

統計研究参考資料

No. 54

アメリカにおけるマイクロデータの提供

1998年1月

法政大学 日本統計研究所
The Japan Statistics Research Institute
Hosei University

はじめに

これまでわが国では、統計は集計結果表の形で主として提供されてきた。これにはいくつかの事情がある。第1に、戦後、統計制度の再建当時、集計機械の保有は総理府統計局など一部の政府統計機関に限られ、予め計画された集計項目についてのみ集計が行われ、結果表が公表提供されてきた。第2に、統計の作成過程で被調査者(調査客体)からあつめられた情報(統計原情報)には個人や組織の秘密事項に関係することも多く、秘密保護の観点からも結果表での提供が合理的な提供方式とされた。

また、提供されている集計表では、集計分類の制約等から研究や行政面に利用できないこともありうる。このような場合については、特に指定統計に関しては、調査個票の目的外申請制度により、特別な承認を得た者について極めて限定的な形で調査個票の使用が認められてきた。わが国の現行統計制度の根本法規である統計法もこのような統計の作成、提供システムを前提したものとなっている。

ところで、欧米では早いところではすでに1960年代から、匿名化措置を施した個別レコード(匿名標本データ)が提供され、学術面などで積極的に利用されてきた。またこの他にも、北欧諸国を中心に、登録などの異種の業務記録をリンケージすることにより統計を編成する試みも行われている。このような調査に基づかない形での新たな統計の作成は、欧州を中心に近年、急速な広がりを見せている。

一方で、個別レコードからの利用者独自の統計解析利用(独自集計、多変量解析法の適用など)、他方で行政における登録(レジスター)に基づく統計の作成は、いずれも今日の情報処理技術の発展、普及をもって初めて可能となったものである。この点で、わが国の統計制度は、このような技術的展開に対してどのようにシステムの軌道修正を行うべきかについて、いくつかの検討課題をかかえている。

その一つが、非集計データの提供システムの構築である。

これについては、1995年3月に出された統計審議会答申『統計行政の新中・長期構想』でも取り上げられている。そこでは、第4章「調査結果の利用の拡大」の中で、標本データ(個票データから必要に応じて抽出を行い、地域区分や世帯番号等の個体の識別子を消去するなど個体の識別を不可能にしたもの)について、個体の秘密保護を担保しつつ、外国の制度及び提供例、国内外におけるニーズの実態、現行法制度との関係、具体的な提供方策等について専門的・技術的研究を行う必要があるとの提言がまとめられている(『答申』74~75頁参照)。

本書で紹介する2つの論文のうち、まずCox論文は、アメリカのセンサス局その他政府機関からどのような種類のマイクロデータがすでに提供されているか、センサス局でのマイクロデータの提供に当たっての法的枠組み、一般利用ファイルの公開に伴う様々な秘密漏洩防止のための措置、公開の可否を判定する機構としてのマイクロデータ審査パネルの構成ならびにその機能、などについて論じたものである。

なおこの論文には、付録として審査パネルへのマイクロデータ公開請求のための申請書作成マニュアルが載せられている。

一方、Alwin の論文は、アメリカにおける政府機関以外の作成主体による代表的な調査を紹介し、それらの調査から編成された匿名マイクロデータの社会科学研究目的への提供の可能性を論じたものである。

もちろんマイクロデータの場合、いかに匿名化措置が施されていても、集計表に比べれば個体が識別される危険性が高いのはいうまでもない。特に企業や組織に関する個人情報の場合にはそうである。解析利用面での利便性と個人や組織の秘密保護をどうバランスをとるかという点では、多くの対処すべき問題が残されている。また欧米とわが国とでは統計あるいは調査をめぐる風土（あるいは国民性のようなもの）にも相違があるように思われる。それだけに今後どのように日本型のシステムを構築していくかについては、それが調査実施の根底に関わる契機を含んでいるだけに、多面的な検討が必要であるように思われる。その点で、マイクロデータ提供における先進諸国における様々な経験、制度面での工夫については、その適用可能性を含めて色々な角度から検討してみる必要がある。

なお、本書資料の訳出は、坂田幸繁（中央大学）が担当した。

この資料が、今後のシステム構築に対して何らかの示唆を与えることができれば幸いである。

1998年1月

法政大学

日本統計研究所

本資料は、平成9年度文部省科学研究費補助金重点領域研究「統計情報のフロンティアの拡大ーマイクロデータによる社会構造解析ー」の計画研究班A02（ウ）「欧米における調査個票の多角的結合的利用の研究」の成果の一部である。

アメリカにおけるマイクロデータの提供

目 次

匿名マイクロデータの開示に関するセンサス局の実務 …………… (1)
Lawrence H. Cox(アメリカ商務省センサス局)

匿名一般利用サンプルの可能性と展望：社会科学における国民的データ資源 … (22)
Duane F. Alwin(ミシガン大学社会調査研究所)

坂田幸繁 訳

匿名マイクロデータの開示に関するセンサス局の実務

Lawrence H. Cox
商務省センサス局

合衆国統計システム

合衆国の統計システムは分散型である。統計調査を企画し、データを収集し、そして統計的要約やマイクロデータを公開する責任は複数の連邦機関の間で分担される。そこには、労働統計局（BLS）、エネルギー情報部、社会保障総局、内国歳入庁、国立衛生統計センター（NCHS）、国立教育統計センター、統計報告部（農務省）、司法統計局、およびセンサス局が含まれる。センサス局は、それが最大であるという意味で国の基本的な統計機関であり、しばしば他の機関に代ってデータを収集し、それらが保有する一定のデータへ一方的にアクセスできる。

分散的な統計システムにおいて、マイクロデータの配布という概念は注意深く定義されねばならない。マイクロデータは統計諸機関の間で、回答者の負担の軽減、収集が困難なデータの質の改善、および独立した出所から収集された情報の確証などの目的のために共有されるだろう。その場合には、識別可能なデータか、少なくとも匿名化するほど制限の無い十分細部が豊かなデータを共有することが必要であろう。

データ共有は合衆国統計システムの大きな関心事であるが、それはこの論文の範囲を超えている。要点としては、データ共有に関する一方通行型のモデルが合衆国では利用されているというだけで今は十分である。センサス局は他の諸機関が保有する一定のマイクロデータへアクセスできるが、その逆は決してありえない—識別可能なセンサスデータは法律によって外部者（法廷でさえ）に公開しないようにしっかりと保護されている。ここでセンサス局のマイクロデータの公開という場合、一般利用ファイルとしてのマイクロデータ・ファイルの直接的公開か、もしくは第三者への選択的な公開を指している。一般的に第三者とは、マイクロデータ・ファイルがそこから作成される調査の後援者であり、彼らはふつうこれらのデータをさらに配布しようとする。法律が規定するような少数の例外はあるが、公開されるマイクロデータ・ファイルは常に匿名化されている。

Lawrence H. Cox, "The Practice of the Bureau of the Census with the Disclosure of Anonymized Microdata"; in *Nutzung von anonymisierten Einzelangaben aus Daten der amtlichen Statistik : Bedingungen und Möglichkeiten*, Germany (West). Statistisches Bundesamt, W.Kohlhammer, 1987.

他の（センサス局以外の）合衆国統計機関によって公開されるマイクロデータ

本節では、センサス局以外で公開される重要なマイクロデータ・ファイルとその漏洩リスクを軽減するために利用される手段について簡単に述べよう。ここに含まれる資料は、合衆国商務省（1978）とこれらの諸機関の代表者との非公式の議論から得られたものである。Mugge(1984)は NCHS の事情について詳細な説明を与えている。

社会保障総局は、社会保障番号を所持するすべての人々をカバーする継続職歴サンプル（Continuous Work History Sample）を開発した。このファイルから、縦断面雇用－被雇用者データ（LEED: Longitudinal Employee-Employer Data）ファイルと呼ばれる1%抽出サンプルが一般利用のために作成されている。LEED ファイルに含まれる変数には、個人に関する年齢、人種、性別、および収入データに加えて、1957 年まで遡及したその個人の雇用主に関するデータなどがある。1962 年のその最初の公開から、利用目的を明示し、許可を受けないファイルの共有や個人を識別するためのマッチング研究へのファイル利用を禁止する制限付き使用契約のもとで一般に公開された。LEED ファイルは内国歳入庁から得た個人に関する収入データを含んでいるので、ファイルの一般公開は 1976 年租税改革法の通過の帰結として中止された。その改革法の中で税務申告書から得られたデータに関する厳密なデータ公開規準が規定された。

社会保障総局はまた、使用上の制限を置かずに、退職歴（Retirement History Survey）などのマイクロデータ・ファイルを一般利用のために公開している。これらのファイルは小サンプルであり（1%未満）、限られた地理的情報しか含まず、公開前に異常な変数や変数の組合せについて秘匿処置を受ける。

内国歳入庁は、使用制限のない全国推定のための税収モデル（Tax Revenue Model for National Estimates）を一般利用のために公開している。抽出比率は一定の部分集団に対しては高目であるが、このファイルはすべての税務申告書の 0.2%未満の標本からなる。州別推定のための税務モデル（Tax Model for State Estimates）は州レベルで識別されるすべての申告書のおよそ 0.3%の標本を含んでおり、州の税務当局に利用可能である。2つのファイルは個人の税務申告書からのサンプルであり、個人の識別子を除去し、各申告書からの 150 のデータ項目を含んでいる。

国立教育統計センターは、全国の 1972 年高卒時のクラスの 0.7%を標本とする一般利用ミ

クロデータを制限無しに公開している。識別情報は除去され、地理的には地区別、および都市と農村の区別しか含まれていないので、このファイルには、家族的な背景、意識、および将来計画に関して学生が提供したデータばかりでなく、学業成績評点の平均（grade point average）、クラス順位、全国試験結果や勉学分野などの情報も含まれる。回答者によって与えられる追加情報は定期的にファイルに追加される。

NCHS は、衛生面接調査（Health interview Survey）、衛生・栄養試験調査（Health and Nutrition Examination Surveys）、全国外来医療事例調査（National Ambulatory Medical Case Survey）、および退院調査（Hospital Discharge Survey）などの調査と統計プログラムの多くから、一般利用マイクロデータファイルを公開している。とりわけ独特のファイルー出生に関するファイルーが一つあり、それは出生記録の 50%標本(州によっては 100%)を包含している。すべてのファイルについて利用者は、そのファイルを統計的な研究や報告目的のためにしか使用しないという宣誓書の署名を要求される。NCHS はすべての直接識別できる情報を除去し、地理的細分を 100000 人以上の人口地域に限定し、希少な特性を表すカテゴリーを削除もしくは併合したりする。NCHS は、機関の外部で入手できる適切なマッピングファイルがないために、これらの漏洩回避手段で十分であると確信している。

BLS は、ときには利用を制限した上で、限定配布のための少数の特別注文のマイクロデータ・ファイルを公開している。BLS は、消費者支出調査（Consumer Expenditures Survey）からの一般利用マイクロデータ・ファイルのための資金提供者であり、技術的にはその公開機関であるけれども、これらのデータはセンサス局によって収集され、加工され、そして漏洩回避策を施される。BLS はそれ自身の継続的な一般利用マイクロデータの公開プログラムを決して有しないが、それは、BLS が州の機関から厳密な条件付き使用協定のもとでデータの多くを得ているためである。

センサス局が公開するマイクロデータ

センサス局は、国の第 1 義的なデータ収集者としての役割には好都合であるが、対応して回答者の秘匿性を守る責任という点では厳しい法的なフレームワークの中で活動している。ここでその法律の関連部分を提示しておくことも重要であろう。

