

政府統計データにおける匿名化について

——パーソナルデータの利活用における基盤整備との関連を中心に——

伊 藤 伸 介

近年、パーソナルデータの利活用と保護に関する議論が社会的な関心を集めていることから、本稿では、パーソナルデータの利活用における基盤整備との関連で、主として、わが国の政府統計（公的統計）を対象に、政府統計データの匿名化に関する方法的な特徴を明らかにするだけでなく、パーソナルデータと統計データの関係について考察を試みた。政府統計データについては、様々な法制度的及び技術的な措置をとることによって、政府統計の匿名化マイクロデータの作成が可能になっているが、統計データの特性、利用目的や対象となる利用者によって様々な匿名化措置が適用されることを勘案すると、政府統計データにおいても、合理的な匿名化の水準を明示するのは困難なことが明らかになった。このことは、パーソナルデータの第三者提供を可能にするために、パーソナルデータにおける特定性の低減化を議論する上でも有用であることから、パーソナルデータの利活用においては、政府統計データの匿名化に関する事例は参考になりうると思われる。

1. はじめに

本稿は、パーソナルデータの利活用における基盤整備との関連で、政府統計データの匿名化に関する特徴を明らかにすることを課題としている¹⁾。近年、ビックデータやパーソナルデータに対する社会的な関心が高まっている。企業は、企業自身が保有するビックデータやパーソナルデータを対象に、営利目的を指向したビックデータやパーソナルデータの利用可

1) 本稿は、内閣官房高度情報通信ネットワーク社会推進戦略本部（IT 総合戦略本部）「パーソナルデータに関する検討会」技術検討ワーキンググループの第2回会合（2013年10月17日）と第6回会合（2014年5月13日）における筆者の報告、中央大学経済研究所情報環境研究会（2014年3月11日、於中央大学）での研究報告、及び経済統計学会関東支部定例研究会（2014年6月7日、於立教大学）における研究報告の内容に基づいて大幅に加筆・修正したものである。本稿のもとになった報告内容の一部については、榎田直木氏（当時（独）統計センター統計作成支援課長、現農林水産省大臣官房統計部統計企画管理官）と赤谷俊彦氏（当時（独）統計センター経営企画室課長代理、現総務省行政管理局副管理官（業務・システム改革統括））から貴重なコメントをいただいた。記して謝意を表したい。なお、本稿における誤りは筆者のみにある。

能性を追究するだけでなく、ビックデータやパーソナルデータの販売を目指している。こうした中で、ビックデータに含まれる個人情報をどのように保護するかについての議論が注目されている。

個人情報の保護に関する法律（平成15年5月30日法律第57号）（以下、「個人情報保護法」と呼称）第2条第1項によれば、「個人情報」とは、「生存する個人に関する情報であって、当該情報に含まれる氏名、生年月日その他の記述等により特定の個人を識別することができるもの（他の情報と容易に照合することができ、それにより特定の個人を識別することができることとなるものを含む）」と規定されている。個人情報保護法の対象になる「個人情報」に当たらない情報（「非個人情報」）であれば、営利目的での利用や販売が容易になる。そこで、非個人情報としてのパーソナルデータを作成するための個人情報の匿名化の可能性を検討することが、パーソナルデータにおける利活用の重要な論点となる。

パーソナルデータの利活用については、総務省、経済産業省でそれぞれ研究会が設置され、議論が重ねられてきた。総務省では、「パーソナルデータの利用・流通に関する研究会」（2012年11月～2013年6月）が開催された。一方、経済産業省は、「個人情報保護の制度・運用に関する研究会」（2010年11月～2011年3月）、さらには「IT融合フォーラム パーソナルデータワーキンググループ」（2012年11月～2013年4月）を開催することによって、パーソナルデータの利活用を指向した匿名化の方法について議論を行ってきた。さらに、2013年6月には「規制改革実施計画」が閣議決定され、規制改革会議で「ビッグデータ・ビジネスの普及（匿名化情報の取扱い）」が設置された。それを受けた形で、内閣官房と消費者庁が所管官庁となり、合理的な匿名化措置の内容を明確化したガイドラインを策定する方向性が出された。

他方、パーソナルデータに関する利活用と保護に関する新たな制度設計を目指して、2013年9月に内閣官房高度情報通信ネットワーク社会推進戦略本部（IT総合戦略本部）「パーソナルデータに関する検討会」が開催される（座長 堀部政男（一橋大学名誉教授）、全5回開催（2013年9月～12月））。そして、パーソナルデータの利活用についての技術的側面に関してさらなる検討を進めるために、本検討会に「技術検討ワーキンググループ（以下「技術検討WG」と略称、主査 佐藤一郎（国立情報学研究所教授）」が設置され、2013年に技術検討WGが4回開催された（2013年9月～12月）。

わが国では、政府統計のマイクロデータが作成・提供されているだけでなく、政府統計データにおける匿名化措置に関する研究蓄積が少なくない。そのため、パーソナルデータの利活用と保護をめぐる議論においては、統計に関する研究分野からの貢献も期待された。それは、個人情報を「匿名化」あるいは「統計化」することによって、非個人情報にすることは可能かという論点について、主に政府統計を対象に、統計データの匿名化や統計表の作成と

いう観点に立って議論することを意味している²⁾。

こうした背景を踏まえ、本稿では、パーソナルデータにおける利活用と保護の両面を踏まえた形で、政府統計データの匿名化に関する方法的な特徴を洞察するだけでなく、パーソナルデータと統計データの関係についても考察を試みることにしたい³⁾。

2. わが国における政府統計マイクロデータの匿名化に関する法制度的措置

欧米諸国では、1960年代以降、様々な政府統計マイクロデータの提供が広範にわたって行われてきた。マイクロデータの匿名化に関する研究も、1970年代に遡ることができ(伊藤・村田(2013))、人口センサスや労働力調査等、主として世帯・人口系のデータを対象に、政府統計の個票データに匿名化措置を施した匿名化マイクロデータが作成・提供されてきた。一方、わが国においても、統計法(平成19年法律第53号)の全面施行に伴い、平成21年4月より政府統計の匿名データの提供が開始され、現在、国勢調査を含む7調査に関する匿名データが作成・提供されている。

こうした政府統計の匿名化マイクロデータの作成・提供に関する動向、さらにはマイクロデータに対する匿名化措置に関する数多くの研究蓄積は、パーソナルデータの匿名化における議論においても参考情報になりうると考える。その理由は、政府統計の匿名化マイクロデータ

2) 例えば、経済産業省「個人情報保護に関する法律についての経済産業分野を対象とするガイドライン(平成26年12月12日厚生労働省・経済産業省告示第4号)」においては、「個人情報」に当たらない事例の1つとして、「特定の個人を識別することができない統計情報」が記載されている。この場合、特定の個人の識別不可能性と「統計情報」との関連で、統計データにおける匿名化や統計表の作成方法を議論することも考えられよう。

