

資料

保存期間：5年
(令和10事務年度末)
令和6年4月9日

第8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8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8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（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">・ 閲覧方式・貸出方式による提供可能性を検証。利用者利便、早期実現の観点を考慮すると、まずは必要な対策を講じた上で貸出方式での提供開始を検討（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">・ ビジネスデータは、公開情報が多い。→ まずは、パーソナルデータから優先して検討。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（データセット固定方式）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、より保守的な匿名加工が必要と考えられる。➢ サンプリングは必須としつつ、適切なサンプリング割合・レコード数について検討。➢ 個人識別性が高くなる住所情報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（削除、一般化、トップコーディング等）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、学術研究を目的とした保有個人情報の提供（個情法69②四）として、あくまでも学術研究に限定するものとする。

※ 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。

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7回技術検証WG（技術的課題）（令和5年11月28日開催）の議事要旨>

○ サンプリング

- ・ 1%は妥当な数字。1%（約23万レコード）もあれば、実証分析には十分。
- ・ サンプリング方法は、安全性の観点からは非復元抽出より復元抽出の方が好ましいが、利用者のニーズも考慮する必要がある。

○ 識別性の高い情報

- ・ 地域情報は外観識別性が高いため、加工方法は特に注意する必要がある。詳細な地方区分での提供は安全性の観点から困難だろう。一方で、地域によっては、細やかな区分による提供が可能な場合も考えられ、他の情報の提供との組合せを踏まえた、より詳細な検討が必要。
- ・ 業種情報は、分析する上で重要な情報であると考えられるため、分類区分を統合する場合には、慎重な検討が必要であろう。

○ その他の情報

- ・ 定量情報を1円単位で提供することはリスクが高まるため、ラウンディングやグルーピングは行うべきであろう。
- ・ 利用者にとっては、匿名データに攪乱的手法が使われることは好ましくない可能性があるため、まずは攪乱的手法を使わない加工方法を検討してはどうか。

3. 本日まで検討いただきたい内容

<施すべき匿名加工技法について>

以下の情報に係る、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、第7回WGでの検討を踏まえた検証結果を、どのように考えるか。

- ✓ 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>

- ✓ 提供対象年分
- ✓ サンプルング方法
- ✓ サンプルデータ
- ✓ 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。

