

第 16 回統計データの二次的利用促進に関する研究会 議事概要

- 1 日 時：平成 25 年 6 月 25 日（火） 10:00～12:10
- 2 場 所：総務省第 2 庁舎 6 階特別会議室
- 3 出席者：廣松座長、縣委員、椿委員、安田委員（玄田委員は欠席）
平山政策統括官、白岩統計企画管理官、金沢調査官
《説明者等》
総務省統計局統計情報システム課（松下課長補佐）、独立行政法人統計センター
総務部経営企画室（赤谷課長代理）、総務省統計研修所（小林次長）
《オブザーバー》
内閣府（統計委員会担当室、経済社会総合研究所）、総務省統計局、財務省、文部科学省、厚生労働省、農林水産省、経済産業省、国土交通省、日本銀行、独立行政法人統計センター
《事務局》
総務省政策統括官（統計基準担当）統計企画管理官室（中村管理官補佐、山根主査）
- 4 議 題：(1) 「オンサイト利用」及び「プログラム送付型の集計・分析」について
(2) 「統計データ・アーカイブ」について
(3) その他
- 5 議事の概要及び意見等
(1) 議題 1 「オンサイト利用」及び「プログラム送付型の集計・分析」について
事務局から、資料 1（「オンサイト利用」及び「プログラム送付型の集計・分析」の実用化に向けた課題整理と今後の進め方）について説明を行い、説明後、意見交換及び質疑応答が行われた。

(意見交換及び質疑応答の概要)
【全体方針、テストデータについて】
 - オンサイト利用施設の管理者の位置付けはどのようになるのか。現在の試行運用においては、「統計法第 33 条に基づき調査票情報の提供を受けた者（個々の利用について、利用者と共同して研究を行う者）」としているが、本格運用の際には変更されることになるのか。（安田委員）
 - 施設の管理者については、「行政機関から調査票情報の取扱いを委託された者」と整理できないかと考えている。この場合も、施設管理者には統計法に基づく調査票情報の適正管理義務や守秘義務が課されることとなる。（事務局）
 - オンサイトが利用できる者だけに限定するのではなくて、サイト以外、例えば地方の利用者等のためにもプログラム送付を併用するという事についてどう考えるか。（廣

松座長)

- 賛成。二次的利用の方法を広げるということは良い。ただし、どのようなリスクがあるか、あるいはプログラムを送付された側の負荷があるかと考えた時に、どのデータから扱うかは検討が必要。オンサイト利用施設との差別化については、例えば ID を利用してリンケージを行うようなものはオンサイト利用限定とし、プログラム送付の枠組みの中では当面検討中という整理にすることなどが考えられる。(椿委員)
- 利用者の教育・訓練のために第一段階としてテストデータを提供するというのは、審査者側からすれば手間が省けると思う。しかし、テストデータに個票データ(調査票情報)を用いるという点が気になる。テストデータについては、架空のデータや擬似マイクロデータなどの統計法第 33 条の枠外でできるとなれば、利用促進という観点から望ましい。(椿委員)
- オンサイト施設とプログラム送付型の併用については、私も賛成。ただし、プログラム送付型の場合の個票データ(調査票情報)を用いるテストデータと擬似マイクロデータとの差異はどのようになっているか。(廣松座長)
- 擬似マイクロデータについて、現在提供されている全国消費実態調査の例では、全ての調査項目が擬似マイクロデータとして提供されているわけではなく、一部の項目を取り出し、区分も大括りにしているなど元の調査票情報と異なるレイアウトになっている。このため、擬似マイクロデータを使ってプログラムを作成しても、そのまま本番(調査票情報)の分析に用いることはできない。(事務局)
- 統計法改正時に二次的利用の制度設計を検討していたが、簡易な教育用データから統計データの扱いに慣れてもらい、次に、匿名データである程度の分析を経験した上で、最終的に調査票情報を利用するというスキームを考えていた。一橋大学で試行的な匿名データの作成と提供を行ったが、その際、利用者の利便性を考えてのことであったが、調査票情報とは異なるレイアウトとした。このことが一因となり、調査票情報のプログラム送付型の集計・分析はうまく運用できなかったが、データの内容自体ではなく、レイアウトのみを調査票情報と同様に変更して匿名データをテストデータとして活用する方法も考えられる。