3) パーソナルデータの利活用と保護に関連して、パーソナルデータをめぐる様々な議論が展開されてきた。これらの議論については、主として、わが国におけるパーソナルデータの保護法制に関する現状と課題、諸外国におけるパーソナルデータに関する法制度的動向、パーソナルデータにおける匿名化の技術と匿名化されたパーソナルデータの利用可能性、パーソナルデータに関する検討会(及び技術検討WG)における議論の内容とその展開方向に大別することができる。このようなパーソナルデータの利活用と保護に関する議論については、学会誌等で特集が組まれることも少なくない。例えば、「ビックデータの利活用に向けた法的課題—パーソナルデータ保護法制の展望」(『ジュリスト』No.1464)、「パーソナルデータの利活用における技術および各国法制度的動向」(情報処理学会『情報処理』Vol.55, No.12)、「小特集 個人データの活用とプライバシー保護」(『電子情報通信学会誌』Vol.98, No.3)等を参照。これらの特集号の中で、パーソナルデータに関する検討会技術検討WGにおける議論の内容については、佐藤(2015)、森(2014a)、森(2014b)、高橋(2014)、高橋他(2015)を参照されたい。なお、本稿は、技術検討WGにおける議論を踏まえながら、パーソナルデータの利活用における基盤整備との関連で、わが国の政府統計データの匿名化に関する法制度的措置及び技術的手法の方法的特徴を明らかにするだけでなく、政府統計データの匿名化措置におけるパーソナルデータへの適用可能性を模索するものである。

が、統計法制度的な措置及び技術的な方法に基づいて作成・提供されているが、パーソナルデータの匿名化においても、匿名化の法制度的な側面と技術的な側面の両方から具体的に議論する必要があるからである。本節では、わが国の統計法制度の下におけるマイクロデータの匿名化措置について概括することにした。

わが国の政府統計における「調査票情報（個票データ）」と「匿名データ」は、統計法第2条第11項と第12項で定義されている。すなわち「統計調査によって集められた情報のうち、文書、図画又は電磁的記録によって記録されているもの」が、政府統計の「調査票情報」に該当する。一方、政府統計の「匿名データ」も、「統計法上では一般の利用に供することを目的として、調査票情報を特定の個人又は法人その他の団体の識別ができないように加工したもの」と定義されている。

政府統計の調査票情報や匿名データは個人情報を含んでいることから、調査票情報や匿名データの提供及び管理を行うことが統計法制度上で定められている。統計法におけるマイクロデータの提供に関する条文は、統計の作成ないしは統計的研究のための調査票情報の提供（第33条）、匿名データの作成（第35条）、及び匿名データの提供（第36条）である⁴⁾。基幹統計調査の場合、最初に、匿名データの作成について統計委員会に諮問が行われる。つぎに、統計委員会の匿名データ部会で基幹統計調査の匿名データに関する匿名性の検討を行った上で、匿名データ部会の審議内容が統計委員会に報告される。そして、匿名データの作成方法の妥当性に関して、統計委員会から答申が出されることになっている。その一方で、調査票情報や匿名データに含まれる個人情報が漏洩しないように、統計法には、調査票情報や

4) (調査票情報の提供)

第三十三条 行政機関の長又は届出独立行政法人等は、次の各号に掲げる者が当該各号に定める行為を行う場合には、その行った統計調査に係る調査票情報を、これらの者に提供することができる。

- 一 行政機関等その他これに準ずる者として総務省令で定める者 統計の作成等又は統計を作成するための調査に係る名簿の作成
- 二 前号に掲げる者が行う統計の作成等と同等の公益性を有する統計の作成等として総務省令で定めるものを行う者 当該総務省令で定める統計の作成等
(匿名データの作成)

第三十五条 行政機関の長又は届出独立行政法人等は、その行った統計調査に係る調査票情報を加工して、匿名データを作成することができる。

- 2 行政機関の長は、前項の規定により基幹統計調査に係る匿名データを作成しようとするときは、あらかじめ、統計委員会の意見を聴かなければならない。

(匿名データの提供)

第三十六条 行政機関の長又は届出独立行政法人等は、学術研究の発展に資すると認める場合その他の総務省令で定める場合には、総務省令で定めるところにより、一般からの求めに応じ、前条第一項の規定により作成した匿名データを提供することができる。

匿名データの適正管理（第42条）、守秘義務（第43条）、さらには罰則規定（第61条）に関する条項が明記されている。

他方、わが国の統計法制度においては、匿名データの作成と提供に関する詳細な手続き上のルールを定めるために、「匿名データの作成・提供に係るガイドライン（総務省政策統括官（統計基準担当）決定、以下『ガイドライン』と呼称）」が制定されている。『ガイドライン』では、主として学術研究目的や教育目的のための匿名データの利用者の範囲⁵⁾を規定しているだけでなく、匿名データの利用申出の際に、匿名データの利用期間、利用場所・保管場所を明示することの必要性を明記している。なお、調査票情報についても、調査票情報の適正な管理方法や取り扱いの仕方に関する注意事項をまとめた「調査票情報等の管理及び情報漏えい等の対策に関するガイドライン（総務省政策統括官（統計基準担当）決定）」が定められている。

ところで、政府統計マイクロデータに関する秘匿性（data confidentiality）については、マイクロデータに含まれる個体識別リスク（identification risk）の定量的な評価を議論することが少なくない（Müller *et al.* (1995)；伊藤 (2010)；伊藤・星野 (2014) 等）。具体的には、侵入者（intruder）を設定した場合、特定の個体に関して侵入者が持っている外部情報（事前情報（a priori knowledge））とマイクロデータファイルに含まれるレコードがキー変数（key variable）を通じて1対1に対応付けられ、その対応付けられたレコードが特定の個体に関する情報であることが確認できた場合に、個体識別が成立するとみなされる。したがって、「基本的には、調査単位とマイクロデータの対応関係を特定されないようにする」ために、政府統計マイクロデータのデータ特性に応じて、リコーディングやトップコーディング等の各種の匿名化技法が適用されるだけでなく⁶⁾、主として外観識別性の高い属性を対象に、様々な匿名化技法を組み合わせることによって、匿名データの作成が行われる⁷⁾。

3. パーソナルデータに関する検討会 技術検討 WG における匿名化の議論の概要

本章では、パーソナルデータに関する検討会技術検討 WG で取りまとめた『技術検討ワーキンググループ報告書』（2013年12月刊行、以下『2013年報告書』と呼称）に基づきなが

5) 匿名データの提供における申出者と利用者の範囲はつぎのとおりである（『ガイドライン』、12ページ）。

- ・自ら責任を持って学術研究の発展に資すると認められる利用を行い得る者
- ・高等教育の発展に資すると認められる利用を行い得る者
- ・国際比較統計利活用事業を行い得る者

6) 『ガイドライン』、45ページ。

7) 『ガイドライン』、47ページ。

ら、パーソナルデータにおける匿名化における議論の概要を述べる。

本検討会における課題は以下の2つに要約される。第1の課題は、現行の個人情報保護法の解釈論として導入可能な「再識別不可能データ」化（提供事業者において容易照合性のない技術的匿名化）措置の内容に関する検討である。第2は、新たな立法措置を前提とした「合理的な技術的匿名化措置」の内容についての検討である⁸⁾。

