る電力利用データを取得したり、電力利用データを時間単位、日単位に置き換える代わりに属性の加工の粒度を細かくしたりといった加工のあり方も考え得る。

i) 電力情報見える化サービスにおける電力利用データの活用

<サービス概要>

本サービスは、サービス利用者に対して、家族構成等の属性データが近い世帯の電力利用状況を比較して提示することで、サービス利用者の節電を喚起するものである。ランキング形式で提示することによって、サービス利用者同士が競争し、結果としてより大きな節電効果を上げることも期待される。

図表 14 電力見える化サービス



<想定されるリスク>

一般に、履歴たる電力利用データから、生活や家電の利用実態、在／不在情報等を判別することが可能。このため、個人、世帯が識別されるとプライバシーが脅かされるリスクはある。

また、属性ごとのリスクは下記の通りである。

図表 15 属性ごとのリスク

準識別子		リスクの特記事項
基本情報	性別	－
	生年	－
	職業	割合の小さい職業の場合、特定性が高くなる。
住所情報	郵便番号	7桁揃うと町丁目レベルの位置精度があり、本人へのアクセスリスクが比較的高い。
	都道府県	－
	市区町村	人口の小さい町村の場合、個人が特定されるリスクが比較的高い。
	町名・丁目・番地	本人へのアクセスリスクが比較的高い。
	アパート・マンション・ビル	本人へのアクセスリスクが比較的高い。
住居情報	住宅区分	－
	住居形態	－
	構造	構造、竣工年、延べ床面積は、住宅情報サイト等の汎用のデータベースを用いて、住居が特定されるリスクがある。
	竣工年	//
	延べ床面積	//
家族情報	家族の人数	大人数の場合、外部観測性があるため、世帯が特定されるリスクが高い。（外れ値）
	家族構成（職業、生年、性別）	外部から判別できるため、個人が特定されるリスクが比較的高い。
家電情報	推定購入年月	－
	推定購入商品	－

また、本サービスにおいては、サービスの性質上、サービス利用者が、自らと属性データが類似する世帯の電力使用量を見ることができる。このため、繰り返し同じ世帯が表示されると個人を識別されるリスクが高まる。

＜加工方法の例＞

以上を踏まえ、本サービスにおける加工方法としては、例えば以下が想定される。

図表 16 加工方法の例

情報分類	対象情報	加工方法
識別子	契約者 ID	- 仮 ID（匿名加工情報に付する識別子）に変換する（不可逆なものとする） - 例：秘密の文字列（アグリゲータのみが保有）でパディングしてハッシュをとる（秘密の文字列を持たないものには連結不可能にする）。秘密の文字列は 24 ヶ月毎にランダム生成する
	氏名、電話番号、メールアドレス	削除
属性	住所	都道府県、市区町村以外を削除
	職業	削除（見える化サービスでは必要性が薄いため）

情報分類	対象情報	加工方法
	家電情報	削除（見える化サービスでは必要性が薄いため）
	家族の人数	1、2、3、4人以上（4区分）
	家族構成	独居、夫婦のみ、親子、その他（4区分）
	竣工年	5年未満、5年以上10年未満、10年以上20年未満、20年以上30年未満、30年以上（5区分）
	延床面積	20㎡未満、20㎡以上40㎡未満、40㎡以上80㎡未満、80㎡以上（4区分）
	構造、住居形態、住居区分	加工無し
履歴		- 履歴情報については、下記の通り利用期限や特異値について配慮した上で利用する。 - 利用期限を、著しく個人識別性が上昇しない範囲で設定する。（例：過去24ヶ月） - 更に、特異値がある等の個人識別性が上昇する可能性があるデータについては、加工等により個人識別性を低減させることを考慮する。

また、本サービスにおいては、サービス利用者が自らと似た属性を持つ他の世帯の電力利用状況を見ることが可能であり、繰り返し同じ世帯が表示されると個人を識別されるリスクが高まることを踏まえ、顧客に提示される電力利用データが、必ず毎回又は適切な間隔で、複数の(k以上の)顧客からランダムに表示されることを保証する必要がある。

なお、この加工を行ったとしても属性の組み合わせにより個人が特定されるリスクがあるため、次のような加工処理等をすることが考えられる。

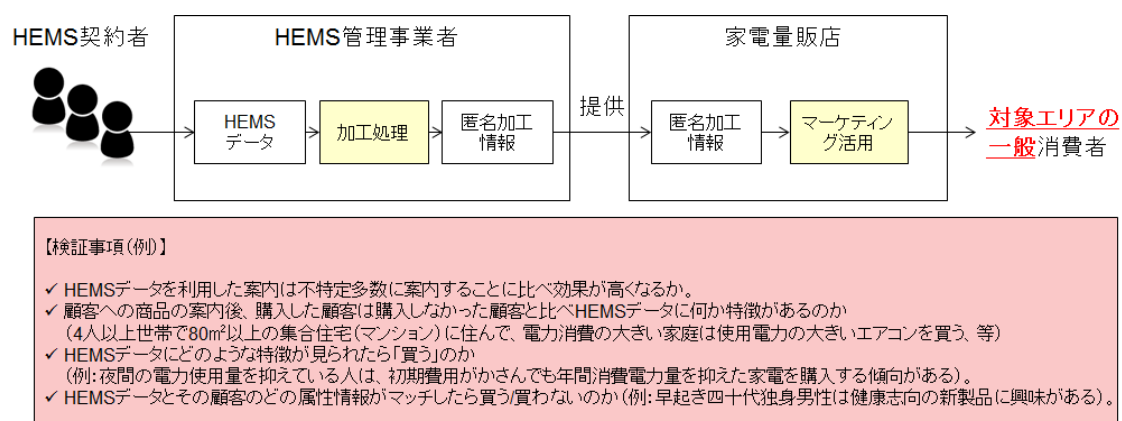
- HEMS普及率が著しく高くはなく、かつ外部からHEMS設置状況が判別できないという仮定を置くことができる際には、適切なサンプル抽出（ランダムサンプリング等）がなされていると見なして個人識別性を確認する。
- さらに曖昧化を行うなど、個人識別性の低減に寄与する妥当な加工処理をする。

ii) 電力利用データと家電購買の推計データを用いたエリアマーケティング

<サービス概要>

本サービスは、家電の販売者（家電量販店等）が、電力利用データと家電購買に係る推計データを踏まえて、特定の地域において、ニーズが高いと想定される家電を推計し、折込チラシやポスティングチラシを送付するサービスである。

