

資料

保存期間：5年
(令和9事務年度末)
令和5年3月14日

第4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

資料内容

1. 本ワーキンググループの経緯・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経緯・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）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

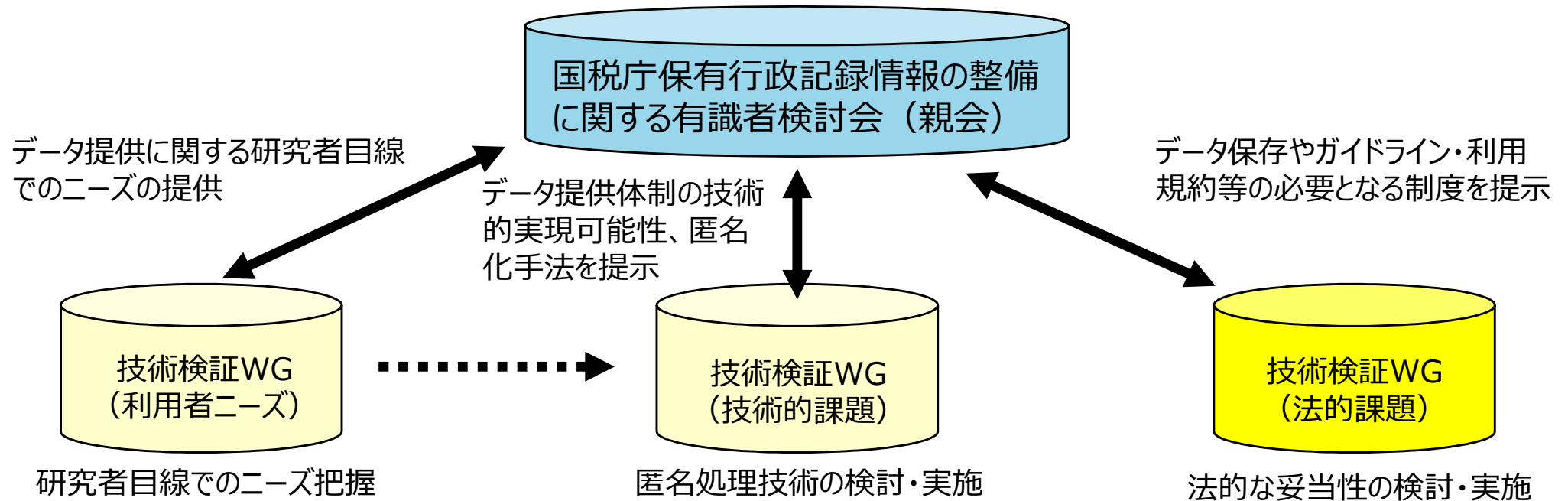
- (2) 座長は必要があると認めるときは、検討会にワーキンググループを置くことができる。
なお、ワーキンググループにおける検討結果は、有識者検討会に報告するものとする。

- 第4回WGでは、主に個人情報保護法制における匿名データ提供の法的位置づけや、匿名加工の基準等について、検討を行う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）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」に対して検討状況を適宜報告することとする。
- 第4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
日置 巴美	弁護士法人三浦法律事務所 弁護士

