

統計マイクロデータの提供の新たな展開方向の一考察

NSTAC

Working Paper No.35

平成 29 年 6 月

独立行政法人 統計センター

製表技術参考資料は、独立行政法人 統計センターの職員がその業務に関連して行った製表技術に関する研究の結果を紹介するためのものである。

ただし、本資料に示された見解は、執筆者の個人的見解である。

目 次

1	はじめに	1
2	匿名データの現状	1
3	新たな匿名データの創設	2
	（1）新たな匿名データの創設（その1）（パブリックユースファイル）	2
	（2）新たな匿名データの創設（その2）（オンサイト施設限定）	4
4	研究室等における探索的研究とオーダーメイド集計（リモートアクセス方式によるオンデマンド集計）	5
5	オンサイト施設における調査票情報のリモートアクセス方式による利用	5
6	おわりに	5

統計マイクロデータの提供の新たな展開方向の一考察

奥積 雅彦*

1 はじめに

統計法（平成19年法律第53号）第32条あるいは第33条に基づく調査票情報の利用の場合、旧統計法（昭和22年法律第18号）における運用（利用目的に高度公益性があることを要件として運用）が現行の統計法における運用に継承されている。さらに、統計法第33条第2号に基づく調査票情報の利用の場合、統計法施行規則第9条第1号から第3号までのいずれかを満たすことを要件としているが、この要件は、旧統計法における運用（利用目的に行政機関等が行うのと同程度の公益性があることを要件として運用）が現行の統計法に基づく省令（統計法施行規則第9条）やガイドラインに継承されている。ただ、行政機関が自ら利用したり提供したりする権限は、旧法下では制度官庁に付与されていたが、「統計法制度に関する研究会」（座長：廣松毅東京大学教授）の提言（平成18年6月「統計法制度に関する研究会報告書」参照）を踏まえ、制度官庁において一元的に調査票の使用について承認を行う必要性は、他の法令との比較においても乏しくなっているとし、統計データの利用の促進のためには、調査票の使用手続の簡素化を図るため、現行の統計法の下では各行政機関に付与されることとなった。

このように現行の統計法では、従来の高度公益目的利用の枠組みは維持しつつ、新たに、一定の公益性があると認められる場合に、委託による統計の作成（オーダーメイド集計）と匿名データの提供の制度が創設された。

本稿では、統計マイクロデータ（調査票情報と匿名データの総称（注））の提供の現状と課題と今後の方向性について私見を述べることにしたい。

(注)統計マイクロデータ：統計の作成のために事実の報告を求める調査によって集められた情報や当該情報を特定の個人や法人等の識別ができないよう加工した匿名データのこと（⇒平成29年5月19日統計改革推進会議最終取りまとめ 脚注1）

2 匿名データの現状

匿名データ（一般の利用に供することを目的として調査票情報を特定の個人又は法人その他の団体の識別（他の情報との照合による識別を含む。）ができないよう加工したものをいう。（統計法第2条第9項の定義規定））は秘密保護のための措置が施されているものではあるが、集計された統計ではないことから、統計制度に対する国民の不安を招かないようにするため、提供対象を一定の公益性が認められる範囲（学術研究、高等教育、国際比較）に限ることとされ、統計法制上、匿名データを広く一般に提供するような運用はなされていない（統計法第36条、統計法施行規則第15条）。

* （独）統計センター 統計情報・技術部次長，職命 統計情報・技術部統計作成支援課長事務取扱

匿名データは、このような特性があることから、提供を受けた者に対し、その情報の適正管理義務（統計法第42条第1項第2号）が課され、提供を受けた目的以外の利用が禁止されている（統計法第43条第2項）。また、凶利目的利用に対しては罰則規定も用意されている（統計法第61条第3号）。

さらに、基幹統計調査に係る匿名データの場合、統計委員会への付議が義務付けられている（統計法第35条第2項）。

なお、現行の匿名データ（総務省統計局所管調査に係る匿名データは、統計センターが統計法第37条により全部委託を受けて提供業務を実施）の提供件数は、年間30件程度と低調である。匿名データは、直接識別情報の除去のほか、リサンプリング、分類区分の集約、トップコーディングなどにより匿名性を担保している。匿名性が強まると情報の質と量の損失につながり、利便性が低下することも想定され、このことも提供件数が低調となる一因であると考えられる。

また、現行の匿名データ（総務省統計局所管調査）は、年次が古く10年以上前のものしか提供されていない場合もあり、このことも提供件数が低調となる一因であると考えられる。一方で、匿名データの諮問・答申手続とその準備段階の負担は大きく、統計局の調査所管課は、本体調査の企画、公表事務に忙殺され、匿名データの作成は二次的利用のサービスであるが故に優先度が低くならざるを得ない。このことも、提供する匿名データの年次が古くなっている一因であると考えられる。

