

世論調査データライブラリーについて

統計数理研究所

名誉教授 鈴木達三

〈はじめに〉

日本世論調査協会に「データライブラリー小委員会」が出来たのは20年以上前のことである。

その頃はまだIBMカードの機械集計が一般的で、それが大型電算機の導入により集計作業は便利になったがまだ高嶺の花という頃である。その頃は報告書ができると個票データの廃棄等のことが日常的であり、集計分析も限られた属性クロスや質問相互のクロスを労力や時間を考慮して慎重にきめ機械集計していた。

その頃のデータライブラリーに対する考え方

と現在とでは大きく異なっていると考えられる。その理由の一つは調査企画、実施、結果整理、集計分析の環境がハード面・ソフト面で大きく変化したことである。調査の手順とハードとの関連を考えてみると、調査企画から調査票作成、調査実施、結果整理、集計等でまず結果整理、集計が電算化され、その後印刷がワープロ入力され、それに加えてソフト面では分析に各種統計ソフトや表計算ソフト、グラフ作成ソフトなどが利用されるようになったことである（表1参照）。

表1 調査手順とそのデータライブラリーとの関連

調査手順	電算機以前	大型電算機代	ワープロ・パソコン時代	データライブラリーとの関連
調査企画	タイプ複写	タゼロックス	ワープロ	文書データベース
調査票作成	タ(活版)イープリ	同 左	ワ(活版)イープリ	文書データベース
標本抽出計画	タイプ印刷	同 左	ワープロ	文書データベース
標本抽出				(調査地点抽出に国勢調査・調査区データベースを利用) (数値データベース)
調査票回収・点検整理				(この部分は従来通りの作業が主なものである)
調査票コーディング				
集計準備	カードパンチ	カードパンチ、キーボード入力	キーボード入力	数値データベース
集計(表)	IBM機械集計	電算集計	電算集計	文書データベース、数値データベース
分析	限られた計算方法の利	分析プログラム、プログラムパッケージ	統計ソフト(各種分析ソフト)	
コード・ブック作成	タイプ印刷	タイプ印刷(電算機作成)	ワープロ	文書データベース
報告書作成	タイプ印刷	印 刷	ワープロ(表グラフ作成ソフト)	文書データベース

データライブラリーは長い間、大型電算機の利用と関連づけて考えられており、初期に考えられていたデータライブラリーでは調査票はローマ字やカナで入力するか英語で入力するか等が話題となる程であった。

このようにデータライブラリーのイメージは技術面だけみてもかなり変化したと考えられる。すなわち、過去には調査データは利用不十分なまま報告書作成後は倉庫にしまわれるか破棄され、“もったいない”ということが先入観としてあったように見える。

現在では調査データは電算入力化され、磁気テープからフロッピーディスクに保存され棄てられる心配はあまりないだろうが、調査データの有効利用あるいは活用という点では、従前とあまり変わらないといえる。さらに調査データにとり込まれていない調査企画、調査実施状況、整理点検資料等、調査データの集計、分析に直接必要な資料を調査データとセットにして保存するということは、あまり考えられていない。しかし、調査票、コードブック、報告書等はワープロ入力されていることが多いので、技術面からみると、〈データベース〉化の基礎は徐々にできつつあるといえる。

すなわち、世論調査のデータベースとして考えられるものは、調査手順として次々に作成準備されてきた〈各種の資料〉を表1に示したように、『文書データベース』あるいは『数値データベース』として作成し、これらのデータベースを、いかに、うまく結びつけて利用しやすいものにするかということであろう。過去の調査については、調査データは割合簡単に『数値データベース』になるが、その他の資料を『文書データベース』にしてとり込むのは容易ではない。しかし最近の調査あるいは今後の調査では、大部分の資料はワープロ

入力化されているので、調査データベースとして、これらの資料の『文書データベース』と調査データの『数値データベース』を作成するのはそれ程困難ではない。であろうから世論調査のデータライブラリーの基礎はできているともいえる。

最近のNHKの『放送研究と調査』(1994年11月号)には“動き出した「世論調査データベース」”と題して、NHK放送文化研究所におけるデータベース化とその運用についての紹介が載っている。NHKの調査についての文研データベースの特徴がいろいろ紹介されている。データライブラリーが現実のものになってきたようにも考えられる。