1. 本ワーキンググループの経緯・位置づけ

- 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。



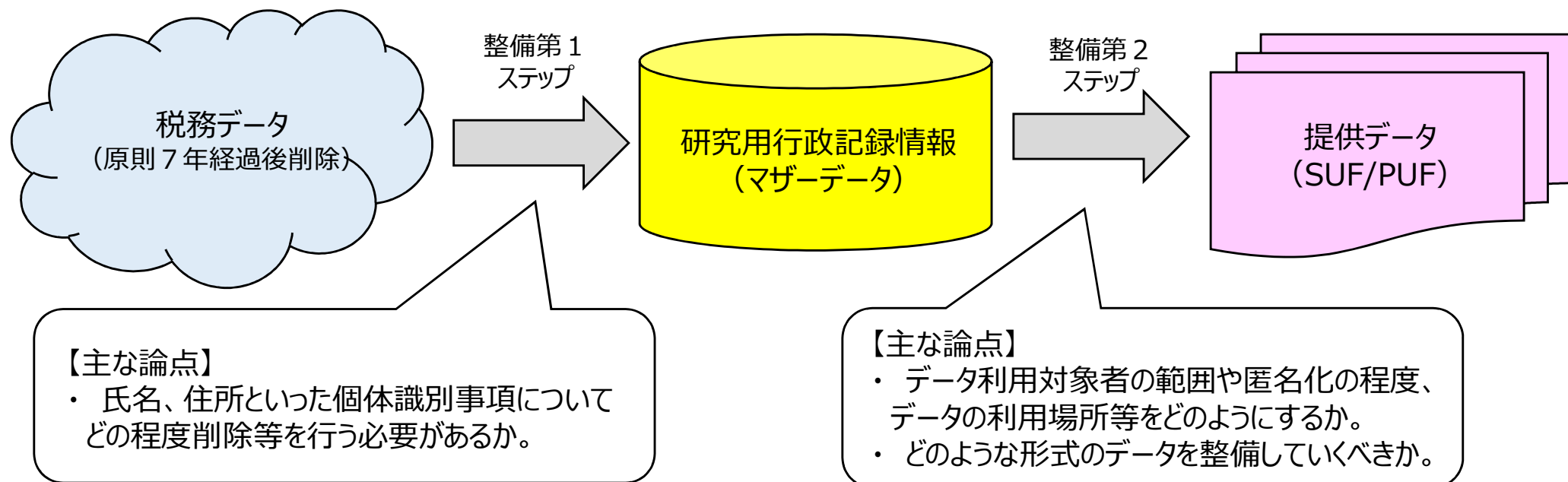
2. これまでの議論（匿名データの保有及び公表の目的等）

● 匿名データの保有及び公表の背景

- 国税庁保有行政記録情報を用いた税務大・大東大との共同研究（以下、共同研究）は、各府省庁が保有するデータは、公開することが適当でない情報であっても、限定的な関係者間での共有を図る「限定公開」とする「オープンデータ基本指針」を踏まえ、国税庁独自に有識者を交え検討を重ねた結果、まずは共同研究という形式から始めることが適切であるという結論が得られた。
- 国税庁の税務データは、申告納税制度の下、納税者の信頼や協力によって集積しているものであることに留意し、適切に取り扱う必要がある。したがって、共同研究において個票データを利用する者は、守秘義務の観点から国家公務員の身分を有する者のみに限定する。
- 一方で、国家公務員の身分を有することなく、かつ、より多くの研究者が税務データを分析することにもニーズがある。
- 現状、共同研究において、分析結果等利用者は国家公務員の身分を有することなく、加工した税務データにアクセス可能であるが、このスキームを参考に、研究者等が加工したデータにアクセスできる仕組み（国税庁版SUF（Scientific Use Files、学術研究用ファイル））の可能性を検討する。

2. 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）



2. これまでの議論（第1回技術検証WG）

<第1回技術検証WG（令和4年3月24日開催）の議事要旨>

○ マザーデータの「個人情報」該当性について

- ・ マザーデータとしてデータを保存する際に、氏名等の情報を削除する等の措置を講じる点については、コンプライアンスリスクを低減させ、保守的にマザーデータを運用する観点で好ましい。
- ・ この場合、マザーデータ内の情報だけで「特定の個人」を識別することが困難であっても、税務データとマザーデータにおいて共通のIDが使用されている場合、容易照合性があるものとして個人情報に該当する。