また、後ほど議題 3 で、統計センターからオンデマンド集計(インターネット上のシステムを利用したオーダーメイド集計の自動提供)の紹介があると思うが、通常の集計表だけであれば、データキューブ(詳細クロス集計表)を準備しておくことで、ほとんどのオーダーに対応できるのではないかと。データキューブを作成した際に、集計結果が少数になるようなセルについては予め秘匿措置を施すような技術開発を進めることは重要である。(安田委員)
- 政府のデータに限らず、研究利用できるデータに対する記述方式が確立していればうまくいくだろう。メタデータが不統一なため、いろいろなことが起きている。メタデータが統一化されていけば利便性も高まる。(椿委員)
- オンサイト利用施設の箇所数・所在地についてどのように考えているか。また、ドイツにおけるプログラム送付型のテストデータはどのようなものか。(縣委員)
- オンサイトの箇所数・所在地について、将来的にはそれぞれの地方ごとに中核となる

拠点があることが望ましい。しかしながら、現時点ではその見通しが立っておらず、まずは東京の拠点をどこにするかということを整理しなければならない。現在は一橋大学と統計数理研究所があるが、国直轄の施設設置の必要性等について、各府省とも検討を進めてきたい。次にテストデータの話だが、ドイツ連邦統計局では、リサンプリングや調査事項ごとのデータスワッピングをすることにより秘匿しているが、この方法であれば事項ごとの分布は保持される。(事務局)

【利用料金について】

- 利用料金については、オンサイト利用のみについての検討課題ではない。現在、オーダーメイド集計については作業量に応じた料金設定をしているところであり、プログラム送付型についても同様に料金の問題が出てくるのではないかと考える。(事務局)
- 現行法令上、匿名データとオーダーメイド集計は手数料を取ると法律上明記されている一方、調査票情報の提供については手数料の記載はないことに留意が必要。その代わりに、調査票情報の提供については利用目的で制限しているとも整理できる。(事務局)
- それはデータ自体の利用についての話であり、オンサイト施設の利用率については異なる考え方ができるのではないか。(安田委員)
- 料金の考え方については、今後引き続き整理が必要と認識。(事務局)

【分析用プログラムの共有について】

- プログラム作成自体が研究テーマという場合もあり、一律に提出するのはどうかと思う。それがオンサイト施設を利用するための条件であり、その代わりに手数料を採らないなどのバスターをするのであれば別だが。提出してもらうのが理想ではあるが、一律に提出(義務がある)と書かない方が良くと思う。プログラム送付型についても、任意で提供してもらうという整理が良いのではないか。(安田委員)
- 例えば、一定期間経過後に共有できるようにする、という考え方はどうか。(廣松座長)
- 2～3年後に自動的に公開するということはあり得る。(安田委員)
- 賛成。直後の公開は難しいと思うが、論文が公表された後の公開は良いのではないか。こういう仕組みができると、論文の査読時にアクセスが可能となる。研究者にとっては重要な知的財産であるため、ある段階までは保護し、その後公開というようにする。手間がかかるかもしれないが、学術コミュニティなどと議論して制度設計することも必要ではないか。(椿委員)
- この点は、オーダーメイド集計に関しても、今後利用者の範囲を広げようとする場合に同じような問題が出てくる可能性がある。今すぐ結論が出るものではないが、学会等と相談をした上で方向性を出すべきではないか。(廣松座長)

【プログラム送付型の対象とするデータについて】

- プログラム送付型について、標本調査かつ世帯調査がこの制度になじむのではないかと説明があったが、事業所・企業系のデータはどうか。(廣松座長)
- 集計方法によっては事業所・企業系のデータも出せる気はするが、もう少し検討が必

要ではないか。(安田委員)

- オンサイトを作るということ自体、オーダーメイド集計や匿名データではできない事業所系のデータを分析するということに設立意義があるともいえる。(廣松座長)
- 事業所・企業系の調査のテストデータに関して、個票データ(調査票情報)に基づいて作ることはリスクが高い。事業所・企業系の統計調査については十分な検討が必要である。(樫委員)
- 世帯対象がなじむというが、統計調査の環境と観点から言えば、企業よりも個人を守れと言う方が親和性があるとも言える。情報保護の観点からの検討も必要。(事務局)

