

オーストラリアにおけるデータアーカイブと マイクロデータの提供*

勝 浦 正 樹・小 林 健太郎

はじめに

統計データアーカイブとは、「統計調査などで得られた個票データを整理、収録し、その散逸を防ぐとともに、当初の利用を超えた二次的な利用のために提供するための機構⁽¹⁾」である。統計データアーカイブ（以下、単にデータアーカイブと呼ぶ）は、後にみるようにイギリス・アメリカなどにおいて大規模な機構としてかなり以前から存在し、政府統計・民間統計を含んだ多種多様なマイクロデータを提供し、多くの利用者がそのデータを利用して様々な分析がなされてきた。

こうした欧米の状況に比して、わが国に本格的なデータアーカイブとして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にSSJデータアーカイブ（Social Science Japan Data Archive）が設立されたのはごく最近になってからで、1996年のことである⁽²⁾。ただし、このデータアーカイブで扱われているのは民間のデータのみであり、民間の調査実施者からデータの寄託を受け、それを保存し、利用希望者に提供することによって、二次的利用を行うための窓口となっている。SSJデータアーカイブの他にも、大阪商業大学、札幌学院大学、家計経済研究所などで、利用できるデータはそれほど多くはないものの、二次的利用のためのデータが提供されている。

しかしながら、政府統計に関しては、わが国で本格的なデータアーカイブは存在しておらず、政府統計の二次的利用が諸外国に比べて積

極的に行われているとはいえない。二次的利用の拡大については、10年以上前に出された総務庁統計局統計基準部（1995）の「統計行政の新中・長期構想」においてすでに指摘され、さらに各府省統計主管部局長等会議（2003）、そして小泉内閣の構造改革路線をも背景とした内閣府経済社会統計整備推進委員会（2005）などでも要請されているところである。試験的には、政府統計についても、いくつかのプロジェクトで匿名標本データを中心に提供されてはいる⁽³⁾。また、こうしたプロジェクトを通じてではなく統計データの「目的外使用」として、直接、総務省に申請し、個票データを利用する方法もある。しかしながら、申請手続きの煩雑さや審査期間の長さなど様々な制約があり、必ずしも利用勝手がよいとはいえない。このような背景から、わが国においても、政府統計に関する本格的なデータアーカイブの設立が望まれているのである。

もしわが国にこうしたデータアーカイブが設立されるとしたら、どのような形態が望ましいのであろうか。もちろん、わが国の統計制度に適ったものがよいだろうが、既に存在している諸外国のデータアーカイブを参考にすることも必要であらう。本論文の目的は、わが国のデータアーカイブの設置に参考となるであろう国の1つとしてオーストラリアを選び現状を概説した上で、それを参照しながらわが国のデータアーカイブの望ましい形を展望することである。オーストラリアを選定した理由は、マイクロデータの提供の体系が、イギリスやアメリカな

どに比べて比較的小規模で、これから政府統計のデータアーカイブの設置を考えるわが国にとって、参考にしやすいと思われるからである。

本論文の構成は以下の通りである。第1節で、諸外国のデータアーカイブに関してごく簡単に要約する。第2節では、オーストラリア統計局 (Australian Bureau of Statistics; ABS) によって提供されている政府統計のマイクロデータである CURFs (Confidentialised Unit Record Files) に関して、いくつかの側面から現状をみていく。さらに第3節では、オーストラリアの唯一のデータアーカイブとして ASSDA (Australian Social Science Data Archive; オーストラリア社会科学データアーカイブ) を取り上げ、その設立の経緯やマイクロデータの提供状況について概説する。そして第4節では、それらを踏まえた上で、わが国においてどのようなデータアーカイブやマイクロデータの提供が望ましいのかに関してコメントする。

1. 諸外国のデータアーカイブ

本節では、データアーカイブの先進国である諸外国におけるデータアーカイブについて簡単に紹介することにする。なお詳細は、統計研究会 (2003, 2005) や内閣府経済社会統計整備推進委員会 (2005) などを参照されたい。

1.1 イギリス

イギリスには、いくつかのデータアーカイブがあるが、その中で有名なのは、マンチェスター大学内にある CCSR (Cathie Marsh Center for Census and Survey Research) やエセックス大学内にある UKDA (The U. K. Data Archive) である。CCSR では、センサスのマイクロデータが提供され、UKDA は労働力調査をはじめとする政府統計ならびに民間統計を提供している。CCSR で提供されるセンサスのマイクロデー

タは、SARs (Samples of Anonymised Records) と呼ばれ、センサスの全データでなく、1-2%程度が抽出されたデータである。

この他にも、Longitudinal データを提供するロンドン大学内にある CeLSIUS (Centre for Longitudinal Study Information and User Support) などもあり、さらに国家統計局 (The Office for National Statistics; ONS) もマイクロデータを提供している。

1.2 アメリカ

アメリカのデータアーカイブとしては、ICPSR (Inter-university Consortium for Political and Social Research) が最も有名であろう。この機関は大学等の連合体 (コンソーシアム) であるが、センサス、CPS (Current Population Survey) といった政府統計ならびに民間統計を多数提供している。また、データの提供だけでなく、サマーセミナーを実施し、マイクロデータの利用方法や実際の分析などに関する教育活動も行っている。マイクロデータの提供についてはこの他にも、たとえば DataWeb において、専用のブラウザを通じて、政府統計のマイクロデータにアクセスすることができる。