ところで、データライブラリーの必要性に関連して調査データが有効利用されずにいて“もったいない”という考え方と並んでよくいわれることは、「研究者は研究費が少ないので小規模な調査しかできない、それで全国調査など大規模な調査の結果を利用できるように考えてもらいたい」ということであった。このような考えも理解できるが、後にのべるような理由から、このようにして科学研究費等で小規模に独自の調査を計画するよりは、アメリカの社会科学研究者達が“National Data Program for the Social Sciences”として1972年から毎年実施している「一般社会調査 (General Social Surveys)」あるいは、このドイツ版であるALLBUS (ドイツの一般社会調査 1980年より隔年実施) あるいは1985年より計画実施されている国際比較調査計画であるISSP (International Social Survey Programme) の調査等の研究のあり方が参考になるように見える。

〈データライブラリー設立準備委員会について〉

さて、世論調査データライブラリーについてのその後の経過は相変わらずであったが、1989年にこれまでの経緯をふまえてデータライブラリーの設立に向けて新しい活動を始め小委員会から、データライブラリー設立準備委員会と名称も改められた。1992年3月「世論調査データ機構（仮称）設立に関する基本構想」をまとめデータライブラリー設立をひろく一般社会に働きかけることになった。

データライブラリー設立準備委員会は1993年から、一般参加のシンポジュームを2回、研究会を3回およびドイツのセントラルアーカイブ（ZA）前所長ショイヒ教授の講演会を開いて設立の基礎をかためてきた。

第1回シンポジュームでは、世界における主要データライブラリーの現状紹介、データライブラリー利用のメリット、データライブラリーの設立における問題点が議論され、第2回は、継続調査データの事例として次の3調査を

統計数理研究所の「日本人の国民性調査」
NHKの「日本人の意識調査」
総理府内閣広報室の
「国民生活に関する世論調査」

とり上げ、それぞれの調査結果だけでなく、継続調査から得られる時系列データの分析結果を示し継続調査データを利用する利点について紹介した。

これはデータライブラリーが出来たらこのような分析が可能になるということを示す意味があった。

研究会は、まずドイツのセントラルアーカイブ（ZA）の内容紹介、第2回は学術情報センターの内容紹介と、数値データベースを取り扱うとしたときのシステム構想およびイン

ターネットによる諸外国データライブラリーへのアクセスについて紹介された。第3回はこれを受けて、インターネットのGopherシステムによる、アメリカ（ICPSR）、ドイツ（ZA）等のデータライブラリーへのアクセスの紹介とパソコンによるインターネット接続実験をおこなった。これによりインターネットによるデータライブラリーの（開設の）可能性について考えることができる。

1994年の9月にはドイツの前ZA所長ショイヒ教授の講演会をおこなった。ショイヒ教授はデータライブラリーの特徴として、アメリカのローパーセンターはデータの保管に重点がありドイツのZAはコンピュータ検索可能な質問項目ごと、調査研究ごとの2つのデータベースをもち、検索利用の便がよいこと。データの利用には専門家ではない一般の研究者向け利用レベル（詳細なコードブック、報告書付き）研究者レベル、将来の利用可能性を考えて保管するレベルの3段階のこと。またデータ保護とデータライブラリーの関連では2つの考え方があり、①質問ごとにその回答を秘密にするか否かを弁別する方式と②データの散逸（データの勝手な利用）を規制する方式のうち、①はフランス式で実際の調査に適用するときの問題がある、②はドイツ式で社会科学的な立場から発案され、フランス式の問題点を解決したもの。

実際に必要なことはデータが勝手に使われるのを規制することで、匿名性が保たれるときやデータ収集者が利用するときは何も規制しない。調査は回答全体から得られる傾向を分析することが主な目的で個々の回答者には関心がない。したがって匿名性に関する問題点は、回答者名簿のファイルと回答データのファイルを分離し、別に保管し、若しその2つのファイルを結合する必要があるときは第