ところで、「匿名化（anonymization, disclosure control (limitation)）」は、政府統計の分野では用語として定着していると思われるが、パーソナルデータにおいて「匿名化」という用語を用いる場合、分野によって異なるイメージを持っていることが考えられる。そこで、技術検討WGにおいては、「匿名化」という用語を使用せず、「識別」と「特定」という用語を用いて、「識別特定情報」、「識別非特定情報」と「非識別非特定情報」という3つの情報を区別している。「識別特定情報」とは、個人が（識別されかつ）特定される状態の情報であり、個人情報にそれが該当する。識別特定情報に対して、名称や住所等の直接的な識別子の削除（「非特定化」）を行うことによって、「一人ひとり識別されるが、個人が特定されない状態の情報」に変換することが可能である。このような「非特定化」を行った情報は「識別非特定情報」と呼称される。さらに、「識別非特定情報」にk匿名化（同じ属性値の組み合わせを持つレコードがk個以上存在すること）といった処理（「非識別化」）を行うことによって、「一人ひとり識別されない（かつ個人が特定されない）状態の情報」への変換がなされる。それは、「非識別非特定情報」と呼ばれる⁹⁾。図3-1は、匿名化に関する本WGで定義した用語の関係性を示したものである。

パーソナルデータに関する検討会の課題である「再識別不可能データ」化のための措置、及び「合理的な技術的匿名化措置」については、技術検討WGの議論においてつぎの3つの結論が得られている¹⁰⁾。第1に、いかなる個人情報に対しても、識別非特定情報や非識別非特定情報となるように加工できる汎用的な匿名化の技術は存在しないことである。第2に、特定の情報に対して識別非特定情報または非識別非特定情報に加工するための技法の適用可能性は、パーソナルデータの特性に応じて判断されるべきだということである。そして、第3に、「(仮称)法第23条1項適用除外情報」¹¹⁾というカテゴリーが設定されるのであ

8) 『2013年報告書』、1ページ。

9) 前掲書、10-11ページ。

10) 前掲書、20-21、23ページ。

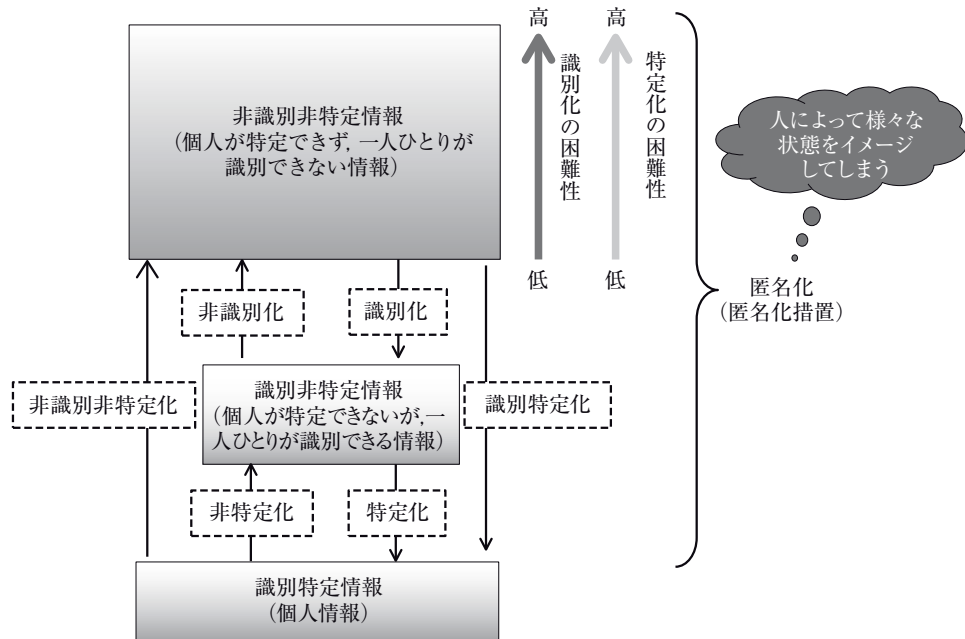
11) 個人情報保護法における「法第23条1項」とは以下の条文である。

(第三者提供の制限)

第二十三条 個人情報取扱事業者は、次に掲げる場合を除くほか、あらかじめ本人の同意を得ないで、個人データを第三者に提供してはならない。

一 法令に基づく場合

図 3-1 匿名化に関する本 WG で定義した用語の関係性



(出所) 『2013年報告書』, 12ページ。

れば、情報の受領者（第三者）への提供において、第三者による特定化・識別化が禁止されることを要件として、個人特定性が低減された個人情報の第三者提供が容認されることである。

技術検討 WG においては、(1)乗降履歴、(2)購買履歴や視聴履歴、(3)統計データ、(4)画像、映像、音声、文章などのデータとインターネットのそれぞれについて、匿名化措置に関する技術的な考察が行われた。その中で、統計データにおける匿名化に関しては、『2013年報告書』においてつぎの3点が指摘されている¹²⁾。

- (1) わが国の統計制度においては、匿名データは、統計法と「匿名データの作成・提供に係るガイドライン」に基づいて作成されており、基幹統計調査の匿名データに関しては、統計委員会への諮問が行われた後に、匿名データの作成方法の妥当性に関する統計委員会からの答申に基づいて匿名データの作成・提供が可能になること。
- (2) 政府統計の匿名データにおいては、様々な法制度的な措置及び技術的な手法を適用す

(以下省略)。

12) 前掲書, 17-18ページ。

ることによって、調査客体が特定される可能性が十分に低く、匿名性が十分に確保されていると判断された場合に、政府統計の匿名データの作成・提供が可能になること。

- (3) 匿名データの作成において適用される技術的な手法の概要は、統計委員会への諮問において公開されているが、統計委員会の匿名データ部会における審議資料や議事内容の詳細に関しては、調査客体の特定化のリスクを減らすために、非公開になっている場合があること。

これらの点を踏まえ、『2013年報告書』は、統計データにおいても、統計制度の下での匿名データの作成・提供の観点から考察した場合、合理的な匿名化の水準を汎用的に示すことは困難であることを結論付けている。

技術検討WGによる『2013年報告書』が刊行された後に、2013年12月に「パーソナルデータの利活用に関する制度見直し方針」（以下「見直し方針」と略称）が出された。「見直し方針」では、パーソナルデータの利活用に関する制度見直し事項として、(1) パーソナルデータにおける利活用とプライバシー保護の両面に立った第三者機関（プライバシー・コミッション）の体制整備、(2) パーソナルデータを加工して個人が特定される可能性を低減したデータの個人情報及びプライバシー保護への影響に留意した取扱いの必要性の2点を指摘している。このことは、パーソナルデータの利活用と保護の両面に立って匿名化措置の適用可能性を追究する必要性を示唆している。さらに、「見直し方針」では、2014年6月まで個人情報保護法の法改正の大綱を取りまとめた上で、2015年1月に改正案を国会に提出するスケジュールが提示されている。

4. パーソナルデータの匿名化に向けたさらなる展開

パーソナルデータに関する検討会は、2014年3月から再開され（2014年1月より座長 宇賀克也（東京大学教授））、「見直し方針」を踏まえた形で、保護されるパーソナルデータの範囲の明確化とパーソナルデータを加工することによって個人が特定される可能性を低減したデータの取り扱い、さらには、第三者機関の設置を中心に議論を行ってきた。本検討会で新たに出された検討課題は、①「(仮称) 準個人情報」の定義や考え方の整理、②「(仮称) 個人特定性低減データ」の定義と「個人情報」や「(仮称) 準個人情報」から加工する場合の最低限かつ具体的な加工方法であって、これらの課題は技術的な議論を要することから、技術検討WGにおいてさらに検討を進めることとなった（2014年4月～5月にかけて2回公式会合が開催）。これらの検討の成果として、2014年5月に『技術検討ワーキンググループ報告書～「(仮称) 準個人情報」及び「(仮称) 個人特定性低減データ」に関する技術的観点からの考察について～』（以下『2014年報告書』と呼称）が刊行された。本節では、『2014

