

紹介

データアーカイブの役割と SSJ データアーカイブの現状

——実証研究における再現性を担保するために

佐藤 朋彦

(元・東京大学社会科学研究所助教授)

佐藤 博樹

(東京大学社会科学研究所教授)

目次

- I はじめに
- II データアーカイブの機能と意義
- III SSJ データアーカイブの現況と管理・運営上の課題
- IV ICPSR の現況
- V 日本における労働分野の実証研究とデータ公開の現状
- VI おわりに

I はじめに

日本では、毎年数多くの統計調査や社会調査さらに世論調査などが実施され、分析結果が報告書として刊行されている。しかし、報告書が刊行された後に、調査された個々のデータ（以下、マイクロデータと記述）が、どのように取り扱われているかについては、ほとんど注目されていない。その背景には、日本では、これらのマイクロデータに関して重要な社会的な研究資源とする見方がなかったこともあり、マイクロデータを収集・整理・保存し、提供のために提供する組織が、10年ぐらい前までは存在していなかったことがある。

本稿では、これら調査の個々のデータを保存し、かつ公開することの社会的な意義と、マイクロデータの収集・保存・提供の作業を行うデータアーカイブの役割について考察するとともに、東京大学社会科学研究所日本社会研究情報センターが運営している Social Science Japan データアーカイブ

（以下、SSJ データアーカイブと記述）の現状と課題について紹介する。また、SSJ データアーカイブとの比較の観点から、世界最大級のデータアーカイブでアメリカのミシガン大学（the University of Michigan）の社会調査研究所（the Institute for Social Research）に事務局がある The Inter-university Consortium for Political and Social Research（以下、ICPSR と記述）の現況についても解説する。

最後に、労働研究の分野においてマイクロデータを活用した研究として『日本労働研究雑誌』に掲載された論文を取り上げ、当該データを第三者が入手しその内容を検証できる条件があるかどうかの検討を行い、実証研究において「再現性」を担保することの重要性を指摘する。

II データアーカイブの機能と意義

1 基本的機能

データアーカイブの機能とは、統計調査や社会調査のマイクロデータのうちコンピュータ処理可能なデータ、すなわち電子的又は磁氣的に記録されたデータを収集・整理・保存し、データの再分析、すなわち二次的な分析を行う利用者に提供することにある。また、マイクロデータの二次的な利用を可能にするため、調査の概要、標本設計、回答状況、調査票、データのクリーニング方法等のドキュ

メント類（以下、メタデータと略記）もマイクロデータとあわせて利用者に提供する。

このようなデータアーカイブによるマイクロデータの収集・保存・提供の機能は、極めて重要なものであるが、それ以上に重要な機能がある。それは、社会科学におけるマイクロデータに基づく実証研究の「再現性」を担保する研究環境を提供することである。「再現性」とは、第三者が、同一のマイクロデータを利用して同一の手続き（例えば統計分析）を行えば、同一の結果が得られることを意味する。「再現性」の担保は、実証研究を支える条件であり、海外で刊行されている学術雑誌には、第三者がアクセスできないデータに基づく論文は受理しないと明記しているものもある。これは、分析に利用したマイクロデータに第三者がアクセスできない論文では、その内容を検証できず、「再現性」が担保されないことによる。日本の社会科学における実証研究では、この「再現性」の担保については、不十分な点が多い。これについては、「V 日本における労働分野の実証研究とデータ公開の現状」で検討する。

2 社会的意義

次に、社会科学におけるデータアーカイブの社会的意義について確認しておこう。

第一の意義は、マイクロデータが二次分析のために広く提供されることによって、新しい研究が可能となり、新たな知見が得られることである。

統計調査や社会調査によって、通常、質問項目に関して回答が得られる。これらをすべて組み合わせた集計は極めて膨大になるため、一般には、調査実施者の問題意識や研究目的に即して、特定の組合せによる集計や統計的な分析が行われる。ところが、異なった問題意識や研究目的からは、異なる集計や分析が必要になる。しかし、当該マイクロデータが二次分析のために提供されていない場合には、新しい調査を自ら実施するほかないのである。データアーカイブにマイクロデータが寄託され、二次分析に提供されるようになれば、既存データに基づいて新しい研究が可能となり、新たな知見や有益な情報を得ることが期待できる。

例えば、最近、学校卒業後に雇用期間が定めら

れた有期契約の仕事に従事する者、いわゆるフリーターとなる若者が増加し、政策的な対応が課題となっている。フリーターの就業問題を検討する場合、フリーターの定義に該当するとみられる若者の属性や就業実態、さらには就業意識などを把握する必要がある。また、過去の調査データを用いて、現在のフリーター概念に基づいて、データを再分析し、現在の就業実態や就業意識などと比較できれば、一時点における横断的なデータ分析よりも、より有効性のある研究が可能となる。しかしこうした分析のためには、過去のマイクロデータが保存されていなければならないのである。つまり、過去のマイクロデータを二次的に利用できるように収集・整理・保存することは、過去のマイクロデータの利用価値を高めることになる。当然のことであるが、過去のデータを利用するニーズが生じて、コンピュータ処理可能なマイクロデータが存在しなければ、またそれを利用するためのメタデータが存在しなければ、データの再分析はできないのである（Collins 1996, 前田 2004）。

第二の意義は、調査対象者の回答負担を軽減することである。統計調査や社会調査を行うには経費と時間が必要となるだけでなく、多くの調査対象となった人や事業所等の協力を得て行われるものである。もし、既存の調査のマイクロデータが利用できないために新しい調査を実施したとすれば、経費や時間の無駄というだけでなく、調査対象者の回答負担を増やし、その結果、調査対象者の調査への協力度を低下させるおそれがある。見方を変えて言えば、マイクロデータの保存と再分析へ提供することは、総体として調査対象者の回答負担を減らすことに貢献するものであり、調査実施者の社会的な責務であるとも言える。