三者の部門がその対応のみを教えるシステムをとること（パネル調査、追跡調査の場合）、また、属性別の分析は ALLBUS（ドイツの一般社会調査）では職業を ILO（世界労働機構）の職業分類で、居住地は調査地点の街区番地等で細かく分類しているので、両方のデータを使いクロス集計すれば回答者を特定できる可能性がある。一般の人にデータを開放するときには、職業・居住地のどちらかの分類を粗くして提供している。ZUMA（社会調査分析研究センター）の研究では原則として2,000人程度の調査では回答者に関する各属性カテゴリに50サンプル以上、クロス集計したとき各マスに5サンプル以上入るように考えた分類をとっている。

また、データライブラリーの社会的貢献としてこれまでいわれていることのほか、データライブラリーが国際比較調査プログラム (ISSP 等) の受け皿の役割をすること、国際比較調査の共通データファイルの作成で属性分類の対応が国によって異なること^{*)}、たとえば教育レベル、職業の区分も同じ言葉で表されたカテゴリが国によって意味が違うこと、データライブラリーはこのような知識を蓄え普及させるのに適した組織である。

現在では多くの機関で調査データを収集している。調査データは一部の研究者のものではない。収集したデータを他と共有し、多くの人が利用してデータの価値が高まる。多くの人がデータを利用するることは社会調査研究が一般の支持を得ることにつながる。

さらにデータライブラリーに適した組織として、大学などの大規模な組織に付属した組

^{*)} 「意識の国際比較方法論」の研究でも同

様の結論が得られている。たとえば鈴木達三 (1992) 「国民性比較の方法論」、
統計数理、40巻2号 243-53頁

織ではなく、独立機関とした方がうまく機能する。いろいろの国の経験からデータライブラリーはいくつかの組織のネットワークの要として存在すると一番うまく機能する等、ZAを中心とした活動の経験をふまえた講演であった。

これらのデータライブラリー設立準備委員会のこれまでの活動状況と、基本構想の概要については、近い将来印刷される予定の「データライブラリー研究会 1994年度報告」に詳しいのでそれを参考にして下さい。

〈データライブラリーの利用について〉

これらの活動を通して明らかになった利点、問題点は以下のようである。

まず、利用についてのイメージはデータの再分析ができる。とくに時系列分析あるいは同種の調査の比較分析等に興味があり、個票データが利用できれば、継続調査の統一的仕様による再分析や質問相互の関連分析等これまで個々になされていたことを統一的にできる。また、たとえば、各自治体で繰り返し実施されている県民意識調査や市政に関する調査などは、地域による比較の視点、あるいは時点による比較の視点をとり入れることによって調査結果より得られる情報がより拡がり、よくなる等であり、また、多くの人の利用のイメージとして、先にもふれたように、調査データは表面的な分析のみで報告書が作成され深い分析もなされないまま放棄されていると考えられ、したがって調査された個票データを集積して再分析できる状態にすれば、よりよい分析・利用ができるのではないかと考えがちである。ここでいくつかの問題がある。

その1つは、誰がそれを（すなわち、全然調査の実情を知らないで、分析だけしたい人の分析の準備を）するのかという点で意見の一

致をみない。また、一方では『集計するまでにいろいろ準備をするのだから、その資料を活用できれば誰にも利用できると考えられる』という意見がある。しかし、それまでいろいろ資料を準備して調査を実施し、結果を出すまで苦労した当事者ることはどう考えるかという問題、つまり調査の著作権の問題があるだろう。

また、調査実施や回答者について考えると、直接調査結果の統計数字を出すための利用以外の利用については承諾しているかどうかが問題になるかも知れない。いわゆるプライバシーの保護に関連して、データライブラリーがどのような対策を考えているかも議論のあるところである（プライバシーの問題については、1994年9月22日のZA前所長ショイヒ教授の講演要旨参照）。

〈調査データの利用のあり方〉

（仮に世論調査データライブラリーに調査データがあるとして）既存の調査データを利用するにはそれなりの利用の仕方があると考える。たとえば、これまでの調査結果をふまえて次の調査を企画する時には、それ以前の調査データの再分析は実施した方がよいと考える。既存の調査データの再分析もしないで次の調査を独自の視点から計画するのはいささか無謀ともいえる。

すなわちデータライブラリーの存在意義として、上にのべた利用のイメージより重要なことは、新しい調査を企画するときに、データライブラリーを利用して、調査にとり上げる項目について、あるいは標本設計について既存調査データの再分析をおこない、その情報を利用することではないかと考える。