年報告書』を踏まえながら、技術検討 WG の公式会合において筆者が行った報告をもとに、パーソナルデータにおける匿名化に向けた議論のさらなる展開について論じることにしたい。

技術検討 WG での検討課題の 1 つである「(仮称) 準個人情報」とは、「特定の個人が識別されていない情報であっても、特定の個人が識別されるおそれのある情報」と定義される。図 4-1 に示されるように、氏名、住所、顔画像といった「個人情報」としての明示的な位置付けが可能な情報とは別に、携帯端末 ID、顔認識データ、位置データ、メールアドレスといった特定の個人を識別するのではないが、その取扱いによっては本人に権利侵害をもたらす可能性が高い情報が「(仮称) 準個人情報」と位置付けられている。

当初、事務局(内閣官房 IT 総合戦略本部)による案では、つぎの 3 つのタイプの情報が「(仮称) 準個人情報」の対象となっていた。すなわち、

「個人情報に該当するものを除き、生存する個人に関する情報であって、次に例示するもの及びこれに類するものを含む情報

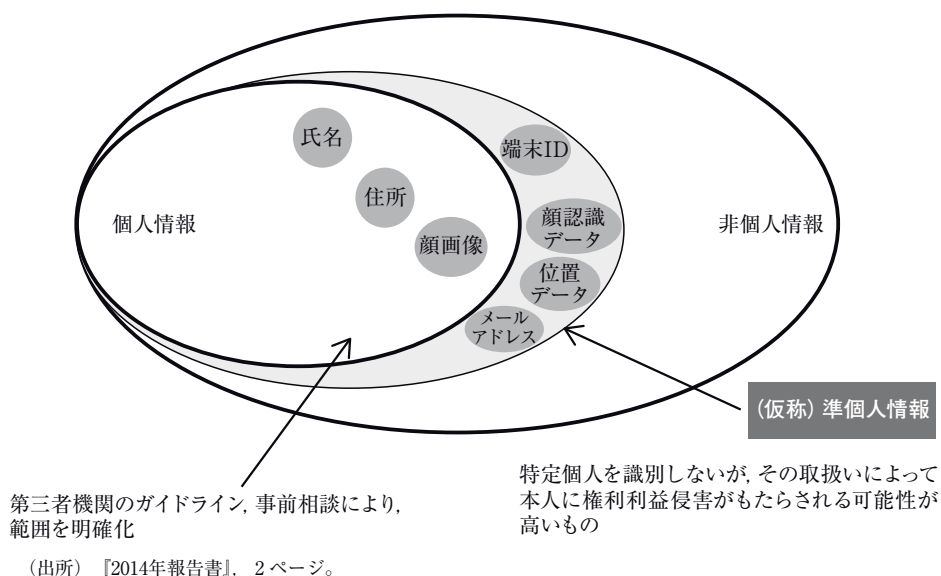
- ① パスポート番号、免許証番号、IP アドレス、携帯端末 ID 等の個人又は個人の情報通信端末(携帯電話端末、PC 端末等)等に付番され、継続して共用されるもの
- ② 顔認識データ、遺伝子情報、声紋並びに指紋等、個人の生体的・身体的特性に関する情報で、普遍性を有するもの
- ③ 移動履歴、購買履歴等の特徴的な行動の履歴¹³⁾

一方、「(仮称) 準個人情報」に匿名化措置を施すことによって個人の特定性が低減されたデータ(「個人特定性低減データ」)の作成を行う上では、上記の事務局案で示された「(仮称) 準個人情報」のタイプごとに、個人特定性低減データに加工するための技法を検討する必要がある。具体的には、上記の 3 タイプの「(仮称) 準個人情報」のそれぞれに関して、次の 3 点を指摘することができる。

第 1 に、「個人又は個人の情報通信端末等に付番され、継続して共用される情報」については、本人確認のために使われる情報か否かが、その情報の特定性に作用すると考えられる。これらの「個人又は個人の情報通信端末等にかかわる」個体識別情報(名前、住所等の直接的な識別子)を任意の一連番号に置き換えたデータとしても、変数の組み合わせによって一意となる場合が生じる。

13) 『2014年報告書』, 9-10ページ。

図 4-1 「準個人情報」の位置付け



第2に、「個人の生体的・身体的特性に関する情報」は、「顔認識データ」のように「それだけで」一意に特定することが可能な情報と、人種のように（例えば、地域を狭めることによって）条件付きで一意の特定が可能な情報に類別可能だと考えられる。このような「個人の身体的特性にかかわる」情報に関しても詳細な地域情報が含まれる場合を勘案すると、変数の削除だけでは十分でない場合があると思われる。さらに、属性によっては、直接的な識別子の削除を含む変数の削除のような手法を用いた場合には、情報量損失はむしろ大きくなる。こうしたタイプの「(仮称) 準個人情報」を個人特定性低減データに加工するためには、直接的な識別子の削除を含む変数の削除だけでなく、属性情報に応じてさらなる加工が必要と思われる。

第3に「特徴的な行動の履歴」については、（例えば）移動履歴に限らず、一意に該当しかつ特異なレコードも「特徴的」なレコードと考えられる。「特徴的な行動の履歴」については、「特殊な一意 (special uniques)」に該当するレコードがそれに該当する可能性がある¹⁴⁾。そのため、特殊な一意に該当するレコードの探索及び特殊な一意に対する加工方法の

14) 特殊な一意とは、「K 個のキー変数の集合において標本一意であるだけでなく、K の部分集合である k 個 (のキー変数の集合) においても標本一意となること」によって、「疫学的に特異であるために、本質的に (intrinsically) まれな属性群の組み合わせを有する」レコードは「特殊な一意」だと考えられる (Elliot (2001))。このことから、少数のキー変数の組み合わせにおいても、一意になるレコードは、特殊な一意に該当するとみなされる (伊藤・星野 (2014), 3 ページ)。

検討が求められる。さらに、「特徴的な行動の履歴」には、位置データと日時を含む移動履歴のような情報も含まれるが、こうした情報に対しては、時点(時間)情報や地点(地域)情報を粗くすることによって、特定性低減の程度を高める必要がある。

一方、個人特定性低減データを作成するための加工方法については、政府統計マイクロデータに適用される技術的な匿名化手法は参考になると考えられる。政府統計マイクロデータに用いられる匿名化技法は、非攪乱的な手法(non-perturbative method)と攪乱的な手法(perturbative method)に大別される。非攪乱的な手法については、例えば、レコードの一部抽出(サンプリング、リサンプリング)、再符号化(リコーディング)、トップ(ボトム)・コーディング、レコード削除が該当し、攪乱的な手法については、スワッピング、ノイズの付加、マイクロアグリゲーション等の手法が含まれる¹⁵⁾。

