日本でのミクロデータ公開の展望

永山 貞則（日本統計協会）

1．はじめに

ミクロデータ公開のための基礎的な情報は，今回の特定領域研究を通じて相当程度蓄積されてきた。その一つは理論的な研究であり，もう一つは諸外国の情報である。理論的な問題は，ミクロデータの公開によって個体のデータが暴露される危険と，個体の秘密保持のために失われる情報の損失とのバランスの問題である。個体データの安全性を高めようとすれば当然に情報のロスは多くなり，情報のロスを少なくすれば暴露のリスクが増す。この両者の関係についての理論的な研究とその実験が精力的に進められた。

また実際のミクロデータの提供に当たって，諸外国ではいかに対応しているかについて多くの情報が集められ，特に米国，英国，ドイツ，オーストラリア，カナダについては実地調査も行われた。安全性と情報損失のバランスの問題は理論だけでは解決できない。実際の提供に当たってどういう問題が起こったか，また起こらなかったか（これも重要である）についての諸外国の経験は，今後，提供を始める我が国にとって極めて有力な情報である。

しかしこの理論的な研究と諸外国の情報だけで日本のミクロデータの公開が一挙に進むわけではない。そのためにはこの 2 種類の情報を結合させて具体的な提供方法のシナリオを画き，その法制的な取扱についての合意が必要である。その戦略については以前にも書いたが，ここで今後の展開について整理してみよう。

なお用語についてであるが，現段階では，国民が果たしてミクロデータ利用の重要性と，秘密保持の安全性を意図通りに理解してくれるかどうかの懸念がある。したがって今後の提供を進めるに当たっては，無用な誤解をされないように細心の注意が必要である。ミクロデータの公開というと，国民は個票データを一般に公開するという意味に解され易いので，ここでは「匿名標本データの提供」とする。提供されるミクロデータは，個体の識別ができないものであり，センサスの場合でも全部ではなく一部の標本である，という理解を容易にするためである。

2．個体の識別可能性－絶対的秘匿と事実上の秘匿－

個体データの秘密漏洩の問題は，主として個体が識別されるかどうかの可能性の問題で

ある。氏名，住所等の識別子が消去されている場合には，母集団一意(population unique)な対象が標本の中から識別される可能性の問題である。理論的研究では，個体の情報を得ようとする侵入者が，母集団に実在する特定の個体を個票データセットの中に探す場合と，個票データセットの中の特定レコードに対応する個体を外部に探す場合について，いろいろな条件を想定して検討されている。[竹村 1] 例えば識別可能性の観点から家計調査の安全性を確かめた実験例では，「キー変数のうち”世帯人員”を削除し，その他のキー年数を global recording した上で，抽出率 0.5 の部分標本をとればよい」という結果が示されている。[加納 2] しかし一方，利用者の立場からみれば，家計調査における世帯人員は重要な変数であり，これを除いたデータでは利用価値は薄いものになってしまう。

このように理論的に安全性を保持しようとすれば情報のロスが大きくなりすぎる場合には，諸外国ではこの問題にどう対応したかの経験が役立つ。

ドイツでは，1980 年連邦統計法第 11 条第 5 項で，科学研究への調査個票の提供を”絶対的秘匿”を条件に許可してきたが，多くの情報が失われるためほとんど利用されなかった。そこで 1987 年連邦統計法第 16 条第 6 項では，申告個票が非常に大きな時間，経費及び労力の支出によってだけ識別できる場合に限り許可されることに改正された。つまり膨大な手間をかけて探索しない限りは漏れないという”事実上の秘匿”に転換したのである。[濱砂 6]

連邦統計法では事実上の秘匿の定義しか与えられていないが，その後，秘匿の効果を実際のデータで検証した結果が報告されている。[ミューラー 11] そこでは共通の標識を使って外部ファイルと一対一の対応ができたものが，調査の実際の名簿と照合すると一致した数ははるかに小さいという結果が示された。確率論的な一意性のリスクは，実際よりも著しく過大に評価されていることが明らかとなった，とされている。

この検討は重要な意味を持つ。もし前述の家計調査における個体識別の可能性が，実際にははるかに低いものであれば，世帯人員の変数を除くことは情報のロスを必要以上に大きくしたことになる。”事実上の秘匿”という概念を弾力的に運用することによって，不要な情報のロスを防ぐことができる。