第三の意義は、わが国の統計調査や社会調査の質の維持・向上に役立つことである。マイクロデータの二次分析が可能になると、新しい調査を企画する際に、既存の調査との比較可能性について検討したり、調査票にどのような質問事項を入れるべきか参考にしたりすることができる。また、マイクロデータの再分析を通じて、当該データの内容に関して有益な示唆を得ることも期待される。さらに、マイクロデータを第三者へ公開することを前

提に調査を実施することによって、調査実施者は企画段階から集計、分析まで調査研究の質の維持、向上のために、コーディング方法などの文書類の整備を含め、意識的に管理することにもなる。

以上のように、データアーカイブは、マイクロデータの二次分析を可能とするだけでなく、調査対象者の回答負担の軽減や統計調査、社会調査の質の向上に貢献できるものなのである。

Ⅲ SSJ データアーカイブの現況と管理・運営上の課題

1 現況紹介

〈沿革〉

はじめにSSJデータアーカイブの設立の経緯について、簡単にふれておこう。1996年に東京大学社会科学研究所の附属施設として日本社会研究情報センターが設立された。同センターは、その事業のひとつとして、民間の調査研究機関や大学等の研究者が実施した社会調査等のマイクロデータの収集を担うものとしてSSJデータアーカイブを開設した。同データアーカイブは1998年4月より二次分析を希望する研究者にデータの提供を開始し、現在では、日本社会研究情報センターの主たる事業となっている。

〈公開データセット数と利用状況〉

SSJデータアーカイブが保存、提供しているデータは、これまでに約60の民間あるいは公的な調査・研究機関、学会や大学の研究者から寄託されたものであり、2006年2月末現在で760データセット(572調査分、ひとつの調査が複数のデータセットから構成されているものがあるため、調査数よりもデータセット数が多い)を研究者などに提供している(図1参照)。主な寄託機関を紹介すると次のようになる。公的な調査・研究機関では、独立行政法人労働政策研究・研修機構、国民生活金融公庫総合研究所、東京都老人総合研究所などがある。また、民間の調査・研究機関では、リクルートワークス研究所、エンタテインメントビジネス総合研究所、全国大学生生活協同組合連合会、ベネッセコーポレーション、生命保険文化センター、

第一生命経済研究所などがある。さらに、学会や研究者関係では、日本家族社会学会、SSM調査研究会、選挙とデモクラシー研究会などから寄託を受けている。なお、これらの機関などが実施した調査データのすべてが、SSJデータアーカイブに寄託されているわけではない。今後、データ寄託のさらなる促進が課題である。

次に収録データの利用状況を見ると、提供を開始した1998年度は年間の利用申請件数はわずか11件、提供データセット数が22であった。その後、公開データセット数が増えるにつれて、利用申請は急速に増加し、2004年度には利用申請件数が343件、提供データセット数が1735となり、直近の2005年度は2月末現在で利用申請件数が318件、提供データセット数が2411となっている。2005年度に提供したデータセットの内訳を見ると、研究用が全体の42.1%であるのに対して、教育用が57.9%となっている。なお、SSJデータアーカイブが提供するマイクロデータの多くは研究目的(学術目的の二次分析)に限定されており、教育目的で利用できるのは、大阪商業大学比較地域研究所と東京大学社会科学研究所が寄託しているJGSS(Japanese General Social Surveys; 日本版総合社会調査)や第一生命経済研究所が寄託している「今後の生活に関するアンケート」など一部に限られている。

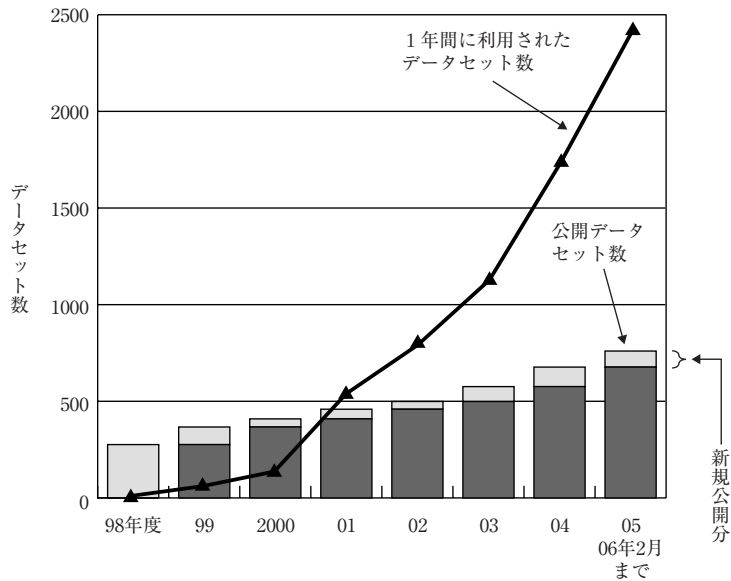
提供したデータセットを利用して作成された論文・著書数は、提供開始から直近の2006年2月末までに合計で262点となっており、2002年度以降は毎年50~60点前後が発表されている(表1)。

このようにSSJデータアーカイブは、わが国の社会科学の実証分析に欠かせない研究資源となりつつある。

〈組織体制と業務分担〉

SSJデータアーカイブは、日本社会研究情報センターのSSJデータアーカイブ部会(2006年4月からデータアーカイブを主として担当する調査基盤研究に加えて、社会調査研究、計量社会研究の3つの分野から構成)が運営している。SSJデータアーカイブの構成員は、教授・助教授(兼任を含む)が合計5名で、担当助手2名および特任専門職員

図1 公開データセットの累積数と1年間の利用データセット数の推移



が1名である（このほかに国内客員の教授と助教授が各1名で二次分析研究会を担当）。また、データアーカイブのさまざまな業務を支援するために、事務補佐員やリサーチ・アシスタント(大学院生)等も非常勤スタッフとして参画しており、これら非常勤スタッフの1週間当たり実働人員は約16.0人（1日当たりでは約3.2人）となっている。

