

第 4 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 企画課データ活用推進室

資料内容

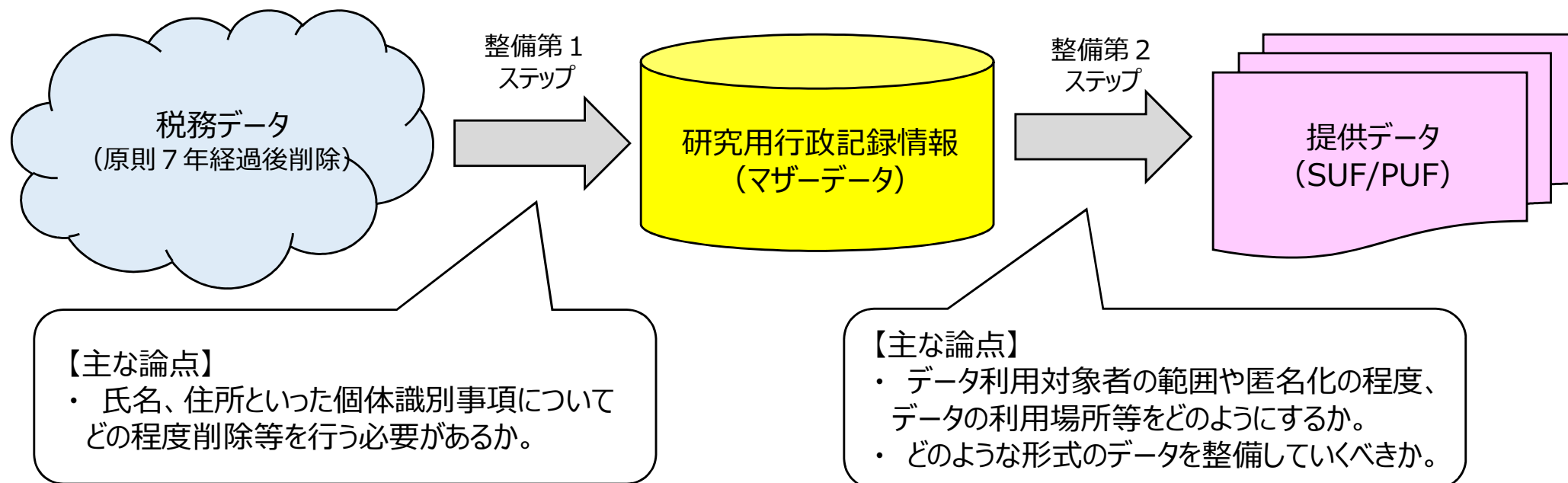
1. これまでの議論

2. 今後の方向性(案)

