

資料

保存期間：5年
(令和9事務年度末)
令和5年5月29日

第3回 国税庁保有行政記録情報の 整備に関する有識者検討会

国税庁 企画課

資料内容

1. これまでの議論

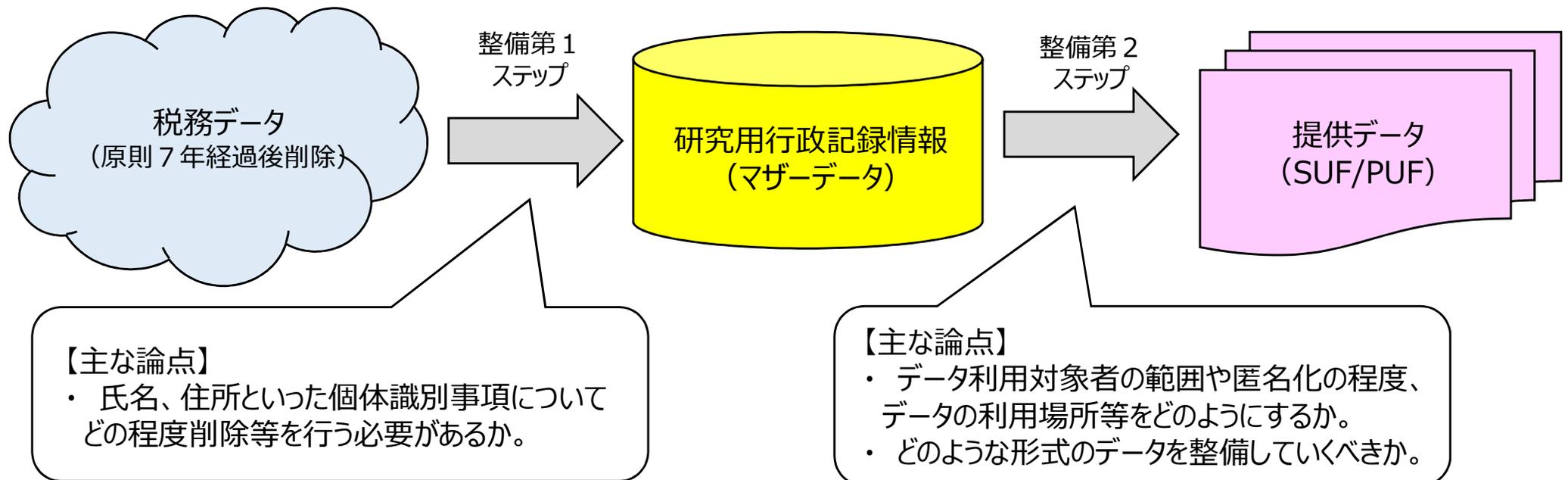
2. 今後の方向性(案)

3. 本有識者検討会と各種WGの位置付け

4. 今後のスケジュール

1. これまでの議論（整備ステップ）

- 国税庁がシステム内で保有する税務データは、現状、原則 7 年経過後に削除することとしている。
- 令和 3 事務年度においては、提供データ（SUF/PUF※）の整備に先立って、長期間保存が可能となる、研究用行政記録情報（マザーデータ）を整備するに当たっての検討を進めてきたところ（整備第 1 ステップ）
 - （※） SUF : Scientific Use File、学術研究用ファイル、PUF : Public Use File、一般公開型ファイル
- 令和 4 事務年度においては、より具体的なデータ提供に向けて、どのような提供データを整備するか議論を進めているところ（整備第 2 ステップ）



1. これまでの議論（データの提供形態について）

- 提供形態として、①データ貸出（CD-R等の媒体にデータを格納して貸出し、使用后返却）、②データ閲覧（国税当局の施設に来訪し閲覧・利用）のいずれかが考えられる。
- コンプライアンスリスクに応じて、例えば、ガイドライン・利用規約における制限や、税目によっては提供形態を限定する等の対応も考えられる。

	データ貸出方式	データ閲覧方式
利用者の利便性	高い	低い (利用者は国税当局の施設に往訪する必要)
国税当局側の負担	比較的低い (閲覧場所等の整備は不要、貸出作業は発生)	高い (閲覧場所を整備するなどの対応が必要)
受入可能件数	広く受け入れることが比較的可能 (データの貸出事務のみが発生)	広く受け入れることは困難 (閲覧場所のスケジュール管理等が必要)
コンプライアンスリスク	高い	低い

1. これまでの議論（データ項目抜粋）

個人課税関連	法人課税関連
確定申告書	確定申告書
青色申告決算書・収支内訳書	法人税申告書別表ファイル
各種届出書	財務諸表（貸借対照表）
個人事業者の消費税申告書	財務諸表（損益計算書）
資産課税関連	連結グループ情報
相続税申告情報	各種届出書
贈与税申告情報	個人事業者の消費税申告書

