

# 日本におけるマイクロ政府統計活用の新しい夜明け

松田 芳郎

今日は、「日本におけるマイクロ政府統計活用の新しい夜明け」と題し、これまでの研究史の回顧で基調報告にさせて頂き、この後の皆さんの統計データの二次利用に関する議論の素材として頂けると幸いです。今度の統計法の改正で研究者に、マイクロ統計の利用が大きく広がることを期待しているからです。周知のように広く利用されるマクロ統計、特にSNAでは、一次統計調査データのように規模分布状態は判りません。その種の情報が判るマイクロデータの利用が、社会経済の実態解明には不可欠ですから、今後研究者のマイクロ統計の利用が増加し、それが統計法の施行規則の改正でより円滑になることを期待しているからです。

## 1. 新統計法におけるマイクロデータの利用について

新統計法では、従来の旧統計法第 15 条2項のいわゆる目的外使用の許可条件より著しく緩やかな条件が設定されています(第 32-42 条)。しかしその内容だけでなく旧法での実態もいかに研究者の中でも知られていないかは、「日本経済新聞」の新法でのマイクロデータ公開可能性の記事を巡ってのネット上で飛び交っている議論を一瞥するだけでも知ることができます。

ここにお集まりの皆さんには周知の様に、この「試行的提供」プロジェクト自体は、旧法の下で実施されてきております。このプロジェクトに辿り着くまで、旧法の段階から、多くの研究者と政府の統計部局関係者の努力で、「目的外使用」の解釈の拡大を通して、すこしづつ政府マイクロ統計の利用が拡大してきた結果生み出されたものであります。

新統計法のこの部分は、まだ完全に運営規則・細則が定められてはいません。それがどのように書かれるかで、実際の運用は大きく異なります。今度のプロジェクトの成果が、それらの制定に反映されることを期待しています。

## 2. 日本におけるマイクロデータ活用の歴史

最初に旧法でのマイクロデータ活用の状況を総括しましょう。基本的には、中央政府、当該調査実施省庁であっても、包括承認を受けていなければ、改めて目的外の手続きが必要です。次に多い利用は他の省庁の利用であり、地方自治体(諸外国の地方政府に相当)がこれに続きます。

研究者の利用は、二つの細い経路で可能でした。ひとつは省庁の依頼による研究会によって行われるもので、関東周辺の研究者が依頼される中心でした。より広い地域の研究者が参加したのは、当時の経済企画庁経済研究所の客員研究員の研究です。これには、多くの関西圏の研究者が参加していた。(研究成果は『経済分析』誌上で公表)。ただこの場合も原則、経済企画庁の行政目的が中心であった。しかし経済研究所長を行政官と研究者の出向と交互に任用するという慣行により、かなり大幅に研究者の意向が取り入れられていたのも事実です。

これをより抜本的に研究者の意向を汲み取るようにするのは、目的外使用の要件の一つである公益性は、行政の必要性だけでなく、学問研究そのものが公益性の基準を満たすという意見を採用することで実現できるのではないかという考えが成り立ちます。ただ問題はそれをどの様にして実現できるかです。今回の統計法改正ではその方向に大きく舵をきったといえます。改正に至るまでに、この方向に大きく舵を切るきっかけになったのは、統計審議会の二つの建議と答申です。すなわち、1985年の「統計行政の中長期構想」(森口繁一会長)と1995年の「統計行政の新中長期構想」(中村隆英会長)です。