図表 17 エリアマーケティングサービス



<想定されるリスク>

一般に、履歴たる電力利用データから、生活や家電の利用実態、在／不在情報等を判別することが可能。このため、個人、世帯が識別されるとプライバシーが脅かされるリスクはある。

また、属性ごとのリスクは下記の通りである。なお、本事例においては、家電に係る情報をユーザーから取得することとしており、法律上、匿名加工情報取扱事業者として、匿名加工情報からの個人の識別は禁止されているものの、サービス（家電量販店等）の持つデータと照合することで個人が特定されるリスクがあると言える。

図表 18 属性ごとのリスク

準識別子		リスクの特記事項
基本情報	性別	－
	生年	－
	職業	割合の小さい職業の場合、特定性が高くなる。
住所情報	郵便番号	7桁揃うと町丁目レベルの位置精度があり、本人へのアクセスリスクが比較的高い。
	都道府県	－
	市区町村	人口の小さい町村の場合、個人が特定されるリスクが比較的高い。
	町名・丁目・番地	本人へのアクセスリスクが比較的高い
	アパート・マンション・ビル	本人へのアクセスリスクが比較的高い。
住居情報	住宅区分	－
	住居形態	－
	構造	構造、竣工年、延べ床面積は、住宅情報サイト等の汎用のデータベースを用いて、住居が特定されるリスクがある。
	竣工年	//
	延べ床面積	//
家族情報	家族の人数	大人数の場合、外部観測性があるため、世帯が特定されるリスクが高い。（外れ値）
	家族構成（職業、生年、性別）	外部から判別できるため、個人が特定されるリスクが比較的高い。
家電情報	推定購入年月	（再識別は禁止されていても）サービスのデータと照合すると個人が特定されるリスクが比較的高い。
	推定購入商品	//

<加工方法の例>

以上を踏まえ、本サービスにおける加工方法としては、例えば以下が想定される。

図表 19 加工方法の例

情報分類	対象情報	加工方法
識別子	契約者 ID	仮 ID（匿名加工情報に付する識別子）に変換する（不可逆なものとする） 例：秘密の文字列（アグリゲータのみが保有）でパディングしてハッシュをとる（秘密の文字列を持たないものには連結不可能にする）。秘密の文字列は 12 ヶ月毎にランダム生成する
	氏名、電話番号、メールアドレス	削除
属性	住所	都道府県、市区町村以外を削除（※）
	職業	削除（マーケティングでは必要性が薄いため）
	家電情報	- 推定購入年月を削除 - 利用している家電情報は残す
	家族の人数	1、2、3、4 人以上（4 区分）

情報分類	対象情報	加工方法
	家族構成	独居、夫婦のみ、親子、その他（4 区分）
	竣工年	5 年未満、5 年以上 10 年未満、10 年以上 20 年未満、20 年以上 30 年未満、30 年以上（5 区分）
	延床面積	20 m²未満、20 m²以上 40 m²未満、40 m²以上 80 m²未満、80 m²以上（4 区分）
	構造、住居区分、住居形態	加工無し
履歴		- 履歴情報については、下記の通り利用期限や特異値について配慮した上で利用する。 - 利用期限を、著しく個人識別性が上昇しない範囲で設定する。（例：過去 12 ヶ月） - 更に、特異値がある等の個人識別性が上昇する可能性があるデータについては、加工等により個人識別性を低減させることを考慮する。

※ダイレクトメールの送付を考えると町名程度の住所の精度とすることが望ましい。

なお、本事例においては、仮想的に想定したサービス内容にとって必要な部分のみに限定するため、家電情報については推定購入年月を削除し、利用している家電が分かるのみの情報にとどめた。

また、この加工を行ったとしても組み合わせにより個人が特定されるリスクが残るため、次のような加工処理等をすることが考えられる。

- **HEMS** 普及率が著しく高くはなく、かつ外部から **HEMS** 設置状況が判別できないという仮定を置くことができる際には、適切なサンプル抽出（ランダムサンプリング等）がなされていると見なして個人識別性を確認する。
- さらに曖昧化を行うことなど、個人識別性の低減に寄与する妥当な加工処理をする。

4.2 購買データの事例

本事例では、購買データとして、クレジットカードの所有者情報である「顧客属性情報」、クレジットカードの利用に伴って蓄積される購買履歴である「カード利用明細データ」を例にしたケースを取り上げる。

カード利用明細データには、利用日や利用加盟店名、支払方法、利用金額が記録されているが、購入対象の商品・サービスの具体的な情報は含まれていない。このため、クレジットカードのデータのみでは、顧客の商品・サービスに対する嗜好やニーズを十分に把握することはできない。