わが国の政府統計の匿名データについては、技術的な手法としては、最低限必須の加工方法である直接的な識別子の削除だけでなく、レコードの一部抽出(サンプリング、リサンプリング)、再符号化、レコード削除といった非攪乱的な手法が基本的な匿名化技法として用いられてきた。これらの手法は、必須ではないと思われるが優先度は高い。そして、国勢調査の匿名データにおいては、上記の非攪乱的な手法だけでなく、攪乱的な手法としてスワッピングが適用された。

わが国の政府統計の匿名データの作成において、レコードの一部抽出(サンプリング、リサンプリング)は、有力な匿名化技法の1つとして特に用いられてきた。「個人情報」及び「(仮称)準個人情報」から個人特定性低減データへ加工するための方法の1つとして、レコードの一部抽出(サンプリング)を行うことによって、個人特定性の低減が可能になると思われる。さらに、レコードの一部抽出を行う場合、抽出率や抽出方法を公開しないことも有効な措置になりうる。

他方、匿名化技法が適用されていない政府統計の個票データにおいては、わが国でもインターネット回線に接続されないオンサイト施設で個票データに直接アクセスするシステムや、大学の研究室からでもリモートアクセスによってログインし、個票データにアクセスすることが可能なシステムが議論されてきた。したがって、パーソナルデータの利活用においても、特定性の困難性の程度が低い個人特定性低減データについては、データアクセス・提供のシステムに関する議論も検討に値すると考える。

技術検討WGにおいて様々な議論が行われた結果、『2014年報告書』において、「(仮称)準個人情報」と「(仮称)個人特定性低減データ」は、つぎのように定義されることとなった。「(仮称)準個人情報」は、(1)個人情報ではないこと、(2)生存する個人に関する情報に

15) 伊藤・星野(2014)、1ページ。

含まれる識別子又は識別子に該当するものであって、①一意性／単射性を有する（識別子と個人が1対1の対応関係にある）もの、②共用性を有する（同一の識別子が複数の事業者で保有されるかあるいは複数のサービスで共通に利用される）もの、③変更あるいは利用停止が容易ではないもの、又は不変性を有する（時間や状態によって動的に変化しない）ものであること、(3)(1)と(2)のすべての条件を満たす識別子が識別子に該当するものであって、密接性を有するものとして、①個人又は個人が使用する通信端末機器等に関するもの、②個人の身体的特性に関するもの、③上記の2項目のほか、特定の個人の識別につながる多量又は多様な情報の収集を可能にするものと定義されている¹⁶⁾。一方、「(仮称)個人特定性低減データ」は、「個人情報」に対して「特定の個人を識別させる情報に加工を施して、特定の個人を識別することができないようにしたもの」であり、「(仮称)準個人情報」に対しても「当該データに含まれる識別子（又は識別子に相当する機能を有するもの）に加工を施して特定の対象を識別することができないようにしたもの」と明記されている¹⁷⁾。

さらに、『2014年報告書』は、「(仮称)個人特定性低減データ」の作成については、「すべてのデータに対して有効となるような具体的な加工方法を提示することは困難で」あって、「最低限の加工方法を定義することはできないことから、特定の個人の識別性を低減させることと利活用のニーズとのバランスを考慮し、事業者自らの判断と責任において、適切な加工を施すことが必要である」と結論付けている¹⁸⁾。

5. パーソナルデータと統計表

パーソナルデータの利用・販売において、「統計（データ）」という形に加工して、パーソナルデータの利用促進を図る可能性が考えられる。『2013年報告書』では、統計データにおける匿名化を議論するために、統計に関する定義を行っているが、それによれば、統計は、「一定の条件（時間・空間・標識・単位）で定められた集団について調べた（あるいは集めた）結果を集計・加工して得られた数値」とされている¹⁹⁾。このような「集計・加工して得られた数値」である統計データは、「統計表」の形で一般に認識されているが、パーソナルデータの議論において、統計表はどのように整理できるであろうか。本節では、パーソナルデータと統計表との関係について考察することにした。

竹内啓編（1989）『統計学辞典』によれば、統計表は、「いくつかの変数値区分の組み合わせの各セルごとに、対応する観測単位数、あるいは観測単位のもつ値の平均値や比率などの

16) 『2014年報告書』、15ページ。

17) 前掲書、25ページ。

18) 前掲書、24-26ページ。

19) 『2013年報告書』、18ページ。

統計量を表示した、表形式の集計データ」²⁰⁾と定義されている。一方、わが国の統計法では、統計ないしは統計表の定義は明文化されていないが、公表される統計表（あるいは集計結果表）は、一般的に基本的と考えられる調査事項をクロスさせた集計表であり、低次元の集計表として調査票情報から作成されると考えられている²¹⁾。

こうした統計表に関する定義に基づくのであれば、統計制度の下で公的統計の調査票情報から作成された集計結果表は、調査票情報そのものとは異なると考えることができる。さらに、匿名データは、統計法36条によれば、「調査票情報を加工したもの」と規定されており、わが国の統計法制度の下では、匿名データは調査票情報の一種と捉えられる。このことから、集計結果表は作成方法の点で匿名データとは基本的に異なるということができよう。

一方、データ構造の観点から見れば、集計結果表と匿名化マイクロデータが対応関係を持つ場合がある。図5-1は、個票データとマイクロアグリゲートデータ（個票データにマイクロアグリゲーションを適用することによって作成されたデータ）との関係を示したものである。図5-1では、例として、質的属性として性別、雇用形態と週間就業時間を、量的属性として年間収入をそれぞれ含む個票データが設定されている。性別、雇用形態と週間就業時間を集計事項としたクロス集計表を作成することが可能であるが、このクロス集計表におけるセルの度数と、性別、雇用形態と週間就業時間の質的属性について同じ属性値を持つレコード群（以下「同質属性値レコード群」と呼称）内のレコード数は一致する。つぎに、同質属性値レコード群内における年間収入の合計を同じレコード群内におけるレコード総数で割ることによって質的属性における分類区分の組合せ（図5-1の例では、性別が1：男、雇用形態が1：正規の職員・従業員、週間就業時間が4：60時間以上）、及び量的属性に関する平均値（図5-1では年間収入が1800万円）から構成される数値群が得られる。各レコードが持つ属性値群は、あくまで集計値として位置付けられているが、これは、質的属性値群と量的属性の平均値群から構成された個票データに準じたデータとみなすことができる。これらのレコードのおのおのについて該当する総数だけレコードを「複製」することによって、マイクロアグリゲートデータが編成される²²⁾。

図5-1では、4変数のみを含む個票データを用いて議論しているが、政府統計の個票データの場合は、政府統計の個票データが持つすべての属性（調査事項）群を集計事項とした多重クロス集計表を作成し、その集計表からマイクロアグリゲートデータを作成することが理論的には可能である。それは、個票データが有するすべての属性群を集計事項の対象とした上

20) 竹内編（1989），284ページ。

21) 山口他（2013），5-6ページ。

22) 伊藤（2009），211ページ。

図 5-1 個票データとミクロアグリゲートデータとの関係

(1) 個別データ (属性群として性別、雇用形態、時間と年間収入のみが配列されていると想定)

