

第13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7月23日（月） 10:00～12:00
- 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）
伊藤政策統括官、千野統計企画管理官、浜東調査官
小林調査官（統計研修所）
《オブザーバー》
内閣府（統計委員会担当室）、総務省統計局、財務省、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（佐藤管理官補佐ほか1名）
- 4 議題：(1) 公的統計マイクロデータのオンサイト利用について
(2) 統計データの二次的利用促進に関する研究会 平成23年度報告書（案）について

5 議事の概要及び意見等

○ 議題1 諸外国における二次的利用の現状について

小林調査官（総務省統計研修所）から、資料1「公的統計マイクロデータのオンサイト利用 その現状と課題」の説明が行われた。

（主な意見、質疑応答）

- ・ 資料4ページに施設利用時に管理者の同室が必要とあるが、管理者とは具体的にはどのような者か。（廣松座長）
- ・ 一橋大学経済研究所附属社会科学統計情報研究センターのマイクロデータ分析セクションの准教授及び助教授の2名であり、私自身もそうであったが、総務省からの出向者である。（小林調査官）
- ・ 利用実績はどれくらいか。（廣松座長）
- ・ 現在の運用は試行的なものであり、統計局との共同研究として行っているものである。技術的な点も含めて様々な検証を行う必要があることから、共同研究者というかたちで、2名に利用してもらっている。試行運用の中で、共同研究者には施設運営者側の管理下であえて情報セキュリティ上の不正な操作を試みてもらい、それに対して技術上どのように手当てをすればよいか、また運用上利用者にとってどのようなやり方がよいのかという検証も行うことにしている。特にお願いして研究を行っているものであり、利用者と言っても一般公募ではない。（小林調査官）
- ・ 利用者は2名だが、単純に実績が2件あるというよりも、むしろ、リスク管理を含めた技術的な実践研究を行っているということか。（廣松座長）
- ・ そのように理解していただきたい。かつて、旧法時代に実施されたマイクロデータの試行的提供では、公募という形で協力者を求めていた訳だが、私自身の考えでは、まだそこまでの段階に至っていないと考えている。（小林調査官）
- ・ インターネットで検索すると、大学共同利用機関法人情報・システム研究機構の統計数

理研究所でもオンサイト利用に関する取組を行っているようであるが、実績はどのように
なっているか。(縣委員)

- 今年の3月31日に統計センターから認定を受けたが、公的統計についての利用実績はない。厚生労働省のレセプトの匿名データのオンサイト利用について先月許可が下りたので、今年9月以降、島根大学医学部の利用が予定されている。(樫委員)
- ホームページの記載を見ると、既に本格運用しているとの印象を受けたが、特定の目的に限定した試行の段階という理解でよいか。(縣委員)
- そのとおり。本格運用はまだである。ホームページの記載は詳細が省略されている場合もあり、必ずしも真意が伝わらない面があるかもしれない。(小林調査官)
- オンサイト利用について、今後、試行段階から本格運用の段階になったとき、資料8頁にあるような今までできなかった探索型の分析(事後チェック方式(調査事項の選択が試行錯誤的に自由にでき、施設外に成果物を持ち出す時に使用した調査事項・作成した統計等の承認を得る方式)の下での研究)というものには、非常に可能性があると感じている。

厚生労働省のレセプトデータに関して、昨年度、探索型の研究が何件か提案されたが承認されず、このことが逆にレセプトの匿名化情報の利用の方向に動き出したという経緯もある。当時、探索型の研究は認められなかった。当初は、国の政策的なテーマの共同研究に限定していくということが望ましいものと考えているが、いずれは、政策的な共同研究の中でも、オンサイト利用であれば色々と発見的なことができるという仕組みにつながっていけば学と官の距離が近づくこととなり、期待している。また、オフサイトの利用形態について今後どのようにしていくかということは、既に利用者もおり、大変であると思うが、ある程度オフサイトでできることについての制限が必要であると感じている。(樫委員)