○ マザーデータに対する本人開示請求等への対応

- ・ マザーデータが個人情報として整理される場合、個人情報保護法等に基づく開示請求等があった際は、個人を特定できる範囲において対応することが求められる。

○ 情報公開法上の開示請求への対応

- ・ 氏名などにつき全部または一部を削除しているとはいえ、マザーデータ内の情報から図らずも個人が特定される場合、情報公開請求に対しては不開示となる。

2. 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。

2. 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。

2. 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

2. 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
贈与申告情報	個人事業者の消費税申告書

2. 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。

2. 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）

3. 本日までご検討いただきたい内容

① 匿名データ提供に関する法的位置づけ等について

(ご検討に際してのポイント)

- 個人情報保護法上の法的位置づけ等

② 匿名加工技法について

(ご検討に際してのポイント)

- 個人情報保護法に定められた加工基準

(参考) 行政機関等匿名加工情報

- 行政機関等匿名加工情報とは、行政機関・独立行政法人等が保有する個人情報の特定の個人を識別することができないように加工し、かつ、当該個人情報を復元できないようにした情報。
- 行政機関等匿名加工情報の作成方法の基準を個人情報保護委員会規則で定める。



- 行政機関等は、行政機関等匿名加工情報の利用に関する提案募集のため、次の3要件を満たす個人情報ファイルを公表。
 - ① 個人情報ファイル簿に掲載された個人情報ファイル（個人情報ファイル簿として公表されるもの）
 - ② 情報公開請求があれば全部又は一部開示されるもの
 - ③ 行政の適正かつ円滑な運営に支障のない範囲内で加工が可能なもの
- 事業者等から提案があった場合には、これを審査の上、行政機関等匿名加工情報を提供。



(参考) 行政機関等匿名加工情報の利用に関する提案からその提供までの主なフロー

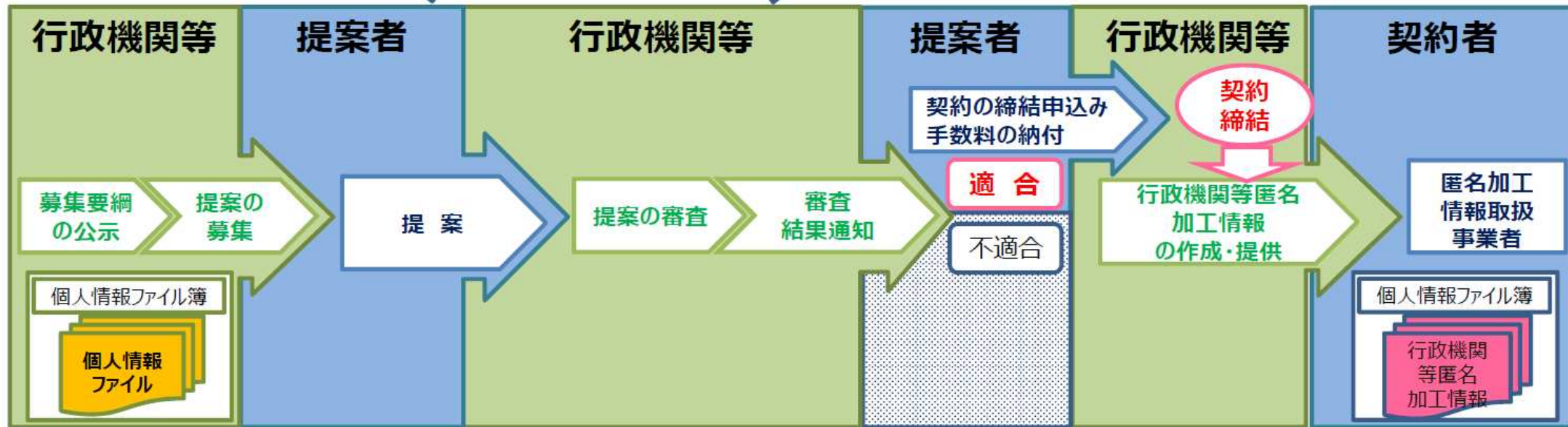
- ▶ 毎年度 1 回以上、30 日以上の期間を定めて、提案の募集を実施
- ▶ 提案の募集前に、行政機関等のウェブ等で募集要綱を公示
- ▶ 提案に必要な一定の書類の様式はウェブ等で入手可

