

第15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3月26日（火） 15:00～17:15
- 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、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
《説明者等》
（三菱総合研究所）佐藤研究員、森研究員、（総務省統計研修所）小林調査官
《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、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) 諸外国における二次的利用の状況について
(2) 前回の議論を踏まえた論点整理について
(3) 研究会における今後の検討について

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状況について

三菱総合研究所（森研究員）から、資料1「現地ヒアリング調査（ワシントン DC、オタワ）報告」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）

- ・ NCHS（アメリカ国立保健統計センター）は必ずしも人的リソースが十分とは言えない。チェック体制について、どれだけの人的、予算的規模であるのか。（玄田委員）
→ NCHSは詳細には把握できていないが、外部の人間も呼んで会議を行っているとのこと。アメリカセンサス局を中心に調査したので、その内容をご報告する。アメリカセンサス局においては、チェック機関であるDRB（Disclosure Review Board：公開審査委員会）が9名のセンサス局員によって構成され、週1回会合を行っている。議題は毎回10程度あり、新たなPUF（Public Use File）の公開等となっている。センサス局はかなり明確なチェックリストを作成し、チェックを行っている。（森研究員）
- ・ 不承認（Reject）となった事例はあるのか。（玄田委員）
→ 相当程度ある。また、条件付き承認となる場合もある。不承認となるのは、個人の特定のおそれがある場合などである。（森研究員）
- ・ DRBでの判断基準は秘匿性のみか。（縣委員）
→ DRBのチェックは秘匿性の観点からのみとなる。^{（注）}（森研究員）
 - ・ （注）DRBによるものではないが、利用申請時のレビュー（研究プロジェクトについての公益性、実行可能性、RDC利用の必要性等の観点からのチェック）についても、別途、研究テーマや利用するデータセットに関係する内部・外部専門家が参加して実施されている。
- ・ カナダ統計局の事例において、RDC（Research Data Center：オンサイト施設）となる大学の選定はどのように行われるのか。（玄田委員）

- 基本は立候補である。設置基準や経済的側面（5年間施設を維持するための基盤が十分に整っているか）、ニーズなどがその後の選定要件となる。経済的側面については、州政府からの補助金やパートナーシップを始めとするサポートが不可欠であることから、大学も主要なところに自然と限定されてくる。（森研究員）
- ・ カナダのRDCが設置されている大学の構成や設置数はどのようになっているか（公立、私立の別など）。また、そもそもの背景として、RDC設置校に限らず、カナダにおける大学の公立、私立の別はどのようになっているか。（縣委員）
- 大学全体の公立、私立の構成割合については把握していないが、RDC設置校については、公立のいわゆる有力大学が多い^{（注）}。縮小版RDCであるサテライト機関も存在し、これについては中規模なところもある。それなりの規模がないと運営も難しいようである。また、設置数については、サテライト機関も含め27箇所（カナダ統計局・同支局に設置されているものを除く）である。（森研究員）
- （注）
- ・ RDCのうち26箇所は大学に設置されているが、すべて公立大学である。これは公立大学を重視して選定した結果ではなく、カナダの大学の多くは公立校となっていることによると考えられる（カナダ大学協会に加盟する97大学のうち、私立大学は10大学程度）。
 - ・ アメリカセンサス局のRDC設置数は15箇所、NCHSのRDC設置数は20箇所（多くはアメリカセンサス局保有のRDCに相乗り）。
- ・ カナダにおいては大学のコンソーシアムにおいてデータ提供について審査を行う仕組みがあったと記憶しているが、現在でも機能しているのか。（安田委員）
- 大学はCRDCN（Canadian Research Data Centre Network）というネットワークに入っており、これがご指摘のコンソーシアムに該当すると考えるが、審査機関には当該コンソーシアムは含まれていないと思われる。（森研究員）
- ・ プログラム送付型の二次的利用については、回帰分析等も許容しているのか。（樫委員）
- そうである。（森研究員）
- ・ カナダにおけるリモートアクセスについては、データの部分的なダウンロードも許容されているのか。（樫委員）
- ダミーデータ等の参照のみが可能であり、本物のデータのダウンロードはできない。また、使えるデータセットについては限りがある。そのため、それ以上のものがあればRDCに行かなければならない。他方で、米国センサス局は公開されるPUFかRDCかしかないので、リモートアクセスはその間を埋めることになるだろう。カナダのRTRA（Real Time Remote Access）は、まだ取り扱っているデータセットの種類が少ない。（森研究員）
- ・ 事業所系のPUFの事例はあるのか。（樫委員）
- 平成23年度に行った文献調査では事例があるとされていたが、今回のヒアリングで確認したところ、研究段階であり公開には至っていないと聞いている。（廣松座長）
- ・ データにアクセス可能な人は国内に限定されているのか。（縣委員）
- OECDの専門家会合の場で聞いたところによると、カナダにおいては、リモートアクセスは海外からのアクセスも排除されないとしているが、RDCは国内の人間に限定されるとのことである。（事務局）
- 国籍も問われないとのことであるが、訴追の可能性も考慮して各種要件を利用のために