【今後の進め方について】

- まだかなり詰めるべき点が残っている。今後の進め方だが、年度内を目処に統計法第33条の運用に関するガイドラインの改定を目指すということだが、委員から指摘のあった問題に関して、法的な観点も含めて検討頂きたい。ガイドラインを改定して、新年度から運用を開始するのか。(廣松座長)
- 改定を目指すと書いているが、実際の問題としてオンサイト利用施設の整備等について、具体的な目処は立っていない。内容について整理はするものの、運用について新年度からすぐにとというのは実際には難しいものと認識している。(事務局)
- この問題については一般に関心が高い。ガイドラインの改正についてはパブリックコメント等の手続きが必要になるかもしれない。また、学会等の意見も聞く必要がでてくるかもしれない。(廣松座長)
- ガイドラインの改正については、利用条件を変更する場合など、内容に応じてパブリックコメントの必要性があり得る。(事務局)
- オンサイト利用施設やプログラム送付型について、これまでこの研究会でも何度か議論があった。それらを踏まえて、ガイドライン等で考え方を整理するということについては評価したい。一方で、実行するためには、各省の協力が重要である。(廣松座長)

※ なお、研究会における議論を踏まえ、研究会の配布資料(論点ペーパー)について、以下の観点から一部修正を行うこととなった。

○プログラム送付型集計・分析で利用可能なデータについて

プログラム作成の容易性から「標本調査かつ世帯対象調査がなじむと考えられる」としていたが、調査対象の秘密保護等の観点も重要であり、対象データについて今後検討が必要である旨に修正。

○オンサイト利用について

引き続きオンサイト利用のための要件の具体化の検討を進めていく必要はあるが、ガイドラインの改定により直ちに実運用が開始できる状況にはなく、各府省の理解と協力等も必要であることから、年度内を目途にガイドラインの改定を目指す旨の記述を削除。

(2) 議題2 「統計データ・アーカイブ」について

事務局から、資料2-1(「統計データ・アーカイブ」の今後の検討方向について(修正版))、資料2-2(調査票情報等の管理状況について)、資料2-3(公文書管理制度の概要)及び資料2-4(過去の研究会等の報告において示された「統計データ・アー

カイク」に関する記述) について説明を行い、説明後、意見交換及び質疑応答が行われた。

(意見交換及び質疑応答の概要)

- 「世界最先端 I T 国家創造宣言」の中で、オープンガバメントを宣言している。こちらの構想するデータの公開というのは、「統計データ・アーカイブ」の対象としているものの概念より広いと受け取っている。統計委員会の範囲で、「統計データ・アーカイブ」を作るという考えは良いが、オープンガバメントとの連携がどのようになるのかは非常に難しい。よく議論すべき。(県委員)
- 現時点では、個人情報の保護を前提として、既に公表しているデータを機械で利用しやすくしたり、統計の分野で言えば、既に e-Stat で公表している集計表をもっと使い勝手の良いようにできないかという視点で議論されている。一方、「統計データ・アーカイブ」で想定しているものは、いわゆる「二次的利用」であり、センシティブな情報についてどのように扱うのかといった内容であり、対象となるデータの範囲が違うのではないかと考えているところ。(事務局)
- オープンデータ政策が最終的に目指しているものはもっと広い概念と思われ、それによって位置付けも変わってくるのではないかと。政府全体の構想の中で、この任務(「統計データ・アーカイブ」)の位置付けを考えることが重要ではないか。(県委員)
- 今期の基本計画を策定する際にオープンデータという言葉はなかったが、I T 戦略本部が、ほぼ近いようなことを言っていた。そういう環境の中で、統計サイドとしては、基本計画で「統計データ・アーカイブ」を現行のような形(統計データの二次的利用の観点からの構想)で取り上げた。その意味で、より広い意味での「アーカイブ」や「オープンデータ」政策に関して、そこまで統計サイドが面倒をみるべきなのか、個人的には疑問を持つ。この研究会も統計委員会も検討すべきことは、あくまで足下の統計法の対象となる統計調査および統計であって、それから先をどのように広げるかについては、委員会全体の機能に関わることであり、大変難しい問題をはらんでいる。

このため、当研究会として、今期の基本計画の中で残された課題としての「統計データ・アーカイブ」について、25 年度末までに何らかの結論を得るということが要求されている制約の下で、あまり広げすぎのではなく、意義のある結論・対応としては、資料 2-1 の「a-2 適切な収集、整理、保管の確保」に特化するのが良いのではないかと考える。(廣松座長)
- 資料 2-1 の趣旨は、原案として、3 つの次元の異なる機能について、この研究会として注力するのは、「a-2 適切な収集、整理、保管の確保」ということにして、「a-1 一元化された専門機関の設置」は現状から見るとほとんど実現可能性はないであろうということである。また、それと密接に関連した問題として、調査票の原票そのものについてどうするかという問題がある。この点について御意見いただきたい。(廣松座長)
- 原票そのものについてはないが、二次的利用の観点からは、利用者側としては、調査票情報だけでなく、それに付随するメタデータが重要。具体的には、e-Stat でも既に掲載されている「調査の概要」がある。また、一般的には公表されていないが、「調査のた