合衆国法典（1976 年 12 月 31 日修正）第 13 編から：

第 8 条(b)・・・連邦政府の省庁、機関、および事業所、コロンビア特別区行政府、本編中の 191(a)項において言及される任意の領地あるいは地域の行政府（その中の政治的下位部局を含む）、州あるいは地方機

関、あるいは他の公的および民間の人間と諸機関に対して、長官は、任意の特定の回答者、もしくはそれに代わって報告される情報を開示しない諸表や他の統計資料の写しを提供することができ、そして特別な統計的編集や調査を行うことができる……。

……(c)本条項の下で提供される情報はどんな場合においても、この権利の申し立てによる侵害に関する訴追の場合を除いて、任意の回答者に、もしくはそのような情報が関係する他の人間に不利益を与えるように使用されてはならない。

第9条。 秘匿される情報；例外

(a) 商務省長官、もしくは商務省あるいはその中の部局や諸機関のすべての職員や被雇用者は、本編第8条に規定されるものを除いて、以下のことをしてはならない。

(1) 情報がそのために供給される統計目的以外の任意の目的に対して、本編の規定下にある情報を提供すること；あるいは、

(2) 本編の下で任意の特定の事業所、もしくは個人によって提供されたデータが識別されるような公表を行うこと；あるいは、

(3) 商務省あるいはその中の部局や諸機関の宣誓職員と被雇用者以外の者に、個人記録を調査させること。

……

(b) 本条項のうち特定の個人と事業所に関するデータの秘匿扱いに効力を及ぼす下位項(a)の諸規定は、政府の諸センサスには適用されない……。

それゆえセンサス局は、公開された各マイクロデータファイルが回答者に不利益を与えることのない有益な統計的目的に役立つこと、そしてファイルは個々の回答者に属すデータを識別するためには使用できないことを保証しなければならない。法律言語は変わらないのに、“不利益を与えること”および“識別する”というその操作的用語は統計的には厳密でなく、再識別の新しい方法、そのデータで達成される分析の範囲、および再識別に使われる代替的なデータ資料の存在などのような諸要因を考慮して、新しいデータ公開毎に再評価されねばならない。われわれは以下にマイクロデータの公開と漏洩回避に関するセンサス局の状況を要約する。Cox, et. al(1985)はセンサス局に影響を与える秘匿性問題全体の概観を与えた。

1960年の10年センサスを始めとして、センサス局は人口と住宅に関する10年センサス調査からの一般利用マイクロデータ・ファイル（いわゆる“一般利用サンプル”）を公開した。漏洩回避の要求とデータ公開に関する他の統計基準を満たすことを条件に、これらのファイルは10年センサス調査から入手できるすべての特性値を含んでいる；各個人の年齢、人種、学歴、職業、所得など、加えて世帯に関する住宅特性。1963年に公開された最初のサンプルは、地理的情報（センサス地域区分と地域コードのタイプ）をほとんど含まず、センサスの詳細調査票を受け取る合衆国の全世帯の1/4抽出標本を元に、合衆国の世帯全

体の 0.1%標本しか代表していない。0.01%部分サンプルもまた一般に公開された。これらのファイルの購入者は、無許可のファイル共有を禁止し、そのファイルから得られたデータを含む公表物すべての写しをセンサス局が受け取ることを要求する同意書に署名した。明らかにこれらの同意書が守られなかったことは明白である：65 のコピーファイルが販売されたが、センサス局は 200 以上の機関でそのコピーの存在を確認することができた。

1970 年センサスの一般利用サンプルは、1/5 センサス詳細調査票サンプルから作成した全世界帯の 1%抽出サンプルを標本とし、少なくとも 250000 人（このときのセンサス・マイクロデータ公開のための人口区分閾値）を含む人口地域まで細分した地理的情報を与えた。6 個の相互に排他的な 1%一般利用サンプルが 1970 年センサスから公開された。それは 2 種類の質問票から引出されたサンプルを標本とし、3 つの異なる地理的分類（大都市／大都市以外、都市／農村、および郡別グループ）別のデータを提供した。それらの利用に制限は置かれなかった。これらのファイルは利用者に極めて人気があり、数千のコピーが直接センサス局によって販売された。

1980 年一般利用サンプルもまた、センサス詳細調査の質問票の全体として 1/5 サンプル（大部分の地域では 1/6）から作成された合衆国の全世界帯の 1%サンプルに基づいた。1970 年と同様に、3 つの別個の一般利用ファイルが公開され、各ファイルは異なる地理的分類に従っている。

センサス局は毎年一般利用のために、経常人口調査(CPS)の 3 月次追加調査から作成された年次人口ファイル (ADF: Annual Demographic File)を公開している。内容と構成において一般利用サンプルと同様であるが、これらのファイルは 1968 年まで遡及した時系列データを与える。CPS の標本抽出設計は標準地理分類と重なっているため、ADF における人口閾値は注意深く監視されねばならない：人口閾値以下の任意の地域を識別するデータは表示できない。したがって例えば、調整した中間サンプル・コードや都市／農村あるいは市街／大都会などの地理区分は、人口閾値にぶつからないことを保証する州法に則して秘匿されるだろう。

センサス局の他のほとんどすべての世帯調査についてもマイクロデータ・ファイルが公開されている。これらのファイルには通常、サンプル全体が含まれる。大雑把な数字では、これらのサーベイの平均標本抽出率は 1/1500 である。センサス局のセンサスもしくはサーベイ調査の一般利用マイクロデータファイルに関しては、利用上の制限は置かれていない。

1983 年にセンサス局ではある新しい貴重な調査、所得・プログラム参加に関する調査 (SIPP: Survey of Income and Program Participation)を開始した。これは個人、家族、および

世帯に関する縦断面調査であり、所得、投資、および他の会計に関する詳細なデータを収集する。最初の 2 回の調査からの横断面データを含む一般利用マイクロデータファイルが公開され、第 3、4 次に続き、第 5 次調査からのファイル公開もまもなく予定されている。第 4 次調査で収集された時事的モジュール（資産と負債）から作成されたマイクロデータファイルもすでに公開された。第 5 次ファイルは個人別に（横断面的に）加工されるけれども、それらは、一組として処理され、個人に関する 15-20 ヶ月間のデータを包含する縦断面マイクロデータを表示する。SIPP をもとに真に縦断面的なマイクロデータファイルを作成する作業が進行中である。この調査は一つの新しい挑戦—詳細な縦断面調査における漏洩回避—を意味する。これらの含意は 1985 年 11 月の社会科学研究会議の後援によるセミナーで議論された（報告書近刊）。

産業や製造業企業の高度に歪んだ分布のために、センサス局は経済センサスや調査からのマイクロデータを公開しない。しかしながら、センサス局はマイクロデータ・ファイル、縦断面事業所データファイル（LED：Longitudinal Establishment Data file）を作成した。それは、1960 年代まで遡及して製造業センサスと年次製造業サーベイ調査（Annual Survey of Manufactures）データから作成された。このファイルはかなり経済学者の関心をひき、そのアクセスと使用上の方針は現在開発中である。検討中の 1 つの選択肢は、LED ファイルから類似した事業所の小さな部分集合全体の記録を集計することによって、みせかけのマイクロ集計値からなるマイクロデータ・ファイルを構成することである。このファイルは、LED の構成、および分析の機会と特性を大規模な利用者に教えるために使われるであろう。実際の LED ファイルに基づき、より現実的な分析を求める人々は、センサス局が遂行するコスト弁済可能な作業にたいする具体的な提案をセンサス局に行なうことができるであろう。

センサス局は、他の政府諸機関や組織に代って多数の統計データの収集と加工を遂行する。そのような調査は、請負い調査の後援者のためにコスト弁済可能なベースで行われ、通常法典第 13 編の下で遂行される。論理的帰結として、これらのファイルはセンサス一般利用ファイルと同じ漏洩回避策と処理を受ける。このことは、調査データの収集に金を払うが、検閲もしくは要約された形態でしかそれを受け取る資格がない調査の後援者にとっては非常な驚きである。

センサス局は、マイクロデータファイルが公開されるべきか否か、そしてどんなフォーマットで公開すべきかを決定する確固とした審査過程をもっている。次節で論じるが、これらの審査は公開を予定されるすべてのマイクロデータファイルに一様に適用される。1985 年中に公開のために審査されるマイクロデータファイルのリストは以下のようである。

公開承認済みのもの：

修正経常人口調査

1984 年科学者およびエンジニアに関する調査

1980-81 年消費者支出—四半期面接調査

1982-83 年消費者支出調査日次記録(日報)

1982-83 年消費者支出—四半期面接調査第 3 次 1984 年所得とプログラム参加に関する調査 (SIPP)のための時事的モジュール

1982 年トラック在庫と利用調査

プエルトリコ一般利用マイクロデータサンプル

SIPP のための第 4 次時事的特殊モジュール

職業訓練縦断面調査終了者制御カード業務データファイルへの追加
追加された州法に関する若者の作業経験に関する調査

青少年拘留、矯正および保護の民間施設に関するセンサス

未決定：

職業訓練縦断面調査面接ファイル修正 1982-83 年日次調査

マイクロデータ審査パネル

センサス局マイクロデータファイルの公開の決定は、マイクロデータ審査パネル (Microdata Review Panel) によって行われる。当パネルは 1981 年にそのために作られ、それは審査が相互に整合的であることを保証する。パネルは、マイクロデータ公開のための基準を開発し、マイクロデータ漏洩回避の原則と最小限の許容規準を満たすマイクロデータ・ファイルの設計に関する指針を後援者とそのセンサス局の対応者に与えるというさらに進んだ目的をもっている。前節最後のリストは 1985 年中のパネルの仕事の負担量を表わしている。

パネルは人口・経済・統計的基準と方法論の理事会 (the Demographic, Economic and Statistical Standards & Methodology directorates), データ利用者サービス部, およびプログラム・政策開発課の上席代表者から構成される。パネルのメンバーは順番に議長となる。

例えば ADF などのマイクロデータの公開継続は、ファイル構造や調査デザイン、もしくは他の重要なファクターに変化がない場合には、パネルによる承認を 1 度だけ受ける。他のファイル、とりわけ SIPP の時事的モジュールから作成される弁済可能な調査やファイルは、その提案時点で審査される。パネルは後援者のセンサス局側代表者を教育し、一緒に作業を進めることに成功を収め、その結果パネルの初期の数年間を除き不幸な“驚き”はほとんどなくなった。審査は、キーとなる識別可能性の脆弱性に関する標準的なチェックリ

ストに答えて後援者が提供する情報、ファイル上の地理表示のレベルに関して要求される情報、およびパネルが要求する特定の表に基づいて行われる。パネルは審査の前後でも、その期間中でも関係者と面会することができる。事実上パネルは、どちらかといえば協定が署名され、調査とその結果物に関するデザインが完成する前の、できるだけ早い段階でのそのような非公式の接触を推奨している。後になって秘匿性維持のために犠牲になるようなデータの収集や公表は、普通こうして避けられる。

パネルが強要する堅固でしっかりした 2 つのルールがある - 100,000 人という人口閾値 (1981 年のパネル創設時の 250,000 人から減少) と上位 0.5% の所得の記号化である。後者は現在 100,000 ドルであり定期的に修正される (1985 年の 75,000 ドルから修正)。その他の判断はすべて、利用可能な情報、すなわちそのファイルと照合できる人口記録の存在と構成、一般的な政策配慮、確固とした前例、および類似ファイルに関する経験に基づいて行われる。審査の基本的な道具はチェックリストであり、それは付録に掲載している。

パネルはマイクロデータの公開を承認したり、否定したりするだけでなく、拒否されたファイルについては、パネルの判断として、もしそれが適用されるならば、公開を許可できるほど十分に漏洩リスクを減少させる漏洩回避の方法を提案する。これらの方法には以下のものが含まれる：