1.3 その他の国

アメリカやイギリス以外でも、多くの国でデータアーカイブによって、マイクロデータが提供されている。

カナダにおいては、カナダ統計局 (Statistics Canada) によって管轄される PUMFs (Public Use Microdata Files) や DLI (Data Liberation Initiative) などといった形でマイクロデータが提供されている。またドイツにおいては、連邦統計局 (Federal Statistical Office Germany) や ZUMA (Centre for Survey Research and Methodology) によって、センサスを含んだマイクロデータが提供されている。この他にもオー

ストリア、オランダ、フランス等々多くの国で、規模や提供されるデータに違いはあるもののデータアーカイブが設置され、政府統計のマイクロデータが提供されている。

本論文では、これらのうち、わが国のデータアーカイブやマイクロデータの提供の将来的な展望を行うために、オーストラリアを取り上げる。オーストラリアを選択した理由は、民間統計に関しては、わが国と同様に1つの主要なデータアーカイブが存在していること、政府統計に関しては、オーストラリア統計局(Australian Bureau of Statistics; ABS)が専ら提供しており、わが国の今後の実現可能と思われる状況との対応関係がみやすいであろうことの2点である。

2. ABS と CURFs

よく知られているように、オーストラリアの統計機構は、わが国のような分散型とは異なり、集中型である。集中型としてのオーストラリアの統計制度の中心となる機関は、ABSである。もちろん、様々な統計が他機関によっても調査されているが、センサスをはじめ主要な統計調査はABSによって実施、提供されている。

データアーカイブとは、はじめに述べたように、政府による統計調査や民間のアンケート調査など様々な調査について、その個票データならびに調査票などの関連する資料(メタデータ)の寄託を受け、必要であれば秘匿などの措置を施した上で、それらを保管し、二次的な利用のためにそれらを第三者に提供する組織である。このような寄託も含めた定義からすると、集中型統計機構の中心であるABSをデータアーカイブに位置づけることには、若干の違和感があるかもしれない。しかし、ABSはCURFsと呼ばれるマイクロデータの提供を自ら行っている。そこで、本節では、データアーカイブとしての

ABSの機能、すなわちABSによるマイクロデータの提供状況についてみていくことにする⁽⁴⁾。

2.1 CURFs とは⁽⁵⁾

ABSは、オーストラリアの政府統計の中心として、センサスをはじめ、家計調査、労働力調査、住宅調査など主要な調査を実施している。それらのデータは、報告書の形だけでなく、インターネットを通じてもちろん公表されるようになっている。さらにABSでは、主要な調査のマイクロデータをCURFs(Confidentialised Unit Record Files; 秘匿化された個票ファイル)と呼び、一定の制約のもとで提供を行っている。CURFsでは、その名前からもわかるように、当然のことながら秘匿のため名前や住所など個人の特特定性が可能となるような情報は取り除かれており、さらにその他の情報に関して、秘匿のために必要に応じて制限されたり、変更されるといった処理も行われている。イギリスのデータアーカイブでは、秘匿化されたマイクロデータを前述のようにSARs(Samples of Anonymised Records)などと呼んでいるが、CURFsはそれに対応するものと考えればよいだろう。

CURFs提供の法的根拠となっているのは、オーストラリアの「センサスおよび統計法1905(Census and Statistics Act 1905)」の第13条第1項、ならびに「統計省令1983(Statistics Determination 1983)」の第7条である。前者が統計情報の開示のための法令を別途定めることを示し、後者ではより具体的に、個票データから氏名や住所などを除去し、個人を特定できないようにした上で統計情報を開示できることを明示したものである⁽⁶⁾。このように、マイクロデータの提供に関して法的な裏付けが存在しているのは、わが国の状況とは大きく異なっている。

2.2 CURFs で提供されるマイクロデータ

ABS の CURFs のうち主要なものが、表1に示されている。これらのデータは、基本的に ABS によって実施されている調査であり、センサス (Census; 国勢調査) をはじめ、住宅調査 (Housing Survey)、家計調査 (Household Expenditure Survey)、所得・住宅費調査 (Income and Housing Costs Survey)、労働移動調査 (Labour Mobility)、国民健康調査 (National Health Survey)、生活時間調査 (Time Use

Survey) などがあり、州に関するデータも含まれている。

各調査で提供されているデータ等の内容 (変数や回答項目など) は、ウェブサイトのそれぞれの調査のリンクをたどることによって確認することができる。なお、センサスに関しては、世帯標本ファイル (Household Sample File) が提供されている⁽⁷⁾。その標本は、民間住居から1パーセントの世帯が無作為に抽出され、その構成要員が対象となるとともに、さらに非民間