課しており、結果的に国内の人に限られてしまうというのが実態である。(森研究員)

- ・ 利用のための教育は必修か。今後日本における運用を想定し、参考にしておきたい。(玄田委員)
- 必修である。主に秘匿についての知識であり、そのほかにもシステム関係の研修も行われる。研修は各RDCで受けることになる。シンクライアントのアカウントセットアップも各RDCで行われるため、利用に先立ってRDCに足を運ばなければならない。(森研究員)
- 関連してだが、情報の漏えいがあった場合、大変厳しい罰則がある。(廣松座長)
- 統計データの不正利用自体でなく関連する詐欺行為等も含めての刑罰だが、収監の措置が実際に執行されている例もある。(森研究員)
- 米国センサス局においては、データに関する研修、データセキュリティ、スタッフの教育について力を入れているとのことであった。(廣松座長)

○ 議題2 前回の議論を踏まえた論点整理について

事務局から、資料2-1「前回の議論を踏まえた論点整理(オンサイト利用)」及び資料2-2「前回の議論を踏まえた論点整理(擬似マイクロデータ)」の説明が行われた。また、それぞれの説明後に、意見交換及び質疑応答が行われた。

(主な意見、質疑応答)

【①オンサイト利用について】

- ・ 統計法第33条のフォローアップとして、調査票情報の提供を受けた者を対象とした監査を実施しており、管理状況について改善を要するような事例も確認されている。問題意識としては、実地調査とは言え、利用者側の申告・説明に基づいた検証となる面が大きいことは否定できず、管理の徹底には限界がある。特に悪意のある者、虚偽の報告や隠ぺいを図ろうとする者に対しては、十分に対応できないのではないかと考える。(総務省統計局)
- ・ 一橋大学におけるオンサイト利用の試行運用状況についても報告する。現在2名の研究者の協力を得て試行運用を行っているが、まだ研究途上の段階である。そのため、最終的な報告ではないが、現時点で、研究者に対しヒアリングした際に出された意見を紹介する。
まず、外部から完全に遮断されているため、プログラムのマニュアルを参照することができず不便であるとのことであった。また、公表資料との整合性についてもチェックできないとの意見があった。e-Stat等が閲覧できるPCを別途用意してほしいとのことである。入退室については、複数回チェックしてはどうかとのことであった。これについては、常駐する職員も統一された基準でチェックが今後必要ではないかとの意見もあった。作業環境については、PCによる作業以外の作業スペースもほしいとのことであった。また、データの出力や持ち出しの審査については、必要であるとの認識であった。(総務省統計局)
- ・ オンサイト利用の前提として、統計法第33条との関係はどのように考えているのか。(縣委員)
- 統計法第33条の運用手段として考えているということである。現在はCD-R等の媒体に調査票情報の必要部分をコピーして、研究者の手にデータを置くことを許している状況であるが、これをオンサイト利用に代えることでセキュリティ上安全な運用が行えるということが期待される。(事務局)
- ・ オンサイト機関で供されるデータの秘匿性はどのようになっているか。(縣委員)

- 統計法第 33 条の対象である調査票情報と同等の内容である。(事務局)
- ・ アメリカ合衆国やカナダのように一定の秘匿処理がなされているデータの利用方法ということではないということか。(縣委員)
- 委員のご指摘は、秘匿性の低い(加工処理の度合いが低い)匿名データの利用方法についてのもと思われるが、そうした点も含めて今後検討を行っていきたいと考えている。(白岩管理官)
- ・ 過去長い経緯があり現行の利用方法となっているが、このまま続けてよいのかという論点がある。また、オンサイトでの利用によって、異なる調査間のデータリンケージの作成と活用が進む可能性もあり、これは大きな利点となるであろう。(廣松座長)
- ・ 分析プログラムの共有については、分析結果の復元可能性や再構築の観点から重要である。他方で、プログラムの最終的な所有権はどうなるのか。各種学会においてはプログラムの提出を求められることがあるので、所有権が完全に研究者に帰属されるのか、オンサイト機関の運営側も関係するのかが気になるところ。(玄田委員)
- ・ 本事項について意見出しをした立場として考え方を述べるが、所有権はあくまで一義的には研究者にあるが、研究者の同意の下、他の者でも活用できるようにするという趣旨。学会においてデータやプログラムの提出が要求されることは確かにあるが、再現性の保証が大きな目的であると承知している。(安田委員)
- ・ 学会における審査を前提にしたプログラムの公表ということであるが、秘匿性について問題のある研究は行われていないという保証にもなるのではないか。(椿委員)
- ・ 走りながら考えないといけないが、秘匿性について事故(調査対象の秘密漏えい)があると大変なので、様々なケースを想定して、慎重に検討をする必要がある。(廣松座長)
- ・ 日本での今後の運用主体や運用形態については、どのようにするか決まっているか。(縣委員)
- 現時点では、方針が決定しているものではない。(事務局)
- ・ 現時点で詳細を全て決めておく必要があるという趣旨ではないが、目指すべきところを見据えて議論することが重要である。アメリカ合衆国やカナダの提供形態にはそれぞれメリット・デメリットがある。日本の行政構造や文化的要素を見て利点の多いものを選択していただきたいと考える。(縣委員)