### 3. 欧米の利用と日本の利用との落差

1980年代の初頭には、すでに欧米先進国では、それぞれの国の統計部局に大学の研究者を迎え入れて一緒にマイクロデータ解析をする、いわゆる on-site data handling が始まり、そのためのマイクロ・データ・ハンドリングの技法は、匿名化した再抽出標本データを利用して事前に十二分に利用経験を積むという時代に入っていました。私は、1981年に一橋大学経済研究所附属日経済統計文献センターに在職しており、一橋大学後援会の3ヵ月の短期留学の枠をもらい、残りの月日は私費で1981-2年と、小樽商科大学時代に Fulbright 留学生として過ごした古巣のHarvardにある日本研究所に行きました。昔の先生の Kuznets, Bergson, Leontief 達は、あるいは世を、あるいはハーヴァード大学を去っていましたが、Dorfman 先生はまだ健在でしたし、新鋭の Griliches 教授の演習にも出席させていただきました。この時には、すでに1981年3月に統計審議会情報処理専門委員会委員に指名されていたので、統計基準部長の紹介状を持って、アメリカ商務省のダンカン統計基準部長に会いました。生憎会ったのは、彼が、レーガン大統領府の統計縮小に腹を立てて、辞表を出した事務所閉鎖のパーティの日でした。再会を約し、後にニューヨークで、色々話を聞くことができ、さらに彼の紹介で、Yale大学のRichard Ruggles 教授の教えを乞うようになりました。ラグルス夫妻がどのようにして商務省の中にマイクロの企業・事業所解析の共同研究の仕組みを作ったかを教わり、さらに紹介されて、Internal Revenue Office を皮切りに、隣国のカナダ統計局にも足をのばして、マイクロデータの活用状況を知る機会に恵まれました。

私自身の関心は、企業データが中心でしたので、アメリカを例に取りますと、ラグルスのデザインのように、商務省センサス局の実験用ブランチで、守秘義務を守る宣誓職員として利用することが中心になりました。ただ規模分布が歪んでいない世帯データはもっと広く利用されており、公開されているリサンプリング匿名標本データで手元で分析した上で、より抽出率の高いデータで多面的な利用を行うときに、上記のブランチに行くのが判りました。利用資格は、分析計画が、現在の日本の学術振興会科学研究費助成金に類似の National Science Foundation (NSF) の研究費を獲得することで、それが利用計画の科学的研究であることをNSFで裏書するとのことでした。

これに対して、1980年代までの日本では、先に述べたようにいわゆる目的外使用の経路を通してのみ利用可能であり、極めて限定した人々、ある意味で省庁に個人的なつながりがあるか、省庁が解析して欲しい課題でないとなかなか難しいといっても過言ではありません。

またそのような研究する能力を備えた研究者の層が極めて薄いだけでなく、共同研究を必要と

する実証的計量経済研究自体を行う研究室も稀でした。しかも多くの大学では大学紛争の後では、on-the-job-training の要素の強いマイクロ・データの実証分析は、教室制は徒弟修業だと排斥され、共同研究も崩壊して、この分野の新規参入者もまた少ないという状況でした。わずかに大学紛争に困惑してアメリカの大学に留学した学生の中から補給される可能性があるだけでした。しかし、それも、数理経済学の方は、英語の能力を相対的に必要としないので、理論に傾斜する学生が多く、まれに実証研究を志しても、アメリカ等のモデルを闇雲に日本のデータに当てはめるといふタイプの研究が横行しました。これは今に至る傾向ではないでしょうか。

このような実証研究の衰退状況を憂えて、宍戸駿太郎筑波大学教授は文部省の重点領域研究「統計データバンク」(MUSE, multi-use-socio-economic , Data Bank Project)を組織し、インターユニバーシティ研究として開拓しました。私も溝口教授の歴史統計のグループで参加しました。しかし、この中からは、マイクロデータの利用を組織化する動きはまだみられませんでした。

マイクロデータ利用は、猪瀬博東大教授を研究代表者とする理系を中心とする「高度情報化社会の研究」に浅野長一郎九州大学教授・丘本正大阪大学教授に社会科学系を組織して欲しいと誘われて、倉林義正一橋大学教授にマイクロ・マクロリンケージの班を組織していただき、古瀬大六東北大学教授(私の小樽商科大学時代の恩師の一人)に書誌情報の計算機処理班を組織していただき、それぞれの事務局を勤めました。これが、私にとって、理系に比べて遅れていた社会科学系での大規模な共同研究の組織化のKnow-howの蓄積になりました。