- 現在、統計センターのサテライト機関(連携協力協定を締結した学術研究機関)は4機関あるが、そのうち、法政大学と神戸大学については、オンサイト施設を利用した調査票情報の提供は行っていないという理解でよいか。(廣松座長)
- そのとおりである。オンサイト施設として、統計センターの技術的な認証を受けているのは、一橋大学と統計数理研究所の2つである。(小林調査官)
- 補足であるが、(事務の委託を受けた)統計センターから技術的な認証を受けるということと、(調査実施者である)統計局の承認を受けることは別のものである。

また、現にオフサイトの利用形態がある中で、オンサイト利用ならではの特色がないと、オンサイト利用へのインセンティブが働かない。オフサイト利用について制限をかけるか、又は、諸外国のように調査票情報についてオフサイト利用は一切認めないというかたちで線引きをきちんとしない限り、利用者側からみた場合の探索型の定義を明確にすることは難しいのではないかと思われる。

現状の一橋大学の取組では、ただ単に利用者個人の研究室の代わりにオンサイト施設を利用でき、利用場所の審査等を受けなくてよいという程度のメリットしかない。そのためにそれをどうクリアするか、何が問題であるかという洗い出しをしたいというのがこの試行の目的であり、希望者が直ちにオンサイト利用できるというかたちにはなっていないものである。(安田委員)

- 旧法下においては、調査票情報の目的外利用は、統計センターではなく、政策統括官の承認がないと行えなかったが、統計センターが一括して申請する形でマイクロデータの試行が実施されていた。法制度的に課題がある側面もあったが、調査票情報ではなく匿名化さ

れたデータを使っていたため法令違反ではないとされた。新法になり、各調査実施者の判断で調査票情報の提供が行えることとなったが、現在のオンサイト施設の運用方法（統計センターと各サテライト機関との協定に基づいて実施する方法）は、法令上の根拠のない事実上の協定であり、本来であれば、統計局が締結すべきことを統計センターが行っており、適当でないと考える。一橋大学の例では、両者共同協定となっており、二次利用の申請を一橋大学の名で行っていただいているのでよいが、そうでないとすると、統計法上の秘密保護規定がかからなくなってしまう。安田委員が言われた様に、技術面の課題と法制度面の課題を解決する必要があるが、現状では、法制度上想定しているオンサイト施設はまだなく、あくまで試行の段階であると認識している。

これらの課題の解決をしないままに、事実上のオンサイト利用が広がるというのは問題があり、オンサイト利用についてどうあるべきか、早急にこの研究会で議論してもらいたい。

また、オーダーメイド集計であれば、ある程度詳細に集計したものを自由に使うということも考えられるし、匿名データとして整備されたものであれば、調査票情報の利用と同程度の厳しい規制が必要であるのかということが考えられる。新法制度下での三つの利用形態に応じた安全管理の在り方というものを考えていかないと、安田委員が言われるとおり魅力のないものになってしまうので、この辺りについて十分に議論する必要がある。

次に、報告書には書いていないが、本来は、統計局か統計センター内においてオンサイト施設をつくるべきと考える。サテライト機関での整備が先というのは本末転倒ではないか。予算がないという事情はあるかもしれないが、諸外国でも自らの出先機関を用いて行っているものであり、民間等に委託をやたらと広げていくということではないと思う。本来は、統計委員会等から、政府全体の方針として、まずは統計センターに整備すべきということが示されるべきではないかと思われる。また、統計センターにつくった場合、当初は、委託を受けて自ら調査票情報を有している統計調査しか利用できないが、各省それぞれで、オンサイト施設の運用を行うのも無駄であるから、一つにまとめていくことが正しいやり方だと思われる。分散型統計機構における調査票情報の二次的利用の在り方について考えなければならない。法制度上、個々の統計調査についての利用の承認申請は各調査実施者に対して行う必要があるが、オンサイト施設の運用という点では、各省が統計センターやあるいは同一のサテライト機関を共通の委託先とすることが考えられる。