【②擬似マイクロデータについて】

- ・ 取扱いや解釈に極力グレーゾーンを作らないことが重要。擬似マイクロデータはリアルでもリアリスティックでもなく、フィクションであると整理すると明解になる。そのため、教育用・訓練用には使えるが、研究用ではないことを明記すべきではないか。一方で、擬似マイクロデータ作成者に用途を限定させる権利を与えることをはっきりさせると、論点として挙げられている多くの課題は解決するのではないかと考えている。(玄田委員)
- ・ 研究用の用途については、あまり強調しない方がよいという趣旨か。(廣松座長)
- ・ そうである。(玄田委員)
- ・ 集計結果は再現するが、データ自体は異なるものとして擬似マイクロデータを作成するというのは可能である。当面はその運用をどうするかということを検討すれば良いのではないか。(椿委員)
- ・ 外国の PUF とは内容も性格もまったく違うものなのか。(縣委員)

- ・ 作成手法が異なっているという点があるが、求められる役割としてはP U Fに近いのではないかと思われる。(事務局)
 - ・ レプリカは秘匿化とはまったく違う措置。匿名データと異なり、レプリカデータは公表しても問題ないが、公表することの社会的意味を考えると利用をある程度限定するのは必要なのではないかと考える。(樫委員)
 - ・ 統計法第 33 条に基づく調査票情報の提供、匿名データの提供も含め、制度全体としては諸外国の二次的利用と同じ効果をもたらすのか。(縣委員)
 - ・ そのような理解で差し支えない。例えだが、擬似マイクロデータと匿名データ（リアルデータ）は、理科室にあるプラスチックの人骨と実際の人骨との違いとイメージしても良いのではないか。(廣松座長)
 - ・ 国がフィクションのデータを公表することについては慎重になるべきではないか。国や独立行政法人でもなく、民間が作ったものであれば誤解も生じないが、公的機関が作成したとなると、「実は本物のデータなのではないか」などと思われてしまうおそれがある。(玄田委員)
 - ・ 海外のP U Fは調査票情報を秘匿化して作ったデータであるが、擬似マイクロデータは、詳細な集計結果を合成して新たに作り出されるものであり、やはりP U Fとは別物と考えた方が良い。合成データであるという点を明確にしておけば良いのかもしれないが、それでも国が作成した場合は、国が数字の真実性を保証した印象が強く、玄田委員が指摘するような誤解を生じさせるおそれがある。一方で、合成して新たに作成したデータであれば、著作権は作成者側にあるため、民間が作成した場合にこの権利を国がコントロールできるのかという問題がある。そういう意味では、民間が自由に作成し、国としては関与しないという整理がすっきりしていると言える。(安田委員)
 - ・ 日本の擬似マイクロデータはフィクションのデータであり、諸外国のP U Fに該当するのは日本ではないと理解してよいか。また、現実を表している（集計結果が、調査票情報で集計を行う場合と近似した傾向を再現できる）が、データに個別性はなく秘匿性の観点から問題がないと言われる擬似マイクロデータを整備するか否かについて判断する必要が出ているということか。(縣委員)
- 日本の匿名データについては、P U Fに近い位置付けと考えて良いのではないか。一方で、擬似マイクロデータについては、利点もあるが、あくまで仮想的なものであるため国が作成主体になるべきかという論点が出てきたところである。(樫委員)
- ・ 論点を整理すると、(オーダーメイド集計は別として) 調査票情報の直接利用、オンサイト利用、リモートアクセス、匿名データに加えて、5番目の類型として合成データ（擬似マイクロデータ）を新たに制度的に位置付けるか否かという構造になっていると理解してよいか。(縣委員)
- そうした理解でよいと考える。(廣松座長)
- 昨年度も諸外国における二次的利用制度について文献調査を行ったが、その結果、P U Fについては各国ともデータの内容がまちまちであり、それに伴い利用制限の範囲も異なっている状況であった。日本における匿名データは一般の利用に供することを目的としており、その意味ではP U Fであるが、現在の運用では、研究目的に適したものとなるように匿名性（加工処理の度合い）は低く抑えている一方で、情報管理等の面から一定の制限をかけている状況。利用者の声を聞くと、確かに、諸外国のP U Fの様にもっと自由に使