#### 4. 旧統計審議会の果たした役割

統計審議会の果たした役割は、行政改革で建議機能等を失ったいわゆる法施行型審議会に移行前と移行後では、大きく異なっているのではないかと思います。

移行前は、学識委員による開発部会は一応統計体系の整備に関して議論する場でもありました。より特徴的なのは、主題別部会と国民経済計算部会の他には、1)調査技術開発部会、2)情報処理部会が省庁横断型の検討の場として存在しました。この1)と2)はマイクロデータの利用に地ならしをするのに大きな役割を果たしました。これらは法施行型に移行に伴って廃止されました。マイクロ・データが提供できるようになる基礎条件は、まず各省庁でデータファイルが磁気媒体に変換され適切に保存されていることです。そのための統一化から情報処理部会は作業しました。その上で集計原表や中間集計表の保存の考えを迫り、多様な集計表が公開を求められても対応できる状況を、マイクロデータ公開にいたる前準備として、検討しました。これは「統計行政の中長期構想」に生かされました。情報処理部会で検討したもので、マイクロデータの活用に関係したのは、統計調査の統合・整合性の確保のための共通識別コードの検討でした。これは、事実上、現在の事業所・企業統一識別コードの全関連統計調査への適用を意図していました。また政府のマイクロ統計データを実際に使用してみて、別な意味で重要な役割を果たしていたのに気がついたのは、私も専門委員として参加した調査技術開発部会での検討でした(奥野忠一、竹内啓の両教授が正委員で浅井晃教授生が専門委員の一人として活躍)。最大のものは、数年かけて行った各省庁のエラー・チェック・システム実情調査とその拡張でした。これが結果的には、

省庁がマイクロデータの目的外使用の拡大に消極的な隠された理由のひとつと判りました。それは、データクリーニングに掛ける人手が都道府県レベルだけでなく中央省庁でも不足してきたことに起因してエラー・チェックがパッケージ化を生み出した可能性があり、集計表様式が定まっているときには、表章区分に必要な変数に関しては、十分なデータクリーニングがされていても、それ以外の変数はかならずしもそうではない可能性があるからです。したがって、目的外使用で提供した相手が、予想外の集計様式を意図したとすると、データクリーニングが十分でない変数が表章区分に選ばれ、結果精度に誤解を招く可能性があることです。また、全数調査型の調査で、昔であるとヴェテランの地方の統計担当者が、前年または前回と比較して、桁ずれ等による異常値を析出できたのに、それが出来るエキスパートが居なくなっており、他方中央省庁では 縦断的 (longitudinal) データファイルを編成していないためそのようなチェックが可能でないことです。このデータクリーニングとエラー・チェック・システムの開発は、実際のデータを使用して周防節雄関西大学教授の協力で行い、この経験がマイクロ・データの公開の制度設計に生かされたと思います。

## 5. 文部省科学研究費特定領域研究:マイクロ統計データの果たした役割

先に述べましたように、「統計行政の中長期構想」の段階では、調査結果の磁気媒体での保存の拡大と中間集計表の利用を可能にするという段階に留まりましたが、「統計行政の新中長期構想」では、「二-三年を目睹として」マイクロデータの利用可能性を検討すべしという項目を明記できました。しかし、「統計行政の中長期構想」の様々な項目が実現されることなく、結果を見ずに消え去った経験があるだけに、今度は研究者の側からの後押しが必要だと考えました。

日本学術会議の経済統計研究連絡委員会(委員長大屋祐雪教授)に因って、文部省科学研究費重点領域を申請するための準備的な研究費総合研究(B)を申請することになりました。関係学会との連携の下準備は順調に進行し、2度目の申請で、重点領域は 1996 年度発足で採択されました。

主題は、1) 各国のマイクロデータ使用の実態と問題点の検討、2) マイクロデータ解析の統計的手法の検討、3) 社会経済事象に適用でした。あらかじめ組織された総括班、公募班に分かれていました。2)と 3)に関しては、指定統計の目的外使用申請をする。ただし原則、いわゆる order-made 集計の変形方式(集計プログラムは研究者が自分で書く)を採用し、事務局で一括集計する。この方式は、Luxemburg Income Study の組織形態を調査した時の経験を生かしたものです。この研究者は、原則、目的外使用申請の研究計画書を提示し、申請が認められるとSASによるプログラムを事務局に送付し、事務局は使用を許されたデータにプログラムを適用して、結果を研究者に送付する。研究期間は3年間であったので、すくなくとも2回の試行錯誤による再申請による計画の書き直しが可能でした。ここで原則と記したのは、通商産業省や厚生省は研究班によっては直接データハンドリングを認めてくれた。ただそれらの班はすでに当該省庁からこれまで目的外使用申請の許可を得た経験者が代表であるグループに限られました。