3. 今後のスケジュール

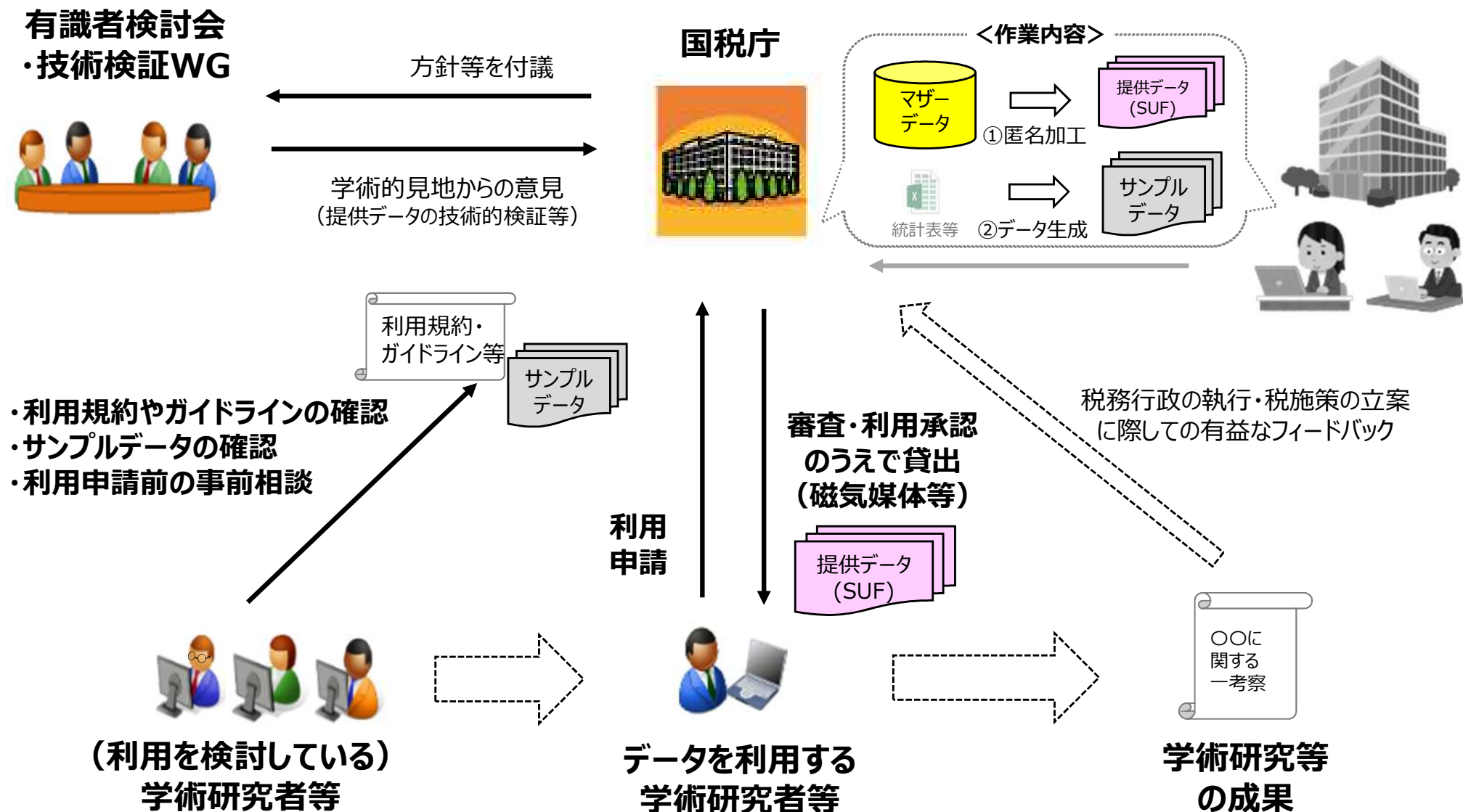
1. これまでの議論（整備ステップ）

- 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 事務年度以降においては、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（匿名データ提供に関する全体像（イメージ））

- 学術研究での利用を前提に、利用者は事前に利用規約等・サンプルデータを確認し、利用申請が承認されれば、**匿名データを貸し出す**ことにより、提供。
- データ作成については、**匿名加工の安全性確保を考慮**し、国税庁での内製に加え、外部事業者への委託も検討。