- グルーピング

- ・ 定義可能なカテゴリーの結合 (消滅) を含む集計
- ・ トップコーディング
- ・ 再コード化もしくは範囲によるコード化
- ・ 地理的な細区分の減少

- あるデータ項目をすべて秘匿とする

- データ攪乱

- ・ 誤差付与 (量的変数)
- ・ データ・スワッピング (質的変数)

- データを丸める

パネルには、完全な漏洩リスクの分析 (製表、マッチング、シミュレーション、および再分析のための一般化されたソフトウェア) を行うために自由に使える必要な道具もスタッフもない。したがって、上述した情報と専門的知識に基きながらも、その決定は主観的なものにとどまっている。この状況を救済するために進行中のものは、(1) マイクロデータにおける漏洩リスクを最小化するための広範な、統計的に防御可能なガイドライン集を、(2) センサス・マイクロデータの結果における漏洩リスクについての一貫した統計的な定

義あるいは基準，およびその測定のための方法と道具を基礎として，制定するために必要な知識と専門的技術を開発することである。そのために必要な知識の多くは今までに知られていないものであろう。1985 年後期にこれらの問題を調査し，適当なソフトウェアを開発するための一つのプロジェクトが始まった。この種の作業の多くは，Paass 博士とドイツ国立情報処理研究所(GMD : Gesellschaft für Mathematik und Datenverarbeitung)の彼の同僚たちによって報告されており，この検討においてはわれわれは彼らと密接に作業を進めたいと考えている。

プロジェクトは始まったばかりであり，具体的な発見事実はまだ得られていない。しかしながらその仕事の範囲は明快であり，以下のリストにある研究領域が含まれる予定である。

- 最高のマッチング法
- 特定化のさいに利用できる人口記録の存在と構造
- ミクロデータにおける漏洩リスクに関する統計的定義や規準，およびその測定方法の開発
- 現代的あるいは新しい漏洩回避法
- ファイル分割
- ミクロ集計値とみせかけのミクロデータ
- データファイルの質，完備性，および有効性に関する代替的な漏洩回避法の影響

結論

歴史的にマイクロデータの秘匿性を保持する機関の方針は，合衆国においては反作用的に展開した。諸機関は表形式で与えられるものより内容豊かなデータを求める利用者の需要とニーズに応えた。その初期の時代（1960 年代）にはコンピュータの計算能力とファイル・マッチングの手法は原始的であったので，単純な漏洩回避手法で十分であった。これらの初期の公開手続きから前例がつけられ，マイクロデータを求める利用者の渴望は増大した。この状況は 1970 年代へと続き，公開されているマイクロデータの量は，より包括的な漏洩回避政策と方法に対して理解されていた，その時の現実のニーズをはるかに超えていた。1980 年代に経験したコンピュータ処理能力の劇的な増大によって，ファイル・マッチングの方法と，したがって特定化の方法はより強力なものとなった。また部分集団に関する人口記録も対応して増大し，それらに含まれるデータ細目の正確性と水準も向上した。必然的に社会内部におけるプライバシー問題へと焦点が集まった。この状況は分散的な合衆国統計システムの内部では錯綜している。なぜならば脅威となるこれらの人口記録の所有者であるのはふつう他の政府機関であり，そのいくつかは統計作成機能と調整機能の両面を

有するからである。

責任をもってこれらの問題を処理する最初のステップは制度上のものであり、マイクロデータ審査パネル（MRP）は初期の有効な事例であった。第 2 のステップがいま着手されている：マイクロデータに関する漏洩リスクの全体的評価であり、それはマイクロデータ公開のための事情に通じた政策、ガイドライン、および解析的ソフトウェアへと導く。

参考文献

Cox, Lawrence, Johnson, Bruce, McDonald, Sarah-Kathryn, Nelson, Dawn, and Vazquez, Violeta (1985): “Confidentiality Issues at the Census Bureau”, Bureau of the Census – *Proceedings of the First Annual Research Conference*, Reston, Va., 199-218.

Mugge, Robert (1984): “Issues in Protecting Confidentiality in National Health Statistics”, *Review of Public Data Use*, 12, 289-294.

United States Department of Commerce (1978): *Statistical Policy Working Paper 2: Report on Statistical Disclosure and Disclosure-Avoidance Techniques*, U. S. Government Printing Office, Washington, D. C.

付録

マイクロデータファイル公開のための承認要求の提案書を提出するための指示 (チェックリスト)

プロジェクト管理者は以下に記載される資料の各々8部をマイクロデータ審査パネル(MRP)議長に、承認が必要とされる時点の少なくとも1ヶ月前に提出しなければならない。議長はその提案書を議論するためのパネル会議を召集する。プロジェクト管理者は通常、会議への出席か、もしくは提案書と当該調査に知識を有する代表者を送ることを求められる。ファイルの公開を承認するか、拒否するかというパネルの決定は、適切な部課長に覚書きの形で文書報告される。

必要資料(各8部コピー)：

1. 部課長からの概要メモ：調査の目的と設計、および他の関連する情報、例えば承認が必要とされる日付などに関する簡単な記述を含む。
2. テープのレコード構成(ふつうデータベース・ディレクトリーと呼ばれる)：公開を提案するファイル上にある変数と具体的なカテゴリーに関する他の情報のすべてを記載する。時間や費用のために初期のMRP審査用のテープのレコード構成が事前に準備できない場合には、書込み記入事項がコード化される方法に関する説明、およびにテープ上に予定されるすべての他の情報のリスト(例えば、サンプル・ウェイトや地理的情報)とともに、テープに関して提案される項目を示すよう書き加えた調査票を提出せよ。
3. 記入済みの“チェックリスト”：そのファイルの潜在的漏洩可能性を評価するためにパネルが必要とする情報を与える。チェックリストは、あなたが秘匿を理由として削除、もしくは変更を提案する項目、およびそのような取扱いが必要か否か不確かである項目について尋ねている。チェックリストのコピー1部が添付される：このコピーを使って、必要ときには追加コピーが作成される。
4. 公開のために提案されるファイルで識別されるすべての地理的区分 area の人口は少なくとも100,000人の居住者を有することを立証する表。この要求は、1981年2月にセンサス局によって発行された“一般利用マイクロデータを開示するための規準”のなかで制定された。これらの規準では最小限人口数の要求からの選択的な免除は許容されない；しかしながら、通常より漏洩可能性が大きなファイル内容については、それを埋め合わせるためにより高い人口閾値が必要であるとパネルは決定できる。

パネルは、識別される各区域が最新時のセンサスに関する標本抽出に属する区域（例えば第 1 次抽出単位 (PSU)）において 100,000 人の居住者数を有するならば、最小限人口数の要求が満たされると考える。意図された地理がセンサスでだけ集約される都市／農村、場所の規模、もしくは他のカテゴリーを含んでいないならば、希望する場合、調査日により近い人口推定値が使用できる。他の資料、例えば予備センサスからの人口データの利用はパネルによって予め承認を受けねばならない。

ファイル上に表示されるすべての地理的識別子を使ってクロス集計された標本抽出地区 (PSU) の総人口がその表には示されねばならない（付属資料の例を参照）。この目的のためには、地理的情報は一般的には回答者によって与えられる情報、例えば農地の状態などは含まない。その表で使われる人口数字の出所は明示されねばならない。このファイルが異なる地理的識別子によって以前に公開されていた（あるいは再び公開される予定である）場合には、表は 2 つの地理的な組合せで表示されねばならない。識別される地区あるいはその他の地区に関する表の各セルは人口数 100,000 人以上となるべきである；任意のセルが 100,000 人未満の場合には、地理的な分類は条件を満たすように修正され、パネルに提出される表に反映されねばならない。サンプルが PSU ベースで設計された場合には、表内のすべてのセルの合計は抽出された PSU の総人口となるはずである。サンプルが指定標本区域内から選出されていない場合（例えば PSU ではなく、住所録からのサンプル）には、すべてのセルの合計は標本抽出に属すすべての区域（しばしば合衆国全体）の総人口となるはずである。サンプルが PSU ベースで設計されていない場合の表の作成法に関する追加情報については、パネルの代表者に連絡して下さい。

付属資料

人口表の準備

事例：都市／農村，中心都市／非中心都市／大都市圏以外，地域，および選出された SMSA を区別するファイルについては，表は以下のように提示される：

表 1

	大 都 市 圏			非大都市圏	
	中心 都市	他の都 市部	農村部	都市部	農村部
北東部：					
識別された SMSA a					
識別された SMSA b					
その他の地域					
北部中央：					
識別された SMSA c					
識別された SMSA d					
識別された SMSA e(一部)					
その他の地区					
南部：					
識別された SMSA e(一部)					
識別された SMSA f					
西部：					
計：					

総合計= (PSU あるいは他のサンプル地区の合計人口)。

出所：1980年センサス (PC80-1-A レポートから転記)

マイクロデータ・ファイルの漏洩可能性に関するチェックリスト

調査タイトル： _____ 日付： _____
 プロジェクト管理者氏名 _____ 部課 _____ 局 _____ 電話 _____
 資金提供機関： _____

(注意：回答のためにもっとスペースが必要な場合は，続きを別紙で添付して下さい。その質問の節と番号を確実に明示して下さい。)

報告者の負担軽減のために，地理的情報に変更がなく，新しい調査事項も導入されず，かつ最初のファイルで承認された開示手段がその後のファイルにも実施されている限りにおいて，例えば CPS あるいは AHS などの反復的調査についてはその発行毎に別個の申請書を作成する必要はない。以下の申請カテゴリーをチェックせよ。

- この申請書は単独ファイルに関するものである。
- この申請書は実質的に同じ内容をもつ連続ファイルに関するものである。今後のファイルの公開予定までの期間を明示せよ。_____
- この申請書は、補充データ、もしくは以前は非公開であったデータが追加された承認済みファイルの再公開に関するものである。最初のファイルが MRP に提出された日付を記入しなさい。_____

(回答が現在では異なっているようなチェックリストの質問だけが記入される必要がある。)

A 項 ファイルに関する追加的な地理的情報

ファイル上の自明の地理的識別子に加えて、データ項目、レコード識別子、あるいはファイル構造は、推論によって追加的な地理情報を与えるかもしれない。それゆえ、100,000人という最小限人口規準を満たさない地理的領域が不注意にも識別されることを回避する手段がとられねばならない。潜在的に問題のある領域が以下で議論される。各区域に関して、提案されたファイルの公開前に行われたか、もしくは行われる予定の処置を示して下さい。

1. 第1次サンプル抽出単位 (PSU) あるいは他の地理的情報は通常、内部利用のために作成された制御番号 control numbers に埋め込まれている。
 - この問題はどのようにして公開ファイルにおいて回避されますか：
 - _____ 制御番号は削除されるか、地理的情報を含まない。
 - _____ 制御番号はスクランブルされている；記入_____
 - _____ その他；記入_____
 - _____
 - _____

2. 多くのデータベースのレコードは、最初のケースから番号が最も小さい PSU 順か、もしくはアルファベットによる郡の順番で配列されている。
 - ファイル上のレコードに関してこのような地理的な推論を回避するために、どのようにレコードが配列される予定であるか、簡単に説明せよ。_____
 - _____
 - _____

3. 住居の具体的な地理を示唆するデータ項目は、パネルのために準備された人口表が表示する自明の識別子より多くのことを暴露するかもしれない。事例：その州グループに関する自明の識別子がファイル上にない場合にスペイン系の姓を含むこと（南西部の 5 つの州だけでコード化されている）；大都市-非大都市の区別が地理的識別子の一部に含まれない場合の、大都市区域から非大都市区域への移動を明示する移動コード；ある識別される地理的区域には一つしかないような原子力発電所や空港の X マイル以内の住居；電話の地域コード；あるいは緯度と経度の座標値。