1. これまでの議論（匿名加工の技法について）

- 非識別化の手法は、以下の表のとおり、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）

1. これまでの議論（サンプルデータ）

- サンプルデータの役割については、①データ提供に際して、研究者等が事前にデータ構造を理解することにより、広く利用されるきっかけを提供すること、②大学等におけるデータ分析等の教育用途としても利用可能であり、③将来的には匿名化データの匿名化のノウハウを蓄積する観点から、サンプルデータを提供することとしてはどうか。
- サンプルデータの整備に当たっては、サンプルデータの役割（特に上記①）や実現可能性を考慮し、まずは、乱数を発生させるなどして作成する方法による疑似データを整備する方向性としてはどうか。
- サンプルデータの公表タイミングについては、データ提供の実現時期を考慮する必要がある。

論点	疑似データの特徴
想定される利用目的	・データ分析等の教育目的 ・共同研究、匿名化データ利用への準備
税務データとの関連	なし（※）
研究分析における利用可能性	疑似データであり、論文への引用は不可
保持できる変数 (収入・所得・控除項目等)	制限なし
実現可能性	比較的容易

（※）疑似データの作成にあっても、一定程度税務データと所得分布等が一致していることが求められるか。

1. これまでの議論（第2回技術検証WG）

＜第2回技術検証WG（令和4年10月14日開催）の議事要旨＞

○ データの提供形態について

- ・ 準備に時間がかかって、提供開始が遅れるよりは、まずは出来ることから取り掛かり、徐々に拡大していくべきである。
- ・ リモートエグゼキューションは、送付したプログラムの結果がエラーとなる可能性もあり、実効性が低いと考えられる。

○ 提供データの項目について

- ・ 他の統計情報から観察できない項目から優先して提供してはどうか。
- ・ 税制の研究においては、住所情報は市町村レベルまでは提供されることが研究の正確性を確保するためには望ましい。

○ サンプルデータについて

- ・ データ利用者がプログラムの正確性を確認する材料としては有用である。また、データ分析のニーズも高まっているため、自由な形で使えるデータがある方が望ましい。
- ・ 実際の税務データと関係のない架空のデータが想定されることから、その作成にあたっては、あまりコストや手間をかけるべきものではないように思われる。

○ データを利用できる者・利用目的の範囲について

- ・ リサーチアシスタントにも利用を認めるべきだが、利用者なのか研究協力者なのかによって、データへのアクセスについて整理する必要がある。
- ・ 利用目的の範囲は「税・財政施策に資すること」よりも広い範囲で認めてもよいのではないか。

1. これまでの議論（第3回技術検証WG）

<第3回技術検証WG（令和5年1月10日開催）の議事要旨>

○ データの提供形態について

- ・ 閲覧方式の分析結果持ち出しの安全性審査については、公的統計における調査票情報の利用において、分析結果等の提供に関する標準的なチェック内容が公表されており、審査基準策定に際しての参考となる。
- ・ 貸出方式の場合は、不適切利用が万が一で発生する場合も想定して制度設計や匿名加工を行う必要がある。
- ・ 貸出方式であっても、利用者の範囲を限定することで、コンプライアンスリスクを下げることはできる。

○ 施すべき匿名加工技法について

- ・ パーソナルデータとビジネスデータは、性質が異なるので、それぞれの特徴を踏まえて施すべき技法を検討すべきである。
- ・ ビジネスデータについては、公開情報が多いため、それらも準識別子となり得る。特に貸出方式の場合は、パーソナルデータから議論した方がよいのではないか。その際は個人情報保護法の観点からも検討する必要がある。
- ・ 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。

○ サンプルデータについて

- ・ 研究にも利用できるようにするならば、サンプリングしたレコードのアイデンティティに関する情報（生年月日や性別等）を、他の近似するレコードのそれとスワッピングして、税務データそのものの金額情報は加工しない方法により、サンプルデータを生成することも考えられる。
- ・ サンプルデータは、疑似データや合成データとして捉えられるだけでなく、一般公開可能な点を勘案すれば、作成の仕方によってはPublic Use File（PUF）と位置付けられうる。

1. これまでの議論（第4回技術検証WG）

＜第4回技術検証WG（令和5年3月14日開催）の議事要旨＞

- 匿名加工したデータを、個人情報保護法上の行政機関等匿名加工情報として外部提供するためには、加工元の情報が情報公開法上の開示請求に対して全部又は一部開示できる必要がある。そのため、提供データの加工元であるマザーデータが、情報公開法上の開示請求に対して不開示になるのであれば、提供データを行政機関等匿名加工情報として取扱うことは困難ではないかと考えられる。
- 行政機関等匿名加工情報としての外部提供でなくても、個人情報保護法上、行政機関等は、学術研究の目的のためであれば、保有個人情報を提供することは可能である。
- 個人情報保護法上は提供が可能であっても、行政機関側の守秘義務の観点からも検討が必要となると考えられる。この場合、技術的な匿名加工の要件だけでなく、データ提供に関する法的規律についても更なる検討の必要があると考えられる。

