

非識別加工情報と共通語彙基盤について

2019年3月20日

日本電気株式会社

公共システム開発本部

岩田

1 非識別加工情報(ビッグパーソナルデータ)について

非識別加工情報を入手することのメリット

○非識別加工情報(匿名加工情報)の定義

個人情報を個人情報の区分に応じて定められた措置を講じて、「特定の個人を識別することができないように加工して得られる個人に関する情報」であって、当該個人情報を復元して特定の個人を再識別することができないようにしたもの

○匿名加工情報の利用者の義務(個人情報保護法による規制)

- ・匿名加工情報を第三者提供するときは、項目および提供方法を公表する。
- ・匿名加工情報を利用するときは、元の個人情報に係る本人を識別しない。
- ・匿名加工情報の適正な取扱いを確保するための措置を自ら講じる(努力義務)。



「匿名加工情報」は、本人同意が無くても、様々な目的に利用したり、第三者提供したりすることが可能なパーソナルデータ(ビッグデータ)である。



○公開情報等ではなく非識別加工情報を入手(使用)することのメリット

行政機関等が業務遂行の目的で保有する個人情報をもとに、匿名加工を行うため

- ・情報が悉皆的であり個人の漏れがないこと。
- ・個人に対する情報の種類や蓄積量が多いこと。
- ・行政情報であるため情報が新鮮かつ正確であること。

などがメリットとして考えられる。

2 非識別加工情報の提供に関する制度について

非識別加工情報を入手するための事務手続きの流れ



民間事業者

- 不適格な者は除外
 - ・過去に禁固以上の刑に処され2年を経過しない者
 - ・過去に義務違反があり利用契約を解除され2年を経過しない者
- 提供を受けた場合
 - ・識別行為の禁止
 - ・安全管理措置
 - ・契約内容の遵守
- 手数料を納付
※提案者意外も提供を受けることが可能
- 契約事項を遵守



厚労省のNDBオーダメイド集計は
・加工料金 $5,900 \times$ 加工に要した時間
とされている。

「行政機関個人情報保護法等改正法の施行に伴う関係政令の整備及び経過措置に関する政令」より

(行個法)手数料は以下の合計
・基本料金21,000(事務費)
・加工料金 $3,950 \times$ 加工に要した時間
・加工を委託した場合の委託費
を委員会が定める書面に「証紙」を貼って納付

行政機関・地方公共団体等

- 提案しようとする者への情報提供
- 対象となる個人情報
 - ・個人情報ファイル簿が公表されていること
 - ・情報公開請求があれば部分開示されること
 - ・行政運営に支障を生じないこと
- 提案についての審査
- 非識別加工情報の作成、公表
 - ・基準に基づく適正加工
 - ・個人情報ファイル簿への記載
- 苦情処理

地方公共団体は条例整備が進んでいないので
提案を受付けていない!!



「行政機関個人情報保護法・独法等個人情報保護法の改正に向けた考え方」(平成28年3月7日 総務省)より

http://www.soumu.go.jp/menu_news/s-news/01gyokan06_02000035.html

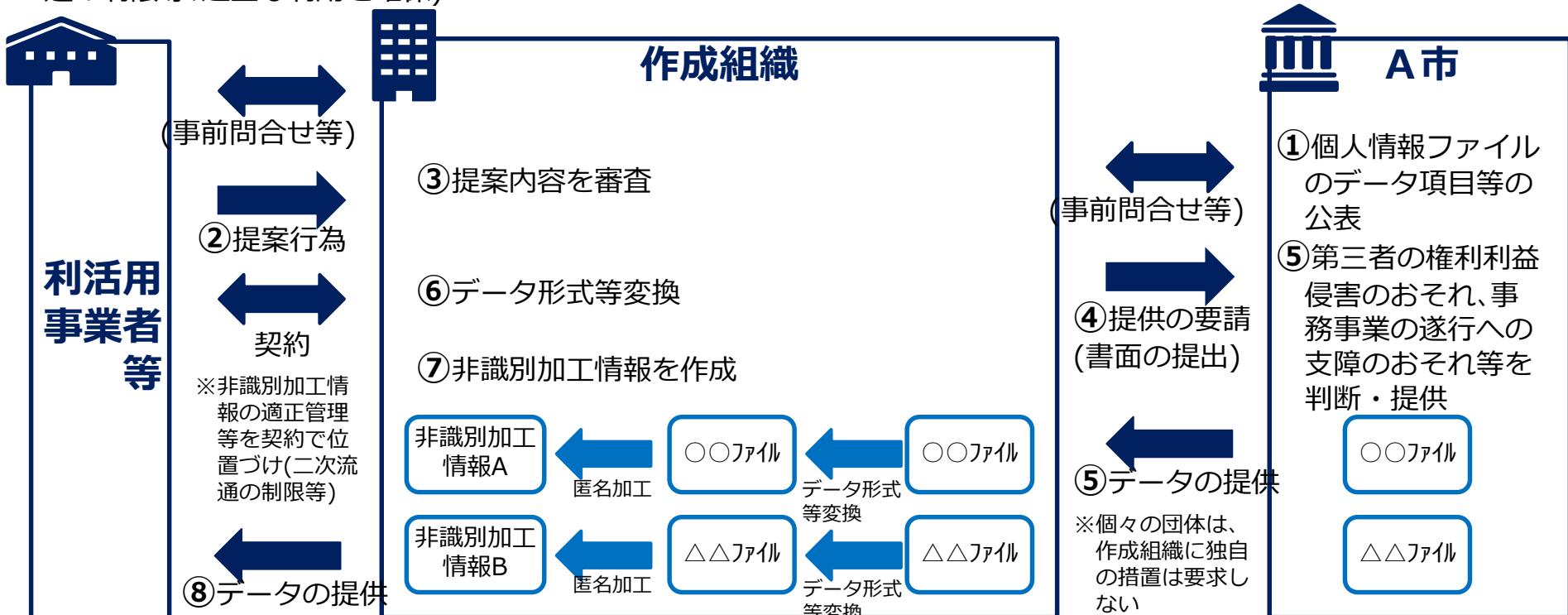
「行政機関の保有する個人情報の保護に関する法律第四章の二の規程による行政機関非識別加工情報の提供に関する規則」(平成29年3月31日 個人情報保護委員会)より