○ この理由のため削除予定の項目すべてを記載せよ： _____

○ 地理的に意味をもつと思われるが、削除すべきか否か決定できないすべての他の項目を列挙せよ： _____

4. 標本抽出情報もいくらかの地理的指標を与えるかもしれない。例えば、一定のウェイトは自己表示の PSU と非表示 PSU との区別を与えるかもしれないし、あるいは計画的に過大に抽出された区域のタイプを識別させるかもしれない。また“Durbin タイプ”、“Hit number”などは地理に関係している。

○ 秘匿を理由に削除される予定のすべての標本抽出情報、もしくはウェイトによって識別されないような部分サンプリング計画を列挙せよ： _____

○ 地理的な意味を有すると考えられるが、削除すべきか決定できないすべての他の標本抽出情報を列挙せよ： _____

B 項 個人の露見という異常なリスクを与えるファイル内容

一般利用マイクロデータに対する開示規準は、提案された内容のいずれかが個人の露見という異常なリスクを与えるか否かを決定するために各ファイルの審査を要求する。MRP は、ファイルから入手できる特性値によって個人を識別する可能性を減少させる複数の手段を確認済みである。その手段が以下で議論され、提案されたファイルに属す関連情報はパネ

ルの審査を助けるために要求される。

1. 氏名、住所、および社会保障、医療保障や医療補助番号のような一意的な数値的識別子は、ファイルから除去されねばならない。
2. 高所得は個人や世帯の可視的な特性値であり、情報のなかでは機密性のある項目と考えられる。それゆえファイル上の各所得数字は、世帯、人、家族にかかわらず、総所得も構成員個人の所得も含めて、最大限 75,000 ドル（現在 100,000 ドル）でその上位をコード化すべきである。一定の条件の下ではこのルールへの例外もありうる；例えば、地理的な細区分がほとんどないケースである。75,000 ドル以上の上位コード化からのずれは、ファイル公開承認のための最終討議に先立ってパネルと十分議論されるべきである。

○ 上位記号化は

_____75,000 ドル以下の点である。

_____75,000 ドルより大きい。額を明示し、パネルとの議論を簡単に要約せよ

3. 所得に加えて、一定の他の特性値が他のものよりある個人を見えやすくするかもしれない：例えば、非常に高い年齢、所有財産の価値や購入価格、賃料、抵当権の金額。間隔変数や順序変数として表示される場合には、ファイルの地理的細区分の表示次第では、これらの項目を上位記号化することが考慮されるべきである。パネルの提案によれば、これらの上位記号化したカテゴリーは、そのファイルが表示している（特性値のタイプに依存）すべての人あるいは世帯の少なくとも 0.5% を含むべきである。例えば 1980 年センサスに関する上位記号化は以下のようなものである：

年齢—90 歳以上

財産価値—200,000 ドル以上

総賃料（設備を含む）—1,000 ドル以上。

○ 上位記号化される項目すべてと対応する上位記号点を列挙せよ： _____

○ 上位記号化の必要性に関して疑問があるすべての他の項目を列挙せよ： _____

-
-
4. 地理次第ではある人を高度に可視的にし、非順序変数として表示され、かつそれゆえ上位記号化できないような他の特性値がある；例えば、使用言語が外国あるいはインディアン部族のものであることを示すコード；エスキモー、アリュート、グアム、あるいはサモア人のような人種的細区分；前の居住地に関するコードなど。これらのケースでは、ファイルの細区分の値はより大きなカテゴリーに統合されねばならないだろう。

○ 秘匿を理由として統合（もしくは削除）される予定の項目をすべて列挙せよ：____

○ 細目を統合する必要性に関して疑問があるその他の項目をすべて列挙せよ：_____

C 項 外部ファイルとの照合可能性と結びついた漏洩リスク

本ファイルのマイクロデータが外部ファイルのデータに対して照合される潜在的可能性を減少させるよう努力しなければならない。なぜならば外部ファイルは通常氏名と住所を含んでおり、したがって調査回答者を識別するために利用できるからである。他の機関や組織が保有する“人口記録”や業務リストにもみられる高度に具体的な特性値がその調査に含まれているような場合には、このような照合が可能であろう。例えば、特定の地理的識別子と結合した自動車の型式、モデル、および年式を含めることは受け入れられない。なぜならばこれらの項目は氏名と住所を含む自動車登録リストと照合できるからである。それらが再コード化され、大きなカテゴリーに括られる場合には、これらの項目はおそらくそのファイルに残すことができる。他のリストや記録としては次のものが含まれる；信用部局のファイル；特定の大型耐久消費財（例えば飛行機）の購入者に関して製造業者が所有するリスト；いくつかの州における選挙人登録リスト；連邦、州、もしくは地方の税務記録；犯罪司法システムの記録；州の狩猟および漁獲許可証名簿；および一定の業者団体の会員登録簿。

ある調査の標本抽出枠がセンサス局外部の資料から作成されている場合にも、照合の可能

性は高まる。とりわけ調査記録にその発生源である機関のファイルからのデータが含まれる場合には、標本抽出枠を提供する機関は調査記録を元の記録と照合できるであろう；扶助プログラム受給金額、プログラムへの参入日。

C.1 項

1. この提案されたファイルにも含まれているデータをもつ人口記録や業務リストを知っていますか？

_____はい—そのリスト名_____

_____いいえ

2. 識別される区域にたった一つしかないような、あるいはそれに関する人口記録を得ることができるような特定のタイプの機関内の住居が、容易に利用できる情報に基づいて、ファイル上の任意のデータ項目から識別できるか？

_____はい—その機関のタイプを示せ_____

_____いいえ

3. 提案されているファイルの中のサンプル・ケースには、センサス局外部の資料によって提供されたリストから選出されたものがありますか？

_____はい—その資料を示し、どのようにして、誰によってサンプル・ケースがそのリストから選出されたか説明せよ_____

_____いいえ

C.2 項

外部ファイルが存在する場合、このファイルに対して調査データを照合する可能性を軽減するためにいくつかの処置がとられる；例えば、選ばれた項目は削除もしくは再コード化されたり、あるいは“ノイズ”（すなわち僅かな確率変動）を付与できる。潜在的照合の可能性を十分に低下させるためにどの措置がとられるべきかを、パネルは予め正確に明示することはできない。しかしながら、外部のデータベースと照合する可能性が存在する場

合に、パネルがファイル公開に伴うリスクを決定するときに考慮するいくつかの要因がある：1) 照合目的に利用できる変数の数、2) 照合を行うために必要な資源、3) データの古さ、4) 外部ファイルのアクセス可能性、信頼性、および完全性、そして 5) データの機密性あるいは一意性。照合をもっと容易にするいくつかの要因が下記に列挙されており、照合の可能性を減少させるために、ファイルの公開前にとられる予定の手段に関する情報が要求される。（注意：たとえ照合に利用される外部ファイルの存在を全く知らない場合でさえも、この情報は必要である。）

つぎの場合、照合は一層簡単である－

a) …任意のデータ項目もしくはそれらの組合せによって、かなり小さくかつ識別可能な人口部分が分離できる場合。非常に小さな人口部分を識別させるコードはその導入を避けるべきである；例えば、高度に限定的な地理と組合されたインディアンの部族やコンドミニアムの状態。もしその変数群が人口記録に一度に現われる可能性がないならば、通常は複数の変数を同時に考慮する必要はない。例えば、性別と年齢は外部ファイルに同時に現われる可能性があるが、出生国と職業はそうではない；したがって、ロシア生まれの建築家といった稀れな事象に対する保護策は必要とはされない。

○ ファイルへの導入を提案する項目のうち、少数の、容易に識別できる人口部分を分離させるデータ項目すべてを列挙せよ： _____

○ この理由のために変更（すなわち削除、再コード化、ノイズ付与）される予定のデータ項目をすべて列挙せよ： _____

b) …ある人口部分のすべての構成員をそのファイルが実質的に含んでいる場合。事例：大手の雇用主、高額個人所得者、医者、特定の種類の科学者、あるいは一定種類の機関の入居者。データ公開の前に一定の層内でのサブサンプリングがしばしば必要とされる。

○ ファイルに該当するものがあれば、これらの人口部分とそれがどのようにサブサンプリングされるか示せ。 _____

- c) …個人の識別子もしくは詳細な地理的情報を有する外部ファイルに対して、その情報がリンクするのに役立つようなレコードや他の情報源から得られた任意の情報をそのファイルが含んでいる場合。事例として含まれるものは、公益事業会社からの燃料消費や費用計算記録；10年センサスからの区画、調査区、あるいは ED 要約特性値；民間や政府機関からの福祉や社会保障データ；警察からの逮捕歴。
- ファイルに関して提案されているなかで、回答者との面接調査では得られていないデータ項目のすべてを列挙せよ： _____

- この理由のために変更されたり、削除されるデータ項目すべてを列挙せよ： _____

- d) …照合のためにしばしば利用されるデータ項目、例えば正確な出生日、性別や人種などをそのファイルが含む場合、あるいは、例えば正確な所得額、不動産税額や他の税額、もしくは政府資金によるプログラムへの参加もしくは終了の日付のように、2つのファイルで一致しているはずの他のデータ項目が含まれている場合。
- 該当していれば、これらのデータ項目を列挙せよ： _____

- この理由のために変更もしくは削除されるデータ項目すべてを列挙せよ： _____

- e) …縦断面データが収集されている場合；すなわち、同一の回答者／単位に関するデータがいくつかの異なる参照期間について収集される場合。基本的な懸念は、外部レコードと潜在的に照合可能な時系列的なデータ項目に関係している；例えば、所得税や雇用レコード。もしデータが複数回、同一の回答者から収集されていれば、面接調査の頻度、任意の1単位がサンプルとなる期間の長さ、およびある時点とその次の時点までのある標本単位を一致照合させる可能性に影響する諸要因を示せ。_____
- _____
- _____
- f) …高度に限定的な地理がファイルに含まれる場合；例えば、州、SMSA など。（この地理については人口表で示されるべきである。）

- g) このファイルを外部記録に対して照合する力を減少させるために、今まで言及されていない配慮があれば説明せよ；例えば、データの非信頼性や自然のノイズ。
-
-

D 項 その他の問題

1. すべてのサンプル・ケースを含むファイルは、その部分サンプルだけを含むファイルよりも漏洩に導く可能性が高い。例えば、ある特定の個人がある特定の調査に協力したことが知られている場合、そのファイルですべてのサンプルケースが利用できると仮定すると、その人のレコードは対応するマイクロデータファイルのなかに見出すことができると推論できる。

○ このファイルは下記のものを含む。-

_____ すべてのケース
_____ ある部分サンプル

2. プロジェクト管理者は、内部用ファイルから特殊な表が作成される場合、それが一般利用ファイルでは除外された細区分を含むならば、秘匿性の問題が生じることがあることを意識すべきである。例えば、そのファイルの複数のデータ項目をクロス集計した表は、一般利用ファイルには含まれない特定の地理を与えることがある。パネルはこれらの表の審査の手続きについてその概要を示すガイドラインを用意している。これらのガイドライン（“一般利用マイクロデータの利用可能性の下での調査集計表の漏洩可能性”）を参考にし、一般利用ファイルでは入手できない細区分を利用する統計表の公開を計画している場合には、パネルに相談して下さい。