3．匿名標本データのモデルセットの作成と検討

我が国でミクロデータの提供が具体化していない理由の一つは，実用的な匿名標本データのセットを作らずに議論だけがなされていたからである。ドイツの例をみても，実験的なデータによる結果と，実際のデータによるチェックとでは大きな差があった。したがって実際に匿名標本データのモデルを作成して検討することが必要である。(なお現在は総務庁統計局で検討が始められている。)

モデルセットでは，二つの検討がなされるべきである。一つは技術的な検討であり，もう一つは安全性のための検討である。

(1) 技術的検討

提供される匿名標本データは，利用者にとって使い易いものでなければならない。検討すべき項目としては，次のようなものが考えられる。

・メタデータの整備：入力形式，項目コード，分類表等のメタデータを，どの程度まで整備し，どういう形式で提供すればよいか。

・標本のサイズをどうするか。秘密保護の観点からいえば，侵入者が特定の個人が標本調査の対象となっていることを知っている場合，提供を例えば 90%の標本に減らせば，一対一の対応ができても特定の個人とは断定できなくなる。

・匿名標本データで提供する標識（変数）の選択

・地域別に抽出率が異なる場合，乗率をどうするか.... 等。

(2) 秘匿方法の検討

諸外国の経験をみても，個体識別不能にするために最もよく用いられる手法は，地域の限定と top cording である。

①地域の範囲の限定：識別できる地域の範囲を限定することは極めて有効な手段であるが，どの位の大きさが適当かは諸外国の例が参考となる。例えば米国では，1960 年センサスからミクロデータが公開されてきているが，当初は最低人口 25 万人未満の地域が識別されてはいけないとなっていた。しかしその後，裾切りが大きすぎるとして，限度を人口 10 万人未満の地域に変更された。このように情報のロスが大きく，かつ開示リスクの問題が実際に起こっていない場合には，その基準を弾力的に変更している。[石田 7]

またデータセットの中の地理的な順序を変えておく工夫も必要である。

②top cording：目立ちやすい項目，例えば高年齢，高所得，多人数世帯等，については top cording で多くが処理される。米国では，例えば所得分布の高所得部分をグルーピングした場合には，最高階級の平均値，中央値，分散等の特性値が併せて提供されている。これは所得分布の分析上，効果的な方法である。[US12]

(3) その他の検討

・ドイツのような外部ファイルとの照合の可能性についての実験。

・global cording, swapping 等の検討

・複数のデータセットの作成.... 等。

4．提供先(利用者)の限定の問題

(1) 一般公開か，限定的提供か

匿名標本データを一般公開(販売)するとなれば，不特定多数の利用者に提供することになり，その場合には悪意の侵入者の可能性も否定できないので，高い安全性が要求される。したがって当然に情報のロスは大きくなる。

一方，提供の対象を分析研究者あるいはそれに類するものに制限し，かつ誓約書を交わす方法をとれば，善意の利用者を前提にリスクの可能性を考えればよく，必要以上の情報のロスをさけることができる。

現段階では国民がミクロデータの利用をどこまで理解してくれるかが分からないので，利用者の範囲を限定した提供を行うのが妥当と考えられる。

(2) 誓約書

調査実施者が匿名標本データを提供する場合は，利用者との間で次のような趣旨の誓約書を交わすことを義務づける。

① 統計的分析目的にのみ使用する。

② 個体が識別可能の恐れがあるような使用及び発表はしない。

③ 発表に当たっては使用データ名を明記する。

④ 他人に譲渡はしない。

⑤ 営利目的には使用しない。

(3) 海外の利用者

米国の公開用ミクロデータの場合は外国人も購入できるが，英国やドイツの場合，海外の利用者については厳しい条件を付けている。我が国も，外国人は日本の研究者と共同研究の場合にのみ利用可能という線が妥当と思われる。

5．法制上の問題点について

現在，政府統計のミクロデータの利用は，調査票の目的外使用の申請をして利用する以外には方法がない。この場合使用を許されるミクロデータは，氏名，住所等の個体識別子は除かれていても，識別の可能性を残したデータの利用であるから，秘密保護の観点から1件毎に厳しい審査を必要とするのはやむを得ない。