通常、調査基盤研究の教授がデータ寄託の依頼を、同研究分野の助教授が業務全体の企画・運営を、助手がデータの利用やホームページの管理を、特任専門職員が寄託されたデータの内容確認（クリーニングや秘匿処理を含む）をそれぞれ担当し、助教授の指示の下で、上述の非常勤スタッフが調査情報の入力のほか、各業務の支援に当たっている。また、データアーカイブ以外の社会科学研究所の研究スタッフや国内客員の教授、助教授も二次分析の普及活動である二次分析研究会や公開セミナーを支援している。

なお、SSJ データアーカイブを運営するための経費は、主として大学運営費交付金で賄われているが、寄託された調査データに関する各種情報の入力に掛かる経費については、1997年度以降毎年、独立行政法人日本学術振興会（当初は文部科学省）より科学研究費補助金（研究成果公開促進費）

表1 SSJDAの公開データセット数と利用状況の推移

年度	公開データセット数	利用申請件数	提供データセット数	発表論文・著書数
1998	276	11	22	3
1999	367	20	72	4
2000	409	51	147	3
2001	459	114	545	34
2002	499	222	796	58
2003	576	224	1,129	49
2004	677	343	1,735	43
2005	760	318	2,411	68
総数	—	1,303	6,857	262

注：2005年度は2006年2月末現在の数値

の交付を受けている。

〈寄託データの保存ファイルと内容確認〉

マイクロデータはさまざまな形式のファイルで寄託される。SSJ データアーカイブでは、これらを統計処理のパッケージソフトであるSPSSに読み込み、同ソフトのファイル形式に変換して保存している。ファイルの変換に当たっては、調査票の質問項目との対応がわかりやすい変数名を付けるとともに、日本語で変数ラベルや値ラベルを付けている。

寄託されるマイクロデータは調査実施者の下で当初の目的である一次分析が行われ、報告書等が出されているものである。したがって、データの確

認作業は手短に済ますことができると一般には思われるかもしれない。しかし、寄託されたマイクロデータを再度単純集計し、報告書等の結果と比較してみると一致しないことも少なくない。そこで、データアーカイブでは、項目ごとに再度単純集計を行い、その結果を報告書と比較しながら、寄託されたマイクロデータの内容確認を行った上で公開している。

なお、確認作業の過程で、報告書と一致していない場合やデータ数が異なる場合には、寄託機関と連絡をとり、データ修正が必要か否かを判断している。しかし、調査実施時点からかなりの時間が経ってから寄託されたデータや、寄託機関と調査や集計を実施した機関が異なる（寄託機関が調査・集計を他へ委託した）場合などでは、確認作業を進める上で必要な情報が寄託機関から得られないこともある。このようなケースでは、SSJデータアーカイブ側で判断し、寄託元の了解を得てデータ修正等を行っている。

また、調査回答者や事業所・企業等が特定されるおそれがある場合には、「事実上の匿名性の概念」（マイクロデータの特定に要する作業量が現実的でないほど大きければ、匿名性は事実上担保されるという考え方）に基づき、対応関係を特定する危険性の高い識別情報の削除やトップコーディング、グルーピングなどの技法を用いて秘匿処理を行っている（濱砂（2000））。

これらの確認結果やデータ修正、秘匿措置については、「データチェック・修正記録票」を設けて詳細に記述するとともに、報告書と一致しないケースについては、データ利用者に提供する際に添付する「readme（利用上の注意）」ファイルに記述している。また、調査ごとにボックスファイルを用意し、そこへ「データチェック・修正記録票」および寄託機関とのやり取りなどをすべて残すこととしている（写真1）。

〈データ利用に関する啓蒙・普及活動〉

データアーカイブや二次分析の意義や可能性についての理解を深めるため、1997年2月より2001年10月まで合計6回にわたり公開セミナーを実施している。

また、寄託されたマイクロデータを利用した二次

分析の普及促進を図るため、2000年度からデータ寄託先から受け入れた国内客員の教授、助教授と社会科学研究所のスタッフで毎年「二次分析研究会」を開催しており、2006年度で7回目となる。「二次分析研究会」では、毎年テーマを決め、そのテーマに合ったデータをSSJデータアーカイブの所蔵データから選定し、関心を持つ研究者・大学院生等が当該データの分析・発表・討議を行いながら研究成果をまとめ、リサーチペーパーとして公表している。

さらに、2005年度にはSSJデータアーカイブが提供するデータを用いた二次分析による研究論文で優秀な論文を表彰する制度（SSJデータアーカイブ優秀論文賞）を新たに設け、二次分析の奨励を図ることとしている。ちなみに、第1回のSSJデータアーカイブ優秀論文賞を受賞したのは、『日本労働研究雑誌』（536号）に掲載された、高橋陽子「ホワイトカラー『サービス残業』の経済学的背景——労働時間・報酬に関する暗黙の契約」である。

2 管理・運営上の課題

〈スタッフの育成とその確保〉

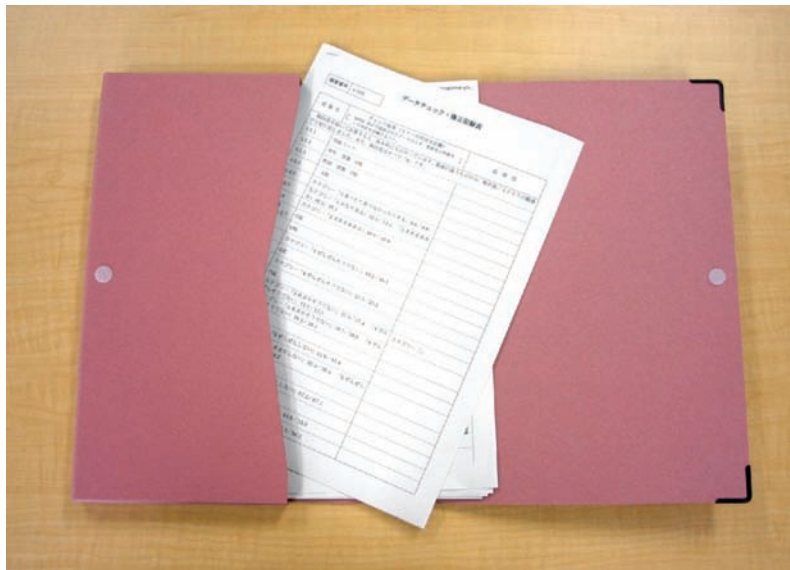
SSJデータアーカイブを支援するスタッフの業務内容は、専門的にかつ多岐にわたるものである。また、最近は業務量も急激に多くなってきている。しかしながら、これらの業務を支えるスタッフの多くは非常勤職員等であるため、人的基盤は脆弱である。