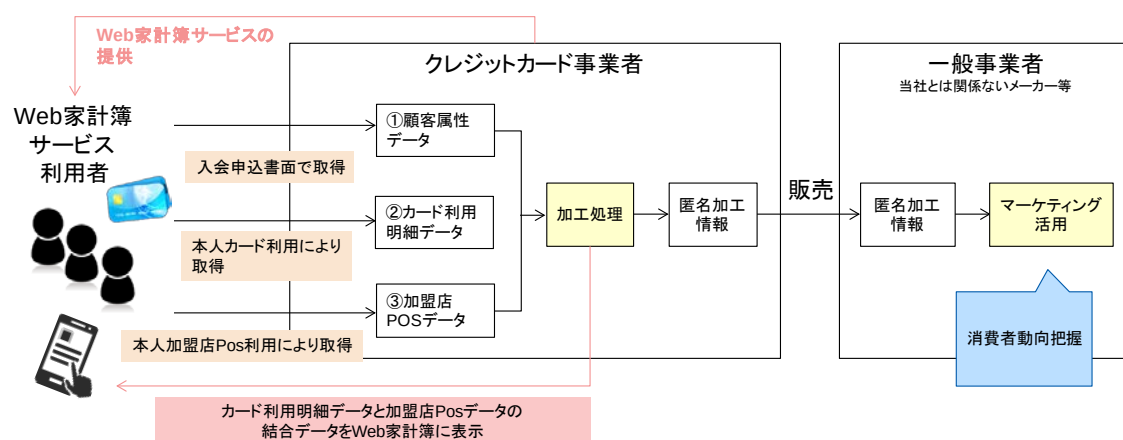
一方、加盟店の中には、POSを導入して商品・サービスの売れ行きを把握している者も多く、POSデータとクレジットカードのデータを組み合わせることで、顧客属性を踏まえた嗜好やニーズの分析が可能になると考えられる。以下では、クレジットカード事業者が、POSデータを取り込んだ上で、匿名加工情報として外部に提供するケースにおける匿名加工のあり方について示す。

1) ユースケースの明確化

クレジットカード会社が提供する「Web 家計簿サービス」では、サービス利用者がカードの自分の属性データや利用明細データを閲覧できることに加え、加盟店の POS データを取り込むことができ、Web 上で、簡便に家計簿をつけることができる。

本ケースは、この Web 家計簿サービスで管理する情報を、匿名加工情報の枠組みを活用して、一般事業者へ提供するというものである。

図表 20 ケースの全体イメージ



本ケースで用いるデータは、①顧客属性データ、②カード利用明細データ、③加盟店 POS データの三種類であり、いずれも契約者 ID（クレカ番号の変換番号）によって、リンクされている。なお②カード利用明細データと③加盟店 POS データとは、利用者本人がリンクを許可していることを前提とする。

またデータセットの規模は、100 万件以上（会員数 100 万人、1 ヶ月あたりの平均利用回数：10 回とするとデータ件数は 1,000 万超）とする。

図表 21 データのレイアウトサンプル

①顧客属性データ												
契約者ID (クレカ番号の 変換番号)	氏名	性別	生年 月日	電話 番号	勤務先	年収	決済金融 機関	住所				
								郵便 番号	都道府 県	市区 町村	町名・丁 目・番地	アパート・マン ション・ビル
53012602	野村太郎	男性	1987年 10月12日	03-222-5555	AA商事	450万円	みずほ	100-0005	東京都	千代田区	丸の内1-6-5	丸ノ内北口ビル
53597201	山田花子	女性	1990年 5月23日	090-444- 7777	BB保険	270万円	三井住友	116-0002	東京都	荒川区	荒川2-3-6	ハイツ荒川
81567824	佐藤一郎	男性	1968年 8月19日	03-123-4567	C区役所	530万円	スルガ	111-1111	東京都	港区	六本木1-2-3	ビバリー六本木

↑

システムのにもともとリンク ←—— 利用者本人が加盟店POSデータのリンクを許可 —— ↓

②カード利用明細データ				
契約者ID	利用日	利用加盟店名	支払方法	利用金額
53012602	2015年10月15日	丸の内店	1回	12,000
53012602	2015年10月16日	荒川店	2回	50,000
53012602	2015年10月17日	六本木店	5回	250,000

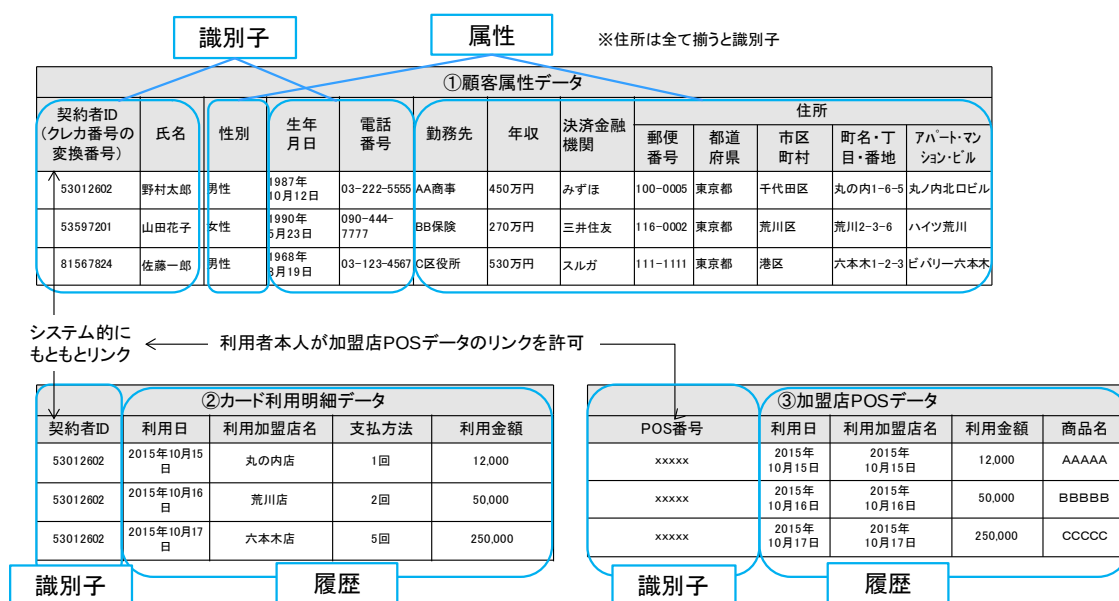
③加盟店POSデータ				
POS番号	利用日	利用加盟店名	利用金額	商品名
xxxxxx	2015年 10月15日	2015年 10月15日	12,000	AAAAA
xxxxxx	2015年 10月16日	2015年 10月16日	50,000	BBBBB
xxxxxx	2015年 10月17日	2015年 10月17日	250,000	CCCCC