https://www.ppc.go.jp/files/pdf/gyoukohou_kisoku.pdf

3 非識別加工情報の提供に関する制度の方向性について

「地方公共団体の非識別加工情報の作成・提供に係る効率的な仕組みの在り方に関する検討会(第4回)」(平成30年12月13日 総務省)より編纂

- ①地方公共団体は、個人情報ファイルに記録されるデータ項目等について公表(国によるポータルサイトの構築も検討)
- ②利活用事業者等は、作成組織に対して非識別加工情報の作成・提供に関する提案を実施
- ③作成組織において、②提案内容について、利用目的や適正管理等の内容を審査
- ④作成組織より、地方公共団体に対し、②提案に対応する個人情報の提供を要請(書面において、利活用事業者・利用目的・適正管理等を明示)
- ⑤地方公共団体は、④を受けて個人情報の提供を判断し、提供(提供の際、作成組織に独自の措置は要求しない)
- ⑥作成組織は、必要に応じて、地方公共団体から提供されたデータの形式等を整理
- ⑦作成組織において、非識別加工情報を作成(地方公共団体から提供を受けた個人情報ファイル毎に匿名加工を実施)
- ⑧利活用事業者等に対して、非識別加工情報を提供(作成組織と利活用事業者間の契約において、非識別加工情報の二次流通の制限等、適正な利用を確保)



4 非識別加工情報のプロファイリングへの活用について

マーケティングにおけるプロファイリングの基準変数への活用

顧客特性

X

消費行動特性

デモグラフィック属性(人口統計的要因)

性別、年齢、職業、収入、家族構成、ライフステージ、人種、教育水準、世代、社会階層など

ジオグラフィック属性(地理的要因)

居住地域、国の違い、文化、気候、人口密度など

サイコグラフィック属性(心理的要因)

ライフスタイル、価値観、性格、宗教など

ビヘイビアル属性(消費行動的要因)

消費パターン(ライトユーザ、ヘビーユーザ)、ロイヤルティ、価格感度、特売反応など

非識別加工情報(行政機関)内に、
医療・介護等はレセプトが存在す
るが、他業務では消費要素が存在
しないことが課題。

住民票(性別、年齢、続柄etc)、
農地基本台帳(耕作主、耕作物etc)、税(収入、職業、自動車所有etc)…

住民票(住所、前住所、国籍etc)、戸籍(出生地etc)、固定資産(土地、家屋etc)…

住民票(国籍、特別永住者etc)、
住民アンケート、健診(成人、乳幼児)・検診(内科、歯科)問診、健診・検診結果、飼犬…

医療・介護レセプト(診療科目、
介護内容etc)、施設利用、図書貸出、生涯学習、ボランティア、各種ログ…

5 非識別加工情報のデータマイニングへの活用について

研究開発における機械学習・データマイニングへの活用

■ 教師あり学習 (例:科学的介護)

入力データ数=出力データ数



仮説に基づく幾多のパラメータから、因果関係を評価

出力データ

介入後介護度データ

○○による要介護度○○の人々、
○○サービスを提供すると、
要介護度が○○に良化するから
○○サービスを強化しよう…
逆に○○サービスは効果がないからやめよう…

■ 半教師あり学習 (例:データヘルス)