表1 ABSで提供される主なCURFs

調査名	年次	提供内容	調査名	年次	提供内容
Australian Housing Survey	1994	Basic	Household Expenditure Survey and the Survey of Income and Housing	2003-04	Basic & Expanded
	1996	Basic			
Census — Household Sample File —	1981	Basic	Income and Housing Costs Survey	1990	Basic
	1986	Basic		1994-95	Basic
	1991	Basic		1995-96	Basic
	1996	Basic		1996-97	Basic
	2001	Basic & Expanded		1997-98	Basic
2001	Basic & Expanded	1999-00		Basic	
Child Care Survey	1993	Basic	Jan-00	Basic & Expanded	
	1996	Basic	Labour Mobility	1984	Basic
	1999	Basic & Expanded		1991	Basic
	2001	Expanded		1994	Basic
2005	Expanded	National Health Survey	1977-78	Basic	
Disability, Ageing and Carers	1993		Basic	1983	Basic
	1998		Basic	1989-90	Basic
	2003		Basic	1995	Basic
Education and Training Experience	1989		Basic	2001	Basic & Expanded
	1993	Basic	2004-05	Basic & Expanded	
	1997	Basic	National Health Survey, Indigenous	2001	Expanded
Education and Training	2001	Basic		Rental Investors' Survey	1993
	2005	Basic	1997		Basic & Expanded
Education and Work	2001	Basic	Time Use Survey	1992	Basic
	2003	Basic		1997	Basic
	2005	Basic			
Household Expenditure Survey	1975-76	Basic			
	1984	Basic			
	1988-89	Basic			
	1993-94	Basic			
	1998-99	Basic			

(出典) ABSのウェブサイト

住居より1パーセントの無作為標本も抽出され、対象に加えられる。そして、住居に関する情報(住居のタイプ、自動車の数など)、世帯に関する情報(世帯収入、家族構成など)、個人に関する情報(性別、年齢、出生地、労働状態など)がマイクロデータとして提供される。

また、表1に示されているようにCURFsには、Basic CURFsとExpanded CURFsの2種類がある⁽⁸⁾。表にみられるように過去の調査ではほとんどBasic CURFsしか利用できないが、最近の調査については、BasicとExpandedの両方が利用可能になっている場合が多い。

BasicとExpandedの違いは、Expandedの方がより詳細な情報を含むということであるが、両者の違いは調査ごとに異なっており、相対的なものである。表2では、2001年のセンサスを例にとり、両者の違いをいくつかの変数について示してある。年齢が両者の違いを示す典型的な変数の1つであり、Basic CURFsでは25歳以上は5歳階級別であるが、Expanded CURFsでは1歳階級となっている。また、たとえば家族構成は、Expanded CURFsではもとの調査項目がそのまま利用できるが、Basic CURFsでは一部の項目が統合されている。両者の違いは、標本の大きさや利用できる変数の数ではなく、変数の中でどの程度詳細な情報が得られるかである。

2.3 CURFsの利用条件

2.3.1 申請

CURFsの利用は、所定の申請書類(誓約書を含む)を提出し、ABSによる審査を通過することによって可能となる。審査結果は2週間以内で申請者に連絡される。また、利用者はCURFsを12か月間保有することができ、延長のオプションもある。申請書類は、CURFsのウェブサイトからダウンロードできるようになっている。

ここで申請者は、自らの連絡先や利用するCURFs名などの基本的な情報を提示するとともに、統計目的以外で利用しない、個人や組織を特定するような試みを行わない、他人に情報を提供しない、CURFsのトレーニング・マニュアル(Responsible Access to ABS Confidentialised Unit Record Files (CURFs) Training Manual)の内容に従って利用する、などを誓約することになる。

また、上記に誓約事項に違反する行為(個人や組織を特定したり、他人へ情報を漏えいした場合)に対しては、前述の「センサスおよび統計法1905」の第19条第2項に基づき、5000豪ドル以下の罰金、または2年以内の懲役(またはその両方)の罰則が用意されている。

表2 Basic CURFsとExpanded CURFsの主な違い(2001年センサス)

変数	Basic CURFs	Expanded CURFs
地域	48地域(人口250,000以上)	63地域(人口124,000以上)
世帯収入	最高階級は1500豪ドル以上	最高階級は2,000豪ドル以上
家族構成	項目を統合	項目は統合なし
年齢	24歳までは1歳階級 25-84歳は5歳階級 85歳以上は1つの階級	84歳までは1歳階級 85歳以上は1つの階級
職業	1ケタ分類	2ケタ分類

(出典) Trewin (2003).

2.3.2 User - Contact Officer - Responsible Officer

CURFs が提供され、実際にそれを利用するのはもちろんユーザー (User) であるが、CURFs を申請・利用するにあたっては、ユーザーの上に、Contact Officer (連絡者) と Responsible Officer (責任者) という2つの役割を果たす要員をユーザーの所属する機関の中から選び、申請しなければならない。ただし、CURFs を利用する主要な大学や政府機関などでは、これら Contact Officer と Responsible Officer はすでに決まっている。