海外のデータアーカイブでは、専門職として「Data Librarian」や「Data Archivist」が育成、配置されており、データアーカイブは専門職によって運営されている（図書館などの司書に対応する専門職）。SSJデータアーカイブを将来にわたって管理・運営、さらに発展させていくには、専門性を持った人材の育成と確保が欠かせない。また、データアーカイブを担う人材には、専門性が必要であることを一般に認知してもらう必要もあろう。

〈提供事務の簡素、合理化〉

データの利用が増えれば、当然のことながらデータを提供する事務も増加することになる。この業務を軽減するには、データ利用数の適正化とデー

写真1 記録保存用のボックスファイル



タ提供の事務の簡素・合理化が必要となる。データの利用状況を見てみると、これまでは異なる3つ以上の調査データを一度に申請する者や、短期間に異なる調査を何回も申請する者も見られた。

そこで、2006年4月から一度に利用申請できる調査数を原則として2つの調査データに限定することとした。また、データ提供後、半年以内は原則として新たな利用申請を受けない方式に改めた。ただし、複数年にわたる調査などで調査番号が異なっても、同一の調査名であれば1調査と数えることとし、時系列的な研究に支障がないようにしている。

また、データの提供方法については、現行のCD-Rに複製して郵送する方式からインターネットを通じて直接ダウンロードする方式にすることも検討していく必要がある。ただし、インターネットを通じたダウンロード方式への移行については、データ転送のセキュリティを確保するほかに、データの散逸や不正な利用を防止するため、マイクロデータ利用上の基本的なルールを利用者、特に学生や院生の間に浸透させる教育が不可欠となろう。

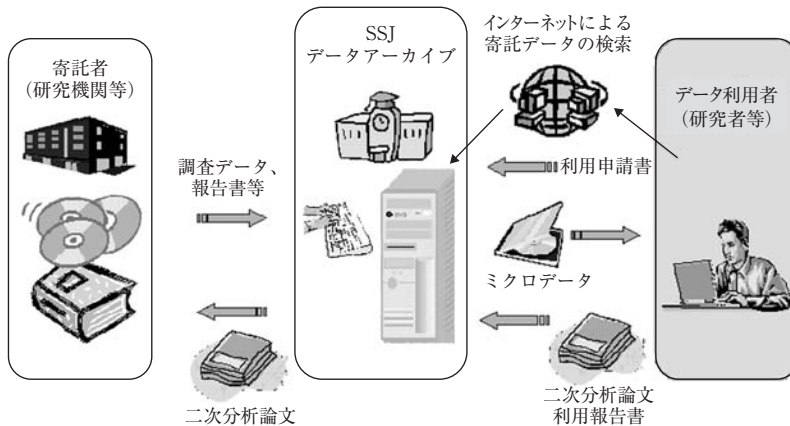
〈寄託データの確保と寄託機関への研究成果の還元〉

データアーカイブのシステムは、調査実施機関等からのマイクロデータの寄託とマイクロデータの提

供を受けた利用者の研究成果の還元で成り立っている(図2)。先に示した図1を見るとわかるように、この数年でSSJデータアーカイブから研究者等へ提供されるデータセット数は飛躍的に伸びている。その一方でそれを利用して得られた新たな研究成果である論文や著作の数は、それに比べて伸びていない状況にある。これにはつぎの二つの要因が考えられる。

第一は、データ利用者が利用するデータについて、十分な情報を得ないまま分析を開始するために、成果が得られない場合である。そこでSSJデータアーカイブではウェブ上に提供している各調査の概要(一部は英文でも提供)を掲載しているほか、調査報告書についても社会科学研究所の図書室を通じて、閲覧(一部複写)できるようになっている。また、申請予定のデータに計画している分析に十分なデータ数が含まれているか否かを見極めることができるよう、リモート集計システムを構築し、その利用を進めているところである。しかし、調査報告書を読まずに分析にすぐ着手する研究者も一部にはいるようである。一方、稲葉(2005)も指摘しているように、各調査項目の信頼性や妥当性に関する情報が不十分なケースも多いと見られる。こうした情報はデータ・クリーニングの作業から得られることも多く、当然のこ

図2 SSJデータアーカイブとデータ寄託者およびデータ利用者との関係



とながらデータの作成機関しか持ちえない。今後はデータの作成機関からはこれらの情報をあわせて寄託してもらい、利用者に開示する必要がある。

第二は、利用者のマイクロデータ利用に関する基本的なルールに関する理解度の低さの問題もあるのではないかと考えられる。新たなデータ寄託を促進する上でも、研究成果の寄託機関への還元はきわめて重要である。しかし、データ利用者の中にはデータの利用期限がきてもデータアーカイブ側が督促するまで、論文を提出しないだけでなく、利用報告書さえ提出されてこない場合もある。今後は利用報告書と論文等研究成果の提出に関する管理を厳格化することを予定している。SSJ データアーカイブでは、提供事務の簡素、合理化を図る観点もあって、データ申請・提供関係のデータベースシステムを新たに構築し、2005年度末より運用を開始した。データベースを効果的に用いることで、個々のデータ利用の状況をさらに詳細に把握することが可能となる。

IV ICPSR の現況

ここでは、SSJ データアーカイブとの比較の観点および SSJ データアーカイブの将来を考える上から、社会科学に関する調査のマイクロデータをアメリカ国内だけでなく、世界各国や国際組織から収集、保存し、それらを学術目的での二次分析