一連番号	性別	雇用形態	週間就業時間	年間収入(千円)
1	1	1	4	1500
2	1	1	2	2300
3	1	1	4	2100
4	1	3	1	1500
5	1	3	2	2700
6	1	3	3	1800
7	2	2	3	3600
8	2	4	4	4000
9	2	2	3	2800

(2) 性別、雇用形態と週間就業時間に関する同質属性レコード群

一連番号	性別	雇用形態	週間就業時間	年間収入(千円)
1	1	1	2	2300
2	1	1	4	1500
3	1	1	4	2100
4	1	3	1	1500
5	1	3	2	2700
6	1	3	3	1800
7	2	2	3	3600
8	2	4	4	2800
9	2	2	4	4000

性別 1:男 2:女
 雇用形態 1:正規の職員・従業員 2:パート・アルバイト 3:派遣・契約社員
 週間就業時間 1:35時間未満 2:35~48時間 3:49~59時間 4:60時間以上

(3) 性別、雇用形態、週間就業時間別クロス集計表

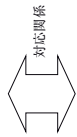
性別	男				女				計
	正規の職員・従業員	パート	アルバイト	派遣社員	正規の職員・従業員	パート	アルバイト	派遣社員	
雇用形態	0	0	1	0	0	0	0	0	0
週間就業時間	1	0	1	0	0	0	0	0	0
35時間未満	0	0	1	0	0	0	0	0	0
35~48時間	1	0	1	0	0	0	0	0	0
49~59時間	0	0	1	0	0	2	0	0	0
60時間以上	2	0	0	0	0	0	0	0	0
計	3	0	3	0	0	2	0	1	9

性別、雇用形態、週間就業時間：総数(N) 年間収入の総計

性別	雇用形態	週間就業時間	総数(N)	年間収入の総計
1	1	2	1	2300
1	1	4	2	3600
1	3	1	1	1500
1	3	2	1	2700
1	3	3	1	1800
2	2	3	2	6400
2	4	4	1	4000

(4) ミクロアグリゲートデータ

ミクロアグリゲーション後の一連番号	性別	雇用形態	週間就業時間	年間収入
1	1	1	1	2300
2	1	1	4	1800
3	1	1	4	1800
4	1	3	1	1500
5	1	3	2	2700
6	1	3	3	1800
7	2	2	3	3200
8	2	2	2	3200
9	2	4	4	4000



(出所) 伊藤 (2009), 210ページ。

で作成される n 次元の多重クロス集計表（以下、「超高次元クロス集計表」と呼称）であって、あらゆる属性の組合せが集計事項として設定可能な集計表である（図5-2）。また、図5-2に見られるように、超高次元クロス集計表は、次元から n 次元までのあらゆる次元のクロス集計表を包含している。さらに、超高次元クロス集計表において設定される集計事項に関する分類区分の統合を行うことによって、新たな集計表の作成も可能である²³⁾。このように超高次元クロス集計表に基づいて作成されるマイクロアグリゲートデータに関しては、統計表に含まれるセルとの対応付けを行うことができる。

このことは、匿名化マイクロデータと統計表においてデータ構造の観点から明示的な線引きを行うのは困難であることを意味している²⁴⁾。

ところで、『2013年報告書』で議論されたように、パーソナルデータにおいては、個人情報としての識別特定情報は、個人の氏名や住所等の直接的な識別子の削除といった非特定化を行うことによって、識別非特定情報に変換される。『2014年報告書』では、識別非特定情報に位置付けられる「(仮称)準個人情報」に対して加工を施すことによって、より識別化が困難な「(仮称)個人特定性低減データ」が作成される（図5-3）。特定化あるいは識別化の困難性の観点からいえば、個人特定性の低減を指向した各種の加工を行うことによって、識別化の困難性の程度が異なる様々な個人特定性低減データの作成が可能になる。

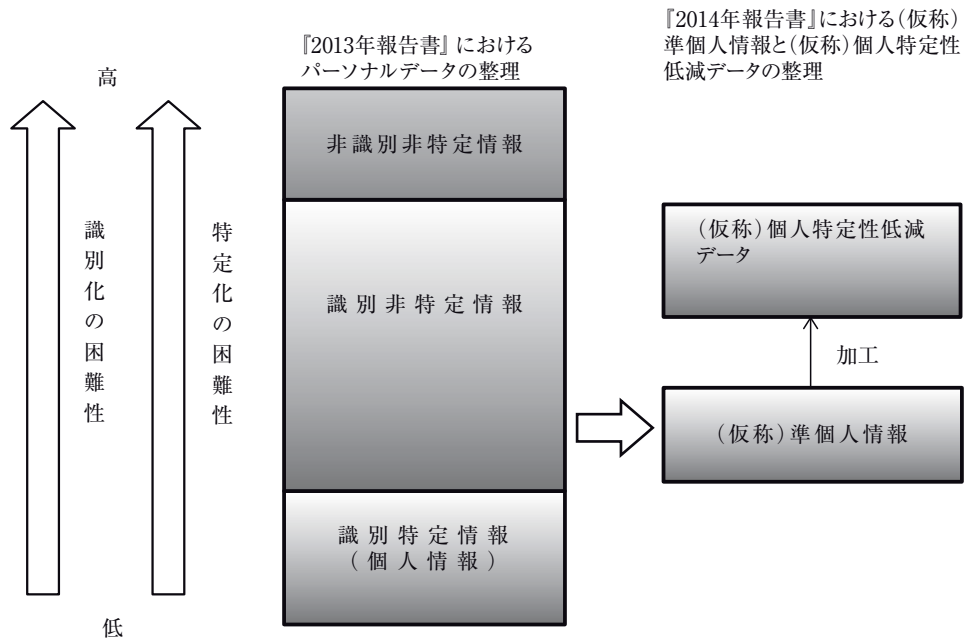
個人特定性低減データと統計データとの関連性については、つぎのように考えることができる。『2013年報告書』によれば、統計データに関しては、「過去の実績・経験等も考慮すると、統計においては個人が特定される（又は個人が識別される）可能性は非常に低くなっているものの、さらに必要に応じて統計に加工が施されることもある」。さらに、「可能性という観点からいえば、他の多種多様なデータと組み合わせられる可能性があるという現状においては、どういったデータとの照合も一切不可能であると理論的に証明できるものではないが、統計における個人の特定化又は識別化は、極めて困難である」²⁵⁾。このことは、統計においては、特定化あるいは識別化が事実上「困難」であるが、「不可能」とは言えないことを意味している。匿名化マイクロデータと統計表においてデータ構造における明確な区別が困難だとみなされるのであれば、統計表は、個人特定性低減データの中で、識別化の困難性の程度が非常に高いデータとして位置付けることが可能である。さらに、統計データにおい

23) 同上、211-212ページ。

24) 超高次元クロス集計表の枠組においては、リコーディングやノイズといったマイクロアグリゲーション以外の匿名化技法を適用した場合でも、匿名化マイクロデータに含まれるレコードの属性値と個票データにあらかじめリコーディングやノイズを適用した上で作成した超高次元クロス集計表のセルの数値は1対1に対応すると考えられる。