このため、匿名データの諮問・答申手続に係る統計法の規制は、緩和（手続きの簡素化など）する余地があるか検討する必要があると考えられる。

3 新たな匿名データの創設

現行の匿名データは、平成19年に全部改正された統計法により、創設されたものであり、一定の効果は認められるが、社会の情報基盤としての統計マイクロデータの利活用の促進の観点からは、現行の匿名データとは別に、利用場所や利用目的に応じた秘匿性の強弱を考慮した新たな匿名データの創設が必要と考えられる。

(1) 新たな匿名データの創設（その1）（パブリックユースファイル）

統計改革推進会議最終取りまとめ（平成29年5月19日）において、社会全体における統計等データの利活用を促進する観点から、「一般の人が利用できる匿名データについて、必要な法制面、技術面から検討し、提供を開始。その際、提供の早期化、手続きの簡素化も検討」することについて提言がなされた。

これは、統計制度に対する国民の不安を招かないようにするため、ことごとくダミーデータに置き換える強度の加工を加えた匿名データでパブリックユースファイルとして、広く一般に提供をするような匿名データを創設することも前向きに検討するべきであると考えられる。データ自体に生々しさがなければ、統計法による規制も緩和する余地があるか検討する必要があると考えられる。

これが実現すれば、調査票情報に直接、触れないで、オーダーメイド集計の提供申出に際

しての準備行為、効率的なオンサイト利用に際しての準備行為に利用することができ、ユーザーにとっての利便性が向上することが期待できる。

研究室において調査票情報に直接触れないで、一般の人が利用できる匿名データ（パブリックユースファイル）で探索的研究を行った上で、オーダーメイド集計に展開していくことも想定される。利便性の向上の観点からは、オンデマンド集計やプログラム送付型のオーダーメイド集計の実現が望ましいと考えられる。

また、探索的研究に際しては、一般の人が利用できる匿名データ（パブリックユースファイル）をソースとするオンデマンド集計のニーズがあれば、検討すべきではないか。

【挿話】 一般用マイクロデータとの関係

一般の人が利用できる匿名データ（パブリックユースファイル）の検討に際しては、一般用マイクロデータ（集計表から作成するなど、調査票情報を直接的に用いない方法により作成する擬似的なマイクロデータ…公的統計の整備に関する基本的な計画（Ⅱ期）25頁の注11）との関係を整理する必要がある。

一般用マイクロデータは、そのデータが疑似データであることから、学术论文のバックデータとして利用することは適当ではないとされているが、①高等教育における演習や②学術研究利用等におけるオンデマンド集計やオンサイト利用に際しての事前準備には有効であると考えられる。また、一般用マイクロデータは統計法の規制を受けないことからパブリックユースファイルとしてリーズナブルに利用できるというメリットがある。

統計センターでは平成27年度の年度目標において、主務大臣から、一般用マイクロデータの提供に向けた研究を行うことを指示され、平成21年全国消費実態調査の特別集計により公表された集計表（基本統計量など）を用い、調査票情報を直接的に用いない方法により平成28年3月に量的変数に係る一般用マイクロデータの提供を開始し、ユーザーや有識者の意見等を踏まえ、同年12月に相関係数を考慮した改訂版の提供を開始した。その後、ユーザーや有識者から詳細版の提供を希望する意見もあり、試行提供を経て、平成29年6月までに本格提供に向けて準備しているところ。さらに質的変数に係る一般用マイクロデータの提供に向けた研究に着手しているところ。

また、統計法の規制を受けずに自由に利用できるパブリックユースファイルである擬似的なマイクロデータに対する産官学のニーズは根強く、匿名データの利用により公表された集計表（基本統計量など）を用い、匿名データを直接的に用いない方法により擬似的なマイクロデータを開発し公開している事例もある。

いずれにしても、一般の人が利用できる匿名データ（パブリックユースファイル）の制度設計（法的規制や手数料の徴収も含む）が、リーズナブルに利用できないようなものであれば、一般用マイクロデータの存続もありうると考えられる。

(2) 新たな匿名データの創設(その2)(オンサイト施設限定)

さらに、統計改革推進会議最終取りまとめにおいては、「社会全体における統計等データ(注)の利活用を促進する観点から、秘匿性の高い統計等データであっても、その一部でも提供できないか、匿名化して提供できないか、匿名化が困難な場合についてオーダーメイド集計やオンサイト施設での利用ができないか等、総合的かつ前向きな検討を行うこと」について提言されている。

(注) 統計等データ：統計、統計マイクロデータ及び統計的な利活用を行うために用いられる行政記録情報をいい、それらのデータの利用や解釈を行うために必要な関連情報(メタデータ)を含む。

(⇒統計改革推進会議最終取りまとめ)

オンサイト施設においてのみ利用可能な匿名データで、調査票情報から直接識別情報(調査区番号、個体識別番号など)を削除し、配列順を変える加工のみを施したもの(以下「オンサイト施設限定の匿名データ」という。)を創設することも考えられる。他の情報との照合による個体識別の防止は、オンサイト施設からの成果物の持ち出し審査により担保することが可能であると考えられる。