Responsible Officer は、その機関に属する CURFs の利用者に関してまさに責任を負う者であり、大学では学長・副学長クラス、政府機関では各省の長官・事務次官クラス、民間機関では CEO クラスの人などが原則としてこの役にあたる。このような人たちが Responsible Officer になるということは、CURFs を利用するのに所属機関のトップクラスの人の承認が必要となることを意味する。しかし、当然のことながら、こうした Responsible Officer が CURFs のユーザーの利用状況を直接監視するわけではない。むしろ、地位の高い人を Responsible Officer におくことには、CURFs の利用には重大な責任が伴うことをユーザーに知らしめるという側面が強いのであろう。Responsible Officer は、次に示す Contact Officer から定期的にその機関における CURFs の利用状況などの報告を受けることになっているが、実際には必要な署名を行う程度の役割であると思われる。

一方、CURFs のユーザーと直接の接点をもち、ABS とも密接に関わるのが、Contact Officer である。たとえばある大学で Contact Officer となっている人は、CURFs の利用可能なファイルリストを管理したり、所属機関における CURFs 利用希望者の申請書をまとめて

ABS に提出したり、CURFs の CD-ROM の保管・配布をしたり、利用終了した場合にそれらを ABS に返却したりなどといった実務的な役割を担う。また、後に説明する RADL (Remote Access Data Library) を利用するためのパスワードなどを ABS から受け取り、管理したりもする。このように Contact Officer は、簡単に言うと、ユーザーと CURFs の提供者である ABS との仲介役・連絡役である。

2.3.3 利用者の条件

ABS の CURFs は、利用に関して違法な行為さえしなければ、官庁職員や大学等の研究者以外でも、利用することが可能である。もちろん、申請書類がきちんと作成され、違法な利用をしないという誓約をした上で、さらに Contact Officer らを立てる必要がある (すでに Contact Officer 等がいる機関であれば必要はない)、その上で審査にパスすれば問題はないはずである。研究目的だけに限らず、商業目的での利用も認められている。

このように CURFs の利用は、申請が適切なものであれば、誰でも可能であると思われるが、申請者は国内に限られる。海外からの申請は原則として認められないが、ケース・バイ・ケースで考慮される場合もある。海外からの申請でも考慮されるのは、オーストラリアの大学や政府機関または各国の統計局と密接な提携関係にある機関の申請者の場合である。ただし、たとえそのような関係にあったとしても、申請が認められるという保証はない。

2.4 料金と提供形態

CURFs の利用料金は、表3に示されている⁹⁾。なお、料金には税金 (GST; Goods and Service Tax) が含まれている。さらに、Basic から Expanded への変更、Expanded から Basic への変更にも、1,320 豪ドルの追加料金

表3 CURFsの料金

CURFsへのアクセス方法	基本料金
CD-ROMによるBasic CURFs	1,320 豪ドル
RADLによるBasic CURFs	1,320 豪ドル
RADLによるExpanded CURFs	1,320 豪ドル
Basic CURFs (CD-ROMまたはRADL)とExpanded CURFs (RADL)	1,980 豪ドル

(出典) 表1に同じ。

が必要となる。

表からわかるように Basic CURFs と Expanded CURFs で料金の差はない。つまり、詳細なデータだから高いというわけではない。CURFs の作成や諸手続きにかかる費用はほとんど変わらないことを反映しているものと思われる。

CURFs のデータへは、以下の3つの形でアクセスすることができる。

- CD-ROM
- RADL (Remote Access Data Library)
- ABSDL (ABS Site Data Laboratory)

CD-ROM については、Basic CURFs のみで利用可能である。もちろん CD-ROM であるから、自らの計算環境で自由に扱えるという大きな利点がある。一方、RADL (Remote Access Data Library) は、インターネットを通じて大学などから ABS の IT 環境にある CURFs にアクセスするものであり、Basic CURFs および Expanded CURFs が利用可能である。より詳細なデータである Expanded CURFs は、RADL での利用に限られる、すなわち CD-ROM では利用できない。上で述べたように RADL 利用者は、Contact Officer を通じて、パスワードなどが与えられ、CURFs にアクセスすることになる。RADL では、システム内で提供される SAS と SPSS で CURFs のデータを分析することになっているが、出力の大きさや性質などに制約があり、ユーザーのアクセス状

況はモニターチェックされる。また、ABSDL は ABS 内での利用のみで可能な特殊な利用形態であり、他のデータセットをも統合して利用できるようになっている。

なお、CURFs のデータの提供書式は、SAS または SPSS に対応した形式で提供される。

2.5 学術機関での利用

大学での CURFs の利用については、特別な協定がある。ABS-AVCC CURF Agreement と呼ばれるもので、AVCC とは、Australian Vice Chancellors Committee (オーストラリア副学長委員会) のことである。

この協定には現在 32 の大学が参加しており、研究・教育目的で CURFs を利用する場合に限って、この協定のもとで、CURFs を「無料」で利用することができる。これらの大学には、当然、Contact Officer がおり、利用者は Contact Officer を通じて、申請を行うことになる (2.3.2 参照)。ただし、たとえその大学に属していても商業ベースのプロジェクトやコンサルティングに CURFs を利用しようとする場合には、この協定のもとでの利用はできず、所定の価格 (表3) で CURFs を購入しなければならない。

3. ASSDA によるマイクロデータの提供

はじめに述べたように、オーストラリアにおいて本格的なデータアーカイブとしてマイクロデータを提供しているのは、ABSを別にすれば、ASSDA (Australian Social Science Data Archive) が唯一の機関である。以下では、ASSDA の概要とマイクロデータの提供状況を見ていくことにする。なお、ASSDA の概要については、Breusch and Holloway (2004) に適切にまとめられており、参考になる。

