

第14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24年12月19日（水） 10:00～12:00
- 2 場所：総務省第2庁舎 3階統計センター第1会議室
- 3 出席者：廣松座長、縣委員、玄田委員、椿委員、安田委員
平山政策統括官、白岩統計企画管理官、浜東調査官
《説明者》
坂下課長（独立行政法人統計センター統計技術研究課）、小林調査官（総務省統計研修所）
《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（中村管理官補佐ほか1名）
- 4 議題：(1) オンサイト利用について
(2) 擬似マイクロデータについて

5 議事の概要及び意見等

○ 議題1 オンサイト利用について

事務局から、資料1-1「オンサイト利用に関する仕組みの整備に当たっての論点」の説明が行われ、引き続き、小林調査官（総務省統計研修所）から、資料1-2「オンサイト利用に関する提案」の説明が行われた。説明後、意見交換及び質疑応答が行われた。

（主な意見、質疑応答）

- ・ 私は、統計そのものを専門としていない立場であるが、諸外国の状況を見て、日本の位置付けを知った際に驚いた。オンサイト利用の仕組みについては、是非、日本でも整備してほしいというのが最初の思いである。現状では、調査票情報はどのような形で管理され提供されているのか、また、オンサイト利用によりどのような形での管理・提供を目指すのかイメージがあればいただきたい。（縣委員）
- ・ 現状では、調査票情報を研究者へ提供し、それぞれの研究室において利用していただいている。ガイドラインにより保管場所や利用場所等のセキュリティに関する基準を定めているが、原則として、申請書における書面審査となっており、実際の利用者における管理実態の把握は、監査を行わない限り難しい。一方、オンサイト利用では、監視カメラや監視員、パソコンのログ（利用履歴）の保存等により、調査票情報の利用状況を的確に管理・把握できる。（事務局）
- ・ 扱われるデータの形態はどうか。デジタル化されているのか。今後は変わるのか。（縣委員）
- ・ 基本的に現状でもデジタル化されている。また、全ての調査項目が提供されるわけではなく、申請時に利用目的・分析方法を詳細に報告してもらい、利用目的等から見て必要最小限の範囲に限り提供している。利用者側からみると、探索型の研究（試行錯誤的な研

究)ができないというデメリットがあるが、オンサイト利用に限定して、データセット(全体)の利用を可能にするということも考えられる。(事務局)

- ・ オンサイト利用については、利用者を特定の場所に閉じ込めるというものではなくて、調査実施者と利用者の両方にメリットがあるような使い方を目指すべき。もちろん、そのためにセキュリティの確保が前提になる。(廣松座長)
- ・ ルクセンブルク・インカム・スタディ(LIS)については、資料1-1(参考①)の表では、どこに位置付けられるか。(玄田委員)
- ・ LISはプログラム送付集計型となる。(事務局)
(注): LISのプログラム送付集計型サービス(LISSY)は、匿名データを用いた参照不可能型に分類される。
- ・ オンサイト利用が広がるのが重要と認識しているが、その場合の大きな原則として、3点ぐらいあるのではないかと考えている。もちろん安全性が最大の原則であり、公益性と公平性も重要な基準になる。オンサイトの施設が広がり、地理的な条件がクリアされていくとともに、利用者の条件をどのようにするかについて、公平性の観点からもう少し踏み込んだ議論が必要ではないかと思う。直接利用型については、オンサイトとは別の議論も必要。今まで、直接利用型が日本で使われてきた特殊な背景については、別途考えなければならない。オンサイトや新統計法ができる前の段階的な措置であったところが大きいと思うが、オンサイト利用が公平性の観点で何ら問題がないとなった場合に、初めて直接利用型について、資料1-2では凍結という言葉も出ているが、どうするか議論できるのではないか。直接利用がどのような背景で利用されているか、様々な背景があると思うが、直接利用とオンサイト利用については、切り離して、慎重に議論した方が良いのではないかという印象を受けた。(玄田委員)
- ・ その意味では、利用実態を把握することは大変重要で議論の前提となる。(廣松委員)
- ・ 運用主体・運用形態に関して、情報・システム研究機構においてオンサイト施設を立ち上げるということを考えたとき、各府省が共通の基準を設けるというのはありがたい。情報・システム研究機構では、公的統計のオンサイト利用については、統計センターの監査を受けるという形で進んでいるが、厚生労働省のレセプトサンプリングデータという匿名データと同様に秘匿すべき機密性の高いデータについては、公的統計とは別のセキュリティ基準で、別個に監査を受けるという形で進んでいる。今後、オンサイト利用の仕組みを整備するに当たり、各府省がそれぞれ独自に監査を行うということになれば、監査を受ける側の負担だけでなく、行う側の手間も問題である。各府省共通の基準を設けることは、利用者側・提供者側の両者にとって有益であるという印象をもっている。(椿委員)
- ・ 今後考えるべき重要な観点であり、その際に、総務省政策統括官室がイニシアティブをとることを期待する。(廣松座長)
- ・ 今の点は、全く同感であるが、例えば、ドイツの場合は集権型の統計行政機構となっており、統計行政そのものがある程度標準化されているという前提がある。オンサイトの整備を契機に、ある程度の統計行政の統一を図っていくのか、逆に統一が図られるまで待つのかという議論がある。(縣委員)
- ・ 現状では、資料1-1にあるとおり、試行的な施設が2か所あるのみで、厚生労働省のレセプトデータの利用は別として、各省が個別にやっているところはないと認識している。(廣松座長)