- ▶ 提案者は、行政機関等匿名加工情報を事業の用に供しようとする者であれば、個人・法人は問わない
- ▶ 未成年者等や禁固刑等の刑に処せられてから 2 年を経過しない者等一定の欠格事由に該当する者は提案不可
- ▶ 提案前の事前相談可

- ▶ 審査基準の適合性審査
 - ① 欠格事由の該当の有無
 - ② 一定の加工基準に合致
 - ③ 事業が新産業の創出等に資すること
 - ④ 漏えい防止等の安全管理措置が適切であること等
- ▶ 審査結果は個別に通知

- ▶ 審査基準に適合した者には、審査結果通知書に契約の締結の申込書・契約書を同封
- ▶ 手数料を納付し、上記書類に必要事項を記入提出することで契約可

- ▶ 契約の締結後、行政機関等が行政機関等匿名加工情報を作成・提供
- ▶ 利用目的の範囲で事業の用に供することができる



(参考) 行政機関等匿名加工情報の審査基準及び加工基準

○ 審査基準

- ① 欠格事由に該当しないこと。
- ② 行政機関等匿名加工情報の本人の数が**1000人以上**かつ個人情報ファイルを構成する保有個人情報の**本人の数以下**であること。
- ③ 特定の個人を識別できず、また、保有個人情報を復元できないように以下の**加工基準に従い加工**すること。
- ④ 事業の目的及び内容が**新たな産業の創出又は活力ある経済社会若しくは豊かな国民生活の実現に資する**ものであること。
- ⑤ 行政機関等匿名加工情報の利用期間が、**利用の目的及び方法からみて必要な期間を超えない**こと。
- ⑥ 行政機関等匿名加工情報の**利用の目的及び方法並びに安全管理の措置**が当該行政機関等匿名加工情報の**本人の権利利益を保護するために適切なもの**であること。
- ⑦ 行政機関等匿名加工情報を作成する場合に、当該行政機関の**事務の遂行に著しい支障を及ぼさないもの**であること。

○ 加工基準

行政機関等匿名加工情報の作成方法に関して、次の措置を講ずることを求める。

- ① **特定の個人を識別することができる記述等**（例：氏名）の**全部又は一部を削除**（置換を含む。以下同じ。）
- ② **個人識別符号の全部を削除**。
- ③ 個人情報と他の情報とを**連結する符号**（例：委託先に渡すために分割したデータとひも付けるID）を**削除**。
- ④ **特異な記述等**（例：年齢116歳）を**削除**。
- ⑤ 上記のほか、個人情報と保有個人情報ファイル内の他の個人情報との差異等の**性質を勘案し、適切な措置を講ずる**。

事務対応ガイド（行政機関等向け）では、個人情報保護委員会規則で定められた行政機関等匿名加工情報の作成方法に関する上記の基準等について、具体的な事例等も交えて、分かりやすく示している。

(参考) 個人情報保護に関する法律 (抄)

(定義)

第六十条

3 この章において「行政機関等匿名加工情報」とは、次の各号のいずれにも該当する個人情報ファイルを構成する保有個人情報の全部又は一部(これらの一部に行政機関情報公開法第五条に規定する不開示情報(同条第一号に掲げる情報を除き、同条第二号ただし書に規定する情報を含む。)又は独立行政法人等情報公開法第五条に規定する不開示情報(同条第一号に掲げる情報を除き、同条第二号ただし書に規定する情報を含む。))が含まれているときは、これらの不開示情報に該当する部分を除く。)を加工して得られる匿名加工情報をいう。