そこで個体の識別可能性を除いた匿名標本データの提供の場合には，統計法ではどう扱えばよいかについて考えてみよう。

(1) 統計法の関係条項

問題点の一つは，統計法第14條(秘密の保護)と，第15條の解釈にある。

第15條　何人も指定統計を作成するために集められた調査票を，統計上の目的以外に使用してはならない。

② 前項の規定は，総務庁長官の承認を得て使用の目的を公示したものについては、これを適用しない。

（なお，第 15 條の 2 は届出統計及び報告調整法による統計報告について同様に定めたものであるが，その但し書きに「前項の規定は，届出統計調査又は報告徴集の実施者が，被調査者又は報告を求められたものを識別することができない方法で調査票または統計報告を使用し，又は使用させることを妨げるものではない。」という規定がある。）

この法の趣旨は，① 申告義務を課すのに対応して秘密の保護を義務づけたものであり，その関連で調査票の使用を制限したものであって，② 調査票の内容，すなわち”集められた統計情報”の活用を制限するのが趣旨ではないはずである。目的外使用の承認の条項は，秘密保護が守られるならば調査票の活用を認めていると解釈できる。

(2) 統計法の調査票の定義と匿名標本データ

問題を匿名標本データの提供に限定した場合には，統計法上ではどう考えればよいのだろうか。

調査票の具体的な定義は統計法の中にはないが，行政管理庁長官「事務処理要領」（昭和 40 年 2 月 26 日）の中で次のように定められている。

「調査票」とは，個々の調査対象ごとにその申告内容が判別できるような形で統計の申告が記載された調査関係書類をいう。したがって，照査表，中間集計表，けん孔カード，磁気テープも「調査票」に該当することがある。

匿名標本データは個体識別不能であるから，上記の定義からみれば，統計法上の調査票には該当しない。それは中間集計材料と考えればよい。

統計法制定当時は，個人が集計・分析できるようになるとは予想されていなかったので，調査票以外の中間集計材料の提供については何も触れてはいなかった。しかし利用の多様化した現在では，匿名標本データの形で提供することは，統計を活用する観点から充分”統計目的に沿う”ものと考えられる。

6．法制的な対応について

それでは匿名標本データの提供には，どのような手続きが必要だろうか。

(1) 統計法の改正は必要か

現行の統計法でも調査票の目的以外の使用が認められているので，調査の実施者は調査票の目的以外の使用申請を行って匿名標本データを作成する。作成したデータセットは調査票ではないから，これを提供することは（もちろん必要な手続きをした上で）現在の統計法でも可能と考えられる。特に上述のように特定の分析目的の利用者に限定して認める場合には，統計法の改正の必要はないと考えられる。

しかしながら将来，一般公開，すなわち不特定多数の利用者に対して販売するするような場合は，誤解をさけるためにも法的に明確にしておくべきだという考えもある。

外国の例をみると，ドイツの規定と米国やカナダの規定ではニュアンスに差がある。ドイツの連邦統計法（1987 年改正）は前述のように「事実上の秘匿があればミクロデータの使用を許可する」という意味の肯定的な規定である。

一方，米国の 1976 年に成立したセンサス法では，商務省長官（及びその職員）は..... 以下のことはしてはならない。という項目の中に，「(2) この法律のもとでいかなる特定の事業所あるいは個人から提供されたデータも識別できる形で公表すること」という規定がある。またカナダの 1971 年に改正された統計法では，「この法律に基づいて得られた情報が，個票から得たいかなる事項も特定の個人，企業事業所，その他の団体に結びつけることが可能な形で公開(disclose)してはならない」と規定されている [石田 6]。このように米国もカナダも肯定的に認めるのではなく，識別可能でないならば disclose が許されるという，否定の否定という形で公開を認めていると解される。

我が国の場合，もしドイツ流にすれば，第 15 条の 2 の届出統計に関する規定の用語を使って，第 15 条の③ に，

「調査実施者は分析目的のために，調査票の一部について被調査者を識別することができない方法で標本データを作成し(以下匿名標本データという)，提供することができる。」という趣旨の条文を入れることが考えられる。

しかし米国，カナダ流に考えれば，現行の統計法は”個人・事業所を識別可能な”調査票の使用を禁じているのであるから，否定の否定として匿名標本データの提供は許されており，あえて統計法を改正しなくてもよいという解釈もできる。