のために提供している世界最大級のデータアーカイブである ICPSR の現況を見ていくことにしよう。

〈加盟機関〉

ICPSR は 1962 年にアメリカのミシガン大学社会調査研究所に創設された組織で、その名称からも明らかなように複数の大学等の機関が連携した組織である。ICPSR にはアメリカだけでなく世界各国の大学等が加盟しており、2005 年度には 16 の機関が新たに加盟し、加盟機関合計は 514 機関となっている。日本では 1999 年までは各大学が個々に ICPSR に加盟していたが、ICPSR のデータを効率的にかつ安価に利用するため、1999 年 9 月に東京大学の SSJ データアーカイブを中心に ICPSR 国内利用協議会を組織した。現在、この国内利用協議会が ICPSR の規約に基づきナショナルメンバーシップを得て、国内利用協議会として ICPSR に加盟している。日本国内の大学は、ICPSR 国内利用協議会の会員となることで、ICPSR のデータを利用できることになる。2005 年度末現在で、国内利用協議会には 22 の大学が加盟している。

〈マイクロデータの提供〉

ICPSR が収蔵しているマイクロデータは、政治、社会、歴史、経済、高齢者、犯罪、公衆衛生、教育、法律および国際関係と広範な学問分野に及んでおり、その数は 6200 以上となっている。これらの中にはアメリカの人口住宅センサスなど政府

統計のマイクロデータも含まれている。ICPSR から送られてきたメールを基にデータの追加および更新状況を集計してみると、2005 年 1 年間で 275 件のデータが新たに公開され、151 件のデータが寄託機関やデータ利用者などからの情報により更新されている。ICPSR に収蔵されているマイクロデータはアメリカだけでなく、世界各国の調査実施者から寄託されたものであり、日本からからも旧・内閣官房広報室の世論調査（1960 年代のものが主で、寄託は政治学者の三宅一郎氏の努力による）や JGSS（英語版）などのデータが寄託されている。なお、提供されるデータの利用は、大学の学部生も利用できるが、利用は学術研究および教育目的に限定されている。

ICPSR のマイクロデータは、日本の場合、3 年ほど前までは国内利用協議会のハブ機関である SSJ データアーカイブを通じて、同協議会加盟機関の利用者に提供されていた。しかし、最近では同協議会加盟機関の IP アドレスを ICPSR に登録すれば、インターネットを通じて直接マイクロデータをダウンロードすることができる。また、ダウンロードが行われた場合には、各加盟機関の代表者（OR: Official Representatives）に実行した者の登録情報とダウンロードされた調査名等がメールで送られてくる仕組みとなっている。ICPSR の年次報告書によると、ダウンロード方式が導入されたこともあり、2005 年 1 年間に提供された総データセット数は 32 万 1134 件、データの総容量は 125 万 5092MB と、2001 年に比べて 4 倍以上となっている。

〈その他の活動〉

ICPSR では、マイクロデータの分析手法に関して毎年夏にサマープログラムを開講している。このプログラムでは、社会科学のための研究方法、統計手法およびデータ分析について、基礎的な方法論から特定分野の高度な研究手法に至るまで、幅広くかつ総合的なカリキュラムが設定されている。日本からも毎年 15 名前後の大学院生等が受講している。なお、このサマープログラムは加盟機関以外に所属する者も受講可能であるが、受講料は加盟機関の受講者の 2 倍となる。

また ICPSR は、提供している一部のマイクロデー

タについて、データをダウンロードすることなしに、ウェブ上でデータ分析ができるシステム（DAS: Online Analysis Using the Data Analysis System）を提供している。このシステムは、度数分布、クロス集計、回帰計算などの統計処理のほか、既存の変数を演算して新しい変数を作成するなどの変数操作の機能も持っており、統計分析ソフトを持っていない利用者にとっては有益なものである。

V 日本における労働分野の実証研究とデータ公開の現状

〈マイクロデータを利用した労働研究の現状〉

労働研究の分野では、マイクロデータを用いた実証研究が盛んになっている。例えば、『日本労働研究雑誌』に掲載された論文にもこうした研究が多い。表 2 は、2003 年から 2005 年の 3 年間に『日本労働研究雑誌』に掲載された論文の中から、マイクロデータを利用した研究を抜き出したものである。同表は、職員録や会社四季報などから作成したデータセットによる分析は除外されている。これらの研究もマイクロデータを利用した分析とすれば、さらに論文数が増加することになる。表 2 の「データの種類欄」の「二次分析」は、筆者らが収集した調査データ以外のマイクロデータを利用した研究を意味する。政府統計を分析したものやデータアーカイブから入手したデータを分析したものなどが該当する。第三者の入手可能性は、論文の内容を検証するためにマイクロデータにアクセスできるかどうかを意味する。「不明」と記されていても、筆者に問い合わせることで入手できる可能性もあるし、論文執筆後に何らかの形でマイクロデータを公開していることもありうる。あくまでも論文の記載内容による判断である。なお、論文執筆時点で入手できなくても、現時点で SSJ データアーカイブに寄託されているものは、SSJDA と記載している。

表 2 にあげた論文が利用したマイクロデータは、a) 筆者らが独自に調査を実施したもの、b) 筆者らが企業の人事部などから入手したもの、c) 目的外利用申請による政府統計、d) データアーカイ

ブ等から研究用に提供されているもの（SSJ データアーカイブの所蔵データや家計経済研究所のパネルデータ）などに分類できる。なお、d)のようにマイクロデータが研究者に公開されるようになったことが、マイクロデータを利用した実証的な研究を盛んにしている要因のひとつである。

〈分析に利用したマイクロデータへ再分析の可能性：再現性の担保〉

研究に利用したマイクロデータを上記の4つに分けた場合、c)やd)による研究は、当該マイクロデータを入手できる可能性があるため、論文の内容を第三者が検証することが可能である。その意味では、再現性が担保されている。ただし、政府統計の場合、すでに公表されている論文の内容を検証することだけを目的として目的外利用申請を行った場合に、申請が許可されるかどうかは不確かである。