25) 『2013年報告書』、19ページ。

図 5-3 技術検討 WG におけるパーソナルデータの整理

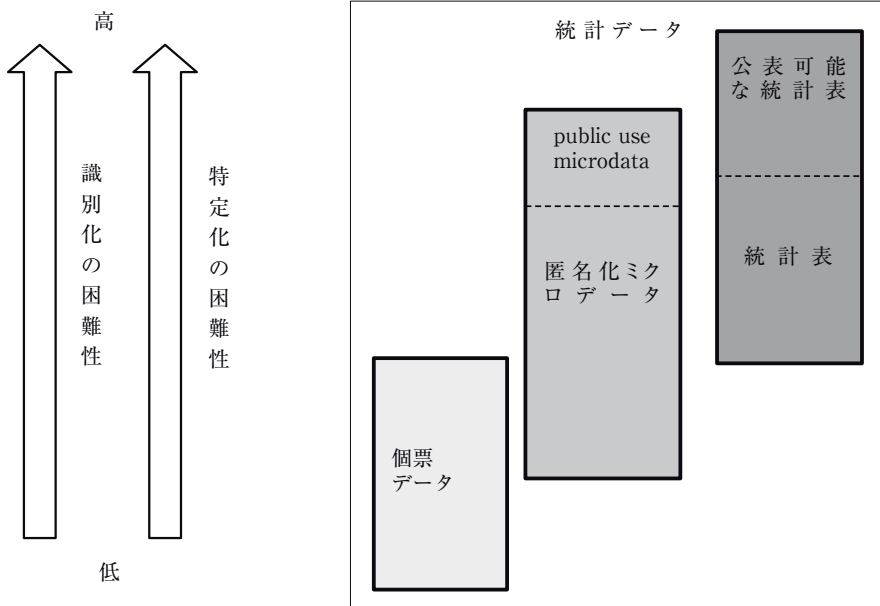


では、集計事項として用いられる属性の数や種類、さらには属性に含まれる分類区分の数によっては、識別化の困難性がより一層高い、非識別非特定化情報に該当する統計表の作成も可能になると考えられる。

こうした識別化・特定化の困難性の視点に立って、統計法制度の下で作成された政府統計の個票データ（調査票情報）、匿名化マイクロデータと統計表との関係を位置付けたのが図 5-4 である。個票データに対して匿名化措置を施すことによって作成される匿名化マイクロデータは、利用目的や対象となる利用者によって、秘匿性の強度が低く、学術研究目的を指向した形で作成・提供される匿名化マイクロデータと、利用目的に制限がなく一般に公開することが可能な public use microdata に大別される。匿名化マイクロデータが public use microdata として提供できるかどうかは、それが公開可能なレベルの秘匿性を確保しているかどうかにかかっている。この「公開可能なレベルの秘匿性」とは、個体情報が露見されるのが事実上困難であり、露見リスクの制御が可能なレベルの秘匿性であって、露見リスクがゼロになることとは異なると考えられる²⁶⁾。一方、統計表も、集計表に含まれるセルに対する秘匿処

26) 例えば、カナダ統計局では、人口センサスの一般公開型マイクロデータ（Public Use Microdata Files）を提供する上で、露見リスク（disclosure risk）をゼロにするという考え方ではなく、露見リスクをコントロールする（disclosure control）という考え方に基づいて、個票データに対する匿名化措置を行っている（赤谷他（2014）、4-5ページ）。

図 5-4 統計データにおける位置付け



理（例えば一次秘匿，二次秘匿等）が施された公表可能な統計表と，公表することが困難であり中間生成物に位置付けられる高次元の統計表に区別される。公表される統計表は，統計表の集計事項として用いられる属性の数だけでなく，属性の種類や分類区分の制約を受ける。さらに，統計表の秘匿性の担保は，セルに含まれる結果数値（度数，平均値，比率，重み付きの度数等）にも依存している。こうした点を踏まえると，パーソナルデータから統計表を作成する場合にも，こうした条件も考慮した上で公表可能性を検討する必要があるように思われる。

ところで，わが国の統計法制度の下で，基幹統計調査の調査票情報から統計表を作成・集計するにあたって，その根拠となる条文は，基幹統計調査の承認に関する統計法第9条，及び総務大臣による基幹統計調査の承認の基準が明記された第10条である²⁷⁾。

27) (基幹統計調査の承認)

第九条 行政機関の長は，基幹統計調査を行おうとするときは，あらかじめ，総務大臣の承認を受けなければならない。

2 前項の承認を受けようとする行政機関の長は，次に掲げる事項を記載した申請書を総務大臣に提出しなければならない。

- 一 調査の名称及び目的
- 二 調査対象の範囲
- 三 報告を求める事項及びその基準となる期日又は期間

基幹統計調査の実施にあたって、統計作成部局は、最初に、統計調査票で捕捉される調査事項のすべてが作成する統計表の集計事項として使用されているかどうかについて、総務省政策統括官室(統計基準担当)からの審査を受けた上で、総務大臣から基幹統計調査の承認を得るために、第9条第2項に関する必要事項が記載された申請書類を総務大臣に提出する。申請書類が提出されると、総務大臣は、統計委員会に対して当該基幹統計調査に関する諮問を行う。統計委員会は、「当該基幹統計の作成の目的に照らして必要かつ十分なものである」という視点から、第9条第2項の「二 調査対象の範囲」から「六 報告を求める期間」について審議を行った後に、当該基幹統計の承認に関して、総務大臣に答申を出すことになっている。その一方で、「七 集計事項」は、総務大臣が当該基幹統計の作成の妥当性を判断するための要件には含まれていない。そのため、わが国においては、統計作成部局が作成する統計表の公表において、統計表における個人情報への露見リスクという観点からの審査は行われていない²⁸⁾。このことから、わが国における現行の統計法の下では、個人情報の露見リスクが十分に低減された、個人情報の特定化が事実上極めて困難な統計表が、統計作成部局の責任の下で公表されている。わが国の統計法制度上では、調査票情報から作成された統計表を公表可能にするための秘匿処理の方法は、統計作成部局における秘匿の考え方、さらには対象となる統計データの特性によって異なる。その意味では、わが国の公表された統計表においても、汎用的な匿名化の水準を示すのは困難だと考えられる。

四 報告を求める者

五 報告を求めるために用いる方法

六 報告を求める期間

七 集計事項

(中略)

- 4 総務大臣は、第一項の承認の申請があったときは、統計委員会の意見を聴かなければならない。ただし、統計委員会が軽微な事項と認めるものについては、この限りでない。

(承認の基準)

第十条 総務大臣は、前条第一項の承認の申請に係る基幹統計調査が次に掲げる要件のすべてに適合していると認めるときは、同項の承認をしなければならない。

- 一 前条第二項第二号から第六号までに掲げる事項が当該基幹統計の作成の目的に照らして必要かつ十分なものであること。

(以下、省略)。

28) アメリカセンサス局の開示評価委員会 (disclosure review board) は、提供の対象としているすべてのマイクロデータと集計表について、露見制御の方法や手続きに関する審査を行っていること (Zayatz (2007), p.255) を勘案すると、わが国の現行の統計法制度の下でも、統計表における公表可能性について審査する仕組みが議論されてもよろしいような気がしている。