1. これまでの議論（データ提供の整備方針）

<データ提供の整備方針>

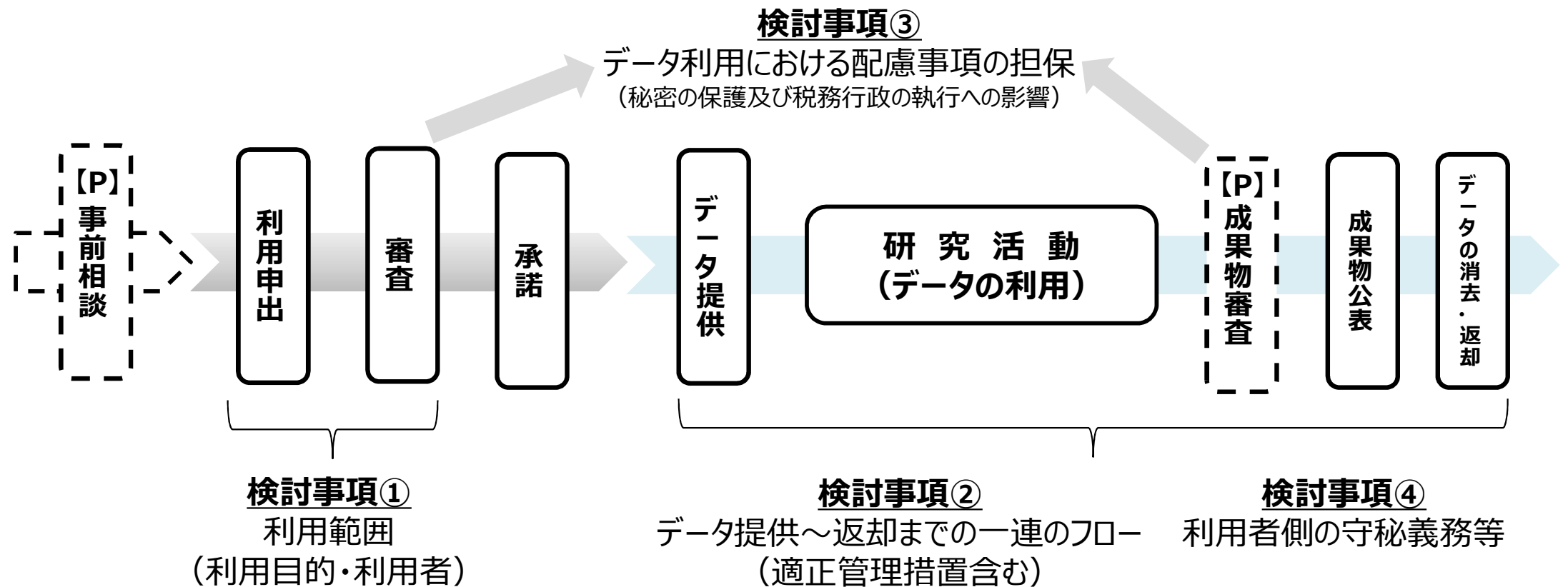
- 貸出方式によりデータを提供する。
- 対象データについては、パーソナルデータから優先する。
- 匿名加工については、サンプリングや住所情報の加工方針から優先的に検討する。
- 個人情報保護法における法的位置づけを踏まえ、利用目的は学術研究に限定する。

検討事項	検討結果
①提供形態	・ <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	・ <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	・ 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）	・ 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。

1. これまでの議論（データの提供方法・法的課題の整理）

- データの提供方法について、税務大学校との共同研究や独立行政法人統計センターによる公的統計の提供方法を参考に、①利用者範囲（利用者・利用目的の範囲）、②データ提供から返却までの一連のフロー（データの適正管理措置）、③利用申出時と研究成果物の審査、④利用規約等で利用者に課すべき守秘義務等を議論（第5回WG）。



- また、データ提供に係るガイドラインや利用規約類の検討において必要な法的課題（情報公開法と守秘義務）を検証し、ガイドライン・利用規約（素案）を議論（第6回WG）。

1. これまでの議論（匿名加工技法）

- 提供データの識別情報（名前や整理番号など）、サンプリング、識別性の高い情報（住所情報など）、その他の情報（課税所得金額等、所得控除、申告納税額など）について、学術的価値の保存及びデータの安全性の観点から、施すべき匿名加工技法を検討（第7回WG及び第8回WG）。

税務データ

匿名加工

提供データ

整理番号	名前	住所	生年月日	性別	業種	課税所得金額等	所得控除	申告納税額
×××	国税太郎	〇〇〇〇	×年×月	M	〇〇業	××××	××××	××××
×××	国税花子	〇〇〇〇	×年×月	F	〇〇業	××××	××××	××××
×××	国税次郎	〇〇〇〇	×年×月	M	〇〇業	××××	××××	××××
×××	国税三郎	〇〇〇〇	×年×月	M	〇〇業	××××	××××	××××

<匿名加工(イメージ)>

整理番号	名前	住所	生年月日	性別	業種	課税所得金額等	所得控除	申告納税額
×××	国税太郎	グルーピング等によって保守的に加工				× トップコーディングやラウンディングによって加工 ×		
×××	国税花子							
×××	国税次郎	サンプリングの対象外(未抽出)						
×××	国税三郎							