2) 識別子、属性、履歴の仕分け

データのレイアウトサンプルを基に、識別子、属性、履歴を仕分けした結果を次図に示す。

住所は全て揃うと識別子となるが、ここでは分割して用いることを念頭に、属性として分類している。

図表 22 識別子、属性、履歴の仕分け結果



3) 個人識別に係るリスクの抽出

個人識別に係るリスクとして、以下の4つの観点から抽出した。データの利用や第三者提供による「不安感（気持ち悪さ）」は個人差があるため、評価の対象外としている。

（評価対象とする個人識別に係るリスク）

- 個人が特定されるリスク
- 個人の属性が推定されるリスク
- データが他の情報と照合されるリスク
- データを用いて本人へアプローチされるリスク

抽出した識別子、属性、履歴ごとのリスクは次の通りである。

識別子

- 契約者 ID、氏名、電話番号、生年月日は、単体で個人を特定できたり、本人にアクセスできたりするリスクがある。
- 契約者 ID を変換した番号を、固定で管理すると、時間の経過、履歴の積み重ねによって、提供先事業者において、個人が特定されるリスクが高まる。一方、ワンタイムとすると、データの価値が大きく損なわれてしまう。

属性

- 住所は全て揃うと単体で個人を特定できたり、本人にアクセスできたりするリスクがある。
- 年収は、プライバシー性が比較的高い。
- 勤務先は、個人が特定されるリスクが比較的高い。
- 決済金融機関は、外部観察性は小さいものの、他の情報と組み合わせられると個人識別性が高くなる。

履歴

- カード利用明細データは、カード利用頻度の低い加盟店の場合、特定個人の識別性が高まる。
- 加盟店 POS データは、特定の商品の購入日が特定されると、本人が特定されるリスクがある。

4) 個人識別に係るリスクを踏まえた加工方法の検討

前ステップで抽出したリスクに応じて、検討した加工方法を以下に示す。

識別子

- 契約者 ID は、32 ビットの鍵付きハッシュによって変換する。この匿名加工情報に付する識別子の有効期間は、活用に必要な限度の期間とする。例：24 ヶ月
- 氏名、電話番号は削除する。
- 生年月日は、年代（10 代、20 代、30 代、40 代、50 代、60 代、70 代以上）の 7 区分にカテゴライズ化する。
- 外れ値に該当する利用者の ID は、提供対象から除外する。
例：年間利用金額が極めて大きい、短期間の決済利用回数が極めて多い、極めて高額な商品を決済した 等。

属性

- 住所は、都道府県、市町村を残し、残りを削除する。
- 年収は、100-300 万、300-500 万、500-800 万、800 万以上の 4 区分にカテゴライズ化する。
- 勤務先は、カテゴライズ化する。
- 加工後の属性は、組み合わせることで個人を識別する可能性があるため、次の方法のいずれか又は組み合わせて加工処理をする。
 - 方法 1：属性全体に対して k-匿名化の加工処理をする。このとき k 値は、データの特性に鑑みて設定する。
 - 方法 2：適切なサンプル抽出をする（※この方法では、k-匿名化はしない）。
 - 方法 3：データのスワッピングを行う。

履歴

- 利用日は、利用月単位に曖昧化する。
- 利用加盟店名は、カード決済頻度が極めて低いものを削除する。
- 年間利用金額が極めて大きい、短期間の決済利用回数が極めて多い契約者の情報を削除（外れ値処理）する。

4.3 移動データの事例

本事例では、鉄道の乗降履歴を例にしたケースを取り上げる。

鉄道の乗降履歴データは、駅エリアの集客力や集客層、潜在商圈の広さ、通勤圏、駅エリアを最寄り駅とする居住者の規模や構成などを把握することで、出店計画や立地評価、広告・宣伝計画などへ活用できることが期待される。

ケース1：イベント多客時流動分析用データ（4日間分）の匿名加工

1) ユースケースの明確化

匿名加工情報の提供先事業者とその利用目的：

イベント時期多客時の旅客流動について分析するため、自治体を含むイベント実行委員会、関連事業者、警備関係者等の第三者に対し、乗降データを提供。

データの使い方：

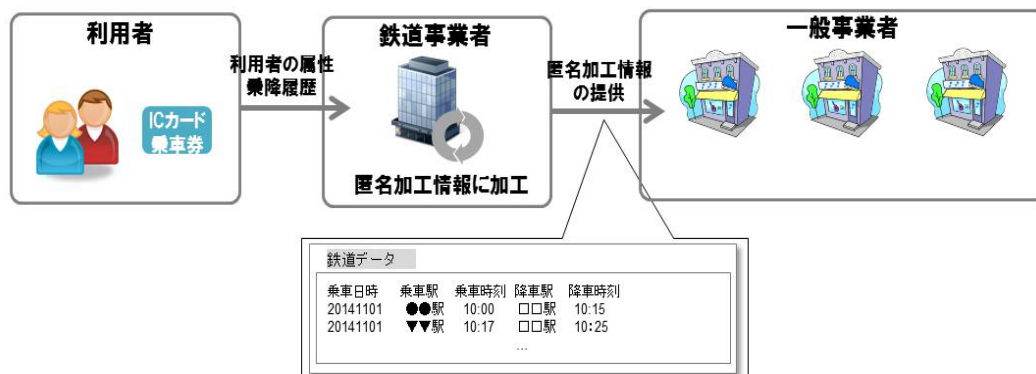
対象となるイベント会場の最寄駅等で降車した利用者の動きについて、当該期間データセットを匿名加工して提供し、以下のような分析に活用する。

- 最寄駅の当日増加人数及び年齢・性別・降車時間・利用者数地域順位分布
 - 駅周辺誘導人員・周辺施設の品揃え・仕入等の検討材料
 - 道路交通等含めた周辺影響範囲の告知や人員配置すべき範囲の検討材料
 - ピーク・閑散時間の把握や撤収時間の検討材料
 - どの地域に広告を出すべきかの検討材料