3.1 ASSDA の組織

ASSDA は、オーストラリア国立大学 (Australian National University; ANU) の社会科学研究科 (Research School of Social Sciences; RSSS) の附属データアーカイブとして1981年に設置された。設立の目的は、社会、政治、経済などに関連するコンピュータで読み取り可能なデータを収集・保存し、二次的な分析が可能となるようにデータを利用者に配布することなどであった。そしてASSDAは2001年1月に組織が変更となり、ANUのACSR (ACSPRI Centre for Social Research; ACSPRI 社会研究センター) の中に組み込まれた。

ここでACSPRIとは、Australian Consortium for Social and Political Research Incorporatedのことで、オーストラリアの大学や政府機関の連合体として1976年に設立され、現在33大学と13政府機関等が加盟している社会・政治研究に関するコンソーシアムである。アメリカにおけるICPSR (1.2参照) を思い浮かべればよいが、機能は異なっている。ACSPRIの当初の目的は、データアーカイブとしての機能、すなわち、様々な調査データの収集・保存・提供、ならびに利用の推進 (データを利用した研究や教育のサポート) などであった。しかしながら、2001年の組織変更以降、ACSPRIはサ

マープログラム・ウィンタープログラムと呼ばれる社会科学におけるデータを利用した研究方法や研究技術に関するセミナーを行い、教育活動を行うことが主要な活動となっている。つまり、データの収集・保存・提供といった狭義のデータアーカイブとしての機能はACSPRIからは分離させられ、ASSDAがそれらの機能に特化している。

このような機能の分離は、大学等の連合体でありANUとは別組織であるACSPRIと、もともとANUという1つの大学の附属機関であるASSDAの管理・運営に関する境界があいまいになってしまったといった経緯によるものである⁽¹⁰⁾。さらにオーストラリアにおける政権が労働党から保守連合に1996年に交代したことをきっかけに、国立大学への補助金の削減などが行われ、ANUとACSPRIのどちらに属しているのかあいまいなASSDAに財政上の問題も生じた。こうしたことから、ASSDAがACSPRIから組織上分離し、ANUの附属機関としてACSRの中に組み込まれている。

もちろん、現在でも両者は密接な関係にあり、ACSPRIのオフィスもオーストラリア国立大学によって提供されているが、両者の機能や管理・運営は区別されており、別個の機関として扱われている。

また、現在ASSDAが組織として組み込まれているACSRは、ACSPRIとANUのRSSSの共同のイニシアチブによって設立され、ACSPRIの寄付やANUの助成等によって財政的な援助を受けている。ACSRの主要な機能は、社会科学等に関して重要な実証的研究を行ったり、それらをバックアップしたりすることであり、研究員がおかれ、さらに大学院生が在籍して実際の研究が進められている。こうした実証研究は、特に政策に関連した研究に重点がおかれ、RSSSとも研究の上で関連しあっている。

これらの3機関は、いずれも場所的にはANUにあるが、上記のようにそれぞれの関係は、過去の経緯などから言ってもかなり複雑でわかりにくい。組織的な関係はともかく、端的に言えば、それらの機能は次のように簡単に要約することができるだろう。

- ACSR：実証研究
- ASSDA：(狭義の)データアーカイブ
- ACSPRI：研究手法等に関するセミナーの実施

これらの組織は、このような明確な役割分担のもと、財政基盤や人的な関係などを密接にさせながら、データアーカイブの機能を包含しており、そこにASSDAが位置しているのである。

3.2 ASSDAで提供されるデータ

ASSDAでは、現在、1500程度のマイクロデータのデータセットが利用可能である。そのうちの約3分の2(1050程度)は、オーストラリアの大学研究者や民間調査機関が行った社会調査や世論調査である。それ以外のデータとして、政府機関等が関わった研究・コンサルティングの結果としてのデータ、オーストラリアの近隣諸国(ニュージーランド、パプアニューギニア、インドネシア、カンボジアなど)のデータなどがある。またABSによるセンサスのデータについては、一時はマイクロデータを提供していた時期もあったが、現在では1966・1971・1976・1981・1986・1991の各年の集計データのみがASSDAによって保存・提供されている。

提供されるデータは、ASSDAのウェブサイトで検索が可能である。主に、データは、

- Australian Studies
- Non-Australian Studies
- Historical Census Data⁽¹¹⁾

と分類され、それぞれについて、人口、経済、教育、労働、政治などに分けられている⁽¹²⁾。

各調査で提供されているデータの概要、すな

わち調査主体、調査概要、サンプリングの方法、サンプル数、変数の数や内容といった情報は、ウェブサイトのそれぞれの調査のリンクをたどることによって確認することができる(後述のようなインターネットを通じた簡単な無料の登録が必要)。ASSDAで保管・提供されているデータはいわゆるメタデータであり、データ本体以外の部分をウェブサイトで確認することができ、さらに基本的な集計表やオンライン上の簡単なデータ分析の結果まで得ることができる。