1. これまでの議論（第5回技術検証WG（利用者ニーズ））

<第5回技術検証WG（令和5年8月23日開催）の議事要旨>

○ 利用範囲について

- ・ 常勤の研究者だけではなく、リサーチアシスタントとしてポストドクターや大学院生にも利用を認めることが望ましい。
- ・ リサーチアシスタントについては、例えば、常勤の研究者の研究室の中でしかデータを利用できないとすることや、常勤の研究者が研究室にいて、常に監督できるような状況でなければ利用できないという厳しい制約を設けることが考えられる。

○ データ提供から返却までの一連のフローについて

- ・ 匿名データを貸出方式で利用することになるため、適切な適正管理措置が求められることは強調しておく必要がある。
- ・ 外部ネットワークから遮断された環境でデータを使わなければならない制限は使い勝手が悪い。最初に利用のハードルを上げて、誰も使わないというよりは、まずは使ってもらうことが重要で、現代的なテクノロジーに合わせて、データの利用中の管理方法を定めてはどうか。

○ 利用申出の審査、研究成果物の審査について

- ・ 利用申出の段階で、税務行政の執行への影響に配慮することや、秘密の保護について利用者に説明し、それらを遵守してデータを使うことを誓約書等で確認することが考えられる。

○ その他の論点について

- ・ 税務大学校との共同研究のガイドラインで求められる守秘義務や、統計法上の匿名データの提供の取組が、匿名化された税務データのガイドラインを作成する上で参考になる。

1. これまでの議論（第6回技術検証WG（法的課題））

＜第6回技術検証WG（令和5年10月3日開催）の議事要旨＞

○ 情報公開法について

- ・ 提供データが行政機関の保有する情報の公開に関する法律第5条1号及び2号の不開示情報に該当するかは、提供データの匿名加工の内容や程度により判断すべきであるが、パーソナルデータについてはかなりの匿名加工がされた粗い情報以外は不開示情報に該当するだろう。
- ・ 匿名データを開示することにより、匿名データ提供の事務の性質上、その適正な遂行に支障を及ぼすおそれがあることが明確に整理されるのであれば、同法第5条6号の不開示情報に該当することになる。

○ 守秘義務について

- ・ 匿名加工した後においても実質秘に該当する可能性があるデータを提供する場合に、その提供行為が国家公務員法及び国税通則法の守秘義務に抵触するかは、当該提供について正当な事由が存在すると言えるかどうかが論点となるだろう。

○ その他の論点について

- ・ 利用者が提供データを漏えいした場合には、法律の適用があるわけではないので、利用者の違反行為が国税庁との契約不履行になるか否かに尽きるだろう。

1. これまでの議論（第7回技術検証WG（技術的課題））

＜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み合わせを踏まえた、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。

1. これまでの議論（第8回技術検証WG（技術的課題））

<第8回技術検証WG（技術的課題）（令和6年4月9日開催）の議事要旨>

○ 識別性の高い情報（定性情報）

- ・ 識別性の高い情報を全て提供する場合は特定リスクが高まるため、いずれかの情報は削除する必要があり、その場合、性別を削除する方法が最も合理的という事務局案に賛成。

○ その他の情報（定量情報）

- ・ 出現率の低い項目については、レコードごと削除されても仕方ないだろう。
- ・ 内訳の金額をグルーピングしても、合計額から推計できるという点は、そこまで慎重になる必要は無いのではないか。その数値や分布から個人が特定されるか否かが問題で、ケースバイケースで考えるしかない。
- ・ 内訳の金額から合計額が推計でき、合計額のトップコーディングが機能しないという点は、内訳の金額を%表示にすれば問題は解決できるが、一方で、実際の税務データは%では保持されていないため、利用者がどのようなイメージを持つか気になる。その点、トップコーデコーディングをした合計額のしきい値から各内訳の金額を割り返すという方法が一案として考えられる。

○ その他（提供対象年分、サンプリング方法）

- ・ 古い年分から提供するという事務局案に賛成。
- ・ サンプリング方法は、最終的には実用性の観点から決めてはどうか。
- ・ サンプルデータは、まず利用者がデータ構造を理解するためのメタデータとして提供するものの、教育目的で使用することも可能ではないか。