図表 23 ユースケースの概要

業務・サービス名	イベント多客時の流動分析用データの提供・販売				
業務・サービスの概要	イベント多客時において、イベント会場最寄駅の利用者の移動データ及び属性情報を提供				
特定される個人の総数	約●●万人				
個人情報	データ種別	データ件数	データ項目	項目	備考
	利用者属性情報	約▲▲万件	カードID	●桁数値	
			性別	男女	
			年齢	値	
			生年月日	yyyymmdd	
			郵便番号	7桁数字	
			都道府県	名称	
			市区町村	名称	
	利用履歴情報	約△△万件	カードID	●桁数値	
			処理名称	出場・入場	
			年月日	yyyymmdd	
			時間	hhmmss	
			利用種別	入場／出場	
			改札口	名称	
			出場駅	名称	※イベント会場 最寄駅
			SF入場駅	名称	
			SF出場駅	名称	
			利用額	値	
			残額	値	
匿名加工情報の利用目的	イベント時の誘導計画、イベント周辺エリアへの影響把握、告知の検討など				

図表 24 ケースの全体イメージ



データの内容：

本ケースで用いるデータは、大きく分けて、IC カード乗車券に任意で記入される利用者の属性情報と、定期券に係る情報、乗降履歴を記録する利用履歴情報の三種類で、これらは各 IC カード乗車券に付されたカード ID によってリンクされている。

図表 25 データセットサンプル

利用者属性情報

カードID	性別	年齢	生年月	郵便番号	都道府県 漢字	市区町村 漢字
α α α	男性	39	197711	xxxxxxx	〇〇県	▲▲市
β β β	女性	33	198305	xxxxxxx	●●県	□□市

定期券発売情報

カードID	定期券 開始年月日	定期券 終了年月日	定期券発駅	定期券着駅	通勤定期 フラグ	通学定期 フラグ
α α α	20150826	20160225	X	Y	1	0
β β β	-	-	-	-	-	-

利用履歴情報

カードID	処理名称	年月日	時間	利用種別	改札口	出場駅	SF入場 駅	SF出場 駅	利用額	残額
α α α	出場	20150912	115244	SF入場SF出場	A2	A	× ×	A	300	2590
β β β	出場	20151118	092225	SF入場SF出場	A4	A	△△	A	500	1420

※定期券の有効区間内の利用履歴情報は無い。

2) 識別子、属性、履歴の仕分け

データのセットサンプルを基に、本ケースに必要なデータを抽出し、それらを識別子、属性、履歴を仕分けした結果を次図表に示す。なお、ここで履歴として分類した乗降履歴のデータは、位置に関する情報であり、個人識別リスクが生じる場合があることに留意する必要がある。

図表 26 識別子、属性、履歴の仕分け結果

利用者属性情報

カードID	性別	年齢	生年月	郵便番号	都道府県 漢字	市区町村 漢字
α α α	男性	39	197711	xxxxxxx	〇〇県	▲▲市
β β β	女性	33	198305	xxxxxxx	●●県	□□市

識別子 属性

利用者履歴情報

カードID	処理名称	年月日	時間	利用種別	改札口	出場駅	SF入場 駅	SF出場 駅
α α α	出場	20160401	072715	SF入場SF出場	A2	A	× ×	A
β β β	出場	20160403	142638	SF入場SF出場	A4	A	△△	A

識別子 履歴

3) 個人識別に係るリスクの抽出

個人識別に係るリスクとして、個人が特定されるリスクを抽出した。

移動データの場合、特に長期間の履歴を対象とする場合には、データを用いて本人へアプローチされるリスクがあるため考慮する必要がある。本ケースは、履歴が短期間に限定されていることから、リスクは小さいものと思量して対象外とした。

(評価対象とする個人識別に係るリスク)

- 個人が特定されるリスク

抽出した識別子、属性、履歴ごとのリスクは次の通りである。

識別子：

- 個人データを構成する加工前の ID を用いると容易に照合される。
- 他の匿名加工情報を作成する際に生成した仮 ID と同一の仮 ID を用いると、復元できるリスクが高くなる。

属性：

- 性別、年齢、住所情報（郵便番号、都道府県、市区町村）は組み合わせると、特定のリスクが高くなる。
- 郵便番号（7 桁）は、市区町村の範囲よりも対象とする領域が小さいため、特定のリスクが高くなる。

履歴：

- 乗降履歴と、外部の情報（イベント参加の事実等）との照合によって、本人が特定される可能性がある。

4) 個人識別に係るリスクを踏まえた加工方法の検討

前ステップで抽出したリスクに応じて、検討した加工方法を以下に示す。

図表 27 個人識別に係るリスクを踏まえた加工方法の例

情報分類	対象情報	加工方法	備考
識別子	カード ID	カード ID は、鍵付きハッシュ化を行い、不可逆的に変換することで、仮 ID を生成する。	- 仮 ID は利用者に対して同一で、利用者の元のカード ID 等が復元できなく、その識別子は期間に制限され、他の匿名加工情報提供時に同一の人が、同一の識別子にならないようにする。 - 有効な方法としてハッシュ化を選択しているが、必ずしもこの方法には限らない。
属性	性別	加工無し。	加工の際、k-匿名化の必要性を考慮する。
	年齢	カテゴライズ化（5 歳刻み、10 歳刻み等）する。	
	生年月	必要性が低いため削除。	
	郵便番号	必要性が低いため削除。	
	都道府県漢字	加工無し。	
	市区町村漢字	加工無し。	

履歴	処理名称、年月日、時間、利用種別、改札口、出場駅、SF 入場駅、SF 出場駅	加工無し。	・特異な値や傾向を持つ履歴は、個人識別性が高いものとして特異値となり得るが、ケースでの該当はなかった。
----	--	-------	---