3.3 利用条件や提供形態

ASSDAからデータの提供を受けようとする者は、まず、氏名、所属、希望データなどといった基本的な情報を申請書類に記入する必要がある。ただし、所属がACSPRIに加盟している機関かどうかによって、申請方法や利用データの制約に違いがある。

これが受け入れられた後、さらに申請者は、誓約書に署名することが要求される。署名によって申請者は、回答者のプライバシーを保護すること、研究成果を公表する際にデータの寄託者・配布者への謝辞を記すこと、公表された出版物を2部提出すること、ASSDAの許可なしにデータを第三者へ配布しないこと、などに同意することが求められる。また、いくつかのデータセットについては、寄託者によって、追加的な制限が設定されることもあり、こうした制限の有無によって利用方法が異なっている。こうした制限はいくつかのレベルに分類されている。

提供されるデータの利用料金としてASSDAが基本的に課すのは、1つのデータセットにつき1,000豪ドルである。ただし、所属する機関がACSPRIに加盟していれば、無料である⁽¹³⁾。データは基本的にインターネットを通じたダウンロードによって提供される。また、UK Data

Archive等によって開発されたウェブ上で稼動する統計分析ソフト (NESSTAR) によって、インターネット上で直接分析することができる。

結局のところ、利用者が ACSPRI に加盟している機関に所属しているかどうか、寄託者によってデータ利用に制限がかけられているかどうかで、利用条件は4通りに分類される。ACSPRI に加盟していない機関に所属している者でも、ASSDA のウェブサイト上で氏名や身分、e-mail アドレスなどの基本的な情報の登録をすることで、利用制限のないデータであれば、無料で NESSTAR を用いた基本集計表やオンラインでの簡単な分析程度は可能となる。ただしこの場合でも、データのダウンロードは有料となるし、制限のあるデータに関してはオンラインでの分析にも料金がかかる。

また、ASSDA で提供されるデータのフォーマットは、基本的に、汎用性の高い SPSS ポータブル (SPSS portable) ファイルである。ただしその他のフォーマット (SAS や Statistica など) でも提供される場合もある。

3.4 データの寄託

ASSDA がデータアーカイブとしての機能を果たすためには、様々なデータが収集されなければならない。そのためには、大学や民間機関等が実施した調査のデータが、実施主体から ASSDA に寄託 (deposit) されることが不可欠である。いかに有用なデータが寄託されるかが、データアーカイブの存立の基盤となるといってもよいだろう。

ASSDA では、積極的に寄託を受けるために、そのサイト内でもかなりのアナウンスをしている。データを寄託することの意義、すなわち研究結果の再現性や妥当性のチェック、寄託前の研究結果のさらなる発展、調査結果の保存等を含めたランニングコストの低減、同様のテーマ

に関心をもつ他の研究者との交流、といったことなどが強調されている。

ただし、データの寄託を受けるに際して、ASSDA では寄託者に対して財政的な援助を与えてはいない。もちろん、寄託に関して、様々な助言を与えることは行っている。また、寄託者は名前や住所といった個人情報を削除した上で、ASSDA にデータを寄託することになるが、これはすべての調査において、通常、調査実施者が回答者に対して、他人に個人情報をみせないということが明示的あるいは暗黙的な前提となっているからである。つまり、調査実施者と回答者からみれば第三者である ASSDA が、個人情報に触れることは好ましくないという立場である。しかしながら、ASSDA にデータが寄託された時点で、氏名や住所といった個人情報に関する変数が残っている場合には、個人の特定化につながる変数の削除といった単純な秘匿措置が ASSDA によって行われる (de-identify)。また、前述のように寄託者は、利用に対して、特定の利用制限を加えることもできる。

3.5 その他

以上のように、ASSDA の機能は、狭義のデータアーカイブ、すなわちデータの収集・保管・配布に特化しており、セミナーや実際の実証研究とは切り離されている。ただし、こうした機能に加えて ASSDA では、International Federation of Data Organisations (IFDO) のオーストラリアの窓口的な役割を果たしており、オーストラリアの研究者に IFDO に加盟するデータアーカイブの情報を伝えたりしている。また、将来的には研究者の要求に応じて加盟国のデータを獲得するような活動も視野に入れているようである。

ASSDA は、その機能は限定的であるものの、データアーカイブとして、イギリスの UK Data

Archive やアメリカの ICPSR に対応するオーストラリアの機関である。しかしながら、UK Data Archive と比較すると、大雑把に言って、予算や人員は 10 分の 1 程度しかない。こうした状況は、政府からの財政的な援助なども大きく関係している⁽¹⁴⁾。

4. わが国におけるデータアーカイブと政府統計のマイクロデータの提供

4.1 オーストラリアと日本のデータアーカイブ

以上、オーストラリアのデータアーカイブとして、ABS の CURFs と ASSDA の現状について概説してきた。両者について簡単に要約・比較したのが、表 4 である⁽¹⁵⁾。

両者の決定的な相違点は、CURFs が、集中型の統計機構をもつオーストラリアの統計局である ABS 自身で、自ら実施している政府統計をマイクロデータとして提供するものであるのに対して、ASSDA は、完全な仲介機関で、他機関が作成したマイクロデータを収集した上で、提供する形をとっていることであり、その結果、民間の統計がほとんどとなっている。わが国でいえば、ASSDA が SSJ データアーカイブに対応している。