私自身は，ドイツ流の肯定的な規定はさらなる議論をよぶ可能性もあるので，賛成ではない。

(2) 匿名標本データの提供の手続き

統計法を改正しない場合でも，匿名標本データの提供を明文化しておくことが，国民から疑念をもたれないためにも必要であろう。

法体系上妥当かどうかの検討が必要だが，たとえば統計法施行令第 6 条（調査票の目的以外の使用の承認の告示の規定）の第 3 号に

「調査票の使用目的が，匿名標本データの作成と提供の場合は，総務庁長官が別に定める手続きに従い承認し，公示するものとする。」

という趣旨の条項を挿入し，「匿名標本データ作成手続規程」を命令等で定める。

規程には，

① 匿名標本データを作成する場合には，データ保護委員会（仮称）の承認を得るものとする。

② 調査実施者が匿名標本データを利用させる場合は，利用者に秘密保護についての

誓約書を提出させるものとする。

として，誓約書の中に前述したような項目を含めておく。そのほかに，利用者の範囲（外国人の場合も含めて），利用の期間等も定める。

7．作成と提供の機関について

匿名標本データの作成は，調査実施機関が調査票の統計目的以外の使用申請を行って作成する。その提供は調査実施機関の責任において，直接または第 3 者機関を介して行うことになる。

匿名標本データの作成はかなりの手間を必要とするので，提供は有料で行うのが妥当と考えられるし，諸外国の例をみても有料が普通である。しかし現在の制度では調査実施機関が手数料を受け取ることができないので，匿名標本データの作成と提供を第 3 者機関に委託することも考えられるが，なお検討を要する問題である。

また Data Archives の設置等については今後の課題である。

8．おわりに

上述のような匿名標本データの提供が実現すれば，利用者は誓約書の提出は義務づけられるもののデータを短期間で入手可能となるし，調査実施機関にとっても統計目的外の使用申請を 1 件毎に審査する負担が軽減される。

ただし実現可能なのは個人・世帯の調査の場合であって，企業・事業所関係の調査ではミクロデータの提供は難しく，諸外国でもあまり例をみない。したがって企業・事業所関係の場合は，調査実施機関または第 3 者機関による受託集計，あるい米国のような宣誓職員制度等を検討して，秘密保護の許す限り積極的なデータの活用を計るべきである。

現在，総務庁では「統計行政の新中・長期構想推進協議会」の第 3 検討委員会で標本データの提供について検討が行われ，その中の「標本データの秘匿措置に関する研究会」で専門的・技術的な研究が行われており，諸外国の資料も集められている。

データの提供そのものは行政の問題であるが，利用するのは研究者であり，研究者も提供方法に関心をもち，その推進に貢献することが望まれる。

【参考文献】

(1)竹村彰通「個票データ開示の理論」重点領研究報告 1997 年 3 月

(2)加納　悟「個票データの開示リスク評価に関する研究」重点領研究報告 1998 年 3 月

(3)永山貞則「ミクロデータの提供への戦略」重点領域研究 1997 年度報告
(4)森　博美「各国におけるミクロデータの提供の現状」日本統計研究所 1996 年 10 月
(5)森　博美「イギリスにおけるミクロデータの提供と利用」統計 1998 年 8 月
(6)濱砂敬郎「ドイツ連邦統計局によるミクロデータの提供」同上
(7)石田　晃「アメリカ・カナダにおけるミクロデータの現状について」敬愛大学研究論集，第 52 号，1997 年 6 月
(8)石田　晃「オーストラリアにおける統計ミクロデータ提供の現状」統計 1998 年 8 月
(9)北田祐幸「我が国のミクロデータ利用の現状と課題」同上
(10)松井　博「官庁統計ミクロデータの利用の経験と今後の課題」重点領域研究 1997 年度報告
(11)W. ミューラー他（濱砂訳)「ミクロデータの事実上の匿名性」日本統計研究所 1997 年 5 月
(12)"Report on Statistical Disclosure Limitation Methodology", Office of Management and Budget, USA. Apr. 1994
(13)"Maintaining the Confidentiality of Data", Office for National Statistics, UK, Apr. 1996