また、オンサイト施設には、セキュリティ対策等に相当の費用が発生するが、施設側の持ち出しだけで費用を負担することになると長続きしない。調査票情報自体の利用に対してはお金をとれなくても、施設の利用料として徴収するという考え方もあるのではないか。こうしたことなど、制度上サテライト機関として認めた場合にはどのようにするのか、まだまだ研究し、決めていかなければならないことが多い。

また、オンサイト施設へのデータ移送については、一般公衆回線の場合データが盗まれる恐れもあり、現状は郵送で行っており、オンラインでは繋がっていない。こうした点も検討する必要がある。

なお、旧法の時代にはICT技術が発達しておらず、調査票情報も手書きであり、必要事項以外を墨塗りして提供するとなると、その作業にかかる行政側の負担も非常に大きく、信頼関係の下、データを丸ごと貸すこともあったが、こっそりコピーをとられる可能性はゼロではない。そのようなデータが外部へ流出したという話は聞いていないが、危機管理

としては、学者の良心にまかせるだけでは不十分である。昨今は民間企業の顧客データの流出事故が見られるが、万が一、統計データが流出した場合、統計調査に一切協力されなくなってしまう。この点については、非常に大事であり、提供する統計データの項目や範囲はきちんと制限するという点で議論していただきたい。(政策統括官)

- ・ 政策統括官から、大変重い課題を示されたものと認識。私も、説明資料の3ページでは、「(独) 統計センターとの連携協力協定に基づく」とされている一方で「統計局との共同研究と位置付け」とあり、ねじれている様な印象を受ける。また、これだけの施設・設備を提供し、維持管理していくのには、かなりコストがかかるはずであり、オンサイト利用について検討する上で重要なポイントと思っている。今日の段階では、これまでとし、今後も議論をしていきたい。(廣松座長)

○ 議題2 諸外国における二次的利用の現状について

事務局から、資料2「統計データの二次的利用促進に関する研究会 平成23年度報告書(案)」の説明が行われた。

(主な意見、質疑応答)

- ・ 本研究会の前身である「統計データの二次利用促進研究会」(旧研究会)でも、報告書(旧報告書)を出しているが、それとの対比において違いや重要な点は何があるか。(縣委員)
- ・ 旧報告書においては、主として、改正統計法の施行に向けて、ガイドラインを策定することや施行令・施行規則をどのように整備していくべきかということが中心であった。一方、改正統計法が施行された後に、統計委員会やIT戦略本部等の各方面から様々な意見・要請が行われており、今回の報告書では、これらの課題にどのように対応していくべきかということを中心に記述している。(事務局)
- ・ 旧報告書での指摘事項については概ね収束し、それが実現した上で、次の段階に入り、現在の研究会において新たな運用上の問題について検討したということによいか。つまり、旧研究会での課題は、大体実現できたということか。(縣委員)
- ・ 旧研究会については、統計法の改正により、調査票情報の提供のほかに、オーダーメード集計及び匿名データの制度が新設されたが、運用開始に当たってガイドライン等の整備が課題となっていたため、これを検討したもの。ガイドラインを策定したことで基本的には対応が済んだということなる。

一方で、現状はというと、匿名データについては技術的な課題があり、遅々としてであるが、進んでいるところである。各方面からの要請に関しては、IT戦略本部関係は、主としてもっと利活用を図るようというものであるが、一方では、調査票情報自体を自由に利用したいといった意見を言われることもあり、国民の信頼の下に成り立っている統計制度を維持するためにも、そのような要望には到底応じられないという説明をしてきているところである。

また、議題1で取り上げたところだが、オンサイト利用についても、まだまだ整理ができていない段階である。

オンサイト利用や擬似マイクロデータなどについては、新しい課題であり、旧報告書においてはその他の留意事項として挙げられているものの主要な論点とはなっておらず、どのように対応すべきかについては、現在の研究会で検討する必要があるというのが流れである。(政策統括官)