めのマニュアル（調査の手引き）」や「集計作業における手順書」がある。直近の調査についてはあまり問題にならないが、過去に遡って時系列的に研究分析する際には、データの誤用を防ぐためにこれらの情報は重要となる。（安田委員）

- 「統計データ・アーカイブ」を何のために作るのかという点が重要。一般的な文書を保存して今後の意思決定の改善のために利用するのは異なり、日常的な決裁とは無関係に将来に向けて蓄積すべきものを蓄積する、そういう意味での二次利用に限定するという理解でよいか。（縣委員）

→ 少なくとも、現行の基本計画における「統計データ・アーカイブ」の部分はそうである。ただし、ご指摘の点を次期基本計画の中で「統計データ・アーカイブ」に関して言及するとすれば、つまりそれを取り込むとなれば、また別の視点、新たな視点が必要となろう。（廣松座長）

- 二次的利用に関して実効性のある計画を立てることと、調査票原票を永年保管することはそもそも別問題であるという印象。統計的なデータ分析（二次的利用）の範囲であれば、現行基本計画の記載が意味するところの説明としてよくわかるが、歴史的な公文書として、個票が100年後、200年後どう利用されるかについてはあまり判断材料がない。

（1）「統計データの収集、整理、保管」に関しては、ペンディング部分は、個票を残す意義があるのかどうかというところに限定した上で、「統計データ・アーカイブ」固有の問題として整理し、それ以外の（2）「統計機関相互のデータ共有・連携」及び（3）「ユーザーへの提供」については、別途それぞれで効果的な計画を作成し対応していくことで、「統計データ・アーカイブ」の問題から分離しておくことの意味は非常にある。（椿委員）

- 調査票原票そのものを残すべきかどうかについては、ここでは判断はせずペンディングとした上で、5ページ（現行基本計画）への対応としては、「a-2 適切な収集、整理、保管の確保」に、当面注力するというところでよろしいか。統計データの提供のワンストップ化や提供する際の目的・資格に関する制約の柔軟化に関しては、「二次的利用」の課題として扱い、「統計データ・アーカイブ」の議論とは質が異なるということで整理したい。（廣松座長）

→ 一同了承。

（3）議題3 その他（統計データのオープン化・透明化の取組（オープンデータ政策関係））

統計局及び独立行政法人統計センターから、参考資料2「統計におけるオープンデータの高度化」についての説明が行われ、説明後、意見交換及び質疑応答が行われた。

（意見交換及び質疑応答の概要）

- 先ほどの「統計データ・アーカイブ」と関連するが、e-Statとは別に「政府統計オンライン調査総合窓口（オンライン調査システム）」というものがある。各統計調査の実施に当たり当該システムを実際に利用するかどうかは各府省の判断によるところがあるものの、オンラインによる統計調査の機能は整備されている。資料2-1の（2）「統計機関相互のデータ共有・連携」との関係でいうと、「e-アーカイブ」というようなものをつ

くるかどうかという論点があり得るし、その利用者の範囲をどのようにするのかという論点があり得る。今はそのような構想は全くないが、将来的にはそうした問題提起もあり得ると思う。(廣松委員)

- 相互関係がよくわからないので、現行の仕組みと、今回議論の対象となっている「統計データ・アーカイブ」との関係のようなものを次回にでも示していただきたい。何を新しく作るべきなのかとか、従来のもとの何をどう組み合わせるのかについて教えていただきたい。(縣委員)

→ 事務局と相談の上、対応したい。本日は必ずしも十分時間を取れなかったため、まだ言い足りないこと等があれば事務局まで意見をお寄せいただければ、対応可能なものについては次回までに資料等を用意して、議論をしていただく材料にしたい。(廣松座長)

(4) 次回開催予定について

- 平成 25 年度は 4 回程度の開催を予定しており、次回会合については、7 月 16 日 (火) 午前中を予定 (※)。

(※) 実際には、7 月 16 日 (火) の研究会は延期となった。次回の開催は秋頃を予定。

以上

≪文責：統計企画管理官付高度利用担当≫