一般に調査企画に当っては、

標本企画：層別多段抽出（確率抽出法）

質問項目の選定：質問↔回答形式の選定が基礎になるが、標本抽出の方は統計的理論によって指針がきまり、実際にどのようにすればよいかその具体的手続きや手順についてきちんと標準化されており、これを基にして実際の調査のときの手順が進められている。標本企画についての留意点も過去の調査データの再分析によって、明らかとなり、さらによい標本企画となる可能性が考えられる（たとえば、内閣総理大臣官房広報室研究報告「世論調査の標本の精度に関する研究」昭和57年3月（1982）、鈴木・高橋宏一（1991）『標本抽出の計画と方法』放送大学教育振興会）。

しかし、質問項目の方はどうかといえば、理論的、経験的な裏付けがあまり豊かとはいえないようにみえる。

ある一つの質問項目をとり上げて考えてみると、大略の質問内容がかたまったところで、具体的な質問の言葉使いはどうすればよいのか？回答選択肢はどのようにすればよいのか？

等について、いろいろの考え方はあるが、
①分析の都合で回答選択肢をきめる場合（たとえば因子分析等をおこなうために5段階のスケールになる形で回答をとる。この際回答の安定性などはあまり考えていない。）
②はっきりした回答がよいと考えて、賛成、反対の2項選択形式でとる。

③質問内容からみて、いくつかの選択肢が考えられるとき、これを集計の便を考えて多项選択型の回答形式でとる場合、あるいは
④自由な発想を重視して自由回答形式でとる等種々の場合が考えられ、どれがよいのか等のことは実はあまりはっきりしたことは出ていない。大きく過去の経験に依存しているのが実情である（これらのこととは狭い範囲の調査対象（大学生など）に対する1回限りの実験調査の結果から推量する等のことは得策と

はいえない)。

このとき、過去の調査資料が利用可能ならば、いろいろの作業仮説のうちからよりよいものを選択することが可能となるかも知れない*)。

これまで過去の資料として通常は自分達が手がけた調査資料の再分析になる。特別な研究調査以外には他の調査システムによる調査データの再分析は不可能であるが、データライブラリーが利用できれば、これまでの調査資料が、調査企画から個票データまでそろって利用可能となり、過去の事例を再吟味してよりよい調査を企画することが可能となろう。とくに、調査システムによる偏りを是正するのに役立つと考えられる。

経験科学においては、これまでに得られた資料は、何事によらず貴重なものである。それが多くの誤差その他をもち、欠点を大なり小なり含むとしても、それを越えて社会の実態の一部を記録している情報的価値は大きく存在するといえる。したがって、そのような情報をとり入れることなく新しい調査を企画することは余り得策とはいえない。

これらのことを考え合わせると、過去に調査経験が豊富で自己調査資料の再分析の容易に実施できるところでは調査企画に当ってそれなりのノウハウの蓄積があるのでよりよい

*) 内閣総理大臣官房広報室研究報告：

「世論調査の質問の仕方と精度に関する

研究」(1983)

内閣総理大臣官房広報室研究報告：

「世論調査の質問形式及び質問内容に関する研究」(1984)

林知己夫・鈴木達三・赤池弘次 (1986)

「統計学特論」放送大学教育振興会

P103 - 118 参照

調査が計画される可能性が高い。一方独自の調査はオリジナリティーがあるかのようにみえるが、いつも実験調査の域を出ず調査結果は必ずしも期待通りにはならないことが多いのではないかと考えられる。

同一調査システムで長期間継続した調査で、他の調査との多面的な比較を経てきた調査データが一応信頼できるのではないかという立場からすれば、上にのべたように過去の調査データを利用しながら調査企画を立てることが現時点では最善の方策であると考えるので、このような観点からデータライブラリーの利用システムを考えるというのが、前記の問題点の解決策の一つではないかと考える。

この方式は独自性を強調し利用に目が向きがちな研究者向きではない。しかし、社会の実態をより一層正しくとらえようという立場の研究者や、社会調査専門家には不可欠なシステムといえる。

