

公的統計データとオープンデータの活用

Role of data for official statistics from the viewpoint of open data

椿 広計[†]
Hiroe Tsubaki

1. はじめに

Statisticには3つの意味がある。

Karl Pearsonによって、「科学の文法」として現象を実証科学的に記述する一連のプロセスを支援する「統計科学(Statistical Sciences)」としての意味が第一である。今日この意味での統計科学は「データ科学(Data Sciences)」とも呼ばれるようになった。第2は、統計科学の一分野である数理統計学で用いられる。すなわち、確率変数としての複数の観測変量の関数としての「統計量」という意味である。推定量や検定統計量が代表的な統計量である。そして最も古いのが、17世紀以来用いられてきた国勢の姿を量的に記述する「国勢学」としての意味である。国勢学においては、「統計は静止した歴史、歴史は動く統計」と標榜されている。Statistics自体の語源はこの国勢学に由来する。主要国は統計部局を集中型(カナダなど)ないしは分散型(日本など)で配置し、国勢の現状を把握してきた。すなわち、各国政府は国民から必要な情報を統計調査によって収集し、集計、推計あるいは加工することで、「公的統計」として公表してきた。

日本の公的統計に関わる法としての統計法は、平成19年に60年ぶりに改正され平成21年に全面施行された。旧法の目的が、「統計の真实性を確保し、統計調査の重複を除き、統計の体系を整備し、及び統計制度の改善発達を図ること」であったのに対し、新法の目的は、「公的統計が国民にとって合理的な意思決定を行うための基盤となる重要な情報であることにかんがみ、公的統計の作成及び提供に関し基本となる事項を定めることにより、公的統計の体系的かつ効率的な整備及びその有用性の確保を図り、もって国民経済の健全な発展及び国民生活の向上に寄与すること」となった。更に、国政全般や社会の意思決定にとって重要な公的統計は、基幹統計として指定された。国民は基幹統計調査には協力義務があり、虚偽の回答を行うことや回答を拒否することに対しては、50万円以下の罰金を課す規定が存在する。

筆者は、個人ないしは法人が政策の運用や社会全体の意思決定に資する正しい情報を納める義務こそ、今日的には所得税や法人税と同様の意義を持つものと考えている。何故ならば、情報を正しく納めないことによって、社会の意思決定はパレート最適なものとならず、機会損失が発生するからである。

特に、Risk and Evidence based Policy Makingの考え方の勃興により、政策に活用するための公的統計という位置づけは益々強化されつつある。国勢のあるべき姿と現状との乖離を問題としてとらえ、政治・行政の方針を定め、適切な政策運用を行うこと、Evidence based Policy Makingの本意であり、その中核として活用されるべき情報が公的

統計なのである。

本発表では、この公的統計を作成するために徴集されるデータ、通常マイクロデータと呼ばれる個人情報ないしは法人情報として秘匿される必要のあるデータについて、その活用の在り方と注意点を議論する。

2. 統計法改正のもたらしたもの

2.1 公的統計データの2次的利用と問題点

新統計法は、公的統計を作成するために収集した個票データ(マイクロデータ)を公益的な統計作成等の研究に利用する可能性を旧法に比して格段と広げた。個人情報ないしは法人情報を含むマイクロデータを統計法33条に基づき目的外利用する方法は、日本学術振興会の科研費など公的機関がその研究意義を認めた学術研究については旧法同様可能である。しかし、個人、法人の秘密情報を個人情報などを保持したまま、研究者にマイクロデータとして提供している国は、OECD諸国で我が国しかないことには注意が必要である。このためデータをどのように加工し、どのように集計結果を出すかについては、目的外利用を申請する時点で明確に記述する必要がある。目的外申請はデータ管理体制を中心に厳格な事前審査が行われるが、実際にデータが渡されたのちに、研究者がどのようにマイクロデータを管理・分析しているかを知るすべはない。

新法では、この統計法33条に基づく目的外申請以外に、基幹統計調査について、統計法35条に基づき調査票情報を加工した「匿名データ」を作成し、36条に基づき、一般の求めに応じて学術研究の発展のためにこの匿名データを提供することが可能となった。基幹統計を所管する府省が作成する基幹統計匿名データについては、その匿名化が個人や法人が特定できてしまうリスクが十分小さくなっているか、匿名化によってデータ提供の有用性が著しく低下していないかといった観点で、内閣府統計委員会が関連府省の諮問を受けて検討し、委員会に配置された匿名データ部会(審議や詳細議事録は匿名化方法を秘匿するために非公開)の検討を基に答申することとなっている。実際の匿名化にあたっては、調査データのリサンプリング、比率の少ない項目の併合あるいは削除、集計結果の変わらない範囲でのスワッピングなどの方法が用いた上で、オリジナルデータの分布が損なわれていないか否かを検証している。匿名データの提供を通じて、これまで公的統計利用では難しかった多変量データ解析など探索的方法による公的統計データの分析も可能になった。しかし、この種の匿名データを用いた実証研究を国際的学術誌がアクセプトするかについては疑問もあり、匿名データで絞り込まれた仮説を33条による目的外申請で検証するというのがあるべき姿である。

平成26年4月現在、総務省関係では国勢調査、労働力調査、住宅・土地統計調査、全国消費実態調査、就業構造

[†] 統計数理研究所 The Institute of Statistical Mathematics

基本調査，社会生活基本調査が，厚生労働省関係では，国民生活基礎調査の匿名データが提供されている。総務省に関わる匿名データの利用については，(独)統計センターとの連携協定を結び認可された，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マイクロデータ分析セクション，法政大学日本統計研究所統計情報ユニット，神戸大学大学院経済学研究科研究助成室データ管理室，情報・システム研究機構新領域融合研究センター統計数理研究所オンライン解析拠点の4機関が協力している。