- ・ 現状では、統計法第 33 条第 1 号に基づく公的機関の利用は別として、第 2 号に基づく研究者の利用は、特定の機関や科学研究費補助金を受けた者に限られており、それほど広く投網をかけなければならない状況にはなっていないのではないかと印象がある。どちらを先にするかという議論もあるが、早いうちに基準を作れば、利用はオンサイトになびいていくのではないかとというのが、実際に現場にいての実感である。

また、現行制度では、各統計の調査実施者ごとに承認を受けざるを得ないということになっているが、ガイドライン等により、基準の統一化を図った方が良い。ただし、この場合、基準はかなり厳しいものにならざるを得ないのではないかと。

また、今の制度では、データ（調査票情報）は、あくまで「利用者」に提供されるものであるが、オンサイト施設側の管理責任についても、問題が発生した場合のペナルティなど、踏み込んだ形で議論する必要がある。そうした観点も含めた検証を行うため、現在一橋大学で試行運用の研究をしている。まだ結論が出ていないところであるが、今後明らかになれば細かい議論もできるのではないかと考えている。（安田委員）

- ・ 管理者（一橋大学）と利用者間で契約書を作っているのか。（玄田委員）
- ・ 今のところ、オンサイト施設の管理者についても、全て利用者を含めている（管理者も、共同利用者の一部として利用申請している。）。現行制度では、利用者でない限り、オンサイト施設の管理者であっても調査票情報を利用できないこととなっており、この点についても運用に当たって今後整理が必要である。（安田委員）
- ・ 関連してだが、情報・システム研究機構では、機密性の高いデータを収集・管理する事業を行おうとしており、民間保有のデータについても検討しているが、契約事項として管理者側の無限責任があったことから上手く進んでいない状況にあることを事例として紹介しておく。（椿委員）
- ・ 先程の縣先生の発言に関連して、ドイツでは、個別の連邦統計庁や連邦雇用庁等の RDC（Research Data Center、オンサイト施設）の経験を踏まえてと言うことらしいが、国としての委員会があり、そこでガイドラインを作成し、RDC の要件等を示しているとのことである。（小林調査官）
- ・ 玄田委員から公平性の観点の重要性が提起されたが、オンサイト施設を整備する場合、当然のことながら国で全部用意することは不可能であることから、何らかの形で、大学等の公的機関との協力は不可欠と思われる。特に地理的なことを考えた場合、地方にある大学等公的機関との協力あるいは契約が必要と思われる。（廣松座長）
- ・ 公平性という観点では、手を挙げたところは全部採択するという考えもあり、それも一つのやり方と思われる。アメリカセンサス局の例では多くのオンサイト施設が設置されているが、大口ユーザーがあるところにオンサイト施設を設置していった結果であり、公平性の観点から、ユーザーの多寡に限らずに各州に一つずつ必要という考え方で設置したものではない。今の統計センターの連携協定のやり方がそうであるが、基本的には、利用者が多くいると思われる大学等が希望し、立候補した場合には拒否せずに認めるというやり方がよいのではないかと。