匿名一般利用サンプルの可能性と展望：

社会科学における国民的データ資源*

Duane F. Alwin

ミシガン大学社会調査研究所

序

かなり重要で大きな関心がもたれている質問として、マイクロレベルの大規模な匿名化された調査データへの実証的な社会学者たちのアクセスに関連する質問がある。このようなマイクロレベルのデータへのアクセスは国によって大きく異なっている。¹合衆国においてマイクロデータへのアクセスは広く普及しており、そのため合衆国にいるわたしの社会科学の同僚たちはしばらくの間、データの配布と個人データの秘匿性の保護の背後にある重大な問題についてあまり知らないまま、そのようなデータ資源をしばしば当然のものと考えている。当然、このタイプのデータへのアクセスはドイツ的文脈においては異なっている。しかし状況は必然的に異なっても、そのようなデータへのアクセスの背後にある複雑な問題に関するわれわれの共通の利害は、それを真剣に検討するきっかけを与えている。経験を共有し、共通する懸念を討議することによって、データの匿名化に関する分野の未解決の研究課題が明確になり、匿名化されたデータベースの配布が促進されるであろう。

この論文では、合衆国の社会学者内部における匿名化されたマイクロデータの配布に関係するいくつかの問題について議論が提供される。このために、合衆国における3つの大規模な社会科学プロジェクトを主に参照する：すなわちハーバード大学のJames A. Davisが企画し、シカゴ大学にある全国世論調査センター (NORC : National Opinion Research Center)

* 著者はこの報告を準備するにあたって、Lawrence Cox, James Davis, Greg Duncan, Mark Abrahamson, David McMillan, Graham Kalton, および Daniel Kasprzyk との会話から多くを得た。

¹David H. Flaherty, *Privacy and Government Data Banks: An International Perspective*. London: Mansell, 1979. を参照。北アメリカとヨーロッパにおけるデータ保護法に関する初期の議論については, Ulrich Dammann, Otto Mallman and Spiros Simitis (Eds.) *Data Protection Legislation: An International Documentation*. Frankfurt am Main: Albert Metzner Verlag GmbH, 1977 を参照のこと。

Duane F. Alwin (1987), "Possibilities and Prospects for Anonymized Public Use Samples : National Data Resources in the Social Sciences" in *Nutzung von anonymisierten Einzelangaben aus Daten der amtlichen Statistik : Bedingungen und Möglichkeiten*, Germany (West). Statistisches Bundesamt , W.Kohlhammer, 1987.

の Davis と Tom W. Smith によって実施された一般社会調査 (GSS : General Social Survey) ; ミシガン大学社会調査研究所の James Morgan と彼の同僚たちが開発した所得動態パネル調査 (PSID : Panel Study of Income Dynamics) ; および Angus Campbell, Philip Converse, Donald Stokes, Warren Miller とその同僚たちが展開した全米選挙研究 (ANES : American National Election Studies) である。これらは公式の政府調査ではないが、社会学者たちによるこれらのデータの広範な利用をめぐる諸問題は、匿名化されたデータベースの配布を議論するには適切である。

これらの課題に関してわたしが提供するの、一般利用データをよく使用し、その早い時期から GSS プロジェクトに相談役的役割で関与してきた一人の社会科学研究者の展望である。²それはおそらく、この種のデータ資源の価値の問題と社会・経済科学におけるこの種のデータに対する需要面に関してもっともよく通じた展望である。データ保護と匿名性の技術的、法的、および倫理的な問題についてはあまり精通したものとはいえないが、簡単にそのような問題のいくつかは以下に議論される。

定義

“一般利用サンプル (Public Use Sample)” という言葉は複数の意味合いを伝える。おそらくもっとも広く普及している使用法は、センサスのサンプルを含めて、いろいろな機構を通して社会・経済科学の研究者が利用できる政府統計調査のマイクロレベル (個人や住居単位) のデータを指す。そのような合衆国の一般利用データの事例としては、経常人口調査 (CPS : Current Population Survey) , 所得とプログラム参加に関する調査 (SIPP : Survey of Income and Program Participation) , 衛生面接調査 (HIS : Health Interview Survey) および一部 10 年センサスからの 1/100 と 1/1000 抽出の一般利用サンプルがある。そのようなデータを社会学者たちにアクセス可能とさせる機構は調査毎に多様であるが、ほとんどの連邦統計調査については審査と配布に関する諸規定が存在する。そのような機構のひとつは機関レベルでの委員会による審査である。ガイドラインと審査過程はデータ保護に関する政府法令についての機関レベルの解釈に任されているという意味において、一般的には、政府統計調査のデータファイルへのアクセスに関する意志決定は機関レベルで行われる。³

² 著者は現在、GSS 管理委員会 (GSS Board of Overseers) の議長職にある。

³ 合衆国センサス局のマイクロレベルデータの公開は議会の法令、“合衆国法典 第 13 編 センサス”，1976 年 12 月 31 日によって規定されている。連邦統計調査におけるデータの秘匿性に関する現代的問題を取り扱った興味ある論文集としては以下を参照せよ。L. Cox, D. Nelson, and V. Vasquez, “Confidentiality Issues at the Census Bureau”, pp.199-218 in *Proceedings of the First Annual Research Conference*, Bureau of the Census。T. Plewes, “Confidentiality: Principles and Practice”, pp.219-226, *Ibid*, W. Griffith, “Discussion: Confidentiality Issues in Federal Statistics”, pp.227-231, *Ibid*.

“一般利用サンプル”という言葉の第 2 の用法は、公開領域 (public domain) におかれた任意の大規模サンプル調査のデータセットを指す。データ保管所 (data archives) を通して社会科学者たちはこれにアクセスできる。そのようなデータセットの収集と管理は政府の資金援助によるものかもしれないが、そのデータはここでは政府目的以外の目的で獲得されており、政府の情報ニーズというより原則的には調査研究のために収集されている。そのようなサンプルは一般的には全国規模のものであり、ミクロレベルの情報の捕捉範囲はしばしば広い。合衆国の 3 つの事例として、ミシガン大学の社会調査研究所によって管理される所得動態パネル調査 (PSID) と全米選挙研究 (ANES)、および NORC によって管理される一般社会調査 (GSS) がある。⁴ これらのデータセットはすべて全国規模であり、多年にわたって調査されている。すべて全米科学基金 (NSF : National Science Foundation) により援助されている。⁵ そのようなデータを社会科学者にアクセスさせるための機構はあまり明確ではない。連邦のガイドラインのなかの諸規定が一般にデータ収集機関による秘匿性の保証に関係し、資金支援機関 (例えば NSF と HHS) は一般的には、関連する諸法規のなかの諸規定を実施するために被験者の保護を処理する地方機関の審査会を頼りにしている。不可避免的に、データの匿名性保護に関する重要な制御の源は社会科学という学科の職業的な倫理規準を当てにしている。⁶

これら 2 つのカテゴリーの一般利用サンプルでは、その調査事項や後援者、その運営と管理の性格、および秘匿性の保証を履行する機構が異なっている。それにもかかわらずそれらの利用や配布に関して多くの共通問題がある。以下においてまず私は社会科学における 3 つの主要なデータ資源、PSID、ANES、GSS を説明し、それらの主要な論点を述べる。それから、必ずしも研究目的を意図して収集されたわけではない調査も含めて、そのようなデータ資源の利用可能性に賛成する論拠を論じる。そして最後に、私の議論はデータ保護と秘匿性に結びつくいくつかの問題点を提起する。それらは、データの出所がどんなものであれ、ミクロレベルのデータへのアクセスを容易にするために満たされねばならない。私の結論は、公衆のプライバシー権と研究者のデータニーズはバランスさせることができるという点にある。要するに、個人の権利と福祉を保護しながら研究を続ける道はある。さて合衆国の 3 つの優れた一般利用データベースを検討し、データの配布と同時に回答者のデータの秘匿性保護に関するそれらの記録をみよう。

⁴ 他の事例がある：全国月次消費者意識調査もまたミシガン大学で実行される；高卒者の全国縦断面調査は NORC によって実行される；および全国労働力参加縦断面調査はオハイオ州立大の Herbert Parnes によって実行されている。しかしながらこれらの調査は社会科学研究者の間ではあまり広く利用されていない。

⁵ 事実上 NSF はここで触れた 3 つの調査に国民的なデータ資源という特別の地位を付与した。それは社会科学の人々にとって重要であるというばかりでなく、その特別な行政上の地位も認めている。

⁶ “修正 ASA 倫理コード” 脚注、1980 年 8 月、p.12 を参照。また全米研究会議 (National Research Council)

一般社会調査

1972 年以来、全米科学基金(NSF)は社会科学のための全米データプログラム (National Data Program for the Social Sciences) , すなわち社会指標研究とデータ・ディフュージョンのための NORC ベースのプログラムを援助してきた。このプログラムの重要な焦点は GSS, つまり約 1500 人の回答者を対象とする年次 (に近い) 全国調査にあった。このプロジェクトの主要な目標は、社会学者たちに大規模で非常に重要な高品質のデータを与えることにあった。GSS は 1972 年から 1985 年までの間で 12 年分実施されている。⁷現在の基金は 1987 年と 1988 年の調査に向けられている。1988 年から先の GSS の継続は、1986 年後期の新しい 5 年周期の資金援助に関して NSF によって審査される予定である。ドイツで GSS に相当するものには、1982 年以来存在する ALLBUS がある。

GSS データは広範囲の変数を含んでおり、それらは社会学者たちにとって常に興味のある多くの領域に触れている。GSS 累積データファイルの中には 700 を超える異なる変数がある。これらの中には標準的な社会-経済的および人口学的変数、そして一連の意識、行動に関する自己申告、および個人の評価が含まれる。意識項目は複数の広範な話題領域をカバーしている。例えば、中絶、犯罪と刑罰、男女の役割、国際関係、国の指導者、国費の優先度、人種問題、宗教上の信条、課税と所得再分配、寛容度、暴力、職業倫理および子供の養育方針などに関する意識である。行動に関する報告は、飲酒と喫煙の程度、銃の所有、組織への会員所属、政治的帰属と投票、社会的相互作用パターン、大きな人生の出来事や悩みを含む。個人の評価は、疎外感あるいはアノミー、総合的な主観的な幸福感、および仕事や結婚、家計、家族、友人、居住場所、余暇時間などの側面についての満足の程度を含む。

GSS のユニークな特徴の一つは反復的な横断面調査の設計にあり、それはさまざまな意識尺度の時間的なトレンドの研究を可能にする。GSS は、予め指定されたローテーション図式にしたがって、それらの質問項目を規則的に繰り返す。加えて、GSS 質問項目の多くはそれ以前の NORC, Gallup, および SRC 調査から選出されており、本調査によってカバーされる論題の多くは GSS の捕捉範囲を超えてトレンド分析が可能である。

方法論的な実験が GSS のもう一つのユニークな特徴である。無作為化実験 (あるいは分割投票) , 特殊尺度, およびテスト-再テスト法の利用によって、GSS データは意識面の測定値に関するいくつかの誤差の源泉を評価することを可能にしている。例えば GSS 調査は、

⁷ “評価研究における個人のプライバシーを保護すること” 1975 年を参照。
⁷ GSS は 1972 年から 1978 年, 1980 年, および 1982 年から 1985 年まで毎年実施された。

質問の言い回しや文脈の影響に関するデータを与えている。⁸GSS テクニカルレポート・シリーズでは、この方法論的な研究の発見物の多くが報告されている。

過去3年においてGSSは、社会科学における重要な本質的問題に焦点を合わせた特殊な1回限りのモジュールを含むようにそのフォーマットを変えた。これらのモジュールの各々がGSS面接調査の約1/4を使用するようにデザインされた。これらの時事的な追加調査の最初のは1985年調査に加えられ、社会ネットワークの問題がその焦点であった。1986年調査は貧困の女性化に関する挿し絵調査を内容とし、そして1987年のモジュールはアメリカ社会における社会-政治参加に関するVerbとNieの1967年調査の反復を内容とする予定である。現行GSS調査の時事的モジュールのデザインは基本デザインのひとつの強化であり、それは、厳密な反復調査という点に加えて、技術革新と新しい測定アプローチの展開を可能にする。