入力データ数≥出力データ数



幾多のパラメータから相関関係を見つけ、因果関係を予想

出力データ

医療・介護レセデータ

(因果関係を導出)
当地域では40歳で血糖値○○以上の人々が、かなりの確立で、○○病になってるから、対象者を集めて教育をしよう…

■ 教師なし学習



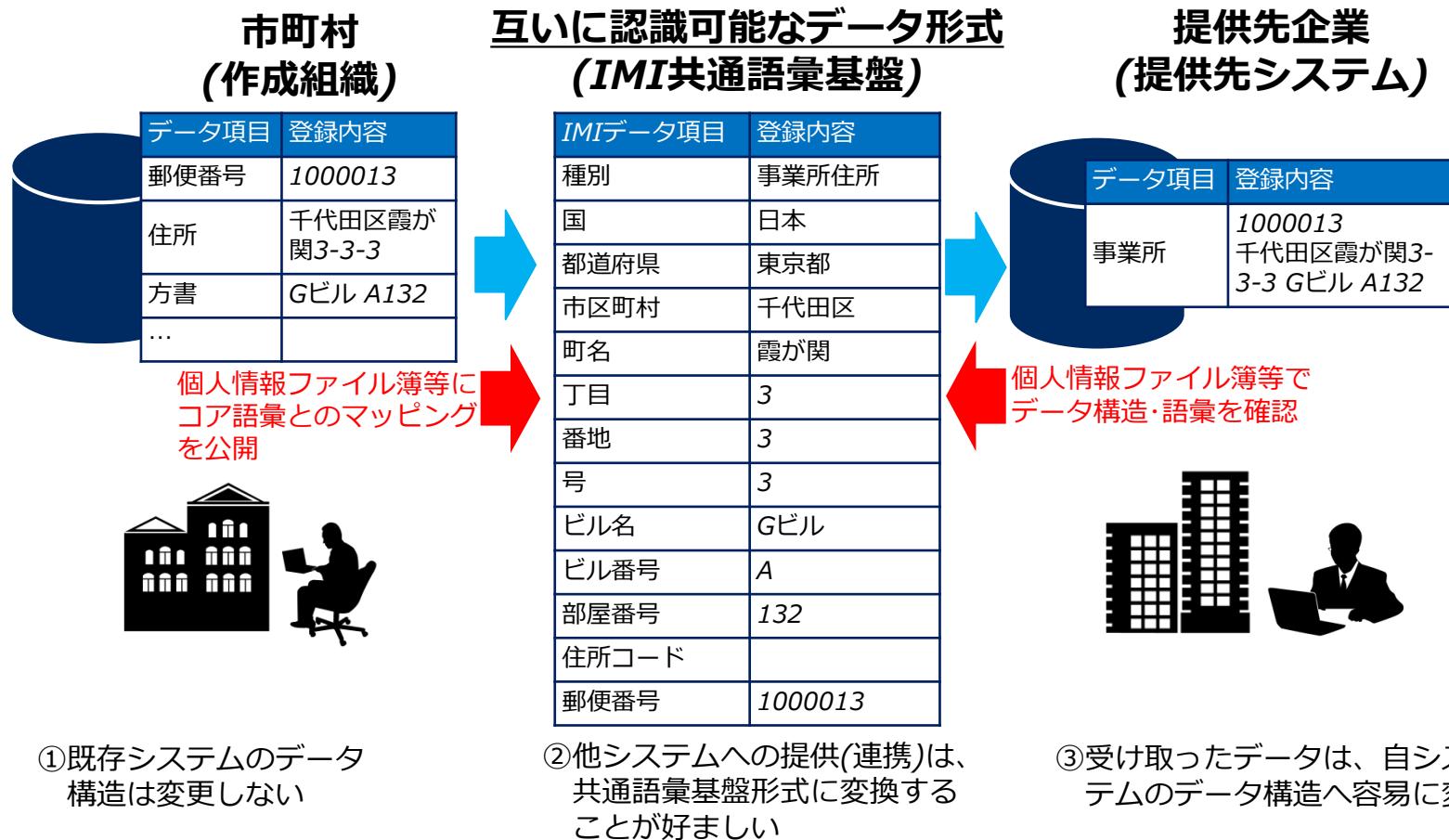
幾多のパラメータから相関関係を見つけ仮説を導出

(相関から仮説導出)
単身世帯で年収が○○円以下だと、
○○状態に陥りやすいという相関がありそうだ。違うデータでも
仮説を検証してみよう…

6 非識別加工情報における共通語彙基盤の活用について

自治体の既存データから共通語彙基盤への変換(マッピング)を意識して、個人情報ファイル簿の公開し、共通語彙基盤に対応した形で非識別加工情報データを提供すべきだが、まだまだ公共データを受けきれるだけのドメイン語彙には至っていないことが課題。。。