6. おわりに

本稿では、パーソナルデータの利活用という観点から、政府統計の匿名化マイクロデータの作成・提供状況も踏まえた形で、政府統計データの匿名化に関する方法的な特徴を洞察した。政府統計データについては、様々な法制度的及び技術的な措置をとることによって、政府統計の匿名化マイクロデータの作成が可能になっているが、統計データの特性、さらには、利用目的や対象となる利用者によって、適用されている匿名化措置が異なることを勘案すると、統計データにおいて合理的な匿名化の水準を明示するのは困難である。このような統計データにおける事例は、パーソナルデータの第三者提供を可能にするために、パーソナルデータにおける個人特定性の低減化を議論する上でも有用だと思われる。したがって、パーソナルデータの利活用と保護においては、こうした政府統計データの匿名化に関する事例を参考にすることも望ましいと考える。

わが国の政府統計の匿名データに関しては、統計委員会の匿名データ部会において匿名データの秘匿性についての審議が行われるが、その審議の内容について統計委員会からの答申が出されると、統計作成部局は匿名データの安全性を確保した上で、匿名データの作成・提供を行っている。その一方で、匿名データの利用者も、統計法制度の下で、匿名データの適正管理の義務を負っている。パーソナルデータにおいても同様に、匿名化措置を施すことによって作成された個人特定性低減データについては、第三者機関による承認²⁹⁾を得ることによって、その提供が可能になるが、パーソナルデータの提供元は、第三者機関による承認を得たとしても、個人特定性低減データという形で提供する際に生じる提供者側の責任を負っていると思われる。他方、個人特定性低減データの提供先（受領者）においては、パーソナルデータの利活用における法制度的な規制というよりも、むしろその安全な利用や節度ある分析について技術的な自主的な考慮や、情報の保管や取扱いなど運用面に関する配慮が求められるだろう。

パーソナルデータに関する検討会 技術検討 WG の議論も踏まえた形で、本検討会において、2014年6月に「パーソナルデータの利活用に関する制度改正大綱（以下『大綱』と略

29) 提供の対象となる個人特定性低減データに関して、第三者機関に届出る項目については、例えば、名称、種類、データが収集された時期（期間）、データに含まれる属性、データに含まれるレコード数等が考えられる。さらに、第三者機関において個人特定性低減データの加工方法を検討する上で、個人特定性低減データ作成のための「最低条件」の適用状況が議論されることもあるだろう。なお、「最低条件」以外の加工方法については、第三者機関へ届け出ることが可能な情報と、届け出するための資料の作成が容易でないといった理由で、第三者機関へ届け出ることが困難な情報があると思われる。

称)」が出された。『大綱』では、本人の同意がなくても第三者へ提供を可能にする「個人の特定性を低減したデータ」について、特定の個人に関する識別の禁止といった取扱いに関する規律や民間団体における自主規制ルールの策定の必要性、さらには、第三者機関の設置による制度の適切な執行の重要性が指摘されている。民間企業が持つパーソナルデータの利活用については、『大綱』に基づいて個人情報保護法の改正をめぐる議論が展開されてきたが、その一方で、総務省「行政機関等が保有するパーソナルデータに関する研究会」が開催され、行政機関や独立行政法人等が有するパーソナルデータに関してもその利用可能性についての議論が行われた。こうした議論の中で個人情報保護法に関する改正作業が進められ、個人情報保護法の改正案が2015年3月に通常国会に提出された(2015年4月執筆時点)³⁰⁾。わが国でもパーソナルデータのさらなる利活用に向けた法制度の整備が、今後より具体的に進展していくであろう。パーソナルデータの利活用と保護における展開方向については、これからも注視していく必要がある。

参考文献

- 赤谷俊彦・荒川智浩・伊藤伸介(2014)「カナダ統計局における政府統計データの提供の動向について」(『ESTRELA』No.241)(財)統計情報研究開発センター、2-9ページ。
- 伊藤伸介(2009)「匿名化技法としてのマイクロアグリゲーションについて」(熊本学園大学『経済論集』第15巻第3・4号合併号)197-232ページ。
- 伊藤伸介(2010)「マイクロデータにおける秘匿性の評価方法に関する一考察」(明海大学『経済学論集』Vol.22, No.2)1-17ページ。
- 伊藤伸介・村田磨理子(2013)「家計調査マイクロデータを用いた攪乱的手法の有効性に関する研究」(『製表技術参考資料』No.22)1-26ページ。
- 伊藤伸介・星野なおみ(2014)「国勢調査マイクロデータを用いたスワッピングの有効性の検証」(『統計学』107号)1-16ページ。
- 佐藤一郎(2015)「パーソナルデータに関わる制度改正方向—パーソナルデータの利活用と保護の両立に向けて—」(『電子情報通信学会誌』Vol.98, No.3)178-187ページ。
- 高橋克己(2014)「個人の移動履歴の保護—プライバシーリスクを明らかにした利活用—」(『情報処理』Vo.55, No.12)1373-1380ページ。
- 高橋克己・正木彰伍・濱田浩気(2015)「個人データの匿名化とその限界」(『電子情報通信学会誌』Vol.98, No.3)193-201ページ。
- 竹内啓編(1989)『統計学辞典』東洋経済新報社。
- 内閣官房高度情報通信ネットワーク社会推進戦略本部(IT戦略本部)パーソナルデータに関する検討会(2013)『技術検討ワーキンググループ報告書』(<http://www.kantei.go.jp/jp/singi/it2/pd/>)

30) 個人情報保護法の改正案では、『大綱』での「個人の特定性を低減したデータ」に該当すると考えられる「匿名加工情報」が条文に規定されているが、「匿名加工情報」の範囲については、今後、具体的な議論が必要になると思われる。

- dai5/siryoun2-1.pdf) (2015年4月17日アクセス)。
- 内閣官房高度情報通信ネットワーク社会推進戦略本部 (IT 戦略本部) パーソナルデータに関する検討会 (2014) 『技術検討ワーキンググループ報告書～「(仮称) 準個人情報」及び「(仮称) 個人特定性低減データ」に関する技術的観点からの考察について～』 (<http://www.kantei.go.jp/jp/singi/it2/pd/dai10/siryoun1-2.pdf>) (2015年4月17日アクセス)。
- 内閣官房高度情報通信ネットワーク社会推進戦略本部 (IT 戦略本部) パーソナルデータに関する検討会 (2014) 『パーソナルデータの利活用に関する制度改正大綱』 (<http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20140624/siryoun5.pdf>) (2015年4月17日アクセス)。
- パーソナルデータに関する検討会の HP (<http://www.kantei.go.jp/jp/singi/it2/pd/>) (2015年4月17日アクセス)。
- 森亮二 (2014a) 「パーソナルデータの匿名化をめぐる議論」(『ジュリスト』 No.1464) 25-31ページ。
- 森亮二 (2014b) 「日本の個人情報保護法改正の状況」(『情報処理』 Vo.55, No.12) 1353-1360ページ。
- 山口幸三・伊藤伸介・秋山裕美 (2013) 「教育用擬似マイクロデータの作成—平成16年全国消費実態調査を例として—」(『統計学』 104号) 1-15ページ。
- Müller, W., Blien, U., Wirth, H. (1995), "Identification Risks of Micro Data: Evidence from Experimental Studies", *Sociological Methods and Research*, Vol. 24, No. 2, pp. 131-157.
- Zayatz, L. (2007), "Disclosure Avoidance Practices and Research at the U.S. Census Bureau: An Update", *Journal of Official Statistics*, Vol. 23, No. 2, pp. 253-265.