これまで、このような粒度の統計マイクロデータを大学の研究者等が利用しようとする場合は、統計法第33条の手続きを経て、その利用目的に、行政機関が行うのと同等の公益性(高度公益性)があると認められる場合に限り、調査票情報の提供を受けることができるものであった。ここで「行政機関が行うのと同等の公益性(高度公益性)」とは、①行政機関からの委託研究、行政機関との共同研究、②科研費などの公的補助金を受けている場合、③行政機関が政策の企画立案等に役立つと認めた場合である(統計法施行規則第9条)。

しかし、オンサイト施設限定の匿名データが創設されれば、一定の公益性(学術研究等)がある場合にも利用可能となり、社会における統計マイクロデータの利活用が促進される。

ところで、学術研究等における統計マイクロデータの利用において、行政機関が行うのと同等の公益性(高度公益性)があると認められる場合かそれ以外(一定の公益性(学術研究等)がある場合)の境界が、科研費などの公的補助金を受けている場合か否かが判断の尺度とされることになっている現行の運用は、見直しの余地があると考えられる。

オンサイト施設限定の匿名データについては、受益者負担の原則の下、手数料を徴収することとなる。ただし、その利用目的が、行政機関からの委託研究、行政機関との共同研究の場合は無償とすることも検討すべきであると考えられる。

この匿名データは、オンサイト施設限定であることから、現行の匿名データのような電磁的記録媒体への出力の手数料は不要となる。一方で、オンサイト施設からの成果物の持ち出そうとする場合は、他の情報との照合による個体識別の防止の観点からの審査が必要となることから、持ち出し審査手数料の創設の検討も必要と考えられる。

この匿名データの提供業務について、統計センターへの全部委託(統計法第37条)も可能ということになれば、統計センターに承諾権が付与され、府省とのやりとりが不要になる。

また、委託元の府省と統計センターとの間で全部委託の場合における、持ち出し審査基準

を取り決め、その範囲内であれば、統計センターの裁量で持ち出し審査を行うことが可能となり、利用者と府省のそれぞれにメリットがあると考えられる。(全部委託しない場合は、統計センターは、行政機関に取り次ぐのみ。)

4 研究室等における探索的研究とオーダーメイド集計（リモートアクセス方式によるオンデマンド集計）

前述の「一般の人が利用できる匿名データ（パブリックユースファイル）」が実現すれば、研究室において、これを用いて、探索的研究を行った上で、オーダーメイド集計を申し込むことが可能となる。利便性の向上の観点からは、調査票情報を直接見ないリアルタイム型のオンデマンド集計（リモートアクセス方式）あるいはプログラム送付型のオーダーメイド集計の実現が望ましいと考えられる。リモートアクセス方式の場合、研究室では、研究の成果物を閲覧することは可能とし、成果物は、一定の条件を満たせば申請により入手可能となるしくみを検討する必要があると考えられる。入手しようとする場合は、他の情報との照合による個体識別の防止の観点からの審査が必要となることから、持ち出し審査手数料の創設の検討も必要と考えられる。

なお、調査票情報を直接見る必要がある場合は、オンサイト施設を利用することを想定している。

【挿話】 調査票情報を直接見る必要がある場合は、オンサイト施設を利用することとなることから、調査票情報を直接見ることはできない研究室において探索的研究をストレスなく行うことができる環境を整備することは、統計マイクロデータの利活用の促進を図る上で重要であると考えられる。

5 オンサイト施設における調査票情報のリモートアクセス方式による利用

研究目的によっては、リモートアクセス方式により調査票情報を直接見る必要がある場合も想定される。また、直接識別情報を含む調査票情報の閲覧や他の情報（行政記録情報）とのデータリンケージをする必要がある場合も想定される。このような場合は、取り扱う情報の性格から、その利用目的が、行政機関からの委託研究、行政機関との共同研究の場合を基本とすることが考えられる（統計法第32条又は第33条による利用を想定）。

なお、成果物における個体識別の防止は、オンサイト施設からの成果物の持ち出し審査で担保することを想定している。

6 おわりに

現行の統計法が公布されてから、10年目を迎える。公的統計の整備に関する基本的な計画もⅢ期を目前に控え、また、統計改革推進会議最終取りまとめが出され、統計行政を取り巻く環境も大きく変化している。そのなかで、法律の改正が必要となるものがある一方で、法律の改正に至らないレベルでも改善できることはあるはずである。

製 表 技 術 参 考 資 料 35

平成 29 年 6 月 発 行

編 集 ・ 発 行 独 立 行 政 法 人 統 計 セ ン タ ー

〒162-8668

東 京 都 新 宿 区 若 松 町 19-1

電 話 代 表 03 (5273) 1200

掲 載 論 文 を 引 用 す る 場 合 は 、 事 前 に 下 記 ま で 連 絡 し て く だ さ い

統 計 情 報 ・ 技 術 部 統 計 技 術 研 究 課

TEL : 03-5273-1368