もう一つの利用形態は，統計法 34 条に基づくオーダーメイド集計である。現在，公的統計の集計結果は政府統計の総合窓口としての e-stat (<http://www.e-stat.go.jp/SG1/estat/eStatTopPortal.do>) からダウンロードすることができる。データの集計結果の中には，「統計 GIS」のように統計調査データを地図上に表示する仕組みも提供されている。これに対してオーダーメイド集計は，政府が集計していない詳細集計を利用者の要求に基づいて行うものである。もちろん，集計頻度が著しく小さくなるセルについては，結果表示を行わないなど，この制度においても匿名化措置は行われている。

2.2 公的統計マイクロデータ活用の今後

新統計法施行以降，日本政府は内閣府統計委員会が中心となり検討を進め，「公的統計の整備に関する基本的な計画（基本計画）」を5年ごとに策定し，閣議決定することとなった。

ここでは，2次利用をはじめとする公的統計データの有効活用についての計画も策定されている。平成26年3月25日新たな基本計画（第Ⅱ期基本計画）が閣議決定（http://www.soumu.go.jp/main_content/000280972.pdf）され，2次利用に関しては，次の3つの計画が定められた。

- ① オーダーメイド集計における利用条件の緩和に向けた検討
- ② 調査票情報の提供におけるリモートアクセスを含むオンライン利用やプログラム送付型集計・分析の実現に向けた整理・検討
- ③ 匿名データの作成及び提供における提供対象統計調査の種類や年次の追加等によるサービスの充実

これらの計画が策定された背景には，政府骨太方針において，「統計データの透明化・オープン化等を，第Ⅱ期基本計画の策定に反映し，その推進を図ること」とされたことが挙げられる。特に，これまで公益性の高い学術研究教育に限定されてきた公的統計マイクロデータ利用について，オーダーメイド集計は民間が活用できるようにすべく，総務省政策統括官室（統計担当）での検討，並びに高次元クロス集計結果の提供様式へ向けた拡充作業が進んでいる。

②については，既に，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マイクロデータ分析セクション並びに情報・システム研究機構新領域融合研究センター統計数理研究所オンライン解析拠点は(独)統計センターからオンライン拠点としての認可を得ている。オンライン拠点では，統計法 33 条に基づく目的外申請のデータを研究者がセキュアに分析できる環境を提供するとともに，研究者の不正行為を抑止するための監視環境並びに個人・法人情報へのア

クセスについての分析記録全保存などの仕組みも整備している。これらを通じて，これまで目的外申請では不可能であった探索的な解析や高度モデリングなど，公的統計マイクロデータ以外の情報とのリンケージなども含め，政策決定の質を上げることに繋がる社会科学実証研究への途が開かれる可能性がある。ただし，オンライン拠点では直接マイクロデータを保持してハンドリング可能とするためにその監視体制などの人員配置は不可避であり，多くの研究機関では実現不可能ではないかとの危惧がある。このため，浮上したのが，(独)統計センターと主要研究機関とを専用回線で繋ぎ，各研究機関はデータを保持することなく，統計センターに存在するマイクロデータを分析できる環境としてのリモートアクセス機関の整備である。これについては，川崎茂前応用統計学会長（国際公的統計学会会長，元総務省統計局長）が中心となり，日本学術会議マスタープランに「公的統計マイクロデータ等の研究活用のための全国ネットワーク整備」として17大学関係部局を協力機関として人文・社会科学融合領域：「エビデンスに基づく政策形成」として提案し，平成26年3月計画番号15として採択され（<http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-22-t188-1-1-1.pdf>），同時期の閣議決定と整合化したものとなっている。その構想を記載したのが図1である。

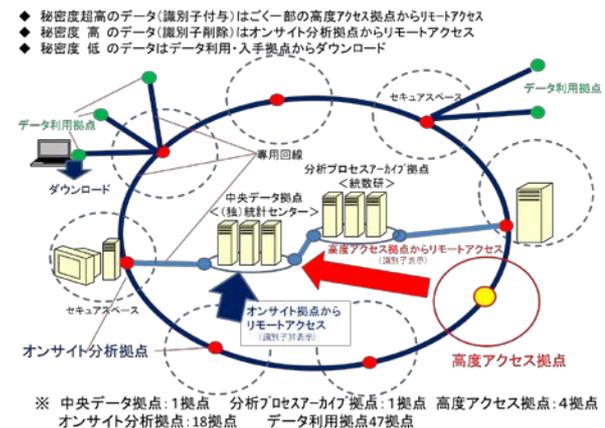


図1 公的統計の全国リモートアクセスネットワーク整備

3. おわりにーオープン化のリスク

本稿で示したように，現在急速に公的統計マイクロデータの活用についての議論が高まっている。(独)統計センターが全府省共通の利用窓口として位置づけられる可能性も高まっている。

しかし，現在の統計法は膨大な情報を持つマイクロデータ利用について公益性の高い学術研究に限っている。実際，税として徴集された情報が，特定の地域，企業，個人が将来的に不利益にさらされる可能性の強い一般目的の決定に利用されれば，公的統計制度の根幹が揺るがされることとなるからである。その意味で，公的統計マイクロデータの民間利用に関しては，それが他者への実質的対抗の意思決定に活用されないための，法的ないしは倫理的制約条件が必要である。また，その利用に関してデータ分析プロセスの第3者による収集・監視のマネジメント・システムの整備も必要と考えている。