えるデータも作成すべきではないかという意見もあるが、一方で、更により詳細な分析が可能となるようなデータの提供を望む意見も見られる。このため、日本における匿名データについて、PUF的なものを目指すのか、それとも現行の匿名データの種類の充実を図るのかという選択肢があるが、我々としてはどちらかと言えば後者を探るべきではないかと考えている。PUF的なものの整備を新たに図るよりは、(独)統計センターにより教育目的の利用のために試行提供されている擬似マイクロデータの活用や公表している集計表の充実を図ることに注力した方が、結果として、世の中のニーズ全体に早く対応することができるのではないかと考えている。(事務局)

→ 県委員の法的位置付けや概念整理についての問題意識は、当方も同様に思うところがある。また、擬似マイクロデータについては、外見的には匿名データと同様の面があり、玄田委員をはじめとする各委員の懸念は非常に重要なポイントと認識している。

いずれにしろ、匿名データに関しては、一義的には、現在提供されているもののような調査票情報に近く研究利用に適した分野での充実を進めるという方向性をもって取り組むべきであり、その上で、更に簡易な形での提供方法もあり得るのではないかと検討を行っている段階であると認識している。

統計法上の新たな利用形態を設けることについては慎重に議論する必要があるが、いわゆるPUF的なものについては、自由に幅広く使えるという役割が重要であって、利用者から見た場合には、それが匿名データや擬似マイクロデータであるかは本質的な問題ではないということに留意が必要である。(白岩管理官)

○ 議題3 研究会における今後の検討について

事務局から、資料3「統計データ・アーカイブ」の今後の検討方向について」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)

- ・ 基本計画では背景事情等について詳細が書かれていないが、本件は1960年代から70年代にかけてオープンリールで記録されたデータが読み込めなくなっていたことの反省を踏まえたものである。具体的には、もし各府省においてデータ管理ができないのであれば、一か所でまとめて行う必要があるという趣旨の日本学術会議の答申を受けて検討が始まったものと承知している。その観点では、今回まとめられた方向性で概ね間違っていないと思うが、検討の一助とするため、次回会合で当時の関係資料を配布してみてもどうか。(安田委員)
- ・ 今回のまとめで違和感はないが、並行して、データを寄託する側にとってのメリット・デメリットや課題も整理することが必要。おそらく一番大きなハードルになる点としては、事後的に何らかのエラー(データの不備)が発覚したときに責任が追及されるようなことになると、データの寄託に非常に躊躇すると思われるので、そうした点の整理についても検討すべきである。(玄田委員)
- ・ 統計データ・アーカイブの出発点は安田委員の指摘のとおりであるが、基本計画を策定した際には、調査票情報そのものの記録としてのアーカイブなのか、二次的利用のために提供したデータのアーカイブなのかは必ずしもはっきりしていなかった。本研究会ではその点についても議論して一つの考え方を示していきたいと考えている。(廣松委員)
- ・ 公文書管理法上の整理はどうなっているのか。(県委員)

- 公文書（行政文書）の定義は、行政機関の職員が組織的に用いるものとしてかなり広くなっているが、そのうち、公文書館に移管するのは「非現用」のものとして行政機関での利用が一旦終わったもので、更に公文書館で保管する歴史的価値を有するものを選択することとなっている。一方で、統計に関するデータは原則として、二次的利用への対応を含めて「現用」のデータとして持ち続ける必要があるものと考えられる。制度上あるいは運用上の整理については今後精査する必要があるが、仮に、概念的に公文書館に移管するものと整理されとしても、統計データについては、非常に大量な情報であり、かつ数的な機械処理が可能な状態で保管されてこそ意味があるデータであるため、現行の公文書館の機能で果たして対応しきれるかどうかなどという問題もある。（事務局）
- 本日は検討のための時間が十分にとれなかったため、今後の検討の方向性については、本日の委員の意見も踏まえ、座長と事務局とで再度検討した上で、次回検討会において改めて修正版として示すこととしたい。（廣松座長）

○ その他（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、次回会合は6月末又は7月に開催見込み。詳細については、追って事務局から連絡を行う。（事務局）

以上