この事務局は当時私の在職していた日本経済統計情報センター(日本経済統計文献センターの拡充改組したもの)に置き、この大量データの処理は新たに科学研究費で購入したサーバ

一で行いました。実際に運用してみると、送られてきたプログラムには各種のプログラムミスもあり、最終的に意図した結果を出すのに事務局は苦労したと聞いてます。目的外使用申請の書類作成作業も大変であり、個別の研究班の申請書を官庁用語に翻訳して意味の通るようにするのに膨大な時間を消費された。このプログラムの修正の努力を次の研究者世代に伝えるために、「目的外使用申請資料集」を部内資料として編集した。目的外使用申請の書類とそれに使用したプログラムの二分冊に分けて毎年度の申請後に印刷しました。

公募班の申請は、ほぼ全国に亘ったが、北海道・東北が極めて少なく、わずかに北海道の旧知のK教授の申請が1件あったに過ぎなかった。何とか、全国に広げたいと思ったが審査会が通らずに残念な思いをしました。

宍戸プロジェクトとの間が、かなり開いていたが、何とか統計的実証分析の後継者の養成に多少は貢献したと思っています。

## 6. 総括班報告書の市販版はなぜあの表紙か:アカデミー・ディ・ダヴィンチの紋章

3か年の研究を終えるまでに、重点領域は特定領域研究と種目の名称が変わった。規定に従って総括班の最終報告書を出しましたが、統計情報研究開発センターの好意により、ハードカバーで索引付きの分厚い市販本を出すことが出来ました。この白地のクロス表紙には、赤で組紐模様を箔押ししてあります(図参照)。扉の裏にアカデミー・ディ・ダヴィンチの紋章と記しましたが、残念ながら、今まで誰にもそれについて質問されたことがありませんでした。

あの万能の才を持ったレオナルド＝ダ＝ヴィンチは、王侯貴族のパトロンの下に過ごしながら精神の自由を保ちたくアカデミーを夢想していて、何種類かの紋章をデザインしていました。このレオナルド＝ダ＝ヴィンチの望んでいたアカデミーの紋章を敢えて表紙に使用したのは、この特定領域の研究も、大学の大小を問わず、国公立を問わず、大学院の有無を問わず、自立した研究者の自由な連合としてありたかったという、領域代表としての私の夢の象徴としてでした。

特定の大学でなければ目的外使用申請に関して、門前払いを食らうのでないように研究者集団を構築したいとの願いです。これが、「統計行政の新中長期構想」でも意図したことでもあると信じていたからです。今度の統計法の改正はこの「国民のための統計」の立場ですから、当然新しい目的外使用申請もこの立場になって運用規則が書かれて欲しいものです。

さて、この特定領域の quasi-order-made 集計は極めて不便であるので、より自由にデータに触れる仕組みが必要だと考えました。1999 年度は特定領域研究終了後のとりまとめ期間として総括班にだけ若干の研究費がつきます。それを活用して、調査の全データから、匿名化再抽出データを作り、リサンプリング・データと呼んで、目的外使用の申請者が手元でデータに触れることが出来るようにする実験的参加者募集をいたしました。無論目的外使用申請の手続きとデータが提供された後のリサンプリングの作業は、事務局が担当しました。

## 7. リサンプリングデータの意図した目的

このリサンプリングデータを作成するという考えは、総括班の一人の大阪大学の伴金美教授の

主張に端を発したものです。すなわち日本の大学で、マイクロ統計のデータの使用方法を教えるのに、アメリカセンサス局の提供するマイクロ統計データを使用しなければならないというのはおかしいという主張です。そのためには、学生と一緒に自由に使用できるマイクロ統計データが必要です。先に述べた、リサンプリングデータの使用はそのための実験と位置づけていました。