所得動態パネル調査

連邦政府が“貧困との戦い”に乗り出した後、1968年にミシガン大学社会調査研究所の経済学者グループは、貧困の性質と個人のライフコースにわたるその持続性に関する基本的な質問への答えを得ることを目的として、縦断面調査に着手した。⁹約5000世帯のサンプルから始まり、PSIDプロジェクトはこれらの第1次世帯のメンバーへの年次面接調査を行うことができた。本調査は第1次世帯のメンバーばかりでなく、大人になった子供たちが家を出たり、あるいは夫婦が離婚したときに形成される分離世帯も追跡した。(1985年に得られた)第18年度のデータではサンプルはおよそ6800の家族から構成されており、そのなかで主要な成人1人が各年面接調査される。

PSIDプロジェクトによって得られる情報は経済的変数と非経済的変数の両方を含む。そして調査は、第1次世帯を去り新しい世帯をつくる個人も追跡しているので、データセットは個人の経歴をその出発点であった家族とリンクさせる可能性を与える(p.51)。PSIDの焦点は所得の源泉と大きさにあり、それがデータ収集努力の大きな特徴である。そこでは、所得データに関する分析者の理解をを助けるような、例えば世帯構成、雇用状態と失業期

⁸ Howard Schuman and Stanley Presser, *Questions and Answers in Attitude Surveys*, New York Academic Press, 1981 参照。

⁹ この要約は、Greg J. Duncan and James N. Morgan, “The Panel Study of Income Dynamics”, pp.50-71 in Glen H. Elder Jr.(ed.), *Life Course Dynamics: Trajectories and Transition, 1968-1980*. Ithaca and London: Cornell University Press, 1985 によって提起された PSID の記述に基いている。また、James N. Morgan, Jonathan Dickinson, Katherine Dickinson, Jacob Benus, and Greg J. Duncan, *Five Thousand American Families-Patterns of Economic Progress, Vol.1*, Ann Arbor, MI: Institute for Social Research, 1974 も参照されたい。本節の参照ページは Duncan and Morgan(1985)の論文による。

間、職業、住宅、支出状態、地理的な移動性、能力の喪失、健康状態や他の背景情報などの内容ある情報の収集が必要とされる(p.55)。本プロジェクトはまた、例えば認知能力、達成への動機づけ、家庭形成、経済的成功に結びつく多様な意識や行動パターンのようなその他の主題に関しても、一回限りでデータを獲得している。人口分析にとって重要であるが、PSIDは家族生活における出来事の歴史を年代順に記録する。例えば出生史、失業期間、住居の移動、仕事や雇用状態の変化、婚姻関係や世帯構成の変化などである。各年の家族データレコードに追加される変数は約600である。¹⁰

PSIDの第1の価値はそのパネル的なデザインにあり、それは個人レベルでの総計としての変化の識別と個人世帯の社会-経済的経験における持続性と変化の分析を可能にしている。加えて、合衆国内外への移住者を除けば、PSIDサンプルは（分枝家族が面接調査されているので）各年でアメリカの家族に対して代表性をもっており、調査スタッフの指摘によれば、“最新時の面接調査の横断面分析は、センサス局の経常人口調査のような新鮮な横断面データによる分析結果とほとんど変わらない”（p.60）。最後に、Elderに従えば、PSIDはエコロジカルな推移によって誘発されるライフコースや経験の軌跡を調査することを可能にしているため、それはライフコースのダイナミクスを分析するのに驚くほど適している。¹¹

全米選挙研究

ミシガン大学の社会調査研究所の調査研究センター/政策研究センターは1952年以来2年毎に、国民の政治参加を追跡するためにアメリカ人の代表的な横断面の面接調査を行っている。¹²大統領選挙の年にはサンプルは秋の選挙前に面接調査され、そして選挙直後に再調査される。米国議会選挙の年には、選挙後だけ面接調査が実施される。

本来、これらの選挙年の研究は主として選挙行動の理解に関心があったのだが、時間と共にその視野は広範囲の社会政策的行動や意識を含むまでに拡大し、いまやこれらの研究は社会科学や行動科学にとって大きな重要性をもつ国民的データ資源の代表である。政治参加や選挙行動に結びつく政治的な党派性や諸要因のトレンドに加えて、これらの研究は、意識形成や変化のプロセスの基本問題や総体としての世論の変化のダイナミクスに結び

¹⁰ 1974年を初めとして、PSIDプロジェクトは（1984年中に）発見事実を年次集成として発刊し、1983年にはPSIDユーザーガイドが出版された。それはプロジェクトの重要な諸側面を要約している。PSIDのデータテープはICPSRデータ保管所を通して入手できる。

¹¹ Glen H. Elder, Jr., “Perspectives on life course”, pp.23-49 in Glen H. Elder, Jr.(ed.), *Life Course Dynamics: Trajectories and Transitions, 1968-1980*, Ithaca and London: Cornell University Press, 1985 参照。

¹² ここでの議論は、Warren E. Miller, “Long-term support for the American National Election Studies”, A proposal to the National Science Foundation, January, 1984 を基礎としている。

つく広範な主題の検討を可能にしている。

ANES は実際、反復的横断面調査の利点とパネル調査的デザインに関する利点とを結合している。というのは、3 つの別個の状況において当初のサンプルが後に再接触調査されるからである。1956 年研究のかなりの部分サンプルは 1958 年と 1960 年に再調査された。1972 年研究の多数の部分サンプルが 1974 年と 1976 年に再調査された。そして 1980 年の研究では、ある部分サンプルは選挙キャンペーン期間中に 3 度面接調査された。

社会科学の人々の一つのデータ資源として、ANES は、その概念の展開が協同モデル (collaborative model) に基礎を置くという点で、調査研究のデザインのための独特なモデルを与える。積極的な利用者たちの共同体はデータベースの方向性を形作り、各々の連続的な選挙年調査の設計に大きく関与している。このモデルは制度化されており、もっと広い研究共同体のために活動する委員会が、核となる反復的な基本質問項目を超えて、その分野の現在の研究ニーズの方法化を行っている。

ANES プロジェクトはまた方法論的な技術革新のリーダーでもある。プロジェクトは、上述のように、複数のタイプの調査デザインを具体化している。例えば 1984 年調査においては、ANES のデザインには、予備選挙期間全体にわたる全国人口の小規模サンプルの週次面接調査が含まれた。また ANES は革新的なデータ収集テクニックの利用に関する実験も行っている。そこには、大規模で複雑なデータ回収努力に代わるコンピュータ支援電話調査の実行能力に関する 1982 年のテストも含まれている。

社会科学における一般利用サンプルの事例

大規模な一般利用サンプルは経験的な社会科学にとって不可欠のものである。そのようなサンプルデータの利用可能性は、人間行動や社会政策の効果に関する知識に貢献する経験的な社会科学という学問の発展の決定的な側面である。そのような大規模データベースには、社会学者に役立つ 3 つの不可欠の機能がある。一つは社会の動向やその原因を監視し、数量化することである；第 2 は社会プロセスや人間行動に関する主要な仮説を検定することである；そして第 3 は国際的な協同や多国間比較の可能性である。¹³ これらの機能のそれぞれを簡単に論じることにしよう。それが適切な場合には研究文献からいくつかの事例を与えよう。

¹³ 一般利用データベースは方法論の展開や新分野の研究のための刺激などの他の重要な機能を与えるが、現在のコメントではここに掲げたものに限定する。

トレンドとその原因を監視すること

社会変化が現に存在するために、社会学者は強制的に、意識と行動の両面の社会的トレンドの監視に関心をもたざるをえない。Toward Social Reporting: Next Steps(1969)の中で、Otis Dudley Duncan は次のように述べている。“社会変化を測定する改善された能力と手腕は社会報告の分野における進歩の土台である。”過去 20 年の間に、反復的な調査デザインを使って社会変化を記録することに重大な関心が生まれてきた。¹⁴ここで論じられた大規模な一般利用データセットの明白な利益はこの点に代表される。そのような研究のなかには、過去の 1 回限りの調査で行われた測定を繰り返す事例がいくつかある。GSS のような他の事例では、測定は年次もしくはそれに近い形で反復される。

GSS は社会科学における反復的横断面的な調査設計の一つの現代的模範例である。GSS によってカバーされる期間は短い—1972 年以来 12 回の調査—けれども、同じ質問項目を使った過去のギャラップや SRC の調査と関係させることは可能である。これらのデータは、年齢別コーホート内部の時間的な変動パターンの分析を可能にする。GSS と他の意識調査を使った研究から得られた重要な主題の概要は以下のようである：

- (1) 大部分の問題についての社会的自由主義の過去 25 年間にわたる漸増；
- (2) 総体的意識の純変化は、交代と同じく重要な転換によって社会のすべての部分に等しく影響した；
- (3) 世代間の伝播が社会経済的地位ばかりか意識や態度をも形成する；
- (4) サブカルチャーのメンバーであることが一定範囲の意識に重大な影響を与える；
- (5) 職業上の流動性は広く多様な意識尺度にほとんど影響しない；
- (6) 教育は社会政治的な意識にもっとも持続的な影響を与える 1 因である；
- (7) 宗教的—民族的グループ間の差異は縮まりつつある。¹⁵

¹⁴ 以下のものを参照されたい。Eleanor Bemet Sheldon and Wilbert E. Moore(Eds.), *Indicators of Social Change: Concepts and Measurements*. New York: Russell Sage Foundation, 1968. Angus Campbell and Philip Converse, *The Human Meaning of Social Change*. New York: Russell Sage Foundation, 1972. Otis Dudley Duncan, Howard Schuman and Beverly Duncan, *Social Change in a Metropolitan Community*. New York: Russell Sage Foundation, 1973. Kenneth C. Land and Seymour Spilerman(eds.), *Social Indicator Models*. New York: Russell Sage Foundation, 1975. Elizabeth Martin, “Surveys as Social Indicators: Problems in Monitoring Trends”. Pp.677-743 in Peter H. Rossi, James D. Wright and Andy B. Anderson, *Handbook of Survey Research*. New York: Academic Press, 1983.