- ・ 半分余談だが、旧研究会では、「二次利用」という用語を名称に用いていたが、法制度上は、当該用語は、限定された意味（調査票情報を直接利用する形態を示し、オーダーメイド集計や匿名データを含めていない）となる。オーダーメイド集計や匿名データも検討の対象であることから、研究会を再開するに当たり、より広義な概念となるよう「二次的利用に関する研究会」に名称を変更している。
- ・ 外国の例でパブリックユースファイルと言うものがあるが、日本にはない、ただ、調査をしてみるとこれに対するニーズはあまりないかもしれないという言及があった。他方、擬似マイクロデータというものがある。内容的によくわからないので確認したいが、擬似マイクロデータは教育目的で利用されているとのことであるが、作成上は似ているものであるか。もし、そうであるなら、擬似マイクロデータを、海外で言うパブリックユースファイルとして提供を拡大する意義はあるのではないかとと思われる。（県委員）
- ・ パブリックユースファイルというものは、例えば国勢調査の中から一定のルールでデータを抽出し、また、所得が極めて多いなど特異的なデータを除外するなどの匿名化処理を大幅に施したものである。しかし、事実と異なるデータとなるので、どのような処理をしたかわからないままそれを用いた研究成果については、科学的な証明力がどこまで認められるのかという議論もある。一方、擬似マイクロデータは、例えば国勢調査で一調査区単位で全ての詳細集計を行ってしまうもの。だが、私としては、作り方によっては集計結果と言えなくなってしまうのではないかと疑問を感じており、教育用擬似マイクロデータについて、統計局は詳細集計結果であるとの説明をされているようだが、実際にどのような処理を行っているのか詳細を確認してみないと判断できない。場合によっては、集計結果を基に作成したデータと言える場合もあれば、匿名データの一つであるパブリックユースファイルと言うべきものである場合もあり、今のところは、位置付けは明確になっていない。

その意味で言うと、報告書の案文について修正した方が適切と考えている箇所があり、31ページの記載について、「個人・法人等が特定されるおそれはない」としているが「…少ない」とし、「利用目的に制限なく利用することができる」については、そのように「考えられる」に修正したい。

皆の共通理解として確定的な位置付けは決まっていないが、一般論として、外国の例のイメージで言えば、大幅な匿名化処理をしているため、科学的な証明としてでなく、大学等での統計演習など統計教育目的に使用するのであれば問題はないということになる。しかし、技術進歩は常に進んでおり、擬似マイクロデータについても極端な話、子供がたくさんいるところだとか、職業で言えば、漁民ばかりのところや一人公務員がいれば個人の特定がされてしまうおそれがあり、理屈で言っている以上に非常に作成方法が難しいものと考えている。こうした点について統計委員会への説明をしないまま取組を進めるのは問題があるのではないかと考えている。（政策統括官）
- ・ 擬似マイクロデータという用語ないし概念は新しいものであり、二次的利用として法制度的には、第33条第二項に基づく調査票情報の利用、第34条の委託による集計（いわゆるオーダーメイド集計）、そして第35条および第36条による匿名データ、があるのみである。擬似マイクロデータが匿名データと、どの程度同じ、あるいは異なるかという点については政策統括官が言うように、まだ公式見解がある訳ではない。また、パブリックユースファイルとの関係でもまだ整理ができていない。（廣松座長）
- ・ そうした位置付けが決まっていないものについて、報告書にあえて記載することの意義

についてだが、議論の喚起や精緻化を促すためという理解でよいか。(県委員)

- ・ 本研究会のまとめとして、新しい方向性の一つを示すという意味で書いておくべきと判断している。(廣松座長)
- ・ 法制度上は、調査票情報の二次利用、オーダーメイド集計、匿名データ、そして、当然のこととして明記されていないが、集計結果の利用の4類型しかない。擬似マイクロデータについては、作成方法をよく確認し、新しく第5の類型とするのか、4類型のどれかに含まれるものであるのか整理をする必要がある。