他方、a)やb)のマイクロデータによる研究は、分析に用いられたマイクロデータが、データアーカイブなどに寄託されるなどして第三者が入手可能な状況にない場合は、論文の内容を第三者が検証することができない。その意味では、再現性が担保されていないことになる。実際、a)やb)に分類される論文をみると、データアーカイブなどにマイクロデータが寄託されているものもあるが、第三者が入手できないマイクロデータがかなりを占める現状にあることがわかる。もちろん、すでに述べたように、筆者らに問い合わせることで、データを入手し、検証の作業を行うことができる可能性もあろう。なお、個別企業の人事データの場合には、分析者がデータアーカイブに寄託を希望しても、企業の許可が得られない可能性もある。しかし、こうした場合を含め、実証研究である以上、再現性を担保するための仕組みに関して検討が必要となろう。

〈マイクロデータの寄託促進など今後の課題〉

表2を作成する際に、論文の記載内容からは、マイクロデータを筆者自身が調査から作成したのか、あるいは第三者の調査によるマイクロデータを利用したのかを、判断できないものもあった。論文を執筆する際には、分析対象となるマイクロデータの作成者がわかるようにすべきであろう。また、明

らかに第三者が作成したマイクロデータを再分析した論文でも、データの入手方法に関して記載のないものもあった。データの入手方法を記載することは、第三者が当該論文の内容を検証するために不可欠な情報であることから記載が求められるものである。

さらに、筆者らが独自に実施した調査によるマイクロデータによる論文の場合に、当該マイクロデータが公開されていない場合がかなりを占めることを指摘した。マイクロデータに基づく実証研究の再現性を担保するだけでなく、マイクロデータを社会科学の共有財産とするためには、マイクロデータの公開を義務づけることが必要となろう。とりわけ科学研究費による研究や公的な機関が実施したマイクロデータに関しては、そのことが緊急の課題である（日本学術会議政府統計の作成・公開方策に関する委員会（2006）など）。また、『日本労働研究雑誌』が論文を掲載する際に、マイクロデータによる研究に関しては、第三者がそれにアクセスできること、つまり再現性の担保を原則的な掲載条件とすることも検討に値しよう。

VI おわりに

データアーカイブの基本的な機能はマイクロデータの収集・保存・提供にあるが、それ以上に重要な機能は、社会科学におけるマイクロデータに基づく実証研究の「再現性」を担保する研究環境を提供することにある。「再現性」の担保だけでなく、マイクロデータが寄託されることによって、同一のデータセットに基づいてより説明のある理論を構築できるようになるのである。他方、労働研究の分野においてマイクロデータを活用した研究として『日本労働研究雑誌』に掲載された論文を参考にすると、マイクロデータを利用した論文が増加しているものの、当該データを第三者が入手し、その内容を検証できる条件を満たしていないと判断できる論文も少なくないことが明らかにされた。こうしたことから、マイクロデータに基づく実証研究の再現性を担保するだけでなく、マイクロデータを社会科学の共有財産とするために、マイクロデータの寄託を義務づけることを検討することが必要

表2 ミクロデータを利用した論文（『日本労働研究雑誌』に2003年から2005年の3年間の掲載分）

号	著者	所属・職位（論文掲載時による）	論文名	使用データ名	データの種類	第三者の入手可能性
511	松繁寿和	大阪大学教授	大手証券倒産後の再就職——年齢と技能の役割の検証	1997年に自主廃業した証券会社の元社員の就職先リスト	筆者が企業の人事部から入手	不明
511	竹内英二	国民生活金融公庫総合研究所上席主任研究員	廃業経験者による開業の実態	国民生活金融公庫総合研究所「2度目の開業に関するアンケート」	筆者らが実施した調査	○SSJDA
512	篠崎武久 石原真三子 塩川崇年 玄田有史	学習院大学大学院博士課程 城西大学講師 (株)労働実務「労基旬報」編集部 東京大学社会科学研究所助教授	パートが正社員との賃金格差に納得しない理由は何か	日本労働研究機構「職場における多様な労働者の活用実態に関する調査」	執筆者のひとりである玄田が参加した研究会が実施した調査の二次分析	○SSJDA
514	大木栄一	日本労働研究機構副主任研究員	企業の教育訓練投資行動の特質と規定要因	「業績主義時代の人事管理と教育訓練投資に関する調査研究」	筆者が参加した研究会の調査	不明
514	黒澤昌子	政策研究大学院大学助教授	公共職業訓練の収入への効果	東京都立職業訓練校修了者の「修了生実態調査」	二次分析	○目的外利用申請による
515	宮本大	同志社大学大学院経済学研究科博士課程	NPOの労働需要——国際および環境団体の雇用に関する実証分析	平成11～13年科学研究費基盤B研究（研究代表中田喜文）「エマージング労働市場の研究」のNPO団体調査とNPOの個人調査	筆者も調査メンバーと考えられるが、論文のみからは確定できない	不明
515	西村智	関西学院大学経済学部専任講師	育児による女性の生涯所得減少効果の日仏比較	フランス国立統計経済研究所(INSEE) “Enquête emplois du temps de ménages de 1998-1999” (生活時間調査) 日本労働研究機構「女性と仕事に関するアンケート」	二次分析、ただし日本のデータを含め、データの入手方法が論文には未記載	○調査名が異なるが日本労働研究機構の「女性の就業意識と就業行動に関する調査」の再分析と考えられ、同調査はSSJDAに寄託済。フランスのデータの利用方法は不明。
516	太田聰一 照山博司	名古屋大学経済学部助教授 京都大学経済研究所教授	フローデータから見た日本の失業——1980～2000	「労働力調査」の個票	政府統計の二次分析	○目的外利用申請による
516	久田満 高橋美保	東京女子医科大学教授 元信州大学附属病院臨床心理士	リストラが失業者および現役従業員の精神健康に及ぼす影響	調査名不明（2つのアンケート調査）	筆者らが実施した調査	不明
517	宮本大	同志社大学大学院経済学研究科博士課程	非営利団体（NPO）の賃金は低いのか——営利企業（PO）との比較を通して	平成11～13年科学研究費基盤B研究（研究代表中田喜文）「エマージング労働市場の研究」のNPO団体調査とNPOの個人調査	筆者も調査メンバーと考えられるが、論文のみからは確定できない	不明
517	大谷剛 梅崎 修 松繁寿和	大阪大学大学院法政大学 大阪大学	若年期の蓄積とキャリアの終盤——大学教育、移動、生え抜き役員昇進	国立大学社会科学系学部の卒業生調査	筆者らが実施した調査と考えられるが、論文のみからは確定できない	不明
518	石原真三子	城西大学専任講師	パートタイム雇用の拡大はフルタイムの雇用を減らしているのか	「雇用動向調査」事業所票	政府統計の二次分析	○目的外利用申請による
518	原ひろみ	労働政策研究・研修機構研究員	正規労働と非正規労働の代替・補完関係の計測——パート・アルバイトを取り上げて	生命保険センター「企業の福利厚生制度に関する調査」	二次分析	○SSJDA