ケース 2：イベント多客時流動分析用データ（1 ヶ月間分）の匿名加工

ケース 1 の期間（4 日間分）を長くしたケース（ケース 2：1 ヶ月分）についても検討を行った。その結果、加工方法に大きな違いがなかったため、ケース 2 においては、2）識別子、属性、履歴の仕分け～4）個人識別に係るリスクを踏まえた加工方法の検討の記載については省略する。

1）ユースケースの明確化

匿名加工情報の提供先事業者とその利用目的：

イベント時期の多客時の旅客流動について分析するため、自治体を含むイベント実行委員会、関連事業者、警備関係者等の第三者に対し、乗降データを提供。

データの使い方：

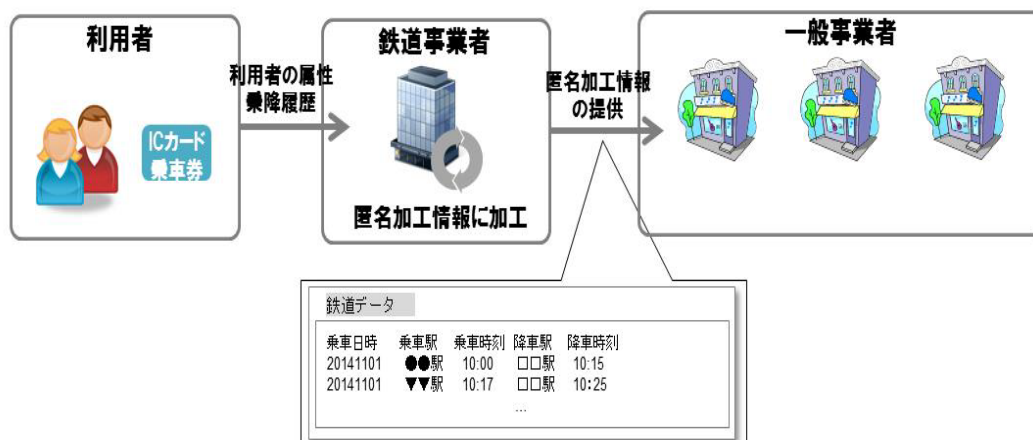
対象となるイベント会場の最寄駅等で降車した利用者の動きについて、当該期間データセットを匿名加工して提供し、以下のような分析に活用する。

- ・ 最寄駅の当日増加人数及び年齢・性別・降車時間・利用者数地域順位分布
 - 駅周辺誘導人員・周辺施設の品揃え・仕入等の検討材料
 - 道路交通等含めた周辺影響範囲の告知や人員配置すべき範囲の検討材料
 - ピーク・閑散時間の把握や撤収時間の検討材料
 - どの地域に広告を出すべきかの検討材料

図表 28 ユースケースの概要

業務・サービス名	イベント多客時の流動分析用データの提供・販売				
業務・サービスの概要	イベント多客時において、イベント会場最寄駅の利用者の移動データ及び属性情報を提供				
特定される個人の総数	約●●万人				
個人情報	データ種別	データ件数	データ項目	項目	備考
	利用者属性情報	約▲▲万件	カードID	●桁数値	
			性別	男女	
			年齢	値	
			生年月日	yyyymmdd	
			郵便番号	7桁数字	
			都道府県	名称	
			市区町村	名称	
	利用履歴情報	約△△万件	カードID	●桁数値	
			処理名称	出場・入場	
			年月日	yyyymmdd	
			時間	hhmmss	
			利用種別	入場／出場	
			改札口	名称	
			出場駅	名称	※イベント会場 最寄駅
			SF入場駅	名称	
			SF出場駅	名称	
			利用額	値	
			残額	値	
匿名加工情報の利用目的	イベント時の誘導計画、イベント周辺エリアへの影響把握、告知の検討など				

図表 29 ケースの全体イメージ



データの内容：

本ケースで用いるデータは、大きく分けて、IC カード乗車券に任意で記入される利用者の属性情報と、定期券に係る情報、乗降履歴を記録する利用履歴情報の三種類で、これらは各 IC カード乗車券に付されたカード ID によってリンクされている。

図表 30 データのセットサンプル

利用者属性情報

カードID	性別	年齢	生年月	郵便番号	都道府県 漢字	市区町村 漢字
α α α	男性	39	197711	xxxxxxx	〇〇県	▲▲市
β β β	女性	33	198305	xxxxxxx	●●県	□□市

定期券発売情報

カードID	定期券 開始年月日	定期券 終了年月日	定期券発駅	定期券着駅	通勤定期 フラグ	通学定期 フラグ
α α α	20150826	20160225	X	Y	1	0
β β β	-	-	-	-	-	-

利用履歴情報

カードID	処理名称	年月日	時間	利用種別	改札口	出場駅	SF入場 駅	SF出場 駅	利用額	残額
α α α	出場	20150912	115244	SF入場SF出場	A2	A	× ×	A	300	2590
β β β	出場	20151118	092225	SF入場SF出場	A4	A	△△	A	500	1420

※定期券の有効区間内の利用履歴情報は無い。

ケース 3：観光活性化施策検討用データ（長期間）の匿名加工

1) ユースケースの明確化

匿名加工情報の提供先事業者とその利用目的：

観光地（観光エリアに含まれる 7 駅：A 駅～G 駅）のさらなる活性化施策を検討するにあたり当該観光エリアの駅の特徴を把握し外部からの観光客を誘致するため、自治体や関連事業者等に対し、乗降データを提供する。