例:住所のデータ変換手続



6 非識別加工情報における共通語彙基盤の活用について

行政データを扱う上での課題の一例(共通語彙基盤も同様)

<氏名表記の問題>

- ・住民票の氏名には、外国人の場合は、原則としてアルファベット表記だが、漢字圏の外国人については、在留カード等の記載に倣いアルファベットと漢字(正字)を併記して設定される場合がある。
例：「ZHANG YULIAN 張玉蓮」
- ・また、外国人住民の「氏名以外で国内における社会生活上通用している呼称」で、(表札や宛先など)居住関係の公証のための氏名は、住民票上では通称という別の欄に記載されている。なお、日常生活上では、通称が氏名として使われる。
例：「氏名：ZHANG YULIAN 張玉蓮、通称：山田 花子」

<住所表記の問題>

- ・日本には地番(土地登記簿上の一筆を示す表記で登記所が指定)と住所(建物を示す表記で市町村が指定)の表記が混在しており、用途(手続き)によって、どちらを使用するのかが異なるため、そこを整理する必要がある。
例：「六本木ヒルズの住居表記：東京都港区六本木6丁目10番1号、地番表記：東京都港区六本木6丁目502番地1」
- ・住所は、総務省の事務連絡により平成23年からは「住居表示」または「地番」の記載のみでは住所が明らかでない場合には、「アパート名」、「居室の番号」までを記載するとされたが、届出された方書が正確かどうかを把握する術がなく、平成23年以前から住民異動をされていない方は、現在でも方書の無い住民票が存在する。
- ・複数の土地にまたがって建つ建物に対する家屋登記の「家屋番号(建物の敷地の地番に棟番号を加えたもの)」に記載する地番は、土地にかかる建物の床面積の大きいほうの「地番」に紐付ける形で登記するため、土地の「地番」と、家屋の「家屋番号」といった登記上の記載(データ)だけでは、土地と建物が正確に紐づかない。

7 非識別加工情報×GISマーケティング

GISマーケティングにおけるデータ活用について

国・地方を通じた厳しい財政状況と、急速に進む少子高齢化に加えて、地方から関東圏への人口流動といったように、持続可能な社会保障の基盤づくりにかけない自治体が抱える問題は、地域によってさまざまで、これまでのような全国統一的な施策では課題解決が困難な状況が進んでいます。

自治体が自ら保有するデータを分析して地域における課題を抽出し、課題を克服するための施策に財政を集中的に投入するといった、メリハリのきいた地域課題解決型の政策であるEBPMが求められ、地域特有の課題を発見し解決することへ財政を投入するといったメリハリの効いた行政が必須となってきています。

EBPMで思いつくのは、クロス集計などに代表される統計的なデータ分析の手法ですが、統計は社会保障や税といった個にアプローチした課題の分析には有効ですが、震災や水害といったような面からのアプローチには向きです。

昨今、各地で発生している震災によって、課題を面から捉えようとしたときに有効な手段である空間情報、すなわちGIS(地理情報システム)が再び注目されてきており、EBPMにおいても、個と面からの複合的なアプローチによって課題の抽出を効率的に行う「GISマーケティング」が注目されてきています。

7 非識別加工情報×GISマーケティング

一般企業

店舗開発

- ・新規出店計画・立地評価
- ・既存店評価・活性化
- ・販売店適正配置計画

顧客データ分析

- ・会員データ、POSデータの有効活用

地方公共団体に置き換えると

販売促進

- ・広告・宣伝計画
- ・小売・流通業商圈分析
- ・市場分析

顧客データ管理

- ・営業業務支援
(顧客属性データを投入し、顧客分布と地域毎の顧客シェアを把握等)

たとえばこんな使い方

購買履歴を可視化。
地域の特性（売れ筋）
把握と品揃え計画！

ネット顧客居住地把握
によるリアル店舗
への誘導計画（DM等）

人口統計・顧客分布・
競合店を重ねて出店
計画！



地方公共団体

政策・インフラ

- ・新規計画・存続評価
- ・既存計画評価・活性化
- ・適正配置・適正配分計画

地域住民データ分析

- ・住民情報系業務データの有効活用(需要分析)
- ・財務データの有効活用(供給分析)

政策・協働促進

- ・政策における圏域分析
- ・環境・需要分析
- ・ポスティング計画

地域住民データ管理

- ・EBPM、PDCA支援
(住民情報系業務データを投入し、住民分布と地域ごとの特徴を把握)

たとえばこんな使い方

ピロティ構造、旧建築基準、木造家屋等を可視化し、震災・火災時の避難経路の有効性を検討

独居老人、在宅療養者等、単独での移動困難者を可視化し、避難支援策を検討

商業施設・店舗等の情報を重ねて災害時の商品供給の可能性を検討