一方、ABS の CURFs は、提供されているデータがそれほど多くないという問題はあるかもしれないが、主要大学から無料で利用可能であることなども含めて、日本のマイクロデータの提供・利用の現状に比べると、風通しのよいものになっていることは確かであろう。これは、

表 4 オーストラリアのデータアーカイブの比較

	ABS (Australian Bureau of Statistics) のCURFs	ASSDA (Australian Social Science Data Archive)
設立時期・マイクロデータ提供開始時期	・1984年(国民健康調査のデータ)が最初 ・本格的な提供は1990年代から	・1976年設立(RSSSの附属機関) ・2001年よりACSRに吸収される
提供されるマイクロデータ	・ABSで調査される諸調査(表1参照)で、センサスをはじめとする60調査程度 ・基本的なデータ(Basic)とより詳細なデータ(Expanded)が選択可能	・大学や民間機関等の社会調査や世論調査 ・近隣諸国の諸調査 など1500程度
利用条件	・ABSによる2週間以内の審査を通過した者が利用可能(原則として国内からのみ申請可能、例外あり)。 ・申請には、Contact OfficerとResponsible Officerを定め、統計目的以外に利用しないなどと誓約	・申請書を提出した後、誓約書を提出する。ほとんど誰でも利用可能 ・調査データの寄託者から利用者に対する付加的な制約があるデータもある
提供形態	・CD-ROM ・RADL(インターネット経由で直接分析、各大学で利用可能) ・ABSDL(ABS内のみで利用可能)	・インターネット経由でダウンロード ・NESSTARによる直接の分析も可能
提供フォーマット	・SAS ・SPSS ・RADL, ABSDLではSASやSPSSを直接利用する	・SPSSポータブルが中心
料金	・1,320豪ドル ・ABS-AVCC CURF協定に加盟している大学からの研究目的の利用は無料	・1,000豪ドル ・ACSPRIの加盟機関の場合は無料
データの寄託	・基本的にABSの実施する調査	・ASSDA外の大学研究者や調査機関からの寄託を募る

2.1で述べたように、秘匿処理を行ったデータの開示に関して法的に積極的な裏づけがあることの影響が大きいと思われる。一方、わが国の政府統計に関しては、現状では統計法の制約があり、はじめに述べたように、そのマイクロデータはオーストラリアや諸外国のように自由に利用できない。しかし、マイクロデータの提供に関して、もともとの個票データではなく、それを匿名化したCURFsのようなマイクロデータ（匿名標本データ）であっても利用したいという要望は強く⁽¹⁶⁾、実際に目的外使用の申請も継続的になされており、マイクロデータの提供や利用が促進されるように統計法が改正されることが望まれる。この点に関しては、総務省で「統計法制度に関する研究会」が設置され、統計データの二次的利用が推進されるような統計法制度の検討が行われ、その成果は統計法制度に関する研究会（2006）で取りまとめられている⁽¹⁷⁾。統計データの二次的利用に関して、法制度が整備され、同報告書で指摘されるような方向に進んでいくことが期待される。

4.2 望ましいデータアーカイブ

統計に関する法制度が二次的利用を推進するように整備された場合、どのようなマイクロデータの提供形態が望ましいのであろうか。もちろん、統計法の内容にもよるであろうが、後述するようにデータアーカイブが設置された上での提供となるであろう。わが国にとって、データアーカイブの設置は、二次的利用のためのデータの提供という機能だけでなく、メタデータの保存という意味でも重要である。現在のところ、個票データの保存に関しては、必ずしも適切に行われていない。

オーストラリアと違ってわが国の統計制度は、総務省、経済産業省、農林水産省等々の複数の省庁に散らばった分散型である。したがって、データアーカイブもそれぞれの統計作成機

関ごとに設置するという方法もある。もちろん、オーストラリアのABSのように1つの機関という方向も考えられる。統計法制度に関する研究会（2006）では、どちらがよいのかに関しては、今後も検討していく必要があるとしている⁽¹⁸⁾。

しかしながら、それぞれの機関に対応したデータアーカイブが設置されたのでは、様々な面（費用や場所だけでなく、専門家の育成等の人的な問題も含む）で効率が悪く、何よりも利用者からみれば、不便である。したがって、美添（2004）でも述べられているように、できれば単一のデータアーカイブであった方が望ましい。このように考えると、アメリカやイギリスなどよりも、オーストラリアのような形態、すなわち政府統計に関しては、ABSのみが提供するという単一の政府統計に関するデータアーカイブが、参考になるだろう。しかし、単一のデータアーカイブを設置する場合でも、分散型と集中型の違いを考慮すれば、わが国でABSに対応する総務省統計局自身が直接マイクロデータを提供することは望ましくなく、単一のデータアーカイブをそれに付随する機関や独立した機関（独立行政法人なども含む）として設置するといった形態の方が望ましいと考えられる。