一 第七十五条第二項各号のいずれかに該当するもの又は同条第三項の規定により同条第一項に規定する個人情報ファイル簿に掲載しないこととされるものでないこと。

二 行政機関情報公開法第三条に規定する行政機関の長又は独立行政法人等情報公開法第二条第一項に規定する独立行政法人等に対し、当該個人情報ファイルを構成する保有個人情報が記録されている行政文書等の開示の請求(行政機関情報公開法第三条又は独立行政法人等情報公開法第三条の規定による開示の請求をいう。)があったとしたならば、これらの者が次のいずれかを行うこととなるものであること。

イ 当該行政文書等に記録されている保有個人情報の全部又は一部を開示する旨の決定をすること。

ロ 行政機関情報公開法第十三条第一項若しくは第二項又は独立行政法人等情報公開法第十四条第一項若しくは第二項の規定により意見書の提出の機会を与えること。

三 行政機関等の事務及び事業の適正かつ円滑な運営に支障のない範囲内で、第百十四条第一項の基準に従い、当該個人情報ファイルを構成する保有個人情報を加工して匿名加工情報を作成することができるものであること。

(利用及び提供の制限)

第六十九条 行政機関の長等は、法令に基づく場合を除き、利用目的以外の目的のために保有個人情報を自ら利用し、又は提供してはならない。

2 前項の規定にかかわらず、行政機関の長等は、次の各号のいずれかに該当すると認めるときは、利用目的以外の目的のために保有個人情報を自ら利用し、又は提供することができる。ただし、保有個人情報を利用目的以外の目的のために自ら利用し、又は提供することによって、本人又は第三者の権利利益を不当に侵害するおそれがあると認められるときは、この限りでない。

一 本人の同意があるとき、又は本人に提供するとき。

二 行政機関等が法令の定める所掌事務又は業務の遂行に必要な限度で保有個人情報を内部で利用する場合であって、当該保有個人情報を利用することについて相当の理由があるとき。

三 他の行政機関、独立行政法人等、地方公共団体又は地方独立行政法人に保有個人情報を提供する場合において、保有個人情報の提供を受ける者が、法令の定める事務又は業務の遂行に必要な限度で提供に係る個人情報を利用し、かつ、当該個人情報を利用することについて相当の理由があるとき。

四 前三号に掲げる場合のほか、専ら統計の作成又は学術研究の目的のために保有個人情報を提供するとき、本人以外の者に提供することが明らかに本人の利益になるとき、その他保有個人情報を提供することについて特別の理由があるとき。

3 前項の規定は、保有個人情報の利用又は提供を制限する他の法令の規定の適用を妨げるものではない。

(参考) 個人情報保護に関する法律 (抄)

(行政機関等匿名加工情報の作成及び提供等)

第七十条 行政機関の長等は、この節の規定に従い、行政機関等匿名加工情報(行政機関等匿名加工情報ファイルを構成するものに限る。以下この節において同じ。)を作成することができる。

2 行政機関の長等は、次の各号のいずれかに該当する場合を除き、行政機関等匿名加工情報を提供してはならない。

一 法令に基づく場合(この節の規定に従う場合を含む。)

二 保有個人情報を利用目的のために第三者に提供することができる場合において、当該保有個人情報を加工して作成した行政機関等匿名加工情報を当該第三者に提供するとき。

3・4 (略)

(提案の募集)

第九十条 行政機関の長等は、個人情報保護委員会規則で定めるところにより、定期的に、当該行政機関の長等の属する行政機関等が保有している個人情報ファイル(個人情報ファイル簿に前条第一号に掲げる事項の記載があるものに限る。以下この節において同じ。)について、次条第一項の提案を募集するものとする。

(行政機関等匿名加工情報をその用に供して行う事業に関する提案)

第一百条 前条の規定による募集に応じて個人情報ファイルを構成する保有個人情報を加工して作成する行政機関等匿名加工情報をその事業の用に供しようとする者は、行政機関の長等に対し、当該事業に関する提案をすることができる。