データの使い方：

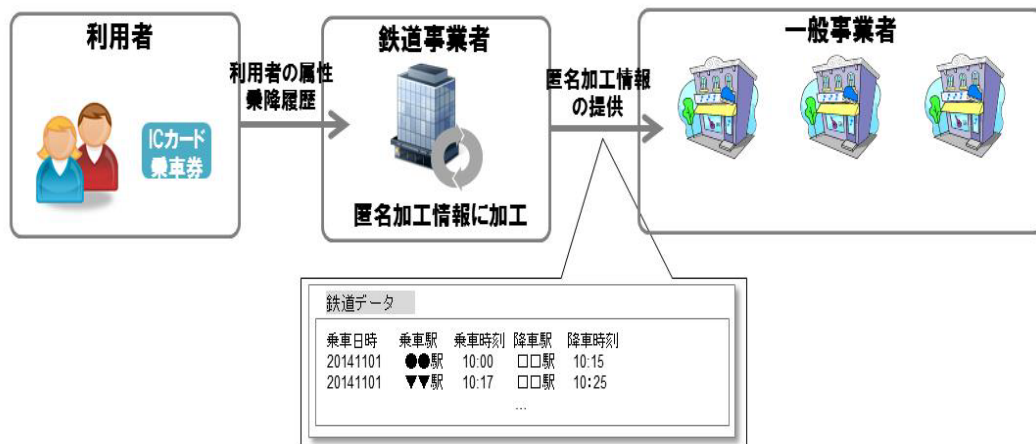
当該観光エリアの駅の利用者について、中長期にわたって特徴を把握できるデータセットを匿名加工して提供し、以下のような分析に活用する。

- 観光エリア内の駅利用者数の年変動分布、及び年齢・性別分布
 - 現状把握の検討材料
- 観光エリア内の駅利用者数に関して、一定期間内のリピーター数分布
 - 駅別の魅力度確認や活性化策等の検討材料
- 観光エリア内流動・滞在時間
 - エリア内の混雑平準化、回遊策等の検討材料
- 利用者数地域順位分布
 - どの地域に広告を出すべきかの検討材料

図表 31 ユースケースの概要

業務・サービス名	観光地における観光活性化施策検討用のデータ提供・販売				
業務・サービスの概要	観光地エリアの駅利用者数の推移や利用者属性の変化などから、各駅の特徴を把握し、観光客誘致の施策検討材料として提供				
特定される個人の総数	約●●万人				
個人情報	データ種別	データ件数	データ項目	項目	備考
	利用者属性情報	約▲▲万件	カードID	●桁数値	
			性別	男女	
			年齢	値	
			生年月日	yyyymmdd	
			郵便番号	7桁数字	
			都道府県	名称	
			市区町村	名称	
	定期券発売情報	約■■万件	定期券開始年月日	yyyymmdd	
			定期券終了年月日	yyyymmdd	
			定期券発駅	名称	
			定期券着駅	名称	
			通勤定期フラグ	値	
			通学定期フラグ	値	
	利用履歴情報	約△△万件	カードID	●桁数値	
			処理名称	出場・入場	
			年月日	yyyymmdd	
			時間	hhmmss	
			利用種別	入場／出場	
			改札口	名称	
			出場駅	名称	
			SF入場駅	名称	
			SF出場駅	名称	
			利用額	値	
			残額	値	
匿名加工情報の利用目的			観光エリア内のエリア内流動の現状把握、混雑平準化施策の検討、回遊施策の検討、告知検討など		

図表 32 ケースの全体イメージ



データの内容：

本ケースで用いるデータは、大きく分けて、IC カード乗車券に任意で記入される利用者の属性情報と、定期券に係る情報、乗降履歴を記録する利用履歴情報の三種類で、これらは各 IC カード乗車券に付されたカード ID によってリンクされている。

図表 33 データのセットサンプル

利用者属性情報

カードID	性別	年齢	生年月	郵便番号	都道府県 漢字	市区町村 漢字
γγγ	男性	39	197711	xxxxxxx	〇〇県	▲▲市
δδδ	女性	33	198305	xxxxxxx	●●県	□□市

定期券発売情報

カードID	定期券 開始年月日	定期券 終了年月日	定期券発駅	定期券着駅	通勤定期 フラグ	通学定期 フラグ
γγγ	20150826	20160225	X	Y	1	0
δδδ	-	-	-	-	-	-

利用履歴情報

カードID	処理名称	年月日	時間	利用種別	改札口	出場駅	SF入場 駅	SF出場 駅	利用額	残額
γγγ	出場	20150912	115244	SF入場SF出場	B1	B	××	B	300	2590
δδδ	出場	20151118	092225	SF入場SF出場	C1	C	△△	C	500	1420

※定期券の有効区間内の利用履歴情報は無い。

2) 識別子、属性、履歴の仕分け

データのセットサンプルを基に、本ケースに必要なデータを抽出し（住民情報に関するデータは抽出しない）、それらを識別子、属性、履歴を仕分けした結果を次図表に示す。なお、ここで履歴として分類した乗降履歴のデータは、位置に関する情報であり、個人識別リスクが生じる場合があることに留意する必要がある。

図表 34 識別子、属性、履歴の仕分け結果

利用者属性情報						
カードID	性別	年齢	生年月	郵便番号	都道府県 漢字	市区町村 漢字
γγγ	男性	39	197711	xxxxxxx	〇〇県	▲▲市
δδδ	女性	33	198305	xxxxxxx	●●県	□□市
識別子			属性			

定期券発売情報						
カードID	定期券 開始年月日	定期券 終了年月日	定期券発駅	定期券着駅	通勤定期 フラグ	通学定期 フラグ
γγγ	20150826	20160225	X	Y	1	0
δδδ	-	-	-	-	-	-
識別子			属性			

利用履歴情報								
カードID	処理名称	年月日	時間	利用種別	改札口	出場駅	SF入場 駅	SF出場 駅
γγγ	出場	20150912	115244	SF入場SF出場	B1	B	××	B
δδδ	出場	20151118	092225	SF入場SF出場	C1	C	△△	C
識別子				履歴				

3) 個人識別に係るリスクの抽出

個人識別に係るリスクとして、“個人が特定されるリスク”を抽出した。移動データの場合、特に長期間の履歴を対象とする場合にはデータを用いて本人へアプローチされるリスクがあるため考慮する必要がある。

(評価対象とする個人識別に係るリスク)