また、昔のいわゆる大型計算機センターを活用するというのも一つの手かもしれないが、この場合、利用料金をとることを認めることが必要である。（安田委員）

- ・ かつて各大学で整備されていた、いわゆるメインフレーム（大型コンピュータ）を備えた電気計算機センター等について、現在はどうなっているのか。（廣松座長）

- ・ システム的には、メインフレームからワークステーション（ネットワークを通じて相互に接続する分散型のシステム）に変わってきているが、情報処理教育センターや総合情報処理センターとして、現在も各大学で施設・設備は有している。一部屋だけでも利用できればオンサイト施設として運用できるので、オンサイト施設としての基準をきちんと整備さえすれば、公平性の観点からも割と機能するのではないか。（安田委員）
- ・ 二次的利用の目的については、学術研究目的であり公益に資するものに限られるという前提があるが、将来的に二次的利用の目的が拡大されることがあった場合には、大学等の学術研究機関に設置された場合にも、学術研究目的以外の利用にも開かれるということが必要ではないか。また、公平性を重視するのであれば、日本では現実的ではないかもしれないが、ドイツではオンサイト施設の運営を政府機関（州政府）が行っており、都道府県の関係部署が運営するという方法もある。ただし、この方法が現実的ではないということであれば、大学法人等に設置し、利用の門戸は広げておくという方法で良いのではないかと考える。（縣委員）
- ・ 大学においても、現在、社会的貢献が求められており、利用目的の拡大があった場合に、利用の門戸を広げるということは、大学の社会貢献という方針にも沿ったものと言えるのではないか。（廣松座長）
- ・ 旧統計法の時代に、LIS を参考にプログラム送付集計型の研究をしたことがあるが、利用者側の技能不足が原因で、送付されたプログラムが動かないことが多かった。時間をかけてやり取りすれば最終的には上手くいくと思われるが、限られた期間では困難であった。いくつかについては、受け付けた側でプログラムを手直しすることにより上手く計算結果を出すことができたが、手直しする側の負担が大きい。プログラム送付集計型については、事前研修も含めた利用者の訓練を行うことが必須である。LIS においても、現在は変わってきているが、以前は1週間程度の研修を受けて認定を受けることが必要であった。プログラム送付集計型については、利用者側に一定以上の技能が求められ事前研修及び認定制度が必要となるが、日本で導入する場合には、この利用条件についての反発が大きく、現状では難しいのではないかとと思われる。ただし、決して、この方法が将来的にも一切使えないということではなく、例えば、大学院生等を主な対象として活用していくことは一つのアイデアとしてあり得る。（安田委員）
- ・ 今日の議論で、とりあえずオンサイト利用の意義・利点については認められたと考えるが、運用主体や運用形態については、今日の意見を整理し、より具体的な形で提示した上で改めて議論したい。（廣松座長）

○ 議題2 擬似マイクロデータについて

坂下課長（独立行政法人統計センター統計技術研究課）から、資料2-1「擬似マイクロデータについて」の説明が行われ、引き続き、事務局から、資料2-2「擬似マイクロデータの制度上の位置付けについての論点（素案）」の説明が行われた。説明後、意見交換及び質疑応答が行われた。

（主な意見、質疑応答）

- ・ 現在、試行提供している擬似マイクロデータについては、いわゆる「事実上の匿名化」であるか、それとも「絶対的な匿名化」であるか。（縣委員）
- ・ 乱数を用いて作成しており、元のデータと全く違ったものとなっており、「絶対的な匿名

化」と考えていただいてよい。(坂下課長)

- ・ そうであれば、現在、第三者への再配布を禁止している理由は何か。(縣委員)
- ・ 内部でも議論があった点であるが、本物のデータではないため再配布しても構わないという考え方もあるが、第三者から見た場合に、調査票情報が漏えいしたと誤解を生じさせるおそれがあるということと、擬似マイクロデータがどのような使われ方をされるのかを把握しておきたいということがあり、少なくとも現時点では再配布は認めない取扱いとしている。(坂下課長)
- ・ ドイツの例では、「事実上の匿名化」のデータを、日本で言うところの匿名データ (SUF : Scientific Use Files) として学術研究目的のみで提供し、「絶対的な匿名化」のデータについてはパブリックユースファイル (PUF : Public Use Files) として、利用目的を問わず提供している。これと対比した際に、日本の擬似マイクロデータについても、パブリックユースファイルとして利用の汎用性を最高度に高めても良いのではないかと思われる。(縣委員)
- ・ 擬似マイクロデータについては、あくまで、擬似的 (リアリスティック) なものであり、リアルなデータ (実在するデータ) ではない。一方、匿名データについては、ある程度情報量が落ちていてもリアルなデータであり、この点を踏まえて議論をしなければならない。先程、坂下課長が述べられたとおり、擬似マイクロデータがあたかもリアルなデータであると誤解を生む点については、社会に与える影響が大きいと考える。擬似マイクロデータ提供される前提として、利用上の注意がユーザーにおいても提供する側においても徹底がされていることが必要である。