¹⁵ 社会的トレンドに関する発見事実はほんの簡単にしか要約できない。重要な参考文献には次のものがある： J. Davis (1978), “Trends in NORC General Social Survey Items 1972-1977”, GSS Technical Report No.9. J.Davis(1980), “Conservative weather in a liberalizing climate: change in selected NORC General Social Survey Items, 1972-1978”, *Social Forces* 58: 1129-1156. J.Davis(1982) “Achievement variables and class cultures: family, schooling, job and forty-nine dependent variables in the cumulative Gss”, *American Sociological Review* 47: 569-586. J.Davis(1975) “Communism, conformity, cohorts, and categories: American tolerance in 1954 and 1972-73”, *American Journal of Sociology* 81: 491-513. J.Davis “New money, an old man/lady and ‘two’s company’: subjective welfare in the NORC General Social Surveys, 1972-1982”, Unpublished GSS report. J.Davis “Up and down

社会プロセスに関する主要な仮説を吟味すること

ほとんど疑問の余地がないことであるが、重要な社会的、経済的変数に関する大規模な一般利用サンプルの大きな利益の一つは即座に多くの変数間の関係を吟味できることにある。これらが全国範囲の調査であり、そのサンプルサイズも大きいことは、多変量モデルによる社会プロセスの検討を可能にする。おそらくそのもっとも有名なアメリカの例は Peter Blau と Otis Dudley Duncan による世代間および世代内の職業上の流動性に関する現代的—古典的研究であり、それは世代の職業変化と呼ばれ、1962年3月の経常人口調査(CPS)に基いている。¹⁶ドイツにおける対応する研究は Karl Ulrich Mayer と Walter Müller によって1970年マイクロセンサスを基礎に実施され、アメリカの研究は Robert Hauser と David Featherman によって1973年CPSにおいて繰り返された。

パネル調査的デザインであるからこそ可能である PSID プロジェクトのもっとも重要で実質的な発見事実のひとつは、子供たちによってその生活環境が受継がれていく相対的に同質で、安定的な家族集団という“貧困者”に関する通俗的な観念に疑問を投げかけたことである。¹⁷これは、世代内および世代間の経済的安定性に関する一般的な“貧困の文化”理論を与え、とりわけ興味深い。家族収入における安定性の相対的な欠如に関するこれらの PSID の新発見とともに、家族構成や労働力の活動/参加などの変化と結びつく諸要因が適切に分析された。

ANES は、その開始以来、政治学の多くの分野にとって選挙行動の分析のための頼み綱であった。¹⁸上記のように、このデータベースは世論や意識の変化と密着した問題の分析においても重要である。最近数年間のアメリカ政治学内部の中心的な論争の一つは、アメリカ人の政治意識の堅固さに関係している。1960年代と1970年代の ANES データを使って、Philip Converse は、政治意識が安定的で有力な行動指示的な気質を示す程度について、政治意識は非常に多様であるという広く影響力のある論証を与えた。¹⁹これらのデータが示

opportunity's ladder”, *Public Opinion* 5: 11-15, 48-51. J. Davis and T. W. Smith(1982), “Have we learned anything from the General Social Survey?” *Social Indicators News-letter* 17: 1-2, 8-11. J. Davis(1976), “Background characteristics in the U.S. adult population 1952-1973: a survey-metric model”. *Social Science Research* 5:349-83. J. Davis(1983) “Counting your change for a ten: America from 1972 to 1982 as reflected in the NORC General Social Survey”, GSS Technical Report No.43.

¹⁶ Peter Blau and Otis Dudley Duncan, *The American Occupational Structure*. New York: Wiley, 1967 を参照せよ。GSS はこれらのアメリカの流動性に関する調査を一層確証させる証拠を与えた。

¹⁷ Greg H. Duncan, *Years of Poverty, Years of Plenty*. Ann Arbor, MI: Institute for Social Research, 1984; Glen H. Elder, Jr., *Life Course Dynamics: Trajectories and Transitions, 1968-1980*. Ithaca: Cornell University Press, 1985 を参照せよ。

¹⁸ 例えば、Angus Cambell, Philip E. Converse, Warren E. Miller and Donald E. Stokes, *The American Voter*. New York: Wiley, 1960 を参照。

¹⁹ Philip E. Converse, “The nature of belief systems in mass publics”. In D. Apter(ed.), *Ideology and Discontent*. New

すところによれば、政党や政治家に関する意識などのいくつかのタイプの意識は安定的でかつ情緒的に強い方向性をもって反映されるが、その他の意識、とくに特定の政府の施策に関する意識は、人間の毎日の生活から隔たっており、大部分の人間にはかなり変化しやすく、それゆえ非常に強い思想的な性向はみられない。支持政党に関するコーホート vs. 時間効果についての Converse の分析と Knoke と Hou による体系的な分析は、ANES が可能にしたもう一つの重要な貢献である。²⁰

多国間比較の可能性

比較研究の価値は大いに確立され、立証されている。多国間比較から得られる知識は大規模な一般利用サンプルのもう一つの利益を代表する。しかしながらそのような作業の成功は、協同者が国家横断的な厳密に等価的なデザインをすすんで築こうとするか否かに依存する。²¹これは GSS に関しては成功した。最初の GSS との国際的協同は 1982 年のドイツ全国社会調査 (ALLBUS) において調査・方法・解析センター (ZUMA: Zentrum für Umfragen, Methoden und Analysen) と共同で行われた。それ以来、多数の国の社会学者たちが GSS モデル、イングランド、ウェールズ、スコットランドにおける英国社会意識調査 (British Social Attitude Study)、およびオーストラリア国立大学によって実施された年次調査などに乗り込んできた。比較可能な多国間の一般利用データセットの存在は、多様な社会的文脈の中での発見事実の交換や同一仮説/理論の反復実験を行う機会を提供し、システム・レベルでの差異をよりよく理解することを可能にする。

そのような国際比較を促進するために、GSS データは、合衆国ではローパー・センター (Roper Center) にあるデータ保管所とミシガンにある ICPSR、イギリスではエセックス大学の社会科学研究会議調査保管所 (Social Science Research Council Survey Archive)、ドイツではケルン大学の経験的社会研究中央保管所 (Zentralarchiv für Empirische Sozialforschung) を通じて国際的な社会科学共同体の人々には入手可能である。さらに NORC スタッフは個人ベースで海外の学者にデータとコードブックを提供することを厭わない。PSID と ANES は両者とも国際的な構成部分をもっており、これらのデータは国際的なデータ保管所によって入手できる。

York: Free Press, 1964; および "Attitudes and non-attitudes: continuation of a dialog". In E. R. Tuft (ed.), *The Quantitative Analysis of Social Problems*. Reading, MA: Addison-Wesley, 1970 を参照。

²⁰ Philip E. Converse, *The Dynamics of Party Support: Cohort Analyzing Party Identification*. Beverly Hills, CA: SAGE, 1976. David Knoke and Michael Hout, "Social and demographic factors in American political party affiliations, 1952-1972." *American Sociological Review*, 1974, 39: 700-13.

²¹ 過程的等価 vs. 機能的等価に関する議論については E. Martin (脚注 16 参照) をみられたい。

一般利用サンプルに対する需要の評価

一般利用データベースに対する需要はあるのか。わたしの議論の決定的な部分のひとつは、社会科学の研究共同体内部において、一般利用サンプルに対する 2 つのタイプの需要が現に存在するという点である。明らかに、そのような質問に対する答えは社会的歴史的な事情に依存し、そのためどんな一般的な答えも経験を共有する特定の状況に限定される。われわれの経験では一般の人々もいるけれども、わたしはそのようなデータに対する自分の国の需要を評価する準備しかない（そしてこのことに限定する）。

過去 20 年間、合衆国における統計諸機関は、研究や統計上の目的のためにマイクロデータへのアクセスを求めて増大する圧力にさらされてきた。²²経済学者、社会学者および人口学者たちは多年にわたって、研究目的のためにセンサス局によって用意された一般利用サンプルに依存していた。研究者たちとその学生たちはセンサス・マイクロデータを定期的に利用する。1979 年に Flaherty (p.293)は、200 人以上の研究者がセンサス局からの一般利用データを使用していると推定し、そして彼は、“少なくとも 1000 人の大学院生と学部生がセンサスからの一般利用サンプルを使用しており、少なくとも 10000 人以上の学生がその存在を知っている”と示唆する 1 人の事情通の第 3 者の言葉を引用している。おそらく 100 本の社会学の論文はセンサスからの一般利用サンプルの分析を基礎に準備されたと彼は推定した。このことは今からほとんど 10 年前のことであり、データへの需要は増大するであろうと予測された。より最近の情報はセンサス局からほとんど確実に入手できる。センサス局は率先してデータ配布への需要に応じており、それはマイクロデータへのアクセスから生じる利益のもっとも強力な例証の一つを与えるけれども、同様の圧力は国立衛生統計センターや社会保障総局も経験している。

政府以外の大規模な一般利用データに対する需要の評価はかなり困難である。なぜなら、そのようなデータはデータ保管所を通して広い範囲からアクセスでき、このようなデータ利用を測定する試みをわたしは知らないからである。したがってこの領域における情報は一般に限定的である。GSS という特定の事例に関しては、そのデータベースを使用する出版物についての非常によい文書がある。GSS のスタッフは、主要な利用者と社会科学雑誌に関する調査から編成されたリストと、社会科学摘要と引用文献集のコンピュータ利用による検索に基いて、研究目的や出版のために GSS がかなり利用されていると報告している。

23

²² この議論は、上に引用した Flaherty(1979:292-300)に大きく依存している。

²³ T. W. Smith and M. Ward, *Annotated Bibliography of Papers Using the General Social Surveys*, 5th edition, NORC (May, 1984) を参照。その著者たちは、1972 年以降、出版物や完成した研究論文に記載されたものとして、知られているものだけで 1000 件以上の利用をリストしている。彼らは、“選定された論文のリストは一般社会調査の利用全体にははるかに届かないと信じられ、それは書籍、提出中の論文、ごく最近

ANES の利用についても比較可能な数字がある。主要な実質的選挙データの利用の全体に関する選挙研究の推定によれば、およそ 82 の書籍、580 の雑誌論文や本の章、156 の会議論文、134 の博士論文が含まれ、加えて 136 の出版物はよくそのデータを使用している。教育などに関係する相当数の利用がこれらの指標には反映されないため、この種の使用件数の計算は明らかにこのようなデータ資源に対する需要を過小推計している。

これらのデータ資源の広範な利用を示す指標は他にもあり、わたしは、かなり印象主義的な情報とこれらのスケッチ的な指標から、現代社会科学の内部では、一般利用データに対する相当大きな需要があると結論づける。さてここで一般利用データの広範な配布制度の下でのプライバシー、秘匿性、およびデータ保護の問題に移ろう。

プライバシー、秘匿性、およびデータ保護

社会歴史家の Barrington Moore は、「プライバシーはどんな社会においても支配的な価値にはなりえない」と述べている。人は社会で生きていかなければならず、社会的関心が優先しなければならない。²⁴古代ギリシャや古代中国に関する彼の分析が示唆することは、“プライベート”や“パブリック”という言葉が存在しており、前者は“反社会的”を暗示していたということである。いろいろな古代社会が社会的関心事の優先性を示しているのに、必ずしもそのような関心事がすべて優先するとはかぎらない。実際、B. Moores の結論は、西洋文明のひとつの大きな成果は、社会的関心事に疑問を投げかけるときのプライバシーという観念の役割であり、すなわち、“社会的権威に逆らう個人を保護する諸権利に関する意識”であるという。

Riecken と Boruch はプライバシー問題と秘匿性問題を区別した。²⁵プライバシー問題は、個人が 1 人であること、あるいは彼/彼女の私的な活動、意識、および主観的評価に関する観察者（調査データの場合は面接調査員）の侵入による観察やデータ収集の過程において発生する。とりわけデータ収集過程が他人と共有したくない情報を個人に漏らすように要求するときには、プライバシーに関する質問が問題となる。通常の調査過程は、もし彼

の著作物をとくに見逃す傾向にある。”と記している。

²⁴) *Privacy: Studies in Social and Cultural History*. London: M.E.Sharpe, Inc., 1984, pp.274-75.。プライバシー問題に関する追加的な議論については以下を参照。David M. O'Brien, *Privacy, Law and Public Policy*. New York: Praeger, 1979. Ferdinand D. Schoeman(ed.), *Philosophical Dimensions of Privacy: An Anthology*. Cambridge: Cambridge University Press, 1984.

