

資料

保存期間：5年
(令和10事務年度末)
令和5年11月28日

第7回 国税庁保有行政記録情報の 整備に関する技術検証WG

国税庁 企画課
データ活用推進室

資料内容

1. 本ワーキンググループの位置づけ

2. これまでの議論

3. 本日も検討いただきたい内容

4. 今後のスケジュール

1. 本ワーキンググループの位置づけ

- 国税庁が保有する行政記録情報のオープン化に向けた検討を効率的に行うため、法的な課題及び技術的な課題に対する具体的な対応方法について検討・確認を行うことを目的として、国税庁保有行政記録情報の整備に関する有識者検討会の下で、本ワーキンググループ（以下「WG」という。）を開催する。

「国税庁保有行政記録情報の整備に関する有識者検討会」開催要綱（抜粋）

3 運営

- (2) 座長は必要があると認めるときは、検討会にワーキンググループを置くことができる。
なお、ワーキンググループにおける検討結果は、有識者検討会に報告するものとする。

- 第7回となる本WGでは、提供データに施す匿名加工技法に関して、**技術的課題を検証**することを目的として開催。
 - WGにおける検討結果については、事務局（国税庁企画課データ活用推進室）において整理の上、「国税庁保有行政記録情報の整備に関する有識者検討会」に対して適宜報告することとする。
 - 第7回WGの構成員は、以下のとおり（敬称略）。
- | | |
|-------|-------------------------|
| 伊藤 伸介 | 中央大学 経済学部 教授 |
| 菅 幹雄 | 法政大学 経済学部 教授 |
| 星野 伸明 | 金沢大学 人間社会研究域 経済学経営学系 教授 |
| 南 和宏 | 統計数理研究所 データ科学研究系 教授 |

2. これまでの議論（匿名加工技法について）

- 非識別化の手法は、下表のとおり、様々な知見の蓄積がある一方、対象データや求めるレベルに応じて、適用すべき技法は様々。
- どの水準まで加工が必要か、技術視点、ユーザー視点、法的視点等から検討する必要。

No	代表的な技法例	技法例	概要
1	属性情報の削除	属性（列）削除	直接個人を特定可能な属性（氏名等）を削除すること。
2		仮名化	直接個人を特定可能な属性またはその組み合わせ（氏名・生年月日）を符号や番号等に置き換えること。例えば、ハッシュ関数。
3	属性情報の一般化	一般化	<ul style="list-style-type: none"> ・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。 ・データ全体に行うものをGlobal Recoding、局所的に行うものをLocal Recodingと呼ぶ。 ・四捨五入や二捨三入などを丸め法（Rounding）と呼ぶ。
4		あいまい化	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。例えば、100歳以上の人は「100歳以上」とする。
5	属性情報の可能技法 ※ 原文ママ	マイクロアグリゲーション	元データをグループ化した後、同じグループのレコードの各属性値を、グループの代表値に置き換えること。
6		ノイズ（誤差）の付加	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。
7		データ交換	カテゴリー属性に対して、レコード間で属性値を（確率的に）入れ替えること。
8		疑似データ作成	元のデータと統計的に疑似させる人工的な合成データを作成すること。
9	その他技法	レコード（行）削除	特に大きい等、特殊な属性（値）を持つレコードを削除する。例えば、120歳以上のレコードは削除する。
10		セル削除	センシティブな属性値等、分析に用いるべきでない属性値を削除する。
11		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。

（出所） 高度情報通信ネットワーク社会推進戦略本部（IT総合戦略本部） パーソナルデータに関する検討会 技術検討ワーキンググループ報告書（2013年）

2. これまでの議論（施すべき匿名加工技法に係る主な意見）

<第3回技術検証WG（技術的課題）（令和5年1月10日開催）の議事要旨>

- ・ 悪意のある利用者への対策として、サンプリングは、全てのレコードが含まれないことのメッセージにもなるため、有効である。
- ・ 住所情報は、外観識別性が高いため、秘匿処理が必要と考えられる。
- ・ まずは、サンプリングの割合と住所情報の加工の方針を固めたうえで、他にも適用すべき匿名加工技法を検討していく必要がある。