一方、擬似マイクロデータについては、乱数を発生させていても、セル型 (一定の塊) になっている場合には、実際のデータと一緒にしてしまうわずかな可能性があるものの、逆に、匿名化措置が困難と考えられている企業情報や年齢各歳別などの細かくなっているデータについては、特定化されるリスクが小さくなっていると思われる。また、匿名データでは作成が不可能である複数の統計データをリンケージさせたデータも作成できる可能性があることがメリットとして考えられる。こうした点では、教育目的以上に、研究目的でも、本格的な研究の準備段階利用として活用できる可能性がある。(椿委員)

- ・ 擬似マイクロデータの有用性については高く評価する。その上で、匿名データを基にしたデータの作成方法に注力することが重要であると考えている。先程のプログラム送付集計型の議論と関わっているが、匿名データを利用する前段階のチェックとしてプログラム送付集計型を活用することを考えた場合、匿名データとレイアウトを同様にした擬似マイクロデータを作成することは大変有意義であり現実的でないか。

また、擬似マイクロデータという名称については、混乱を招くおそれがあり、私も反対である。代替案がある訳ではないが、強いて言うなら「パイロット版データ」等とした方が良いのではないか。パイロットという用語には、「試験的である。本物ではない」というある程度の社会的なコンセンサスがあると思う。「擬似」という用語よりは、「パイロット」や「テスト」という用語を用いた方が良いのではないか。リアルなデータではないということをはっきり認識してもらうために、名称についてはもう少し検討をした方が良いのではないか。(玄田委員)

- ・ 擬似マイクロデータと言うかレプリカデータと言うかは別として、このようなデータを開発したそもそもの目的として、一つには、訓練用データとしてフォーマットも含め同一の

データを作成するということがあり、もう一つの目的としては、授業に用いる教育用データとして作成するということがあり、この場合には、ある程度それらしい集計量が出ればそれで良いため、もっと変数や件数をコンパクトにしたデータでも良いのではないかと考えられる。さらに、レプリカデータの本当の目的として、研究のテストデータ、プログラム送付集計型のテストデータとしての使用が想定され、この場合、データは詳細であった方が良い。こういった形態での住み分けができれば一番良いが、最初から全ての変数の分布を再現しようとするのは目指さない方が良いと思われる。(安田委員)

- ・ 匿名データに対しては、全て対応するレプリカデータを作るということが大事ではないかと思う。匿名データの利用の前にレプリカデータで事前検証するということが、匿名データの正確な利用につながると思われる。(玄田委員)
- ・ 統計法上の匿名データの定義・解釈について質問するが、いわゆる「事実上の匿名化」のデータは匿名データであるが、「絶対的な匿名化」のデータは匿名データではないと考えてよいか。(県委員)
- ・ 統計法上は匿名化の程度で区別しているわけではなく、調査票情報から直接作成されたマイクロデータは匿名データとして捉えることになる。一方で、擬似マイクロデータは集計表から作成するものであり、統計法上の利用の制限等と切り離して考えることができるのではないかというのが論点としてある。(事務局)
- ・ 擬似マイクロデータについては、「絶対的な匿名化」のデータというわけではなく、全くの虚構データと考えてしまった方がわかりやすい。また、現行の匿名データについては、「事実上の匿名化」に該当する。(安田委員)
- ・ 日本の匿名データについては、ドイツで言うところの SUF に相当するものと位置付けられる。(事務局)
- ・ 事業所・企業を対象とした統計調査については、おそらく匿名データは作成できないであろうが、事業所・企業を対象とした擬似マイクロデータを作成する意義についてはどのように考えるべきか。(廣松座長)
- ・ できれば、作成する方向で研究を進めてほしい。(椿委員)
- ・ 名称については、当面、今の擬似マイクロデータを仮称として用いることとして、内容については調査票情報とは切り離されたものであり、現行の制度上で問題はないという研究会としての合意がとれたものと考えて。その上で、今後、どのようなものを作成していくべきかという点については、この研究会等で意見を聞いて提言できればと考えている。調査票情報を基にして作成すべきか、又は匿名データを基にして作成すべきかという点も論点かと思う。また、すでに多く利用されており反響もあるようだが、今後も潜在的ユーザーも含めて利用者のニーズも検討していく必要があるのではないか。(廣松座長)
- ・ 匿名データを基にした擬似マイクロデータを作成したことがあるが、その際は、あくまで教育用データを想定したものであり、擬似マイクロデータの利用方法の可能性を広く考えた場合には、方向性が若干異なってくるのではないかという印象がある。(小林調査官)
- ・ 統計センター以外の者が作成を行う場合のことも今後検討する必要がある。その他にも意見があれば事務局に連絡していただきたい。(廣松座長)

○ 次回の研究会は、来年2月中旬以降、年度内目途に開催予定。

以上