出来ることならば、単に匿名化だけでなくスワッピング等の処理を施して、実際のデータに基づいているが個別の回答者とのリンケージ不可能なデータ、これを我々はレプリカ・データと呼びましたが、それで目的外の使用条件の外で大学院生にも自由に使えるデータにしたかったのです。この我々の意図は、当時も今も実現していません。この間に、欧米先進諸国のマイクロデータ開示の状況は、総括班の浜砂敬郎九州大学教授、森博美法政大学教授たちの手によって精力的に進められ、その全容がいかにか日本の現状とかけ離れているかが明らかになりました。

この特定領域のメンバーと重複しますが、独立の研究グループによるアジア地域諸国のマイクロデータ利用の動きが進んでいました。一つの源泉は、長年かかって実施されていた溝口敏行一橋大学教授を代表者とした所得資産分布の研究プロジェクトです。これは、アジア諸国の集計量データによるもので、寺崎康博長崎大学助教授や私を中核メンバーとし、その成果はすでに1997年に上梓されていました。さらに、これらの一連の所得分布研究の遺産を受けて、寺崎(すでに東京理科大学経営学部教授ですぐ議論できるようになっておりました)と一緒に、マイクロ・データによる研究に進化させようとしていました。すなわち、スリランカ統計局からデータの提供を受けてマイクロ・データの解析がその始まりでした。またフィリピンの統計基準部に相当する組織からすでにマイクロデータの提供が進行していることを知りました。いまひとつは、辻井博京都大学教授と一緒に中国農業部の農家固定点調査のマイクロデータの提供を受けてパネル・データを編成する実験を行っていました。

このような状況から判断すると、日本は単に欧米の先進国だけでなく、アジア諸国の発展途上国にも取り残されかかった状態であるといえます。

## 8. 一橋大学のセンターの後続プロジェクトと日本学術会議の意見表明

この特定領域による取りまとめ研究は、マイクロ統計活用研究会による科学研究費研究成果公開研究費助成を申請して、データベース作成費の申請を行い、採択されました。ただ私は一橋大学を定年で去り、東京国際大学に移っていましたので、統計情報研究開発センターの助力を得て、その計算機システムを利用して井出満理事長が、大阪産業大学教授として、継続申請されました。井出教授の定年退職後は、森博美法政大学教授が研究代表者で、今日まで継続しています。

これはリサンプリングデータを作成する目的として、目的外使用申請をします。ただ、その際利用者も公募することにしました。残念ながら、リサンプリングデータを編成するところまでを目的外使用申請にして、その後の利用は目的外の縛りからはずそうというのが目標でしたが、それは実現せず、公募した研究の計画調書を含めて目的外使用申請の対象になりました。この申請事務は統計情報研究開発センターが分担してリサンプリングデータを作成してくれただけでなく、公募班の計算プログラムを点検の上、全データで同じ計算をして、リサンプリングデータのユーザビリ

データの検証作業を分担してくれました。

このリサンプリング・データの抽出率は最初 20%とし、さらに試行錯誤を経て 50%の抽出率にすることができました。この精度検証の作業は、一応の試行実験の終了を機会に、「Sinfonica 研究叢書」の一冊として纏められました。しかし当初目標にしたレプリカ・データ・データベースとして公開することは出来ませんでしたので、課題は残されております。

この間、私は日本学術会議の第 18-19 期(平成 12-17 年)に学会推薦会員として、経済統計研究連絡委員会委員長となり、第3部経済学の代表として学術基盤情報常置委員会委員長に推挙され運営会議に陪席するようになりました。経済統計研究連絡委員会の協力を得て2本の対外報告と1本の要望をまとめることができました。その対外報告がお役に立って、総務省統計研修所にマイクロ統計解析室が設置され、一橋大学経済研究所附属日本経済統計情報センターが社会科学統計情報研究センターに拡充改組する切っ掛けを作りました。

この拡大したセンターが総務省統計局の依頼により「匿名標本データの試行的運用」プロジェクトを引き受けることになったわけです。このプロジェクトは、総務省統計局と 21 世紀COE[現在のグローバルCOE]プログラムの統計分析拠点としての一橋大学との共同研究としてなされたものです。私は直接運営には関係していませんので、今回各種の資料を拝見して、大きく前進したと感じました。目的外使用の申請手続きも、恐らく匿名標本化ということすなわち全データを提供しないということで、多少簡略になったに違いありません。リサンプリングの抽出率は 80%ということでもまず満足できる規模であると思います。