もちろん、データアーカイブが設置されれば、データの収集・保存・提供といったデータアーカイブの基本的な機能だけでなく、利用者に対するセミナーの実施や、匿名化の方法等データアーカイブに関する研究や実証研究に対する要求も出されるであろう。もちろん、そうした機能を含めた規模の大きいデータアーカイブが設置されれば非常に望ましいが、オーストラリアのASSDA等の場合のように、機能をいくつかの機関に分散させるというのも1つの方法である。

いずれにせよ、オーストラリアをはじめとする諸外国の例も参照しながら、わが国でも政府

統計に関するデータアーカイブが設置され、政府統計を二次的に利用しやすくなる環境が整えられることが望まれている。しかも、できるだけ早い実現が、統計利用者の強いニーズである。

注

- (*) 本稿を作成するにあたり、ASSDA (Australian Social Science Data Archive) の Sophie Holloway 氏より多くの有用な助言をいただいた。記して謝意を表したい。もちろん本稿における誤りは、すべて筆者のみに帰するものである。なお本稿は、統計研究会 (2004) に所収の勝浦・小林「オーストラリアにおけるデータアーカイブの現状」を加筆・修正したものである。
- (1) 美添 (2004), 17 ページ。
- (2) SSJ データアーカイブおよび統計データの二次的利用全般については、佐藤他 (2000) に詳しい。
- (3) 統計研究会 (2006), 1 ページ。
- (4) 本稿の以下の記述は 2006 年 7 月時点での情報をもとにしており、その後の変更等は反映されていないので注意されたい。
- (5) CURFs については、ABS のウェブサイトやそこに掲載されている諸文書を参照した。
- (6) 統計研究会 (2003), 68 ページ。また、具体的な条文は、統計研究会 (2005), 86-88 ページを参照されたい。
- (7) Household Sample File の詳細については、Trewin (2003) を参照のこと。
- (8) Basic と Expanded 以外に、Specialist というレベルの CURFs も存在する。これは 2.4 で述べる ABSDL でのみ利用可能な、Expanded よりも詳細な情報が得られる特殊なデータである。
- (9) 以前は、提供されるデータの調査時期によって料金に差がついていた (新しい調査の方が高かった)。
- (10) Breusch and Holloway (2004), p. 228.
- (11) ASSDA のサイトをみると、センサスのデータに関しても、他のマイクロデータと並列されているためマイクロデータが提供されているように思えるが、提供されているのは集計データのみであることに注意されたい。
- (12) Breusch and Holloway (2004) pp. 224-226 では、データを申請する多くの研究者が興味をもっている

代表的な 4 つの分野 (Australian Election Studies, National Drug Strategy Survey, Negotiating the Life Course, Longitudinal Survey of Australian Youth) について概説されている。

- (13) Breusch and Holloway (2004) で示されている料金体系とは、最近では異なっているので注意されたい。
- (14) Breusch and Holloway (2004), p. 227.
- (15) 諸外国のデータアーカイブを比較するための同様の表が統計研究会 (2004) 91-98 ページに掲載されている。
- (16) はじめに述べたいいくつかのプロジェクトによるマイクロデータの提供は、匿名標本データである。また、こうしたデータや二次的利用全般、データアーカイブなどに関する要望は、統計研究会 (2006) で実際に利用者にアンケート調査され、それらに対するニーズが大きいことが明らかにされている。
- (17) 同報告書では二次的利用以外にも、統計調査の民間委託の推進なども検討されている。
- (18) 統計法制度に関する研究会 (2006), 35 ページ。

参考文献

- Breusch, T. and S. Holloway (2004), "Australian Social Science Data Archive," *Australian Economic Review*, 37 (2), 222-229.
- 各府省統計主管部局長等会議 (2003) 『統計行政の新たな展開方向』(各府省統計主管部局長等会議申合せ)。
- 内閣府経済社会統計整備推進委員会 (2005) 『政府統計の構造改革に向けて』(経済社会統計整備推進委員会報告書)。
- 佐藤博樹・石田浩・池田謙一編 (2000) 『社会調査の公開データ 2 次分析への招待』東京大学出版会。
- 総務庁統計局統計基準部監修 (1995) 『統計行政の新中・長期構想—統計審議会答申—』全国統計協会連合会。
- 統計法制度に関する研究会 (2006) 『統計法制度に関する研究会 報告書』(統計調査の民間委託・統計データの二次利用の促進について)。
- 統計研究会 (2003) 『国の統計調査に係るデータ・アーカイブに関する研究報告書』財団法人統計研究会。
- 統計研究会 (2005) 『「統計データアーカイブに関する調査研究」研究報告書』財団法人統計研究会。

統計研究会 (2006) 『統計データの二次的利用の仕組みの在り方に関する調査研究』研究報告書』財団法人統計研究会。

Trewin, D (2003), "Census of Population and Housing, Household Sample File, Australia 2001," Australian Bureau of Statistics Technical Paper 2037.0.

美添泰人 (2004) 「統計データアーカイブの現状と官庁統計における今後の課題」『青山経済論集』第56

巻 第2号, 17-38 ページ。

参考ウェブサイト (以下の URL より必要なリンクを辿っていく必要がある)

ABS

<http://www.abs.gov.au/>

ASSDA

<http://assda.anu.edu.au/>