²⁵ Henry W. Riecken and Robert F. Boruch(1974), *Social Experimentation: A Method for Planning and Evaluating Social Intervention*. New York: Academic Press(pp.255-69)。また Robert F. Boruch, *Assuring the Confidentiality of Social Research Data*, 1979 を参照のこと。

／彼女が望むならば、与えられた質問に個人が回答しない権利に配慮している。秘匿性問題は、論理的にはデータ収集の後に、潜在的に個人が識別され、彼／彼女の身元がその情報にリンクされるような方法で、調査で得られた情報を漏洩するように要求されるときに生じる。

このとき明らかに、個人のプライバシーの権利と、公共財とみなされるもの、すなわち政府の情報ニーズや研究上の調査へのこの情報の利用可能性との間にはある緊張が存在する。プライバシーに対する市民の諸権利という観点から、これらの問題についての議論は、政府の真意や秘匿性の保証と結びつくその責任に関する懸念を引き起こす。疑問の余地はほとんどなく、これらのプライバシー問題は極めて重要であり、秘匿の原則はデータの配布において妥協されるべきではない。

社会科学者の見地からの一般利用データへのニーズに関する議論は、情報の秘匿性に関していくつかの問題を引き起こしている。第 1 に、一般利用サンプルにおける漏洩リスクとは何か、そしてどんなタイプの情報が決定的にそのような危険性を産み出すのか。第 2 に、調査レコードのプライバシーの保証が満たされていることを確信させる有効な審査機構とは何か。第 3 に、識別可能なプライベートな情報の漏洩リスクを減少あるいは除去する最高のテクニックとは何か。

漏洩リスク

一般利用データセットに含まれる個人情報の秘匿性を保証することへの懸念は、人間に関する研究を管理する諸政策の反映される。例えば、プライバシー法の焦点は以下のようである：

- (1) レコードの研究利用と行政利用とを区別すること、
- (2) レコードが行政利用から研究利用へと流れるような条件を注意深く規定すること、
- (3) 研究利用から行政利用への逆の流れを禁止すること²⁶。

さらに、それは生医学研究における個人保護のために第 1 義的に展開されたのだけれども、研究上の被験者の保護をめざす諸政策は、NSF のような重要な政府機関によって支援される調査についてはその回答者の秘匿性を保護することを要求している。²⁷

²⁶ Bradford Gray, "Developing issues in the ethics of social research on health". Health Survey Research Methods, 2nd biennial conference, Williamsburg, Virginia. National Center for Health Statistics Research. U. S. Dept. of health, Education and Welfare, 1977. DHEW Publication No. (PHIS) 79-3207.

²⁷ The Federal Register, "Final Regulations Amending Basic HHS Policy for the protection of Human Research Subjects", vol.46, no.16, pp.8366-8392 を参照。

合衆国において研究上の利用のために公開される一般利用サンプルは、個人の識別子が機械可読な形態では支払われないという意味で、一般的には匿名化されている。氏名や社会保障番号、および街区住所という形態の識別情報は、一般利用データセットのレコードからは除去される。したがってこのような処理の下では、一般利用データセットを得た研究者によって氏名や住所から個人が識別される危険はない。政府以外での研究において個人をそのデータにリンクさせるレコードは、一般的に破壊されているか、あるいはデータを収集する機関の管理下で別個に保存されている。それでは一般利用マイクロデータの公開から、どんな問題領域が個人に属すデータの漏洩リスクと関係するのか。

政府以外の調査における一つの潜在的な懸念は、法的プロセスに反する研究上のデータの秘匿性の保護である。刑事法廷が個人に関する研究上の記録を召喚する可能性は、回答者の秘匿性にとって脅威であり、研究者は法律上それを保護することができないかもしれない。Riecken と Boruch (1974) は、研究者たちには名誉をかけて秘匿性を守る道義的な責任と職業的な責務があるのに、研究者たちにとってそれを守ることが困難、もしくは不可能であるような状況があるかもしれないと記している。これらの状況の多くは単なる仮説的なものかもしれないけれども、合衆国では立法府委員会や司法当局が研究の過程で獲得された情報を入手しようとした複数の事例がある²⁸。Riecken と Boruch (1974, p.256) は、政府当局が拘禁すると脅して研究者から情報をうまく獲得した一例を引用している。これらはまれな事例ではあるが、秘匿性を侵す潜在的可能性をわれわれに思い出させる。このような潜在的な危険性のために、研究者たちはすべての識別情報をすすんで除去したり、もしくは国外にレコードリンケージを貯蔵し、法的な救済を求めて努力した。

秘匿性の領域における第 2 の懸念は、匿名化されたデータを使う研究者やそのスタッフの一人がファイルに含まれる地理的情報や他の情報を帰納的に利用して、個人個人の身元にたどり着くかもしれないという点である。特定の回答者の身元が統計レコードの中の回答の一意の組合せから推論できるような場合には、大変な注意力と管理力が働かされねばならない。合衆国においてこのような事例は決して存在しないことを知っているが、それが現実となる可能性があるため、一般利用サンプルの調達者はそのファイルのコード化された変数を審査するさいに極端に用心深くならざるをえない。例えば、GSS はその累積ファイルにおいては合衆国の地理的な区分 (9 カテゴリー) だけをコード化しており、州コードは与えていない。ある所与の調査の標本抽出枠についてあまり多く知られていない場合には、もっと詳細な地理的な情報がなければ、一定の回答者が識別されることは事実上不可能であろう。他の調査、例えば ANES では住居の状態が与えられており、例えば PSID のようないくつかの事例では、より詳細な地理的情報が与えられる。これによって一定の回

²⁸ Bradford Gray(1977)は 1 つの研究を引用しており、それは、研究上の記録に対して発された少なくとも 17 の召喚状と政府の情報需要に関する 26 の他の事例を報告している。

答者の身元が暴露される可能性は増大する。²⁹このような事はほとんど可能性としてありそうもないが、一般利用データを配布する機関や組織は審査にあたってその可能性に注意を払わねばならない。

一般利用サンプルにおける個人データの秘匿性を保護するためにいつも非常な努力が払われるけれども、社会の認識はしばしば異なっている。全米科学アカデミーの資金援助による合衆国での1979年全国調査は、いくつかのかなり驚くべき結論を与えている。³⁰彼らは以下の質問を尋ねた：「氏名と住所によって識別される個人の調査記録は合衆国センサス局のファイルの中に保存されている。これらの記録には職業、所得、人種および年齢などの事項に関する情報が含まれる。あなたは、これらの記録がそれを見たい人には誰にでも公開されているのか、そうではなく非公開なのか、ご存知ですか。」³¹回答者のほんの35%がそのようなファイルは非公開であると答えた；18%はそれらは公開であると答えた；そして46%は公開か非公開か知らなかった。さらに他の政府機関によるセンサスレコードへのアクセスに関して調べると、センサスレコードは完璧に秘匿されていると実際に信じていたのは、全回答者中たった5%にすぎなかった。80%の人は、センサスレコードは秘匿されていない、もしくは他の政府機関が、実際にやろうとすればセンサス局から情報を獲得できるだろうと答えた。

明らかに、センサスレコードの匿名性に関する社会の意識は、法的な要請と現実の実務から全く乖離している。秘匿性の保証と機関の実務を規制する法令に関して社会をもっと教化する必要性を指摘することに加えて、この状況に関する社会の誤認こそは、個人データの秘匿性の保護に非常な関心を払う一層強い理論的根拠となる。

リスクの除去

秘匿情報の漏洩が生じないことを保証するために、審査と監督のための実質的な責任はデータを配布する機関あるいは組織の側にある。公表されるその種の情報をモニターする機関の審査パネルの機構が、現実には重要な防護策となる。そのときどんな情報を公開すべきかの決定は、ガイドラインと審査過程が秘匿性の保護に関する政府法令の当該機関レベルでの解釈に依存しているという点で、分散化している。政府の援助によって獲得された非政府系の一般利用データの場合には、秘匿性の保証に付随する重大な規制がある。しか

²⁹ PSID データは匿名化された形態で公開されているけれども、居住地の州と郡はコード化され、そして“回答者によって報告される環境情報は、その州の雇用保障職員から得られる失業水準、単純労働の賃金率、市場の需要状態についての郡レベルのデータによって補足される。”(p.55)

³⁰ Privacy and Confidentiality as Factors in Survey Response. National Academy of Sciences, Washington D.C., 1979. 参照。

³¹ もちろんそのようなレコードをセンサス局以外の誰かに漏らすことは犯罪である；それらは他の連邦諸

しこれらのケースでは、資金援助機関は一般には、被験者の保護を処理する地方機関の審査委員会を頼りにしている。先述したように、データの秘匿性保護をコントロールする重要な源泉は、職業上の倫理基準に依存する。

ここで述べた審査機関に加えて、データ漏洩の潜在的危険性を減少もしくは除去するという観点から、これらの審査パネルがもつ代替策を簡単に考察しておくことも重要である。個人識別情報の除去に加えて、個別データを一層匿名化するために複数の周知の機械的な技法が存在する。³²一つの解決は、潜在的可能性をもつ識別情報をすべて単純に削除することである。もし個人が、他の変数と組合せて彼らが住むセンサス調査区の知識から識別される可能性があれば、調査区情報は統計レコードからおそらく除去されるであろう。もう一つの解決は、攻撃されそうな変数をもっと粗いカテゴリーに再コード化することである。その結果、公開される変数は原データファイルほどには詳細ではなくなる。3番目の解決は、データに一定量のランダムな攪乱を導入することである。その結果個人レコードは誤差を含んでいるが、統計的推定値は誤差レベルを修正できる。そして最後に、データは“マイクロ集計値”として公開される。そこではデータは小集合に関して平均され、データはマイクロ集計値を構成する一人の個人としてより、むしろこれらの“みせかけ”の個人としてだけ公開される。

データの秘匿性保護の問題に対するこれらの機械的な解決に加えて、Riecken and Boruch (pp.259-61) は、非政府系の調査におけるデータ保護にはとりわけ、現実には法律上の他の解決が究極的には必要であると主張する。ひとつのそのような解決は、研究調査者が1種の、僧侶や弁護士によって享受される種類の“宣誓特権 (testimonial privilege)”を有することである。これは司法や議会の行動から生じるプライバシー侵害を防ぐだろう。第2の法律的な解決は、回答者の身元を漏洩する研究者やそのスタッフの一員に対して厳しい法的な処罰を法律上課すことである。最終的に明らかなことは、研究調査者やそのスタッフは、秘匿性の倫理基準、およびデータ利用と秘匿性の保護を規制する法律や政府の規制に関して不断に教化されることが必要である。

機関には入手可能ではない。

³² この議論は主に Riecken and Boruch(1974)に依拠している。

統計研究参考資料(最近発行分)

35	外国人の地域分布	91/09/30
36	ソ連産業連関表とその利用	92/02/28
37	EC統計局型アメリカ合衆国産業連関表	92/12/31
38	日系ブラジル人就労・生活実態調査	93/02/28
39	『女性と統計』関連主要文献	93/07/31
40	インストローと女性に関する統計	93/09/30
41	中国統計制度論集	93/12/31
42	ジェンダー統計の現状	94/01/31
43	ロシアにおける統計制度・政策の改革	94/03/31
44	EU型日本産業連関表(改訂版)	94/12/31
45	国連(1984年)『女性の状況に関する統計と指標のための概念と方法の改善	95/12/31
46	フィリピンの農業統計	95/12/31
47	ロシアにおける統計制度・政策の改革(Ⅱ)	96/03/31
48	統一価格中国日本産業連関表 1985, 1987, 1990年	96/03/31
49	国連(1995年)『世界規模のジェンダー統計に関するワークショップ』(翻訳)	96/04/30
50	英国統計制度関係資料	97/03/31
51	インストローとジェンダー統計	97/06/30
52	アメリカ合衆国労働省 労働統計局『製造業生産労働者の時間当たり報酬	97/07/31
53	わが国における外国人の死亡特性	97/12/15

統計研究参考資料 No.54

アメリカにおけるマイクロデータの提供

1998年1月

発行所 法政大学日本統計研究所

194-02東京都町田市相原町4342

Tel. 0427-83-2325

Fax. 0427-83-2332

E-mail jsri@mt.tama.hosei.ac.jp

発行人 伊藤 陽一