<第3回有識者検討会（令和5年5月29日開催）の議事要旨>

- ・ 利用者の研究目的によってニーズは異なるため、例えば特異値が想定されない研究においては、最初から特異値が含まれるレコードを除いた範囲からサンプリングを行うなどの方法も考えられる。
- ・ 全数ではなく標本抽出することはやむを得ない。標本抽出の検討に際しては、抽出割合よりも標本サイズの方が重要である。
- ・ 匿名加工については、研究者側のニーズも踏まえながら、加工の程度を考える必要がある。
- ・ 年齢と性別と地域の情報があると、対象が絞られる可能性が高くなるため、いずれかは提供しない又は情報の粒度を粗くする必要がある。
- ・ 学術研究の目的のために提供するのであれば、その学術研究の範囲について検討する必要がある。その検討に当たっては「何が学術研究に含まれるか」よりも「誰が行う学術研究を対象とするか」がより重要な論点ではないかと考えられる。

2. これまでの議論（現在の方向性）

<データ提供の整備方針>

- 貸出方式によりデータを提供する。
- 対象データについては、パーソナルデータから優先する。
- 匿名加工については、サンプリングや住所情報の加工方針から優先的に検討する。
- 個人情報保護法における法的位置づけを踏まえ、利用目的は学術研究に限定する。

検討事項	検討結果
①提供形態	<ul style="list-style-type: none">・ <u>閲覧方式・貸出方式</u>による提供可能性を検証。<u>利用者利便、早期実現の観点</u>を考慮すると、まずは必要な対策を講じた上で<u>貸出方式での提供開始</u>を検討（閲覧方式の可能性も引き続き検討。）。➢ 閲覧：利用者の利便性、必要となる体制整備等が課題。➢ 貸出：リスクを低減するための方策（利用者の範囲・利用目的の限定や匿名加工の度合い）や不適切利用時のペナルティ等については要検討。
②データ項目	<ul style="list-style-type: none">・ <u>ビジネスデータは、公開情報が多い</u>。→ まずは、<u>パーソナルデータから優先</u>して検討。ビジネスデータの提供可能性は、提供形態も含めて、引き続き検討を進める。→ 提供データはあらかじめ固定（<u>データセット固定方式</u>）するが、今後、提供するデータ項目の範囲が拡大する場合はオーダーメイド方式の可能性も検討する。
③匿名加工技法	<ul style="list-style-type: none">・ 貸出方式であることを踏まえると、<u>より保守的な匿名加工</u>が必要と考えられる。➢ <u>サンプリングは必須</u>としつつ、適切なサンプリング割合・レコード数について検討。➢ <u>個人識別性が高くなる住所情報</u>の加工方針については、必要性も踏まえて、特に保守的に検討。➢ その他の技法（<u>削除、一般化、トップコーディング等</u>）の適用については、データの性質を踏まえて詳細を検討。
④個人情報保護法 （利用者の範囲や利用目的）	<ul style="list-style-type: none">・ パーソナルデータを前提とした場合の個人情報保護法における法的位置づけについては、<u>学術研究を目的とした保有個人情報の提供</u>（個情法69②四）として、あくまでも<u>学術研究に限定</u>するものとする。

※ 上記を踏まえ、①データ提供に係るガイドライン・利用規約類の策定、②詳細な匿名加工手法の検討も進めている。

3. 本日も検討いただきたい内容

<施すべき匿名加工技法について>

下記について、学術的価値の保全及びデータの安全性の確保の観点から講ずべき匿名加工技法について、どのような措置が必要と考えられるか

- ✓ 識別情報（名前や整理番号など）
- ✓ サンプルング
- ✓ 識別性の高い情報（住所情報など）
- ✓ その他の情報（所得金額、所得控除、申告納税額など）

<その他>

- ✓ パーソナルデータの範囲（所得税の一表と三表）の妥当性
- ✓ その他、匿名化を施すうえで必要な技術的課題

4. 今後のスケジュール

- 令和5事務年度（令和5年7月～）においては、データ提供に向けた準備（提供データの整備や、ガイドライン・利用規約類の策定）を本格化させることとし、令和6年度中に、準備が整い次第、対外的に行政記録情報の提供を開始することを目指す。
- 各WGにおける検証も踏まえつつ、提供するデータ、方式及び場所に関しては、有識者検討会において議論の上、決定する。