518	佐藤博樹 佐野嘉秀 原ひろみ	東京大学社会科学研究所教授 東京大学助手 日本労働研究機構研究員	雇用区分の多元化と人事管理の課題——雇用区分間の均衡処遇	「雇用管理の現状と新たな働き方の可能性に関する調査研究委員会」が実施したアンケート調査	筆者らが実施した調査	○SSJDA
518	西本万映子 今野浩一郎	湘南国際女子短期大学講師 学習院大学教授	パートを中心にした非正規社員の均衡処遇と経営パフォーマンス	厚生労働省「パートタイム労働者の均衡処遇と経営パフォーマンスに係る調査」	筆者らが参加した研究会で実施した調査	○SSJDA
520	今井保次	(財)社会経済生産性本部メンタル・ヘルス研究所研究主幹	健康な職場の実現	メンタルヘルス研究所のメンタルヘルスのデータ	メンタルヘルス研究所が実施した調査の二次分析	不明
524	高山与志子	ジェイ・ユーラス・アイアール(株)マネージングディレクター	米国における非正規雇用の実態	米国労働局「Contingent and Alternative Work Supplements to the Current Population Survey」	二次分析、ただし論文にはデータの入手方法が未記載	○ICPSR 国内利用協議会を通じて ICPSR からデータを入手可能
524	馬駿	富山大学経済学部助教授	日本企業の昇進昇給システムにおける「年功」——ある大手企業のブルーカラー従業員に対するコーホート分析	大手製造業X社Y製作所のブルーカラー従業員の人事データ	筆者が人事部から入手	不明
526	木村琢磨 佐野嘉秀 藤本真 佐藤博樹	東京大学大学院経済学研究科 東京大学社会科学研究所客員助教授 労働政策研究・研修機構研究員 東京大学社会科学研究所教授	製造分野における請負企業の事業戦略と人事管理の課題	生産職場における人材活用・人材ビジネスに関する研究会「第2回請負企業の経営戦略と人事戦略に関するアンケート調査」	筆者らが実施した調査	○SSJDA に寄託済み、2006年秋に公開予定
527	加藤隆夫	コルゲート大学経済学部教授	従業員代表制の経営参加度とその決定要因——計量分析	HRM Survey of Japanese Firms と日経財務データ	筆者らが実施した調査	不明
527	西本真弓	阪南大学専任講師	育児休業取得とその取得期間の決定要因について	(財)連合総合生活開発研究所「仕事と育児に関する調査」	二次分析	○SSJDA
527	武内真美子	大阪大学大学院博士課程	女性就業のパネル分析——配偶者所得効果の再検証	(財)家計経済研究所「消費生活に関するパネル調査」	二次分析	○家計経済研究所に申請すれば可能
528	小原美紀	大阪大学大学院助教授	雇用保険制度が長期失業の誘引となっている可能性	(株)構造計画研究所「雇用政策の有効性に関するシミュレーション：実査データ」(経済産業省の委託調査)	二次分析	不明(調査を実施した構造計画研究所に目的外利用を申請すれば可能か不明)
528	岩田正美	日本女子大学人間社会学部教授	誰がホームレスになっているのか?——ポスト工業社会への移行と職業経験等からみたホームレスの3類型	都市生活研究所「平成11年路上生活者実態調査」	著者が調査を行った研究会に参加していたかどうか不明	不明
528	大竹文雄	大阪大学社会経済研究所教授	失業と幸福度	「くらしと社会に関するアンケート」	筆者が実施した調査	不明
529	久保克行	早稲田大学商学部専任講師	合併に伴う人事制度の統合と雇用・処遇の変化——個人データによる分析	1999年に合併したA社の人事データ	筆者らが入手した人事データ	不明
531	西岡由美	湘北短期大学助手	技能実習生の活用実態と日本人社員との代替関係について	財団法人雇用開発センター「日系人と技能実習生の就労実態等について」	筆者が調査を行った研究会に参加していたか不明	不明
532	吉田恵子	大阪大学大学院博士課程	自己啓発が賃金に及ぼす効果の実証分析	(財)家計経済研究所「消費生活に関するパネル調査」	二次分析	○家計経済研究所に申請すれば可能