なお、擬似マイクロデータは、小規模の調査では作成できず、統計局が一部の大規模調査について取組んでいる状況である。報告書でも言及しているが、統計調査の性質、対象、規模に応じて、本当は一つ一つ丁寧にみる必要があるのに、十把一絡げで議論されがちな風潮が感じられる。

新しい技術進歩による取組については、それが役に立つことであれば大いに結構なことであるが、法制度を無視して進めていくことはできないものとする。(政策統括官)

- ・ 擬似マイクロデータについては、これから検討するという段階であり、これをパブリックユースファイルとするかどうかは、その後の議論であると思う。

以前、統計審議会の時代にデータリンケージ関係の研究会があった際、その中で擬似マイクロデータを作成したが、そのデータ自体を公開するという段階には至らなかった。

先程話があったとおり、擬似マイクロデータの作成方法については様々なバリエーションがあり、このため検討が必要なのは言ったとおりだが、リアルデータ(調査票情報)そのものから匿名データを作成する場合と比較して、ある種のリスクが回避されるというのは事実である。(樫委員)

- ・ 擬似マイクロデータとは少し性格が違う話となるが、いわゆる擬似パネルデータというものも新しい提供の仕方としてはあり得る。(廣松座長)
- ・ 報告書案において、二次的利用に関する民間の意見について、意見に対するコメント(考え方)を記述しているが、前回の研究会の資料では記述していなかったものであり、意見があればお伺いしたい。(政策統括官)
- ・ 29ページから30ページにかけてのオンサイト利用についての「ニーズはないのではないか」という意見については、あえて記載する必要はないのではないか。おそらく、オンサイト利用についてニーズを有しているのは民間企業ではなく研究者ではないか。(廣松座長)
- ・ オンサイト利用のニーズについて、否定的な意見が民間企業からあったのは事実としてある。(政策統括官)
- ・ 民間企業へのヒアリングの中で意見があったものなので、事実として記載するのはやむを得ないと考えるが、日本の現状について率直に申し上げると、民間企業にせよ研究者全体にせよ、モデリングやデータ解析技術の力量が、海外と比較するとまだ常にレベルが低い。

こうした状況がある中で、ある種きちんとした利用ができるデータがあれば、技術力も促進されるという面があるのではないかと。ヒューマンリソースの面について、諸外国と異なっており、米国では統計的なモデリングができるいわゆるスタティスティシャンと呼ばれる者が、大学学部卒業レベルで、2万5千人が職業を持っているということがある中で、そういうことについても考えていく必要がある。(樫委員)

- ・ 報告書のスタイルとして、どの項目についても「こうした意見やこうした状況にある」

ということを記載しているが、委員の趣旨を踏まえるなら、民間企業のこういう意見の背景として、「利用能力が十分でないから」と記載した方がよいだろうか。(政策統括官)

- ・ ということ自体を言いたいわけではないが、ただ、率直に申し上げて、まだまだ、民間のアナリストと言われる方々でも、集計結果を利用して分析をやっているというのが現状である。便宜の提供として、公表集計結果の充実やオーダーメイド集計の実施は、貴重でありありがたいことだと思うが、一方でこういった側面があることも事実と感じている。(樫委員)

- ・ 例えば、23 ページを見ても、民間企業には十分な利用能力がないことが見て取れる。(廣松座長)

- ・ 背景として、民間企業において利用能力が十分でないということがあるが、当該記載箇所へ追記することが適当か検討したい。(政策統括官)

- ・ 28 ページの「二次的利用の推進に向けた取組の方向性」における「基本的な考え方」の③はあえてここで明示する必要があるのだろうか。念を押ししたいという趣旨か。(縣委員)

- ・ 基本的な考え方として①から⑤について書いておきたい。ただし、③については、今の案文は言葉足らずなので「様々な統計調査があり、また、行政情報により作成される統計もある。」のあとに「一概に論ずることはできず、それぞれの特徴を踏まえつつ議論する必要がある」旨の文章を追記する必要があると思われる。