社会全般の動きや個々人の意見の動向は一個人の発想とか個人の論理で計画された、ただ1回の独自の調査ではばかり知ることが多分できないだろうが、皆が智恵を出し合って計画し実施された調査資料、調査データの総体をとり込んで計画される調査は、非常に進み方はおそいかも知れないが、一步一步未知の領域を明るみに出すように、まぎれのない社会の実態を浮かび上がらせる利点があるだろう。

データライブラリーのこれ以外の利用としては、全く新しい分析方法が開発され、過去のデータから現在までのデータを同一方法で分析し、その分析方法の可否を歴史的に確認する等のことが考えられる。これは研究の過程で過去のデータの利用があれば、その出所、その他が通常の引用文献と同様の扱いを受けるものと考えられ、調査担当者、研究者の著

作権はそれなりに保護されると考えられる。

また、過去のデータのみで従前からある分析方法による再分析のみの論文は、それなりの評価しか受けないだろう。ここに新しい知見があるとすればそれは結構なことではないかといえる。

〈まとめにかえて〉

いずれにしてもデータを利用する、利用させるといつても利用する側に、「データのもつ情報的価値を正しく判断できること」が必要であり、「データのとられたもとのところ、データをとる時の諸状況等を考慮できること」が必要である。

第1回シンポジュームで、林知己夫日本世論調査協会会長が総括したように、データライブラリーはデータのもつ情報的価値をよく理解している人が利用する形になればよいわけで、データライブラリーとしてはデータの利用と共にデータを扱うためのデータの科学、データ科学の考え方を十分身につけてもらうようになる。ものをみると実証的な立場をとり、データの中からどういうふうにして妥当な情報を探っていくか、データの中から何をとり出していくか、こういうことをしっかりと身につけてもらう。これらを実行する技術、どのようにしてデータを作るか、どのようにデータを収集するか、それからデータの解析の3本柱が重要である。このような教育訓練の場をそなえることもデータライブラリーとして大事なことである。

データライブラリーの機能として

◎調査法に関するノウハウの発見と蓄積により次の調査に役立つ作業仮説の提供、あわせて、調査企画部門、調査実施部門に対してコメントや作業仮説の提供ができる。

◎調査実施について、例えば質問項目ごとの

回答状況のチェックにより調査状況等の実態に関するノウハウが得られる。

◎これらの知識を生かした調査の実施が可能になる。

このようなことがつけ加えられる。

これらは、どのようにすればよりよい質問が作れるかということにつながり、情報はどうのようになるとくるかということに関連することである。

これらのこととは、研究者・専門家のみならず調査機関の担当者や調査依頼者（委託者）の大きな関心事であると考える。はじめに述べたように、調査票、コードブック等の「文書データベース」と、調査結果的回答をコード化した「数値データベース」により、データライブラリーの基礎はできているといえる。

しかし、データライブラリーが国内だけでなく国際的にも活動範囲を拡げるときには、日本語による「文書データベース」と同時に、たとえば英語による「文書データベース」が必要になり、質問文やコードブックも日本語版と英語版の両方が必要になる^{*)}。このとき質問文の英語訳、キーワードの英語訳あるいは逆に外国のものならばその日本語訳が「定訳」となっていないと、データベースとして将来問題を生じる可能性がある。

この点から考えると、データライブラリーが国際比較調査計画の受け皿の役割を引き受けることが重要な意味をもつ。ショイヒ教授が指摘するように、国際比較調査のノウハウの蓄積がなくては、日本語以外の言語の「文

^{*)} ドイツのセントラルアーカイブでは日本から送られたSSM調査（日本社会学会が1955年、1965年、1975年、1985年の10年ごとに実施している社会的成層とその移動の調査）の調査資料が手つかずのままになっているという。

「書データベース」にとり組むのは容易なことではない。

ここまでみてくると、インターネットと接続できる環境で「Gopher」システムの端末を立ち上げてメニューを開き、そこにデータライブラリーの看板を出せばデータライブラリーが稼働したことになるともいえるが、そのための準備には相当の覚悟が必要ということになり、やはりデータライブラリーの設立は国

家的事業ではないかと考える次第である。

注) データライブラリーの考え方は詳しく
はデータライブラリー設立準備委員会
の「世論調査データ機構（仮称）設立
に関する基本構想」平成4年3月（1992）、
その後の経過についてはデータライブ
ラリー研究会の「研究会報告」（1995）
印刷予定を参照して下さい。