紹介 データアーカイブの役割とSSJ データアーカイブの現状

532	原ひろみ 佐藤博樹	労働政策研究・ 研修機構研究員 東京大学社会科学 研究所教授	労働組合支持に何が影響 を与えるのか——労働者 の権利に関する理解に着 目して	(財) 連合総合生活開発研究所 「第5回勤労者の仕事と暮らし についてのアンケート調査」	筆者らが参加した研究会 会で実施	○SSJDA
533	安達智子	大阪教育大学講 師	大学生のキャリア選択 ——その心理的背景と支 援	正社員としての就労経験を持た ない18～24歳の学生を対象にし た調査	筆者が実施した調査か どうかの記載がなく不 明	不明
534	佐藤博樹	東京大学社会科 学研究所教授	若年者の新しいキャリア としての「未経験者歓迎」 求人と「正社員登用」機 会	(財) 連合総合生活開発研究所 「若年者の職業選択とキャリア 形成に関する調査」	筆者が参加した研究会 が実施した調査	○SSJDA
534	井川静恵	大阪大学大学院 国際公共政策研 究科	制度改定による賃金構造 の変化——企業内人事マ イクロデータによるパネ ル分析	ある企業の人事マイクロデータ	筆者が人事部から入手 したデータ	不明
534	小川浩	関東学園大学経 済学部助教授	賃金制度を少子化から評 価する	「国民生活基礎調査」の個票	政府統計の二次分析 (厚生科研「社会保障 政策が育児コストを通 じて出生行動および消 費・貯蓄行動に及ぼす 影響に関する研究」内 で行われたもの)	○目的外利用申請による
535	岩澤美帆 三田房美	国立社会保障・ 人口問題研究所 主任研究官 国立社会保障・ 人口問題研究所 主任研究官	職縁結婚の盛衰と未婚化 の進展	国立社会保障・人口問題研究所 「出生動向調査」1982～1992年 の個票データ	政府統計の二次分析 (社人研の研究員として 分析)	○目的外利用申請による
535	酒井正 樋口美雄	慶應義塾大学大 学院 慶應義塾大学商 学部教授	フリーターのその後—— 就業・所得・結婚・出産	慶應義塾大学経商連携21世紀 COEプログラム「市場の質に 関する理論形成とパネル実証 分析」第1回慶應パネル調査	筆者らが参加した研究会 会で実施した調査	不明
535	川口章	同志社大学政策 学部教授	結婚と出産は男女の賃金 にどのような影響を及ぼ しているのか	(財)家計経済研究所「消費生活 に関するパネル調査」	二次分析	○家計経済研究所に申 請すれば可能
536	高橋陽子	学習院大学大学 院	ホワイトカラー「サービ ス残業」の経済学的背景 ——労働時間・報酬に関 する暗黙の契約	(財) 連合総合生活開発研究所 (2002)「働き方の多様化と労働 時間等の実感調査」	二次分析	○SSJDA
538	阿部正浩	獨協大学経済学 部助教授	男女の雇用格差と賃金格 差	「就業構造基本調査」1982年と 1997年 「賃金構造基本統計調査」平成 13年 「女性雇用管理基本調査」平成 13年	政府統計の二次分析	○目的外利用申請による
540	小倉一哉	労働政策研究・ 研修機構副主任 研究員	長期休暇が企業経営に与 える影響——因果関係に 関する一考察	労働政策研究・研修機構 (2004)「労働者の働く意欲と雇 用管理のあり方に関する調査」	二次分析	○SSJDA
540	大木栄一	職業能力開発総 合大学校助教授	長期休暇と能力形成	厚生労働省委託調査「能力開発 基本調査報告(平成14年度)」、 富士総合研究所(2000)「能力 開発等の活動に取り組むための 長期休暇制度の導入促進に向け た調査研究」	筆者が参加した調査 (前者)と協力した調 査(後者)、後者は二 次分析	不明
541	青島矢一	一橋大学イノー ベーション研究セ ンター助教授	R&D人材の移動と技術 成果	応用物理学会の会員5000人を対 象に行った調査(1999年実施)	筆者が実施した調査	不明
542	原ひろみ	労働政策研究・ 研修機構研究員	新規学卒労働市場の現状 ——企業の採用行動から	労働政策研究・研修機構 (2004)「若年者の採用・雇 用管理の現状に関する調査」	筆者が参加した研究会 で実施した調査	○SSJDA

543	緒方一子	東京地下鉄(株)臨床心理士	キャリア・コンサルティングと個人情報保護法の活用と保護	キャリアコンサルタント300人を対象に行った調査(2005年実施)	筆者が実施した調査	不明
543	開本浩矢	兵庫県立大学経営学部助教授	成果主義導入における従業員の公正感と行動変化	富士ゼロックス総合研究所(2001)「成果主義的人事制度に関する社員意識調査」	筆者が実施した調査か記載がない	不明
545	梅崎 修 中嶋 哲夫	法政大学キャリアデザイン学部専任講師 人事教育コンサルタント	評価者負担が評価行動に与える影響 ——「人事マイクロ・データ」と「アンケート調査」の統計分析	B社(生産財製造業)で実施した従業員アンケートとB社企業内人事マイクロ・データ	筆者らが入手した個別企業人事データと当該企業が実施したアンケート調査のデータの二次分析	不明

注:1) 二次分析=筆者以外が調査を実施したデータを筆者が分析したもの。
2) SSJDA=SSJ データアーカイブ。

となろう。

日本語文献

佐藤博樹・石田浩・池田謙一編(2000)『社会調査の公開データ』東京大学出版会。
美添泰人(2005)「統計データの保存と再利用の体制」『統計』2005年6月号, pp.32-37, 日本統計協会。
濱砂敬郎(2000)「2.3 事実上の匿名性の原則」『講座マイクロ統計分析①統計調査制度とマイクロ統計の開示』日本評論社。
東京大学社会科学研究所(2005)『東京大学社会科学研究所附属日本社会研究情報センター 外部評価報告書』。
前田幸男(2004)「世論調査データの行方——データ・アーカイブの役割」『中央調査報』(No.558)中央調査社。
稲葉昭英(2005)「社会調査および公共利用データをめぐっていくつかの問題——全国家族調査(NFRJ)からの問題提起」『中央調査報』(No.572)中央調査社。
日本学術会議政府統計の作成・公開方策に関する委員会

(2006)『政府統計の改革に向けて——変革期にある我が国政府統計への提言』日本学術会議。

英語文献

Collins, Patrick T. (1996) "Establishing Data and Documentation Standards for Investigators Who Are Required to Archive Research Data." *IASSIST Quarterly* 20, no. 3: 21-23.
ICPSR (2005) "Building Partnerships & Leadership." *2004-2005 Annual Report*.

さとう・ともひこ 現・総務省統計局統計調査部消費統計課課長補佐。最近の主な著作に廣松毅・佐藤朋彦・高木新太郎・木村正一『経済統計』(新世社, 2006年)。
さとう・ひろき 東京大学社会科学研究所教授。最近の主な著作に佐藤博樹・佐藤厚編著『仕事の社会学』(有斐閣, 2004年)。