統計調査は、国の行政機関が行っているものだけで年間で200件以上も実施されており、議論の際に、誰がどのように作っている統計であるか知らないままに十把一絡げで議論されるケースが多く、非常に危惧している。(政策統括官)

- ・ 今の政策統括官の説明と逆で、③の案文では、「すべての統計調査が使えない」というように見える。(縣委員)

- ・ 個人・法人等の秘密保護が欠ける場合はすべて利用が認められないというのは、書いてあるそのとおり。ただ、先程言ったように言葉を書き分ける必要があり、上の三行(前半部分)は、様々な統計調査があり十把一絡げの議論は良くないということであり、後半部分の秘密の保護が必要であるということは全部の統計に通じる話である。(政策統括官)

- ・ そうであれば、あえてここで明記するべきかということが考えられる。(縣委員)

- ・ 明記した方が良いと考える。一部では、かなり乱暴な議論がなされている。(廣松座長)

- ・ そういう危惧があるということであれば、了解した。(縣委員)

- ・ 例えば、IT関係では、ビックデータの話など、壮大な構想もあるが、この辺のことは、ほとんど抜けているというのが現状であり、統計の分野としては念を押ししておくべきと考えている。(廣松座長)

- ・ 補足だが、経験上、社会学者の方は個人を同定した分析をやりたいと言われることが多いが、統計的分析でなく、こうした個人を特定した社会的分析に用いることは認められないということを説明してきている。こうしたこともあるので、私も書いた方がよいと思う。(安田委員)

- ・ 統計と統計調査という用語が混在しているが、整理をした方がよいのではないか。使い分けについて、もう少し精査した方がよい。(廣松座長)

- ・ ③の記述については、「これらすべての統計を作成するための調査は」が正しいかと思う。(政策統括官)

- ・ 同じく28ページの「基本的な考え方」の④について、「統計の誤用を防ぐために」とあ

るが、統計自体の誤りというより、その利用により国民等利用者に誤解を与えることを避けなければならないという趣旨が重要である。(樫委員)

- ・ 「統計の誤用を防ぐために」の前に「国民に誤解を与えるような」と追記する等の修文をしたい。(政策統括官)
- ・ 本報告書が決定した後は、どのような運びとなるか。(縣委員)
- ・ 本報告書は、23年度報告書として、24年度以降に取り組むべき事項を整理したものであり、本報告書の取りまとめをもって研究会が終了するというのではなく、引き続き研究会を開催し、残った課題について議論していただくこととなる。(事務局)
- ・ 直近としては、本報告書について、統計委員会へ報告するため、そこで議論があるかもしれない。
- ・ e-Stat の充実や二次的利用の対象となる統計調査の拡大など具体的な施策が挙げられているが、具体的な具現化はどのようになるのか。(縣委員)
- ・ e-Stat については、今後、研究会の間でも統計局から説明していただくことも可能であるし、オンサイト利用のガイドラインに関しては、当室で案を作成の上、説明し、今後ご議論いただくことになる。(政策統括官)
- ・ 報告書を受けてすぐに動き出すということではないという理解か。(縣委員)
- ・ 事項によって濃淡があり、二次的利用の対象統計調査の拡大は、比較的、技術的な課題のみであり粛々と進めることができるであろうが、オンサイト利用や擬似マイクロデータについては、まだ制度面での整理を含めてかなり詰めなければならないところがあるので、研究会で議論し提言をし、統計委員会やあるいは一般から意見をいただくということになるのではないかと考えている。
報告書については、今回の意見を踏まえて修文をした上で公表したいと思うが、具体の修文については、私(座長)に一任してほしい。勿論、確認のため、公表前に、改めて各委員へお示しする。(廣松座長)
- ・ 次回の開催日程については、事務局から改めて連絡する。(事務局)

以上