- ・ 個人が特定されるリスク

抽出した識別子、属性、履歴ごとのリスクは次の通りである。

識別子：

- 個人データを構成する加工前の ID を用いると容易に照合される。
- 仮 ID は、属性と履歴の同一カードに同一の仮 ID を付与するため、長期間同一の ID で管理すると、履歴の積み重ねによって、特定の個人を識別できるリスクが高まる。
- 他の匿名加工情報を作成する際に生成した仮 ID と同一の仮 ID を用いると、復元できるリスクが高くなる。

属性：

- 性別、年齢、住所情報（郵便番号、都道府県、市区町村）は組み合わせると、特定のリスクが高くなる。
- 郵便番号（7 桁）は、市区町村の範囲よりも対象とする領域が小さいため、特定のリスクが高くなる。

履歴：

- 乗降履歴と、外部の情報（イベント参加の事実等）との照合によって、本人が特定される可能性がある。
- 乗降履歴を解析すると、パターン性が判別され、定期的に通う駅が推定され当該エリア居住者等の情報が特定されるリスクが高まり、さらにアプローチされるリスクの生じる可能性がある。
- 乗降履歴を長期間解析すると、住民の最寄り駅と職場の最寄り駅が推定される可能性がある。
- 乗降履歴データを継続的に提供する場合、積み重ねられた乗降履歴の一意性から、履歴をつなげられてしまい、個人を識別されるリスクがある。
- 着駅を観光エリア 7 駅に絞っているため、着駅を絞らない場合よりは特定されるリスクは低いもののランダム抽出ではないため、サンプリング情報よりはリスクが高い。
- 元のデータから作り出ている履歴は、長さが長ければ照合されるリスクが高くなる。
- ある特異な履歴レコード（の利用者）が識別されてしまうと、それ以降の同一の仮 ID が振られた全ての履歴を追跡されてしまうリスクがある。

4) 個人識別に係るリスクを踏まえた加工方法の検討

前ステップで抽出したリスクに応じて、検討した加工方法を以下に示す。

図表 35 個人識別に係るリスクを踏まえた加工方法の例

情報分類	対象情報	加工方法	備考
識別子	カード ID	・カード ID は、鍵付きハッシュ化を行い、不可逆的に変換することで、仮 ID を生成する。 ・長期間にわたる場合は利用目的に応じて ①適当な時間毎に更新する ②他の ID とスワッピングする 等の加工を検討する。	・仮 ID は利用者に対して同一で、利用者の元のカード ID 等が復元できなく、その識別子は期間に制限され、他の匿名加工情報提供時に同一の人が、同一の識別子にならないようにする。 ・有効な方法としてハッシュ化を選択しているが、必ずしもこの方法には限らない。
属性	性別	加工無し。	・加工の際、k-匿名化の必要性を考慮する。
	年齢	カテゴライズ化（5 歳刻み、10 歳刻み等）する	
	生年月	必要性が低いため削除。	
	郵便番号	必要性が低いため削除。	
	都道府県漢字	加工無し。	
	市区町村漢字	加工無し。	

	定期券情報	定期券開始年月日、定期券終了年月日、定期券発駅、定期券着駅、通勤定期フラグ、通学定期フラグ	加工無し。	・対象エリア内の定期券は乗降履歴が記録されないため、乗降推定用に有効な定期券発行枚数を用いる。 ・定期券情報は、最低半年ごとに書き換えられるものであり、履歴的な性質を有している。
履歴	処理名称、年月日、時間、利用種別、改札口、出場駅、SF 入場駅、SF 出場駅	・7 駅に限定されているため、着駅を絞らない場合よりはリスクは低いものの、 ①仮 ID を制御して他の仮 ID と紐付かないようにする ②ランダム化（一部の情報をランダム化する事も有効） ③履歴の長さを見て一部を削除する 等の加工をすることでよりリスクが軽減される。	・住民情報と判断されうる（7 駅のエリア内で繰り返しの乗降履歴がある等）乗降履歴は削除。 ・特異な値や傾向を持つ履歴は、個人識別性が高いものとして特異値となり得るが、本ケースでの該当はなかった。 ・乗り越しで乗降する定期券情報は履歴の中に含まれないので、推定情報として加算されるため、個人を特定されるリスクはきわめて少ない。	

参考資料 匿名加工方法に係る用語説明

図表 36 匿名加工方法に係る用語説明

番号	用語	解説
1	k-匿名化	単独では個人を識別できないが、複数を組み合わせることで個人を高い確率で識別することが可能な属性（例えば、性別、年齢、居住地、職業等）について、どの属性値の組み合わせでも、対象とするデータ中に必ず k 件以上存在する状態にすること。
2	仮名化	削除又は復元することのできる規則性を有しない方法により他の記述等に置き換えること。
3	一般化	属性の値を上位の値や概念に置き換えること。例えば、10 歳刻み、キュウリ→野菜にするなど。
4	トップ（ボトム）コーディング	数値属性に対して、特に大きい、もしくは小さい属性値をまとめること。例えば、100 歳以上の人を「100 歳以上」とするなど。
5	ノイズ（誤差）付加	数値属性に対して、一定の分布に従った乱数的なノイズを加えること。
6	スワッピング（データ交換）	カテゴリ属性に対して、レコード間で属性値を（確率的に）入れ替えること。データスワップとも呼ぶ。
7	サンプリング	元データ全体から一定の割合・個数でランダムに抽出すること。
8	グルーピング	属性や履歴において、詳細な項目を一定のまとまりや区分に置き換えること。

出所) 1、3、5、6：内閣官房「技術検討ワーキンググループ報告書」（平成 25 年 12 月 10 日）を基に作成

4、6：内閣官房「技術検討ワーキンググループ報告書」（第 13 回パーソナルデータに関する検討会配布資料（平成 25 年 12 月 10 日）、「匿名化技術の現状について」（パーソナルデータに関する検討会 第 1 回技術検討ワーキンググループ 高橋構成員提出資料（平成 25 年 9 月 27 日））を基に作成

8：瓜生和久「一問一答 平成 27 年改正個人情報保護法」（商事法務、平成 27 年）を基に作成