特筆すべきことは、2点あります。1) 統計局所管の世帯調査のうち、国勢調査以外の、いわゆる大型標本調査4種類が、マイクロデータとして保存され利用可能な過去の時点に遡って提供されたこと、2) 利用者が北海道を含めてほぼ全国に広がったことです。特定領域の時には空白地帯であった北海道に、関東地区のマイクロ統計活用研究会で利用した人が、職を移してきて申請していました。これは、抽出率の低い相対的に小規模データで実験後、80%データに挑んだという意味で、それぞれのプロジェクトが住み分けられてきたともいえます。この上は、さらに全データを使用する必要のある研究は、総務省統計研修所にマイクロ統計解析室に使用希望出すことによってより明確な分業関係が作れると思います。

問題は、第1に、総務省統計局以外の他の省庁のデータをどのようにして利用可能な仕組みを作るかということ、規模分布の関係でリサンプリングになじまない、事業所や企業に関する調査データをどのようにして利用可能なシステムを作るかです。第2に、解析結果をどのように蓄積するかということ、解析結果明らかになった統計調査結果の抱えている問題点が明らかになったときにどのようにして、調査官庁にフィードバックするかです。

## 9. 在りうべきセンター像

在りうべきセンター像としては、

- 1) レプリカ・データ・データベースの推進かマイクロ統計活用研究会のデータベースが大学院生を含めて簡易な手続きで教育用に使用できる。
- 2) 教育用データベースで試行を試みた

後に、例えば社会科学統計情報研究センターのような組織で、匿名化標本データで研究が出来る。それに必要な経費は別枠の科学研究費で賄えるようなシステムにする。3) 匿名化標本データでもまだ十分でないことが明らかにされた研究に関してまた匿名標本化データにはなじまない事業所・企業等に関する統計調査に関しても解析できるように、例えば総務省統計研修所にマイクロ統計解析室のような所で全省庁のデータも解析できる体制を作る。それに必要な経費は、やはり別枠の科学研究費を制度化する。同時に、総務省統計研修所にマイクロ統計解析室を例にとると、利用のハード・ソフトの両面で改善したものにする。4) これらの多面的解析に耐えうるようなデータのクリーニングや分類コード等の符号付けの再格付けが可能なようにメタ情報を整備する。さらに、経済統計に関する基礎的知識が体系的に教授される教育体制の整備が急務である。

こういうことが出来るならば、日本のマイクロ統計データ活用の後進性も少しは改善されるのではないだろうか。

まつだよしろう（青森公立大学教授、一橋大学・東京国際大学名誉教授）

## 参考引用文献

- 統計審議会[1995]『統計行政の新中長期構想』
- 日本学術会議[2005]「電子媒体学術情報の恒久的な蓄積・保存・利用体制の整備・確立(要望)」
- \_\_\_\_\_学術基盤情報常置委員会[2001]「情報化社会における政府統計の一次データの提供形態のあり方について(対外報告)」
- \_\_\_\_\_・\_\_\_\_\_ [2005]「政府統計・世論調査等の一次データ(含む個票データ)の体系的保存と活用・公開方策について(対外報告)」
- 稲葉由之[2008]『官庁統計の収集・公開・利用のための理論的問題の検討』(平成16-19年度科学研究費補助金(基盤研究(A))研究成果報告書)
- 松井 博[2005]「政府統計マイクロデータの利用」*Estrela* #137.
- \_\_\_\_\_ [2008]『公的統計の体系と見方』(日本評論社)
- 松田芳郎(編)[1999]『統計情報のフロンティアの拡大:マイクロデータによる社会構造解析』(統計情報研究開発センター)
- \_\_\_\_\_・他(共編)[1999-2001]『講座マイクロ統計分析』(第1-4巻既刊、5未刊)(日本評論社)
- 山口幸三[2008]「政府統計の個票利用と統計法改正:試行的提供の経験を踏まえて」『経済研究』59-2.