2 前項の提案は、個人情報保護委員会規則で定めるところにより、次に掲げる事項を記載した書面を行政機関の長等に提出しなければならない。

一～四 (略)

五 提案に係る行政機関等匿名加工情報の利用の目的及び方法その他当該行政機関等匿名加工情報がその用に供される事業の内容

六～八 (略)

3 (略)

(提案の審査等)

第十二条 行政機関の長等は、第一百条第一項の提案があったときは、当該提案が次に掲げる基準に適合するかどうかを審査しなければならない。

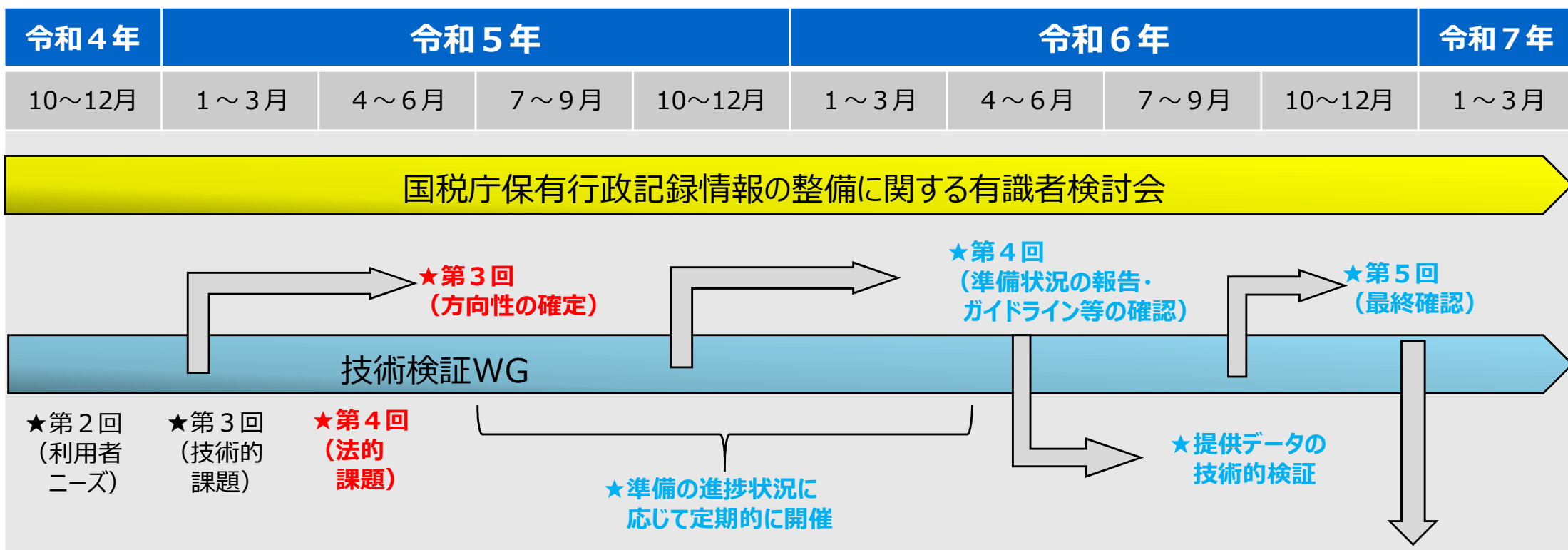
一～三 (略)

四 第一百条第二項第五号の事業が新たな産業の創出又は活力ある経済社会若しくは豊かな国民生活の実現に資するものであること。

五～七 (略)

4. 今後のスケジュール（案）

- 令和5年5～6月頃開催予定の有識者検討会において、匿名化データの整備方針（案）を議論の上で決定し、令和5事務年度（令和5年7月～令和6年6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・利用規約類の整備）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。



準備が整い次第、提供を開始する。