2. 今後の方向性（ガイドライン・利用規約類の策定方針（案））

＜ガイドライン・利用規約類の策定方針（案）＞

- 国税庁にとっての先例である税務大学校との共同研究におけるガイドラインや、総務省の独立行政法人統計センターによる公的統計の匿名データ利用の手引等を参考にして策定。

これまでの検討事項	WGでの検討結果
①利用者範囲	・ <u>申出者は、大学や公的機関等に所属する常勤の研究者に限定</u> し、当該研究者の指導監督の下、博士研究員、大学院生（ただし、博士課程に限る）も補助者として利用可能。 <u>利用目的は税・財政施策の改善・充実に係る統計的研究に限定。</u>
②適正管理措置	・ <u>研究用匿名データの利用場所</u> （施錠可能な物理的な場所（国内）に限定するなど）や <u>研究用匿名データの利用可能な端末や環境を制限</u> （外部ネットワークから遮断された環境であり、本人認証等の不正操作対策を義務化など）。
③利用申出の審査、研究成果物の審査	・利用者が国税庁に <u>利用の申出を行う際、国税庁において配慮事項（特に公序良俗に反しないか）を審査し、成果物の公表前にも、国税庁で配慮事項（税務データに関する秘密の保護や公序良俗に反しないか）を審査。</u>
④利用者に対する守秘義務の負担	・利用者に研究用匿名データを利用させる際、当該データの開示や漏えい、目的外の利用が無いよう、 <u>利用者に誓約書を提出させ、守秘義務を負わせる。</u> 情報漏えいなど <u>契約違反があった場合の措置も併せて規定</u> （違反者及び代表者等の氏名等の公表、今後の研究用匿名データの利用禁止など）。
⑤情報公開法 （不開示情報の該当性）	・ <u>①「利用目的」や②「適正管理措置」を講じないまま公にした場合、提供事務の遂行に支障を及ぼすおそれがある（情公法5六）</u> ことから、 <u>不開示情報に該当。</u>

⇒ 上記の策定方針及びWGでの検討結果を踏まえ、ガイドライン・利用規約類を策定

2. 今後の方向性（匿名加工技法の方向性（案））

匿名加工技法の方向性

- 提供データは、学術的価値を有する情報である必要がある一方で、データ提供の安全性を考慮し、他の情報と照合しても個人等が特定されない匿名化処理を行った情報（※）とする。

※社会通念上、一般人の判断力や理解力をもって、具体的な人物と情報との間に同一性を認められない程度に匿名加工された情報。

⇒サンプリングは必須としつつ、個人識別性が高くなる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。その他の項目については、データの性質を踏まえて、検討。

⇒最終的には、データの安全性を評価するとともに、利用者のニーズも考慮のうえ、総合的に適否を判断する。



WGでの検討結果（匿名加工技法（案））

○サンプリング：申告書（所得税の一表と三表）データ全件ではなく、その一部(約1%)を抽出して提供。

○名前、マイナンバー、整理番号：個人を直接的に特定できる情報であるため、削除。

○住所、性別、業種、生年月日：情報の組合せによっては、個人を特定される可能性が排除できないため、グルーピング等によって、保守的に匿名加工。

○その他の情報（所得や税額）：出現頻度が低い情報のレコード削除、
極端に大きな値のトップコーディング、
ラウンディング（数字を丸める） 等

⇒ **上記の方向性及びWGでの検討結果を踏まえ、提供データを作成**

3. 今後のスケジュール（案）

- 本日の有識者検討会において、ガイドライン・利用規約類の策定方針（案）及び匿名加工技法の方向性（案）を議論したうえで、令和6事務年度（令和6年7月～）においては、実際に提供するデータの技術的検証を行い、有識者検討会において議論した上で、提供準備が整い次第、対外的に行政記録情報の提供を開始する。
- データの外部提供後は、当面は提供体制の整備・充実化を図ることとし、その後、提供対象税目（所得税以外の税目）の拡大等も併せて検討する。