2. 今後の方向性（案）

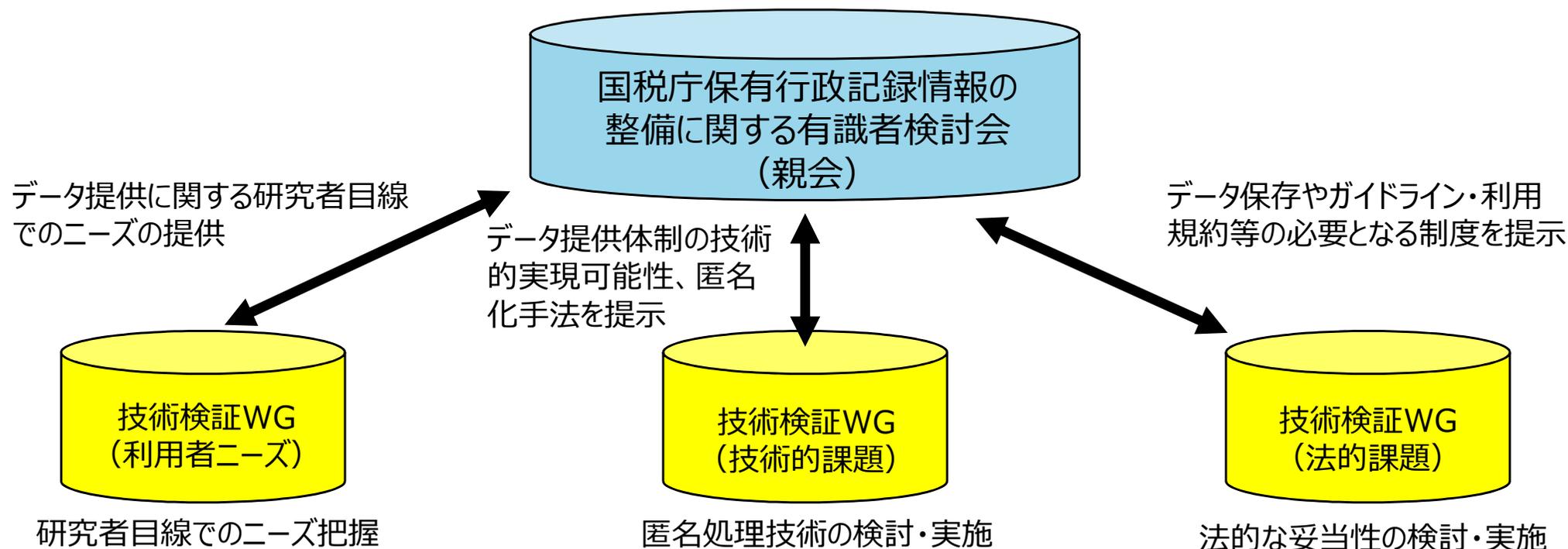
<WGでの議論を踏まえた、データ提供の整備方針（案）>

- まずは、**貸出方式**による提供形態から開始し、データ提供の対象は**パーソナルデータから優先**することとしてはどうか。また、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。
- 今後、**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**の検討を進める。

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うえ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。）<ul style="list-style-type: none">➢ 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。<ul style="list-style-type: none">→ 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匿名加工手法	<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。<ul style="list-style-type: none">➢ 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。

3. 本有識者検討会と各種WGの位置付け

- 国税庁保有行政記録情報の整備に関する有識者検討会は、統計学、経済学、法律の各専門家から構成され、全体の方向性を検討することを主な役割とする。
- 技術検証WGは、データ提供に関する研究者目線でのニーズを把握するための利用者ニーズの把握を目的としたもの、そのうえで匿名化を施すうえでの技術的課題の検証を目的としたもの、さらに、議論の進展に応じて、データ利用に際しての法的規律を検討する法的課題の検証を目的としたものの開催を検討する。なお、WGの検討内容は有識者検討会へ報告する。
- 前述の①データ提供に係るガイドラインや利用規約類は、主に利用者ニーズ・法的課題に係るWG、②詳細な匿名加工手法の検討は主に技術的課題に係るWGにおいて、更に詳細な検討を進める。



4. 今後のスケジュール（案）

- 本日の有識者検討会においてデータ提供の整備方針（案）を議論したうえで、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。

